

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DERİN ÖĞRENME YÖNTEMLERİ İÇİN DOĞRUSAL
OLMAYAN AKTİVASYON FONKSİYONLARININ
GELİŞTİRİLMESİ**

**Hazırlayan
Serhat KILIÇARSLAN**

**Danışman
Doç. Dr. Mete ÇELİK**

Doktora Tezi

**Eylül 2021
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DERİN ÖĞRENME YÖNTEMLERİ İÇİN DOĞRUSAL
OLMAYAN AKTİVASYON FONKSİYONLARININ
GELİŞTİRİLMESİ
(Doktora Tezi)**

**Hazırlayan
Serhat KILIÇARSLAN**

**Danışman
Doç. Dr. Mete ÇELİK**

**Eylül 2021
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu tez çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Serhat KILIÇARSLAN

İmza:



YÖNERGEYE UYGUNLUK

“Derin Öğrenme Yöntemleri için Doğrusal Olmayan Aktivasyon Fonksiyonlarının Geliştirilmesi” adlı Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi 'ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan

Serhat KILIÇARSLAN

Danışman

Doç. Dr. Mete ÇELİK



Bilgisayar Mühendisliği ABD Başkanı

Prof. Dr. Veysel ASLANTAŞ

İmza

TEŞEKKÜR

Bu tezin hazırlanması sırasında değerli fikirleri ve engin bilgisi ile bana yol göstererek yardım ve desteğini hiçbir zaman esirgemeyen kıymetli hocam Sayın Doç. Dr. Mete ÇELİK'e, tezi izleme komitesinde bulunarak çalışmamıza katkıda bulunan ve yardımlarını her zaman yakından hissettiğim değerli hocalarım Doç. Dr. Serkan ÖZTÜRK'e ve Dr. Öğretim Üyesi Nurettin ÜSTKOYUNCU'ya teşekkür ederim.

Hayatım boyunca verdikleri tüm desteklerden dolayı haklarını ödeyemeyeceğim aileme, destekleriyle her an yanımda olan sevgili eşim Sabire KILIÇARSLAN'a ve hayattaki motivasyon kaynaklarım sevgili oğlum Yusuf Taha KILIÇARSLAN'a ve kızım Elif Nisa KILIÇARSLAN'a teşekkür ederim.

Serhat KILIÇARSLAN

KAYSERİ, Eylül 2021

DERİN ÖĞRENME YÖNTEMLERİ İÇİN DOĞRUSAL OLMAYAN AKTİVASYON FONKSİYONLARININ GELİŞTİRİLMESİ

Serhat KILIÇARSLAN

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü

Doktora Tezi, Eylül 2021

Danışman: Doç. Dr. Mete ÇELİK

ÖZET

Derin öğrenme modellerinin çalışma hızının artırılması, yerel minimuma takılmaması, doğruluk performansının artırılması gibi özellikleri göz önüne alınarak aktivasyon fonksiyonları geliştirilmektedir. ReLU, sigmoid ve tanjant aktivasyon fonksiyonlarının kaybolan gradyan ve negatif bölge gibi problemlerinden dolayı derin öğrenme mimarilerini eğitmek zordur. Bu tez çalışmasında, kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelmek için yeni sabit parametrelili RSignELU aktivasyon fonksiyonları, çekirdek tabanlı hibrit KAF+RSignELU aktivasyon fonksiyonları ve eğitilebilir parametrelili P+RSignELU aktivasyon fonksiyonları önerilmiştir. Ayrıca, derin öğrenme modellerinde hiper-parametre değerleri optimize edilmeden kullanıldığında model başarı performansı çok düşük olmaktadır. Hiper-parametre optimizasyonu, derin öğrenme mimarilerinin, hesaplama maliyetlerinin düşürülmesini ve doğruluk performanslarının artırılmasını sağlamaktadır. Tez çalışmasında, derin öğrenme mimarilerinin hiper-parametrelerinin optimizasyonu için sezgisel yöntemler kullanılmıştır. Çeşitli veriler üzerine gerçekleştirilen deneysel değerlendirmede, hiper-parametre optimizasyonu yapılmıştır ve literatürde bilinen ReLU ve swish aktivasyon fonksiyonlu derin öğrenme mimarilerinin sınıflandırma başarıları değerlendirilmiştir. Önerilen sabit parametrelili RSignELU, çekirdek tabanlı KAF+RSignELU ve eğitilebilir P+RSignELU aktivasyon fonksiyonları ile literatürdeki çalışmalara göre daha yüksek başarı elde edilmiştir. Ayrıca hiper-parametre optimizasyonu için kullanılan yöntemler ile derin öğrenme modellerinin başarıları artırılmıştır.

Anahtar Kelimeler: Derin öğrenme yöntemleri, Hiper-parametre optimizasyonu, Aktivasyon fonksiyonları,

DEVELOPMENT OF NONLINEAR ACTIVATION FUNCTIONS FOR DEEP LEARNING MODELS

Serhat KILIÇARSLAN

Erciyes University, Graduate School of Natural and Applied Sciences

PhD Thesis, September 2021

Supervisor: Assoc. Prof. Dr. Mete ÇELİK

ABSTRACT

Activation functions are developed by taking into account the features of deep learning models, such as increasing the working speed of the model, not getting stuck in the local minimum, and increasing the accuracy performance. Deep learning architectures are difficult to train due to the problems of ReLU, sigmoid and tangent activation functions such as vanishing gradient and negative region. In this thesis study, new fixed parameter activation functions (RSigELU), kernel-based hybrid activation functions (KAF+RSigELU), and trainable P+RSigELU activation functions are proposed to overcome these problems. In addition, when the hyper-parameter values are not optimized in deep learning models, the model success performance is very low.. Hyper-parameter optimization enables deep learning architectures to reduce computational costs and increase accuracy performance. In the thesis study, heuristic methods have been used for optimization of hyper-parameters of deep learning. In the experimental evaluations on various datasets, hyper-parameter optimization was performed and classification success of ReLU and swish activation functions was evaluated. The proposed fixed parameter RSigELU, kernel-based KAF+RSigELU, and trainable P+RSigELU activation functions outperformed other activation functions. In addition, the methods, used for hyper-parameter optimization, increased the performance of the deep learning models

Keywords: Deep learning methods, Hyper-parameter optimization, Activation functions

İÇİNDEKİLER

DERİN ÖĞRENME YÖNTEMLERİ İÇİN DOĞRUSAL OLMAYAN AKTİVASYON FONKSİYONLARININ GELİŞTİRİLMESİ

BİLİMSEL ETİĞE UYGUNLUK	ii
YÖNERGEYE UYGUNLUK.....	iii
TEŞEKKÜR.....	v
ÖZET	vi
ABSTRACT.....	vii
İÇİNDEKİLER	viii
KISALTMALAR.....	xii
TABLolar LİSTESİ.....	xiv
ŞEKİLLER LİSTESİ.....	xviii
GİRİŞ	1

1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR TARAMASI

1.1. EVRİŞİMSEL (KONVOLÜSYONEL) SİNİR AĞLARI.....	8
1.2. YIĞINLANMIŞ OTOKODLAYICI	10
1.3. DESTEK VEKTÖR MAKİNELERİ.....	12
1.4. RELİEFF BOYUT AZALTMA YÖNTEMİ	14
1.5. TEZ KAPSAMINDA KULLANILAN VERİ KÜMELERİ.....	15
1.5.1. Nutrisyonel Anemi Veri Kümesi.....	15
1.5.2. Mikrodizi Veri Kümeleri.....	16
1.5.3. MNIST, Fasion MNIST, CIFAR-10, CIFAR-100 ve IMDB Movie Veri Kümeleri	17
1.6. LİTERATÜR TARAMASI	20

1.6.1. Veri Madenciliği Yöntemleri ile Anemi Verilerin Sınıflandırılması Üzerine Yapılan Çalışmalar.....	20
1.6.2. Veri Madenciliği Yöntemleri ile Mikrodizi Verilerinin Sınıflandırılması Üzerine Yapılan Çalışmalar.....	22
1.6.3. Sabit ve Eğitilebilir Aktivasyon Fonksiyonları Üzerine Yapılan Çalışmalar	24
1.6.4. Parametrik Olmayan Aktivasyon Fonksiyonları Üzerine Yapılan Çalışmalar	27

2. BÖLÜM

RELU VE SWISH AKTİVASYON FONKSİYONLU HİBRİT DERİN ÖĞRENME YÖNTEMLERİ KULLANARAK ANEMİ VERİLERİN SINIFLANDIRILMASI

2.1. GİRİŞ.....	29
2.2. YÖNTEM.....	32
2.3. DENEYSEL DEĞERLENDİRME	38
2.4. SONUÇLAR	45

3. BÖLÜM

RELU AKTİVASYON FONKSİYONLU HİBRİT DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK MİKRODİZİ VERİLERİNİN SINIFLANDIRILMASI

3.1. GİRİŞ.....	46
3.2. YÖNTEM.....	47
3.3. ÖNERİLEN HİBRİT SINIFLANDIRMA YÖNTEMLERİ	49
3.4. DENEYSEL DEĞERLENDİRME	50
3.5. SONUÇLAR	56

4. BÖLÜM

RSİGELU: DERİN SİNİR AĞLARI İÇİN DOĞRUSAL OLMAYAN AKTİVASYON FONKSİYONLARI

4.1.	GİRİŞ.....	58
4.2.	YÖNTEM.....	60
4.3.	ÖNERİLEN RSİGELU AKTİVASYON FONKSİYONLARI	61
4.3.1.	RSigELUS Aktivasyon Fonksiyonu.....	65
4.3.2.	RSigELUD Aktivasyon Fonksiyonu	66
4.4.	DENEYSEL DEĞERLENDİRME	67
4.5.	SONUÇLAR	78

5. BÖLÜM

KAF+RSİGELU: DERİN SİNİR AĞLARI İÇİN DOĞRUSAL OLMAYAN VE ÇEKİRDEK TABANLI AKTİVASYON FONKSİYONLARI

5.1.	GİRİŞ.....	79
5.2.	YÖNTEM.....	81
5.2.1.	MNIST ve Fashion MNIST Veri Kümeleri için Kullanılan ESA Mimarisi.....	81
5.2.2.	CIFAR-10 Veri Kümesi için Kullanılan ESA Mimarisi	82
5.3.	ÖNERİLEN ÇEKİRDEK TABANLI KAF+RSİGELU AKTİVASYON FONKSİYONLARI	82
5.4.	DENEYSEL DEĞERLENDİRME	85
5.5.	SONUÇLAR	97

6. BÖLÜM

P+RSİGELU: DERİN ÖĞRENME ALGORİTMARI İÇİN EĞİTİLEBİLİR PARAMETRİK AKTİVASYON FONKSİYONLARI

6.1.	GİRİŞ.....	100
------	------------	-----

6.2.	YÖNTEM.....	100
6.3.	ÖNERİLEN EĞİTİLEBİLİR P+RSİGELU AKTİVASYON FONKSİYONLARI	100
6.3.1.	Eğitilebilir P+RSigELUS Aktivasyon Fonksiyonu.....	103
6.3.2.	Eğitilebilir P+RSigELUD Aktivasyon Fonksiyonu	103
6.4.	DENEYSEL DEĞERLENDİRME	104
6.5.	SONUÇLAR	113

7. BÖLÜM

TARTIŞMA, SONUÇ VE ÖNERİLER

7.1.	TARTIŞMA VE SONUÇ.....	114
7.2.	ÖNERİLER.....	116
	KAYNAKLAR	117
	EKLER.....	132
	ÖZGEÇMİŞ.....	133

KISALTMALAR

AEN-CMI	: Koşullu Karşılıklı Bilgiye Sahip Uyarlanabilir Elastik Ağ
ALL	: Akut Lenfoblastik Lösemiden
AML	: Akut Miyeloid Lösemi
APL	: Uyarlanabilir Aktivasyon Fonksiyonu
BLU	: Bükülebilir Doğrusal Birim
CNS	: Merkezi Sinir Sistemi
ÇO	: Çaprazlama Oranı
DGA	: Diferansiyel Gelişim Algoritması
DGA-ESA	: Diferansiyel Gelişim Algoritması ile Güçlendirilmiş Evrimsel Sinir Ağları
DGA-YOK	: Diferansiyel Gelişim Algoritması ile Güçlendirilmiş Yığınlanmış Otokodlayıcı
Doğ_Doğ.	: Doğrulama Doğruluk
Doğ_Hata	: Doğrulama Hata
DTW	: Ayrık Dalgacık Dönüşümü
DVM	: Destek Vektör Makinesi
DVM+YOK	: Destek Vektör Makineleri ile Yığınlanmış Otokodlayıcı
DWFS	: Dinamik Ağırlıklı FS
ESA	: Evrimsel Sinir Ağları
FLNN	: Fonksiyonel Bağlantılı Sinir Ağı
GA	: Genetik Algoritma
GA-ESA	: Genetik Algoritması ile Güçlendirilmiş Evrimsel Sinir Ağları
GA-YOK	: Genetik Algoritması ile Güçlendirilmiş Yığınlanmış Otokodlayıcı
KAF	: Çekirdek Tabanlı Aktivasyon Fonksiyonu
KAF+RSigELUD	: Çekirdek Tabanlı Çift Parametrelili RSigELU
KAF+RSigELUS	: Çekirdek Tabanlı Tek Parametrelili RSigELU
k-NN	: K-En Yakın Komşuluk

LReLU	: Leaky ReLU
MeLU	: Mexican ReLU
MWT	: Hareketli Pencere Tekniği
NB	: Naive Bayes
ÖF	: Ölçekleme Faktörü
P+RSigELUD	: Çift Parametrelili Eğitilebilir Rsignet
P+RSigELUS	: Tek Parametrelili Eğitilebilir Rsignet
PFTS	: Parametric Flatten-T Swish
PSO	: Parçacık Sürü Optimizasyonu Algoritması
PSO-ESA	: Parçacık Sürü Optimizasyonu Algoritması İle Güçlendirilmiş Evrimsel Sinir Ağları
PSO-YOK	: Parçacık Sürü Optimizasyonu Algoritması ile Güçlendirilmiş Yığınlanmış Otokodlayıcı
RBF	: Radyan Tabanlı Fonksiyon
RSigELUD	: Çift Parametrelili Rsignet
RSigELUS	: Tek Parametrelili Rsignet
YAKA	: Yapay Arı Kolonisi Algoritma
YAKA-ESA	: Yapay Arı Kolonisi Algoritması ile Güçlendirilmiş Evrimsel Sinir Ağları
YAKA-YOK	: Yapay Arı Kolonisi Algoritması ile Güçlendirilmiş Yığınlanmış Otokodlayıcı
YOK	: Yığınlanmış Otokodlayıcı
YSA	: Yapay Sinir Ağları

TABLOLAR LİSTESİ

Tablo 1.1. Çekirdek Fonsiyonları.....	14
Tablo 1.2. Nutrisyonel anemi veri kümeleri parametreleri	16
Tablo 1.3. Mikrodizi veri kümeleri.	16
Tablo 2.1. Hiper-parametre aralığı.....	33
Tablo 2.2. ESA modeli hiper-parametreleri.....	33
Tablo 2.3. Tez çalışmasında kullanılan ESA mimarisi	34
Tablo 2.4. YOK modeli hiper-parametreleri ve mimarisi.....	34
Tablo 2.5. Farklı çaprazlama ve mutasyon değerine göre GA'nın değerlendirme sonuçları	35
Tablo 2.6. Farklı hızlandırma sabitleri ve başlangıç ağırlık sabiti değerine göre PSO değerlendirme sonuçları	36
Tablo 2.7. Farklı limit parametreleri için YAKA değerlendirme sonuçları.....	37
Tablo 2.8. Farklı ölçekleme faktörü ve çaprazlama oranı için DGA değerlendirme sonuçları	37
Tablo 2.9. GA ile ESA mimarisinin sonuçları	39
Tablo 2.10. GA ile YOK mimarisinin sonuçları	39
Tablo 2.11. PSO ile ESA mimarisinin sonuçları	40
Tablo 2.12. PSO ile YOK mimarisinin sonuçları	40
Tablo 2.13. YAKA ile ESA mimarisinin sonuçları	41
Tablo 2.14. YAKA ile YOK mimarisinin sonuçları	42
Tablo 2.15. DGA ile ESA mimarisinin sonuçları	43
Tablo 2.16. DGA ile YOK mimarisinin sonuçları	43
Tablo 2.17. Tüm sezgisel yöntemlerin deneysel değerlendirme sonuçları	44
Tablo 3.1. YOK modelinin hiper-parametreleri.....	47
Tablo 3.2. ESA modelinin hiper-parametreleri.....	48
Tablo 3.3. Çalışmada kullanılan ESA modeli ve parametreleri	48
Tablo 3.4. Yumurtalık mikrodizi veri kümesi için deneysel sonuçları	51
Tablo 3.5. Lösemi mikrodizi veri kümesi için deneysel sonuçları	52
Tablo 3.6. CNS mikrodizi veri kümesi için deneysel sonuçları.....	53

Tablo 3.7. Mikrodizi veri kümeleriyle ilgili çalışmaların karşılaştırılması	56
Tablo 4.1. Çalışmada kullanılan VGG mimarinin parametreleri	61
Tablo 4.2. MNIST veri kümesi kullanılarak β parametresinin belirlenmesi.....	67
Tablo 4.3. Fasion-MNIST veri kümesi kullanılarak β parametresinin belirlenmesi.....	67
Tablo 4.4. CIFAR-10 veri kümesi kullanılarak β parametresinin belirlenmesi.....	68
Tablo 4.5. IMDB veri kümesi kullanılarak β parametresinin belirlenmesi.....	68
Tablo 4.6. MNIST veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları	68
Tablo 4.7. MNIST veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları	69
Tablo 4.8. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları...	69
Tablo 4.9. Fashion MNIST veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları.....	71
Tablo 4.10. Fashion MNIST veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları.....	71
Tablo 4.11. Fashion MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	72
Tablo 4.12. CIFAR-10 veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları.....	73
Tablo 4.13. CIFAR-10 veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları.....	73
Tablo 4.14. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	74
Tablo 4.15. IMDB Movie veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları.....	75
Tablo 4.16. IMDB Movie veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları.....	76
Tablo 4.17. IMDB Movie veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	76
Tablo 5.1. MNIST veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranları	86

Tablo 5.2. MNIST veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları.....	86
Tablo 5.3. MNIST veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları.....	87
Tablo 5.4. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları...	87
Tablo 5.5. Fashion MNIST veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranı.....	89
Tablo 5.6. Fashion MNIST veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları.....	90
Tablo 5.7. Fashion MNIST veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları.....	91
Tablo 5.8. Fashion MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	91
Tablo 5.9. CIFAR-10 veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranı	93
Tablo 5.10. CIFAR-10 veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları.....	94
Tablo 5.11. CIFAR-10 veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları.....	95
Tablo 5.12. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranı	95
Tablo 6.1. MNIST veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları	104
Tablo 6.2. MNIST veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları	105
Tablo 6.3. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları.	105
Tablo 6.4. CIFAR-10 veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları.....	107
Tablo 6.5. CIFAR-10 veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları.....	107
Tablo 6.6. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	108

Tablo 6.7. CIFAR-100 veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları.....	110
Tablo 6.8. CIFAR-100 veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları.....	110
Tablo 6.9. CIFAR-100 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları	110



ŞEKİLLER LİSTESİ

Şekil 1.1. ESA modelinin blok diyagramı	9
Şekil 1.2. Yığılanmış otokodlayıcı model.....	12
Şekil 1.3. Doğrusal olamayan verilerin DVM ile sınıflandırılması	13
Şekil 1.4. MNIST veri kümesi örnekleri.....	17
Şekil 1.5. Fashion MNIST veri kümesi örnekleri	18
Şekil 1.6. CIFAR-10 veri kümesi örnekleri	19
Şekil 1.7. CIFAR-100 veri kümesi örnekleri	19
Şekil 1.8. IMDB Movie veri kümesi örnekleri	20
Şekil 3.1. Önerilen hibrit modelin blok diyagramı.	50
Şekil 3.2. Yumurtalık mikrodizi veri kümesi için algoritmaların ROC eğrisi	54
Şekil 3.3. Lösemi mikrodizi veri kümesi için algoritmaların ROC eğrisi	55
Şekil 3.4. CNS mikrodizi veri kümesi için algoritmaların ROC eğrisi.....	55
Şekil 4.1. VGG tabanlı ESA mimarisi	60
Şekil 4.2. Sigmoid fonksiyon	62
Şekil 4.3. Hiperbolik tanjant fonksiyon	63
Şekil 4.4. Literatürdeki bazı aktivasyon fonksiyonlarının davranışları	63
Şekil 4.5. Önerilen (a) RSigELUS ve (b) RSigELUD fonksiyonlarının davranışları.....	64
Şekil 4.6. MNIST veri kümesi aktivasyon fonksiyonlarının ortalama doğrulama (validation) hata (Doğ. Hata) sonuçları.....	69
Şekil 4.7. Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama.....	72
Şekil 4.8. CIFAR-10 veri kümesi için öne aktivasyon fonksiyonlarının ortalama doğrulama hata (Doğ. Hata) sonuçları	74
Şekil 4.9. IMDB Movie veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama hata (Doğ. Hata) sonuçları	77
Şekil 5.1. MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları	88
Şekil 5.2. Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları	92

Şekil 5.3. CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları	96
Şekil 6.1. Önerilen (a) P+RSigELUS ve (b) P+RSigELUD aktivasyon fonksiyonlarının davranışı	102
Şekil 6.2. MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları	106
Şekil 6.3. CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları	109
Şekil 6.4. CIFAR-100 veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları	111

GİRİŞ

Derin öğrenme, günümüzdeki popüler makine öğrenmesi yöntemlerinden olup sınıflandırma, doğal dil işleme, görüntü işleme, zaman serileri analizi gibi birçok alanda aktif olarak kullanılmaktadır [1]. Derin öğrenme yöntemleri yapay sinir ağları (YSA) yapısına benzer şekilde hesaplamalara dayanmaktadır. Geçmiş yıllarda, bilgisayar donanımlarının kısıtlı olmasıyla klasik YSA'nın eğitilmesi zordu ve dolayısıyla, YSA yapısındaki katmanların sayısı az tutulmaktaydı. Derin öğrenme yöntemlerin sahip olduğu çok fazla katman sayısı ile karmaşık problemlerin çözümüne önemli katkıları olmuştur. Derin öğrenme yöntemlerinden evrimsel sinir ağları (ESA) ve Yığılanmış otokodlayıcı (YOK) üzerine literatürde çok sayıda çalışmalar gerçekleştirilmiştir.

Sinir ağı mimarilerinin eğitim aşamasında, modelin çalışmasının hesaplama maliyetini azaltma, yerel minimuma takılmama, doğruluk oranını artırma gibi amaçlar için aktivasyon fonksiyonları kullanılmaktadır. Sinir ağlarının değişkenler arasındaki ve katmanları arasındaki karmaşık ilişkiyi modelleyebilmesi ve öğrenebilmesi için aktivasyon fonksiyonları önemlidir. İyi tasarlanmış aktivasyon fonksiyonu, modellerin performanslarında etkilidir. Derin öğrenme yöntemlerinde, aktivasyon katmanlarında ağa gelen giriş bilgisinin işlenmesi ve bu girdilere karşılık gelen çıktının üretilmesi için aktivasyon fonksiyonları kullanılmaktadır. Literatürde veriye ve probleme uygun çeşitli aktivasyon fonksiyonları önerilmiştir. Derin öğrenmede kullanılan aktivasyon fonksiyonlarının, doğrusal olmama ve türevinin alınabilir olması beklenmektedir [2].

Aktivasyon fonksiyonları doğrusal olduğunda tek boyutlu YSA özelliğine sahip doğrusal regresyon gibi davranmaktadır. Ancak derin öğrenme mimarileri, yaygın olarak karmaşık gerçek dünya problemleriyle ilgilendiğinden, doğrusal aktivasyon fonksiyonu başarılı sonuç verememektedir. Bundan dolayı doğrusal olmayan aktivasyon fonksiyonları çok katmanlı derin sinir ağlarında tercih edilerek karmaşık gerçek dünya

verilerinden anlamlı özellikleri öğrenebilmektedir [2]. Derin sinir ağlarında geri yayılım algoritması kullanılarak parametrelerin güncellenmesi sağlanır. Güncelleme işleminde türev değeri geri döndürüldüğünden, mimari içerisindeki kullanılan aktivasyon fonksiyonunun türevinin alınabilir olması gerekmektedir. Ayrıca, derin sinir ağları mimarilerinde daha derin katmanlara doğru gidildikçe nöronlarda güncellemeler zorlaşmaktadır. Bundan dolayı kaybolan gradyan problemi başlamaktadır. Kaybolan gradyan problemi, derin sinir ağlarının derinleşerek öğrenmesinin önünde büyük engel oluşturmaktadır [6]. Bu engel mimarinin yerel minimuma takılma ihtimalini yükseltmektedir. Çünkü sinir ağları derinleştikçe kaybolan gradyan ve negatif bölge problemi ortaya çıkmaktadır. Literatürdeki bilinen problemlerin üstesinden gelmek için aktivasyon fonksiyonları doğrusal ve doğrusal olmayan olmak üzere ikiye ayrılabilir. Bu problem, yaygın olarak kullanılan ReLU [7], sigmoid [4] ve hiperbolik tanjant [5] aktivasyon fonksiyonlarında görülmektedir.

Literatürde, mevcut problemlerin üstesinden gelmek için, ReLU [7], LReLU [3], SELU [8], ELU [9], PReLU [10], Hexpo [11], LISA [12], ALISA [12], PELU [13], Parametri Flatten-T Swish (PFTS) [14] ve Mexican ReLU (MeLU) [15] gibi sabit ve eğitilebilir aktivasyon fonksiyonları önerilmiştir.

Kaybolan gradyan probleminin üstesinden gelebilmek için önerilen ReLU aktivasyon fonksiyonu [7], pozitif bölge değerlerine göre işlem yaparken negatif bölge değerlerini sıfır olarak ayarlamaktadır. Bu durumda negatif çıktıların türevi alınamamakta ve negatif bölgedeki değerlerin öğrenme sırasında göz ardı edilmesine neden olmaktadır. Bu problemin üstesinden gelmek için literatürde LReLU [3], ELU [8], PReLU [10] gibi aktivasyon fonksiyonu çalışmaları yapılmıştır. Ayrıca, literatürde çekirdek tabanlı aktivasyon fonksiyonunda (kernel activation function - KAF), mevcut problemlerin dışında, ön yargı kayması, esnek olmama ve uyarlanamama problemleri ortaya çıkmaktadır [17]. Önyargı kayması problemi, pozitif bölgelerde çalışan algoritmalarda negatif değerlerin pozitif değerlere indirgenmesidir. Bu durumda fonksiyonun çıkış değerlerinin ortalamasında negatif sonuç alınamaz. Önyargı kaymasından dolayı, modelin yakınsama hızı ve performansı doğrudan etkilenmektedir. [16]. Uyarlanamama problemi, önyargı kayması olduğunda eğitim sırasında modelin esnek bir şekilde çalışmasına izin verilmemesidir. Bu durumda, aktivasyon fonksiyonlarının doğrusal

olmayan temsil kabiliyeti zayıflatılmaktadır [14]. Esnek olmama problemi, aktivasyon fonksiyonları üzerinde negatif değerlerin göz ardı edilmesiyle, modelin ifade gücünü ve performansını zayıflatmaktadır [14]. KAF parametrik olmayan aktivasyon fonksiyonlarıdır ve az sayıda gizli katman içeren derin sinir ağlarının doğruluğunu geliştirmek için önerilmiştir.

Bu tez kapsamında, iki tür çalışma gerçekleştirilmiştir. Birincisi derin öğrenme yöntemlerinin hiper-parametreleri, sezgisel yöntemlerden parçacık sürü optimizasyon algoritması (PSO) [18], genetik algoritma (GA) [19], diferansiyel gelişim algoritması (DGA) [20] ve yapay arı kolonisi algoritması (YAKA) [21] kullanılarak belirlenmiştir. Belirlenen hiper-parametrelerinden sonra, literatür de bilinen ReLU ve swish aktivasyon fonksiyonlu derin öğrenme yöntemleri kullanılarak anemi ve DNA mikrodizi [22] veri kümeleri üzerinde deneysel değerlendirmeler gerçekleştirilmiş ve aktivasyon fonksiyonlarının başarı karşılaştırılması yapılmıştır. İkincisi, kaybolan gradyan, negatif bölge, önyargı kayması, uyarlanamama ve esnek olmama problemlerinin üstesinden gelmek amacıyla yeni sabit parametrelili aktivasyon fonksiyonları (RSigELUS ve RSigELUD), çekirdek tabanlı hibrit aktivasyon fonksiyonları (KAF+RSigELUS ve KAF+RSigELUD) ve eğitilebilir aktivasyon fonksiyonları (P+RSigELUS ve P+RSigELUD) önerilmiştir. Önerilen yeni tek ve çift parametrelili RSigELU aktivasyon fonksiyonları, ReLU, sigmoid ve ELU aktivasyon fonksiyonlarının birleşiminden meydana gelmektedir. Önerilen RSigELU aktivasyon fonksiyonlarının türevleri, pozitif aktivasyon bölgesinde ReLU ve sigmoid fonksiyonlarının birleşimi davranışını, negatif aktivasyon bölgesinde ELU fonksiyonu davranışını ve doğrusal aktivasyon bölgesinde ise doğrusal aktivasyon fonksiyonu davranışını sergilemektedir [7]. Bu özellikleri ile önerilen aktivasyon fonksiyonlarının türevleri alınabilmekte ve bu fonksiyonlar kaybolan gradyan, negatif bölge, önyargı kayması, uyarlanamama ve esnek olmama problemlerinin üstesinden gelerek öğrenme işlemini sürekli devam ettirebilmektedir. Önerilen RSigELU aktivasyon fonksiyonları ile MNIST, Fashion MNIST, CIFAR-10, CIFAR-100 ve IMDB Movie veri kümeleri [23-26] üzerinde deneyler gerçekleştirilmiştir ve başarılı sonuçlar elde edilmiştir.

Tezin Amacı ve Katkısı

Bu tez amacı, kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelmek için yeni sabit parametrelili RSignELU aktivasyon fonksiyonları, çekirdek tabanlı hibrit KAF+RSignELU aktivasyon fonksiyonları ve eğitilebilir parametrelili P+RSignELU aktivasyon fonksiyonları geliştirmektir. Ayrıca, derin öğrenme yöntemlerinin sezgisel yöntemler ile hiper-parametre optimizasyonu yapmaktır. Tez kapsamında ele alınan aktivasyon fonksiyonları ve hiper-parametre optimizasyonu yaklaşımları ve yapılan çalışmaların literatüre katkıları aşağıda listelenmiştir.

1. ReLU ve swish aktivasyon fonksiyonlu hibrit derin öğrenme yöntemleri kullanarak anemi verilerin sınıflandırılması: Bu çalışmada derin öğrenme yöntemlerinin hiper-parametrelerinin sezgisel yöntemlerden parçacık sürü optimizasyon algoritması (PSO) [18], genetik algoritma (GA) [19], diferansiyel gelişim algoritması (DGA) [20] ve yapay arı kolonisi algoritmasıyla (YAKA) [21] belirlenmiştir. Belirlenen hiper-parametrelerinden sonra, literatür de bilinen ReLU ve swish aktivasyon fonksiyonlu derin öğrenme yöntemleri ile nutrisyonel anemi hastalığının tahmin edilmesi üzerine çalışma yapılmıştır. Yapılan çalışmanın detayları ve deneysel değerlendirmeler Bölüm 2’de sunulmuştur. Elde edilen sonuçlar derlenerek birisi yayımlanmış [27] ve biriside revizyon aşamasında olan iki yayın haline getirilmiştir.

2. ReLU aktivasyon fonksiyonlu hibrit derin öğrenme yöntemleri kullanılarak mikrodizi verilerinin sınıflandırılması: Bu çalışmada mikrodizi veri kümelerinden yumurtalık, lösemi ve merkezi sinir sistemi (CNS) sınıflandırılması için, boyut azaltma algoritması olan ReliefF ve yığınlanmış otokodlayıcı (YOK) ile destek vektör makinesi (DVM) ve ESA kullanılarak hibrit bir yaklaşım önerilmiştir. Üç mikrodizi veri kümesine uygulanan yöntemler arasında hibrit ReliefF+ESA yöntemi ile literatürde bilinen en iyi performans elde edilmiştir. Yapılan çalışmanın detayları ve deneysel değerlendirmeler Bölüm 3’de sunulmuştur. Elde edilen sonuçlar derlenerek bir dergide yayınlanmıştır [22].

3. RSignELU: Derin sinir ağları için doğrusal olmayan aktivasyon fonksiyonları: Bu çalışmada, derin öğrenme algoritmalarında meydana gelen kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmek için tek ve çift parametrelili doğrusal olmayan yeni sabit parametrelili RSignELU aktivasyon fonksiyonları önerilmiştir. Bu tez

kapsamında RSignELUS ve RSignELUD aktivasyon fonksiyonları, kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmektedir ve pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktif çalışabilmesi için önerilmiştir. Yapılan çalışmanın detayları ve deneysel değerlendirmeler Bölüm 4’de sunulmuştur. Elde edilen sonuçlar derlenerek bir dergide yayınlanmıştır [28].

4. KAF+RSignELU: Derin sinir ağları için doğrusal olmayan ve çekirdek tabanlı aktivasyon fonksiyonları: Bu çalışmada, çekirdek tabanlı aktivasyon fonksiyonu (KAF) ile RSignELU aktivasyon fonksiyonları birleştirilerek yeni çekirdek tabanlı hibrit aktivasyon fonksiyonları (KAF+RSignELUS ve KAF+RSignELUD) önerilmiştir. KAF aktivasyon fonksiyonu, belirli değerler altında veri kümeleri üzerinde öğrenme işlemini gerçekleştirmeyip, negatif bölge, negatif bölgeden kaynaklı olarak önyargı kayması (bias shift), kaybolan gradyan gibi problemleri ile yüzleşmektedir. Bu problemlerin üstesinden gelinebilmek için KAF+RSignELU aktivasyon fonksiyonları önerilmiştir. Yapılan çalışmanın detayları ve deneysel değerlendirmeler Bölüm 5’de sunulmuştur. Elde edilen sonuçlar derlenerek bir dergiye gönderilmiş ve revizyon aşamasındadır.

5. P+RSignELU: Derin öğrenme algoritması için eğitilebilir parametrik aktivasyon fonksiyonu: Bu çalışmada, eğitilebilir yeni özgün aktivasyon fonksiyonları (P+RSignELUS ve P+RSignELUD) önerilmiştir. Sabit parametrelili RSignELU aktivasyon fonksiyonlarının en ideal eğim parametrelerini belirlemesi için model üzerinde çok sayıda deneme yapılmaktadır. Önerilen P+RSignELU aktivasyon fonksiyonları ile, model ilk çalıştırmada aktivasyon fonksiyonuna en uygun eğim parametresini belirleyerek öğrenme sürecini en kısa sürede gerçekleştirebilmektedir. Yapılan çalışmanın detayları ve deneysel değerlendirmeler Bölüm 6’de sunulmuştur. Elde edilen sonuçlar derlenerek bir dergiye gönderilmiştir.

Tezin Organizasyonu

Bu tez çalışması şu şekilde organize edilmiştir:

- Giriş bölümünde, tez konusunun tanıtımı yapılmış, aktivasyon fonksiyonlarında hangi problemlerin olduğu ifade edilmiş, tez kapsamında yapılan çalışmaların literatüre katkıları sunulmuş ve tezin takip eden bölümleri tanıtılmıştır.
- Birinci bölümde, tez çalışmasında kullanılan yöntemler ve materyaller detaylı şekilde tanıtılmıştır. Daha sonra literatürde yapılan çalışmalardan bahsedilmiştir.
- İkinci bölümde, derin öğrenme yöntemlerinin hiper-parametrelerinin sezgisel yöntemlerle belirlenmesi üzerine çalışma yapılmıştır. Belirlenen hiper-parametre değerlerinden sonra, literatürdeki ReLU ve swish ile nutrisyonel anemi hastalığının tahmin edilmesi üzerine çalışmalar yapılmıştır. Bölüm sonunda, önerilen yöntemin başarımı test edilerek sunulmuştur.
- Üçüncü bölümde, mikrodizi veri kümesinin boyutunu indirgemek için kullanılan ReliefF ile YOK algoritmaların yanında, ESA mimarisi ile sınıflandırma işlemini ReLU aktivasyon fonksiyonu kullanılarak gerçekleştirilmiş ve deneysel değerlendirme sonuçları sunulmuştur.
- Dördüncü bölümde, sabit parametrelili RSigELU aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonlarının çalışma mantığından bahsedilmiştir. Bölümün sonunda, önerilen aktivasyon fonksiyonlarının başarımı test edilerek sunulmuştur.
- Beşinci bölümde, çekirdek tabanlı KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonlarının çalışma mantıklarından bahsedilmiştir. Bölümün sonunda, önerilen aktivasyon fonksiyonlarının başarımı test edilerek sunulmuştur.
- Altıncı bölümde, eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonlarının çalışma mantıklarından bahsedilmiştir. Bölümün sonunda, önerilen aktivasyon fonksiyonlarının başarımı test edilerek sunulmuştur.

- Yedinci bölümde ise tez kapsamında yapılan çalışmalardan elde edilen sonuçlar değerlendirilmiş ve gelecekte yapılabilecek çalışmalara dair önerilere yer verilmiştir.



1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR TARAMASI

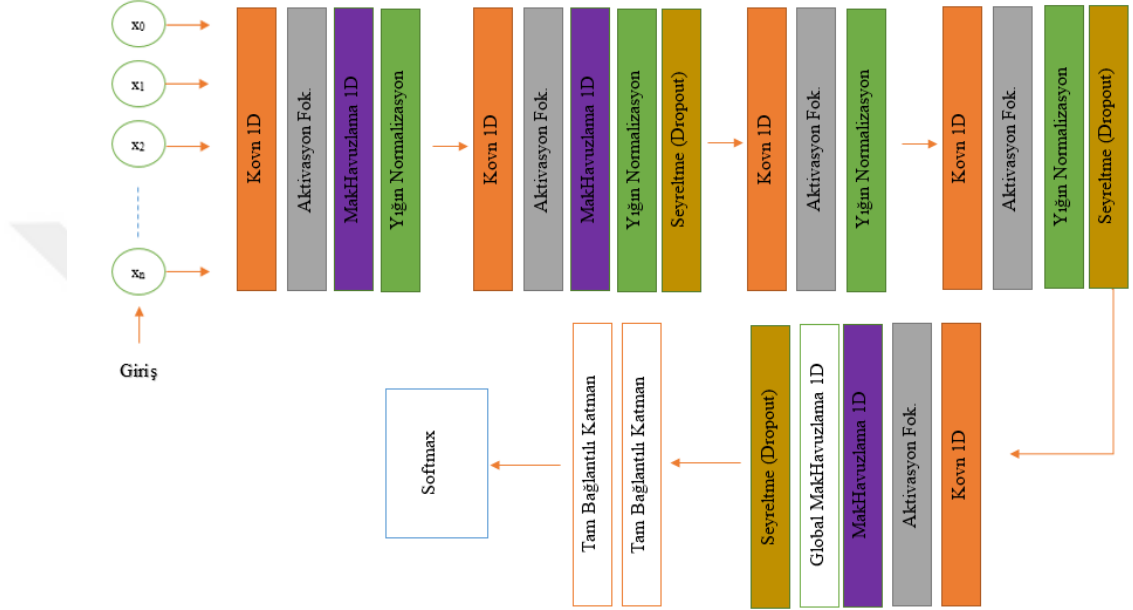
Bu bölümde, ilk önce derin öğrenme yöntemlerinden ESA ve YOK, klasik veri madenciliği yöntemlerinden destek vektör makinesi (DVM) ve ReliefF algoritmaları sunulmuştur. İkinci olarak bu tez kapsamında kullanılan veri kümeleri hakkında bilgi verilmiştir. Son olarak da literatürde yapılan çalışmalardan bahsedilmiştir.

1.1. Evrişimsel Sinir Ağları

Evrişimsel sinir ağları (ESA) Yan liu tarafından önerilmiştir [29]. Biyolojiden esinlenerek üretilen çok katmanlı sinir ağlarıdır [30, 31]. ;ESA yaygın olarak görsel algılama ve nesne tanıma uygulamalarında görüntü işleme görevlerini yerine getirebilmek için tasarlanmıştır. Temel olarak YSA gibi, ESA öğrenebilen ağırlıklar ve önyargı parametrelerinden oluşmaktadır. ESA iki temel kısımdan meydana gelmektedir [29]. Bu kısımlardan birincisinde özellik çıkartma işlemi evrişim ve havuzlama katmanlarında gerçekleşirken, ikinci kısımda sınıflandırma işlemi gerçekleştirilmektedir. Birinci kısım, konvolüsyon, aktivasyon fonksiyonu, normalizasyon ve havuzlama katmanlarından oluşur. İkinci kısım ise tam bağlantı ve softmax katmanlarından oluşmaktadır. ESA modeli içerisinde doğrusal olmayan veri kümelerindeki gizli kalmış ilişkileri ortaya çıkarmak için tam bağlantılı katman kullanılmaktadır [29]. Şekil 1.1’de ESA modelinin blok diyagramı görülmektedir.

Şekil 1.1’deki ESA modelinde evrişim işlemi farklı boyuttaki filtrelerin (3×3 , 5×5 , 7×7) ham veri kümesi üzerinde dolaşması sonucunda Eşitlik 1.1’deki gibi bir özellik haritası oluşturmaktadır. ESA’daki adım (stride) değeri, konvolüsyon işlemi sırasında matris yapısına dönüştürülen veriler üzerinde bir veya daha fazla bitsel kaydırma gerçekleştirmek için kullanılır. Bu süreç, doğrudan çıktı boyutunu etkiler. Seçilen filtre

boyutu ile üretilen özellik haritasının boyutu belirlenmektedir. Filtrenin dolaşmasının sonucunda elde edilen özellik haritasıyla veri kümesine ait özelliklerin keşfedilmesi sağlanmaktadır. Eşitlik 1.1’de σ aktivasyon fonksiyonunu, w_i ağırlıkları veya filtreleri, x_i giriş verisinin yerel bölgesini, b_i önyargı değerini ve y_i elde edilen matris çarpımının sonucunu ifade etmektedir [29, 32].



Şekil 1.1. ESA modelinin blok diyagramı

$$y_i = \sigma(w_i x_i + b_i) \quad (1.1)$$

Evrişim işleminin ardından temel olarak ESA’nın işlem yükünü ve giriş boyutunu azaltmak için maksimum havuzlama katmanı kullanılır. Boyut azaltma işlemi ile konvolüsyon katmanından gelen özellik haritasının boyutunun belirlenen havuzlama filtrelerine (3×3 , 5×5 , 7×7) ve adım parametrelerine göre azaltılması yapılmaktadır. Maksimum havuzlama yönteminde, belirlenen filtre boyutu ile elde edilen özellik haritası üzerinden dolaşımlar yapılmaktadır. Dolaşım sırasında havuzlama filtresiyle mevcut özellik haritası içerinden maksimum değerler alınmasıyla yeni özellik haritası oluşturulur [33]. Verilerin normalizasyonu için yığın normalizasyon katmanı kullanılır. Bu katman, eğitim sürecini hızlandırma, modeli düzenli duruma getirme ve başlangıçta kullanılan değerlere bağımlılığı azaltma amacıyla kullanılır [22, 34]. Normalizasyondan sonra, öğrenme sırasında ezberlemenin önüne geçmek için seyreltme katmanı (dropout),

modelin her çalışmasında bağlantılarının seyreltilmesi amacıyla kullanılır. Seyreltme işleminin ardından sınıflandırma gerçekleştirilmesi için tam bağlantılı katman ile tüm katmanların birbiri ile bağlantısı sağlanır. Daha sonra sınıflandırma için softmax veya sigmoid aktivasyon fonksiyonları yardımıyla verilerin hangi sınıfa ait olduğu hesaplanmaktadır [34-36]. Softmax, lojistik regresyonun genişletilmiş bir versiyonudur ve iki veya daha fazla sınıf problemi için olasılığa dayalı bir doğrusal sınıflandırıcıdır. ESA'da ağırlıkların doğru şekilde güncellenmesi öğrenme süreci için önemlidir. Ağırlıkları güncellemek için geri yayılım algoritması kullanılmaktadır.

1.2. Yığınlanmış Otokodlayıcı

Yığınlanmış otokodlayıcı (YOK) denetimsiz bir öğrenme olup, YSA'ların bir çeşidi olarak giriş, gizli ve çıkış katmanlarından oluşan bir derin öğrenme algoritmasıdır. Otokodlayıcılar çok karmaşık veri kümelerinde YSA'lardan daha iyi başarı oranı elde etmek ve mevcut ağırları iyileştirmek için kullanılmaktadır [37,38].

Otokodlayıcılar giriş ve çıkış katmanındaki nöron sayısı, gizli katmana göre daha fazladır ve giriş ve çıkış katman sayısı birbirine eşit olarak ayarlanmaktadır. Öğrenme işlemi gizli katman aracılığıyla gerçekleşmektedir. Gizli katmandaki nöron sayısı otokodlayıcının performansını etkileyen en önemli faktördür. Giriş katmanından gelen veri kümesi gizli katmana ulaşırken, veri kümesini sıkıştırma işlemi Eşitlik 1.2'deki gibi gerçekleştirilerek daha küçük boyutta veri kümesi elde edilir. Eşitlik 1.2 ve Eşitlik 1.3'de girdi katmanı olan x_a/y_b 'nin a/b 'inci nöron değeri, gizli katmandaki y_b 'nin b 'inci nöronuna aktarılan değeri, n nöron sayısı, girdi katmanındaki w_{ab}/w_{ba} 'nın a nöronundan b nöronuna giden ağırlık değeri ve f ise aktivasyon fonksiyonunları (tanh, sigmoid, softmax) temsil etmektedir [39]. Modeli sıkıştırma aşamasında girdilerin yeniden yapılandırılması için eğitilebilir ve modelin gizli katman aracılığıyla girdilerin en iyi temsilcilerini öğrenmesi sağlanır. Bunun için sıkıştırma işlemi eğitim aşamasında **ölçekli eşlenik eğim algoritması** ve **çapraz entropi maliyeti** fonksiyonları kullanılmıştır. Gizli katmandaki sıkıştırılmış olan veri kümesi çıkış katmanına ulaşırken Eşitlik 1.2 ile hesaplanır. Eşitlik 1.3 ile gürültüden arındırılmış giriş verilerine çok yakın yeni veri kümesi elde edilir. Yapılan bu işlemlere sınıflandırma yapmadan veri kümesine kodlayıcı (encoder) ve kod çözücü (decoder) işlemleri adı verilmektedir. Otokodlayıcının asıl amacı sınıflandırma yapmak

değildir. Otokodyayıcılar, belli boyuttaki veri kümesinin öznitelik sayısını daha küçük bir boyuttaki öznitelik sayısına düşürmek için kullanılır. Giriş sinyallerinin toplam değeri 0 olduğunda öğrenme gerçekleşmeyecektir. Böylece, bu problemin üstesinden gelebilmek için toplam değere önyargı eklenerek öğrenmeyi hızlandırmakta ve yerel optimuma takılmanın önüne geçilebilmektedir [37, 40-41].

$$y_b = f\left(\sum_{a=1}^n x_a \times w_{ab}\right) + \text{önyargı} \quad (1.2)$$

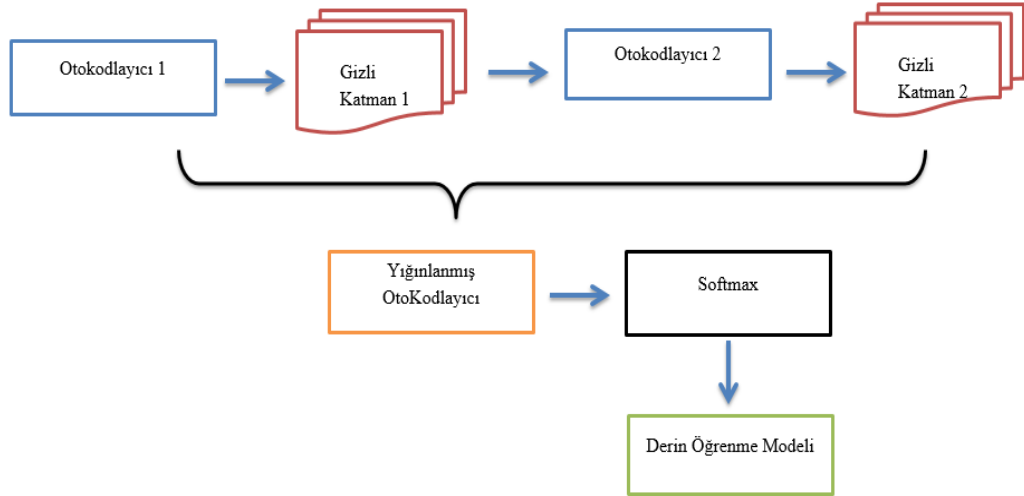
$$x'_a = f\left(\sum_{b=1}^m y_b \times w_{ba}\right) + \text{önyargı} \quad (1.3)$$

$$\text{OKH}(x_a, x'_a) = \|x_a - x'_a\|_2^2 = \left\|x_a - \left(\sum_{b=1}^m y_b \times w_{ba}\right)\right\|_2^2 \quad (1.4)$$

Eşitlik 1.4'de gürültüden arındırılmış ve girdi katmanındaki değere yakın değerler oluşturulabilmesi için x'_a ve x_a iki değer arasındaki farkın karesini küçültmek amacıyla **ortalama karesel hata (OKH)** yöntemi kullanılır. Ayrıca algoritmada kullanılan parametrelerin yapılandırılması için **Levenberg-Marquardt** algoritması kullanılır [42].

YOK ve türevlerinin yapısında aşırı öğrenme sorunu bulunmaktadır. Aşırı öğrenme, bir modelin eğitim kümesindeki verilerde yüksek başarı oranı gösterirken, eğitim dışında kullanılan test veri kümelerinde düşük başarı oranı göstermektedir. Aşırı öğrenme sorununun üstesinden gelmek için Eşitlik 1.5'deki düzenlileştirme (regularization) yöntemi olan **L2WeightRegularization** kullanılır.

$$\min\left[\left(\sum_{b=1}^n (x'_a - x_a)^2\right) + \gamma L(w)\right] \quad (1.5)$$

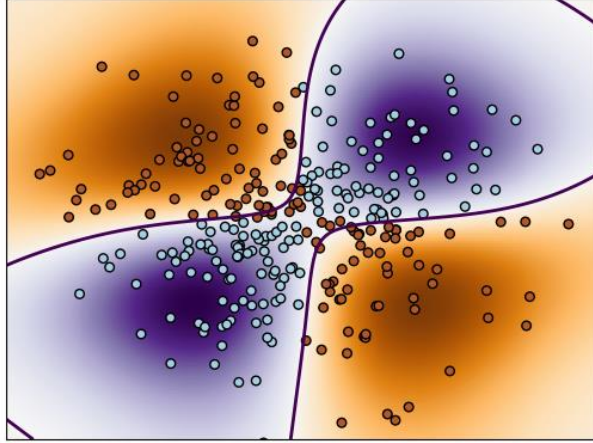


Şekil 1.2. Yığınlanmış otokodlayıcı modeli [43]

Şekil 1.2’de görüldüğü gibi karmaşık verilerin sınıflandırılması için arka arkaya oto kodlayıcıların bağlanmasıyla YOK mimarisi elde edilmektedir. Birinci otokodlayıcının gizli katmanı ikinci oto kodlayıcının girişi olarak arka arkaya bağlanarak boyut azaltma işlemi gerçekleştirilmektedir. Böylece veri kümesindeki gürültüler yok edilerek yeni veri kümesi elde edilmektedir. Boyut azaltma işleminden sonra sınıflandırma işleminde başarı oranını arttırabilmek için lojistik regresyon temelli softmax katmanı yardımıyla yapının denetimli bir öğrenme modeli olması sağlanmış olmaktadır [43].

1.3. Destek Vektör Makineleri

Destek vektör makineleri (DVM) Vapnik tarafından istatistiksel öğrenme teorisine dayanılarak geliştirilmiş, sınıflandırma ve regresyon için kullanılabilen bir yöntemdir [44, 45]. DVM, zaman serileri tahmini, canlı türlerinin sınıflandırılması, hastalıkların teşhis edilmesi, bankacılık, lojistik gibi birçok alanda sınıflandırma amacıyla kullanılmaktadır [48]. DVM ilk olarak iki sınıflı ve doğrusal verilerin sınıflandırılması için kullanılırken, zamanla çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması için kullanımı yaygınlaşmıştır (Şekil 1.3). DVM, karar fonksiyonunun bulunması ve verileri birbirinden ayırabilen hiper-düzlemin tespit edilmesi ile sınıflandırmayı gerçekleştiren yöntemdir. DVM sınıflandırma işlemlerini kareler (squared) optimizasyon yöntemlerine dönüştürerek çözer ve öğrenme basamağında işlem sayısını azaltarak diğer algoritmalara göre daha hızlı çözüm sunabilir [46].



Şekil 1.3. Doğrusal olmayan verilerin DVM ile sınıflandırılması

Farklı sınıf özelliklerine sahip veri kümelerini birbirinden ayırmak için Şekil 1.3’deki gibi bir ayıraç çizilerek hiper-düzlemde sınıflandırma yapılmaktadır. DVM denetimli bir öğrenme algoritması olup, veri kümesini eğitim ve test olarak ayrılmasıyla öğrenme işlemini gerçekleştirmektedir [47].

DVM’de sınıflandırma işlemlerinin gerçekleştirilme sürecinde uygun çekirdek fonksiyonunun seçilmesi önemli rol oynamaktadır. Çekirdek fonksiyonları iki gruba ayrılmaktadır. Birinci grupta doğrusal çekirdek fonksiyonu olarak doğrusal fonksiyon kullanılmaktadır. İkinci grupta doğrusal olmayan çekirdek fonksiyonu olarak sigmoid, radyan tabanlı fonksiyon (RBF), polinomsal gibi fonksiyonlar kullanılmaktadır [46]. Çoğu uygulamada sınıfların doğrusal çekirdek fonksiyonu ile ayrılması mümkün olmamaktadır. Bundan dolayı doğrusal olmayan çekirdek fonksiyonları kullanılmaktadır (Tablo 1.1) [47]. Tablo 1.1’de verilen fonksiyonlar içerisinde kullanılan parametrelerden a alfa eğimini, c sabit terimi ayırıcısı hiper düzlemin düzgünlüğünü ve verilerin doğru sınıflandırılma dengesini kontrolünü, d polinom derecesini ve γ çekirdek boyutunu ifade etmektedirler [49].

Tablo 1.1. Çekirdek Fonsiyonları

Çekirdek Tipi	Fonksiyon
Doğrusal	$k(x, y) = x^T y + c$
Sigmoid	$k(x, y) = \tanh(ax^T y + c)$
RBF	$k(x, y) = \exp(-\gamma \ x - y\ ^2)$
Polinomsal	$k(x, y) = (ax^T y + c)^d$

1.4. ReliefF Boyut Azaltma Yöntemi

ReliefF, Kira ve Rendell tarafından boyut azaltma ve ikili sınıflandırma için geliştirilmiştir [50]. ReliefF modeli veri kümesindeki gürültü direncini artırarak ve eksik verileri göz ardı ederek çok sınıflı problemlerde başarılı şekilde çalışması amacıyla geliştirilmiştir [45, 51]. Boyut azaltma yöntemleri, veri kümesindeki gereksiz niteliklerin kaldırılmasına yardımcı olur. Bu yöntemler verilerin sıkıştırılmasına yardımcı olarak depolama alanından tasarruf sağlar. Ayrıca boyut azaltma işlemi, hesaplama süresini de azaltır ve algoritmaların çalışma süresini kısaltır. Bu yöntemleri kullanmanın dezavantajı, nadiren de olsa önemli veri kaybına neden olmalarıdır [70]. ReliefF aynı sınıftaki örneklerle yakınlığı ve farklı sınıflara uzaklığı ele alan bir model geliştirerek veri kümesindeki önemli nitelikleri tanımlayabilmektedir [52].

ReliefF modeli, farklı niteliklere sahip örnekler arasında birbirine en yakın nitelikleri seçer. Bu modelde, veri kümesi eğitim ve test verileri olarak iki kısma ayrılmıştır. Model, eğitim kümesindeki örneklerden R_i rasgele örnekleri seçer ve seçilen R_i örneğine en yakın komşuları bulmak için aynı ve farklı sınıfların en yakın komşuluğunu hesaplamak üzere Eşitlik 1.6'daki fark fonksiyonu *diff* kullanılır [53].

$$diff(A, I_1, I_2) = \frac{|değer(A, I_1) - değer(A, I_2)|}{mak(A) - min(A)} \quad (1.6)$$

Eşitlik 1.6, örnekler arasındaki en yakın mesafe A ve iki farklı örnek arasındaki mesafe farkı hesaplamak için I_1 ve I_2 kullanılır. Aynı sınıftaki en yakın komşu mesafesi H ve farklı sınıflardaki en yakın komşu mesafesi M seçilir. R_i , M , H ve veri kümesinin ağırlıklandırma vektörünün değerlerine bağlı olarak, bitişik örnek A_f 'nin sınıftaki ve

sınıflar arasındaki mesafesi karşılaştırılır. Karşılaştırma sonucunda, W_f parametresi uzak özelliklere daha az ağırlık verilerek hesaplanır [54]. Bu işlemler her özellik için m kez tekrarlanır ve her özelliğin ağırlık değerleri hesaplanır. Eşitlik 1.7 ağırlıkları güncellemek için kullanılır [45,51].

$$W_{new}[A_f] = W_{old}[A_f] + \frac{\text{diff}(A_f, R_i, M)}{m} - \frac{\text{diff}(A_f, R_i, H)}{m} \quad (1.7)$$

1.5. Tez Kapsamında Kullanılan Veri Kümeleri

Tez çalışmasında, deneysel değerlendirmelerde kullanılan nutrisyonel anemi, DNA mikrodizi, MNIST, fashion MNIST, CIFAR-10, CIFAR-100 ve IMDB movie veri kümeleri hakkında bilgi aşağıda verilmiştir.

1.5.1. Nutrisyonel Anemi Veri Kümesi

Tez çalışmasında kullanılan anemi hastalığı verisi Tokat Gaziosmanpaşa Üniversitesi, Tıp Fakültesi'nden temin edilmiştir. Veri 2013-2018 yılları arasındaki 5 yıllık aralığında 15300 hastanın tam kan sayım test sonuçlarını içermektedir [27]. Hamileler, çocuklar ve kanser hastaları çalışma kapsamının dışında tutulmuştur. Veri kümesindeki gürültüler temizlenmiş ve anemi hastalığı teşhisinde önemli olmadığı uzmanlar yardımı ile belirlenen parametreler veri kümesinden çıkartılmıştır. Tez çalışmasında 24 öznitelik kullanılmıştır (Tablo 1.2). Veri kümesinde 10379 kadın ve 4921 erkek hastanın verileri bulunmaktadır. Veri kümesinde, anemi hasta sayısı 1019 (%7), demir eksikliği anemisi hasta sayısı 4182 (%27), B12 eksikliği anemisi hasta sayısı 199 (%1), folat eksikliği anemisi hasta sayısı 153 (%1) ve anemi olmayan hasta sayısı 9747 (%64) olarak tespit edilmiştir. Veri kümesinde transferin saturasyonu “SDTSD” özniteliği elde edilirken uzman doktor tarafından belirlenen Eşitlik 1.8 kullanılmıştır.

$$SDTSD = \left(\frac{SD}{TSD} \right) \times 100 \quad (1.8)$$

Tablo 1.2. Nutrisyonel anemi veri kümeleri parametreleri

Parametre	Açıklama	Parametre	Açıklama
CİNSİYET	Kadın Erkek Cinsiyet	NE	Nötrofil
WBC	Beyaz Kan Hücreleri	LY	Lenfosit
RBC	Kırmızı Kan Hücreleri	MO	Monositler
HGB	Hemoglobin	EO	Eozinofiller
HCT	Hematokrit	BA	Bazofil Sayısı
MCV	Ortalama Eritrosit Hacmi	SERUM DEMİRİ (SD)	
MCH	Ortalama Hemoglobin Hacmi	TOPLAM SERUM DEMİRİ (TSD)	
MCHC	Kandaki Ortalama Alyuvar Hemoglobin Yoğunluğu	SDTSD	
RDW	Kırmızı Kan Hücreleri Dağılım Genişliği	FERRİTTİN	
PLT	Platelets	FOLAT	
MPV	Ortalama Trombosit Hacmi	B12	
PCT	Kandaki Trombositlerin Kana Oranı		
PDW	Kandaki Trombositlerin Kana Oranı		

1.5.2. Mikrodizi Veri Kümeleri

Tez çalışmasında derin öğrenme yöntemlerinin performanslarını değerlendirmek için Tablo 1.3'de verilen üç genel gen mikrodizi veri kümesi kullanılmıştır [55]. Mikrodizi veri kümeleri, kanser hastalarının sınıflandırılması için biyomedikal alanında toplanmaktadır. Boyut azaltma işlemi, mikrodizi veri kümelerindeki çok sayıda özellik ve az sayıda örnek nedeniyle modellerin eğitim aşamasında bir zorunluluktur.

Tablo 1.3. Mikrodizi veri kümeleri.

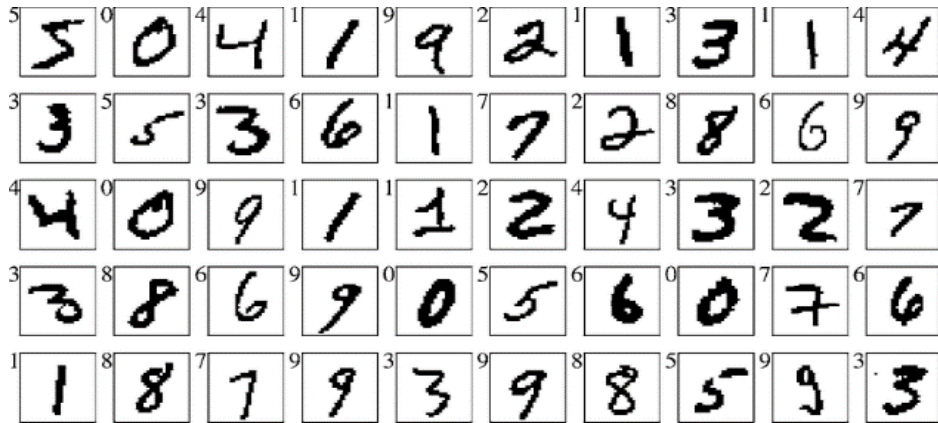
Veri Kümeleri	Gen sayısı	Örnek sayısı	Sınıf sayısı
Merkezi Sinir Sistemi (CNS)	7129	60	2
Lösemi	7129	72	2
Yumurtalık	15154	253	2

Merkezi sinir sistemi (CNS) veri kümesi, embriyonal tümör hastalığı nedeniyle başvuran hastalardan elde edilmiştir [56]. Veri kümesi 60 örnek, 7129 nitelik ve 2 farklı sınıf içermektedir. Veri kümesi hayatta kalan 21 kişiden (sınıf 1 etiketli) ve ölen 39 kişiden (sınıf 0) oluşmaktadır [56]. Çalışmada kullanılan ikinci mikrodizi veri kümesi olan lösemi veri kümesi, 72 hastanın 47'sinde akut lenfoblastik lösemiden (ALL) oluşan 72 örnek, 7129 nitelik ve 2 farklı sınıf ve 25 akut miyeloid lösemi (AML) içermektedir [57]. Yumurtalık kanseri mikrodizi veri kümesi, 91'i kontrol ve 162'si yumurtalık kanseri olan 2 farklı sınıftaki 253 hasta ve 15154 özniteliği içermektedir [58]. Tez çalışmasında kullanılan tüm veri kümelerinin sayısal değerleri 0 ile 1 arasında normalleştirme yapılarak deneylerde kullanılmıştır.

1.5.3. MNIST, Fasion MNIST, CIFAR-10, CIFAR-100 ve IMDB Movie Veri Kümeleri

Bu tez çalışmasında kullanılan MNIST, Fasion MNIST, CIFAR-10, CIFAR-100 ve IMDB Movie veri kümelerinin detayları aşağıda verilmiştir [23-26].

MNIST veri kümesi 0 ile 9 arasında toplamda 70.000 el yazısı rakamlarından oluşmaktadır [23]. Tez çalışmasında, veri kümesinde bulunan görüntülerin 60.000 tanesi eğitim için kullanılırken 10.000 tanesi test amacıyla kullanılmıştır. MNIST veri kümesi 10 sınıfa sahiptir ve her bir resim 28×28 gri ölçekdir. Şekil 1.4'de MNIST veri kümesinin örnekleri görülmektedir.



Şekil 1.4. MNIST veri kümesi örnekleri[23]

Fasion MNIST veri kümesi Zalando araştırma ekibi tarafından MSINT veri kümesine alternatif olarak 70.000 kıyafet resiminden oluşmaktadır [24]. Tez çalışmasında, veri kümesinde bulunan görüntülerin 60.000 tanesi eğitim ve 10.000 tanesi test verisi olarak kullanılmıştır. Kullanılan veri kümesi 10 sınıfa sahiptir ve her bir resim 28×28 gri ölçektir. Şekil 1.5’de Fashion MNIST veri kümesinin örnekleri görülmektedir.



Şekil 1.5. Fashion MNIST veri kümesi örnekleri [24]

CIFAR-10 veri kümesi 60.000 renkli resimlerden meydana gelmektedir [25]. Veri kümesi 10 sınıftan oluşmakta olup her bir resim 32×32 piksel boyutundadır. Tez çalışmasında, verilerin 50.000 tanesi eğitim verisi ve 10.000 tanesi ise test verisi olarak kullanılmıştır. Şekil 1.6’da CIFAR-10 veri kümesinin örnekleri görülmektedir.



Şekil 1.6. CIFAR-10 veri kümesi örnekleri[25]

CIFAR-100 veri kümesi 60.000 renkli resimlerden meydana gelmektedir [25]. Veri kümesi 100 sınıftan oluşmakta olup her bir resim 32×32 piksel boyutundadır. Veri kümesinde 100 sınıf ve her sınıf 600 görüntü içerir. Tez çalışmasında, verilerin 50.000 tanesi eğitim verisi ve 10.000 tanesi ise test verisi olarak kullanılmıştır. Şekil 1.7’de CIFAR-100 veri kümesi görülmektedir.



Şekil 1.7. CIFAR-100 veri kümesi örnekleri [25]

IMDB Movie veri kümesi 50.000 film kaydından meydana gelmektedir. Tez çalışmasında, veri kümesinin 25.000 tanesi eğitim verisi (12.500 pozitif ve 12.500 negatif etiketli) ve 25.000 tanesi test verisi (12.500 pozitif ve 12.500 negatif etiketli) olarak

kullanılmıştır. Ayrıca 1 ile 10 arasındaki ölçeğe filmlere kullanıcı puanları verilmiştir [26]. Şekil 1.8’de IMDB Movie veri kümesinin örnekleri görülmektedir.

Season	Episode #	Title	Rating	# of Votes	Air date	Description	
187	9	23	Finale	9.8	9150	16 May 2013	One year later, Dunder Mifflin employees past ...
137	7	21	Goodbye, Michael	9.8	6796	28 Apr. 2011	As the office gets ready for Michael's final d...
77	5	13	Stress Relief	9.7	6965	1 Feb. 2009	Dwight's too-realistic fire alarm gives Stanle...
186	9	22	A.A.R.M.	9.5	3357	9 May 2013	Dwight prepares for a marriage proposal and hi...
59	4	9	Dinner Party	9.4	4757	10 Apr. 2008	Michael invites Jim and Pam and Andy and Angel...
132	7	16	Threat Level Midnight	9.4	4181	17 Feb. 2011	Michael finally finishes his movie "Threat Lev...
27	2	22	Casino Night	9.4	4139	11 May 2006	The Dunder Mifflin crew holds a casino party a...
94	6	4	Niagara: Part 1	9.4	4010	8 Oct. 2009	The Office travels to Niagara Falls to celebra...
95	6	5	Niagara: Part 2	9.4	2448	8 Oct. 2009	Pam is forced to deal with an embarrassing inj...

Şekil 1.8. IMDB Movie veri kümesi örnekleri [26]

1.6. Literatür Taraması

Literatürde, derin öğrenme mimarilerinde bulunan aktivasyon fonksiyonlarındaki kaybolan gradyan ve negatif bölge problemlerine odaklanılmıştır. Bu tez çalışmasında kaybolan gradyan, negatif bölge, önyargı kayması, uyarlanamama ve esnek olmama problemlerinin üstesinden gelmek amacıyla yeni sabit parametrelili aktivasyon fonksiyonları (RSigELUS ve RSigELUD), eğitilebilir aktivasyon fonksiyonları (P+RSigELUS ve P+RSigELUD) ve çekirdek tabanlı hibrit aktivasyon fonksiyonları (KAF+RSigELUS ve KAF+RSigELUD) önerilmiştir. Ayrıca, derin öğrenme yöntemlerinden ESA ve YOK’un hiper-parametrelerini belirlemek için sezgisel tabanlı optimizasyon yöntemlerinden sonra, literatürde bilinen ReLU ve swish aktivasyon fonksiyonları kullanılarak nutrisyonel anemi ve DNA mikrodizi veri kümeleri kullanılarak sınıflandırma üzerindeki etkileri incelenmiştir.

1.6.1. Veri Madenciliği Yöntemleri ile Anemi Verilerin Sınıflandırılması Üzerine Yapılan Çalışmalar

Bu bölümde, veri madenciliği yöntemleri ile anemi hastalığı verilerinin sınıflandırma doğruluğunu arttırmak için yapılan çalışmalardan bahsedilmiştir. Daha sonra derin

öğrenme yöntemlerinin hiper-parametrelerinin optimizasyonu için literatürde yapılan çalışmalardan bahsedilmiştir.

Literatürde, Jatoi ve arkadaşları tam kan sayımı ile anemi hastalığını veri madenciliği yöntemleri ile tahmin için kullanılmıştır [70]. Yu ve arkadaşları anemi hastalığını tahmin etmek için YSA kullanılmıştır [71]. Abdullah ve Aswari yaptıkları çalışmada, anemi hastalıklarının türlerinin sınıflandırmak için karar ağaçları ve DVM algoritmalarını kullanmışlardır [72]. Hasani ve Hanani yaptığı çalışmada, hemogram test sonuçlarını kullanarak, demir eksikliği anemisi, β -talasemi ve α - talasemi hastalıklarını tahmin etmek için VOTE algoritması kullanılmıştır [73]. Azarkhish ve arkadaşlarının yaptığı çalışmada, demir eksikliği anemisini tespit etmek için uyarlamalı ağ tabanlı bulanık çıkarım sistemi (ANFİS) ve YSA algoritmaları kullanılmıştır [74]. Kaya ve arkadaşlarının yaptığı çalışmada, 18 yaşından küçük çocuklarının anemi hastalıklarını tespit etmek için YSA algoritması kullanılmıştır [75]. Shouval ve arkadaşlarının yaptığı çalışmada, transplantasyon sonucunu ve donör seçimini öngörmek için veri madenciliği tekniklerinden karar ağacı ve YSA algoritmaları kullanılmıştır [76]. Elshami ve Alhalees yaptıkları çalışmada, CBC sonuçlarını veri madenciliği teknikleri yardımıyla talasemi hastalığını tespit etmek için YSA algoritması kullanılmıştır [77]. El-Haleas ve Shurrab yaptıkları çalışmada, CBC verilerinden anemi hastalığını tahmin etmek için derin öğrenme yöntemi H2O algoritması kullanılmıştır [78]. Cil ve arkadaşlarının yaptığı çalışmada, β -talasemi ve demir eksikliği anemisini saptamak için bir aşırı öğrenme makinesi (ELM) algoritması kullanılmıştır [79].

Literatürde hiper-parametre optimizasyonu üzerine yapılan çalışmalardan, Gülcü ve Kuş yaptıkları çalışmada, ESA ve YOK algoritmaları kullanılarak GA, PSO, DGA ve Bayes optimizasyon yöntemleri aracılığıyla seyreltme oranı, öğrenme oranı gibi hiper-parametrelerin optimizasyonunu incelemiştir [80]. Bochinski ve arkadaşlarının yaptığı çalışmada, ESA ve YOK algoritmalarının yapısındaki parametrelerden, filtre, nöron sayıları gibi hiper-parametrelerini optimize etmek için GA kullanılmıştır [81]. Rincon ve arkadaşlarının yaptığı çalışmada, ESA algoritmasının filtre boyutu ve filtre sayısı hiper-parametrelerin optimizasyonu için GA kullanılmıştır [82]. Baldominos ve arkadaşlarının yaptıkları çalışmada, ESA algoritmasında öğrenme oranı, optimizasyon metodu, aktivasyon metodu, yığın boyutu gibi hiper-parametrelerinin belirlenmesi için GA

kullanılmıştır [83]. Silva ve arkadaşları, akciğer kanseri hastalığını sınıflandırmak için ESA mimarisinin filtre ve nöron sayısı hiper-parametrelerinin belirlenmesi için GA'yı kullanmıştır [84]. Yamasaki ve arkadaşları, ESA algoritmasındaki filtre sayısı, filtre boyutu gibi hiper-parametrelerin optimizasyonu için PSO kullanılmıştır [85]. Silva ve arkadaşlarının yaptıkları çalışmada, akciğer kanserini tespit etmek için ESA algoritmasında filtre boyutu, filtre sayısı, yığın boyutu gibi hiper-parametrelerin ayarlanması için PSO kullanılarak hastalık tespiti yapmışlardır [86]. Wang ve arkadaşlarının yaptıkları çalışmada, ESA algoritmasında filtre sayısı, filtre boyutu, nöron sayısı gibi hiper-parametreleri belirlenmesi için PSO kullanılmıştır [87]. Wang ve arkadaşları yaptıkları çalışmada, ESA algoritmasında kullanılan hiper-parametrelerin optimize edilmesi için DGA kullanılmıştır [88]. Sun ve arkadaşlarının yaptıkları çalışmada, YOK algoritması içerisinde kullanılan hiper-parametrelerin optimizasyonu için PSO kullanılmıştır [89]. Li ve Yang yaptıkları çalışmada, rüzgar enerjisi tahmini için YOK ile seyrek bayes regresyon algoritmalarından hibrit bir model oluşturulup, hiper-parametrelerin ayarlanması için YAKA kullanmışlardır [90]. Zhu ve arkadaşlarının yaptıkları çalışmada, MNIST veri kümesini sınıflandırmak için ESA algoritmasında nöron boyutu, öğrenme oranı, filtre sayısı gibi hiper-parametrelerini belirlemek için YAKA kullanmışlardır [91].

1.6.2. Veri Madenciliği Yöntemleri ile Mikrodizi Verilerinin Sınıflandırılması Üzerine Yapılan Çalışmalar

Literatürde mikrodizi veri kümelerinde sınıflandırma doğruluğunu geliştirmek için çeşitli özellik seçimi ve/veya boyut azaltma ve sınıflandırma algoritmaları üzerine çalışmalardan bahsedilmiştir. Çalışmaların çoğunda Meme [93], Kolon [76, 77], Yumurtalık [58], CNS [76, 79] ve Lösemi [92] mikrodizi veri kümeleri kullanılmıştır.

Literatürde, Wang ve arkadaşlarının yaptıkları çalışmada, kolon ve lösemi kanseri mikrodizi verilerini sınıflandırmak için, koşullu karşılıklı bilgiye sahip uyarlanabilir elastik ağ (Adaptive Elastic Net with Conditional Mutual Information) (AEN-CMI) yöntemini önermişlerdir [94]. Bennet ve arkadaşlarının yaptıkları çalışmada, mikrodizi verilerini sınıflandırmak amacıyla ayırık dalgacık dönüşümü (discrete wavelet transform) (DTW) ve hareketli pencere tekniği (moving window technique) (MWT) yaklaşımları ile

en yakın komşu, naif bayes ve DVM algoritmalarına dayanan hibrit bir yaklaşım kullanmışlardır [95]. Sun ve arkadaşları yaptıkları çalışmada, fonksiyonel bağlantılı sinir ağı (functional link neural network) (FLNN) ve özellik seçim yöntemi olarak t-testi ile mikrodizi veri kümelerini sınıflandırmak üzere hibrit bir çalışma önermişlerdir [96]. Kumar ve arkadaşlarının yaptığı çalışmada, mikrodizi veri kümelerinde özellik seçme yöntemi olarak dinamik ağırlıklı FS (datasets using dynamic weighted FS) (DWFS) ve KNN algoritması ile hibrit yöntem kullanılmıştır [97]. Diaz ve arkadaşlarının yaptıkları çalışmada, mikrodizi verilerini sınıflandırmak için rastgele orman algoritmasını ve Ludwig ve arkadaşları gen ekspresyon verilerini sınıflandırmak için bulanık bir karar ağacı algoritması kullanmışlardır [98, 99]. Medjahed ve arkadaşlarının yaptıkları çalışmada, mikrodizi verilerinden özellik seçme işlemi için destek vektör makineleri ile Yığınlanmış otokodlayıcı (DVM-YOK) elde edilen öznitelikleri sınıflandırma için ikili yusufçuk (BDF) algoritması kullanılmıştır [100]. Liu ve arkadaşlarının yaptıkları çalışmada, üç mikrodizi veri kümesini sınıflandırmak için ESA ve YOK hibrit modelini kullanılmıştır [101]. Rani ve arkadaşlarının yaptıkları çalışmada, kolon, akciğer ve yumurtalık mikrodizi veri kümelerini sınıflandırma üzerine iki aşamalı MI-GA ve DVM yöntemi kullanılmıştır [102]. Zeebaree ve arkadaşları, mikrodizi verilerini sınıflandırma için ESA kullanmıştır [103]. Kılıçarslan ve arkadaşlarının yaptıkları çalışmada, mikrodizi prostat kanseri veri kümesinde PSO ve DVM hibrit yöntemi kullanılmıştır [104]. Thara ve arkadaşları yaptıkları çalışmada, gastrik kanser veri kümesini sınıflandırmak için sürüye dayalı zeka temelli özellik seçim yöntemi kullanılmıştır [105]. Amirreza Rouhi ve arkadaşları yaptıkları çalışmada, bilinen beş mikrodizi veri kümesinin sınıflandırma için çift karınca kolonisi ile naive bayes sınıflandırma algoritmaları hibrit kullanılmıştır [106]. Venkataramana ve arkadaşlarının yaptıkları çalışmada, lösemi kanseri mikrodizi veri kümesi sınıflandırmak için karma bir özellik seçme yöntemini kullanmışlardır [107]. Jain ve arkadaşları korelasyon tabanlı ve ikili PSO ile özellik seçimi gerçekleştirildikten sonra, naif bayes algoritması ile sınıflandırılması gerçekleştirilmiştir [108].

1.6.3. Sabit ve Eğitilebilir Aktivasyon Fonksiyonları Üzerine Yapılan Çalışmalar

Bu bölümde, derin öğrenme algoritmalarında meydana gelen kaybolan gradyan ve negatif bölge problemlerine çözüm üretebilmek amacıyla literatürde önerilen sabit ve eğitilebilir aktivasyon fonksiyonlarının çalışmalarından bahsedilmiştir.

Literatürde ReLU [7] aktivasyon fonksiyonu kaybolan gradyan problemiyle başa çıkmak için önerilmiştir. Ayrıca ReLU, girişe negatif değer geldiğinde herhangi bir değişiklik gerçekleştirilmeden pozitif değer döndürerek kaybolan gradyan probleminin üstesinden gelmektedir (Eşitlik 1.9). Ancak ReLU aktivasyon fonksiyonunun dezavantajı, negatif bölgeye gelen değeri sıfıra döndürmesidir. Ayrıca, ReLU aktivasyon fonksiyonunun sadece pozitif bölgede çalıştığından dolayı negatif giriş değeri olan nöronların ölmesine neden olmaktadır. ReLU’da meydana gelen negatif değer probleminin üstesinden gelmek için Leaky ReLU (LReLU) [3], SELU [8], ELU [9], PReLU [10], GELU [59], DReLU [60] vb. aktivasyon fonksiyonları önerilmiştir. Maas ve arkadaşları tarafından önerilen LReLU aktivasyon fonksiyonu (Eşitlik 1.10), nöronun giriş değerinin negatif olması durumunda sabit bir α katsayısı ile (0.01 gibi) çarpılır. Ayrıca LReLU’nun, ReLU ve sigmoid fonksiyonlarında var olan kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmektedir. PReLU aktivasyon fonksiyonu (Eşitlik 1.11), LReLU aktivasyon fonksiyonundaki α parametre değerini kullanmak yerine, veriye göre eğitilebilir parametre kullanılarak kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmektedir [10]. Clevert ve arkadaşlarının önerdikleri ELU aktivasyon fonksiyonunda (Eşitlik 1.12) üstel yapı kullanılarak, negatif değer ve kaybolan gradyan problemlerinin üstesinden gelinmiştir [9]. Trottier ve arkadaşlarının geliştirmiş olduğu PELU aktivasyon fonksiyonu (Eşitlik 1.13), ELU fonksiyonundaki üstel yapının ayarlanması için geliştirilmiştir [13]. Klambauer ve arkadaşları tarafından önerilen SELU aktivasyon fonksiyonu (Eşitlik 1.14), gradyan kayma, negatif değer ve ELU’nun kötü başlangıç değeri atama problemlerinin üstesinden gelinmesi için önerilmiştir. SELU aktivasyon fonksiyonu, ELU aktivasyonundan farklı olarak, sabit bir β parametresi eklenerek bahsedilen problemlerin üstesinden gelebilmektedir [8]. Godin ve arkadaşları tarafından önerilen DReLU aktivasyon fonksiyonunu (Eşitlik 1.15), derin modelleri eğitmek için geliştirilmiştir. DReLU aktivasyon fonksiyonu, ELU aktivasyon fonksiyonuna benzer şekilde hem pozitif hem de negatif bölgede çalışabilmektedir.

Ayrıca DReLU aktivasyon fonksiyonu, hem ReLU hem de hiperbolik tanjant aktivasyon fonksiyonlarının özelliklerini sergilemekte ve kaybolan gradyan probleminin üstesinden gelebilmektedir [60]. Qiumei ve arkadaşları tarafından yapılan çalışmada ELU ve RELU aktivasyon fonksiyonlarının sınıflandırma doğruluğunu ve hesaplama hızlarını arttırmak amacıyla FELU aktivasyon fonksiyonunu önerilmiştir. Önerilen FELU aktivasyon fonksiyonu (Eşitlik 1.16), kaybolan gradyan probleminin üstesinden gelebilmektedir. Ayrıca FELU, negatif bölgede hesaplamayı hızlandırabilmek için üstel bir fonksiyon kullanılarak ağırlık hızlı şekilde çalışmasını sağlamaktadır [61]. Ramachandran ve arkadaşları tarafından önerilen Swish aktivasyon fonksiyonu, ReLU ve sigmoid aktivasyon fonksiyonlarını barındırmasıyla kaybolan gradyan probleminin üstesinden gelinmektedir [62]. Eşitlik 1.9-1.16’de gösterilen parametrelerden x giriş verisi, α, β sabit öğrenme parametresi, Eşitlik 1.13 ve 1.16’de bozulma b ve a doyma noktasını düşürülebilen $a, b > 0$ değerlerdir.

$$f(x) = \begin{cases} 0, & \text{eğer } x < 0 \\ x, & \text{eğer } x \geq 0 \end{cases} \quad (1.9)$$

$$f(x) = \begin{cases} 0.01x, & \text{eğer } x < 0 \\ x, & \text{eğer } x \geq 0 \end{cases} \quad (1.10)$$

$$f(x) = \begin{cases} \alpha x, & \text{eğer } x < 0 \\ x, & \text{eğer } x \geq 0 \end{cases} \quad (1.11)$$

$$f(x) = \begin{cases} \alpha(e^x - 1), & \text{eğer } x < 0 \\ x, & \text{eğer } x \geq 0 \end{cases} \quad (1.12)$$

$$f(x) = \begin{cases} \alpha(e^{\frac{x}{b}} - 1), & \text{eğer } x < 0 \\ \frac{a}{b}x, & \text{eğer } x \geq 0 \end{cases} \quad a, b > 0 \quad (1.13)$$

$$f(x) = \beta \begin{cases} \alpha e^x - \alpha, & \text{eğer } x < 0 \\ x, & \text{eğer } x \geq 0 \end{cases} \quad (1.14)$$

$$f(x, y) = \begin{cases} 0, & \text{eğer } x \leq 0 \text{ ve } y \leq 0 \\ x, & \text{eğer } x > 0 \text{ ve } y \leq 0 \\ -y, & \text{eğer } x \leq 0 \text{ ve } y > 0 \\ x - y, & \text{eğer } x > 0 \text{ ve } y > 0 \end{cases} \quad (1.15)$$

$$f(x) = \begin{cases} x, & \text{eğer } x > 0 \\ a(2^{x/\ln(2)} - 1), & \text{eğer } x \leq 0 \end{cases} \quad (1.16)$$

Kaybolan gradyan problemini çözmek için önerilen ReLTanh [63] aktivasyon fonksiyonu (Eşitlik 1.17), hiperbolik tanjant aktivasyon fonksiyonunun pozitif ve negatif bölgelerinin birbirleriyle değiştirilmesi ve türevlerinin alınması ile her iki bölgede de öğrenme sağlanmaktadır. Kong ve arkadaşları tarafından önerilen Hexpo aktivasyon fonksiyonu (Eşitlik 1.18), kaybolan gradyan problemini çözmek için önerilmiştir [64]. Bawa ve Kumar tarafından önerilen sabit LISA ve eğitilebilir ALISA aktivasyon fonksiyonları (Eşitlik 1.19), kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelmesi için önerilmiştir [12]. Sharma tarafından önerilen ReS aktivasyon fonksiyonu (Eşitlik 1.20), kaybolan gradyan probleminin üstesinden gelmek için önerilmiştir [65]. Softplus aktivasyon fonksiyonu (Eşitlik 1.21), sigmoid ve hiperbolik tanjant aktivasyon fonksiyonlarının yerine alternatif olarak Liu ve arkadaşları tarafından önerilmiştir [66]. Glorot ve Bengio tarafından önerilen softsign aktivasyon fonksiyonu (Eşitlik 1.22), sigmoid ve tanjant aktivasyon fonksiyonlarının türev karmaşıklığını azaltmak için önerilmiştir [34]. Yun tarafından geliştirilen eğitilebilir sigmoidal aktivasyon fonksiyonu kaybolan gradyan probleminin üstesinden gelmek için önerilmiştir [30]. Önerilen bu aktivasyon fonksiyonu sürekli fonksiyonların yerine kesintili fonksiyon kullanılmaktadır. Kesintili fonksiyonları, sürekli fonksiyonlara göre negatif bölge problemde daha etkilidir [30]. Chieng ve arkadaşları tarafından geliştirilen Parametric Flatten-T Swish (PFTS) aktivasyon fonksiyonu (Eşitlik 1.23), negatif bölge probleminden kaynaklı olarak önyargı kayması (bias shift) ve esnek olmama problemlerinin üstesinden gelmesi için önerilmiştir [14]. Godfrey tarafından geliştirilen parametrik bükülebilir doğrusal birim (BLU) aktivasyon fonksiyonunu, PReLU ve PELU aktivasyon fonksiyonlarına alternatif olarak sınıflandırma doğruluğunu arttırmak amacıyla önerilmiştir [68]. Eşitlik 1.17-1.23'de kullanılan x giriş verisi, a_1 , a_2 ve t sabit öğrenme katsayısı, $Tanh'$ tanjant hiperbolik birinci türevini a , b , c ve $d > 0$ sabit katsayıları ifade etmektedir

$$f(x) = \begin{cases} \text{Tanh}'(\lambda^+)(x - \lambda^+) + \text{Tanh}(\lambda^+), & x \geq \lambda^+ \\ \text{Tanh}(x), & \lambda^- < x < \lambda^+ \\ \text{Tanh}'(\lambda^-)(x - \lambda^-) + \text{Tanh}(\lambda^-), & x \leq \lambda^- \end{cases} \quad (1.17)$$

$$f(x) = \begin{cases} -a \left(e^{\frac{-x}{b}} - 1 \right), & \text{eğer } x \geq 0 \\ c \left(e^{\frac{x}{d}} - 1 \right), & \text{diğerleri} \end{cases} \quad (1.18)$$

$$f(x) = \begin{cases} a_1 x - a_1 + 1, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ a_2 x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (1.19)$$

$$f(x) = \begin{cases} 0, & \text{eğer } x \leq 0 \\ \frac{x}{1 + e^{-x}}, & \text{eğer } x > 0 \end{cases} \quad (1.20)$$

$$f(x) = \ln(1 + e^x) \quad (1.21)$$

$$f(x) = \frac{x}{(1 + |x|)} \quad (1.22)$$

$$FTS = \begin{cases} \frac{x}{1 + e^x} + t & \text{eğer } x > 0 \\ t & \text{eğer } x < 0 \end{cases} \quad (1.23)$$

1.6.4. Parametrik Olmayan Aktivasyon Fonksiyonları Üzerine Yapılan Çalışmalar

Bu bölümde, derin öğrenme algoritmalarında kullanılmak üzere literatürde önerilen parametrik olmayan aktivasyon fonksiyonları çalışmalarından bahsedilmiştir.

Literatürde, Agostinelli ve arkadaşlarının yaptıkları çalışmada, uyarlanabilir aktivasyon fonksiyonu (piecewise linear activation - APL) önerilmiştir (Eşitlik 1.24) [31]. APL, parçalı doğrusal aktivasyon fonksiyonu ile gradyan iniş fonksiyonunu kullanılarak her nöron üzerinde bağımsız öğrenebilme özelliğine sahiptir. Scardapanea ve arkadaşları tarafından çekirdek tabanlı aktivasyon fonksiyonu (kernel based activation function - KAF) önerilmiştir (Eşitlik 1.25) [17]. KAF, çekirdek tabanlı fonksiyonlardan (gauss,

polinom vb.) yararlanarak öğrenme işleminde yüksek doğruluk oranı elde edilmesi için kullanılır. Ancak parametrik olmayan aktivasyon fonksiyonu KAF, öğrenme sürecinde parametrelerde negatif bölgeden kaynaklanan önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama problemleri ile karşı karşıya kalmaktadır.

$$g(s) = \max\{0, s\} + \sum_{i=1}^s a_i \max\{0, -s + b_i\} \quad (1.24)$$

$$KAF = \sum_{i=1}^D a_i K(s, d_i) \quad (1.25)$$

Bu tez çalışmasında, literatürde var olan aktivasyon fonksiyonlarının eksiklikleri gidermek ve tahmin doğruluklarını arttırmak için yeni sabit parametrelili RSigELU aktivasyon fonksiyonları, çekirdek tabanlı hibrit KAF+RSigELU aktivasyon fonksiyonları ve eğitilebilir P+RSigELU aktivasyon fonksiyonları önerilmiştir.

2. BÖLÜM

RELU VE SWİSH AKTİVASYON FONKSİYONLU HİBRİT DERİN ÖĞRENME YÖNTEMLERİ KULLANARAK ANEMİ VERİLERİN SINIFLANDIRILMASI

2.1. Giriş

Derin öğrenme mimarilerinde öğrenme işleminin başlatılabilmesi için en yaygın problem hiper-parametrelerin belirlenmesidir [114]. Çünkü derin öğrenme mimarilerinde iyi belirlenmiş hiper-parametre değerleri olmadan başarı performansı istenilen düzeyde olmayacaktır [115]. Bundan dolayıda, optimize edilmeyen hiper-parametreler uygulamacıları tarafından el ile değiştirilir ve en ideal hiper-parametre değeri bulunana kadar denemeler devam eder [116]. Günümüz teknolojisinin ilerlemesiyle, kullanılan derin öğrenme yöntemlerinin hesaplama maliyetinin düşürülmesi ve başarı oranının yükseltilmesi için hiper-parametre optimizasyonu yapılmaktadır. Ancak her problemin yapısı aynı olmadığından dolayı ortak derin öğrenme ve ortak hiper-parametre değerlerinin kullanılması zordur [172].

Derin öğrenme modellerinin tasarım aşamasında iki türlü hiper-parametre optimizasyonu gerçekleşmektedir. Birincisi eğitim işlemi sırasında veri üzerinden elde edilen parametreler ile modeli oluşturanlar tarafından deneme yanılma veya rastgele arama yöntemleriyle hiper-parametreler belirlenmektedir [117]. İkincisi, derin öğrenme yöntemlerinde, eğitim işlemi sırasında hiper-parametrelerin optimizasyonu sezgisel tabanlı algoritmalar kullanılmaktadır. Sezgisel tabanlı optimizasyon algoritmaları, rastgele ve olasılık temelli algoritmalara göre hem hızlı, hem de rastgele arama alanlarında daha geniş arama yapabilmektedir. Bu durumda sezgisel tabanlı algoritmalar tarafından belirlenen hiper-parametreler ile derin öğrenme modelinin yerel ve global en

iyi çözüm uzayına ulaşmasına imkan tanır. Sezgisel tabanlı optimizasyon algoritmalarında aramaya tek bir nokta üzerine değil bir nokta ailesinden başlar ve yerel optimuma takılmadan optimizasyon işlevini yerine getirir. Hiper-parametreler kullanılan algoritmalara göre, öğrenme oranı, momentum oranı, filtre sayısı, filtre boyutu gibi parametrelerdir [118]. Tez çalışmasında, nutrisyonel anemi veri kümelerininin sınıflandırılması için kullanılan derin öğrenme yöntemlerinden ESA ve YOK algoritmalarının hiper-parametreleri belirlenerek, yüksek başarı oranları elde edilebilmesi için sezgisel tabanlı algoritmalar kullanılmıştır [119]. Bu tez çalışmasında, sezgisel tabanlı algoritmalarından parçacık sürü optimizasyon algoritması (PSO) [18], genetik algoritma (GA) [19], diferansiyel gelişim algoritması (DGA) [20] ve yapay arı kolonisi algoritması (YAKA) [21] kullanılmıştır. Hiper-parametreleri belirlenen derin öğrenme algoritmaları, literatürde bilinen ReLU ve swish aktivasyon fonksiyonları ile birlikte kullanılarak anemi hastalığının sınıflandırılması üzerindeki etkisi incelenmiştir.

Anemi hastalığı, günlük kullanım terimi ile “kansızlık”, vücudumuzdaki kanın içindeki hemoglobin miktarının cinsiyete, beslenmeye ve benzeri durumlara göre belli değerlerin altında kalmasıyla anemi meydana gelmektedir. World Health Organization'un (WHO) raporuna göre dünyadaki nüfusun yaklaşık %20'sinde anemi hastalığı bulunmaktadır. Anemi, bir başka ifadeyle kandaki alyuvar sayısındaki azalma olarak ifade edilir. Kandaki alyuvar sayısının azalması, hemoglobin miktarının düşmesiyle ortaya çıkmaktadır. Anemiyi tespit edebilmek için tam kan sayımı (hemogram) testi yapılarak hemoglobin yoğunluğuna bakılmaktadır. Hemoglobin değeri, Blanc'a göre kadınlarda <12 gr/dl, hamilelerde <11 gr/dl ve erkeklerde <13 gr/dl, demir değeri <20 µg/L, ferritin değeri <15 µg/L, B12 değeri <200 pg/mL ve folat değeri <4 ng/mL altında ise nutrisyonel anemidir [109].

Nutrisyonel anemi, demir eksikliği anemisi, B12 eksikliği anemisi ve folat eksikliği anemisi olarak üçe ayrılabilir. Demir eksikliği anemisi, erkeklerde ve kadınlarda gastrointestinal kanamalardan dolayı en sık görülen anemidir. Ayrıca ihtiyaç duyulan demir oranının yiyeceklerden alınamaması, bağırsaklardan emiliminin az olmasıyla ve demir eksikliğin eritropoyezi azaltması sonucu meydana gelmektedir. Demir eksikliği anemisinin görülme sıklığı, Avrupa'da %3, Amerika'da %4, Afrika'da %29, Asya'da %40 civarındadır [109]. 2015 yılında yapılan araştırmalara göre, anemiden kaynaklı yaklaşık

54.000 ölüm vakası meydana gelmiştir [110]. Demir eksikliği anemisini erken teşhis edilemezse, baş ağrısı, ellerde ve ayaklarda uyuşma, sinirlilik hali, saçların dökülmesi, çarpıntı, huzursuz, bacak sendromu gibi belirtiler görülmektedir. B12 eksikliği anemisi (pernisisyöz), özellikle hayvansal proteinlerin eksik alınımından dolayı az gelişmiş ülkelerde sık görülmektedir. B12 eksikliği anemisinin ana nedeni, ince bağırsakta ilgili vitaminin emiliminin yapılamamasından kaynaklanmaktadır. B12 eksikliği anemi hastalığı sinsi olarak başlayarak, hastalarda baş dönmesi, kişilerde nörolojik sorunlar, pankreas hastalıkları ve bazı kanser hastalıklarıyla karşımıza çıkarabilmektedir. Folat anemisi ise vücuttaki folik asit eksikliğinden dolayı kan hücrelerinin üretimini azaltır. Eğer folik asit eksikliği ve kan hücre üretiminin azlığı tespit edilemezse, ishal, diş eti iltihabı, depresyon, unutkanlık ve sinir sistemi rahatsızlıklarına sebebiyet verecektir. Doğru tanı koyulması ve etyolojik faktörün saptanması için gerekli çalışmalar yapılmalıdır. Doğru ilaç ile yeterli sürede tedavi edilmesi gerektiği akıldan çıkarılmamalıdır [111,112]. HGB anemisi, vücutta organlara ve dokulara yeterli miktarda oksijen taşımak için gerekli olan sağlıklı kırmızı kan hücreleri olmadığında ortaya çıkar. Kırmızı kan hücrelerinin sayısı azaldığında HGB-anemi oluşur. Genellikle vücudun sağlıklı alyuvar üretme yeteneği diğer hastalıklardan etkilendiğinde veya kırmızı kan hücresi yıkımı veya kaybı anormal şekilde arttığında ortaya çıkar. HGB-anemi hastalığı yorgunluk ve enerji kaybı, nefes darlığı, göğüs ağrısı, uykusuzluk ve bacaklarda kramplar gibi semptomlarla kendini gösterebilir [113]. Yapılan araştırmalarda nutrisyonel anemilerinin tedavilerinin gerçekleştirilebilmesi için erken teşhisin önemli bir etken olduğu görülmüştür.

Tez çalışmasında, HGB anemi, nutrisyonel anemiler (demir eksikliği, B12 eksikliği ve folat eksikliği) ve anemi olmayan hastaların tahmini ve sınıflandırılması için PSO [18], GA [19], DGA [20] ve YAKA [21] algoritmaları ile YOK ve ESA derin öğrenme algoritmalarını kullanan hibrit modeller önerilmiştir. Ayrıca, literatürde bilinen ReLU [7] ve swish [62] aktivasyon fonksiyonları kullanılarak performans karşılaştırılması yapılmıştır. Sezgisel tabanlı optimizasyon algoritmalarıyla, seyreltme (dropout) değeri, öğrenme oranı, momentum değeri, l2 regülasyon değeri, optimizasyon yöntemi, çıkış aktivasyon yöntemi, yığın boyutu ve tekrar değerlerin optimizasyonu gerçekleştirilmiştir. Optimum hiper-parametre değerleri belirlendikten sonra, ESA için en iyi başarı PSO ve

DGA’da, YOK için en iyi başarı GA’da gözlemlenmiştir. Literatürde, yaygın olarak ESA hiper-parametrelerinin optimizasyonu üzerine çalışmalar görülmektedir. Yapılan tez çalışmasında, ESA ve YOK her iki algoritmanında hiper-parametrelerinin optimizasyonu ile literatürde bilinen ReLU ve swish aktivasyon fonksiyonlarının anemi hastalığı üzerindeki performans değerlendirmesi yapılmıştır. Yapılan çalışmanın sonuçları derlenerek iki yayın sunulmuştur. Yayınlardan birisi bir dergide yayınlanmıştır [27] ve diğeri ise bir dergide revizyon aşamasındadır..

2.2. Yöntem

Bu tez çalışmasında, anemi ve nutrisyonel anemi hastalığına ait verileri kullanarak derin öğrenme yöntemlerinin hiper-parametrelerinin optimizasyonu ve literatürde bilinen ReLU ve swish aktivasyon fonksiyonları kullanarak sınıflandırma işlemi gerçekleştirilmiştir. Kullanılan yöntemler ESA ve YOK hakkında detaylı bilgiler Bölüm 1.1 ve 1.2’de verilmiştir. ESA ve YOK hiper-parametreleri Tablo 2.2, 2.3 ve 2.4’de gösterilmiştir. Tez çalışmasında kullanılan DGA, PSO, YAKA ve GA optimizasyon algoritmalarının kontrol parametreleri için Tablo 2.5-2.8’deki gibi deneysel çalışmalar gerçekleştirilerek belirlenmiştir. Kullanılan sezgisel tabanlı algoritmalar her zaman en iyi çözümü elde etmek üzere çalıştırılmaktadır. Bunun yanında sezgisel algoritmalar bütün problemlerde aynı başarıyı vermemektedir. Bundan dolayı, hangi algoritmanın en iyi başarı oranını elde ettiğini belirlemek için deneyler gerçekleştirilmiştir. Optimizasyon algoritmalarının kontrol parametreleri geliştiricileri tarafından varsayılan (default) değer olarak kullanılmaktadır. Ancak belirlenen bu değerler her bir problem için uygun olmadığından dolayı, bu tez çalışmasındaki problem için en ideal kontrol parametrelerinin belirlenmesi önem arz etmektedir. Yapılan deneylerde, optimizasyon algoritmalarının kontrol parameteleri anemi veri kümesine üzerinde, YOK ve ESA mimarileri kullanılarak en iyi başarı sonucu veren değerlerden elde edilmiştir. Belirlenen kontrol parametrelerinden en iyi sonucu veren değerler ile ESA ve YOK’un hiper-parametrelerinin daha gerçekçi şekilde tespit edilmesi sağlanmıştır. Algoritma içindeki aktivasyon ve optimizasyon fonksiyonları sabit tutulmuştur. Tablo 2.1 hiper-parametre aralıklarını gösterilmektedir.

Tablo 2.1. Hiper-parametrelerin aralıkları

Hiper-Parametreler	Aralık
Seyreltme (Dropout)	[0,1-0,4]
Momentum	[0,9 - 1]
Öğrenme Oranı	[0,003 - 0,1]
L2Regularization(düzenleyici)	[0,003 - 0,1]
Mak-Tekrar	[5, 10, 15, 20]
Eğitim-Tekrar	[50, 75, 100, 125, 150]
Optimizer	[ADAM, SGD, RMSprop, Nadam]
Çıkış Aktivasyon Fonk,	[Sigmoid, Softmax]

Tablo 2.1’de optimize edilecek değerlerden momentum, derin öğrenmede meydana gelebilecek gürültülerin salınımını azaltmak amacıyla kullanılır. Derin öğrenmede öğrenme oranı parametresi modelin yerel optimumda takılmasını engelleyip global optimuma ulaşmasına sağlamak için kullanılır. Diğer bir hiper-parametre olan L2Regularization parametresi, modelin karmaşıklığını azaltmak ve başarı oranını arttırmak için kullanılır. En son olarak da optimize edilen hiper-parametrelerden seyreltme (dropout), modelde aşırı öğrenmenin önüne geçebilmek için modelin her çalışmasında bağlantılarının seyreltilmesi amacıyla kullanılır. Derin öğrenme modellerinde öğrenme işleminin gerçekleşebilmesi ve optimum değeri bulabilmesi için optimizasyon yöntemi kullanılmaktadır. Optimizer, sinir ağının hızlı bir şekilde eğitmesi ve aynı zamanda yerel minimuma takılmasını önlemek amacıyla kullanılır. Bu tez çalışmasında, SGD, ADAM ve RMSprop optimizasyon yöntemleri kullanılmıştır.

Tablo 2.2. ESA modeli hiper-parametreleri

Hiper-Parametreler	Model 1	Model 2	Model 3
Mak-Tekrar	40	35	40
Yığın boyutu	32	16	24
Optimizer	SGD	SGD	ADAM
Çıkış Aktivasyon Fonk.	Softmax	Sigmoid	Softmax
Hiper-Parametreler	Model 4	Model 5	Model 6
Mak-Tekrar	10	20	5
Yığın boyutu	8	16	32
Optimizer	ADAM	RMSprop	RMSprop
Çıkış Aktivasyon Fonk.	Sigmoid	Softmax	Sigmoid

Tablo 2.3. Tez çalışmasında kullanılan ESA mimarisi

	ESA Katmanlar	Parametreler	
Konv_Block_1 Konv	Giriş Verisi		
	Konv1	16@3×3 filtre	Padding = same
	Konv1	16@3×3 filtre	Padding = same
	Yığın Normalizasyon		
	Aktivasyon Fonksiyon		
Konv_Block_2 Konv	Seyreltme		
	Konv2	32@3×3 filtre	Padding = same
	Konv2	32@3×3 filtre	Padding = same
	Yığın Normalizasyon		
	Aktivasyon Fonksiyon		
Konv_Block_3 Konv	Seyreltme		
	Konv3	48@3×3 filtre	Padding = same
	Konv3	48@3×3 filtre	Padding = same
	Yığın Normalizasyon		
	Aktivasyon Fonksiyon		
Konv_Block_4 Konv	Seyreltme		
	Konv4	64@3×3 filtre	Padding =same
	Konv4	64@3×3 filtre	Padding = same
	Konv4	64@3×3 filtre	Padding = same
	Yığın Normalizasyon		
Konv_Block_5 Konv	Aktivasyon Fonksiyon		
	Seyreltme		
	Konv5	96@3×3 filtre	Padding =same
	Konv5	96@3×3 filtre	Padding = same
	Konv5	96@3×3 filtre	Padding = same
	Yığın Normalizasyon		
	Aktivasyon Fonksiyon		
	Seyreltme		
	Tam Bağlantılı Katman	1024	
	Tam Bağlantılı Katman	1024	

Tablo 2.4. YOK modeli hiper-parametreleri ve mimarisi

Hiper-Parametreler	Model 1	Model 2	Model 3
Ae1 Gizli Katman Nöron Sayısı	20	20	20
Ae2 Gizli Katman Nöron Sayısı	10	10	10
Ae3 Gizli Katman Nöron Sayısı	8	8	8
Ae4 Gizli Katman Nöron Sayısı	4	4	4
Eğitim- Tekrar	150	125	125
Yığın Boyutu	32	64	64
Optimizör yöntemi	Nadam	Adam	Adam
Aktivasyon Fonksiyonu	Swish	Swish	ReLU

Tablo 2.5. Farklı çaprazlama ve mutasyon değerine göre GA'nın değerlendirme sonuçları

Çaprazlama	Mutasyon	Doğruluk	F-Skor	Kesinlik	Duyarlılık
0,5	0,1	97,69	97,64	97,62	97,69
0,6	0,1	97,90	97,90	97,89	97,91
0,7	0,1	97,95	97,98	97,92	97,99
0,8	0,1	97,64	97,66	97,62	97,67
0,9	0,1	97,93	97,97	97,91	98,01
0,5	0,2	97,68	97,70	97,64	97,69
0,6	0,2	97,50	97,52	97,45	97,65
0,7	0,2	98,12	98,16	98,10	98,20
0,8	0,2	98,50	98,53	98,48	98,60
0,9	0,2	98,35	98,37	98,32	98,33
0,5	0,3	97,79	97,81	97,75	97,85
0,6	0,3	97,64	97,65	97,60	97,70
0,7	0,3	98,02	98,10	97,98	98,12
0,8	0,3	98,10	98,12	98,02	98,14
0,9	0,3	97,88	97,92	97,80	97,94
0,5	0,5	97,58	97,62	97,56	97,66
0,6	0,5	97,16	97,20	97,10	97,21
0,7	0,5	98,12	98,16	98,05	98,16
0,8	0,5	97,45	97,50	97,40	97,46
0,9	0,5	97,62	97,64	97,58	97,65
0,5	0,7	97,58	97,60	97,54	97,62
0,6	0,7	97,90	97,95	97,92	97,96
0,7	0,7	97,25	97,28	97,22	97,25
0,8	0,7	97,73	97,76	97,70	97,76
0,9	0,7	98,05	98,09	98,03	98,06
0,5	0,9	97,53	97,55	97,50	97,54
0,6	0,9	97,47	97,52	97,40	97,50
0,7	0,9	97,73	97,76	97,70	97,76
0,8	0,9	97,62	97,66	97,60	97,66
0,9	0,9	97,64	97,66	97,62	97,68

Tablo 2.6. Farklı hızlandırma sabitleri ve başlangıç ağırlık sabiti değerine göre PSO değerlendirme sonuçları

Hızlandırma Sabiti ($c1=c2$)	Başlangıç Sabiti (w)	Doğruluk	F-Skor	Kesinlik	Duyarlılık
0,4	0,8	0,9856	0,9854	0,9852	0,9856
0,5	0,8	0,9871	0,987	0,9869	0,9871
0,6	0,8	0,9854	0,9856	0,9852	0,986
0,7	0,8	0,9863	0,9862	0,9861	0,9863
0,8	0,8	0,9834	0,9836	0,9824	0,9849
0,9	0,8	0,9856	0,9856	0,9856	0,9856
0,4	1,0	0,9858	0,986	0,9858	0,9863
0,5	1,0	0,9856	0,9858	0,9856	0,9861
0,6	1,0	0,9854	0,9855	0,9854	0,9856
0,7	1,0	0,9861	0,9859	0,9858	0,9861
0,8	1,0	0,9813	0,9816	0,9806	0,9826
0,9	1,0	0,9852	0,9851	0,9847	0,9854
0,4	1,2	0,9858	0,9859	0,9858	0,9861
0,5	1,2	0,9786	0,9796	0,9784	0,9808
0,6	1,2	0,9854	0,9851	0,9847	0,9854
0,7	1,2	0,9861	0,9859	0,9856	0,9862
0,8	1,2	0,9863	0,9863	0,9863	0,9863
0,9	1,2	0,9861	0,9862	0,9861	0,9863
0,4	1,4	0,9834	0,9837	0,9828	0,9847
0,5	1,4	0,9867	0,9867	0,9867	0,9867
0,6	1,4	0,9819	0,9801	0,9767	0,9836
0,7	1,4	0,9839	0,9839	0,9832	0,9845
0,8	1,4	0,9854	0,9854	0,9854	0,9854
0,9	1,4	0,9760	0,9760	0,9752	0,9769
0,4	1,6	0,9845	0,9843	0,9841	0,9845
0,5	1,6	0,9858	0,9857	0,9856	0,9858
0,6	1,6	0,9852	0,9852	0,985	0,9854
0,7	1,6	0,9819	0,9819	0,9819	0,9819
0,8	1,6	0,9623	0,9626	0,9619	0,9634
0,9	1,6	0,9845	0,9842	0,983	0,9854

Tablo 2.7. Farklı limit parametreleri için YAKA değerlendirme sonuçları

YAKA Limit	Doğruluk	F-Skor	Kesinlik	Duyarlılık
10	0,9848	0,9847	0,9846	0,9848
15	0,9843	0,9841	0,9838	0,9845
20	0,9827	0,9827	0,9824	0,983
25	0,9853	0,9854	0,9851	0,9856
30	0,9854	0,9853	0,9850	0,9856

Tablo 2.8. Farklı ölçekleme faktörü ve çaprazlama oranı için DGA değerlendirme sonuçları

Ölçekleme Faktörü	Çaprazlama Oranı	Doğruluk	F-Skor	Kesinlik	Duyarlılık
0,2	0,15	0,9760	0,9759	0,9758	0,9760
0,4	0,15	0,9612	0,9614	0,9607	0,9620
0,6	0,15	0,9718	0,9718	0,9716	0,9720
0,8	0,15	0,9465	0,9456	0,9394	0,9522
1,0	0,15	0,9747	0,9743	0,9727	0,9759
0,2	0,25	0,9540	0,9539	0,9525	0,9553
0,4	0,25	0,9679	0,9673	0,9657	0,9689
0,6	0,25	0,9749	0,9751	0,9760	0,9743
0,8	0,25	0,9649	0,9651	0,9647	0,9655
1,0	0,25	0,9773	0,9775	0,9768	0,9776
0,2	0,35	0,9511	0,9256	0,8901	0,9654
0,4	0,35	0,9758	0,9756	0,9753	0,976
0,6	0,35	0,9714	0,9718	0,9686	0,9751
0,8	0,35	0,9736	0,9734	0,9729	0,9738
1,0	0,35	0,9688	0,9680	0,9690	0,9670
0,2	0,45	0,9653	0,9632	0,9666	0,9599
0,4	0,45	0,9572	0,9563	0,9572	0,9555
0,6	0,45	0,9549	0,9558	0,9538	0,9579
0,8	0,45	0,9673	0,9668	0,966	0,9676
1,0	0,45	0,9538	0,9518	0,9448	0,9592
0,2	0,55	0,9625	0,9625	0,9627	0,9623
0,4	0,55	0,9762	0,9761	0,9745	0,9779
0,6	0,55	0,9716	0,9718	0,9716	0,9721
0,8	0,55	0,9740	0,9742	0,9745	0,9740
1,0	0,55	0,9409	0,9403	0,9420	0,9386

GA, PSO, DGA ve YAKA'nın kontrol parametrelerinin etkisini analiz etmek için anemi verisi üzerinde ESA yöntemiyle deneysel değerlendirme yapılmıştır. Daha sonra, her 5 çalışma için elde edilen ortalama doğruluk, f-skor, kesinlik, duyarlılık değerlerinin sonuçları verilmiştir. Tablo 2.5-2.8'e kadar olan deneyler ile optimizasyon algoritmaları için en ideal kontrol parametre değeri elde edilmiştir.

Hiper-parametrelerin optimizasyonu için kullanılan algoritmaların anemi veri kümesi üzerinden kontrol parametre değerleri belirlenmiştir. Tablo 2.5'de GA için 0,8 çaprazlama değeri, 0,2 mutasyon değeri ile %98,50 en iyi başarı sonucu elde edilmiştir. GA'nın popülasyon sayısı 10 ve nesil sayısı 5 sabit tutularak deneyler gerçekleştirilmiştir. Tablo 2.6 incelendiğinde, PSO algoritması için kontrol parametresi olan hızlandırma sabiti ve başlangıç ağırlık sabiti değerini belirlemek üzere deneyler gerçekleştirilmiştir. En iyi hızlandırma sabiti 0,5 ve başlangıç ağırlık sabiti 0,8 değeri için %98,71 ile en iyi başarı sonucu elde edilmiştir. PSO'da popülasyon sayısı 20 ve nesil sayısı 5 olarak belirlenerek deneyler gerçekleştirilmiştir. Tablo 2.7 incelendiğinde, YAKA algoritması kontrol parametresi olan farklı limit değerleri üzerine deneyler gerçekleştirilmiştir. Limit değeri 30 için %98,54 ile en iyi başarı sonucu elde edilmiştir. YAKA algoritmasında başlangıç popülasyonu olarak 30 ve nesil sayısı 10 sabit tutularak deneyler gerçekleştirilmiştir. Ayrıca YAKA'da limit değeri büyüdükçe başarı oranının arttığı gözlemlenmiştir. Son olarak DGA algoritmasında kontrol parametresi olarak çaprazlama oranı ve ölçekleme faktörü belirlemek üzerine deneyler yapılmıştır (Tablo 2.8). Gerçekleştirilen deneylerde, ölçekleme faktörü 1,0 ve çaprazlama oranı 0,25 olduğunda %97,73 ile en iyi başarı sonucu elde edilmiştir. Deneylerde 20 popülasyon ve 5 nesil değerleri sabit belirlenmiştir. Tüm optimizasyon algoritmaları kontrol parametrelerinin deneylerinde ESA + ReLU aktivasyon fonksiyonu mimarisi ile anemi veri kümesi kullanılarak gerçekleştirilmiştir. Tablo 2.5-2.8 arasındaki sonuçlar incelendiğinde en iyi sonucun PSO ile elde edildiği gözlemlenmiştir.

2.3. Deneysel Değerlendirme

Derin öğrenme algoritmalarından YOK ve ESA hiper-parametrelerinin değerleri GA, PSO, YAKA ve DGA ile optimizasyon işlemi anemi veri kümesi ile gerçekleştirilmiştir. Deneysel değerlendirmeler için kullanılan anemi veri kümesi Bölüm 1.5.1'de

açıklanmıştır. Elde edilen her bir optimizasyon yöntemine göre derin öğrenme algoritmalarının doğruluk, f-skör, kesinlik ve duyarlılık sonuçları Tablo 2.9-2.17’de gösterilmektedir. Deneylerde, sabit mimariler kullanılarak farklı hiper-parametrelerin değerlerine göre deneyler gerçekleştirilmiştir.

Tablo 2.9. GA ile ESA mimarisinin sonuçları

	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg,	Doğruluk	F-Skor	Kesinlik	Duyarlılık
%70 Eğitim- %30 Test	GA-ESA_Model 1	0,20	0,0017	0,9642	0,0206	0,9843	0,9857	0,9745	0,9812
	GA-ESA_Model 2	0,10	0,0058	0,9846	0,0442	0,9830	0,9833	0,9812	0,9882
	GA-ESA_Model 3	0,15	0,0031	0,9000	0,0314	0,9833	0,9822	0,9838	0,9857
	GA-ESA_Model 4	0,30	0,0015	0,9811	0,0788	0,9813	0,9817	0,9780	0,9865
	GA-ESA_Model 5	0,20	0,0013	0,9808	0,0786	0,9795	0,9801	0,9799	0,9814
	GA-ESA_Model 6	0,25	0,0017	0,9295	0,0080	0,9823	0,9825	0,9824	0,9826
%60 Eğitim- %40 Test	GA-ESA_Model 1	0,25	0,0025	0,9194	0,0150	0,9763	0,9777	0,9725	0,9798
	GA-ESA_Model 2	0,20	0,0073	0,9214	0,0111	0,9795	0,9794	0,9778	0,9797
	GA-ESA_Model 3	0,15	0,0050	0,9221	0,0135	0,9826	0,9820	0,9835	0,9847
	GA-ESA_Model 4	0,10	0,0020	0,9677	0,0369	0,9781	0,9788	0,9701	0,9801
	GA-ESA_Model 5	0,30	0,0020	0,9650	0,0743	0,9732	0,9725	0,9745	0,9778
	GA-ESA_Model 6	0,25	0,0059	0,9146	0,0673	0,9738	0,9737	0,9738	0,9738

Tablo 2.10. GA ile YOK mimarisinin sonuçları

Test	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg,	Doğruluk	F-Skor	Kesinlik	Duyarlılık
%70 Eğitim- %30 Test	GA-YOK_Model 1	0,20	0,0020	0,9641	0,0019	0,9602	0,9600	0,9603	0,9649
	GA-YOK_Model 2	0,25	0,0014	0,9731	0,0014	0,9400	0,9412	0,9314	0,9375
	GA-YOK_Model 3	0,15	0,0025	0,9260	0,0025	0,9390	0,9391	0,9378	0,9402
%60 Eğitim- %40 Test	GA -YOK_Model 1	0,20	0,0020	0,9630	0,0020	0,9555	0,9557	0,9493	0,9600
	GA -YOK_Model 2	0,15	0,0014	0,9268	0,0016	0,9330	0,9338	0,9287	0,9320
	GA -YOK_Model 3	0,10	0,0016	0,9630	0,0020	0,9396	0,9384	0,9101	0,9402

Tablo 2.9 ve 2.10’da GA ile ESA ve YOK modellerinin hiper-parametre deneyleri gözlemlenmektedir. Yapılan deneylerde GA için kontrol parametresi Tablo 2.5’deki en iyi sonucu veren çaprazlama değeri 0,8 ve mutasyon değeri 0,2 kabul edilerek deneyler gerçekleştirilmiştir. Deneyler incelendiğinden ESA için en iyi başarı oranı %70 eğitim ve %30 test yapısında GA_ESA_Model_1 modeliyle %98,43 olarak bulunmuştur.

Tablo 2.11 ve 2.12’de PSO ile yapılan deneylerin sonuçları görülmektedir. Deneylerde, ESA ve YOK modellerinin hiper-parametrelerinin optimizasyonu için PSO kullanılmış olup, Tablo 2.6’daki farklı hızlandırma sabiti (HS) ve başlangıç sabiti (BS) için en iyi başarı HS=0,5 ve BS=0,8 değerleri ile bulunmuştur. Deneysel sonuçlar ESA için incelendiğinde, en iyi başarı oranı %70 eğitim ve %30 test yapısında PSO-ESA_Model_3 modeliyle belirlenen seyreltme oranı 0,25, öğrenme oranı 0,0929, momentum 0,9503 ve LR2 reg. değeri 0,0381 ve en iyi başarı oranı %98,67 olarak bulunmuştur. YOK için en iyi başarı oranı %70 eğitim ve %30 test yapısında PSO-YOK_Model_4 modeliyle %95,93 ile elde edilmiştir. En iyi sonucun elde edildiği seyreltme oranı 0,15, öğrenme oranı 0,001961, momentum 0,98737 LR2 reg. değeri 0,0010 olarak belirlenmiştir.

Tablo 2.13. YAKA ile ESA mimarisinin sonuçları

Test	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg,	Doğruluk	F-Skor	Kesinlik	Duyarlılık
%70 Eğitim- %30 Test	YAKA-ESA_Model 1	0,20	0,0051	0,9022	0,0051	0,9854	0,9853	0,9851	0,9856
	YAKA-ESA_Model 2	0,25	0,0163	0,9138	0,1622	0,9843	0,9838	0,9832	0,9847
	YAKA-ESA_Model 3	0,15	0,0157	0,9131	0,0166	0,9854	0,9853	0,9852	0,9854
	YAKA-ESA_Model 4	0,10	0,0013	0,9891	0,0895	0,9865	0,9863	0,9869	0,9857
	YAKA-ESA_Model 5	0,30	0,0011	0,9030	0,0069	0,9815	0,9815	0,9815	0,9815
	YAKA-ESA_Model 6	0,15	0,0027	0,9128	0,0163	0,9832	0,9833	0,9832	0,9834
%60 Eğitim- %40 Test	YAKA-ESA_Model 1	0,15	0,0078	0,9051	0,0089	0,9822	0,9824	0,9820	0,9828
	YAKA-ESA_Model 2	0,20	0,0181	0,9156	0,0192	0,9722	0,9577	0,9626	0,9535
	YAKA-ESA_Model 3	0,25	0,0025	0,9494	0,0514	0,9835	0,9831	0,9817	0,9846
	YAKA-ESA_Model 4	0,30	0,0026	0,9471	0,0467	0,9735	0,9734	0,9735	0,9732
	YAKA-ESA_Model 5	0,10	0,0024	0,9181	0,0214	0,9817	0,9806	0,9784	0,9828
	YAKA-ESA_Model 6	0,10	0,0025	0,9203	0,0235	0,9565	0,9519	0,9413	0,9631

Tablo 2.14. YAKA ile YOK mimarisinin sonuçları

Test	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg.	Doğruluk	F-Skor	Kesinlik	Duyarlılık
%70 Eğitim- %30 Test	YAKA-YOK_Model 1	0,10	0,0016	0,9694	0,0016	0,9444	0,9461	0,9383	0,9543
	YAKA-YOK_Model 2	0,20	0,0025	0,9376	0,0025	0,9322	0,9296	0,9121	0,9485
	YAKA-YOK_Model 3	0,15	0,0012	0,9815	0,0012	0,9342	0,9332	0,9268	0,9397
	YAKA-YOK_Model 4	0,20	0,0031	0,9206	0,0031	0,9314	0,9315	0,9205	0,9432
%60 Eğitim- %40 Test	YAKA-YOK_Model 1	0,30	0,0013	0,9786	0,0131	0,9374	0,9403	0,9299	0,9513
	YAKA-YOK_Model 2	0,20	0,0022	0,9497	0,0022	0,9342	0,9346	0,9225	0,9471
	YAKA-YOK_Model 3	0,20	0,0013	0,9811	0,0013	0,9317	0,9325	0,9209	0,9446
	YAKA-YOK_Model 4	0,15	0,0017	0,9649	0,0017	0,9562	0,9567	0,9538	0,9597

Tablo 2.13 ve 2.14’de YAKA ile ESA ve YOK üzerine yapılan deneylerin sonuçları görülmektedir. Deneylerde ESA ve YOK modellerinin hiper-parametrelerinin optimizasyonunu için YAKA kullanılmış olup, Tablo 2.7’deki limit kontrol parametre değeri 30 için en iyi başarı bulunmuştur. Deneysel sonuçlar ESA için incelendiğinde, en iyi başarı oranı %70 eğitim ve %30 test yapısında YAKA-ESA_Model_4 modeliyle elde edilmiştir. Elde edilen en iyi sonuç %98,65 değerine seyreltme oranı 0,10, öğrenme oranı 0,0013, momentum 0,9891 ve LR2 reg. değeri 0,0895 olduğunda ulaşılmıştır. YOK için, en iyi başarı oranı %60 eğitim ve %40 test yapısında YAKA-YOK_Model_4 ile %95,62 elde edilmiştir. Bu başarı oranı için seyreltme oranı 0,15, öğrenme oranı 0,0017, momentum 0,9649 ve LR2 reg. değeri 0,0017 olarak belirlenmiştir.

Tablo 2.15. DGA ile ESA mimarisinin sonuçları

Test	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg	Doğruluk	F-Skor	Keskinlik	Duyarlılık
%70 Eğitim- %30 Test	DGA-ESA_Model 1	0,25	0,0264	0,9336	0,0731	0,9867	0,9866	0,9867	0,9865
	DGA-ESA_Model 2	0,20	0,0451	0,9337	0,0866	0,9755	0,9752	0,976	0,9744
	DGA-ESA_Model 3	0,15	0,0929	0,9503	0,0381	0,9856	0,9855	0,9854	0,9856
	DGA-ESA_Model 4	0,20	0,0241	0,9222	0,0099	0,9675	0,9678	0,9666	0,9689
	DGA-ESA_Model 5	0,30	0,0019	0,9280	0,0813	0,9830	0,9831	0,983	0,9832
	DGA-ESA_Model 6	0,20	0,0022	0,9714	0,0780	0,9821	0,9808	0,9797	0,9819
%60 Eğitim- %40 Test	DGA-ESA_Model 1	0,20	0,0082	0,9248	0,0700	0,9556	0,9549	0,9521	0,9578
	DGA-ESA_Model 2	0,15	0,0237	0,9222	0,0099	0,9725	0,9712	0,976	0,9668
	DGA-ESA_Model 3	0,20	0,0082	0,9248	0,0701	0,9636	0,9639	0,9623	0,9656
	DGA-ESA_Model 4	0,10	0,0085	0,9248	0,0699	0,8758	0,8303	0,7603	0,9175
	DGA-ESA_Model 5	0,25	0,0059	0,9822	0,0672	0,9716	0,9714	0,9716	0,9713
	DGA-ESA_Model 6	0,20	0,0013	0,9248	0,0703	0,9807	0,9808	0,9802	0,9814

Tablo 2.16. DGA ile YOK mimarisinin sonuçları

Test	Model	Seyreltme	Öğrenme	Momentum	LR2 Reg	Doğruluk	F-Skor	Keskinlik	Duyarlılık
%70 Eğitim- %30 Test	DGA-YOK_Model 1	0,20	0,0013	0,9776	0,0023	0,9503	0,9498	0,9423	0,9577
	DGA-YOK_Model 2	0,15	0,0019	0,9873	0,0010	0,944	0,9437	0,9375	0,9503
	DGA-YOK_Model 3	0,20	0,0011	0,9373	0,0023	0,9346	0,9351	0,9272	0,9433
	DGA-YOK_Model 4	0,25	0,0013	0,9776	0,0023	0,9466	0,9455	0,9316	0,9617
%60 Eğitim- %40 Test	DGA-YOK_Model 1	0,20	0,0011	0,9373	0,0023	0,9503	0,9491	0,9417	0,9568
	DGA-YOK_Model 2	0,10	0,0011	0,9373	0,0023	0,9356	0,9353	0,9268	0,9442
	DGA-YOK_Model 3	0,30	0,0012	0,9396	0,0026	0,9261	0,9257	0,9124	0,9398
	DGA-YOK_Model 4	0,20	0,0013	0,9776	0,0023	0,9425	0,9415	0,9307	0,9529

Tablo 2.15 ve 2.16’da DGA deneylerinin sonuçları görülmektedir. Deneylerde ESA ve YOK’un hiper-parametrelerinin optimizasyonunu için DGA kullanılmış olup, Tablo 2.8’deki çaprazlama oranı (ÇO) ve ölçekleme faktörü (ÖF) kontrol parametreleri değeri ÇO= 0,25 ve ÖF=1,0 ile en iyi başarı bulunmuştur. Sonuçlar ESA için incelendiğinde, en iyi başarı oranı %70 eğitim ve %30 test ile DGA-ESA_Model_1 modelinde elde edilmiştir. Elde edilen en iyi sonuç %98,67 değerine seyreltme oranı 0,25, öğrenme oranı 0,0264, momentum 0,9336 ve LR2 reg. değeri 0,0731 olduğunda ulaşılmıştır. YOK için, en iyi başarı oranı %70 eğitim ve %30 test yapısında DGA-YOK_Model_1 ile %95,03 elde edilmiştir. Bu başarı oranı için seyreltme oranı 0,20, öğrenme oranı 0,0013,

momentum 0,9776 ve LR2 reg. değeri 0,0023 olarak belirlenmiştir.

Tablo 2.17. Tüm sezgisel yöntemlerin deneysel değerlendirme sonuçları

Model	Doğruluk	F-Skor	Kesinlik	Duyarlılık
GA-ESA	0,9843	0,9857	0,9745	0,9812
GA-YOK	0,9602	0,9600	0,9603	0,9644
PSO-ESA	0,9867	0,9867	0,9865	0,9872
PSO-YOK	0,9593	0,9569	0,9427	0,972
YAKA-ESA	0,9865	0,9863	0,9869	0,9857
YAKA-YOK	0,9562	0,9567	0,9538	0,9597
DGA-ESA	0,9867	0,9866	0,9867	0,9865
DGA-YOK	0,9503	0,9498	0,9423	0,9577

Tez çalışmasında, aynı anda iki farklı derin öğrenme ESA ve YOK algoritmaları üzerinde hiper-parametre optimizasyonu gerçekleştirilmiştir. Gerçekleştirilen deneyler, birden fazla optimizasyon tekniği kullanılması ve kullanılan optimizasyon yöntemlerinde kontrol parametrelerinin optimize edilmesinden dolayı literatürdeki diğer çalışmalara göre bu özelliğiyle ön plana çıkmaktadır. Yapılan deneysel çalışmalarda, derin öğrenme mimarilerinde hiper-parametrelerin optimize edilmemesi modelin öğrenme gücünü olumsuz etkilemektedir. Çünkü, iyi hiper-parametre değerleri olmadan kullanılan derin öğrenme algoritmalarında başarı performansı çok düşük olacaktır [120]. Hiper-parametre optimizasyon yöntemleriyle birlikte ESA ve YOK algoritmaları kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Gerçekleştirilen deneylerde Tablo 2.17’de en iyi hiper-parametre optimizasyonunun yanında, sınıflandırma başarısı olarak ESA algoritması için PSO ile DGA optimizasyon algoritmaları ile %98,67 başarı oranı elde edilmiştir. YOK için gerçekleştirilen deneylerde ise en iyi hiper-parametre optimizasyonunun yanında sınıflandırma başarısı olarak GA ile %97,02 elde edilmiştir. Sezgisel optimizasyon yöntemleri başarı oranı olarak iyi sonuçlar vermesine rağmen hesaplama süresinin fazla olması dezavantajlarının arasındadır.

2.4. Sonuçlar

Bu tez çalışmasında, PSO, GA, DGA ve YAKA algoritmaları ile YOK ve ESA derin öğrenme algoritmalarını kullanan hibrit modelleri anemiye doğru bir şekilde sınıflandırmak ve teşhis etmek üzere geliştirilmiştir. Hiper-parametre optimizasyonu son zamanlarda derin öğrenme yöntemleri için önem arz etmeye başlamıştır. Derin öğrenme mimarilerinde iyi başarının elde edilebilmesi için hiper-parametre optimizasyonu yapılması önemlidir. Çünkü her zaman aynı parametreler ile modellerde istenilen sonuçlar elde edilememektedir. Optimizasyon yapılmış hiper-parametrelili mimarilerin hesaplama maliyeti düşürülerek ve başarı oranları iyileştirilir. Tez çalışmasında, GA, PSO, YAKA ve DGA sezgisel algoritmaları kullanılarak derin öğrenme mimarilerinin hiper-parametreleri optimize edilmeye çalışılmıştır. Elde edilen optimize değerine göre literatürde bilinen ReLU ve swish aktivasyon fonksiyonları ile modelin başarımının nasıl değiştiği incelenmiştir. ESA derin öğrenme modeli için PSO ve DGA'nın en iyi başarı oranını verdiği gözlemlenmiştir. YOK derin öğrenme modeli için GA'nın en iyi başarı oranını verdiği gözlemlenmiştir.

3. BÖLÜM

RELU AKTİVASYON FONKSİYONLU HİBRİT DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK MİKRODİZİ VERİLERİNİN SINIFLANDIRILMASI

3.1. Giriş

Son yıllarda, doku ve hücre örneklerinden mikrodizi verileri üretmenin ve mikrodizi verilerini analiz etmenin önemi artmıştır. Mikrodizi teknolojisi, araştırmacıların binlerce genin ifade düzeylerini aynı anda ölçmelerine imkan verir. Mikrodizi verilerinin sınıflandırılması, hastalık teşhisi veya tümör türlerinin ayırt edilmesi için önemlidir [57]. Ayrıca mikrodizi teknolojisi, normal hücre ve hastalıklı hücre genlerinin modellerini ortaya çıkarmanın etkili ve kolay bir yolu olarak kullanılmaktadır. Mikrodizi teknolojisi, biyoinformatik alanındaki araştırmacılar tarafından geniş çapta incelenmiştir. Mikrodizi teknolojisi, yüzlerce genin eşzamanlı analizini sağlamaktadır. Böylece kansere bağlı hastalıklı genleri ortaya çıkarmak mümkündür. Mikrodizi verileri genellikle 100'den az örneğe sahipken, özellik sayısı 100.000'e kadar çıkabilir. Verilerdeki bu yüksek boyutluluk ve dengesizlik nedeniyle, veri madenciliği yöntemleri mikrodizi verilerinin analizinde çok çekici hale gelmektedir [57].

Mikrodizi verilerinin çeşitli nedenlerle sınıflandırılması zordur. Birincisi, veri kümesinin elde edilme zorluğundan dolayı az sayıda örneklem büyüklüğüne sahiptir. İkincisi, çok sayıda özellik içerir ve bazı özellikler sınıflandırma ile ilgili olmayabilir. Bu nedenle, sınıflandırma doğruluğunu iyileştirmek için boyut azaltma veya özellik seçme algoritmaları gerekmektedir. Üçüncüsü, veride deneysel komplikasyonlardan dolayı gürültü ve değişkenlik içerebilir. Dördüncüsü, mikrodizi verilerinin sınıflandırılması hesaplama açısından karmaşıktır ve dolayısıyla verimli ve hızlı

sınıflandırma algoritmaları gerektirmektedir. Bu sınırlamaların üstesinden gelmek için, literatürde, mikrodizi verilerini analiz etmek için çeşitli özellik seçimi, boyut azaltma ve sınıflandırma yaklaşımları önerilmiştir. Araştırmalar, özellik seçimi, boyut azaltma ve sınıflandırma yaklaşımlarından hiçbirinin tüm mikrodizi veri kümelerinde [121] iyi çalışmadığını ve bu nedenle yeni özellik çıkarma/boyut azaltma ve sınıflandırma yaklaşımlarını keşfetmek için çalışmalar yapılması gerektiğini göstermektedir.

Sınıflandırma için boyut azaltma yöntemleri ve veri madenciliği veya derin öğrenme yöntemleri birlikte kullanılır. Öte yandan yetersiz örnek sayısı ve veri kümesindeki dengesizlik nedeniyle sınıflandırma başarıları istenilen düzeyde değildir. Bu zorluğun üstesinden gelmek için farklı algoritmalar hibrit bir şekilde kullanılmaktadır [122]. Bu tez çalışmasında, ReLU aktivasyon fonksiyonu ile YOK - ESA ve ReliefF - ESA boyut azaltma ve derin öğrenme yaklaşımlarını kullanan iki hibrit model, mikrodizi veri kümelerini kullanarak kanseri teşhis etmek için önerilmiştir. Literatür incelendiğinde önerilen modelleri kullanan mikrodizi veri kümelerine ilişkin herhangi bir çalışma tespit edilmemiştir. Yapılan çalışmanın sonuçları derlenerek bir dergide yayınlanmıştır [22].

3.2. Yöntem

Bu tez çalışmasında, derin öğrenme yöntemleri kullanılarak mikrodizi veri kümesi üzerinden hastalıkları tespit etmek üzerine çalışma gerçekleştirilmiştir. Kullanılan yöntemler ESA, YOK, DVM ve ReliefF hakkında detaylı bilgileri sırasıyla, Bölüm 1.1, 1.2, 1.3 ve 1.4’de verilmiştir.

Deneylerde kullanılan modellerin hiper-parametreleri ve mimarileri Tablo 3.1, 3.2 ve 3.3’de verilmiştir. En iyi hiper-parametre değerleri, deneyler ile belirlenmiştir.

Tablo 3.1. YOK modelinin hiper-parametreleri

Parametreler	Değerler
AE1 gizli nöron	500
AE2 gizli nöron	200
İterasyon	100
L2 Reg	0,01
Sparse Reg	4

Tablo 3.2. ESA modelinin hiper-parametreleri

Hiper-Parametreler	Model
Tekrar	100
Momentum	0,9
Öğrenme Oranı	0,001
Yığın Boyutu	64
Optimizer	SGD
Aktivasyon Fonksiyonu	ReLU
Çıkış Fonksiyonu	Softmax

Tablo 3.1’de, YOK modelinde kullanılan parametreler deneysel değerlendirmeler sonucunda belirlenmiştir. YOK modelinde, ilk otokodlayıcı 500 nöron ve ikinci otokodlayıcı 200 nörondan oluşur. Mikrodizi veri kümelerinin geniş boyutu ve veri kaybını önleme nedeniyle YOK mimarilerinde nöron boyutu büyük kullanılmıştır. Mikrodizi veri kümelerinin [56-58] sınıflandırılması için DVM ve ESA modelleri ayrı ayrı YOK ile hibritlenerek deneysel değerlendirmeler yapılmıştır. DVM modelinde, çekirdek fonksiyonu olarak polinom fonksiyonu kullanılmıştır. Mikrodizi veri kümeleri üzerinde kullanılan ESA modelinin yapısı ve parametreleri Tablo 3.3’de gösterilmiştir.

Tablo 3.3. Çalışmada kullanılan ESA modeli ve parametreleri

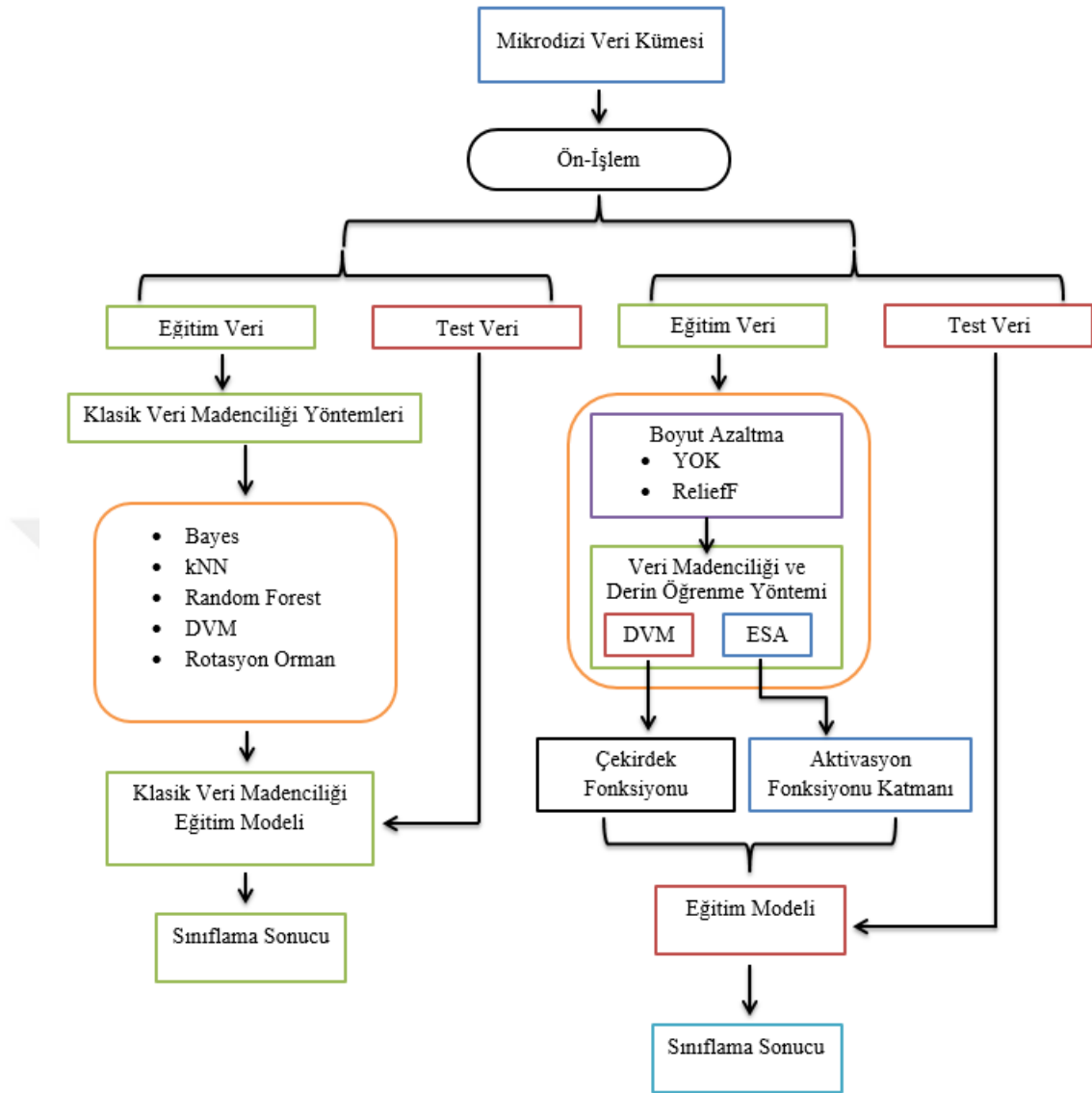
SESA Katmanlar	Parametreler
Giriş Verisi	15154×1×1
Konv1	8@3×3 filtre Adım=2 Padding=0
MakHavuzlama1	Havuz boyutu=2 Adım =2
Konv2	16@3×3 filtre Adım =2 Padding=0
MakHavuzlama2	Havuz boyutu =2 Adım =2
Konv3	32@3×3 filtre Adım =2 Padding=0
MakHavuzlama3	Havuz boyutu =2 Adım =2
Konv4	64@3×3 filtre Adım =1 Padding=1
MakHavuzlama4	Havuz boyutu =2 Adım =2
Tam bağlantı katmanı	2

Tablo 3.3’de, sınıflandırma işlemi için her mikrodizi [56-58] veri kümesi 4 farklı evrişim ve havuzlama katmanlarına tabi tutularak deneyler gerçekleştirilmiştir. Evrişim katmanında filtre boyutları 3×3 olarak ve adım (stride) değerleri 1 veya 2 olarak seçilmiştir. Adım işlemi, katmanlar arasındaki veri kaybını en aza indirmek için

kullanılmıştır. Bu aynı zamanda örnekleme katmanı olan havuzlama katmanı için de geçerlidir. Havuzlama işlemi için havuz boyutu ve adım sayısı da 2 olarak seçilmiştir.

3.3. Önerilen Hibrit Sınıflandırma Yöntemleri

Tez çalışmasında, mikrodizi veri kümelerini sınıflandırmak için ReLU aktivasyon fonksiyonu tabanlı ReliefF-ESA ve YOK-ESA hibrit modelleri önerilmiştir (Şekil 3.1). Modellerde öncelikle veri kümesine ön işlem uygulanmıştır ve eksik kayıtlar silinmiştir. Veri kümesinden anlamlı özellikleri ortaya çıkarmak için boyut azaltma algoritmaları olan ReliefF ve YOK uygulanmıştır. Boyutu azaltılmış veri kümesinin sınıflandırma sonuçlarını iyileştirmek için, ReLU aktivasyon tabanlı ReliefF-ESA modeli ile deneysel değerlendirmeler gerçekleştirilmiştir. Ayrıca ReliefF-ESA modelinde tamamen bağlantılı katman kullanılırken, YOK-ESA için Softmax katmanı kullanılmıştır.



Şekil 3.1. Önerilen hibrit modelin blok diyagramı.

3.4. Deneysel Değerlendirme

Tez çalışmasında, mikrodizi veri kümelerini sınıflandırmak için ReLU aktivasyon tabanlı ReliefF-ESA ve YOK-ESA hibrit modelleri önerilmiştir. Her bir mikrodizi veri kümelerinin sınıflandırılma performanslarını karşılaştırmak için doğruluk, kesinlik ve duyarlılık kriterleri kullanılmıştır. Algoritmaların sonuçları yumurtalık mikrodizi veri kümesi için Tablo 3.4'de, lösemi mikrodizi veri kümesi için Tablo 3.5'de ve CNS mikrodizi veri kümesi Tablo 3.6'da gösterilmektedir.

Tablo 3.4. Yumurtalık mikrodizi veri kümesi için deneysel sonuçları

Boyut Azaltma	Yöntemler	Veri Kümesi	Duyarlılık	Kesinlik	Doğruluk
Düz	DVM	Test1(60-40)	95,22	94,75	95,02
		Test2(70-30)	95,96	95,17	95,65
		Test3(80-20)	96,51	95,86	96,14
		<i>Ortalama</i>	95,89	95,26	95,6
	ESA	Test1(60-40)	95,89	95,34	95,62
		Test2(70-30)	96,32	95,67	95,96
		Test3(80-20)	96,65	96,02	96,32
		<i>Ortalama</i>	96,29	95,68	95,97
YOK	DVM	Test1(60-40)	95,66	95,12	95,41
		Test2(70-30)	96,21	95,49	95,86
		Test3(80-20)	96,45	95,94	96,2
		<i>Ortalama</i>	96,11	95,52	95,82
	ESA	Test1(60-40)	96,05	95,37	95,8
		Test2(70-30)	96,74	96,02	96,31
		Test3(80-20)	97,13	96,49	96,88
		<i>Ortalama</i>	96,64	95,96	96,33
ReliefF	DVM	Test1(60-40)	97,29	96,75	97,05
		Test2(70-30)	97,67	97,01	97,32
		Test3(80-20)	98,46	97,72	98,04
		<i>Ortalama</i>	97,81	97,16	97,47
	ESA	Test1(60-40)	98,75	98,1	98,42
		Test2(70-30)	97,69	97,12	97,37
		Test3(80-20)	100	100	100
		<i>Ortalama</i>	98,81	98,41	98,6

Tablo 3.5. Lösemi mikrodizi veri kümesi için deneysel sonuçları

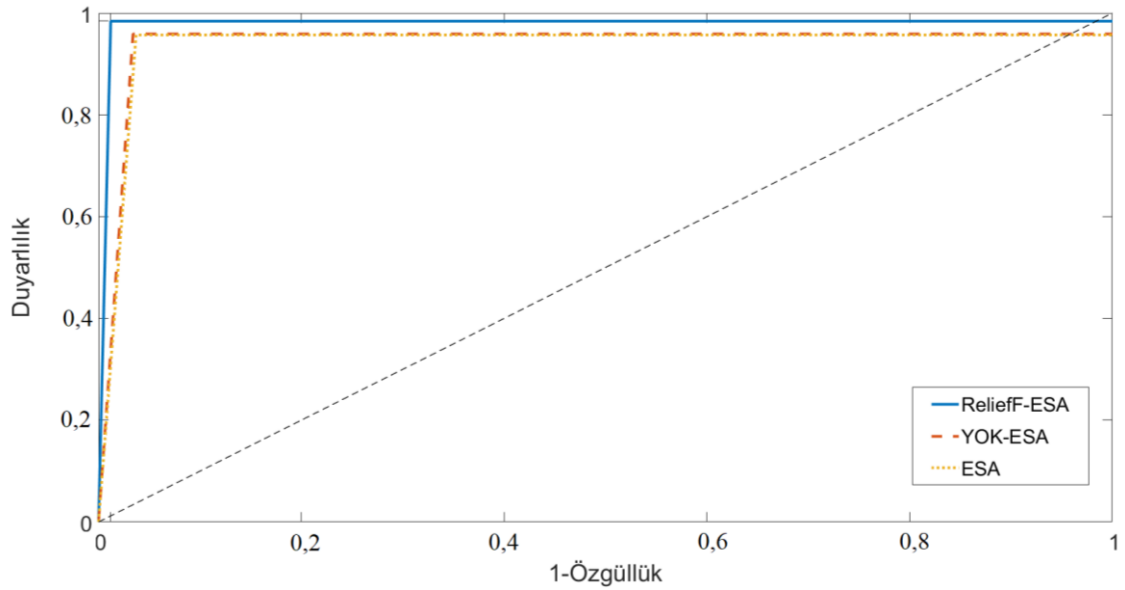
Boyut Azaltma	Yöntemler	Veri Kümesi	Duyarlılık	Kesinlik	Doğruluk
Düz	DVM	Test1(60-40)	94,49	93,76	94,05
		Test2(70-30)	94,61	93,95	94,22
		Test3(80-20)	95,32	94,05	94,83
		<i>Ortalama</i>	<i>94,81</i>	<i>93,92</i>	<i>94,37</i>
	ESA	Test1(60-40)	95,57	94,96	95,24
		Test2(70-30)	95,46	94,87	95,11
		Test3(80-20)	96,12	95,54	95,89
		<i>Ortalama</i>	<i>95,72</i>	<i>95,12</i>	<i>95,41</i>
YOK	DVM	Test1(60-40)	91,05	88,5	89,3
		Test2(70-30)	94,24	92,43	93,32
		Test3(80-20)	95,33	93,52	94,64
		<i>Ortalama</i>	<i>93,54</i>	<i>91,48</i>	<i>92,42</i>
	ESA	Test1(60-40)	93,64	91,36	92,25
		Test2(70-30)	95,76	93,78	94,87
		Test3(80-20)	97,56	95,43	96,37
		<i>Ortalama</i>	<i>95,65</i>	<i>93,52</i>	<i>94,5</i>
ReliefF	DVM	Test1(60-40)	97,86	97,27	97,62
		Test2(70-30)	98,21	97,74	97,95
		Test3(80-20)	98,83	98,36	98,6
		<i>Ortalama</i>	<i>98,3</i>	<i>97,79</i>	<i>98,06</i>
	ESA	Test1(60-40)	99,74	99,41	99,58
		Test2(70-30)	100	100	100
		Test3(80-20)	100	100	100
		<i>Ortalama</i>	<i>99,91</i>	<i>99,8</i>	<i>99,86</i>

Tablo 3.6. CNS mikrodizi veri kümesi için deneysel sonuçları

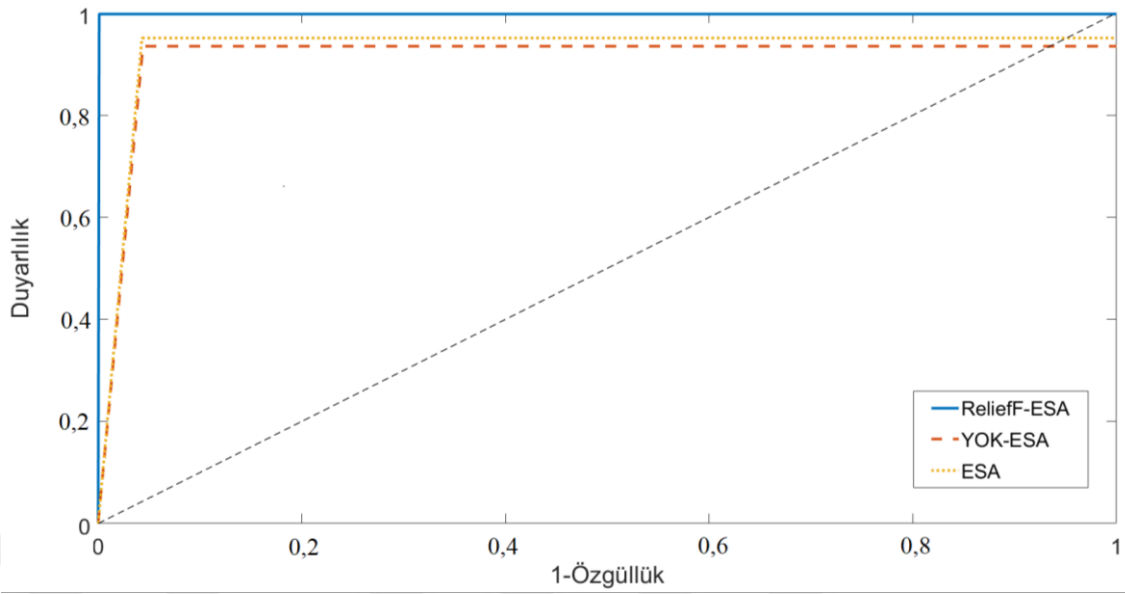
Boyut Azaltma	Yöntemler	Veri Kümesi	Duyarlılık	Kesinlik	Doğruluk
Düz	DVM	Test1(60-40)	59,02	61,24	60
		Test2(70-30)	61,63	63,32	62,54
		Test3(80-20)	64,22	66,23	65
		<i>Ortalama</i>	<i>61,62</i>	<i>63,6</i>	<i>62,51</i>
	ESA	Test1(60-40)	66,45	68,14	67,25
		Test2(70-30)	68,02	69,68	68,74
		Test3(80-20)	69,51	71,22	70,34
		<i>Ortalama</i>	<i>67,99</i>	<i>69,68</i>	<i>68,78</i>
YOK	DVM	Test1(60-40)	68,02	69,52	68,75
		Test2(70-30)	69,75	71,32	70,43
		Test3(80-20)	70,26	72,87	71,94
		<i>Ortalama</i>	<i>69,34</i>	<i>71,24</i>	<i>70,37</i>
	ESA	Test1(60-40)	69,85	71,49	70,56
		Test2(70-30)	71,14	73,87	72,22
		Test3(80-20)	74,05	76,72	75
		<i>Ortalama</i>	<i>71,68</i>	<i>74,03</i>	<i>72,59</i>
ReliefF	DVM	Test1(60-40)	79,05	80,75	79,96
		Test2(70-30)	79,86	81,42	80,54
		Test3(80-20)	80,94	82,23	81,67
		<i>Ortalama</i>	<i>79,95</i>	<i>81,47</i>	<i>80,72</i>
	ESA	Test1(60-40)	82,43	84,11	83,02
		Test2(70-30)	83,95	84,83	83,97
		Test3(80-20)	84,76	85,53	84,85
		<i>Ortalama</i>	<i>83,71</i>	<i>84,82</i>	<i>83,95</i>

Tez çalışmasında, ReLU aktivasyon tabanlı ReliefF-ESA ve YOK-ESA hibrit modelleri ile ESA ve DVM algoritmaları üç farklı mikrodizi [56-58] veri kümesi üzerine uygulanmıştır. Tez çalışmasında, YOK ve ReliefF boyut azaltma algoritmalarının DNA mikrodizi veri kümelerinin üzerindeki sınıflandırma performansları karşılaştırılmıştır. Mikrodizi verilerinin sınıflandırma performanslarını karşılaştırmak için doğruluk, kesinlik ve duyarlılık kriterleri kullanılmıştır. Üç mikrodizi veri kümelerinin hepsinde, ilk olarak DVM ve ESA sınıflandırma yöntemleri doğrudan uygulanmıştır. Mikrodizi veri kümelerinin çok fazla öznelikli yapısından kaynaklı olarak sınıflandırma olumsuz etkilenmektedir. Bu olumsuzluğu kaldırmak için boyut azaltma algoritmaları kullanılmış ve sınıflandırma değerlendirmesi tekrar yapılmıştır. Tablo 3.4'de, DVM ve ESA yöntemleri doğrudan yumurtalık kanseri mikrodizi verilerine uygulandığında, algoritmalar sırasıyla, %95,6 ve %95,97 doğrulukla sınıflandırma gerçekleştirmiştir. YOK boyut azaltma yöntemi uygulanarak 200 nitelik elde edilip, sınıflandırma yöntemleri uygulanarak sırasıyla, %95,82 ve %96,33 aralığında doğruluk değerleri elde

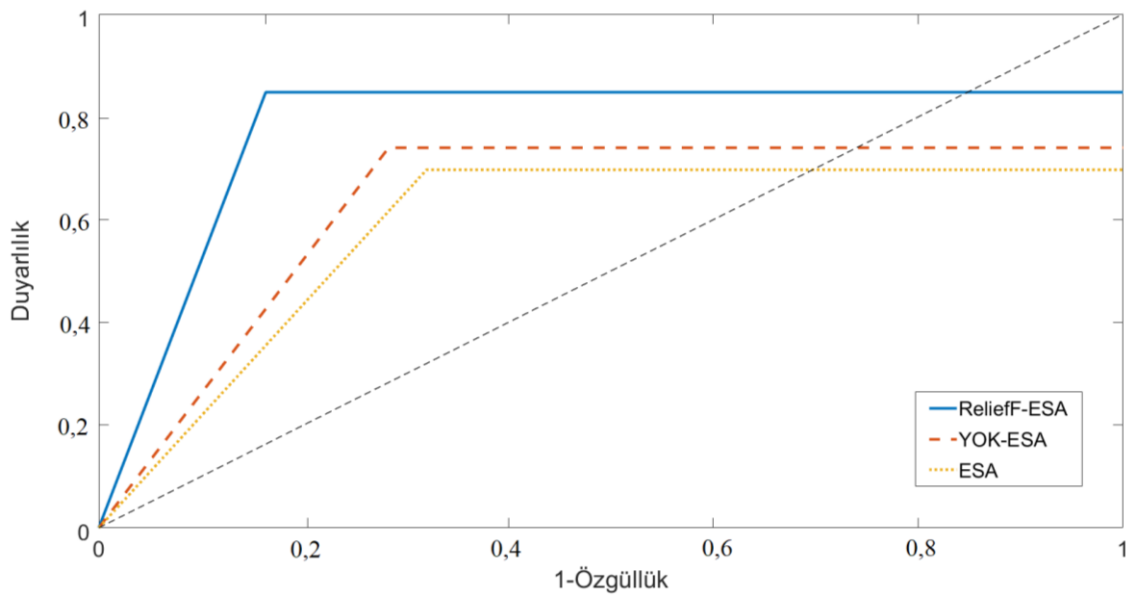
edilmiştir. Boyut azaltma yöntemlerinin performanslarını karşılaştırmak için kullanılan ReliefF yöntemi 36 öznitelik ve sınıflandırma yöntemleri DVM ve ESA uygulandığında sırasıyla, %97,47 ve %98,6 aralığından doğruluk elde edilmiştir. Gerçekleştirilen mikrodizi veri kümeleri üzerindeki deneylerde, ReliefF yönteminin YOK'den daha iyi boyut azaltma sağladığını göstermektedir. Tablo 3.5'de, ReliefF-ESA önerilen hibrit yöntem lösemi kanseri mikrodizi verilerine uygulandığında %99,86 en iyi başarı elde edilmiştir. Tablo 3.6'da, önerilen ReliefF-ESA yönteminin CNS mikrodizi veri kümesine uygulanmasıyla, doğruluk değerleri %65'ten %83,95'e yükselmiştir. ReliefF-ESA yöntemi diğer mikrodizi verilerinde kullanıldığında diğer önerilen yöntemlere göre daha başarılı sınıflandırma sonuçları elde edilmektedir. Derin öğrenme yöntemlerinden YOK'un boyut azaltma işleminin düzgün çalışabilmesi için çok fazla örneğe ihtiyaç vardır. Ancak ReliefF algoritması, az sayıda örnek içeren veri kümesinden gereksiz nitelikleri kaldırır ve küçük veri kümesiyle modelin doğruluğunu artırır [123]. Tablo 3.5'de gösterildiği gibi, en fazla örneğe sahip olan yumurtalık mikrodizi veri kümesi en yüksek başarı oranına sahiptir. Diğer mikrodizi veri kümelerinde, YOK'nin sınıflandırma sonuçları, örnek sayısının az olması nedeniyle önerilen yöntemden daha düşüktü. Ayrıca üç mikrodizi veri kümesini sınıflandırmak için önerilen hibrit yöntemler üzerinden elde edilen ROC eğrileri Şekil 3.2, 3.3 ve 3.4'de verilmiştir.



Şekil 3.2. Yumurtalık mikrodizi veri kümesi için algoritmaların ROC eğrisi



Şekil 3.3. Lössen mikrodizi veri kümesi için algoritmaların ROC eğrisi



Şekil 3.4. CNS mikrodizi veri kümesi için algoritmaların ROC eğrisi

ReliefF-ESA hibrit modeli üç mikrodizi veri kümesine uygulanmasıyla elde edilen ortalama duyarlılık ve özgüllük değerleriyle ROC eğrisi çizilmiştir (Şekil 3.2, 3.3 ve 3.4). Çizilen ROC eğrilerinin önerilen ReliefF-ESA modelinin diğer modellere göre daha başarılı bir sınıflandırma gerçekleştirdiği görülmüştür. Önerilen ReliefF-ESA modeli sonuçları, aynı veri kümesi üzerinde literatürde yapılan çalışmaların sonuçları ile karşılaştırılmıştır (Tablo 3.7).

Tablo 3.7. Mikrodizi veri kümeleriyle ilgili çalışmaların karşılaştırılması

Yazarlar	Yıl	Mikrodizi Veri Kümesi	Sınıf Sayısı	Özellik Sayısı	Doğruluk (%)
Kumar ve ark.[76]	2015	Lösemi	72	7129	97,22
		Yumurtalık	253	15154	98,42
Ludwig ve ark. [78]	2015	Lösemi	72	7129	87,50
		Yumurtalık	253	15154	97,11
Liu ve ark. [175]	2015	Lösemi	72	7129	94,4
Diaz ve ark. [77]	2016	Lösemi	38	3051	91,3
Medjahed ve ark. [79]	2017	Lösemi	72	5147	95,81
		Yumurtalık	253	15154	98,19
Liao ve ark. [80]	2017	Yumurtalık	253	15154	97,3
Wang ve ark.[72]	2019	Lösemi	72	7129	91,05
Shekar ve ark. [93]	2020	Lösemi	72	7129	98
Kang [176]	2019	CNS	72	7129	75
Alannı ve ark. [177]	2019	Lösemi	72	7129	98,61
		Yumurtalık	253	15154	98,6
Önerilen Yöntem	2019	Lösemi	72	7129	99,86
		CNS	60	7129	83,95

Tablo 3.7’de, mikrodizi veri kümelerinde önerilen hibrit ReliefF-ESA yönteminin, literatürdeki çalışmalara göre daha yüksek bir başarı oranına ulaştığı görülmektedir.

3.5. Sonuçlar

Kanser hastalığının dünya çapında yaygınlaşmasıyla birlikte doku örneklerinden mikrodizi verileri üretmenin ve bu verilerin analiz edilmesinin önemi son yıllarda artmıştır. Bu verilerin sınıflandırılmak, hastalığı teşhis etmek veya farklı tümör türlerini belirlemek için çok önemlidir. Bununla birlikte, az sayıda örnek, birçok özellik ve gürültülü veriler nedeniyle mikrodizi verilerini sınıflandırmak zordur. Tez çalışmasında, dünyada yaygın kanser türlerinden alınan mikrodizi veri kümelerinde hastalığı teşhis etmek için ReliefF boyut azaltma yöntemi ve ReLU aktivasyon fonksiyon tabanlı tek boyutlu ESA modeli hibrit olarak kullanılmıştır. Boyut azaltma ve sınıflandırma performanslarını karşılaştırmak için farklı yöntemlerle de analizler yapılmıştır. Deneysel çalışmalarda kullanılan her bir yöntem için üç farklı eğitim / test oranı belirlenmiş ve elde edilen modellerin ortalamaları ve başarı oranları alınarak sonuçlar değerlendirilmiştir. Ayrıca deneysel sonuçlar literatürdeki mevcut tekniklerle karşılaştırılmış ve önerilen

yaklaşımın başarılı sonuçlar ürettiği gözlemlenmiştir. Önerilen hibrit ReliefF - ESA yöntemi yumurtalık, lösemi ve CNS mikrodizi veri kümeleri için, sırasıyla, % 98,6, % 99,86 ve % 83,95 sınıflandırma doğrulukları elde edilmiştir. Önerilen hibrit modeller, kanser teşhisi için tıbbi veri analizi gerçekleştirilmektedir. Bu sayede insan hatalarının önüne geçilmeye çalışılmıştır.



4. BÖLÜM

RSigELU: DERİN SİNİR AĞLARI İÇİN DOĞRUSAL OLMAYAN AKTİVASYON FONKSİYONLARI

4.1. Giriş

Eğitim verilerini kullanarak model oluşturan derin öğrenme algoritmaları sonucu bilinmeyen giriş verilerinin değerlerini tahmin etmede kullanılmaktadır. Sinir ağı mimarilerinin eğitim aşamasında, modelin çalışmasının hesaplama maliyetini alztma, yerel minimuma takılmama, doğruluk oranını artırma gibi amaçlar için aktivasyon fonksiyonları kullanılmaktadır. Sinir ağlarının değişkenler arasındaki ve katmanları arasındaki karmaşık ilişkiyi modelleyebilmesi ve öğrenebilmesi için aktivasyon fonksiyonları önemlidir. İyi tasarlanmış aktivasyon fonksiyonu, modellerin performanslarında etkilidir. Derin öğrenme yöntemlerinde, aktivasyon katmanlarında ağa gelen giriş bilgisinin işlenmesi ve bu girdilere karşılık gelen çıktının üretilmesi için aktivasyon fonksiyonları kullanılmaktadır. Literatürde veriye ve probleme uygun çeşitli aktivasyon fonksiyonları önerilmiştir. Derin öğrenmede kullanılan aktivasyon fonksiyonlarının, doğrusal olmama ve türevinin alınabilir olması beklenmektedir [2].

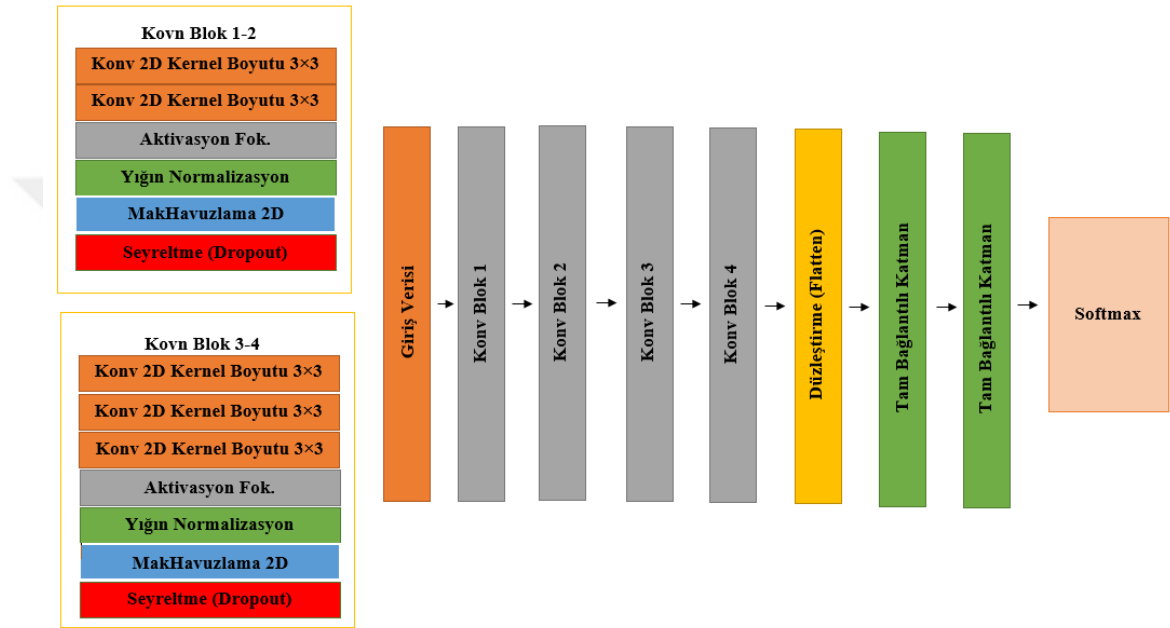
Aktivasyon fonksiyonları doğrusal olduğunda tek boyutlu YSA özelliğine sahip doğrusal regresyon gibi davranmaktadır. Ancak derin öğrenme mimarileri, yaygın olarak karmaşık gerçek dünya problemleriyle ilgilendiğinden, doğrusal aktivasyon fonksiyonu başarılı sonuç verememektedir. Bundan dolayı doğrusal olmayan aktivasyon fonksiyonları çok katmanlı derin sinir ağlarında tercih edilerek karmaşık gerçek dünya

verilerinden anlamlı özellikleri öğrenebilmektedir [2]. Derin sinir ağlarında geri yayılım algoritması kullanılarak parametrelerin güncellenmesi sağlanır. Güncelleme işleminde türev değeri geri döndürüldüğünden, mimari içerisindeki kullanılan aktivasyon fonksiyonunun türevinin alınabilir olması gerekmektedir. Ayrıca, derin sinir ağları mimarilerinde daha derin katmanlara doğru gidildikçe nöronlarda güncellemeler zorlaşmaktadır. Bundan dolayı kaybolan gradyan problemi başlamaktadır. Kaybolan gradyan problemi, derin sinir ağlarının derinleşerek öğrenmesinin önünde büyük engel oluşturmaktadır [6]. Bu engel mimarinin yerel minimuma takılma ihtimalini yükseltmektedir. Çünkü sinir ağları derinleştikçe kaybolan gradyan ve negatif bölge problemi ortaya çıkmaktadır. Literatürdeki bilinen problemlerin üstesinden gelmek için aktivasyon fonksiyonları doğrusal ve doğrusal olmayan olmak üzere ikiye ayrılabilir. Bu problem, yaygın olarak kullanılan ReLU [7], sigmoid [4] ve hiperbolik tanjant [5] aktivasyon fonksiyonlarında görülmektedir. Mevcut problemlerin üstesinden gelmek için literatürde, ReLU [7], LReLU [3], SELU [8], ELU [9], PReLU [10], Hexpo [11] ve LISA [12] gibi aktivasyon fonksiyonları önerilmiştir. Kaybolan gradyan probleminin üstesinden gelebilmek için önerilen ReLU aktivasyon fonksiyonu [7], pozitif bölge değerlerine göre işlem yaparken negatif bölge değerlerini sıfır olarak ayarlamaktadır. Bu durumda negatif çıktıların türevi alınamamakta ve negatif bölgedeki değerlerin öğrenme sırasında göz ardı edilmesine neden olmaktadır. Bu problemin üstesinden gelmek için literatürde LReLU [3], ELU [8], PReLU [10] gibi aktivasyon fonksiyonu çalışmaları yapılmıştır.

Bu tez çalışmasında, kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelmek amacıyla ReLU, sigmoid ve ELU aktivasyon fonksiyonlarının birleşimi olan yeni sabit tek ve çift parametrelili RSigELU (RSigELUS, RSigELUD) aktivasyon fonksiyonları önerilmiştir. Önerilen RSigELU aktivasyon fonksiyonları pozitif aktivasyon bölgesinde ReLU ve sigmoid aktivasyon fonksiyonlarının birleşimi davranışı sergiler, negatif aktivasyon bölgesinde ELU aktivasyon fonksiyonu gibi ve doğrusal aktivasyon bölgesinde doğrusal (linear) aktivasyon fonksiyon gibi davranmaktadır [7]. Deneysel değerlendirmeler MNIST [23], Fashion MNIST [24], IMDB [26] ve CIFAR-10 [25] benchmark veri kümeleri üzerinden gerçekleştirilmiştir. Yapılan çalışmanın sonuçları derlenerek bir dergide yayınlanmıştır [28].

4.2. Yöntem

Tez çalışmasında VGG tabanlı ESA mimarisi, önerilen RSigELU aktivasyon fonksiyonlarını test etmek için kullanılmıştır. ESA yöntemi hakkındaki bilgi Bölüm 1.1’de verilmiştir ve çalışmada kullanılan VGG mimarisi Şekil 4.1’de görülmektedir.



Şekil 4.1. VGG tabanlı ESA mimarisi

Şekil 4.1’de verilen ESA mimarisi, konvolüsyon, düzleştirme, iki adet tam bağlantılı ve softmax katmanlarında oluşmaktadır. Tablo 4.1’de çalışmada kullanılan mimaride birinci konvolüsyon katmanında 32 çekirdek bulunmaktadır. İkinci konvolüsyon katmanında 48 çekirdek, üçüncü konvolüsyon katmanında 64 çekirdek ve en son katmanda da 96 çekirdek bulunmaktadır.

Tablo 4.1. Çalışmada kullanılan VGG mimarinin parametreleri

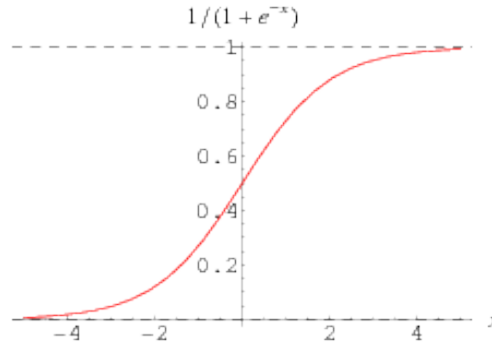
ESA Katmanı	Parametreler		
Girş Verisi			
Konv1	32@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Konv1	32@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Yığın Normalizasyon			
MakHavuzlama1	Havuz Boyutu =2	Adım boyutu=2	
Seyreltme (Dropout)	Oran 0,4		
Konv2	48@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Konv2	48@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Yığın Normalizasyon			
MakHavuzlama2	Havuz Boyutu =2	Adım boyutu =2	
Seyreltme	Oran 0,4		
Konv3	64@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Konv3	64@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Konv3	64@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Yığın Normalizasyon			
MakHavuzlama3	Havuz Boyutu =2	Adım boyutu =2	
Seyreltme	Oran 0,4		
Konv4	96@3×3 filtre	Dolgu =same	Düzenleyici.l2=0,0005
Konv4	96@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Konv4	96@3×3 filtre	Dolgu = same	Düzenleyici.l2=0,0005
Yığın Normalizasyon			
MakHavuzlama3	Havuz Boyutu=2	Adım boyutu =2	
Seyreltme	Oran 0,4		
Düzleştirme (Flatten)			
Tam bağlantılı katman	512		
Tam bağlantılı katman	512		
Softmax			

Deneyler, VGG mimarisi sabit tutularak gerçekleştirilmiştir. Deneylerde kullanılan veri kümeleri hakkındaki bilgiler Bölüm 1.5.3’de verilmiştir.

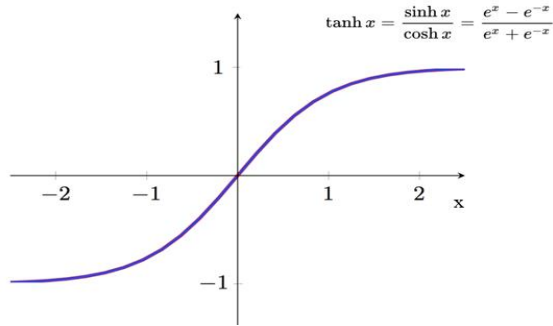
4.3. Önerilen RSigELU Aktivasyon Fonksiyonları

Tez çalışmasında, derin öğrenme modellerinde son zamanlarda gradyanların ortadan kaybolması ile ilgi odağı olan kaybolan gradyan probleminin üstesinden gelinebilmesi için yeni RSigELU aktivasyon fonksiyonları önerilmiştir. Derin sinir ağlarında bu problemlerden dolayı ağlar zorlukla güncellenmekte ya da ölmektedir. Bu işlem ağın derinleşmesine, öğrenmesine engel ve eğitim için çok fazla hesaplama gücüne ihtiyaç duyulmasına sebep olmaktadır. Bu nedenle derin sinir ağlarında yerel minimuma takılma gerçekleşebilmektedir [63]. Kaybolan gradyan problemi Hochreiter tarafından tespit edilmiştir [120]. Bu problem art arda dizili derin öğrenme modellerinde geri yayılım

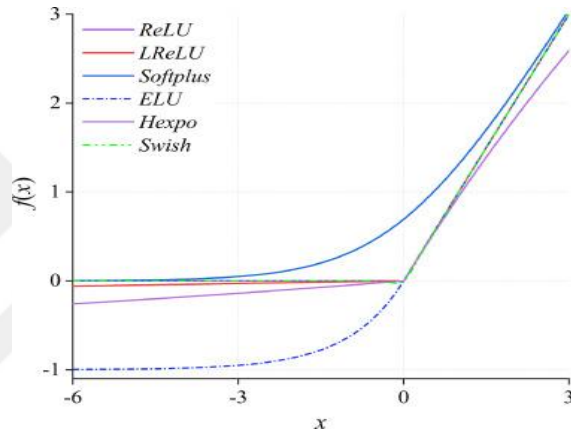
algoritması kullanılarak her işlemde türev alınmasından dolayı ortaya çıkmaktadır. Sinir ağları mimarilerinde, ağırlıkların güncellenmesi amacıyla geri yayılım algoritması kullanılmaktadır. Geri yayılım algoritması her adımda türev işlemi alarak ağırlıklar günceller. Bundan dolayı, model derinleşmeye başlamasıyla ve türevlerin küçülerek 0 olması ile gradyan yok olmakta ve öğrenme çok az gerçekleşmektedir. Bu nedenle, derin ağlarda yaygın olarak kullanılan sigmoid ve tanjant aktivasyon fonksiyonlarından dolayı modelde kaybolan gradyan problemi gözlemlenmektedir. Sigmoid fonksiyonu Şekil 4.2'de $[-5,5]$ değerleri arasında 0-1 aralığında anlamlı çıktı üretmekte ve sonraki işleminde sonucu 0'a yaklaşıp öğrenmeyi kesmektedir. Ayrıca aynı problem hiperbolik tanjant fonksiyonu için de geçerlidir. Şekil 4.3'de görüldüğü gibi $[-1,1]$ değerleri arasında anlamlı çıktılar üretmektedir ve geri yayılım algoritmasındaki türev işleminden dolayı, hiperbolik tanjant aktivasyon fonksiyonu sigmoid aktivasyon fonksiyonu gibi sonucu 0'a yaklaşmaktadır [121]. Bu problemin üstesinden gelebilmek için türevi alınabilen fonksiyonlar tercih edilerek, öğrenme işleminin sürekli olması sağlanmaktadır. Literatürde önerilen bazı aktivasyon fonksiyonlarının davranışları Şekil 4.4'de gösterilmiştir.



Şekil 4.2. Sigmoid fonksiyon



Şekil 4.3. Hiberbolik tanjant fonksiyon

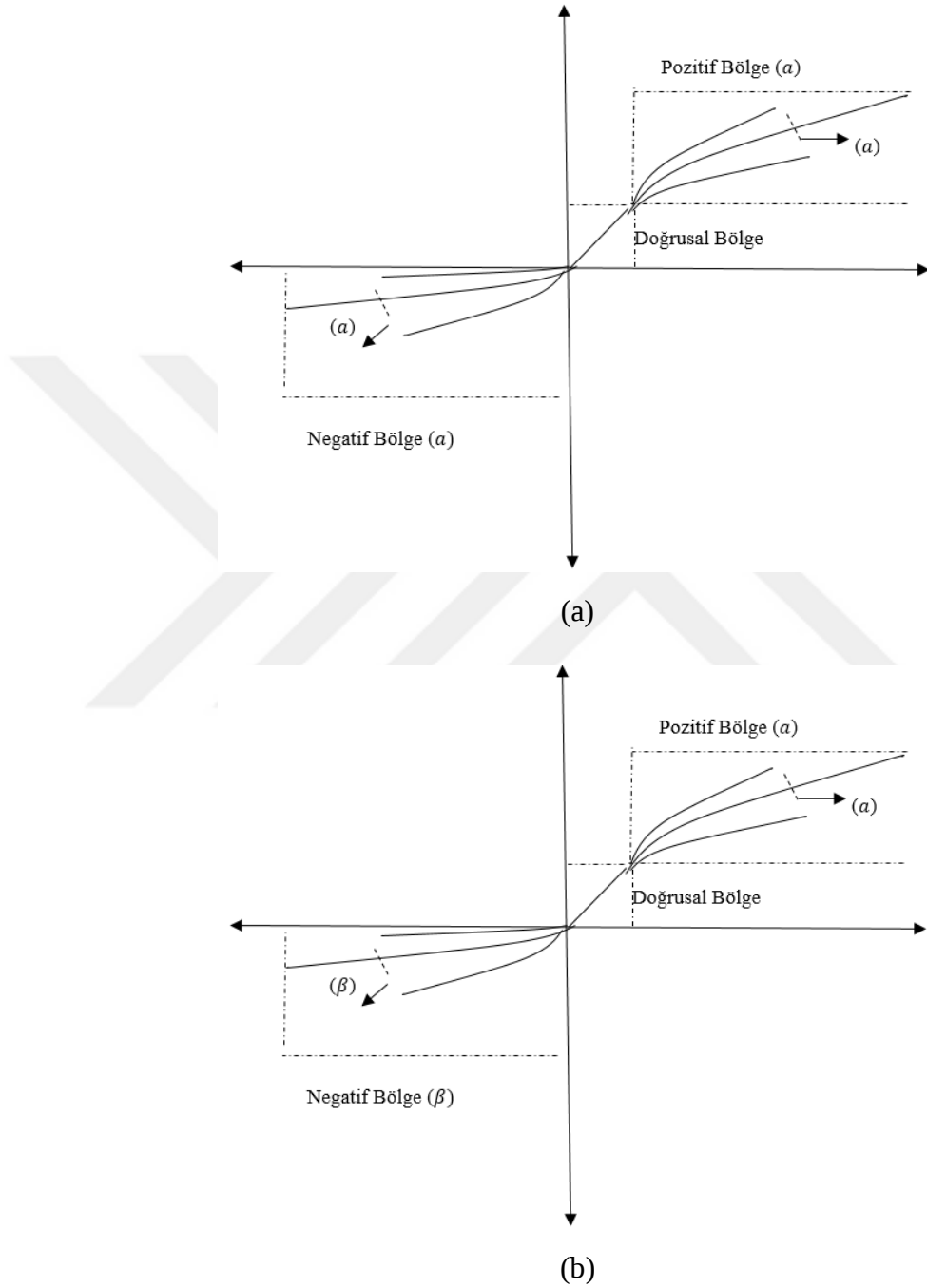


Şekil 4.4. Literatürdeki bazı aktivasyon fonksiyonlarının davranışları [63]

Kaybolan gradyan probleminin üstesinden gelebilmek için ortaya atılan ReLU aktivasyon fonksiyonu [7], pozitif değerlere göre işlem yaparken negatif girdileri sıfır olarak ayarlamaktadır. Böylece negatif çıktıların türev alınamayıp öğrenme işlemi gerçekleşmemektedir.

Bu tez çalışmasında, kaybolan gradyan problemi ve negatif bölge problemlerinin üstesinden gelmek için üç aktivasyon bölgesindeki (pozitif, negatif ve doğrusal bölgeler) veriye uygun çalışabilecek özelliğe sahip yeni RSigELU aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonları tek ve çift parametrelilik olarak test edilmiştir. Önerilen iki parametre her iki bölgenin de karakteristiğini yakalayabilmek için kullanılmıştır. Bu parametrelerden tek ve çift olarak ifade edilen α ve β tanımlanmış eğim katsayılarıdır. Pozitif aktivasyon bölgesinin kontrolü α ile sağlanırken, negatif

aktivasyon bölgesinin kontrolü β ile sağlanmaktadır. Önerilen aktivasyon fonksiyonunun davranışı Şekil 4.5’de gösterilmektedir.



Şekil 4.5. Önerilen (a) RSigELUS ve (b) RSigELUD fonksiyonlarının davranışları

Önerilen RSigELUS ve RSigELUD aktivasyon fonksiyonlarının detayları aşağıdaki bölümlerde verilmiştir.

4.3.1. RSigELUS Aktivasyon Fonksiyonu

Önerilen tek parametrelili RSigELU aktivasyon fonksiyonu (RSigELUS) Eşitlik 4.1’de verilmiştir.

$$f(x) = \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \alpha(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.1)$$

Eşitlik 4.1’deki RSigELUS fonksiyonunun üç farklı aktif olduğu bölge gösterilmektedir. RSigELUS aktivasyon fonksiyonu, diğer tüm eğitim parametreleri sabit tutularak, α eğim değişkeni $0 < \alpha < 1$ arasında değerler alarak deneysel değerlendirme yapılmaktadır. Bu değer pozitif ve negatif bölgeyi kontrol etmek için kullanılmaktadır. Eşitlik 4.1’deki x parametresi fonksiyonunun girişini temsil etmektedir. Önerilen RSigELUS aktivasyon fonksiyonunun türevi Eşitlik 4.3’de gösterilmektedir. Üç farklı bölgede aktif olabilen RSigELUS aktivasyon fonksiyonu Şekil 4.5-a’deki gibi davranış sergilemektedir. Pozitif bölge için sigmoid ile ReLU aktivasyon fonksiyonları hibrit yapı gerçekleştirerek derin modellerde kaybolan gradyan problemi olmadan sürekli öğrenme gerçekleştirebilmektedir. Doğrusal bölge için doğrusal aktivasyon fonksiyonu kullanılmaktadır. Negatif bölge için ise ELU aktivasyon fonksiyonu kullanılmaktadır. Önerilen tek parametrelili RSigELUS aktivasyon fonksiyonu verilen değerlere göre farklı davranış sergileyebilmektedir. RSigELUS aktivasyon fonksiyonunda α parametresi 0 ($\alpha = 0$) değerini aldığıında Eşitlik 4.2’deki gibi ReLU davranışı sergilemektedir.

$$f(x) = \begin{cases} x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ 0, & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.2)$$

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \alpha e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.3)$$

4.3.2. RSigELUD Aktivasyon Fonksiyonu

Önerilen çift parametrelili RSigELU aktivasyon fonksiyonu (RSigELUD) Eşitlik 4.4’de verilmiştir.

$$f(x) = \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \beta(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.4)$$

Eşitlik 4.4’de RSigELUD fonksiyonunun üç farklı aktif olduğu bölge gösterilmektedir. β eğim katsayısı için $0 < \beta < 1$ arasında değerler alabilmektedir. Deneysel çalışmalarda $\beta = 0,2$ olarak belirlenerek tüm deneylerde bu değer kullanılmıştır. RSigELUD aktivasyon fonksiyonunun türevi Eşitlik 4.6’da gösterilmektedir. Önerilen her iki aktivasyon fonksiyonlarının türevlerinin alınabilmesi nedeniyle öğrenme işleminin sürekli olduğu ifade edilmektedir. Eşitlik 4.1’de önerilen aktivasyon fonksiyonu verilen değerlere göre farklı aktivasyon fonksiyonu gibi davranabilmektedir. Örneğin $\alpha = 0$ ve $\beta = 0$ değerlerini aldığında Eşitlik 4.2’deki gibi ReLU aktivasyon fonksiyonu davranışı sergilemektedir. Ayrıca $\alpha = 0$ ve $\beta = 0,2$ değerlerini aldığında Eşitlik 4.5’deki gibi ELU aktivasyon fonksiyonu davranışı sergilemektedir.

$$f(x) = \begin{cases} x, & \text{eğer } 0 \leq x \leq \infty \\ \beta(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.5)$$

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \beta e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (4.6)$$

Sinir ağlarında eğitim sırasında parametrelerin güncellenmesi için geri yayılım mimarisi esas alındığından, kullanılan aktivasyon fonksiyonlarının da türevinin alınabilmesi önemlidir. Önerilen tek ve çift parametrelili aktivasyon fonksiyonları incelendiğinde türev işleminden sonra hem pozitif hem de negatif bölgede aktivasyon fonksiyonlarının aktif olduğu görülmektedir.

4.4. Deneyisel Değerlendirme

Tez çalışmasında önerilen RSigELU aktivasyon fonksiyonlarının MNIST [22], Fasion MNIST [23], CIFAR-10 [24] ve IMDB Movie [25] veri kümeleri üzerine deneyisel değerlendirmeleri yapılmıştır. Çift parametrelili aktivasyon fonksiyonunda β değerinin belirlenmesi için MNIST, Fasion MNIST, CIFAR-10 ve IMDB Movie veri kümeleri üzerine gerçekleştirilen deney sonuçları sırasıyla, Tablo 4.2, 4.3, 4.4 ve 4.5’de görülmektedir. MNIST veri kümesi için önerilen tek ve çift parametrelili RSigELU aktivasyon fonksiyonları ve bilinen diğer aktivasyon fonksiyonlarına ait başarı oranları sırasıyla, Tablo 4.6, 4.7, ve 4.8’de görülmektedir. Deneylerde karşılaştırma için kullanılan aktivasyon fonksiyonlarının parametre değerleri Bawa [136] tarafından yapılan çalışmada verilenlere göre ayarlanmıştır. Gerçekleştirilen deneyler 5’er defa tekrarlanarak sonuçlar elde edilmiştir.

Tablo 4.2. MNIST veri kümesi kullanılarak β parametresinin belirlenmesi

Alpha (α)	Scale (β)									
	0,1		0,2		0,3		0,4		0,5	
	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.
0,1	0,0537	0,9935	0,0465	0,9946	0,0449	0,9945	0,0495	0,9935	0,0651	0,9927
0,2	0,0562	0,9929	0,0390	0,9946	0,0529	0,9941	0,0540	0,9935	0,0596	0,9944
0,3	0,0467	0,9935	0,0356	0,9948	0,0410	0,9931	0,0550	0,9931	0,0489	0,9939
0,4	0,0427	0,9941	0,0360	0,9945	0,0548	0,9936	0,0510	0,9939	0,0548	0,9934
0,5	0,0457	0,9940	0,0458	0,9947	0,0498	0,9938	0,0616	0,9926	0,0491	0,9937

Tablo 4.3. Fasion-MNIST veri kümesi kullanılarak β parametresinin belirlenmesi

Alpha (α)	Scale (β)									
	0,1		0,2		0,3		0,4		0,5	
	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.
0,1	0,5812	0,9250	0,337	0,9252	0,5353	0,9135	0,5558	0,9167	0,5733	0,9190
0,2	0,6023	0,9243	0,4985	0,9256	0,5335	0,9220	0,5265	0,9216	0,5204	0,9190
0,3	0,5748	0,9221	0,6519	0,9257	0,5522	0,9144	0,4610	0,9151	0,4732	0,9205
0,4	0,5898	0,9222	0,4755	0,9255	0,5440	0,9223	0,4468	0,9222	0,4091	0,9209
0,5	0,5566	0,9255	0,4700	0,9244	0,4888	0,9162	0,4296	0,9216	0,5174	0,9155

Tablo 4.4. CIFAR-10 veri kümesi kullanılarak β parametresinin belirlenmesi

Alpha (α)	Scale (β)									
	0,1		0,2		0,3		0,4		0,5	
	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.
0,1	0,6307	0,8353	0,6787	0,8358	0,6524	0,8284	0,6214	0,8327	0,6658	0,8325
0,2	0,6416	0,8323	0,6663	0,8346	0,6837	0,8179	0,6647	0,8343	0,6750	0,8238
0,3	0,6129	0,8318	0,6897	0,8325	0,6124	0,8310	0,6508	0,8291	0,7277	0,8016
0,4	0,6995	0,8317	0,6746	0,8331	0,6466	0,8300	0,6481	0,8326	0,6655	0,8310
0,5	0,7029	0,8161	0,6095	0,8421	0,6405	0,8316	0,6083	0,8377	0,6548	0,8215

Tablo 4.5. IMDB veri kümesi kullanılarak β parametresinin belirlenmesi

Alpha (α)	Scale (β)									
	0,1		0,2		0,3		0,4		0,5	
	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.	Doğ. Hata	Doğ. Doğ.
0,4	0,7906	0,8101	0,8604	0,8411	0,7011	0,8356	0,7247	0,8411	0,7525	0,8422
0,5	0,8547	0,8387	0,8077	0,8407	0,7615	0,8404	0,7208	0,8400	0,7150	0,8327
0,6	0,8920	0,8320	0,7333	0,8330	0,9370	0,8221	0,7603	0,8284	0,7388	0,8208
0,7	0,7920	0,8339	0,6230	0,8357	0,7372	0,8330	0,7617	0,8264	0,7489	0,8353
0,8	0,6197	0,8354	0,7337	0,8372	0,7573	0,8370	0,7540	0,8311	0,7217	0,8326
0,9	0,7038	0,8296	0,6614	0,8392	0,7499	0,8305	0,6997	0,8338	0,7433	0,8266

MNIST, Fasion MNIST, CIFAR-10 ve IMDB Movie veri kümelerinde sırasıyla, Tablo 4.2, 4.3, 4.4, ve 4.5’de gerçekleştirilen deneylerdeki sonuçlara göre tüm çift parametrelili deneylerde β değeri $\beta = 0,2$ ’de en iyi başarı oranları elde edilebilmektedir. Bundan dolayı çalışmada β değeri $\beta = 0,2$ belirlenmiştir.

Tablo 4.6. MNIST veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları

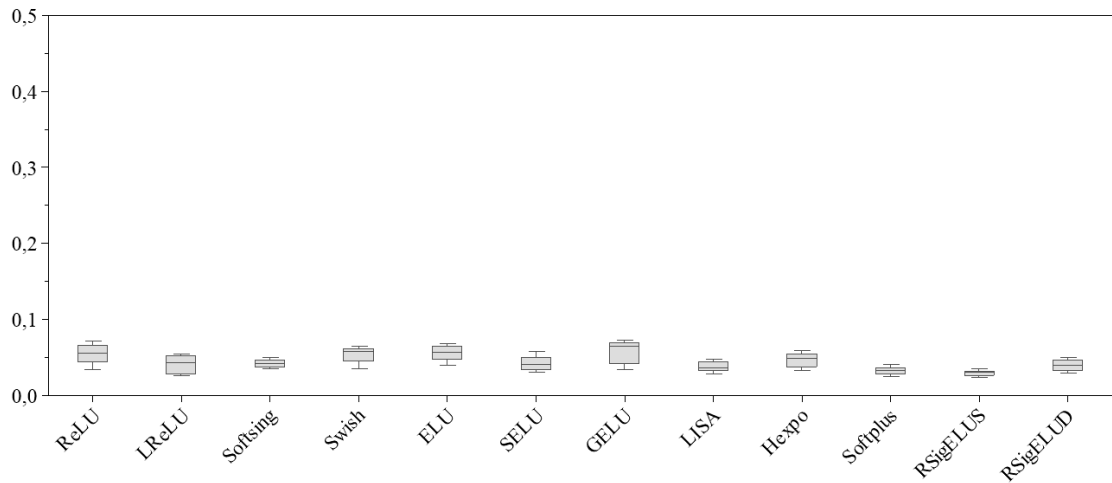
Alpha(α)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,01	40	0,0032	0,9990	0,0284	0,9949	290
0,02	40	0,0026	0,9995	0,0270	0,9953	321
0,03	40	0,0023	0,9994	0,0282	0,9955	325
0,04	40	0,0016	0,9995	0,0264	0,9948	320
0,05	40	0,0026	0,9994	0,0302	0,9944	325
0,1	40	0,0029	0,9991	0,0306	0,9955	480
0,15	40	0,0030	0,9992	0,0297	0,9942	442
0,2	40	0,0033	0,9990	0,0254	0,9957	290
0,25	40	0,0037	0,9989	0,0236	0,9947	661
0,3	40	0,0043	0,9990	0,0326	0,9949	662
0,35	40	0,0052	0,9986	0,0378	0,9938	683
0,4	40	0,0050	0,9987	0,0298	0,9940	334
0,45	40	0,0060	0,9985	0,0306	0,9941	323
0,5	40	0,0081	0,9981	0,0425	0,9936	324

Tablo 4.7. MNIST veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları

Alpha(α)	Scale(β)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,05	0,2	40	0,0038	0,9993	0,0322	0,9953	327
0,1	0,2	40	0,0022	0,9996	0,0465	0,9946	364
0,15	0,2	40	0,0024	0,9996	0,0504	0,9951	360
0,2	0,2	40	0,0024	0,9995	0,0390	0,9945	362
0,25	0,2	40	0,0021	0,9996	0,0320	0,9951	363
0,3	0,2	40	0,0024	0,9995	0,0356	0,9948	362
0,35	0,2	40	0,0028	0,9993	0,0416	0,9947	362
0,4	0,2	40	0,0034	0,9994	0,0360	0,9945	362
0,45	0,2	40	0,0034	0,9994	0,0353	0,9944	363
0,5	0,2	40	0,0023	0,9995	0,0458	0,9947	366

Tablo 4.8. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	P1	P2	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
Relu [7]			40	0,0019	0,9995	0,0326	0,9946	200
Lrelu [3]	0,2		40	0,0027	0,9994	0,0277	0,9951	201
Softsing [34]			40	0,0055	0,9985	0,0417	0,9933	201
Swish [62]			40	0,0024	0,9993	0,0340	0,9946	200
Elu [9]	0,2		40	0,0133	0,9970	0,0614	0,9917	205
Selu [8]			40	0,0037	0,9990	0,0290	0,9951	284
Gelu [59]			40	0,0010	0,9996	0,0396	0,9945	401
LISA [12]	0,15		40	0,0010	0,9998	0,0321	0,9945	324
Hexpo [64]			40	0,0008	0,9999	0,0354	0,9949	326
Softplus [66]			40	0,0075	0,9974	0,0288	0,9936	201
Önerilen RSigELUS	0,2		40	0,0033	0,9990	0,0254	0,9957	290
Önerilen RSigELUD	0,05	0,2	40	0,0038	0,9993	0,0322	0,9953	327



Şekil 4.6. MNIST veri kümesi aktivasyon fonksiyonlarının ortalama doğrulama (validation) hata (Doğ. Hata) sonuçları

Şekil 4.6, MNIST veri kümesi için aktivasyon fonksiyonu doğrulama (validation) hata değerlerini göstermektedir. Şekilde görülebileceği gibi, önerilen RSigELUS aktivasyon fonksiyonunun ortalama doğrulama hata değeri en küçüktür ve en küçük değişkenliğe sahiptir. Önerilen RSigELUD aktivasyon fonksiyonunun ortalama doğrulama hata değeri, LISA, Softplus ve softsing aktivasyon fonksiyonları dışında diğer aktivasyon fonksiyonlarından daha küçüktür. Önerilen RSigELUS ve RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre tutarlı sonuçlar vermiştir.

Tablo 4.6, 4.7, ve 4.8’de, MNIST veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 40 tekrar ile RSigELUS aktivasyon fonksiyonunda $\alpha=0,2$ parametre değeri ile 0,9957 doğruluk ve 0,0254 hata elde edilmiştir. Önerilen RSigELUD aktivasyon fonksiyonu $\alpha=0,05$ ile $\beta=0,2$ değerlerinde doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) sonuçları, sırasıyla, 0,9953 ve 0,0322 gözlemlenmiştir. Literatürde bilinen aktivasyon fonksiyonlarından ise LReLU ve SELU aktivasyon fonksiyonlarında yapılan deneylerde her ikisinin doğrulama doğruluk değeri 0,9951 olduğu görülmektedir. Ancak önerilen aktivasyon fonksiyonları diğer bilinen aktivasyon fonksiyonlarına göre biraz yavaş çalıştığı görülmektedir.

Fasion MNIST veri kümesi için önerilen tek ve çift parametrelili aktivasyon fonksiyonları ve bilinen diğer aktivasyon fonksiyonlarına ait başarı oranları, sırasıyla, Tablo 4.9, 4.10 ve 4.11’de görülmektedir.

Tablo 4.9. Fashion MNIST veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları

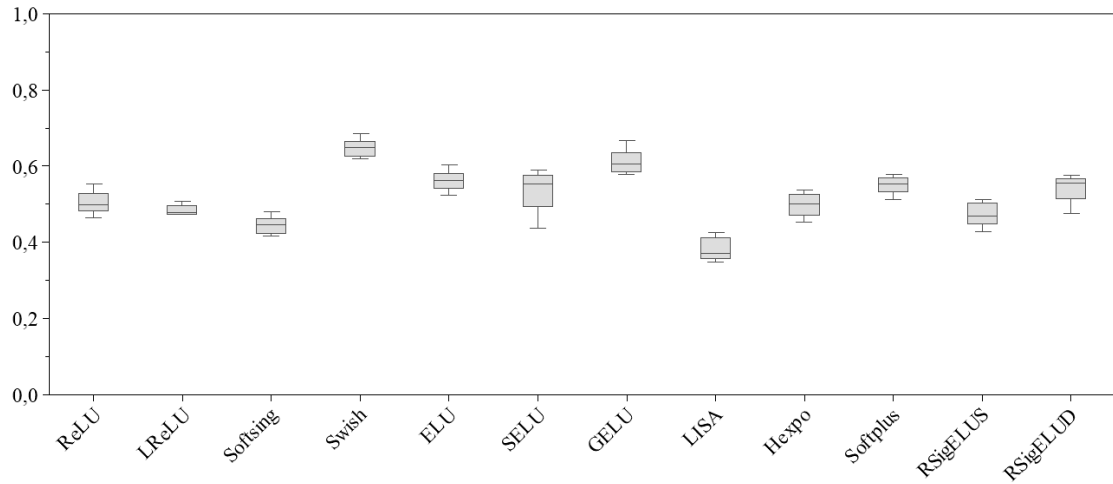
Alpha(α)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,01	40	0,0203	0,9932	0,3504	0,9273	369
0,02	40	0,0200	0,9946	0,5267	0,9260	367
0,03	40	0,0216	0,9936	0,4473	0,9257	368
0,04	40	0,0203	0,9939	0,5279	0,9266	369
0,05	40	0,0224	0,9933	0,4682	0,9277	365
0,1	40	0,0215	0,9938	0,5071	0,9254	370
0,15	40	0,0268	0,9926	0,4005	0,9276	372
0,2	40	0,0293	0,9917	0,5266	0,9259	380
0,25	40	0,0312	0,9904	0,6026	0,9235	374
0,3	40	0,0474	0,9848	0,4272	0,9239	374
0,35	40	0,0420	0,9879	0,4668	0,9239	361
0,4	40	0,0381	0,9884	0,5437	0,9217	362
0,45	40	0,0620	0,9799	0,4346	0,9202	362
0,5	40	0,0505	0,9836	0,4971	0,9220	363
0,001	40	0,0222	0,9931	0,5565	0,9273	362
0,002	40	0,0195	0,9938	0,4250	0,9254	363
0,003	40	0,0199	0,9938	0,5175	0,9244	364
0,004	40	0,0208	0,9936	0,5188	0,9236	364
0,005	40	0,0203	0,9938	0,4510	0,9269	367
0,006	40	0,0218	0,9933	0,5115	0,9242	366
0,007	40	0,0216	0,9934	0,4615	0,9264	369
0,008	40	0,0210	0,9937	0,4994	0,9275	370
0,009	40	0,0252	0,9925	0,5157	0,9237	370

Tablo 4.10. Fashion MNIST veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları

Alpha(α)	Scale(β)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,05	0,2	40	0,0261	0,9926	0,5712	0,9232	372
0,1	0,2	40	0,0300	0,9912	0,337	0,9252	370
0,15	0,2	40	0,0254	0,9930	0,5543	0,9174	370
0,2	0,2	40	0,0275	0,9925	0,4985	0,9256	369
0,25	0,2	40	0,0291	0,9914	0,4694	0,9240	368
0,3	0,2	40	0,0277	0,9917	0,6519	0,9257	368
0,35	0,2	40	0,0273	0,9921	0,562	0,9257	367
0,4	0,2	40	0,0309	0,9911	0,4755	0,9255	366
0,45	0,2	40	0,0343	0,9896	0,5535	0,9234	365
0,5	0,2	40	0,0318	0,9910	0,4700	0,9244	366

Tablo 4.11. Fashion MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	P1	P2	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
Relu [7]			40	0,0206	0,9938	0,4987	0,9243	285
Lrelu [3]	0,2		40	0,0298	0,9913	0,4731	0,9266	288
Softsing [34]			40	0,0276	0,9906	0,4500	0,9190	246
Swish [62]			40	0,0248	0,9924	0,6259	0,9123	286
Elu [9]	0,2		40	0,0201	0,9937	0,5620	0,9231	329
Selu [8]	0,1		40	0,0226	0,9930	0,5647	0,9139	292
Gelu [59]			40	0,0228	0,9932	0,5843	0,9159	411
LISA [12]	0,15		40	0,0379	0,9871	0,4024	0,9234	369
Hexpo [64]			40	0,0238	0,9920	0,4892	0,9178	369
Önerilen RSigELUS	0,05		40	0,0224	0,9933	0,4682	0,9277	365
Önerilen RSigELUD	0,35	0,2	40	0,0273	0,9921	0,5620	0,9257	367



Şekil 4.7. Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama hata (Doğ. Hata) sonuçları

Şekil 4.7, Fashion MNIST veri kümesine yönelik aktivasyon fonksiyonu doğrulama hata değerlerini göstermektedir. Şekilde görülebileceği gibi, önerilen RSigELUS aktivasyon fonksiyonunun ortalama doğrulama hata değeri ReLU, LReLU, Swish, ELU, SELU, GELU, Hexpo, Softplus aktivasyon fonksiyonları ve önerilen RSigELUD aktivasyon fonksiyonundan daha küçük ve LISA ve Softsing aktivasyon fonksiyonlarından daha büyüktür. Önerilen RSigELUD aktivasyon fonksiyonunun ortalama doğrulama hata değeri, ReLU, LReLU, Softsing, SELU, LISA, Hexpo ve önerilen RSigELUS dışında diğer aktivasyon fonksiyonlarından daha küçüktür.

Tablo 4.9, 4.10 ve 4.11 incelendiğinde Fasion MNIST veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile RSigELUS aktivasyon fonksiyonunda $\alpha=0,05$ parametre değeri ile 0,9277 doğruluk ve 0,4682 hata elde edilmiştir. Önerilen RSigELUD aktivasyon fonksiyonu, diğer aktivasyon fonksiyonlarından LReLU göre düşük başarı elde ettiği Tablo 4.11’de görülmektedir. Bu değerler, RSigELUD fonksiyonu için doğrulama doğruluk (Doğ_Doğ.) 0,9257, LReLU fonksiyonu için doğrulama doğruluk (Doğ_Doğ.) 0,9266 değerleri elde edilmiştir.

CIFAR-10 veri kümesi için aktivasyon fonksiyonlarına ait başarı oranları Tablo 4.12, 4.13 ve 4.14’da görülmektedir.

Tablo 4.12. CIFAR-10 veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları

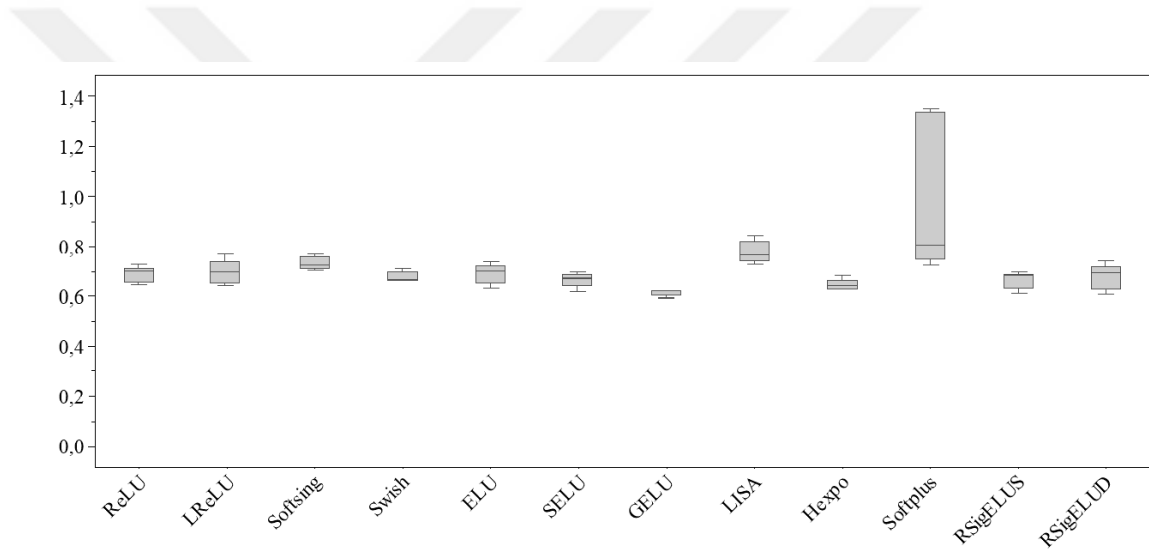
Alpha(α)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,05	40	0,3811	0,9081	0,6772	0,8382	720
0,1	40	0,3839	0,9080	0,6610	0,8365	788
0,15	40	0,4156	0,8979	0,6876	0,8385	880
0,2	40	0,3799	0,9093	0,6764	0,8398	840
0,25	40	0,3807	0,9092	0,6825	0,8372	843
0,3	40	0,3861	0,9056	0,7212	0,8269	875
0,35	40	0,3895	0,9057	0,6866	0,8346	844
0,4	40	0,3968	0,9012	0,6783	0,8307	641
0,45	40	0,5030	0,8608	0,5920	0,8412	588
0,5	40	0,4926	0,8634	0,6256	0,8312	566

Tablo 4.13. CIFAR-10 veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları

Alpha(α)	Scale(β)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,05	0,2	40	0,3177	0,9021	0,5477	0,8430	728
0,1	0,2	40	0,3983	0,9030	0,6787	0,8358	625
0,15	0,2	40	0,3884	0,9059	0,6804	0,8345	620
0,2	0,2	40	0,4034	0,9014	0,6663	0,8346	609
0,25	0,2	40	0,3906	0,9043	0,6980	0,8294	648
0,3	0,2	40	0,3998	0,9012	0,6897	0,8325	566
0,35	0,2	40	0,4016	0,9003	0,6922	0,8283	572
0,4	0,2	40	0,3896	0,9031	0,6746	0,8331	569
0,45	0,2	40	0,3904	0,9018	0,6851	0,8297	648
0,5	0,2	40	0,4811	0,8678	0,6095	0,8421	644

Tablo 4.14. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	P1	P2	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
Relu [7]			40	0,5444	0,8511	0,6858	0,8161	640
Lrelu [3]	0,2		40	0,3782	0,8799	0,5970	0,8202	655
Softsing [34]			40	0,6264	0,8236	0,7167	0,8036	488
Swish [62]			40	0,5108	0,8660	0,6701	0,8281	485
Elu [9]	0,2		40	0,5455	0,8512	0,6409	0,8292	482
Selu [8]			40	0,4870	0,8658	0,6014	0,8387	523
Gelu [59]			40	0,4870	0,8640	0,6113	0,8342	532
LISA [12]	0,15		40	0,6097	0,8240	0,6332	0,8218	562
Hexpo [64]			40	0,5765	0,8411	0,6364	0,8307	602
Softplus [66]			40	0,5640	0,8455	1,7497	0,6113	628
Önerilen RSigELUS	0,45		40	0,5030	0,8608	0,5920	0,8412	588
Önerilen RSigELUD	0,05	0,2	40	0,3177	0,9021	0,5477	0,8430	728



Şekil 4.8. CIFAR-10 veri kümesi için öne aktivasyon fonksiyonlarının ortalama doğrulama hata (Doğ. Hata) sonuçları

Şekil 4.8, CIFAR-10 veri kümesi için aktivasyon fonksiyonu doğrulama hata değerlerini göstermektedir. Şekilde görülebileceği gibi, önerilen RSigELUS aktivasyon fonksiyonunun ortalama doğrulama hata değeri, GELU aktivasyon fonksiyonu dışındaki diğer fonksiyonlardan daha küçüktür. Önerilen RSigELUD aktivasyon fonksiyonunun ortalama doğrulama hata değeri, SELU, GELU, Hexpo ve RSigELUS aktivasyon fonksiyonları dışında diğer aktivasyon fonksiyonlarından daha küçüktür. Şekil 4.8'de görüldüğü gibi, önerilen RSigELUS ve RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre tutarlı sonuçlar vermiştir.

Tablo 4.12, 4.13 ve 4.14’da CIFAR-10 veri kümesi için önerilen tek ve çift parametrelili RSigELU aktivasyon fonksiyonları ve diğer aktivasyon fonksiyonlarına göre başarı ve hata oranları gözlemlenmektedir. Deneylerde en iyi doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile RSigELUD aktivasyon fonksiyonu $\alpha=0,05$ ve $\beta=0,2$ parametre değeriyle, sırasıyla, 0,8430 doğruluk ve 0,5477 hata sonuçları olduğu görülmektedir. Ardından RSigELUS aktivasyon fonksiyonunda $\alpha=0,45$ değer ile doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri sırasıyla, 0,8412 doğruluk ve 0,5920 hata elde edilmiştir. Önerilen aktivasyon fonksiyonlarının dışında en iyi başarı oranı SELU aktivasyon fonksiyonunda olduğu görülmektedir. SELU fonksiyonu için en iyi ile doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri sırasıyla, 0,8387 ve 0,6014 olduğu görülmektedir.

IMDB Movie veri kümesi için aktivasyon fonksiyonlarına ait başarı oranları Tablo 4.15, 4.16 ve 4.17’de görülmektedir.

Tablo 4.15. IMDB Movie veri kümesi üzerinde RSigELUS aktivasyon fonksiyonunun başarı oranları

Alpha(α)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,4	40	0,0300	0,9902	0,7540	0,8338	925
0,45	40	0,0068	0,9976	0,9519	0,8334	521
0,5	40	0,0095	0,9972	0,8737	0,8251	522
0,55	40	0,0178	0,9940	0,8103	0,8357	877
0,6	40	0,0210	0,9929	0,7885	0,8252	878
0,65	40	0,0212	0,9926	0,7764	0,8378	880
0,7	40	0,0195	0,9929	0,7394	0,8427	831
0,75	40	0,0204	0,9932	0,7903	0,8448	820
0,8	40	0,0372	0,9865	0,9033	0,8174	822
0,85	40	0,0332	0,9883	0,7173	0,8372	923
0,9	40	0,0104	0,9966	0,8002	0,8332	520

Tablo 4.16. IMDB Movie veri kümesi üzerinde RSigELUD aktivasyon fonksiyonunun başarı oranları

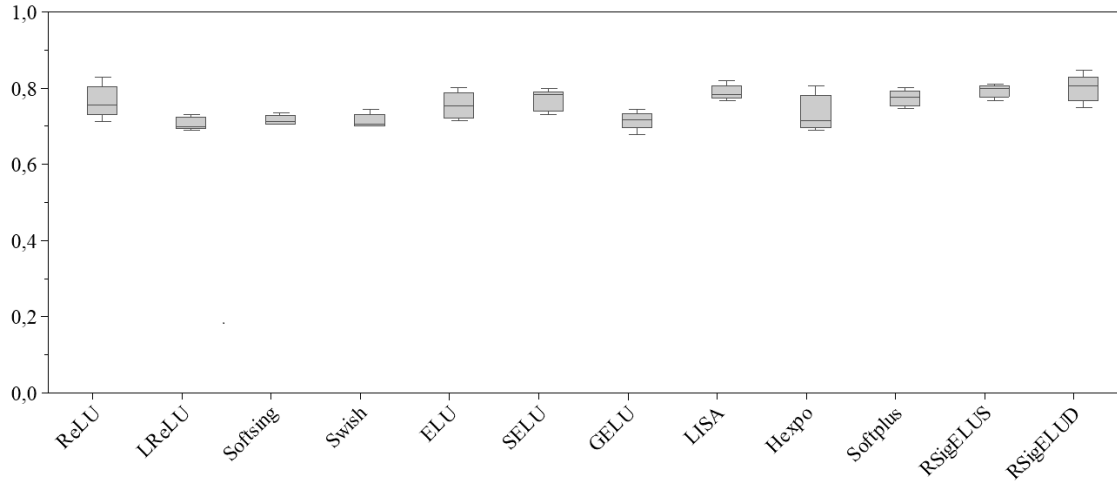
Alpha(α)	Scale(β)	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
0,4	0,2	40	0,0215	0,9925	0,8604	0,8411	885
0,45	0,2	40	0,0241	0,9918	0,7855	0,8375	892
0,5	0,2	40	0,0202	0,9928	0,8077	0,8407	908
0,55	0,2	40	0,0206	0,9933	0,7287	0,8409	918
0,6	0,2	40	0,0218	0,9924	0,7333	0,8330	932
0,65	0,2	40	0,0255	0,9913	0,8020	0,8381	928
0,7	0,2	40	0,0245	0,9918	0,6230	0,8357	925
0,75	0,2	40	0,0211	0,9920	0,7702	0,8394	918
0,8	0,2	40	0,0231	0,9920	0,7337	0,8372	921
0,85	0,2	40	0,0215	0,9922	0,7778	0,8358	932
0,9	0,2	40	0,0215	0,9925	0,6614	0,8392	882

Tablo 4.17. IMDB Movie veri kümesi üzerinde aktivasyon fonksiyonlarının ortalama başarı oranları

Aktivasyon	P1	P2	Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ Hata	Doğ Doğ.	Süre (s)
Relu [7]			40	0,0249	0,9915	0,7651	0,8324	895
Lrelu [3]	0,2		40	0,0234	0,9928	0,6792	0,8322	926
Softsing [34]			40	0,0221	0,9922	0,7312	0,8352	1148
Swish [62]			40	0,0183	0,9940	0,7099	0,8356	1152
Elu [9]	0,2		40	0,0235	0,9919	0,7464	0,8382	1147
Selu [8]	0,1		40	0,0223	0,9923	0,7727	0,8305	1134
Gelu [59]			40	0,0245	0,9916	0,7032	0,8408	1118
LISA [12]	0,15		40	0,0216	0,9926	0,7588	0,8377	1142
Hexpo [64]			40	0,0218	0,9926	0,7453	0,8370	936
Softplus [66]			40	0,0271	0,9914	0,7989	0,8203	1169
Önerilen RSigELUS	0,75		40	0,0204	0,9932	0,7903	0,8448	820
Önerilen RSigELUD	0,4	0,2	40	0,0215	0,9925	0,8604	0,8411	885

Tablo 4.15, 4.16 ve 4.17’de IMDB Movie veri kümesi için önerilen tek ve çift parametrelili RSigELU aktivasyon fonksiyonları ve diğer aktivasyon fonksiyonlarına göre başarı ve hata oranları gözlemlenmektedir. Sonuçlar incelendiğinde en iyi doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile RSigELUS aktivasyon fonksiyonu $\alpha=0,75$ parametre değeriyle, sırasıyla, 0,8448 doğruluk ve 0,7903 hata sonuçları olduğu görülmektedir. Ardından, önerilen RSigELUD aktivasyon fonksiyonu $\alpha=0,4$ ile $\beta=0,2$ değerleri ile doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) sırasıyla 0,8411 doğruluk ve 0,8604 hata elde edilmiştir. Önerilen aktivasyon fonksiyonlarının dışında en iyi başarı oranı GELU aktivasyon fonksiyonunda olduğu görülmektedir. GELU aktivasyon fonksiyonu için en iyi ile doğrulama doğruluk

(Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri sırasıyla 0,8408 doğruluk ve 0,7032 hata olduğu görülmektedir.



Şekil 4.9. IMDB Movie veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama hata (Doğ. Hata) sonuçları

Şekil 4.9, IMDB Movie veri kümesi için aktivasyon fonksiyonu doğrulama hata değerlerini göstermektedir. Şekilde görülebileceği gibi, önerilen RSigELUS aktivasyon fonksiyonunun ortalama doğrulama hata değeri, LISA aktivasyon fonksiyonundan daha küçük ve diğer aktivasyon fonksiyonlarından daha büyüktür. RSigELUD aktivasyon fonksiyonunun ortalama doğrulama hata değeri, diğer aktivasyon fonksiyonlarından daha büyüktür. Şekil 4.9'da görüldüğü gibi, RSigELUS ve RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre tutarlı sonuçlar vermiştir.

MNIST, Fasion MNIST, CIFAR-10 ve IMDB Movie veri kümelerinin farklı özellikleri vardır. Şekil 4.6, 4.7, 4.8 ve 4.9'da görülebileceği gibi, literatürdeki diğer aktivasyon fonksiyonlarının hiçbirisi tüm veri kümeleri için tutarlı doğrulama hata değerlerine sahip değildir. Bununla birlikte, önerilen aktivasyon fonksiyonları, doğrulama hataları açısından MNIST, Fashion MNIST, CIFAR-10 ve IMDB Movie veri kümeleri için diğer aktivasyon fonksiyonlarının çoğundan daha iyi performans göstermektedir.

4.5. Sonuçlar

Derin öğrenme modellerinin günlük hayatımızdaki kavramları öğrenebilmesinde çok önemli bir rolü mevcuttur. Aktivasyon katmanlarında ağa gelen giriş bilgisini işlemek ve bu girdilere karşılık gelen çıktıyı üretmek için aktivasyon fonksiyonları kullanılmaktadır. Derin öğrenme mimarileri yaygın olarak karmaşık problemlerle ilgilendiğinden doğrusal aktivasyon fonksiyonları işlevini yitirmektedir. Bundan dolayı derin öğrenme modellerinde doğrusal olmayan aktivasyon fonksiyonu ortaya çıkmaktadır. Çünkü sinir ağlarının değişkenler arasındaki karmaşık ve sürekli ilişkiyi öğrenmesi ve yaklaşık olarak tahmin etmesi için aktivasyon fonksiyonları temel oluşturur. Ancak yaygın olarak bilinen sigmoid ve hiperbolik tanjant aktivasyon fonksiyonlarında kaybolan gradyan problemi gözlemlenmektedir. Bu problemin üstesinden gelebilmek içinde ReLU aktivasyon fonksiyonu önerildi. Ancak, ReLU aktivasyon fonksiyonunun dezavantajı, negatif bölgeye gelen değeri sıfıra döndürmesidir. Ayrıca, ReLU aktivasyon fonksiyonunun sadece pozitif bölgede çalıştığından dolayı negatif giriş değeri olan nöronların ölmesine neden olmaktadır. ReLU'da meydana gelen negatif değer probleminin üstesinden gelinebilmesi için literatürde aktivasyon fonksiyonları geliştirilmiştir. Tez çalışmasında hem kaybolan gradyan hem de negatif bölge probleminin üstesinden gelebilmek için yeni tek ve çift parametrelili RSiGELU aktivasyon fonksiyonları önerilmiştir. Önerilen tek ve çift parametrelili RSiGELU aktivasyon fonksiyonları, MNIST, Fashion MNIST, CIFAR-10 ve IMDB Movie veri kümeleri üzerinde derin sinir ağları mimarileri kullanılarak diğer aktivasyon fonksiyonlarına göre daha iyi başarı oranını elde edildiği görülmektedir.

5. BÖLÜM

KAF+RSigELU: DERİN SİNİR AĞLARI İÇİN DOĞRUSAL OLMAYAN VE ÇEKİRDEK TABANLI AKTİVASYON FONKSİYONLARI

5.1. Giriş

Giriş bilgilerini işleyerek, işlenmiş olan bu giriş bilgilerine karşılık gelen çıktıyı üretmek amacıyla aktivasyon fonksiyonları önerilmiştir. Derin öğrenmede aktivasyon fonksiyonları, hızlı öğrenmenin gerçekleştirilmesi ve yüksek doğruluk sonuçları elde edilmesi açısından önemlidir.

Derin öğrenme mimarilerinde, geri yayılım algoritması kullanılarak hesaplanan ağırlıkların bir sonraki katmana aktarılıp aktarılmayacağına karar vermek için aktivasyon fonksiyonları kullanılır [17]. Ek olarak, derin öğrenme mimarilerinde aşırı öğrenmeyi önlemek, doğruluk performansını artırmak ve hesaplama maliyetini düşürmek için aktivasyon fonksiyonları kullanılmaktadır [17, 31]. Aktivasyon fonksiyonları doğru seçilmemişse derin öğrenme mimarilerinin öğrenme süreci sınırlı ve yavaş olabilir. Aktivasyon fonksiyonları, öğrenme sürecinde negatif bölgeden kaynaklanan önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama problemleri ile karşı karşıya kalmaktadır [14]. Negatif bölge problemi, negatif değer olması durumunda model çıktılarının sıfıra atanmasıdır. Negatif çıktılar türevlenemeyeceğinden öğrenme sürecinde bu değerler göz ardı edilerek negatif bölge problemi meydana gelir. Önyargı kayması problemi, pozitif bölgelerde çalışan algoritmalarda negatif değerlerin pozitif değerlere indirgenmesidir. Bu durumda fonksiyonun çıkış değerlerinin ortalamasında negatif sonuç alınamaz. Önyargı

kaymasından dolayı, modelin yakınsama hızı ve performansı doğrudan etkilenmektedir. [16]. Uyarlanamama problemi, önyargı kayması olduğunda eğitim sırasında modelin esnek bir şekilde çalışmasına izin verilmemesidir. Bu durumda, aktivasyon fonksiyonlarının doğrusal olmayan temsil kabiliyeti zayıflatılmaktadır [14]. Kaybolan gradyan probleminde sigmoid ve hiperbolik tanjant aktivasyon fonksiyonları belirli bir eşik değerlerinde türev işleminin sonucu için 0'a yaklaşır ve öğrenmeyi sonlandırarak başarısız bir performansa elde ettirmektedir [120, 121]. Uyarlanamama problemi, önyargı kayması olduğunda eğitim sırasında modelin esnek bir şekilde çalışmasına izin verilmemesidir. Bu durumda, aktivasyon fonksiyonlarının doğrusal olmayan temsil kabiliyeti zayıflatılmaktadır [14]. Esnek olmama problemi, aktivasyon fonksiyonları üzerinde negatif değerlerin göz ardı edilmesiyle, modelin ifade gücünü ve performansını zayıflatmaktadır [14]. Derin öğrenmede, önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama problemlerin üstesinden gelmek için en çok kullanılan aktivasyon fonksiyonları ReLU ve türevleridir [5, 122, 123]. Tez çalışmasında, RELU, sigmoid ve ELU aktivasyon fonksiyonlarının kombinasyonu ile RSigELUS ve RSigELUD aktivasyon fonksiyonlarını önerilmiştir [28]. Önerilen RSigELUS ve RSigELUD aktivasyon fonksiyonları, kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmektedir. Ayrıca, önerilen RSigELU aktivasyon fonksiyonları pozitif, negatif ve doğrusal (lineer) aktivasyon bölgelerinde aktif çalışabilmektedir [28].

Çekirdek tabanlı aktivasyon fonksiyonu (KAF), parametrik olmayan bir aktivasyon fonksiyonudur [17]. KAF, derin sinir ağı tasarımının basitleştirilmesine yol açan az sayıda gizli katman ile daha yüksek başarı oranları elde edilebilir. Böylece ağı boyutu, aktivasyon nöron karmaşıklığı ve derin sinir ağlarının aşırı öğrenme riski azaltılabilir.

Literatürdeki aktivasyon fonksiyonları sabit, eğitilebilir ve parametrik olmayan aktivasyon fonksiyonları olmak üzere üçe ayrılabilir. Literatürde, kaybolan gradyan probleminin üstesinden gelmek için sabit ve eğitilebilir aktivasyon fonksiyonları önerilmiştir [17, 124-126]. ReLU, LReLU, ELU, SELU, DReLU, FELU ve RSigELU gibi sabit aktivasyon fonksiyonları, parametreleri sabit olan ve model eğitimi sırasında parametreleri güncellenmeyen fonksiyonlardır [3, 7, 8, 28, 60-62]. Bundan dolayı, literatürde önerilen aktivasyon fonksiyonlarında öğrenme süreci yavaşlayabilir ve sınırlı öğrenme gerçekleşebilir. ALISA, PELU, PRELU, Parametric Flatten-T Swish (PFTS) ve

Mexican ReLU (MeLU) gibi eğitilebilir aktivasyon fonksiyonları, model eğitimi sırasında her yinelemede, parametreleri her bir nörona göre güncelleyen fonksiyonlardır [10, 12-15].

Bu tez çalışmasında, KAF [17] ve RSigELU [28] aktivasyon fonksiyonlarının birleşimi ile çekirdek tabanlı KAF+RSigELUS ve KAF+RSigELUD yeni ve hibrit aktivasyon fonksiyonları önerilmiştir. KAF+RSigELU aktivasyon fonksiyonu içerisinde, çekirdek fonksiyonu olarak gauss kullanılmaktadır. Gauss fonksiyonu, düşük hesaplama maliyeti ve yüksek başarı oranı elde edebilme gibi özellikleri ile ön plana çıkmaktadır [17]. KAF, ReLU ve sigmoid aktivasyon fonksiyonlarının genişletilmiş bir versiyonunu ifade etmektedir [37]. KAF, öğrenme sürecinde, negatif bölgeden kaynaklanan önyargı kayması, kaybolan gradyan, uyarılanama ve esnek olmama problemleri ile karşı karşıya kalmaktadır. Mevcut problemlerin üstesinden gelmek için, önerilen yeni hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları pozitif, negatif ve doğrusal aktivasyon bölgelerinde etkilidir. Önerilen KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının performans değerlendirmeleri MNIST [22], Fashion MNIST [23] ve CIFAR-10 [24] benchmark veri kümeleri üzerinde yapılmıştır. Deneysel değerlendirmeler, önerilen aktivasyon fonksiyonlarının mevcut problemlerin üstesinden geldiğini ve ReLU, LReLU, ELU, PReLU ve düz KAF aktivasyon fonksiyonlarından daha iyi performans göstermektedir. Önerilen hibrit KAF + RSigELU aktivasyon fonksiyonları, basit ESA mimarileri kullanılarak yüksek başarı oranı elde etmiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir ve revizyon aşamasındadır.

5.2. Yöntem

Tez çalışmasında, kullanılan ESA hakkındaki bilgi Bölüm 1.1’de verilmiştir. Önerilen KAF+RSigELU aktivasyon fonksiyonları, KAF aktivasyon fonksiyonunda belirlenen farklı ESA mimarileri ile deneysel değerlendirmeleri gerçekleştirilmiştir [17].

5.2.1. MNIST ve Fashion MNIST Veri Kümeleri için Kullanılan ESA Mimarisi

Tez çalışmasında, MNIST ve Fashion MNIST için basit ESA mimarisi olarak üç evrişim katmanı ile katman içerisinde filtre sayısı 50, filtre boyutu 5×5 kullanılmıştır. KAF için

kullanılan nöron başına parametre sayısı $D=32$ olarak belirlenmiştir. Bu işlemden sonra özellik haritası matris boyutunu azaltmak amacıyla maksimum havuzlama katmanı kullanılır ve filtre boyutu 3×3 olarak kullanılmıştır. Tam bağlantılı katman kullanılarak önceki katman ile tüm nöronlar arasında bağlantı sağlanması için 512 nöron kullanılmıştır. Sınıflandırma işleminin başarılı şekilde gerçekleştirilebilmesi için softmax katmanı kullanılmıştır [33, 44, 62, 127]. ESA mimarisinde kayıp fonksiyonu olarak kategorik_çapraz_antropi (categorical_crossentropy) kullanılmış ve optimizör için adam fonksiyonu tercih edilmiştir.

5.2.2. CIFAR-10 Veri Kümesi için Kullanılan ESA Mimarisi

Çalışmada, CIFAR-10 için basit ESA mimarisi olarak evrişim katmanı ile katman içerisinde filtre sayısı 150, filtre boyutu 5×5 ve adım sayısı 1 olarak kullanılmıştır. Adım, filtrenin giriş verisinin etrafında kaç adım kayacağını belirlemek için kullanılır. Evrişim işleminden sonra, aktivasyon katmanı olarak yeni hibrit KAF+RSigELU aktivasyon fonksiyonları kullanılmıştır. KAF için kullanılan nöron başına parametre sayısı $D=32$ olarak belirlenmiştir. Bu işlemden sonra özellik haritası matris boyutunu azaltmak amacıyla maksimum havuzlama katmanı kullanılır ve filtre boyutu 3×3 ve adım sayısı 2 olarak uygulanmıştır. Havuzlama işleminin ardından, ağ seyrekleştirilerek ezberlemenin önüne geçebilmek için seyrelme oranı 0,25 kullanılmıştır. Elde edilen verileri tam bağlantılı katmana hazırlamak için düzleştirici (flatten) katmanı kullanılmıştır. Tam bağlantılı katman kullanılarak önceki katman ile tüm nöronlar arasında bağlantı sağlanması için 512 nöron tercih edilmiştir. Sınıflandırma işleminin başarılı şekilde gerçekleştirilebilmesi için softmax katmanı tercih edilmiştir [36, 127]. Kayıp fonksiyonu olarak ESA mimarisinde, kategorik_çapraz_antropi (categorical_crossentropy) ve optimizör için sgd fonksiyonu kullanılmıştır.

5.3. Önerilen Çekirdek Tabanlı KAF+RSigELU Aktivasyon Fonksiyonları

Aktivasyon fonksiyonları, veri kümesindeki karmaşık yapıların çözülmesi, hızlı öğrenmenin gerçekleştirilmesi ve yüksek performanslı sonuçlar elde edilmesi açısından önemlidir. Ağa gelen bilgiyi işlemek ve bu girdilere karşılık gelen çıktıyı üretmek için aktivasyon fonksiyonları tercih edilir. Bu tez çalışmasında, hibrit KAF+RSigELUS ve

KAF+RSigELUD aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonlarında gauss çekirdek fonksiyonu kullanılmıştır. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktiftir. KAF aktivasyon fonksiyonu Eşitlik 5.1 ve 5.2'deki gibidir [17].

$$KAF = \sum_{i=1}^D a_i K(s, d_i) \quad (5.1)$$

$$K(s, d_i) = \exp\{-\gamma(s - d_i)^2\} \quad (5.2)$$

Eşitlik 5.1'de, D değerinin seçimine bağlı olarak nöron başına parametre sayısı potansiyel olarak sınırsız büyüyebilir. D , kullanıcı tanımlı bir hiper parametredir. Stokastik bir yapıda, sözlük seçimi için herhangi bir yöntem uygulanmadıkça D 'nin eğitim yinelemelerinin sayısı ile doğrusal olarak büyüyeceği anlamına gelir. D değeri yardımıyla x ekseninde örnekleme yapılabilir [17]. Böylece model üzerinde öğrenme işlemini gerçekleştirecek alt eğitim verileri elde edilir. Elde edilen verilere göre öğrenme sürecinde yüksek düzeyde fayda sağlanabilir [128]. Eşitlik 5.2, gauss fonksiyonunu temsil etmektedir. Eşitlik 5.2'de γ uyarlamalı parametreler yardımıyla dolaylı olarak hareket etmesiyle yöntemin daha iyi davranması sağlanabilir. γ parametresi, çekirdek bant genişliğinin hiper-parametresi olarak adlandırılır. Çekirdek işlevi olarak, kullanılan veri kümeleri doğrusal olmadığından, kurulu modellerin başarısı için gauss çekirdek işlevi kullanılır. Böylece oluşturulan ağlarda ağın boyutu, aktivasyon nöron karmaşıklığı ve aşırı uyum riski azaltılır.

Bu çalışmada, her iki bölgenin (pozitif ve negatif) özelliklerini yakalamak için, sırasıyla, Eşitlik 5.3 ve 5.4'te gösterilen yeni RSigELUS ve RSigELUD aktivasyon fonksiyonları önerilmiştir. Denklemlerde α ve β parametreleri eğim katsayılarını ifade etmektedir. Pozitif bölge α , negatif bölge ise β ile kontrol edilir [39].

$$f(x) = \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \alpha(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.3)$$

$$f(x) = \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \beta(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.4)$$

Tek parametrelili RSigELUS aktivasyon fonksiyonu (Eşitlik 5.3) için eğim değişkeni aralığı $0 < \alpha < 1$ arasında değiştirilerek deneyler yapılır ve diğer tüm derin öğrenme mimari parametreleri sabit tutulur. Eşitliklerdeki x parametresi, aktivasyon fonksiyonunun girdisini temsil etmektedir. Çift parametrelili RSigELUD (Eşitlik 5.4'te) aktivasyon fonksiyonunda, β negatif eğim katsayısı $0 < \beta < 1$ aralığında değerler alabilir ve pozitif eğim katsayısı $0 < \alpha < 1$ aralığında değerler alabilir. RSigELUS ve RSigELUD aktivasyon fonksiyonlarının türevleri sırasıyla, Eşitlik 5.5 ve 5.6'da gösterilmiştir.

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \alpha e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.5)$$

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \beta e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.6)$$

KAF, derin sinir ağlarının tasarımının basitleştirilmesine yol açan az sayıda gizli katmanla daha yüksek başarılar neden olabilir. KAF, öğrenme sürecinde parametrelerde negatif bölgeden kaynaklanan önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama problemleri ile karşı karşıya kalmaktadır. Bu tez çalışmasında, bilinen problemlerin üstesinden gelmek için çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları, sırasıyla, Eşitlik 5.7 ve 5.8'de gösterildiği gibi önerilmiştir. KAF+RSigELU aktivasyon fonksiyonları, gauss çekirdek fonksiyonu

kullanılarak oluşturulmuştur. Bu nedenle oluşturulan ağlarda ağın boyutu, aktivasyon nöron karmaşıklığı ve aşırı öğrenme riski azaltılır. Önerilen KAF+RSigELU aktivasyon fonksiyonları, doğrusal olmayan veriler üzerinde öğrenme süreçleri gerçekleştirir ve pozitif, negatif ve doğrusal bölgelerde aktif olmaktadır.

$$f(x) = KAF + \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \alpha(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.7)$$

$$f(x) = KAF + \begin{cases} x \left(\frac{1}{1 + e^{-x}} \right) \alpha + x, & \text{eğer } 1 < x < \infty \\ x, & \text{eğer } 0 \leq x \leq 1 \\ \beta(e^x - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.8)$$

$$\frac{df(x)}{dx} = \sum_{i=1}^D a_i \frac{dK(s, d_i)}{ds} + \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \alpha e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.9)$$

$$\frac{df(x)}{dx} = \sum_{i=1}^D a_i \frac{dK(s, d_i)}{ds} + \begin{cases} \frac{-\alpha x}{(e^x + 1)^2} + \frac{\alpha x - \alpha}{(e^x + 1)} + \alpha + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \beta e^x, & \text{eğer } -\infty < x < 0 \end{cases} \quad (5.10)$$

Önerilen çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının türevleri, sırasıyla, Eşitlik 5.9 ve 5.10'da gösterilmiştir KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının türev alınma sonrası hem pozitif hem de negatif bölgelerde aktif olduğu görülmektedir. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarında gauss çekirdek fonksiyonu kullanılmıştır. Derin sinir ağlarında parametrelerin geri yayılım algoritması esas alındığından dolayı, önerilen aktivasyon fonksiyonlarının türev alınabilme özelliğine sahip olması önemlidir [28].

5.4. Deneyel Değerlendirme

Önerilen hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının MNIST, Fashion MNIST ve CIFAR-10 veri kümeleri üzerinde deneyel değerlendirmeleri yapılmıştır. Deneyel karşılaştırmaları için ReLU [7], LReLU [3], ELU [9], PReLU [10] ve düz KAF [17] aktivasyon fonksiyonları kullanılmıştır.

MNIST veri kümesi için aktivasyon fonksiyonlarının γ , α ve β optimum parametre değerlerini belirlemek için deneyler yapılmıştır. Düz KAF, KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının deneysel sonuçları, sırasıyla, Tablo 5.1, 5.2 ve 5.3'te verilmiştir. Tablo 5.4'te MNIST veri kümesi için yapılan deneylerin ReLU, LReLU, ELU, PReLU, düz KAF ve önerilen hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının ortalama performans sonuçları sunulmuştur. Gerçekleştirilen deneyler 5'er defa tekrarlanarak sonuçlar elde edilmiştir.

Tablo 5.1. MNIST veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranları

λ	Hata	Doğruluk	Doğ. Hata	Doğ. Doğ.
0,2	0,0123	0,9966	0,0412	0,9917
0,3	0,0038	0,9988	0,0452	0,9893
0,4	0,0037	0,9988	0,0405	0,9906
0,5	0,0070	0,9979	0,0372	0,9897

Tablo 5.2. MNIST veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları

α	λ	Hata	Doğruluk	Doğ. Hata	Doğ. Doğ.
0,1	0,2	0,0122	0,9964	0,0565	0,9906
0,2	0,2	0,0053	0,9984	0,0520	0,9919
0,3	0,2	0,0093	0,9973	0,0700	0,9902
0,4	0,2	0,0079	0,9976	0,0656	0,9907
0,5	0,2	0,0123	0,9965	0,0479	0,9902
0,1	0,3	0,0081	0,9974	0,0488	0,9911
0,2	0,3	0,0089	0,9973	0,0635	0,9891
0,3	0,3	0,0109	0,9969	0,0479	0,9895
0,4	0,3	0,0087	0,9975	0,0460	0,9920
0,5	0,3	0,0063	0,9980	0,0649	0,9878
0,1	0,4	0,0064	0,9968	0,0343	0,9922
0,2	0,4	0,0100	0,9972	0,0459	0,9917
0,3	0,4	0,0096	0,9973	0,0572	0,9886
0,4	0,4	0,0060	0,9981	0,0657	0,9868
0,5	0,4	0,0099	0,9970	0,0537	0,9902
0,1	0,5	0,0062	0,9982	0,0403	0,9915
0,2	0,5	0,0115	0,9966	0,0451	0,9905
0,3	0,5	0,0094	0,9972	0,0516	0,9910
0,4	0,5	0,0071	0,9981	0,0502	0,9899
0,5	0,5	0,0075	0,9978	0,0520	0,9907

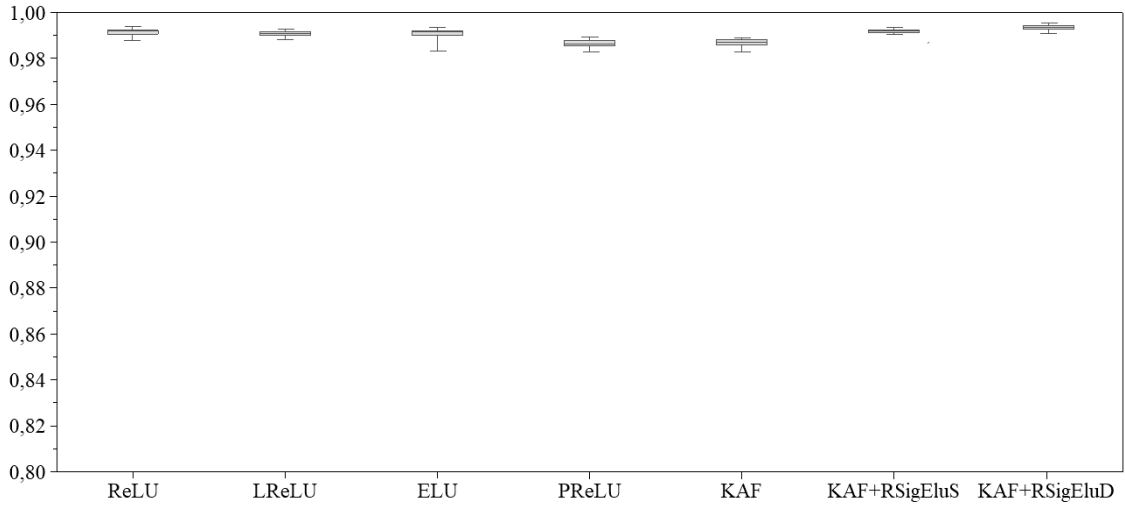
Tablo 5.3. MNIST veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları

α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
0,1	0,2	0,2	0,0470	0,9984	0,0445	0,9914
0,2	0,2	0,2	0,0081	0,9978	0,0446	0,9900
0,3	0,2	0,2	0,0072	0,9979	0,0647	0,9870
0,4	0,2	0,2	0,0074	0,9980	0,0621	0,9898
0,5	0,2	0,2	0,0059	0,9980	0,0484	0,9912
0,1	0,2	0,3	0,0090	0,09974	0,0576	0,9910
0,2	0,2	0,3	0,0103	0,9970	0,0453	0,9899
0,3	0,2	0,3	0,0059	0,9984	0,0393	0,9929
0,4	0,2	0,3	0,0070	0,9977	0,0741	0,9889
0,5	0,2	0,3	0,0102	0,9967	0,0505	0,9913
0,1	0,2	0,4	0,0113	0,9965	0,0546	0,9889
0,2	0,2	0,4	0,0074	0,9978	0,0445	0,9908
0,3	0,2	0,4	0,0057	0,9983	0,0451	0,9904
0,4	0,2	0,4	0,0073	0,9979	0,0682	0,9904
0,5	0,2	0,4	0,0080	0,9976	0,0453	0,9914
0,1	0,2	0,5	0,0078	0,9977	0,0463	0,9915
0,2	0,2	0,5	0,0059	0,9983	0,0419	0,9908
0,3	0,2	0,5	0,0116	0,9967	0,0614	0,9877
0,4	0,2	0,5	0,0062	0,9980	0,0477	0,9905
0,5	0,2	0,5	0,0081	0,9975	0,0525	0,9912

Tablo 5.4. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Metod	α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
ReLU				0,0022	0,9994	0,0487	0,9914
ELU	0,2			0,0064	0,9988	0,0836	0,9911
LReLU	0,001			0,0045	0,9990	0,0916	0,9908
PReLU				0,0109	0,9974	0,0862	0,9894
KAF			0,2	0,0051	0,9985	0,0469	0,9899
Önerilen KAF+RSigELUS	0,1		0,4	0,0102	0,9973	0,0492	0,9916
Önerilen KAF+RSigELUD	0,3	0,2	0,3	0,0064	0,9982	0,0481	0,9924

Tablo 5.1, 5.2, 5.3 ve 5.4’de, MNIST veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 40 tekrar ile KAF+RSigELUD aktivasyon fonksiyonunda $\gamma = 0,3$, $\alpha = 0,3$ ve $\beta = 0,2$ parametre değerleri ile 0,9924 doğruluk ve 0,0481 hata elde edilmiştir. Ayrıca düz KAF aktivasyon fonksiyonu için $\gamma = 0,2$ parametresi ile ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri sırasıyla 0,9899 doğruluk ve 0,0469 hata değerleri elde edildiği görülmüştür.



Şekil 5.1. MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları

Şekil 5.1, MNIST veri kümesi için ortalama doğrulama doğruluk sonuçlarını göstermektedir. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına kıyasla en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) değeri elde edilmiştir. Ayrıca, önerilen aktivasyon fonksiyonlarının doğrulama doğruluk oranı, KAF+RSigELUS için 0,9915 ile 0,9925 ve KAF+RSigELUD için 0,9910 ile 0,9930 arasındadır. Bu durum önerilen çekirdek tabanlı hibrit aktivasyon fonksiyonlarının diğer aktivasyon fonksiyonlarına göre daha kararlı olduğunu göstermektedir. Kutu grafiklerinde, değerleri arasındaki büyük uzunluğu çok küçük olması ve uç değerler arasındaki farkın çok küçük olması algoritmaların kararlılığının bir göstergesidir [129]. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının kutu büyük boyutlarının diğer kutu büyük boyutlarından daha kısa olduğu, büyüklerin kutuya mesafesinin daha yakın olduğu ve ortanca değerinin kutunun ortasında olduğu görülmektedir. Ayrıca Şekil 5.1'de görüldüğü gibi KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının kutularının en düşük değerleri diğer aktivasyon fonksiyonlarının kutularının en düşük değerlerinin üzerindedir. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının kutularının en yüksek değerleri diğer aktivasyon fonksiyonlarının geri kalanının üzerindedir. Sonuç olarak, önerilen aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına kıyasla MNIST veri kümesinde daha kararlı sonuçlar elde edilmektedir. Ayrıca,

KAF+RSigELUS, KAF+RSigELUD, KAF ve LReLU aktivasyon fonksiyonlarının sonuçlarının simetrik olarak dağıldığı da görülmektedir [130].

Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının γ , α ve β optimum parametre değerlerini belirlemek için deneyler yapılmıştır. Fashion MNIST veri kümesi için düz KAF ve önerilen çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının deneysel sonuçları, sırasıyla, Tablo 5.5, 5.6 ve 5.7'de verilmiştir. Tablo 5.8'de Fashion MNIST veri kümesi için yapılan deneylerin ReLU, LReLU, ELU, PReLU, düz KAF ve KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının ortalama performans sonuçları sunulmuştur.

Tablo 5.5. Fashion MNIST veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranı

λ	Hata	Doğruluk	Doğ. Hata	Doğ. Doğ.
0,2	0,0204	0,9931	0,5777	0,9089
0,3	0,0094	0,9972	0,5813	0,8858
0,4	0,0096	0,9973	0,5203	0,8986
0,5	0,0252	0,9920	0,4825	0,8908

Tablo 5.6. Fashion MNIST veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları

α	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
0,1	0,2	0,0334	0,9883	0,5564	0,9095
0,2	0,2	0,0281	0,9905	0,6497	0,9082
0,3	0,2	0,0308	0,9895	0,6992	0,9066
0,4	0,2	0,0317	0,9899	0,6176	0,9034
0,5	0,2	0,0403	0,9861	0,5957	0,9047
0,1	0,3	0,0338	0,9886	0,5821	0,9078
0,2	0,3	0,0247	0,9914	0,5946	0,9061
0,3	0,3	0,0292	0,9899	0,5907	0,9069
0,4	0,3	0,0361	0,9879	0,5937	0,9079
0,5	0,3	0,0307	0,9896	0,6673	0,8982
0,1	0,4	0,0330	0,9883	0,5465	0,9105
0,2	0,4	0,0293	0,9900	0,6334	0,9079
0,3	0,4	0,0257	0,9912	0,6153	0,9086
0,4	0,4	0,0265	0,9915	0,5410	0,9082
0,5	0,4	0,0366	0,9876	0,6094	0,9035
0,1	0,5	0,0263	0,9905	0,5740	0,9086
0,2	0,5	0,0236	0,9920	0,6106	0,9064
0,3	0,5	0,0332	0,9887	0,6048	0,9054
0,4	0,5	0,0285	0,9900	0,5592	0,9071
0,5	0,5	0,0319	0,9891	0,6147	0,9091

Tablo 5.7. Fashion MNIST veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları

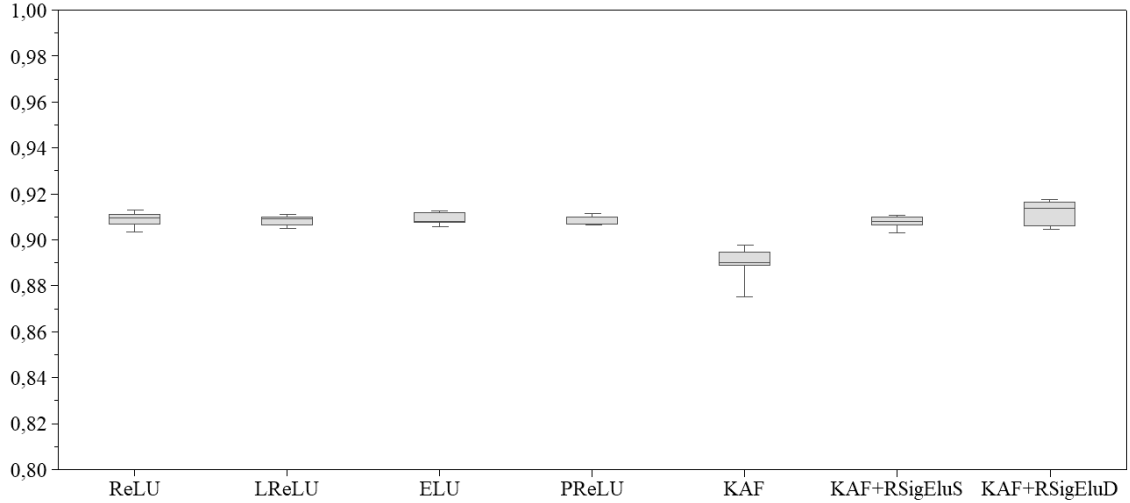
α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
0,1	0,2	0,2	0,0286	0,9902	0,5857	0,9016
0,2	0,2	0,2	0,0320	0,9890	0,5666	0,9083
0,3	0,2	0,2	0,0279	0,9903	0,6258	0,9104
0,4	0,2	0,2	0,0284	0,9905	0,6455	0,9108
0,5	0,2	0,2	0,0329	0,9884	0,6404	0,9053
0,1	0,2	0,3	0,0322	0,9895	0,6186	0,9062
0,2	0,2	0,3	0,0298	0,9892	0,6375	0,9063
0,3	0,2	0,3	0,0356	0,9875	0,5890	0,9054
0,4	0,2	0,3	0,0423	0,9861	0,6425	0,9037
0,5	0,2	0,3	0,0313	0,9892	0,6666	0,8987
0,1	0,2	0,4	0,0335	0,9882	0,6095	0,8991
0,2	0,2	0,4	0,0343	0,9875	0,5794	0,9054
0,3	0,2	0,4	0,0293	0,9900	0,6265	0,9074
0,4	0,2	0,4	0,0293	0,9901	0,6220	0,9058
0,5	0,2	0,4	0,0265	0,9905	0,5612	0,9091
0,1	0,2	0,5	0,0228	0,9923	0,5383	0,9071
0,2	0,2	0,5	0,0223	0,9924	0,5381	0,9150
0,3	0,2	0,5	0,0225	0,9923	0,5994	0,9094
0,4	0,2	0,5	0,0355	0,9880	0,6189	0,9022
0,5	0,2	0,5	0,0299	0,9900	0,5804	0,9090

Tablo 5.8. Fashion MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Method	α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
ReLU				0,0187	0,9942	0,6380	0,9102
ELU	0,2			0,0215	0,9928	0,7184	0,9084
LReLU	0,001			0,0209	0,9936	0,5942	0,9094
PReLU				0,0360	0,9872	0,5592	0,9081
KAF			0,2	0,0272	0,9911	0,5738	0,8933
Önerilen KAF+RSigELUS	0,4		0,4	0,0214	0,9930	0,6939	0,9064
Önerilen KAF+RSigELUD	0,2	0,2	0,5	0,0554	0,9932	0,6616	0,9103

Tablo 5.5, 5.6, 5.7 ve 5.8’de, Fashion MNIST veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 40 tekrar ile KAF+RSigELUD aktivasyon fonksiyonunda $\gamma = 0,5$, $\alpha = 0,2$ ve $\beta = 0,2$ parametre değerleri ile 0, 9103 doğruluk ve 0, 6616 hata elde edilmiştir. Ayrıca düz KAF aktivasyon fonksiyonu için $\gamma = 0,2$ parametresi ile ortalama doğrulama

doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri, sırasıyla, 0,8933 doğruluk ve 0,5738 hata değerleri elde edildiği görülmüştür.



Şekil 5.2. Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları

Şekil 5.2, Fashion MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk sonuçlarını göstermektedir. Önerilen hibrit KAF+RSigELUD aktivasyon fonksiyonu, diğer fonksiyonlara kıyasla en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) performansını vermiştir. Şekil 5.2’de görüldüğü gibi, KAF+RSigELUD aktivasyon fonksiyonunun ortalama doğrulama doğruluk değeri, ReLU, LReLU, ELU, PReLU, KAF ve KAF+RSigELUS aktivasyon fonksiyonlarından daha büyüktür. Ayrıca, KAF+RSigELUS ve KAF+RSigELUD fonksiyonlarının kutularındaki en düşük değerleri, KAF aktivasyon fonksiyonunun üzerinde ve diğer aktivasyon fonksiyonlarına yakındır. KAF+RSigELUD aktivasyon fonksiyonunun, diğer aktivasyon fonksiyonlarından daha iyi bir sonuç verdiği görülmektedir. Ayrıca KAF+RSigELUS, ReLU, LReLU, ELU ve PReLU aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına kıyasla Fashion MNIST veri kümesinde tutarlı sonuçlar vermektedir. Şekil 5.2’de, KAF+RSigELUS ve ReLU aktivasyon fonksiyonlarının sonuçlarının simetrik olarak dağıldığı görülmektedir. KAF aktivasyon fonksiyonu incelendiğinde, özellikle alt bıyıkların kutudan çok uzak olduğu ve ortanca değerin ortadan uzak olduğu

görülmektedir. Dolayısıyla, KAF aktivasyon fonksiyonunun Fashion MNIST veri kümesi için farklı sonuçlar verdiği ve kararsız bir davranış gösterdiği söylenebilir.

CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının γ , α ve β optimum parametre değerlerini belirlemek için deneyler yapılmıştır. CIFAR-10 veri kümesi için düz KAF ve önerilen çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının deneysel sonuçları, sırasıyla, Tablo 5.9, 5.10 ve 5.11'de verilmiştir. Tablo 5.12'de CIFAR-10 veri kümesi için yapılan deneylerin ReLU, LReLU, ELU, PReLU, düz KAF ve önerilen KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının ortalama performans sonuçları sunulmuştur.

Tablo 5.9. CIFAR-10 veri kümesi üzerinde KAF aktivasyon fonksiyonunun başarı oranı

λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
0,2	-	-	-	-
0,3	-	-	-	-
0,4	0,5108	0,8171	1,0238	0,6835
0,5	0,1390	0,9551	1,0971	0,7123

Tablo 5.10. CIFAR-10 veri kümesi üzerinde KAF+RSigELUS aktivasyon fonksiyonunun başarı oranları

α	λ	Hata	Doğruluk	Doğ. Hata	Doğ. Doğ.
0,1	0,2	0,1753	0,9401	1,0071	0,7299
0,2	0,2	0,2074	0,9262	0,9166	0,7445
0,3	0,2	0,2235	0,9225	0,9376	0,7447
0,4	0,2	0,2300	0,9183	0,9581	0,7379
0,5	0,2	0,2555	0,9102	0,9778	0,7275
0,1	0,3	0,1761	0,9398	0,9935	0,7217
0,2	0,3	0,1754	0,9399	0,9665	0,7316
0,3	0,3	0,1999	0,9303	0,9108	0,7494
0,4	0,3	0,2129	0,9267	0,9320	0,7369
0,5	0,3	0,2204	0,9225	0,8949	0,7499
0,1	0,4	0,3037	0,8934	1,1596	0,7344
0,2	0,4	0,3311	0,8864	1,0720	0,7335
0,3	0,4	0,2901	0,8990	1,0370	0,7509
0,4	0,4	0,2266	0,9199	0,9428	0,7391
0,5	0,4	0,2184	0,9225	1,5901	0,6170
0,1	0,5	0,1512	0,9512	0,9169	0,7446
0,2	0,5	0,1641	0,9466	0,9614	0,7321
0,3	0,5	0,1742	0,9410	0,9910	0,7321
0,4	0,5	0,2032	0,9289	0,8859	0,7480
0,5	0,5	0,2068	0,9270	0,8893	0,7475

Tablo 5.11. CIFAR-10 veri kümesi üzerinde KAF+RSigELUD aktivasyon fonksiyonunun başarı oranları

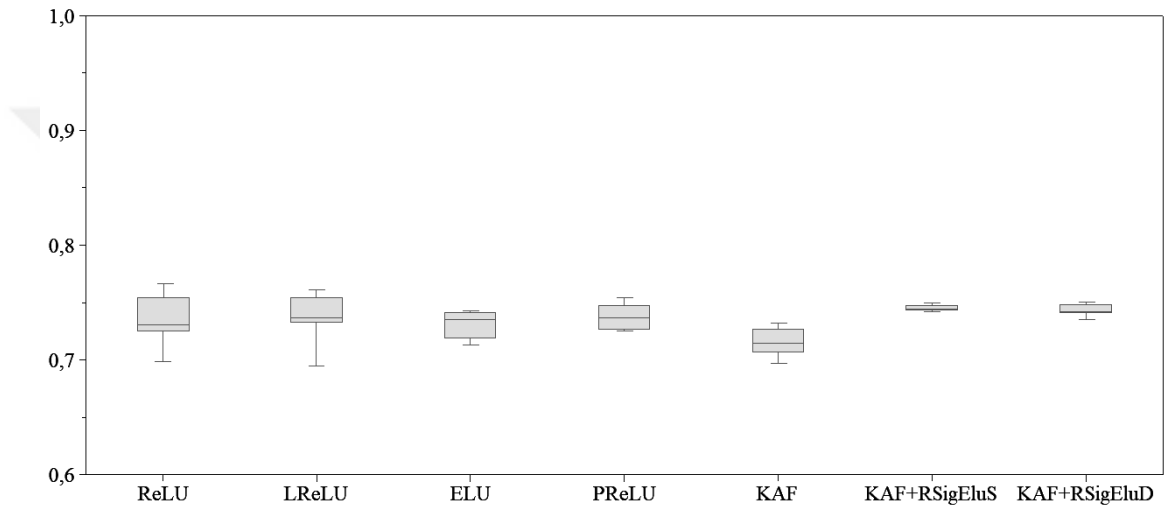
α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
0,1	0,2	0,2	0,1977	0,9325	0,8942	0,7488
0,2	0,2	0,2	0,2080	0,9269	0,9054	0,7436
0,3	0,2	0,2	0,1904	0,9331	0,9370	0,7501
0,4	0,2	0,2	0,2055	0,9285	1,0414	0,7154
0,5	0,2	0,2	0,2326	0,9189	0,9429	0,7430
0,1	0,2	0,3	0,1604	0,9463	1,0293	0,7020
0,2	0,2	0,3	0,1906	0,9355	0,9163	0,7410
0,3	0,2	0,3	0,1965	0,9330	0,9157	0,7441
0,4	0,2	0,3	0,1860	0,9360	1,0773	0,7120
0,5	0,2	0,3	0,2097	0,9280	0,9865	0,7273
0,1	0,2	0,4	0,1587	0,9473	0,9553	0,7414
0,2	0,2	0,4	0,1765	0,9406	0,9049	0,7483
0,3	0,2	0,4	0,1716	0,9420	0,9532	0,7463
0,4	0,2	0,4	0,1864	0,9358	1,0059	0,7348
0,5	0,2	0,4	0,2019	0,9297	0,9722	0,7243
0,1	0,2	0,5	0,1609	0,9463	0,9958	0,7256
0,2	0,2	0,5	0,1667	0,9455	1,0186	0,7179
0,3	0,2	0,5	0,1725	0,9422	1,0432	0,7103
4	0,2	0,5	0,1675	0,9418	0,9515	0,7446
0,5	0,2	0,5	0,1781	0,9389	0,9880	0,7153

Tablo 5.12. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranı

Metod	α	β	λ	Hata	Doğruluk	Doğ_Hata	Doğ_Doğ.
ReLU				0,2798	0,9041	0,9066	0,7366
ELU	0,2			0,2556	0,9133	0,9208	0,7313
LReLU	0,00			0,2700	0,9070	0,8762	0,7385
	1						
PReLU				0,2758	0,9045	0,8575	0,7411
KAF			0,5	0,3465	0,8817	0,9256	0,7166
Önerilen KAF+RSigELUS	0,2		0,4	0,1033	0,9656	1,0829	0,7517
Önerilen KAF+RSigELUD	0,3	0,2	0,2	0,0703	0,9767	1,1159	0,7490

Tablo 5.9, 5.10, 5.11 ve 5.12’de, CIFAR-10 veri kümesi üzerinde gerçekleştirilen deneylerde en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 40 tekrar ile KAF+RSigELUS aktivasyon fonksiyonunda $\gamma=0,4$ ve $\alpha=0,2$ parametre değerleri ile 0,7550 doğruluk ve 1,1043 hata elde edilmiştir. Ayrıca düz KAF aktivasyon fonksiyonu için $\gamma = 0,5$ parametresi ile ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri, sırasıyla 0,7123 doğruluk ve

1,0971 hata değerleri elde edilmiştir. KAF'da $\lambda = 0,2$ ve $\lambda = 0,3$ parametre değerlerinde yetersiz öğrenme gerçekleştirmektedir. Tablo 5.9 incelendiğinde negatif bölge probleminde kaynaklı olarak öğrenme işleminin gerçekleşmediği ve bazı nöronlar öldüğünden önyargı kayması olduğu görülmektedir. KAF aktivasyon fonksiyonu, negatif değerlerin pozitif değere dönüştürmektedir ve bundan dolayı negatif değerleri keşfedememesinden fonksiyonunun esnek bir yapıda olmadığı görülmektedir.



Şekil 5.3. CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (Doğ_Doğ.) sonuçları

Şekil 5.3, CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk sonuçlarını göstermektedir. Önerilen çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına kıyasla en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) performansını vermiştir. Şekil 5.3'de, önerilen aktivasyon fonksiyonlarında diğer aktivasyon fonksiyonlarına göre daha küçük bir değişiklik olmaktadır. Şekil 5.3'de, KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının ortalama doğrulama doğruluk değeri ReLU, LReLU, ELU, PReLU ve KAF aktivasyon fonksiyonlarından daha büyüktür. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının kutu çizim uzunluklarının diğer kutu çizim uzunluklarına göre daha kısa olduğu ve bıyıkların kutuya mesafesinin daha yakın olduğu görülmektedir (Şekil 5.3). Ayrıca Şekil 5.3'de KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarındaki kutuların en düşük değerleri diğer

aktivasyon fonksiyonlarının kutularının en yüksek yakın olduğu görülmektedir. Sonuç olarak KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının diğer aktivasyon fonksiyonlarına göre CIFAR-10 veri kümesinde daha kararlı sonuçlar verdiği görülmektedir. Diğer aktivasyon fonksiyonları incelendiğinde özellikle alt bıyıkların kutudan çok uzak olduğu ve ortanca değerin ortadan uzak olduğu görülmektedir. Bundan dolayı, diğer aktivasyon fonksiyonlarının CIFAR-10 veri kümesi için farklı sonuçlar verdiği ve kararsız bir davranış sergilediği söylenebilir.

Tez çalışmasında, çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları önerilmiştir. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının, MNIST, Fasion MNIST ve CIFAR-10 veri kümeleriyle gerçekleştirilen deneysel değerlendirmelerde, diğer aktivasyon fonksiyonlarına göre en iyi sonuçları verdiği, sırasıyla, Tablo 5.4, 5.8 ve 5.12'de görülmektedir. Önerilen hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları, pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktiftir ve mevcut problemlerden önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama probleminin üstesinden başarıyla gelebilmektedir. KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonlarının kutu boyutunun diğer kutu boyutlarına göre daha kısa olduğu görülmektedir.

5.5. Sonuçlar

Çekirdek tabanlı aktivasyon fonksiyonu (KAF), öğrenme sürecinde parametrelerde negatif bölgeden kaynaklanan önyargı kayması, kaybolan gradyan, uyarlanamama ve esnek olmama problemleri ile karşı karşıya kalmaktadır. Bu tez çalışmasında, mevcut problemlerin üstesinden gelmek için çekirdek tabanlı hibrit KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları önerilmiştir. KAF+RSigELU aktivasyon fonksiyonları MNIST, Fasion MNIST ve CIFAR-10 veri kümeleri üzerinde deneysel değerlendirmeler yapılmıştır ve mevcut problemlerin üstesinden gelerek ReLU, LReLU, ELU, PReLU ve düz KAF aktivasyon fonksiyonlarından daha iyi sonuçlar vermektedir. Aktivasyon fonksiyonlarının kutu grafiğindeki bıyık uzunlukları kısa ve ortanca değer kutu grafiğinin ortasına yakındır. Kutu grafiklerinde düşük doğruluk değerleri diğer aktivasyon fonksiyonlara göre kutuya yakın ve üst doğruluk değeri ise en yüksek değere yakın olduğu görülmektedir. Bu bilgiler doğrultusunda kutu grafikleri incelendiğinde

önerilen aktivasyon fonksiyonlarının diğer aktivasyon fonksiyonlarına göre daha kararlı ve tutarlı sonuçlar verdiği görülmektedir.



6. BÖLÜM

P+RSigELU: DERİN ÖĞRENME ALGORİTMARI İÇİN EĞİTİLEBİLİR PARAMETRİK AKTİVASYON FONKSİYONLARI

6.1. Giriş

Derin öğrenmede yaygın olarak kullanılan ESA modeli, konvolüsyon, havuzlama, seryeltme ve aktivasyon fonksiyonu gibi katmanlardan oluşmaktadır. Aktivasyon fonksiyonları, derin öğrenme mimarilerinin önemli bir parçasıdır. Ayrıca aktivasyon fonksiyonları, girdileri bir sonraki nörona göndererek, kullanılan nöronun aktif olup olmayacağına karar vermede önemli bir rol oynamaktadır. Ek olarak, derin öğrenmede aktivasyon fonksiyonları hızlı öğrenmenin gerçekleştirilmesi ve yüksek doğruluk sonuçları elde edilmesi açısından önemlidir. Literatürde bilinen kaybolan gradyan ve negatif bölge problemlerine çözüm üretebilmek için LReLU, ELU, RSigELU, GeLU, FeLU, PReLU, DReLU, PELU gibi sabit parametrelili ve eğitilebilir aktivasyonları önerilmiştir [7, 8, 10, 13, 28, 60, 61, 62].

Derin öğrenme modellerinin eğitim sırasında optimal parametre değerlerini elde edebilmesi için geri yayılım algoritmasını kullanmaktadır. Geri yayılım algoritması kullanılarak model üzerindeki ağırlıklar ve önyargı değerleri her tekrarda bütün nöronlar üzerinde güncellenir. Bu sayede, modelin eğitim süreci kesintisiz olarak devam edebilmektedir. Eğitilebilir aktivasyon fonksiyonları, eğitim sırasında her bir nörona en uygun değerleri elde etmek için aktivasyon fonksiyonunun eğim parametreleri sürekli olarak günceller. Literatürde, ALISA, PELU, PReLU, PFTS ve MeLU gibi eğitilebilir aktivasyon fonksiyonları önerilmiştir [10, 12-15].

Tez çalışmasında, ESA, her katmanında öğrenme işlemini gerçekleştirebilmesi için eğitilebilir P+RSigELU aktivasyon fonksiyonları önerilmiştir. Böylece önerilen P+RSigELU aktivasyon fonksiyonları eğitim sürecini sürekli olarak devam ettirebilmektedir. RSigELU aktivasyon fonksiyonlarında en ideal eğim (α ve β) parametrelerini belirlemek için çok sayıda deneme yapılmaktadır. Bundan dolayı, gerçek zamanlı problemler üzerinde uygulamalar yapılırken en uygun değeri bulmak çok zaman almaktadır. Ancak önerilen P+RSigELU aktivasyon fonksiyonları, ilk çalıştırmada en uygun eğim (α ve β) parametrelerini belirleyerek öğrenme işlemini en kısa sürede gerçekleştirir. Önerilen tek ve çift eğitilebilir parametrelili P+RSigELU aktivasyon fonksiyonları MNIST [23], CIFAR-10 [24] ve CIFAR-100 [24] veri kümeleri üzerinde test edilmiş ve literatürde bilinen diğer PReLU, PELU ve ALISA aktivasyon fonksiyonlarından daha iyi sonuçlar verdiği görülmüştür. Ayrıca ReLU ve sigmoid aktivasyon fonksiyonlarında bulunan kaybolan gradyan ve negatif bölge problemlerin üstesinden gelinmiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir.

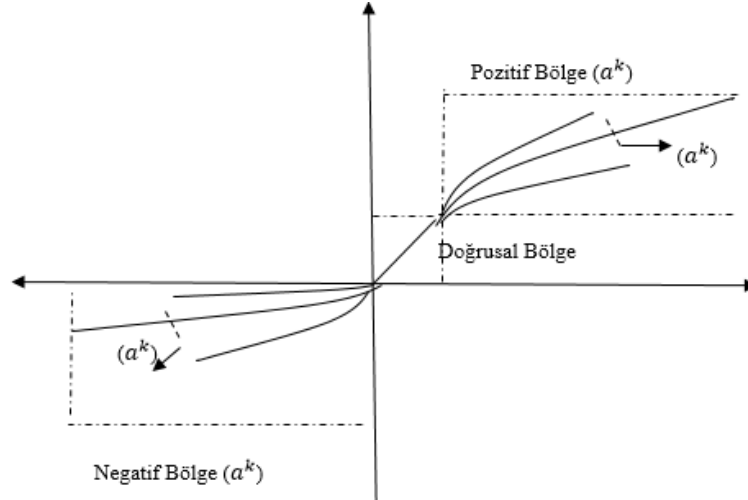
6.2. Yöntem

Tez çalışmasında VGG tabanlı ESA mimarisini, önerilen RSigELU aktivasyon fonksiyonlarını test etmek amacıyla kullanmıştır. Kullanılan ESA yöntemi hakkındaki bilgi Bölüm 1.1’de verilmiştir. Ayrıca, tez çalışmasında kullanılan VGG mimarisi Bölüm 4.2’de verilmiştir. Deneysel değerlendirmeler için kullanılan veri kümeleri hakkındaki bilgiler Bölüm 1.5.3’de verilmiştir.

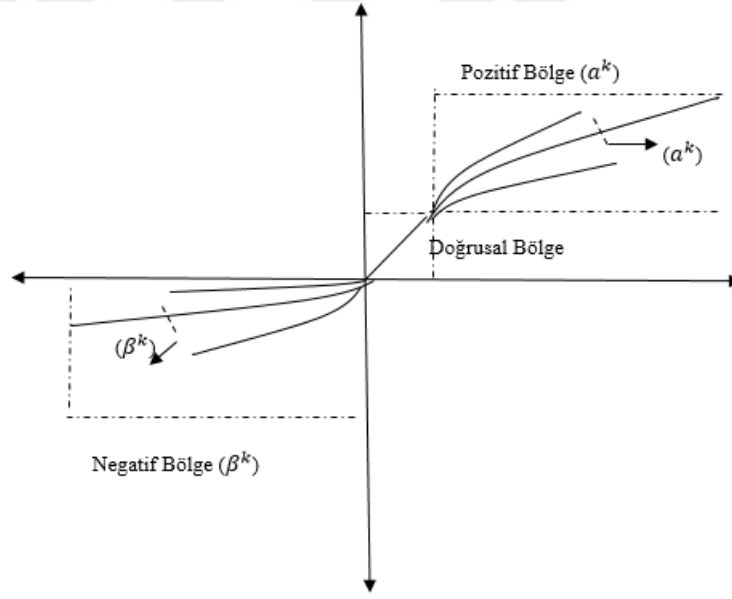
6.3. Önerilen Eğitilebilir P+RSigELU Aktivasyon Fonksiyonu

Tez çalışmasında, sabit parametrelili RSigELU aktivasyon fonksiyonlarını geliştirmek için yeni eğitilebilir tek parametrelili P+RSigELUS ve çift parametrelili P+RSigELUD aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonları, negatif, pozitif ve doğrusal bölgelerde aktif olarak çalışabilmektedir. Negatif ve pozitif bölgelerin özelliklerini yakalamak için tek ve çift eğim (α^k ve β^k) parametreleri kullanılmaktadır [28]. P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının davranışı Şekil 6.1’de gösterilmiştir. RSigELUS ve RSigELUD aktivasyon fonksiyonlarında α ve β eğim

parametrelerinin deęerleri sabit olarak belirlenerek deneyler yapılmaktadır [28]. Eęitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, her nörona en uygun parametre deęerini belirlemek için α ve β eęim parametreleri eęitilebilir hale getirilmesiyle elde edilebilmektedir. Eęitilebilir P+RSigELU aktivasyon fonksiyonlarında, pozitif ve negatif bölgelerde eęim parametreleri aktif şekilde eęitilebilmektedir. Tez alışmasında, önerilen aktivasyon fonksiyonlarının parametreleri dıřındaki kullanılan parametreler sabit tutularak deneyler gerekleřtirilmiřtir. Önerilen eęitilebilir P+RSigELU aktivasyon fonksiyonları, pozitif bölge için sigmoid ve ReLU fonksiyonlarıyla hibrit yapı gerekleřtirerek derin modellerdeki kaybolan gradyan problemi olmadan sürekli öęrenme gerekleřtirebilmektedir. Doğrusal bölge için doğrusal aktivasyon fonksiyonu kullanılmaktadır. Ayrıca, negatif bölge için de ELU aktivasyon fonksiyonu kullanılmaktadır. Önerilen eęitilebilir P+RSigELU aktivasyon fonksiyonlarının eęri (α^k ve β^k) eęilimi temel olarak PSigELU aktivasyon fonksiyonları ile tutarlıdır ve bundan dolayı önerilen aktivasyon fonksiyonları RSigELU aktivasyon fonksiyonlarının doğruluk avantajını deęiřtirmemesini saęlamaktadır. Eęitilebilir α^k ve β^k eęim parametrelerini belirlemek için alpha-beta bařlatıcısı (alpha-beta initializer) olarak HeNormal kullanılmıřtır [10]. HeNormal sayesinde önerilen aktivasyon fonksiyonlarının global minimuma hızlı ve verimli bir şekilde ulařmasına yardımcı olmaktadır. Bundan dolayı, derin öęrenme mimarilerinin daha hızlı ve verimli alışması saęlanmaktadır [10].



(a)



(b)

Şekil 6.1. Önerilen (a) P+RSigELUS ve (b) P+RSigELUD aktivasyon fonksiyonlarının davranışı

6.3.1. Eğitilebilir P+RSigELUS Aktivasyon Fonksiyonu

Bu bölümde, eğitilebilir tek parametrelili P+RSigELUS aktivasyon fonksiyonunu sunulmuştur. P+RSigELUS aktivasyon fonksiyonu Eşitlik 6.1'de tanımlanmıştır.

$$f(x) = \begin{cases} x^k \left(\frac{1}{1 + e^{-x^k}} \right) \alpha^k + x^k, & \text{eğer } 1 < x < \infty \\ x^k, & \text{eğer } 0 \leq x \leq 1 \\ \alpha^k (e^{x^k} - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (6.1)$$

Eşitlik 6.1, eğitilebilir P+RSigELUS aktivasyon fonksiyonu, pozitif, negatif ve doğrusal bölgelerde aktif olduğunu göstermektedir. Eşitlik 6.1'de, α^k pozitif bölge eğim katsayısını ve negatif bölge eğim katsayısını temsil etmektedir. Eşitlik 6.1'deki x^k aktivasyon fonksiyonunun girişini temsil etmektedir. Ayrıca k parametresi, doğrusal olmayan aktivasyon fonksiyonlarında eğim değerlerinin farklı nöronlarda değişeceğini bildirmektedir [67]. P+RSigELUS aktivasyon fonksiyonunun eğitilebilir eğim parametreleri, eğitim sürecinde model parametreleri ile birlikte eğitilir. P+RSigELUS aktivasyon fonksiyonu, Şekil 6.1 (a)'deki gibi davranış sergilemektedir.

Eğitilebilir P+RSigELUS aktivasyon fonksiyonunun türevi Eşitlik 6.2'de verilmiştir.

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha^k x^k}{(e^{x^k} + 1)^2} + \frac{\alpha^k x^k - \alpha^k}{(e^{x^k} + 1)} + \alpha^k + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \alpha^k e^{x^k}, & \text{eğer } -\infty < x < 0 \end{cases} \quad (6.2)$$

6.3.2. Eğitilebilir P+RSigELUD Aktivasyon Fonksiyonu

Bu bölümde, eğitilebilir çift parametrelili P+RSigELUD aktivasyon fonksiyonunu sunulmuştur. P+RSigELUD aktivasyon fonksiyonu Eşitlik 6.3'de tanımlanmıştır.

$$f(x) = \begin{cases} x^k \left(\frac{1}{1 + e^{-x^k}} \right) \alpha^k + x^k, & \text{eğer } 1 < x < \infty \\ x^k, & \text{eğer } 0 \leq x \leq 1 \\ \beta^k (e^{x^k} - 1), & \text{eğer } -\infty < x < 0 \end{cases} \quad (6.3)$$

Eşitlik 6.3, P+RSigELUD aktivasyon fonksiyonunun, pozitif, negatif ve doğrusal bölgelerde aktif olduğunu göstermektedir. Eşitlik 6.3'te, α^k pozitif bölge eğim katsayısını ve β^k negatif bölge eğim katsayısını temsil etmektedir. Ayrıca x^k aktivasyon fonksiyonunun girdisidir ve $f(x)$ aktivasyon fonksiyonunun çıktısıdır. α^k ve β^k eğim katsayısı her bir nöronda eğitilebilir parametre değerini almaktadır. Eğitilebilir P+RSigELUD aktivasyon fonksiyonu, Şekil 6.1 (b) 'deki gibi davranış sergilemektedir.

Eğitilebilir P+RSigELUD aktivasyon fonksiyonunun türevi Eşitlik 6.4'te verilmiştir.

$$\frac{df(x)}{dx} = \begin{cases} \frac{-\alpha^k x}{(e^{x^k} + 1)^2} + \frac{\alpha^k x - \alpha^k}{(e^{x^k} + 1)} + \alpha^k + 1, & \text{eğer } 1 < x < \infty \\ 1, & \text{eğer } 0 \leq x \leq 1 \\ \beta^k e^{x^k}, & \text{eğer } -\infty < x < 0 \end{cases} \quad (6.4)$$

6.4. DeneySEL Değerlendirme

Önerilen P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının MNIST [22], CIFAR-10 [24] ve CIFAR-100 [24] veri kümeleri üzerinde deneysel değerlendirmeleri sunulmuştur. Deneylerde, önerilen aktivasyon fonksiyonlarının deneysel sonuçlarını karşılaştırmak için PReLU [10], PELU [13] ve ALISA [12] aktivasyon fonksiyonları kullanılmıştır. Feneyler 5'er defa tekrarlanarak sonuçlar elde edilmiştir.

MNIST veri kümesi için önerilen eğitilebilir tek ve çift parametrelili P+RSigELU aktivasyon fonksiyonları ile PReLU [10], PELU [13] ve ALISA [12] aktivasyon fonksiyonlarının deneysel sonuçları Tablo 6.1, 6.2 ve 6.3'te verilmiştir.

Tablo 6.1. MNIST veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim_Hata	Eğitim_Doğ.	Doğ_Hata	Doğ_Doğ.
Tekrar 1	40	0,2864	0,9193	0,2160	0,9380
Tekrar 2	40	0,1861	0,9429	0,1562	0,9492
Tekrar 3	40	0,3235	0,9161	0,2086	0,9444
Tekrar 4	40	0,2309	0,9305	0,1745	0,9483
Tekrar 5	40	0,2239	0,9347	0,1475	0,9558
Ortalama		0,2501	0,9287	0,1805	0,9471

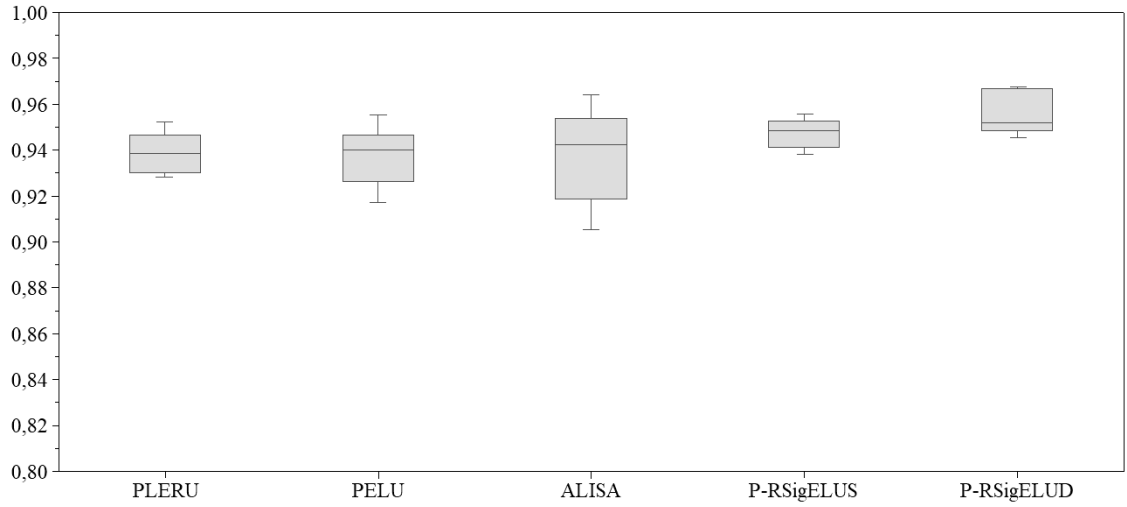
Tablo 6.2. MNIST veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim_Hata	Eğitim_Doğ.	Doğ_Hata	Doğ_Doğ.
Tekrar 1	40	0,2557	0,9246	0,1923	0,9453
Tekrar 2	40	0,2059	0,9335	0,1523	0,9517
Tekrar 3	40	0,1594	0,9546	0,1038	0,9674
Tekrar 4	40	0,1778	0,9459	0,1176	0,9658
Tekrar 5	40	0,2325	0,9314	0,1633	0,9518
Ortalama		0,2062	0,9380	0,1458	0,9564

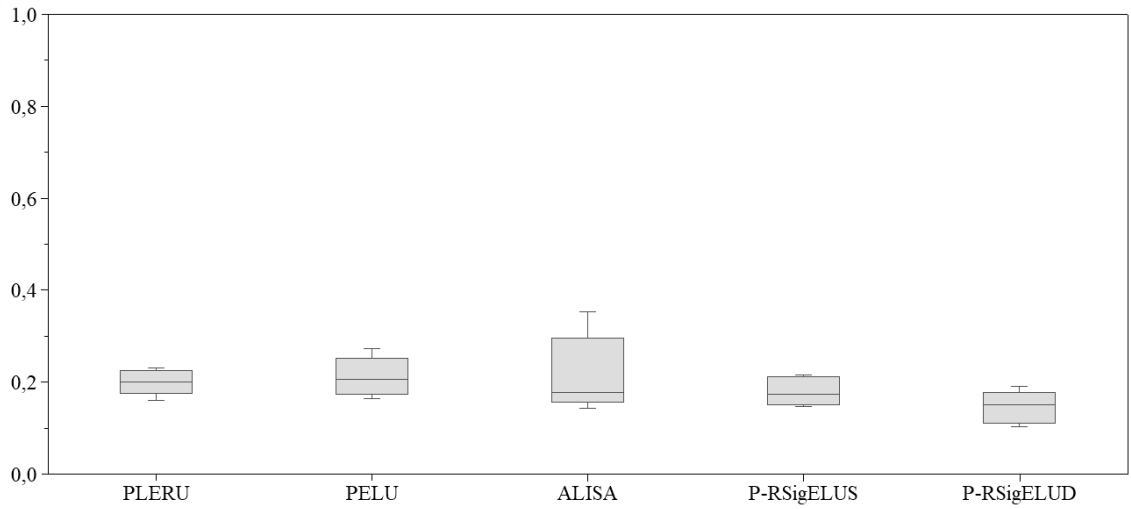
Tablo 6.3. MNIST veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	Eğitim_Hata	Eğitim_Doğ.	Doğ_Hata	Doğ_Doğ.
PReLU	0,2782	0,9139	0,2011	0,9383
PELU	0,2721	0,9187	0,2134	0,9362
ALISA	0,3018	0,9095	0,2275	0,9335
Önerilen P+RSigELUS	0,2501	0,9287	0,1805	0,9471
Önerilen P+RSigELUD	0,2062	0,9380	0,1458	0,9564

Tablo 6.1, 6.2 ve 6.3’de, MNIST veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile P+RSigELUD aktivasyon fonksiyonunda 0, 9564 doğruluk ve 0, 1458 hata elde edilmiştir. Eğitilebilir P+RSigELUS içinde ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri, sırasıyla, 0,9471 doğruluk ve 0,1805 hata elde edilmiştir. Son olarak, ALISA aktivasyon fonksiyonu en düşük ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 0,9335 doğruluk ve 0,2275 hata ile elde edilmiştir.



(a)



(b)

Şekil 6.2. MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları

Şekil 6.2, MNIST veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (a) ve doğrulama hata (b) sonuçlarını göstermektedir. Önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre en iyi ortalama performans değerlerini vermiştir. Ayrıca Şekil 6.2'de P+RSigELUD aktivasyon fonksiyonunun ortalama doğrulama doğruluk değeri, PReLU, PELU, ALISA ve P+RSigELUS aktivasyon fonksiyonlarından daha büyüktür. Önerilen eğitilebilir aktivasyon fonksiyonlarının ortalama doğrulama doğruluk performans aralığı, P+RSigELUS için 0,9380 ile 0,9558 ve P+RSigELUD için 0,9453 ile 0,9674 arasındadır. Ortalama doğrulama hata performans aralığı, P+RSigELUS için 0,1475 ile 0,2160 ve

P+RSigELUD için 0,1038 ile 0,1923 arasındadır. P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, Şekil 6.2 (a) kutu grafiğinin en düşük değerleri, diğer aktivasyon fonksiyonlarının üzerindedir. Bu nedenle, P+RSigELUD aktivasyon fonksiyonu daha iyi sonuç verdiği görülmektedir. Şekil 6.2 (b) P+RSigELUD aktivasyon fonksiyonunun hata değerleri kutu grafiğinde diğer aktivasyon fonksiyonları kutu grafiğine göre en düşük değerler arasındadır. Böylece Şekil 6.2(a-b)'deki kutu grafikleri incelendiğinde P+RSigELUD aktivasyon fonksiyonunun diğer fonksiyonlara göre daha tutarlı sonuçlar verdiği görülmektedir.

CIFAR-10 veri kümesi için önerilen eğitilebilir tek ve çift parametrelili P+RSigELU aktivasyon fonksiyonları ve bilinen diğer aktivasyon fonksiyonlarına ait başarı oranları Tablo 6.4, 6.5 ve 6.6'da görülmektedir.

Tablo 6.4. CIFAR-10 veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ. Hata	Doğ. Doğ.
Tekrar 1	40	0,5458	0,8549	0,6973	0,8168
Tekrar 2	40	0,5517	0,8521	0,6553	0,8259
Tekrar 3	40	0,5491	0,8556	0,6835	0,8164
Tekrar 4	40	0,5585	0,8506	0,6772	0,8175
Tekrar 5	40	0,5533	0,8521	0,7062	0,8093
Ortalama		0,5516	0,8530	0,6839	0,8171

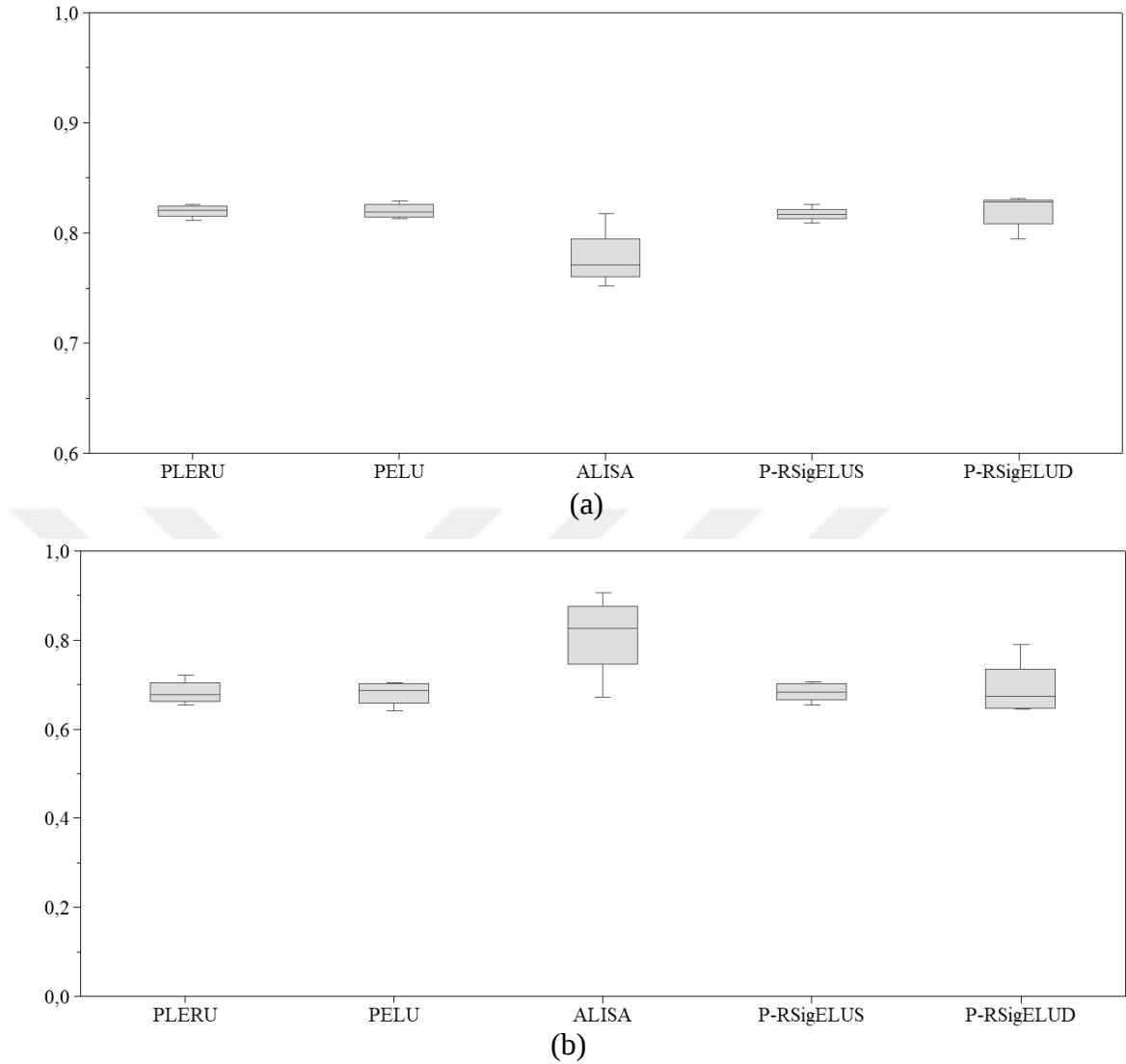
Tablo 6.5. CIFAR-10 veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim Hata	Eğitim Doğ.	Doğ. Hata	Doğ. Doğ.
Tekrar 1	40	0,6596	0,8116	0,7913	0,7947
Tekrar 2	40	0,5289	0,8606	0,6466	0,8306
Tekrar 3	40	0,5445	0,8531	0,6724	0,8215
Tekrar 4	40	0,5228	0,8619	0,6770	0,8282
Tekrar 5	40	0,5215	0,8596	0,6481	0,8285
Ortalama		0,5554	0,8493	0,6870	0,8207

Tablo 6.6. CIFAR-10 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	Eğitim_Hata	Eğitim_Doğ.	Doğ_Hata	Doğ_Doğ.
PReLU	0,5175	0,8624	0,6815	0,8200
PELU	0,5440	0,8544	0,6814	0,8199
ALISA	0,6093	0,8298	0,8135	0,7762
Önerilen P+RSigELUS	0,5516	0,8530	0,6839	0,8171
Önerilen P+RSigELUD	0,5554	0,8493	0,6870	0,8207

Tablo 6.4, 6.5 ve 6.6’de, CIFAR-10 veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile eğitilebilir P+RSigELUD aktivasyon fonksiyonunda 0,8207 doğruluk ve 0,6870 hata elde edilmiştir. Ayrıca PReLU aktivasyon fonksiyonunun ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri sırasıyla, 0,8200 doğruluk ve 0,6815 hata elde edilmiştir. Ayrıca ALISA aktivasyon fonksiyonu, en düşük ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 0,7762 doğruluk ve 0,8135 hata elde edilmiştir.



Şekil 6.3. CIFAR-10 veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları

Şekil 6.3, CIFAR-10 veri kümesi için aktivasyon fonksiyonlarını ortalama doğrulama doğruluk (a) ve doğrulama hata (b) sonuçlarını göstermektedir. Önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre en iyi ortalama performans değerlerini vermiştir. P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının doğrulama doğruluk oranı aralığı, P+RSigELUS için 0,8093 ile 0,8259 ve P+RSigELUD için 0,7947 ile 0,8306 arasındadır. Ayrıca, önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının doğrulama hata oranı aralığı, P+RSigELUS için 0,6553 ile 0,7062 ve P+RSigELUD için 0,6466 ile 0,7913 arasındadır. P+RSigELUD aktivasyon fonksiyonu Şekil 6.3 (a) kutu

grafisindeki en yüksek değer, diğer aktivasyon fonksiyonlarının tepe değerlerinin üzerindedir. Şekil 6.3 (b)'de ALISA aktivasyon fonksiyonu dışında diğer aktivasyon fonksiyonlarının en düşük hata değerleri birbirine yakındır. Şekil 6.3 (a ve b)'de ALISA aktivasyon fonksiyonunun diğer aktivasyon fonksiyonlarına göre daha düşük başarı gösterdiği ve sonuçlar arasında tutarsızlık olduğu görülmektedir. Ayrıca P+RSigELUS aktivasyon fonksiyonunu diğer aktivasyon fonksiyonlarına göre daha küçük kutu boyutu ile daha tutarlı sonuçlar verdiği görülmektedir.

CIFAR-100 veri kümesi için P+RSigELU aktivasyon fonksiyonları ve bilinen diğer aktivasyon fonksiyonlarına ait başarı oranları Tablo 6.7, 6.8 ve 6.9'da görülmektedir.

Tablo 6.7. CIFAR-100 veri kümesi üzerinde P+RSigELUS aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim_Hata	Eğitim_Doğ,	Doğ_Hata	Doğ_Doğ,
Tekrar 1	40	1,3885	0,6531	1,9894	0,5372
Tekrar 2	40	1,4114	0,6458	1,9413	0,5467
Tekrar 3	40	1,3929	0,6540	1,9804	0,5448
Tekrar 4	40	1,3813	0,6563	2,0319	0,5329
Tekrar 5	40	1,3729	0,6548	2,0330	0,5368
Ortalama		1,3894	0,6528	1,9952	0,5396

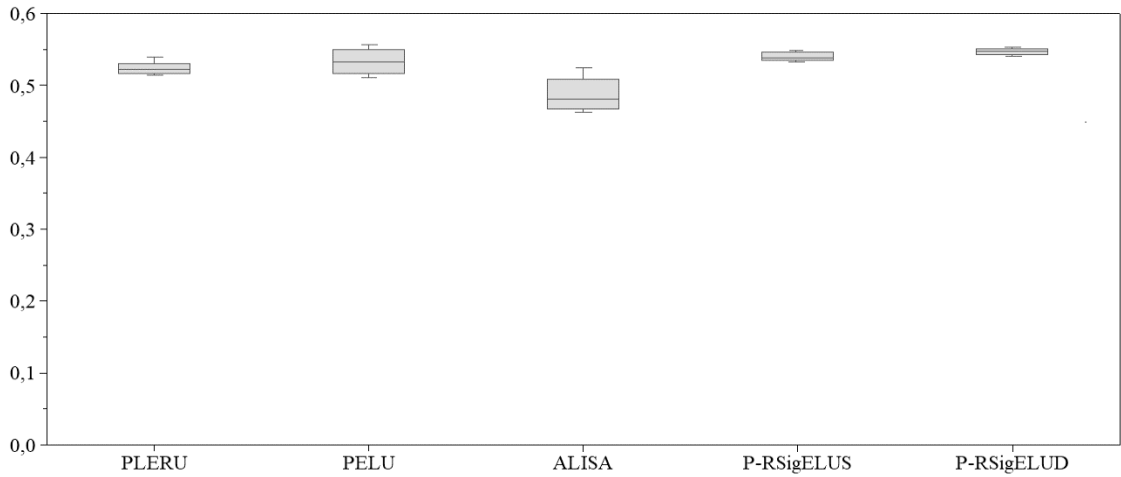
Tablo 6.8. CIFAR-100 veri kümesi üzerinde P+RSigELUD aktivasyon fonksiyonunun başarı oranları

Tekrar	Model Tekrar	Eğitim_Hata	Eğitim_Doğ,	Doğ_Hata	Doğ_Doğ,
Tekrar 1	40	1,3268	0,6685	1,9586	0,5524
Tekrar 2	40	1,3047	0,6732	1,9993	0,5431
Tekrar 3	40	1,3299	0,6647	1,9818	0,5463
Tekrar 4	40	1,2988	0,6760	2,0193	0,5404
Tekrar 5	40	1,3188	0,6689	1,9852	0,5490
Ortalama		1,3158	0,6702	1,9888	0,5462

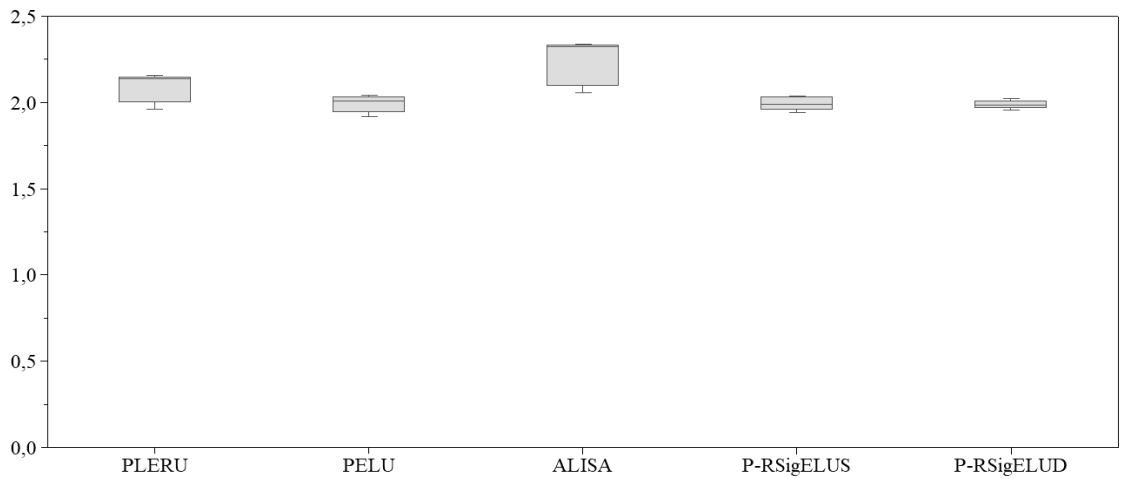
Tablo 6.9. CIFAR-100 veri kümesi üzerinde aktivasyon fonksiyonlarının başarı oranları

Aktivasyon	Eğitim_Hata	Eğitim_Doğ,	Doğ_Hata	Doğ_Doğ,
PReLU	1,3127	0,6717	2,0874	0,5269
PELU	1,4471	0,6341	1,9911	0,5325
ALISA	1,5408	0,6130	2,2368	0,4861
Önerilen P+RSigELUS	1,3894	0,6528	1,9952	0,5396
Önerilen P+RSigELUD	1,3158	0,6702	1,9888	0,5462

Tablo 6.7, 6.8 ve 6.9’de, CIFAR-100 veri kümesi üzerinde gerçekleştirilen deneysel sonuçlarda en iyi ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değeri 40 tekrar ile eğitilebilir P+RSigELUD aktivasyon fonksiyonunda 0,5462 doğruluk ve 1,9888 hata elde edilmiştir. Eğitilebilir P+RSigELUS aktivasyon fonksiyonunda ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) 0,5396 doğruluk ve 1,9952 hata elde edilmiştir. Ayrıca ALISA aktivasyon fonksiyonunda ortalama doğrulama doğruluk (Doğ_Doğ.) ve doğrulama hata (Doğ_Hata) değerleri 0,4861 doğruluk ve 2,2368 hata ile elde edilmiştir.



(a)



(b)

Şekil 6.4. CIFAR-100 veri kümesi için aktivasyon fonksiyonlarının ortalama (a) doğrulama doğruluk ve (b) doğrulama hata sonuçları

Şekil 6.4, CIFAR-100 veri kümesi için aktivasyon fonksiyonlarının ortalama doğrulama doğruluk (a) ve doğrulama hata (b) sonuçlarını göstermektedir. Önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, diğer aktivasyon fonksiyonlarına göre en iyi ortalama performans değerlerini vermiştir. Ayrıca, önerilen P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının doğrulama doğruluk oranı aralığı, P+RSigELUS için 0,5329 ile 0,5467 ve P+RSigELUD için 0,5404 ile 0,5524 arasındadır. Eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonlarının doğrulama hata oranı aralığı, P+RSigELUS için 1,9413 ile 2,0330 ve P+RSigELUD için 1,9586 ile 2,0193 arasındadır. Önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, Şekil 6.4 (a) kutu grafiğindeki en düşük değer, ALISA aktivasyon fonksiyonunun üzerindedir ve diğer aktivasyon fonksiyonlarına yakındır. Şekil 6.4 (a ve b)'de önerilen aktivasyon fonksiyonlarının kutu boyutları diğerinden daha küçüktür. Bu nedenle, önerilen aktivasyon fonksiyonlarının diğer aktivasyon fonksiyonlarına göre daha başarılı ve tutarlı sonuçlar sunduğu görülmektedir.

Tez çalışmasında, eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları önerilmiştir. Eğitilebilir aktivasyon fonksiyonları, eğitim sırasında her bir nörona en uygun değerleri elde etmek için aktivasyon fonksiyonunun eğitim parametrelerini (α^k ve β^k) sürekli olarak güncellemektedir. Böylece önerilen eğitilebilir P-RSigELU aktivasyon fonksiyonları eğitim sürecini sürekli olarak devam ettirebilmektedir. Kutu grafikleri incelendiğinde, önerilen aktivasyon fonksiyonlarının diğer aktivasyon fonksiyonlarına göre daha kararlı ve tutarlı sonuçlar verdiği görülmektedir. Önerilen eğitilebilir P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktif çalışmaktadır. Önerilen aktivasyon fonksiyonları, önyargı kayması, negatif bölge ve kaybolan gradyan problemleriyle başarıyla üstesinden gelebilmektedir. Eğitilebilir P+RSigELU aktivasyon fonksiyonları, eğri eğilimi temel olarak RSigELU ile tutarlıdır, bu da yeni eğitilebilir aktivasyon fonksiyonlarının RSigELU aktivasyon fonksiyonları ile doğruluk avantajını değiştirmemesini sağlamaktadır.

6.5. Sonuçlar

Tez çalışmasında, derin öğrenme mimarisi ESA'nın her nöronunda öğrenme işlemini gerçekleştirebilmesi için eğitilebilir P-RSigELU aktivasyon fonksiyonları önerilmiştir. Eğitilebilir aktivasyon fonksiyonları, eğitim sırasında her bir nörona en uygun değerleri elde etmek için aktivasyon fonksiyonunun eğim parametrelerini (a^k ve β^k) sürekli olarak güncellemektedir. Bu nedenle, önerilen eğitilebilir P-RSigELU aktivasyon fonksiyonları eğitim sürecini sürekli olarak devam ettirebilmektedir. Çalışmada önerilen eğitilebilir tek ve çift parametrelili P+RSigELUS ve P+RSigELUD aktivasyon fonksiyonları, önyargı kaydırma, negatif bölge ve kaybolan gradyan problemlerinin üstesinden başarıyla gelebilmektedir. Önerilen eğitilebilir P+RSigELU aktivasyon fonksiyonları MNIST, CIFAR-10 ve CIFAR-100 veri kümeleri üzerinde test edilmiş PReLU, PELU ve ALISA aktivasyon fonksiyonlarından daha iyi sonuçlar verdiği görülmektedir.

7. BÖLÜM

TARTIŞMA, SONUÇ VE ÖNERİLER

7.1. Tartışma ve Sonuç

Bu tez çalışmasında, sinir ağlarının daha etkin bir şekilde çalışabilmesi için yeni aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonları, herbiri farklı tipteki veri kümelerini daha etkin bir şekilde analiz etmek için tasarlanmıştır. Geliştirilen yöntemler ve literatüre katkısı aşağıdaki şekilde sıralanmıştır.

- Bölüm 2’de, literatürde sık kullanılan ReLU ve swish aktivasyon fonksiyonları kullanılarak, ESA mimarisi ile anemi veri kümesi üzerinde deneysel değerlendirmeler gerçekleştirilmiştir. Yapılan deneysel çalışmalarda, kullanılan derin öğrenme yöntemlerinin hiper-parametrelerini belirlemek için sezgisel tabanlı algoritmalarından, PSO, GA, DGA ve YAKA kullanılmıştır. Yapılan çalışmanın sonuçları derlenerek iki yayın üretilmiştir. Yayınlardan birisi yayınlanmıştır ve diğeri ise revizyon aşamasındadır.
- Bölüm 3’de, mikrodizi veri kümesinin boyutunu indirmek için kullanılan ReliefF ile YOK algoritmaların yanında, ReLU aktivasyon fonksiyonu tabanlı ESA mimarisi ile sınıflandırma işlemini gerçekleştirilmiş ve deneysel değerlendirme sonuçları sunulmuştur. Yapılan çalışmanın sonuçları derlenerek bir dergide yayınlanmıştır [22].

- Bölüm 4’de, derin öğrenme algoritmalarında meydana gelen kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelebilmek amacıyla tek ve çift parametrelili, doğrusal olmayan yeni sabit parametrelili RSigELU aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonları, literatürde bilinen kaybolan gradyan ve negatif bölge problemlerinin üstesinden gelerek öğrenme işlemine devam edebilmektedir. Ayrıca, önerilen RSigELU aktivasyon fonksiyonları, pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktif olarak çalışmaktadır. Deneysel değerlendirmeler, önerilen RSigELU aktivasyon fonksiyonları MNIST, Fashion MNIST, IMDB ve CIFAR-10 benchmark veri kümeleri üzerinden gerçekleştirilmiştir. RSigELU aktivasyon fonksiyonları, sonuçları karşılaştırmak amacıyla kullanılan diğer aktivasyon fonksiyonlarına göre daha iyi başarı oranı elde edildiğini ve etkili şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergide yayınlanmıştır [28].
- Bölüm 5’de, çekirdek tabanlı aktivasyon fonksiyonu (KAF) ile hibrit şekilde RSigELU aktivasyon fonksiyonları birleştirilerek yeni özgün çekirdek tabanlı aktivasyon fonksiyonları (KAF+RSigELUS ve KAF+RSigELUD) önerilmiştir. Önerilen çekirdek tabanlı KAF+RSigELUS ve KAF+RSigELUD aktivasyon fonksiyonları, pozitif, negatif ve doğrusal aktivasyon bölgelerinde aktiftir ve mevcut problemlerden önyargı kayması (bias shift), kaybolan gradyan, uyarlanamama ve esnek olmama probleminin üstesinden başarıyla gelebilmektedir. Çekirdek tabanlı KAF+RSigELU aktivasyon fonksiyonları, MNIST, Fashion MNIST ve CIFAR-10 benchmark veri kümeleri üzerinde literatürdeki diğer aktivasyon fonksiyonlarına göre daha iyi başarı oranı elde edildiğini ve etkili şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir ve revizyon aşamasındadır.
- Bölüm 6’da, eğitilebilir yeni özgün aktivasyon fonksiyonları (P+RSigELUS ve P+RSigELUD) önerilmiştir. Önerilen eğitilebilir P+RSigELU aktivasyon fonksiyonları, MNIST, CIFAR-10 ve CIFAR-100 veri kümesi üzerinde literatürdeki PReLU, PELU ve ALISA aktivasyon fonksiyonlarına göre oldukça etkili şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir.

Tez çalışmasında önerilen tüm aktivasyon fonksiyonları, ortak olarak MNIST ve CIFAR-10 görüntü veri kümeleri üzerinde literatürde önerilen aktivasyon fonksiyonlarına göre en iyi başarı oranına sahiptir. Ancak literatürdeki aktivasyon fonksiyonlarına göre yavaş çalışmaktadırlar. Metinsel IMDB movie veri kümesi üzerinde, literatürdeki aktivasyon fonksiyonlarına göre daha hızlı çalıştığı gözlemlenmiştir.

7.2. Öneriler

- ESA+LSTM (uzun kısa süreli bellek) ve diğer zaman serisi analiz tekniklerini ve farklı aktivasyon fonksiyonları kullanarak anemi veri kümelerinin zaman serisi analizleri yapılabilir. Ayrıca, hiper-parametre optimizasyonu için hesaplama süresini daha kısa gerçekleştirebilecek optimizasyon yöntemleri üzerine çalışmalar yapılabilir.
- Gelecekteki çalışmalarda, mikrodizi veri kümelerinin zaman serisi analizini gerçekleştirmek için LSTM (uzun kısa süreli bellek) gibi derin öğrenme analizleri yapılabilir. ESA mimarisinde yinelemeli hiper-parametreler elde etmek için akıllı optimizasyon yöntemleri ve mikrodizi veri kümelerine uygun aktivasyon fonksiyonlarının geliştirilmesi üzerine yaklaşımlar geliştirilmesi mümkün olabilir.
- Önerilen RSigELU aktivasyon fonksiyonlarının parametreleri sabit kabul edilmektedir. Derin öğrenme algoritmalarının hiper-parametrelerini elde etmek ve önerilen aktivasyon fonksiyonlarını diğer ağ yapılarına uygulamak için akıllı optimizasyon yöntemlerinin kullanımı ile ilgili çalışmalar yapılabilir..
- Önerilen çekirdek tabanlı KAF+RSigELU aktivasyon fonksiyonları tek boyutlu mimari altında çalışmaktadır. Önerilen aktivasyon fonksiyonlarının hesaplama karmaşıklığını daha aza indirmek ve çok katmanlı mimarilerde kullanma kolaylığı sağlanması üzerine çalışmalar yapılabilir.
- Önerilen eğitilebilir P+RSigELU aktivasyon fonksiyonları sabit veri kümeleri üzerinde çalışmaktadır. Eğitilebilir P+RSigELU aktivasyon fonksiyonlarının akan veri kümeleri üzerinde dinamik olarak çalışabilecek tasarımların geliştirilmesi çalışmaları yapılabilir.

KAYNAKLAR

1. Schmidhuber, J., 2015. Deep Learning in Neural Networks: An Overview, **Neural Networks**, **61**:85-117.
2. Bircanoğlu, C., & Arica, N., 2018. A Comparison of Activation Functions in Artificial Neural Networks, *In 2018 26th Signal Processing and Communications Applications Conference (SIU)*, 1-4, IEEE.
3. Maas, A. L., Hannun, A. Y., & Ng, A. Y., 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models, *Proceedings of the 30 th International Conference on Machine Learning*, 3-20.
4. Glorot, X., Bordes, A., & Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks, *In Proceedings of The Fourteenth International Conference on Artificial Intelligence And Statistics*, 315-323.
5. Krizhevsky, A., Sutskever, I., & Hinton, G. E., 2017. Imagenet Classification with Deep Convolutional Neural Networks, *Communications of the ACM*, **60**(6), 84-90.
6. Ebrahimi, M. S., & Abadi, H. K., 2018. Study of Residual Networks for Image Recognition, **Arxiv Preprint Arxiv:1805.00325**.
7. Nair, V., & Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines, *The 27th International Conference on Machine Learning (ICML 2010)*.
8. Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S., 2017. Self-normalizing neural networks, **Advances in Neural Information Processing Systems**, **30**:971-980.
9. Clevert, D. A., Unterthiner, T., & Hochreiter, S., 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (Elus), *In 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 1–14.

10. He, K., Zhang, X., Ren, S., & Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification, *In Proceedings of the IEEE International Conference On Computer Vision*, 1026-1034.
11. Kong, S., & Takatsuka, M., 2017. Hexpo: A Vanishing-Proof Activation Function, *In 2017 International Joint Conference on Neural Networks (IJESNA)*, 2562-2567, IEEE.
12. Bawa, V. S., & Kumar, V., 2019. Linearized Sigmoidal Activation: A Novel Activation Function with Tractable Non-Linear Characteristics to Boost Representation Capability, **Expert Systems with Applications**, **120**:346-356.
13. Trottier, L., Gigu, P., & Chaib-draa, B., 2017. Parametric Exponential Linear Unit for Deep Convolutional Neural Networks, *In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 207-214, IEEE.
14. Chieng, H. H., Wahid, N., & Ong, P., 2020. Parametric Flatten-T Swish: An Adaptive Non-Linear Activation Function for Deep Learning, **Journal of Information and Communication Technology (JICT)**, **20**(1): 21-39.
15. Maguolo, G., Nanni, L., & Ghidoni, S., 2021. Ensemble of Convolutional Neural Networks Training with Different Activation Functions, **Expert Systems with Applications**, **166**: 114048.
16. Zhao, H., Liu, F., Li, L., & Luo, C., 2018. A Novel Softplus Linear Unit for Deep Convolutional Neural Networks, **Applied Intelligence**, **48**(7): 1707-1720.
17. Scardapane, S., Van Vaerenbergh, S., Totaro, S., & Uncini, A., 2019. Kafnets: Kernel-Based Non-Parametric Activation Functions for Neural Networks, **Neural Networks**, **110**: 19-32.
18. Eberhart, R., & Kennedy, J., 1995. A New Optimizer Using Particle Swarm Theory, *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 39-43, IEEE.
19. Holland, J. H., 1992. Genetic Algorithms, **Scientific American**, **267**(1): 66-73.

20. Storn, R., & Price, K., 1997. Differential Evolution—A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, **Journal of Global Optimization**, **11**(4): 341-359.
21. Karaboga, D., & Akay, B., 2009. A Survey: Algorithms Simulating Bee Swarm Intelligence, **Artificial Intelligence Review**, **31**(1-4): 61.
22. Kilicarslan, S., Adem, K., & Celik, M., 2020. Diagnosis and Classification of Cancer Using Hybrid Model Based on Relieff and Convolutional Neural Network, **Medical Hypotheses**, **137**: 109577.
23. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., 1998. Gradient-Based Learning Applied to Document Recognition, **Proceedings of the IEEE**, **86**(11): 2278-2324.
24. Xiao, H., Rasul, K., & Vollgraf, R., 2017. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms, **Arxiv Preprint Arxiv:1708.07747**.
25. Krizhevsky, A., Sutskever, I., & Hinton, G. E., 2012. Imagenet Classification with Deep Convolutional Neural Networks, **Advances in Neural Information Processing Systems**, **25**, 1097-1105.
26. Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C., 2011. Learning Word Vectors for Sentiment Analysis, *In Proceedings of the 49th Annual Meeting of The Association for Computational Linguistics: Human Language Technologies*, 142-150.
27. Kilicarslan, S., Celik, M., & SAHIN, Ş., 2021. Hybrid Models Based on Genetic Algorithm and Deep Learning Algorithms for Nutritional Anemia Disease Classification, **Biomedical Signal Processing and Control**, **63**: 102231.
28. Kiliçarslan, S., & Celik, M., 2021. RSigELU: A Nonlinear Activation Function for Deep Neural Networks, **Expert Systems with Applications**, **174**: 114805.
29. LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep Learning, **Nature**, **521**(7553): 436-

444.

30. Yun, B. I., 2019. A Neural Network Approximation Based on a Parametric Sigmoidal Function, **Mathematics**, 7(3): 262.
31. Agostinelli, F., Hoffman, M., Sadowski, P., & Baldi, P., 2014. Learning Activation Functions to Improve Deep Neural Networks, *In 3th International Conference on Learning Representations*, 1–9.
32. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R., 2018. Deep Learning for Healthcare Applications Based on Physiological Signals: A Review, **Computer Methods and Programs in Biomedicine**, 161: 1-13.
33. Zeiler, M. D., & Fergus, R., 2014. Visualizing and Understanding Convolutional Networks, *In European Conference on Computer Vision*, 818-833, Springer, Cham.
34. Glorot, X., & Bengio, Y., 2010. Understanding the Difficulty of Training Deep Feedforward Neural Networks, *In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249-256.
35. Deng, L., & Yu, D., 2014. Deep Learning: Methods and Applications, **Foundations and Trends in Signal Processing**, 7(3–4): 197-387.
36. Cireşan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J., 2010. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition, **Neural Computation**, 22(12): 3207-3220.
37. Liu, Q., & Furber, S., 2016. Noisy Softplus: A Biology Inspired Activation Function, *In International Conference on Neural Information Processing*, 405-412, Springer, Cham.
38. Nguyen, H. G., Nans, A., Talke, K., Candela, P., & Everett, H. R., 2015. Automatic Behavior Sensing for A Bomb-Detecting Dog, **International Society for Optics and Photonics**, 9468:94680.

39. Bengio, Y., 2007. Learning Deep Architectures for AI, **Foundations and Trends in Machine Learning**, **2** (1): 1-127.
40. Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B., 2020. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN), **Sensors**, **20**(12):3344.
41. Bottou, L., LeCun, Y., & Haffner, P., 1990. Gradient-Based Learning Applied to Document Recognition, **Proceedings of the IEEE**, **79**(11): 22-40.
42. Deperlioğlu, Ö., & Köse, U., 2018. Diagnosis of Diabete Mellitus Using Deep Neural Network, *In 2018 Medical Technologies National Congress (TIPTEKNO)*, 1-4, IEEE.
43. Adem, K., Kiliçarslan, S., & Cömert, O., 2019. Classification and Diagnosis of Cervical Cancer with Stacked Autoencoder and Softmax Classification, **Expert Systems with Applications**, **115**: 557-564.
44. Alladi, K., Andrews, G. E., & Berkovich, A., 2004. A Four Parameter Generalization of Gollnitz's (BIG) Partition Theorem, **Dinacs Series in Discrete Mathematics and Theoretical Computer Science**, **64**: 1-8.
45. LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D., 1990. Handwritten Digit Recognition with A Back-Propagation Network, **In Advances in Neural Information Processing Systems**, 396-404.
46. Osowski, S., Siwek, K., & Markiewicz, T., 2004. MLP and SVM Networks-A Comparative Study, *In Proceedings of the 6th Nordic Signal Processing Symposium*, 37-40, IEEE.
47. Cortes, C., & Vapnik, V., 1995. Support-Vector Networks, **Machine Learning**, **20**(3): 273-297.
48. Ayhan, S., & Erdoğan, Ş., 2014. Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü için Çekirdek Fonksiyonu Seçimi, **Eskişehir Osmangazi**

Üniversitesi İktisadi ve İdari Bilimler Dergisi, 9(1): 175-201.

49. Göldoğan, E., 2017. Çeşitli Çekirdek Fonksiyonları ile Oluşturulan Destek Vektör Makinesi Modellerinin Performanslarının İncelenmesi: Bir Klinik Uygulama, İnönü Üniversitesi, Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, Malatya, s.75.
50. Kira, K., & Rendell, L. A., 1992. A Practical Approach to Feature Selection, **In Machine Learning Proceedings 1992**, 249-256, Morgan Kaufmann.
51. Ng, A., 2010. Cs294a lecture notes: Sparse Autoencoder, Stanford University.
52. Kononenko, I., Robnik-Sikonja, M., & Pompe, U., 1996. ReliefF for Estimation and Discretization of Attributes in Classification, Regression, and ILP Problems, **Artificial Intelligence: Methodology, Systems, Applications**:31-40.
53. Chandrashekar, G., & Sahin, F., 2014. A Survey on Feature Selection Methods, **Computers & Electrical Engineering**, **40**(1): 16-28.
54. Venkataramana, L., Jacob, S. G., Ramadoss, R., Saisuma, D., Haritha, D., & Manoja, K., 2019. Improving Classification Accuracy of Cancer Types Using Parallel Hybrid Feature Selection on Microarray Gene Expression Data, **Genes & Genomics**, **41**(11): 1301-1313.
55. Zhu, Z., Ong, Y. S., & Dash, M., 2007. Markov Blanket-Embedded Genetic Algorithm for Gene Selection, **Pattern Recognition**, **40**(11): 3236-3248.
56. Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., & Golub, T. R., 2002. Prediction of Central Nervous System Embryonal Tumour Outcome Based on Gene Expression, **Nature**, **415**(6870): 436-442.
57. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P. & Lander, E. S., 1999. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, **Science**, **286**(5439):531-537.
58. Petricoin III, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M. & Liotta, L. A., 2002. Use of Proteomic Patterns in Serum to

Identify Ovarian Cancer, **The Lancet**, 359(9306): 572-577.

59. Hendrycks, D., & Gimpel, K., 2016. Gaussian Error Linear Units (Gelus), **Arxiv Preprint Arxiv:1606.08415**.
60. Godin, F., Degrave, J., Dambre, J., & De Neve, W., 2018. Dual Rectified Linear Units (DReLU): A Replacement for Tanh Activation Functions in Quasi-Recurrent Neural Networks, **Pattern Recognition Letters**, 116:8-14.
61. Qiumei, Z., Dan, T., & Fenghua, W., 2019. Improved Convolutional Neural Network Based on Fast Exponentially Linear Unit Activation Function, **IEEE Access**, 7 :151359-151367.
62. Ramachandran, P., Zoph, B., & Le, Q. V., 2017. Searching for Activation Functions, **Arxiv Preprint Arxiv:1710.05941**.
63. Wang, X., Qin, Y., Wang, Y., Xiang, S., & Chen, H., 2019. ReLTanh: An Activation Function with Vanishing Gradient Resistance for SAE-Based DNNs and its Application to Rotating Machinery Fault Diagnosis, **Neurocomputing**, 363: 88-98.
64. Kong, S., & Takatsuka, M., 2017. Hexpo: A Vanishing-Proof Activation Function, *In 2017 International Joint Conference on Neural Networks (IJESNA)*, 2562-2567, IEEE.
65. Sharma, O., 2019. A New Activation Function for Deep Neural Network, *In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 84-86, IEEE.
66. Liu, Q., & Furber, S., 2016. Noisy Softplus: A Biology Inspired Activation Function, *In International Conference on Neural Information Processing*, 405-412, Springer, Cham.
67. Gao, H., Cai, L., & Ji, S., 2020. Adaptive Convolutional ReLUs, *In Proceedings of the AAAI Conference on Artificial Intelligence*, 3914-3921.

68. Godfrey, L. B., 2019. An Evaluation of Parametric Activation Functions for Deep Learning, *In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 3006-3011 IEEE.
69. Kononenko, I., 1994. Estimating Attributes: Analysis and Extensions of RELIEF, *In European Conference on Machine Learning*, 171-182, Springer, Berlin, Heidelberg.
70. Jatoi, S., Panhwar, M. A., Memon, M. S., & Baloch, J. A., 2018. Mining Complete Blood Count Reports for Disease Discovery, **International Journal of Computer Science and Network Security**, **18**(1): 121–127.
71. Yu, C. H., Bhatnagar, M., Hogen, R., Mao, D., Farzindar, A., & Dhanireddy, K., 2017. Anemic Status Prediction Using Multilayer Perceptron Neural Network Model, *3rd Global Conference on Artificial Intelligence, In GCAI*, 213-220.
72. Abdullah, M., & Al-Asmari, S., 2016. Anemia Types Prediction Based on Data Mining Classification Algorithms, *In Communication, Management and Information Technology*, 629-636, CRC Press.
73. Hasani, M., & Hanani, A., 2017. Automated Diagnosis of Iron Deficiency Anemia and Thalassemia by Data Mining Techniques, **International Journal of Computer Science and Network Security (IJCSNS)**, **17**(4):326.
74. Azarkhish, I., Raoufy, M. R., & Gharibzadeh, S., 2012. Artificial Intelligence Models For Predicting Iron Deficiency Anemia and Iron Serum Level Based on Accessible Laboratory Data, **Journal of Medical Systems**, **36**(3): 2057-2061.
75. Kaya, E., Aktan, M. E., Akdoğan, E., & Koru, A. T., 2015. Diagnosis of Anemia in Children Via Artificial Neural Network, **International Journal of Intelligent Systems and Applications in Engineering**, **3**(1): 24-27.
76. Shouval, R., Bondi, O., Mishan, H., Shimoni, A., Unger, R., & Nagler, A., 2014. Application of Machine Learning Algorithms for Clinical Predictive Modeling: A Data-Mining Approach in SCT, **Bone Marrow Transplantation**, **49**(3): 332-337.

77. Elshami, E. H., & Alhalees, A. M., 2012. Automated Diagnosis of Thalassemia Based on Datamining Classifiers, *In The International Conference on Informatics and Applications (ICIA2012)*, 440-445.
78. L-Halees, A. M., & Shurrah, A. H., 2017. Blood tumor Prediction Using Data Mining Techniques, **Health Informatics—An International Journal**,6(2).
79. Çil, B., Ayyıldız, H., & Tuncer, T., 2020. Discrimination of β -Thalassemia and Iron Deficiency Anemia Through Extreme Learning Machine and Regularized Extreme Learning Machine Based Decision Support System, **Medical Hypotheses**, **138**:109611.
80. Gülcü, A., & Kuş, Z., 2019. Konvolüsyonel Sinir Ağlarında Hiper-Parametre Optimizasyonu Yöntemlerinin İncelenmesi, **Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji**, 7(2):503–522.
81. Bochinski, E., Senst, T., & Sikora, T., 2017. Hyper-Parameter Optimization for Convolutional Neural Network Committees Based on Evolutionary Algorithms, *In 2017 IEEE International Conference on Image Processing (ICIP)*, 3924-3928, IEEE.
82. Lopez-Rincon, A., Tonda, A., Elati, M., Schwander, O., Piwowarski, B., & Gallinari, P., 2018. Evolutionary Optimization of Convolutional Neural Networks for Cancer miRNA Biomarkers Classification, **Applied Soft Computing**, **65**: 91-100.
83. Baldominos, A., Saez, Y., & Isasi, P., 2018. Evolutionary Convolutional Neural Networks: An Application to Handwriting Recognition, **Neurocomputing**, **283**: 38-52.
84. da Silva, G. L., da Silva Neto, O. P., Silva, A. C., de Paiva, A. C., & Gattass, M., 2017. Lung Nodules Diagnosis Based on Evolutionary Convolutional Neural Network, **Multimedia Tools and Applications**, **76**(18):19039-19055.
85. Yamasaki, T., Honma, T., & Aizawa, K., 2017. Efficient Optimization of

- Convolutional Neural Networks Using Particle Swarm Optimization, *In 2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, 70-73, IEEE.
86. da Silva, G. L. F., Valente, T. L. A., Silva, A. C., de Paiva, A. C., & Gattass, M., 2018. Convolutional Neural Network-Based PSO for Lung Nodule False Positive Reduction on CT images, **Computer Methods and Programs in Biomedicine**, **162**: 109-118.
 87. Wang, B., Sun, Y., Xue, B., & Zhang, M., 2018. Evolving Deep Convolutional Neural Networks by Variable-Length Particle Swarm Optimization for Image Classification, *In 2018 IEEE Congress on Evolutionary Computation (CEC)*, 1-8, IEEE.
 88. Wang, B., Sun, Y., Xue, B., & Zhang, M., 2018. A Hybrid Differential Evolution Approach To Designing Deep Convolutional Neural Networks for Image Classification, *In Australasian Joint Conference on Artificial Intelligence*, 237-250, Springer, Cham.
 89. Sun, Y., Xue, B., Zhang, M., & Yen, G. G., 2018. An Experimental Study on Hyper-Parameter Optimization for Stacked Auto-Encoders, *In 2018 IEEE Congress on Evolutionary Computation (CEC)*, 1-8, IEEE.
 90. Li, Y., & Yang, R., 2013. A Hybrid Algorithm Combining Auto-Encoder Network with Sparse Bayesian Regression Optimized by Artificial Bee Colony for Short-Term Wind Power Forecasting, **Przegląd Elektrotechniczny**, **89**(2a): 223-228.
 91. Zhu, W., Yeh, W., Chen, J., Chen, D., Li, A., & Lin, Y., 2019. Evolutionary Convolutional Neural Networks Using ABC, *In Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 156-162.
 92. Liu, Y., 2008. Detect Key Gene Information in Classification of Microarray Data **EURASIP ,Journal on Advances in Signal Processing**, **2008**:1-10.
 93. Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., &

- Friend, S. H., 2002. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer, **Nature**, **415**(6871):530-536.
94. Wang, Y., Yang, X. G., & Lu, Y., 2019. Informative Gene Selection for Microarray Classification Via Adaptive Elastic Net with Conditional Mutual Information, **Applied Mathematical Modelling**, **71**: 286-297
95. Bennet, J., Arul Ganaprakasam, C., & Arputharaj, K., 2014. A Discrete Wavelet Based Feature Extraction and Hybrid Classification Technique for Microarray Data Analysis, **The Scientific World Journal**, **2014**:1-15.
96. Sun, X., Liu, Y., Xu, M., Chen, H., Han, J., & Wang, K., 2013. Feature Selection Using Dynamic Weights For Classification, **Knowledge-Based Systems**, **37**: 541-549.
97. Kumar, M., Singh, S., & Rath, S. K., 2015. Classification Of Microarray Data Using Functional Link Neural Network, **Procedia Computer Science**, **57**:727-737.
98. Díaz-Uriarte, R., & De Andres, S. A., 2006. Gene Selection and Classification of Microarray Data Using Random Forest, **BMC Bioinformatics**, **7**(1):3.
99. Ludwig, S. A., Jakobovic, D., & Picek, S., 2015. Analyzing Gene Expression Data: Fuzzy Decision Tree Algorithm Applied to The Classification of Cancer Data, *In 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1-8, IEEE.
100. Medjahed, S. A., Saadi, T. A., Benyettou, A., & Ouali, M., 2017. Kernel-based learning and Feature Selection Analysis for Cancer Diagnosis, **Applied Soft Computing**, **51**: 39-48.
101. Liu, J., Wang, X., Cheng, Y., & Zhang, L., 2017. Tumor Gene Expression Data Classification Via Sample Expansion-Based Deep Learning, **Oncotarget**, **8**(65):109646.
102. Rani, M. J., & Devaraj, D., 2019. Two-Stage Hybrid Gene Selection Using Mutual Information and Genetic Algorithm for Cancer Data Classification, **Journal of**

Medical Systems, 43(8):235.

103. Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M., 2018. Gene Selection and Classification of Microarray Data Using Convolutional Neural Network, *In 2018 International Conference on Advanced Science and Engineering (ICOASE)*, 145-150, IEEE.
104. Kılıçarslan, S., & Çelik, M., 2019. Rotasyon Orman Sınıflandırma Algoritması Kullanarak Kronik Böbrek Rahatsızlığının Tahmini, **Journal of Science and Technology of Dumlupınar University**, 043:21-34.
105. Thara, L., & Gunasundari, R., 2018. Swarm Intelligence Based Feature Selection Algorithms and Classifiers for Gastric Cancer Prediction, *In International Conference on Intelligent Data Communication Technologies and Internet of Things*, 1194-1201, Springer, Cham.
106. Rouhi, A., & Nezamabadi-pour, H., 2016. A Hybrid Method for Dimensionality Reduction in Microarray Data Based on Advanced Binary Ant Colony Algorithm, *In 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, 70-75, IEEE.
107. Venkataramana, L., Jacob, S. G., Ramadoss, R., Saisuma, D., Haritha, D., & Manoja, K., 2019. Improving Classification Accuracy of Cancer Types Using Parallel Hybrid Feature Selection on Microarray Gene Expression Data, **Genes & Genomics**, 41(11): 1301-1313.
108. Jain, I., Jain, V. K., & Jain, R., 2018. Correlation Feature Selection Based Improved-Binary Particle Swarm Optimization for Gene Selection and Cancer Classification, **Applied Soft Computing**, 62: 203-215.
109. Wall, C., & Gillies, N., 2019. Nutritional Anemias, *In Encyclopedia of Pharmacy Practice and Clinical Pharmacy*.
110. Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., & Bell, M. L., 2016. Global, Regional, and National Life Expectancy, All-Cause

- Mortality, and Cause-Specific Mortality For 249 Causes of Death, 1980–2015: A Systematic Analysis for The Global Burden of Disease Study 2015, **The Lancet**, **388**(10053):1459-1544.
111. Beyazal Polat, H., & Ayaz, T., 2019. Metformin Kullanan Tip 2 Diyabetli Bireylerde B12 Vitamin Eksikliğinin Belirlenmesi, **Türkiye Klinikleri Sağlık Bilimleri Dergisi**, **4**(3):237-241.
112. Emen, B., Kılıç Öztürk, Y., Eren, M. A., Özdemir, E., Öztürk, F., Düzenli, E., & Sariaslan, D., 2013. B12 Vitamin Eksikliği Bulunan Hastalarda Etiyolojik Faktörler ile Laboratuvar Verileri Arasındaki İlişkinin Geriye Dönük Değerlendirilmesi, **İzmir Tepecik Eğitim ve Araştırma Hastanesi Dergisi**, **23**(1): 19-23.
113. Salive, M. E., Cornoni-Huntley, J., Guralnik, J. M., Phillips, C. L., Wallace, R. B., Ostfeld, A. M., & Cohen, H. J., 1992. Anemia and Hemoglobin Levels in Older Persons: Relationship with Age, Gender, and Health Status, **Journal of the American Geriatrics Society**, **40**(5): 489-496.
114. Pınar, M., Okumuş, O., Turgut, U. O., Kalıpsız, O., & Aktaş, M. S., 2017. Büyük Veri İçeren Öneri Sistemleri İçin Hiperparametre Optimizasyonu, *Ulusal Yazılım Mühendisliği Sempozyumu*, 22-272.
115. Guo, B., Hu, J., Wu, W., Peng, Q., & Wu, F., 2019. The Tabu_Genetic Algorithm: A Novel Method for Hyper-Parameter Optimization of Learning Algorithms, **Electronics**, **8**(5):579.
116. Bacanin, N., Bezdan, T., Tuba, E., Strumberger, I., & Tuba, M., 2020. Optimizing Convolutional Neural Network Hyperparameters by Enhanced Swarm Intelligence Metaheuristics, **Algorithms**, **13**(3): 67.
117. Bergstra, J., & Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization, **The Journal of Machine Learning Research**, **13**(1): 281-305.
118. Tanyıldızı, E., & Demirtaş, F., 2019. Hiper Parametre Optimizasyonu Hyper

- Parameter Optimization, *In 2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 1-5, IEEE.
119. Liashchynskiy, P., & Liashchynskiy, P., 2019. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS, **Arxiv Preprint Arxiv:1912.06059**.
 120. Hochreiter, S., & Schmidhuber, J., 1997. Long Short-Term Memory, **Neural Computation**, **9**(8): 1735-1780.
 121. Bengio, Y., Simard, P., & Frasconi, P., 1994. Learning Long-Term Dependencies with Gradient Descent is Difficult, **IEEE Transactions on Neural Networks**, **5**(2): 157-166.
 122. Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C., 2011. Learning Word Vectors For Sentiment Analysis, *In Proceedings of the 49th Annual Meeting of The Association for Computational Linguistics: Human Language Technologies*, 142-150.
 123. Szandala, T., 2021. Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks, **In Bio-Inspired Neurocomputing**, 203-224.
 124. Apicella, A., Donnarumma, F., Isgrò, F., & Prevete, R., 2021. A Survey on Modern Trainable Activation Functions, **Neural Networks**, **138**:14-32.
 125. Basirat, M., & Roth, P. M., 2021. S* ReLU: Learning Piecewise Linear Activation Functions via Particle Swarm Optimization, *In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 645-652.
 126. Chung, H., Lee, S. J., & Park, J. G., 2016. Deep Neural Network Using Egitimable Activation Functions, *In 2016 International Joint Conference on Neural Networks (IJESNA)*, 348-352, IEEE.
 127. Zhang, J. G., & Deng, H. W., 2007. Gene Selection for Classification of Microarray

- Data Based on The Bayes Error, **BMC Bioinformatics**, 8(1): 370.
128. Genton, M. G., 2001. Classes of Kernels for Machine Learning: A Statistics Perspective, **Journal of Machine Learning Research**, 2(2):299-312.
129. Krzywinski, M., & Altman, N., 2014. Visualizing Samples with Box Plots, **Nature Methods**, 11(2): 119-120.
130. Karakoyun, M., & Hacıbeyoğlu, M., 2014. Statistical Comparison of Machine Learning Classification Algorithms Using Biomedical Datasets, **Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi**, 16(48):30-42.
131. Trottier, L., Gigu, P., & Chaib-draa, B., 2017. Parametric Exponential Linear Unit for Deep Convolutional Neural Networks, *In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 207-214, IEEE.

EKLER

Anemi hastalığını teşhis etmek için Tokat Gaziosmanpaşa Üniversitesi, Tıp Fakültesi, İç Hastalıkları Anabilim Dalına 2013-2018 yılları arasında gelen hastaların tam kan sayım testleri verileri alınmış. Bu veriler gerekli yasal izinler Tokat Gaziosmanpaşa Üniversitesi Klinik Araştırmalar Etik Kurulundan 18KA EK-206 kayıt numaralı proje no ve 83116987-450 sayı ile alınmıştır.





ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Serhat KILIÇARSLAN

Uyuğu: Türkiye (T.C)

Doğum Tarihi ve Yeri:

Medeni Durum:

e-mail:

Yazışma Adresi:

EĞİTİM

Derece	Kurum	Mez. Yılı
Lise	Gültepe Endüstri Meslek Lisesi, İSTANBUL	2003
Lisans	Marmara Üni. Teknik Eğitim Fak. Bilgisayar ve Kontrol Öğrt., İSTANBUL	2008
Lisans	Kocaeli Üni. Mühendislik Fak. Bilgisayar Müh. KOCAELİ	2017
Yüksek Lisas	Tokat Gaziosmanpaşa Üni. Mühendislik Fak. Mekatronik Müh. ABD, TOKAT	2014
Doktora	Erciyes Üni. Fen Bilimleri Ens. Bilgisayar Müh. ABD, KAYSERİ	2021

İŞ DENEYİMLERİ

Yıl	Kurum	Görev
2008-2011	İBB-İSMEK	Usta Öğretici
2011-Halen	Tokat Gaziosmanpaşa Üni. Pazar MYO. Bilgisayar Programcılığı Bölümü, TOKAT	Öğretim Görevlisi

YABANCI DİL

İngilizce

YAYINLAR

SCI, SSCI, AHCI 'İndekslerine Giren Dergilerde Yayınlanan Makaleler

- Kiliçarslan, S., & Celik, M. (2021). RSigELU: A Nonlinear Activation Function for Deep Neural Networks, **Expert Systems with Applications**, 174, 114805.

- Kılıçarslan Serhat, Çelik Mete, Şahin Şafak (2021). Hybrid Models Based on Genetic Algorithm and Deep Learning Algorithms for Nutritional Anemia Disease Classification, **Biomedical Signal Processing And Control** (Yayın No: 6495232)
- Kılıçarslan Serhat, Adem Kemal, Çelik Mete (2020). Diagnosis and Classification of Cancer Using Hybrid Model Based on Relieff and Convolutional Neural Network, **Medical Hypotheses**, 137(109577), 1-10., Doi: <https://doi.org/10.1016/j.mehy.2020.109577> (Yayın No: 6007355)

TR Dizinli Dergilerde Yayımlanan Makaleler

- Kılıçarslan Serhat, Çelik Mete (2019). Rotasyon Orman Sınıflandırma Algoritması Kullanarak Kronik Böbrek Rahatsızlığının Tahmini, **Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, (43), 21-34. (Kontrol No: 5703020)

Uluslararası Bilimsel Toplantılarda Sunulan Ve Bildiri Kitaplarında (Proceedings) Basılan Bildiriler :

- Kılıçarslan Serhat, Adem Kemal, Çelik Mete (2019). Derin Öğrenme ile Lösemi Kanseri Mikrodizi Verilerinin Sınıflandırılması, *13. Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu (Icits 2019)*, 533-538. (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:5098192)
- Kılıçarslan Serhat, Çelik Mete (2019). Rastgele Alt Uzay Yöntemi Kullanılarak Otizm Spektrum Bozukluğu Tahmini, *13. Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu (Icits 2019)*, 520-526. (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:5098194)
- Kılıçarslan Serhat, Çelik Mete, Şahin Şafak (2019). Rastgele Orman Algoritması ile Demir Eksikliği Anemisi Tahmini (Prediction of Iron Deficiency Anemia Using Rastgele Forest Algorithm), *13. Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu (Icits 2019)* (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:5098190)

SCI Dizinli Dergi Sistemlerine Yklenen İncelemedeki Çalıřmalar

- S. Kılıçarslan, M. Celik, S. řahin, “Determination of Hyper-Parameters of Deep Learning Models with Heuristic-Based Optimization Methods for Prediction of Anemia Disease”, **Medical & Biological Engineering & Computing** (SCI-E dergiye gnderildi), 2021.
- S. Kılıçarslan, M. Celik, “KAF+RSIGELU: A Nonlinear and Kernel-Based Activation Function for Deep Neural Networks, **Neural Computing & Applications** (SCI-E dergiye gnderildi), 2021.

