



FATİH SULTAN MEHMET VAKIF ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI

ÇOK SINIFLI MEDİKAL GÖRÜNTÜ
SINIFLANDIRMASI İÇİN MELEZ DERİN ÖĞRENME
YAKLAŞIMLARI

YÜKSEK LİSANS TEZİ

ZELİHA KAYA AKÇELİK

İSTANBUL, 2024



FATİH SULTAN MEHMET VAKIF ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI

ÇOK SINIFLI MEDİKAL GÖRÜNTÜ
SINIFLANDIRMASI İÇİN MELEZ DERİN ÖĞRENME
YAKLAŞIMLARI

YÜKSEK LİSANS TEZİ

ZELİHA KAYA AKÇELİK
(220221001)

Danışman
(Dr. Öğr. Üyesi Gönül Uludağ)

İSTANBUL, 2024

09/08/2024

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Enstitümüz Bilgisayar Mühendisliği Anabilim/Anasanat Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans programı öğrencisi 220221001 numaralı Zeliha KAYA AKÇELİK'in hazırladığı "Çok Sınıflı Medikal Görüntü Sınıflandırması İçin Melez Derin Öğrenme Yaklaşımları" konulu Yüksek Lisans tezi ile ilgili Tez Savunma Sınavı, 09/09/2024 Pazartesi günü saat 11:00' da yapılmış, sorulara alınan cevaplar sonunda adayın tezinin **Kabulüne Oy Çaldığı**/Oy Birliği ile karar verilmiştir.

Tez adı değişikliği yapılması halinde: Tez adının
.....
şeklinde değiştirilmesi uygundur.

Jüri Üyesi	Karar
1. (Danışman) Dr. Öğr. Üyesi Gönül ULUDAĞ	Kabul.....
2. Prof. Dr. A. Şima UYAR	Kabul.....
3. Doc. Dr. Yusuf YASLAN	Kabul.....
4.	
5.	
6. (İkinci Danışman)*	

*2. Danışman varsa doldurulması gerekmektedir.

ETİK BİLDİRİM

Bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bağlı olduğum üniversite veya bir başka üniversitedeki başka bir çalışma olarak sunulmadığını beyan ederim.

Zeliha Kaya Akçelik

DÜZELTME METNİ

- ResNet-Ağaç temelli ara katmanına boyut indirgeme algoritmalarından UMAP ve t-SNE algoritmaları uygulanmış ve boyut indirgeme algoritmalarıyla ilgili kapsamlı bilgiler eklenmiştir.
- ResNet-Ağaç temelli ara katmanına meta-sezgisel yöntemlerden CMA-ES ve PSO algoritmaları uygulanmış ve bu algoritmalar ilgili kapsamlı bilgiler eklenmiştir.
- Meta-sezgisel yöntemler ile birlikte sınıflandırma performansı değiştiğinden dolayı dördüncü bölümdeki tablolar güncellenmiştir.
- Yukarıdaki düzeltmelerin yanında, tezin genelinde dil ve anlatım hataları düzeltilmiş ve uygun yerlere kaynak eklenmiştir.

TEŞEKKÜR

Tez konusunun belirlenmesinden tamamlanmasına kadar her aşamasında değerli tecrübelerini, zamanını ve motivasyon artırıcı konuşmaları ile desteklerini esirgemeyen değerli danışman hocalarım ve Dr. Öğr. Üyesi Gönül ULUDAĞ'a, Dr. Öğr. Üyesi Berna KİRAZ'a ve destekleri için Dr. Öğr. Üyesi Zeki KUŞ'a ve Dr. Öğr. Üyesi Musa AYDIN'a teşekkürlerimi sunuyorum.

Eğitim hayatım boyunca desteğini ve inancını hiç esirgemeyen anne ve babama tüm fedakarlıkları için teşekkür ederim. Tüm bu desteklere ek olarak sabır, anlayış ve sevgisini hiç eksiltmeyen eşim Kadırcan'a teşekkür ediyorum.

Zeliha Kaya Akçelik

ÇOK SINIFLI MEDİKAL GÖRÜNTÜ SINIFLANDIRMASI İÇİN MELEZ DERİN ÖĞRENME YAKLAŞIMLARI

Zeliha Kaya Akçelik

ÖZET

Bu tez çalışması, farklı özelliklere sahip kan hücrelerinin sınıflandırılması için derin öğrenme (ing: deep learning) mimarileri, meta-sezgisel (ing: meta-heuristic) yaklaşımları, boyut azaltma yöntemleri ve ağaç tabanlı sınıflandırma algoritmalarının birlikte kullanıldığı melez bir çerçeve uygulamasıdır. Bu çalışma, Residual Network (ResNet) ResNet18 ve ResNet34 gibi derin öğrenme mimarilerinin, covariance matrix adaptation evolution strategy (CMA-ES) ve particle swarm optimization (PSO) gibi meta-sezgisel yöntemlerin, principal component analysis (PCA), indepedent component analysis (ICA), uniform manifold approximation and projection (UMAP) ve t-distributed stochastic neighbor embedding (t-SNE) gibi boyut indirgeme yaklaşımlarının ve Random Forest (RF), eXtreme gradient boosting (XGBoost) ve light gradient-boosting machine (LightGBM) gibi ağaç tabanlı algoritmalarının başarımlarının birlikte kullanıldığı melez bir çerçeve sunmaktadır.

Yapılan çalışmalar, görüntü işleme çalışmalarında ResNet18 ve ResNet34 gibi derin öğrenme mimarilerinin oldukça başarılı olduğunu göstermiştir. Ayrıca yapılan literatür çalışmaları RF, XGBoost ve LightGBM gibi ağaç tabanlı algoritmalarının sınıflandırma başarımlarının oldukça yüksek olduğunu göstermektedir. Bu yaklaşımların başarımlarından yola çıkarak, ResNet18 ve ResNet34 yaklaşımları özniteliklerinin çıkarımında kullanılırken, ağaç tabanlı algoritmalar olan karar ağaçları ise sınıflandırma amacı ile kullanılmaktadır.

Aynı zamanda geçmiş literatür çalışmaları göstermiştir ki, aykırı değerlerden arındırarak aşırı uyum (ing: over-fitting) probleminin önüne geçmek ve genelleştirme performansını arttırmak ve/veya hesaplama karmaşıklığını düşürmek amacı ile

sınıflandırma öncesinde uygun öznitelik seçimi ve/veya boyut azaltma yaklaşımlarının kullanılması önem arz etmektedir. Bu bağlamda, derin öğrenme mimarisi ön-eğitim (ing: pre-train) aşamasından sonra meta-sezgisel algoritmalarından CMA-ES ve PSO algoritmaları özniteliklerin seçilmesi amacı ile ve PCA, ICA, UMAP ve t-SNE gibi boyut indirgeme algoritmaları farklı kombinasyonlarda kullanılarak sınıflandırma performansı üzerindeki etkileri araştırılmıştır.

Bu tez çalışmasında önerilen melez yöntemlerin başarımı MedMNIST veri kümesinde bulunan, 8 adet farklı kan hücrelerinin bulunduğu BloodMNIST veri kümesi kullanılmıştır.

Makine öğrenimi yöntemlerinin başarımının hiper-parametre ayarlarına oldukça bağlı olduğu gerçeğinden yola çıkarak her aşamada ResNet mimarisine; RF, Xgboost ve LightGBM sınıflandırıcılarına; CMA-ES ve PSO meta-sezgisellerine özgü hiper-parametreler seçilmiş ve bu parametrelerin en iyi değerleri belirlenmiştir.

Bu tez çalışması kapsamında, en uygun melez çerçevenin belirlenmesi amacı ile, seçilen yaklaşımların farklı kombinasyonlarını içeren melez yöntemlere ait başarımları detaylı incelenmiştir. Bu bağlamda, deneysel çalışmalar 4 farklı melez yöntem üzerinde gerçekleştirilmiştir.

Uygulanan ilk melez çerçeve, BloodMNIST verisine ait öznitelikler ResNet mimarileri kullanılarak ön-eğitim aşamasından geçirilerek çıkarılmıştır. Çıkarılan bu öznitelikler kullanılarak kan hücrelerinin sınıflandırılması için RF, XGBoost ve LightGBM algoritmalarının başarımları araştırılmıştır. Bu incelemede, LightGBM sınıflandırıcı başarımının diğerlerinden daha iyi olduğu görülmüştür.

Çalışmadaki ikinci melez çerçeve, hesaplama karmaşıklığını düşürmek ve/veya aykırı değerlerden arındırarak genelleştirilmiş başarımı arttırmak amacı ile ResNet ve sınıflandırıcı ara katmanına boyut indirgeme yöntemleri eklenerek başarıma olan etkileri araştırılmıştır. Yapılan analizler, PCA ve ICA algoritmalarının başarıyı arttırdığını gösterirken UMAP ve t-SNE algoritmalarının başarıyı düşürdüğü gözlemlenmiştir.

Üçüncü melez çerçeve, daha seçkin özniteliklerin elde edilebilmesi amacıyla ResNet ve sınıflandırıcı ara katmanına ek olarak öznitelik seçme amacı ile meta-

sezgisel yaklaşımlar uygulanarak başarıma olan etkileri araştırılmıştır. Yapılan analizler, meta-sezgisel algoritmalarının özellikle hiper-parametre optimizasyonundan sonra başarıyı arttırdığı gözlemlenmiştir. CMA-ES ve PSO yöntemleri arasından en yüksek başarı PSO algoritması ile elde edilmiştir

Çalışmada uygulanan dördüncü melez çerçevede ise, meta-sezgisel algoritmalarla elde edilen başarıyı artırmak amacıyla meta-sezgisel ve sınıflandırıcı ara katmanına boyut indirgeme algoritmaları uygulayarak sonuçlar üzerindeki etkileri incelenmiştir. Yapılan analizler, meta-sezgisel ve boyut indirgeme algoritmalarının birlikte kullanılmasının başarıyı düşürdüğünü gösterirken, en iyi özniteliklerin meta-sezgisel yaklaşımlar ile seçilmesinin başarıyı arttırdığını göstermiştir.

Bu tez çalışması, ResNet mimarisi ile elde edilen ön-eğitim seti üzerinden PSO gibi meta-sezgisel yaklaşımlar ile elde edilen özniteliklerin LightGBM gibi ağaç-tabanlı yaklaşımlar ile melez kullanımının sınıflandırma başarımını iyileştirdiğini göstermiştir. Bu melez çerçeve, farklı derin öğrenme mimarilerine, öznitelik seçme amacı ile farklı meta-sezgisel yöntemlerin, farklı boyut indirgeme yaklaşımlarının ve farklı sınıflandırma yaklaşımlarının birlikte kullanılabileceği bir yapı sunmaktadır.

Anahtar Kelimeler: Derin Öğrenme, Rastgele Orman, XGBoost, LightGBM, ResNet18, ResNet34, PCA, ICA, Meta-Sezgisel, UMAP, t-SNE, CMA-ES, PSO

HYBRID DEEP LEARNING APPROACHES FOR THE MULTI CLASS MEDICAL IMAGE CLASSIFICATION

Zeliha Kaya Akçelik

ABSTRACT

This thesis presents a hybrid framework that combines deep learning architectures, meta-heuristic approaches, dimensionality reduction methods, and tree-based classification algorithms for the classification of blood cells with different characteristics. The study incorporates deep learning architectures such as Residual Network (ResNet), including ResNet18 and ResNet34, meta-heuristic methods such as covariance matrix adaptation evolution strategy (CMA-ES) and particle swarm optimization (PSO), dimensionality reduction methods like principal component analysis (PCA), independent component analysis (ICA), uniform manifold approximation and projection (UMAP), and t-distributed stochastic neighbor embedding (t-SNE), along with tree-based algorithms like Random Forest (RF), eXtreme gradient boosting (XGBoost), and light gradient-boosting machine (LightGBM).

Research has shown that deep learning architectures such as ResNet18 and ResNet34 are highly effective in image processing tasks. Furthermore, literature studies indicate that tree-based algorithms like RF, XGBoost, and LightGBM achieve very high classification performances. Building on the success of these approaches, the ResNet18 and ResNet34 models are used for feature extraction, while tree-based algorithms such as decision trees are employed for classification.

Additionally, past literature demonstrates that appropriate feature selection and/or dimensionality reduction approaches are important in mitigating overfitting, enhancing generalization performance, and/or reducing computational complexity

before classification. In this context, after the pre-training stage of the deep learning architecture, meta-heuristic algorithms like CMA-ES and PSO are used for feature selection, and dimensionality reduction algorithms such as PCA, ICA, UMAP, and t-SNE are used in different combinations to investigate their effects on classification performance.

The performance of the proposed hybrid methods is evaluated using the BloodMNIST dataset, which contains eight different types of blood cells from the MedMNIST dataset. Acknowledging that the performance of machine learning methods heavily depends on hyperparameter settings, specific hyperparameters were selected for the ResNet architecture, RF, XGBoost, LightGBM classifiers, and CMA-ES and PSO meta-heuristics, and the best values for these parameters were determined.

In this thesis, the performance of hybrid methods, incorporating different combinations of the selected approaches, was thoroughly examined to determine the most suitable hybrid framework. In this context, experimental studies were conducted on four different hybrid methods.

The first hybrid framework involved extracting features from the BloodMNIST dataset using ResNet architectures during the pre-training stage, followed by investigating the classification performance of RF, XGBoost, and LightGBM algorithms using these extracted features. In this examination, it was observed that the LightGBM classifier outperformed the others.

The second hybrid framework investigated the effects of adding dimensionality reduction methods between the ResNet and classifier layers to reduce computational complexity and/or remove outliers to improve generalization performance. The analysis showed that PCA and ICA algorithms improved performance, while UMAP and t-SNE algorithms decreased it.

In the third hybrid framework, meta-heuristic approaches were applied between the ResNet and classifier layers for feature selection to obtain more refined features, and their effects on performance were examined. The analysis revealed that meta-heuristic algorithms, particularly after hyperparameter optimization, improved

performance. Among CMA-ES and PSO methods, the highest performance was achieved with the PSO algorithm.

In the fourth hybrid framework, meta-heuristic algorithms were combined with dimensionality reduction algorithms between the meta-heuristic and classifier layers to further enhance performance. The analysis showed that using meta-heuristic and dimensionality reduction algorithms together decreased performance, while selecting the best features using meta-heuristic approaches improved classification performance.

This thesis demonstrates that hybrid use of feature extraction through ResNet architectures, feature selection with meta-heuristic approaches such as PSO, and classification with tree-based approaches such as LightGBM improves classification performance. The hybrid framework offers a structure that can incorporate various deep learning architectures, different meta-heuristic methods for feature selection, diverse dimensionality reduction approaches, and different classification methods.

Keywords: Deep Learning, Random Forest, XGBoost, LightGBM, ResNet18, ResNet34, PCA, ICA, meta-heuristic, UMAP, t-SNE

ÖN SÖZ

Tez çalışmasında, derin öğrenme ve ağaç tabanlı algoritmalar melez çerçevesinde kan hücreleri verilerinin sınıflandırma başarımları değeri sunulmuştur. Çalışma kapsamında, ilk melez çerçevede derin öğrenme modellerinden ön-eğitim (ing:pre-train) edilmiş ResNet18 ve ResNet34 modelleri ile öznelikler çıkartılmış ve bu öznelikler sınıflandırma algoritmaları için girdi olarak hazırlanmıştır. Sınıflandırma bölümünde ise ağaç tabanlı algoritmalar olan Rastgele Orman, XGBoost ve LightGBM kullanılmış ve bu algoritmaların sınıflandırma başarımının artırılması amacıyla hiper-parametre optimizasyonu yapılmıştır. İkinci melez çerçevede, ResNet-ağaç temelli algoritma melez çerçeve ara katmanına boyut indirgeme algoritmalarından PCA, ICA, UMA ve t-SNE uygulanarak veri seti daha az boyut ile temsil edilmiş ve sınıflandırma başarımına olan etkisi incelenmiştir. Üçüncü melez çerçevede, ResNet-ağaç temelli algoritma melez çerçeve ara katmanına meta-sezgisel algoritmalarından CMA-ES ve PSO algoritmaları uygulanarak daha seçkin özneliklerin seçilmesinin sınıflandırma başarımına olan etkisi incelenmiştir. Dördüncü melez çerçevede ise üçüncü melez çerçeveye ek olarak ara katmana boyut indirgeme algoritmaları uygulanarak sınıflandırma başarımına olan etkisi incelenmiştir.

Kan hücrelerinin sınıflandırılmasında derin öğrenme yöntemlerinin başarısı, doğru hiper-parametre ayarlamalarına (ing: tuning) bağlıdır. Bu çalışma, ön-eğitim (ing: pre-train) edilmiş ResNet18 ve ResNet34 modellerinin birleştirilerek farklı ağaç tabanlı algoritmalar ile sınıflandırılmasının kan hücresi verilerinde nasıl kullanıldığını, hiper-parametrelerin nasıl optimize edildiğini, çerçeveye eklenen meta-sezgisel yöntemlerin doğru öznelikleri yakalayarak sınıflandırma başarımını arttırdığını ve elde edilen sonuçların ne kadar etkileyici olduğunu incelemektedir. Ayrıca, ICA ve PCA algoritmalarının öznelik seçme aşamasında iyileştirme sağladığını ve veri kümesinin daha az boyut ile temsil edilebileceğini ele alırken uygulanan UMAP ve t-SNE algoritmalarının bu melez çerçeve için uygun olmadığını göstermektedir.

Bu tezde sunulan sonuçlar, kan hücrelerinin sınıflandırılmasında derin öğrenme ve ağaç temelli algoritmalar melez çerçevesinde sınıflandırma başarısını, hiper-parametre optimizasyonu algoritmaların etkinliğini, meta-sezgisel algoritmalar ile daha seçkin özneliklerin çıkarımı ve boyut indirgeme algoritmalarının sınıflandırma başarımına olan etkisini vurgulamaktadır. Bu çalışma, gelecekteki araştırmalara ışık tutmak amacıyla gerçekleştirilmiştir.

İÇİNDEKİLER

ÖZET.....	v
ABSTRACT	viii
ÖN SÖZ.....	xi
ŞEKİL LİSTESİ.....	xiv
TABLO LİSTESİ.....	xv
SEMBOLLER	xvii
KISALTMALAR	xix
GİRİŞ	1
BİRİNCİ BÖLÜM.....	4
1. TEMEL KAVRAMLAR	4
1.1. DERİN ÖĞRENME	4
1.1.1. ResNet Modeli	4
1.2. AĞAÇ TABANLI ALGORİTMALAR	7
1.2.1. RF Sınıflandırması (RF Classifier)	7
1.2.2. XGBoost Sınıflandırması (XGBoost Classifier)	7
1.2.3. LightGBM Sınıflandırması (LightGBM Classifier)	8
1.3. BOYUT İNDİRGEME ALGORİTMALARI	8
1.3.1. Temel Bileşenler Analizi (Principal Component Analysis [PCA]).....	8
1.3.2. Bağımsız Bileşenler Analizi (Independent Component Analysis [ICA])	13
1.3.3. Uniform Manifold Approximation and Projection (UMAP).....	14
1.3.4. t- Distributed Stochastic Neighbor Embedding (t-SNE).....	15
1.4. META-SEZGİSEL ALGORİTMALAR.....	16
1.4.1. Covariance Matrix Adaptation Evolution Strategy CMA-ES)	17
1.4.2. Particle Swarm Optimization (PSO).....	20
1.5. DERİN ÖĞRENME VE AĞAÇ TEMELLİ ALGORİTMALAR İLE İLGİLİ LİTERATÜR ÇALIŞMASI	21
İKİNCİ BÖLÜM	23
2. ÖNERİLEN YÖNTEMLER	23
2.1. DERİN ÖĞRENME-AĞAÇ TEMELLİ ALGORİTMA MELEZ ÇERÇEVESİ	23
ÜÇÜNCÜ BÖLÜM	26
3. DENEYSEL ÇALIŞMALAR.....	26
3.1. DENEYSEL TASARIM	26
3.2. VERİ SETİ	26

3.3. DEĞERLENDİRME METRİKLERİ.....	27
3.4. HİPER-PARAMETRE DEĞERLERİ.....	27
DÖRDÜNCÜ BÖLÜM	31
4. BULGULAR VE TARTIŞMALAR	31
4.1. BİRİNCİ MELEZ ÇERÇEVE İÇİN SONUÇLAR.....	31
4.2. İKİNCİ MELEZ ÇERÇEVE İÇİN SONUÇLAR	32
4.3. ÜÇÜNCÜ MELEZ ÇERÇEVE İÇİN SONUÇLAR.....	34
4.4. DÖRDÜNCÜ MELEZ ÇERÇEVE İÇİN SONUÇLAR	36
SONUÇ VE GELECEK ÇALIŞMA	40
KAYNAKÇA	42



ŞEKİL LİSTESİ

	Sayfa
Şekil 1. 1 Artık öğrenmesi (He, 2016).....	5
Şekil 1. 2 ResNet18 Mimarisi (He, 2016).....	6
Şekil 1. 3 ResNet18 Mimarisi (He, 2016).....	6
Şekil 1. 4 CMA-ES algoritmasının akış diyagramı.....	19
Şekil 1. 5 PSO akış diyagramı.....	21
Şekil 2. 1 Birinci melez çerçevenin akış diyagramı	23
Şekil 2. 2 İkinci melez çerçeve için akış diyagramı	24
Şekil 2. 3 Üçüncü melez çerçeve için akış diyagramı.....	24
Şekil 2. 4 Dördüncü melez çerçeve için akış diyagramı	25
Şekil 3. 1 BloodMNIST veri kümesindeki örnek kan hücresi görüntüleri: (a) bazofil, (b) eozinofil, (c) eritrosit, (d) normoblast, (e) lenfosit, (f) monosit, (g) nötrofil, (h) trombosit	27
Şekil 4. 1 İkinci melez çerçevede PCA algoritması ile (a) LightGBM algoritması için hata matrisi, (b) XGBoost algoritması için hata matrisi, (c) RF algoritması için hata matrisi.....	38
Şekil 4. 2 Üçüncü melez çerçevede LightGBM sınıflandırması ile (a) CMA-ES algoritması uygulandığında elde edilen hata matrisi (b) PSO algoritması uygulandığında elde edilen hata matrisi.....	38

TABLO LİSTESİ

Tablo 3. 1: Algoritmalar için bulunan hiper-parametre aralıkları.....	28
Tablo 3. 2: Algoritmalar için bulunan hiper-parametre değerleri	29
Tablo 4. 1: ResNet-Ağaç tabanlı sınıflandırıcı melez çerçevesinin varsayılan değerleri ile elde edilen sonuçlar	32
Tablo 4. 2: ResNet-Ağaç tabanlı algoritma melez çerçevesinde ağaç tabanlı algoritmaların hiper-parametre optimizasyonu sonucunda bulunan en iyi değerler ile elde edilen sonuçlar.....	32
Tablo 4. 3: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PCA uygulanması sonucunda elde edilen sonuçlar	33
Tablo 4. 4: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına ICA uygulanması sonucunda elde edilen sonuçlar	33
Tablo 4. 5: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına UMAP uygulanması sonucunda elde edilen sonuçlar	33
Tablo 4. 6: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına t-SNE uygulanması sonucunda elde edilen sonuçlar	34
Tablo 4. 7: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına CMA-ES uygulanması sonucunda elde edilen sonuçlar	34
Tablo 4. 8: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına hiper-parametre optimizasyonu yapılan CMA-ES uygulanması sonucunda bulunan en iyi değerler ile elde edilen sonuçlar.....	35
Tablo 4. 9: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PSO uygulanması sonucunda elde edilen sonuçlar	35
Tablo 4. 10: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına hiper-parametre optimizasyonu yapılan PSO uygulanması sonucunda bulunan en iyi değerler ile elde edilen sonuçlar.....	35
Tablo 4. 11: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PCA uygulanması sonucunda elde edilen sonuçlar	36

Tablo 4. 12: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına ICA uygulanması sonucunda elde edilen sonuçlar 37

Tablo 4. 13: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına UMAP uygulanması sonucunda elde edilen sonuçlar..... 37

Tablo 4. 14: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına t-SNE uygulanması sonucunda elde edilen sonuçlar 37



SEMBOLLER

t	: Beklenen değer
$H(x)$	: Öğrenilmesi gereken hedef fonksiyonu
x	: Giriş
w	: Katman ağırlığı
N	: Toplam veri noktalarının sayısı
i	: Veri noktalarının indisleri
j	: Veri noktalarının indisleri
s	: Kovaryans
R	: Kovaryans Matrisi
X_s	: Standartlaştırılmış veri matrisi
t_1	: R matrisinin 1. özvektörü
λ_1	: R matrisinin 1. Özdeğeri
y_1	: 1. Temel bileşen
t_2	: R matrisinin 2. Özvektörü
λ_2	: R matrisinin 2. Özdeğeri
y_2	: 2. Temel bileşen
T	: Dönüşüm matrisi
m	: Ortalama
σ	: Adım sayısı
λ	: Popülasyonun büyüklüğü
μ	: Seçilen bireyler sayısı
z	: Normal dağılımdan örneklenen vektör
$v_i(t)$	: Parçacığın mevcut durumundaki hız vektörü.
w	: Parçacığın kendi yönündeki hareket katsayısı
$x_i(t)$	: Parçacığın mevcut durumundaki konum vektörü.
$p_{best,i}(t)$	: Parçacığın geçmiş bilgilerinden elde ettiği en iyi konum.
$g_{best,i}(t)$	: Tüm parçacıklar kapsamında elde edilen en iyi konum.

- c_1 : Parçacığın kendi geçmişinden gelen öğrenme katsayısı.
- c_2 : Parçacığın toplumun geçmişinden gelen öğrenme katsayısı
- r_1 ve r_2 : $(0,1)$ aralığında tanımlanan rastgele değerler



KISALTMALAR

DÖ	Derin Öğrenme
RF	Rasndom Forest
XGBoost	Extreme Gradient Boosting
ark.	Arkadaşları
CNN	Convolutional Neural Network (Evrişimsel Sinir Ağı)
LightGBM	Light Gradient Boosting Machine
ed. veya haz.	Editör/yayına hazırlayan
ing	İngilizce
ReLU	Rectified Linear Unit (Doğrultulmuş Doğrusal Birim)
s.	Sayfa/sayfalar
PCA	Principal Compenent Analysis
ICA	Independent Compenent Analysis
UMAP	Uniform Manifold Approximation and Projection
t-SNE	t-Distributed Stochastic Neighbor Embedding
CMA-ES	Covariance Matrix Adaption Evolution Strategy
PSO	Particle Swarm Optimization
ResNet	Residual Network

GİRİŞ

Kan hücreleri, vücudun sağlıklı işleyişi için hayati öneme sahiptir. Bu hücreler, enfeksiyonlara karşı koruma, vücudu savunma, oksijen taşıma ve kanın pıhtılaşması gibi birçok temel vücut işlevinde kritik roller üstlenir. Kan hücreleri genel olarak kırmızı, beyaz ve trombosit olmak üzere üç ana gruba ayrılmaktadır. Bu hücrelerin doğru bir şekilde sınıflandırılması ve sayılması, çeşitli hastalıkların teşhis ve tedavisinde hayati öneme sahiptir. Ancak, bu hücrelerin manuel olarak sınıflandırılması ve sayılması, yüksek derecede uzmanlık gerektiren ve zaman alıcı bir işlemdir. Geleneksel yöntemlerle yapılan bu analizler, mikroskop altında hücrelerin tek tek incelenmesini ve sınıflandırılmasını içerir. Bu süreç hem yoğun emek gerektirir hem de insan hatasına açık olabilir. Bu nedenle, yüksek hassasiyetle otomatik olarak kan hücrelerini sınıflandırabilen yöntemlere olan ihtiyaç giderek artmaktadır. Otomatik sınıflandırma sistemleri hem zaman kazandırır hem de daha tutarlı ve doğru sonuçlar elde edilmesini sağlar.

Geleneksel yöntemler, mikroskop altında hücre morfolojisinin incelenmesine dayanırken, son yıllarda bilgisayarlı görüntü işleme ve yapay zeka teknikleri bu alanda devrim yaratmıştır. Özellikle derin öğrenme algoritmaları, kan hücrelerini yüksek doğrulukla sınıflandırabilme potansiyeli sunmaktadır. Derin öğrenme, çok katmanlı yapay sinir ağlarının (ing: deep neural networks) kullanıldığı bir makine öğrenmesi alt dalıdır. Derin öğrenme modelleri, özellikle görüntü işleme görevlerinde etkileyici sonuçlar elde etmektedir. Kan hücreleri sınıflandırmasında, derin öğrenme modelleri genellikle öznitelik çıkarımı ve sınıflandırma olarak iki ana aşamada kullanılmaktadır.

Derin öğrenme mimarilerinin tek başına kullanımı genellikle başarılı sonuçlar verirken, birden fazla modelin birleşimi ile performans daha da artırılabilir. Bu teknik ile farklı modellerin güçlü yönlerini birleştirerek hata oranlarını düşürmek amaçlanmaktadır. Bu çalışmanın ilk melez çerçevesinde, ResNet18 ve ResNet34 modelleri ön-eğitim edilmiş ağırlıklarla birleştirilmiş ve öznitelik çıkarımı için kullanılmıştır. Elde edilen bu öznitelik vektörü ağaç tabanlı sınıflandırma algoritmalarına girdi olarak verilmiş ve sınıflandırma işlemi tamamlanmıştır. Her bir

algoritmanın doğruluk, kesinlik, duyarlılık ve F1 skor değerleri karşılaştırılmıştır. İkinci melez çerçevede, modellerin birleştirilmesinden sonra ara katmana hesaplama karmaşıklığını azaltmak ve genel performansı iyileştirmek amacıyla boyut indirgeme algoritmaları uygulanmıştır. Bu süreç, gereksiz veya gürültülü özelliklerin elimine edilmesini sağlar ve modelin genelleme kabiliyetini artırır. Derin öğrenme modellerinde, özellikle ara katmanlarda öznitelik seçimi uygulamak, model performansı için avantaj sağlamaktadır. Üçüncü melez çerçevede, modellerin birleştirilme aşamasından sonra yine hesaplama karmaşıklığını azaltmak ve aşırı öğrenmeyi engellemek amacıyla meta-sezgisel yöntemlerden CMA-ES ve PSO algoritmaları kullanılarak daha seçkin öznitelikler elde edilmesi hedeflenmiştir. Dördüncü melez çerçevede ise, modellerin birleştirilmesi ve meta-sezgisel yöntemler ile daha seçkin özniteliklerin seçilmesi aşamasından sonra ara katmana boyut indirgeme algoritmaları uygulanarak performansa olan etkisi incelenmiştir.

Sınıflandırma ve meta-sezgisel algoritmalarının performansını arttırmak hiperparametreleri optimize etmek ile sağlanmaktadır. Bu çalışmada, sınıflandırma algoritmalarından RF algoritması için tahmin edici sayısı (ing: estimator), maksimum özellik sayısı (ing: maximum features), maksimum derinlik sayısı (ing: maximum depth) ve kriter (ing: criterion) parametreleri optimize edilmiştir. XGBoost algoritması için tahmin edici sayısı (ing: number of estimator), öğrenme katsayısı (ing: learning rate), maksimum derinlik sayısı (ing: maximum depth) ve alt örnek (ing: subsample) parametreleri optimize edilmiştir. LightGBM algoritması için tahmin edici sayısı (ing: estimator), öğrenme katsayısı (ing: learning rate), yaprakların sayısı (ing: number of leaves) ve kriter (ing: criterion) parametreleri optimize edilmiştir. Meta-sezgisel algoritmalarından PSO algoritması için sürü boyutu (ing: swarm size), maksimum iterasyon sayısı (ing: maximum iteration size) ve atalet ağırlığı (ing: inertia weight) parametreleri optimize edilmiştir. CMA-ES algoritması için ise popülasyon boyutu (ing: population size), adım sayısı (ing: step size), maksimum iterasyon sayısı (ing: maximum iteration size) optimize edilmiştir.

Bu çalışmanın amacı, derin öğrenme modellerinden ResNet mimarisiyle elde edilen ön-eğitim seti üzerinden ağaç tabanlı yaklaşımlar ile melez kullanımının sınıflandırma başarısına olan etkisini gözlemlemektir. Karmaşıklığı azaltmak ve

sınıflandırma başarısını arttırmak amacıyla bu melez çerçevenin ara katmanlarına meta-sezgisel ve boyut indirgeme algoritmaları birlikte ve ayrı şekilde uygulanarak sınıflandırma başarısına olan etkileri incelenmiştir.

Yapılan çalışmanın ilk bölümünde, derin öğrenme, sınıflandırma algoritmaları, boyut azaltma algoritmaları ve meta-sezgisel algoritmalarına ait bilgiler içermektedir.

İkinci bölümünde, birinci bölümde bahsedilen yöntemlerden derin öğrenme, meta-sezgisel algoritmalar ve ağaç tabanlı algoritmaların kullanıldığı melez bir çerçeve sunulmuştur ve çalışmanın akışı belirtilmiştir. Litaratürdeki kan hücresi sınıflandırmasındaki melez çalışmalara yer verilmiştir.

Üçüncü bölümde, BloodMNIST veri kümesi üzerinde ResNet18 ve ResNet34, meta-sezgisel ve Ağaç tabanlı algoritmalar melez yönteminin deneysel çalışmaları yapılmıştır.

Dördüncü bölümde ise üçüncü bölümde yapılmış olan deneylerin bulguları ve tartışılması olası durumları anlatılmaktadır.

BİRİNCİ BÖLÜM

1. TEMEL KAVRAMLAR

Bu bölümde derin öğrenme ve ağaç tabanlı algoritmalara ilişkin temel kavramlar ve literatürdeki çalışmalar incelenmiştir.

1.1. DERİN ÖĞRENME

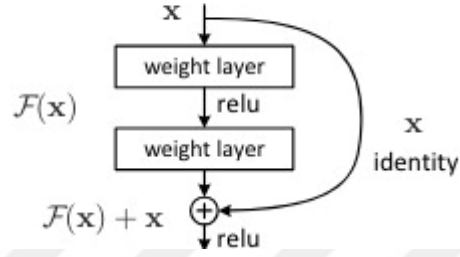
Derin öğrenme makine öğrenmesinin bir alt dalıdır. Derin öğrenme, ardışık katmanlardan oluşan yapay sinir ağlarını kullanarak verileri işler ve karmaşık örüntüleri öğrenir. Bu teknoloji, görüntü tanıma, ses tanıma, doğal dil işleme ve birçok başka alanda devrim yaratmıştır (LeCun ve ark., 2015).

Derin öğrenmede, verilerin farklı özellik seviyelerini veya temsillerini öğrenmeye dayalı bir yapı bulunmaktadır. Alt düzey özelliklerden türetilen üst düzey özellikler, hiyerarşik bir temsil oluşturur. Bu yapı, farklı soyutlama seviyelerine karşılık gelen çeşitli temsil seviyelerini öğrenir (Şeker, 2017).

1.1.1. ResNet Modeli

ResNet (Residual Network), derin öğrenme modellerinde derinliği artırmanın getirdiği zorlukları aşmak için geliştirilmiş bir yapay sinir ağı mimarisidir. İlk olarak (He, 2015) makalesinde tanıtılmıştır. ResNet, özellikle derin sinir ağlarının eğitiminde karşılaşılan sorunları çözmesiyle bilinir ve bu alanda önemli bir dönüm noktası olmuştur. Geleneksel derin sinir ağlarında katman sayısı arttıkça, model karmaşılaşır ve aşırı öğrenme (ing: overfitting) durumu gerçekleşebilir. Bundan dolayı modelin performansı genellikle belirli bir noktadan sonra düşmeye başlamaktadır. Bu durum, kaybolan gradyan (ing: vanishing gradient) ve azalan performans (ing: degrading performance) sorunlarına yol açar. Derin ağların eğitimi sırasında, gradyanların geriye doğru yayılımı zayıflayarak erken katmanların öğrenmesini zorlaştırır. ResNet, bu sorunları aşmak için artık öğrenme (ing: residual learning) adı verilen bir teknik

kullanılır. Bu teknik, her birkaç katmanda bir "atlama bağlantısı" (ing: skip connection) ekleyerek, katmanların çıkışlarını bir sonraki katmana doğrudan iletir. Böylece, derin ağların öğrenmesi kolaylaştırılır ve daha derin ağların eğitimi mümkün hale gelir. Şekil 1.1.'de artık katmanının öğrenmesi gösterilmektedir.



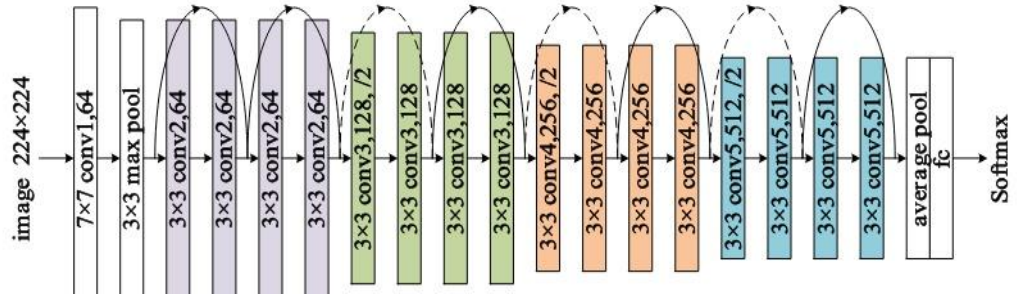
Şekil 1. 1 Artık öğrenmesi (He, 2016)

Bir artık blok denklem 1.1'deki formül ile tanımlanır.

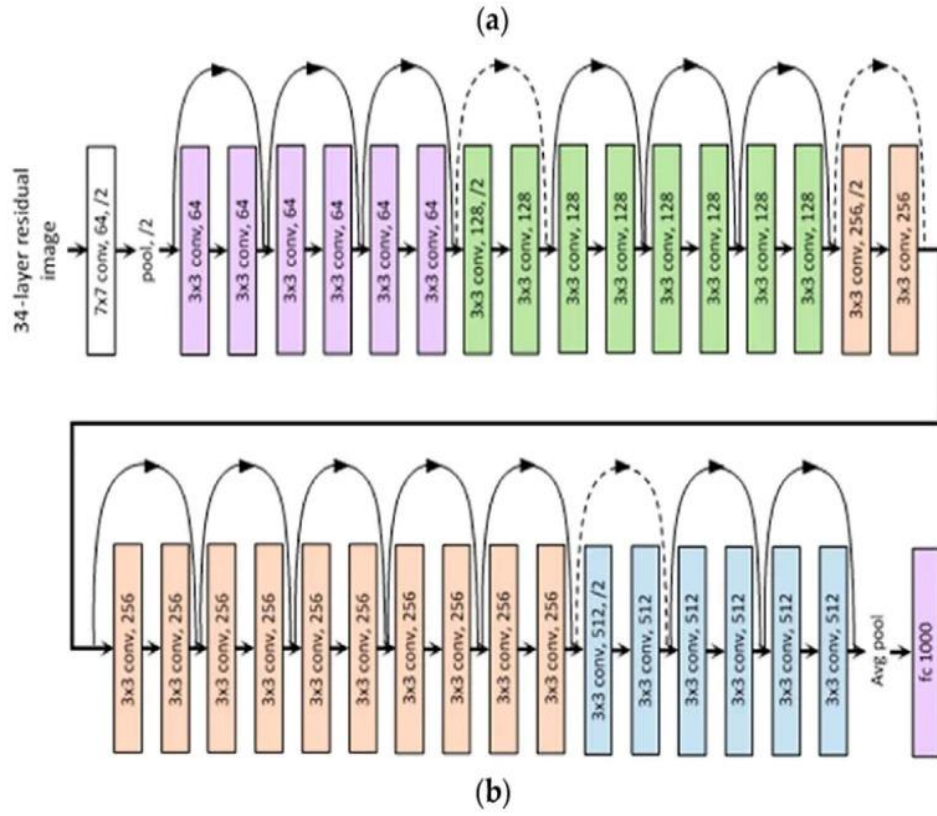
$$H(x) = F(x, \{W_i\}) + x \quad (1.1.)$$

Bu denklemde $H(x)$; öğrenilmesi gereken hedef fonksiyonu, $F(x, \{W_i\})$; öğrenilen artık fonksiyonu, x ; girişi ve $\{W_i\}$; katman ağırlıklarını temsil etmektedir.

ResNet mimarisi, artık blokların üst üste eklenmesiyle oluşturulmaktadır. Temel ResNet modelleri, katman sayısına göre isimlendirilir. ResNet18, 18 katmanlı model, ResNet34, 34 katmanlı model, ResNet50, 50 katmanlı model, ResNet101, 101 katmanlı model ve ResNet152, 152 katmanlı modelden oluşmaktadır. Bu modeller genel olarak ilk konvolüsyon katmanı (Conv1), birkaç konvolüsyon katmanından oluşan artık blokları (ing: residual blocks), özellik haritalarını azaltmak için maksimum havuzlama (ing: max pooling) ve sınıflandırma için kullanılan tam bağlantılı katmandan (ing: fully connected layer) oluşmaktadır. Şekil 1.2.'de ResNet18, Şekil 1.3.'te ResNet34 mimarisi yer almaktadır.



Şekil 1. 2 ResNet18 Mimarisi (He, 2016)



Şekil 1. 3 ResNet18 Mimarisi (He, 2016)

1.2. AĞAÇ TABANLI ALGORİTMALAR

1.2.1. RF Sınıflandırması (RF Classifier)

RF algoritması, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan güçlü bir makine öğrenme algoritmasıdır. İlk olarak (Breiman, 2001) makalesinde tanıtılmıştır. Algoritmanın etkinliği ve esnekliği sayesinde günümüzde hala geniş bir uygulama alanına sahiptir. Rastgele orman algoritması birkaç temel prensip üzerine kurulmuştur. Her bir karar ağacı oluşturulurken, her düğümde rastgele bir alt özellik seti seçilir ve en iyi bölünme bu alt set üzerinde yapılır. Bu, ağaçlar arasındaki korelasyonu azaltarak genel modelin performansını artırır.

Algoritmanın çalışma adımları şu şekildedir; eğitim veri setinden rastgele örnekler seçilir, her ağaç oluşturulurken, her düğümde rastgele bir alt özellik seti seçilir ve en iyi bölünme bu alt set üzerinde yapılır. Sınıflandırma problemi için her bir ağaç, bir sınıf tahmini yapar ve Rastgele Orman algoritması bu tahminleri birleştirerek en çok oy alan sınıfı nihai tahmin olarak seçer. Rastgele Orman, özelliklerin önemini değerlendirmek için kullanılabilir, bu da hangi özelliklerin tahminlerde daha etkili olduğunu belirlemede yardımcı olur (Genuer, Poggi, & Tuleau-Malot, 2010). Tek bir karar ağacının aksine, Rastgele Orman modelleri aşırı öğrenmeye karşı daha dirençlidir (Cutler et al., 2012).

1.2.2. XGBoost Sınıflandırması (XGBoost Classifier)

XGBoost, yüksek performans ve hız sunan bir gradyan arttırma (ing: gradient boosting) çerçevesidir. İlk olarak (Chen & Guestrin, 2016) makalesinde tanıtılmış olup, Kaggle yarışmalarında ve çeşitli endüstriyel uygulamalarda başarıyla kullanılmaktadır. XGBoost, özellikle büyük veri setlerinde ve karmaşık modellere ihtiyaç duyulan uygulamalarda etkili sonuçlar vermesiyle öne çıkmaktadır. Hem sınıflandırma hem de regresyon problemlerinde kullanılıp kullanıcıya geniş bir parametre ayarlama olanağı sunmaktadır (Li et al., 2020). XGBoost, gradyan arttırma (ing: gradient boosting) algoritmasının bir iyileştirmesidir. Paralel hesaplama kullanarak eğitim süresini oldukça azaltmaktadır. Oluşturulan ağaçta aşırı öğrenmeyi engellemek amacıyla budama işlemi yapılır (Kaya, 2023).

Algoritmanın çalışma adımları şu şekildedir; model basit bir tahmin ile başlar. Her iterasyonda yeni bir karar ağacı eklenir ve bu ağaç, mevcut modelin hatalarını düzeltmeye çalışır. Her yeni ağaç, mevcut hatalara göre optimize edilir ve bu süreç, belirli bir iterasyon sayısına veya hata oranı düşüşüne ulaşılan kadar devam eder. Son tahmin, tüm ağaçların birleşimiyle elde edilir. Her ağacın katkısı, belirli bir öğrenme oranı ile ağırlıklandırılır.

1.2.3. LightGBM Sınıflandırması (LightGBM Classifier)

LightGBM, Microsoft tarafından geliştirilmiş, gradyan arttırma (ing: gradient boosting) çerçeve tabanlı bir makine öğrenme algoritmasıdır. Yüksek verimlilik ve düşük bellek kullanımı ile öne çıkan LightGBM, büyük veri setleri ve yüksek boyutlu özellik uzaylarında üstün performans gösterir (Ke et al., 2017). LightGBM, gradyan arttırma yönteminin bir optimizasyonu olup, özellikle endüstriyel uygulamalarda hız ve doğruluk açısından önemli avantajlar sunmaktadır. LightGBM iyileştirme üzerine kurulmuş bir algoritmadır ve diğer, gradyan arttırma algoritmalarına kıyasla daha karmaşık hiper-parametre ayarları gerektirebilir (Guo et al., 2021). Geleneksel ağaç büyütme yöntemlerinde derinlik seviyesine göre bölünme yapılırken, LightGBM, yaprak (ing: leaf) bazlı büyütme yöntemini kullanır. Bu, daha büyük bir kayıp azaltımı sağlar ve daha derin ağaçlar oluşturur.

Algoritmanın çalışma adımları şu şekildedir; veri setindeki özellikler gruplandırılır ve veri örnekleme işlemi optimize edilir. LightGBM, yaprak bazlı büyütme yöntemini kullanarak her iterasyonda yeni ağaçlar oluşturur. Her yeni ağaç, mevcut modelin hatalarını azaltmak için optimize edilir ve model iterasyonlar boyunca güncellenir.

1.3. BOYUT İNDİRGEME ALGORİTMALARI

1.3.1. Temel Bileşenler Analizi (Principal Component Analysis [PCA])

PCA yüksek boyutlu veri setlerinin boyutunu azaltmak ve verideki önemli desenleri ortaya çıkarmak için kullanılan yaygın bir istatistiksel tekniktir. 1933 yılında Karl Pearson tarafından tanıtılan PCA, veri analizi ve makine öğrenme uygulamalarında sıkça kullanılır (Pearson, 1901). PCA, veri setindeki değişkenler

arasındaki korelasyonları kullanarak, en büyük varyansa sahip doğrusal kombinasyonları belirler ve veriyi bu yeni eksenlere projekte eder.

PCA' in üç temel amacı bulunmaktadır. Bunlar;

- Veri boyutlarını azaltmak,
- Tahminleme yapmak,
- Veri setini analizler için görüntülemek.

PCA istatistiksel bir metot olup; veri setinin açıklanması, set içerisindeki örüntünün tanımlanması ve veri içerisindeki farklı ya da benzer örüntülerin tanımlanmasında kullanılabilmektedir. PCA' da boyut azaltma işlemi uygulanırken veri kaybı yaşanmaması çok büyük bir avantaj sağlamaktadır. PCA tekniği bilgisayar bilimlerinde görüntü işleme alanında sıkça kullanılan bir yöntemdir.

PCA uygulaması altı ana adımdan oluşmaktadır. Bunlar aşağıdaki gibi sıralanabilir;

- Veri setlerinin hazırlanması,
- Verileri ortalamadan çıkarma,
- Kovaryans/ Korelasyon matrisini oluşturma,
- Öz değer ve öz vektörlerin bulunması,
- Bileşenlerin seçilmesi ve özellik vektörünün oluşturulması,
- Yeni veri setlerinin oluşturulması

PCA' in daha anlaşılır bir şekilde incelenebilmesi için bazı matematiksel ifade ve işlemlerin bilinmesi gerekmektedir. PCA hesaplamalarında standart sapma ve varyans bulma, kovaryans/ korelasyon matrisleri bulma ve kovaryans/ korelasyon matrislerinin öz değer ve öz vektörlerini hesaplama gibi işlemlere ihtiyaç duyulduğundan bu işlemlerin matematiksel teori ve uygulamalarını bilmek önem arz etmektedir.

Temel bileşenlerin elde edilmesi aşağıda açıklanmıştır.

$$\text{Bireyler} \left\{ \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{(n-1)1} & x_{(n-1)2} & \dots & x_{(n-1)j} & x_{(n-1)p} \\ x_{n1} & x_{n2} & \dots & x_{nj} & x_{np} \end{bmatrix} \right. \quad (1.2)$$

($X=(X_{ij} \text{ n} \times \text{p}$ boyutunda gözlem matrisi, $X_{ij} = i$. birey için j . değişkenin gözlem değeri, $i=1,2,3,\dots,n, j=1,2,3,\dots,p$)

Yukarıdaki matrsten yararlanarak temel bileşen analizinin gerçekleştirilebilmesi için izlenmesi gereken adımlar aşağıdaki gibidir.

- Varyans-kovaryans ya da R-korelasyon matrisinin hesaplanması
- Hesaplanan varyans-kovaryans veya R korelasyon matrislerinin öz değerlerinin ve bunlara karşılık gelen öz vektörlerinin bulunması
- Temel bileşenlerin hesaplanması.

$$Kovaryans \left\{ \begin{bmatrix} \frac{X_{11}-\bar{X}_1}{S_1} & \frac{X_{12}-\bar{X}_2}{S_2} & \dots & \frac{X_{1j}-\bar{X}_j}{S_j} & \frac{X_{1p}-\bar{X}_p}{S_p} \\ \frac{X_{21}-\bar{X}_1}{S_1} & \frac{X_{22}-\bar{X}_2}{S_2} & \dots & \frac{X_{2j}-\bar{X}_j}{S_j} & \frac{X_{2p}-\bar{X}_p}{S_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{X_{(n-1)1}-\bar{X}_1}{S_1} & \frac{X_{(n-1)2}-\bar{X}_2}{S_2} & \dots & \frac{X_{(n-1)j}-\bar{X}_j}{S_j} & \frac{X_{(n-1)p}-\bar{X}_p}{S_p} \\ \frac{X_{n1}-\bar{X}_1}{S_1} & \frac{X_{n2}-\bar{X}_2}{S_2} & \dots & \frac{X_{nj}-\bar{X}_j}{S_j} & \frac{X_{np}-\bar{X}_p}{S_p} \end{bmatrix} \right. \quad (1.3)$$

$$s = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_k), i = 1, 2, \dots, n; k \quad (1.2)$$

$$= 1, 2, \dots, p$$

$$R = r_{ij} = \frac{S_{ij}}{(S_{ii}S_{jj})^{1/2}} \quad (1.3)$$

(X_s : ortalaması 0 ya da varyansı 1 olan ($n \times p$) standartlaştırılmış veri matrisi)

$$Y_{p \times n} = T_{p \times p} Z_{p \times n} \quad (1.4)$$

$$E(Y) = E(T'Z) = T'E(Z) = 0 \quad (1.5)$$

$$Var(Y) = T'E(ZZ')T = T'RT \quad (1.5)$$

$R_{p \times p}$ boyutlu değişkenler arası korelasyon matrisi olmaktadır. Dönüştürülmüş Y matrisinin vektörleri birbirine dik olabilmesi için $Var(Y)$ matrisinin köşegen olması gerekmektedir. Bu matris köşegenlerinde birçok T dönüşüm matrislerinden amaca en

uygun olanının seçilebilmesi için y vektörleri üzerine bazı kısıtlayıcılar konması gerekmektedir:

1. Kısıt: y vektörlerinin birincisi olan y_1 öyle seçilmeli ki varyans değeri en büyük olsun.

$$Var(y_1) = \max \frac{1}{n-1} \sum_{i=1}^n (y_{1i})^2 \quad (1.6)$$

2. Kısıt: y_1 vektörü bulunurken kullanılan t_1 vektörü elemanlarının kareleri toplamı 1 olmalıdır.

$$t_1' t_1 = 1 \quad (1.7)$$

$$y_{1i} = t_1' z_i \quad (1.8)$$

Kısıt 1 yardımı ile;

$$Var(y_1) = \max \frac{1}{n-1} \sum_{i=1}^n (y_{1i})^2 \quad (1.9)$$

$$= \frac{1}{n-1} \sum_{i=1}^n (t_1' z_i)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n t_1' z_i z_i' t_1$$

$$= t_1' \frac{1}{n-1} Z Z' t_1$$

$$= t_1' R t_1$$

Kısıt 2 yardımı ile;

$$\varphi_1 = t_1' R t_1 - \lambda (t_1' t_1 - 1) \quad (1.10)$$

$$\frac{\partial \varphi_1}{\partial t_1} = 2 R t_1 - 2 \lambda_1 t_1 = 0 \quad (1.11)$$

$$(R - \lambda_1 I) t_1 = 0 \dots (1) \quad (1.12)$$

(λ_1 , R matrisinin özdeğeri; t_1 , R matrisinin özvektörü) elde edilmektedir.

Öz değerleri elde etmek için $|R - \lambda I| = 0$ açılımından elde edilen p. dereceden polinom denklemden p tane λ değeri bulunur. R matrisi pozitif tanımlı ve simetrik olduğundan elde edilecek değerlerin hepsi gerçek değerlerdir.

(1) denklemi t_1' ile çarpılırsa $t_1' R t_1 - \lambda t_1' t_1 = 0$ ve $t_1' t_1 = 1$ olduğundan;

$t_1' R t_1 = \lambda_1$ Ve $Var(y_1) = \lambda_1$ olarak bulunmaktadır.

$y_{1i} = t_1' Z$ İle dönüştürülmüş olan vektöre birinci temel bileşen adı verilmektedir. Eğer t_1 vektörü yerine $\sqrt{\lambda_1} t_1$ vektörü kullanılır ise varyans ağırlıklarının katsayılarına etki etmesi sağlanarak birinci asıl temel bileşen bulunmaktadır.

İkinci temel bileşen y_2 bulunurken;

1. Kısıt: y_2 vektörü varyansı y_1 'den sonra en büyük olmalı.
2. Kısıt: t_2 vektörü birim normal vektör olsun

$$t_2' t_2 = 1 \quad (1.13)$$

3. Kısıt: y_2 ve y_1 vektörleri birbirine dik olsun.

$$t_2' t_1 = 0 \quad (1.14)$$

$$y_2 = t_2' Z \quad (1.15)$$

$$Var(y_2) = \frac{1}{n-1} \sum_{i=1}^n y_{1i} y_{2i} \quad (1.16)$$

$$= \frac{1}{n-1} \sum_{i=1}^n (t_1' z_i)(t_2' z_i)'$$

$$= \frac{1}{n-1} \sum_{i=1}^n t_1' z_i z_i' t_2$$

$$= t_1' R t_2$$

$$= R t_i = t_i \lambda_i$$

$$t_1' R t_2 = \lambda_1 t_1' t_1 = 0 \quad (1.17)$$

$$Var(y_2) = E(t_2' Z)^2 \quad (1.18)$$

$$= \frac{1}{n-1} \sum_{i=1}^n (t_2' z_i)(t_2' z_i)' = t_2' R t_2' \text{dir}$$

$$\varphi_2 = t_2' R t_2 - \lambda_2(t_2' t_2 - 1) - \gamma_1(t_2' t_1 - 0) \quad (1.19)$$

$$\frac{\partial \varphi_1}{\partial t_1} = 2R t_2 - 2\lambda_2 t_2 - \gamma_1 t_1 = 0 \quad (1.20)$$

$$2(R - \lambda_2)t_2 - \gamma_1 t_1 = 0 \quad (1.21)$$

$$\gamma_1 t_1 = 0 \quad (1.22)$$

(yani $\gamma_1 = 0$) olmalıdır ki

$$(R - \lambda_2)t_2 = 0 \text{ olsun.} \quad (1.23)$$

$$|R - \lambda_2| = 0 \quad (1.24)$$

Bu sonuca göre λ_2 , R matrisinin ikinci özdeğeri ve t_2 , R matrisinin ikinci özvektörüdür.

j=1,2,3,...,p değerleri için tüm λ_j , t_j , γ_j değerleri elde edilir.

$$\sum_{j=1}^p \lambda_{j=p+1} z R \quad (1.25)$$

$$\prod_{j=1}^p \lambda_j = |R| \quad (1.26)$$

1.3.2. Bağımsız Bileşenler Analizi (Independent Component Analysis [ICA])

ICA, çok değişkenli veri setlerindeki gizli bağımsız kaynakları ortaya çıkarmak için kullanılan güçlü bir istatistiksel tekniktir. Özellikle, biyomedikal ve medikal veri setlerinde önemli uygulamalara sahiptir. ICA, verideki karmaşıklığı azaltarak daha anlamlı ve yorumlanabilir bileşenler elde edilmesine olanak tanır. Bu

teknik, verideki bileşenlerin istatistiksel olarak bağımsız olduğu varsayımına dayanır ve bu bileşenleri ortaya çıkarır. ICA, özellikle veri setinde karışık sinyallerin ayrıştırılmasında ve analiz edilmesinde etkilidir.

ICA uygulaması altı ana adımdan oluşmaktadır. Bunlar aşağıdaki gibi sıralanabilir;

- Veri setlerinin hazırlanması,
- Verileri ortalamadan çıkarma,
- Karışık bileşenlerin bulunması,
- Bağımsız bileşenlerin bulunması,
- Bileşenlerin sıralanması ve yorumlanması,
- Değerlendirme ve yorumlama

ICA, PCA ile benzerlik göstermesine rağmen aralarında farklılıklar yer almaktadır. Bu farklar aşağıda yer almaktadır.

- PCA, veri setindeki değişkenler arasındaki korelasyonu en aza indirmeye odaklanırken, ICA bağımsız bileşenleri (kaynakları) bulmaya çalışır.
- PCA'da, yeni değişkenler genellikle birbirine bağımlıdır ve veri setindeki genel varyansı maksimize etmek için en iyi bileşenler bulunur.
- ICA ise, bağımsız bileşenler oluşturarak değişkenler arasında bağımsızlık sağlar. Her bir bileşen, diğerlerinden mümkün olduğunca bağımsızdır ve veri setindeki gizli kaynakları temsil eder.
- PCA, özellik seçimi, boyut azaltma ve veri sıkıştırma gibi geniş bir uygulama alanına sahipken, ICA genellikle EEG, MEG sinyallerinin analizi, fMRI verilerinin işlenmesi gibi alanlarda kullanılır.

1.3.3. Uniform Manifold Approximation and Projection (UMAP)

UMAP, yüksek boyutlu veri setlerini daha düşük boyutlu uzaylara indirmek için kullanılan bir boyut indirgeme tekniğidir. Bu teknik, özellikle veri görselleştirme, kümelenme ve özellik mühendisliği gibi alanlarda yaygın olarak kullanılır. UMAP,

verilerin manifold yapısını öğrenerek, bu yapıyı koruyan düşük boyutlu bir temsil oluşturmayı hedefler. Bu süreç, verinin hem yerel komşuluk yapılarını hem de genel topolojik özelliklerini anlamayı ve korumayı gerektirir (McInnes, 2018; Becht 2019).

UMAP'ın temel olarak iki aşaması vardır. Yerel komşuluk yapısının öğrenilmesi ve düşük boyutlu uzaya projeksiyon edilmesi. İlk olarak, UMAP yüksek boyutlu uzaydaki her veri noktası için yerel komşuluk yapısını belirler. Bu aşamada, her veri noktasının en yakın komşuları tespit edilir ve bu komşuluk ilişkileri, verinin manifold yapısını koruyacak şekilde hesaplanır (McInnes, 2020). Yüksek boyutlu uzaydaki veriler arasındaki benzerlikler hesaplanır ve bu benzerlikler, daha sonra düşük boyutlu bir uzaya projekte edilir. İkinci aşamada ise yerel komşuluk yapısının korunması hedeflenir. Bu işlem, verilerin düşük boyutlu bir temsiline ulaşmak için bir optimizasyon probleminin çözülmesiyle gerçekleştirilir. UMAP, bu süreçte verinin yerel yapısını koruyarak, kümeleri anlamlı bir şekilde temsil eder. UMAP'ın parametrik esnekliği, kullanıcılara projeksiyonun hem yerel hem de global yapılarını kontrol etme imkanı tanır (McInnes, 2020). Yerel komşuluk sayısı ve minimum mesafe gibi parametreleri ayarlayarak, veri setine uygun projeksiyonlar elde edilebilir.

UMAP, özellikle büyük veri setlerinin görselleştirilmesi ve analiz edilmesi için güçlü bir araçtır. Biyoinformatikte, özellikle tek hücreli RNA dizileme (scRNA-seq) verilerinin analizi için yaygın olarak kullanılmaktadır. Ayrıca, görüntü işleme, doğal dil işleme ve genetik veri analizi gibi çeşitli alanlarda da etkili bir yöntemdir.

1.3.4. t- Distributed Stochastic Neighbor Embedding (t-SNE)

t-SNE, yüksek boyutlu verileri daha düşük boyutlu bir uzaya indirgemek için kullanılan bir tekniktir. Özellikle veri görselleştirme amacıyla tercih edilen t-SNE, veriler arasındaki karmaşık ilişkileri ve kümelenme yapılarını iki veya üç boyutlu alanlarda görselleştirerek daha anlaşılır hale getirir. 2008 yılında Laurens van der Maaten ve Geoffrey Hinton tarafından geliştirilen bu yöntem, makine öğrenimi ve veri bilimi alanlarında yaygın bir kullanım alanına sahiptir (van der Maaten, 2008).

t-SNE'nin çalışma prensibi, veriler arasındaki benzerlikleri öncelikle yüksek boyutlu uzayda, sonrasında da düşük boyutlu uzayda modellendirmektir. İlk olarak,

yüksek boyutlu uzaydaki veri noktaları arasındaki benzerlikler, her bir veri noktasının diğer veri noktalarına olan mesafeleri hesaplanarak belirlenir. Bu aşamada, veri noktaları arasındaki mesafeleri modellemek için Gaussian (normal) dağılımı kullanılır. t-SNE ikinci olarak ise, yüksek boyutlu uzaydaki benzerlikleri koruyarak veri noktalarını düşük boyutlu bir uzaya projekte eder. Bu işlem, veri noktaları arasındaki uzaklıkların düşük boyutlu uzayda da korunacak şekilde optimize edilmesiyle gerçekleştirilir. Bu aşamada, Kullback-Leibler divergansı adı verilen bir ölçüt kullanılarak, yüksek ve düşük boyutlu uzaylar arasındaki olasılık dağılımları arasındaki fark minimize edilir (Linderman, 2019). Son olarak, düşük boyutlu uzayda veri noktaları arasındaki mesafelerin hesaplanmasında, t-dağılımı kullanılır. Bu dağılım, özellikle t-SNE'nin küçük veri noktaları arasındaki mesafeleri doğru bir şekilde temsil etmesine yardımcı olur ve kümelerin daha net bir şekilde ayrışmasını sağlar (van der Maaten ve Hinton, 2008).

t-SNE, veri görselleştirmede güçlü bir araç olmasına rağmen, bazı sınırlamalara da sahiptir. Büyük veri setleri üzerinde çalışırken oldukça zaman alıcıdır ve hesaplama maliyeti yüksektir (Linderman, 2019). Ayrıca, t-SNE projeksiyonlarının tekrarlanabilirliği düşüktür; farklı çalıştırmalarda aynı veri seti için farklı sonuçlar elde edilebilir. Bu nedenle, elde edilen sonuçların dikkatli bir şekilde yorumlanması gerekmektedir.

t-SNE, birçok alanda yaygın olarak kullanılmaktadır. Görüntü işleme alanında yüksek boyutlu görüntü verilerinin düşük boyutlu uzaya indirgenmesi, görüntü sınıflandırma ve nesne tanıma gibi görevlerde t-SNE'nin kullanılmaktadır. (van der Maaten, 2014). Doğal dil işleme alanında ise kelime gömme (ing:word embedding) yöntemleriyle oluşturulan yüksek boyutlu vektörlerin düşük boyutlu uzaya indirgenmesi, kelime ilişkilerini ve anlamlarını görselleştirmede kullanılmaktadır.

1.4. META-SEZGİSEL ALGORİTMALAR

Meta-sezgisel algoritmalar, karmaşık optimizasyon problemlerini çözmek için geliştirilen esnek ve güçlü yaklaşımlardır. "Sezgisel" terimi, bir problemi çözmek için kullanılan basit ve hızlı yöntemleri ifade ederken, "meta-sezgisel" bu basit yöntemlerin bir araya getirilmesiyle oluşturulan daha genel ve kapsamlı algoritmaları tanımlar. Bu

algoritmalar genellikle arama stratejileri ve çözüm iyileştirme mekanizmaları olmak üzere iki temel bileşene dayanır. Meta-sezgisel algoritmaların birçok çeşidi vardır ve her biri farklı problem türleri için optimize edilmiştir. Bu çalışmada en optimum sonuçları elde edebilmek için meta-sezgisel yöntemlerden Parçacık Sürü Optimizasyonu (ing: Particle Swarm Optimization [PSO]) ve Kovaryans Matrisi Uyarlama Evrim Stratejisi (ing: Covariance Matrix Adaptation Evolution Strategy [CMA-ES]) yöntemleri kullanılmıştır.

1.4.1. Covariance Matrix Adaptation Evolution Strategy CMA-ES)

CMA-ES, meta-sezgisel optimizasyon yöntemidir. Özellikle, büyük boyutlu, karmaşık ve çok modlu problemleri çözmekte oldukça etkili bir yaklaşımdır. Bu yöntem, sürekli ve gerçek değerli parametre optimizasyonu için kullanılır ve çoğunlukla geleneksel gradyan tabanlı optimizasyon tekniklerinin yetersiz kaldığı durumlarda tercih edilir.

CMA-ES, evrimsel stratejiler olarak adlandırılan stokastik bir arama yöntemi üzerine kuruludur. Evrimsel stratejiler, biyolojik evrimi taklit eder ve çözüm uzayında daha iyi çözümler bulmayı amaçlar. CMA-ES, bu stratejilerden biri olarak, özellikle kovaryans matrisinin adaptasyonu sağlar. Kovaryans matrisi, popülasyonun dağılımını tanımlayan bir matristir ve bu matrisin adaptasyonu, algoritmanın çözüm uzayında daha etkili arama yapmasını sağlamaktadır (Hansen, 2001). Algoritma 1.1'de CMA-ES algoritmasına ait sözde kod verilmiştir.

Algoritma 1.1 CMA-ES algoritması sözde kodu

1.Adım: Başlangıç parametrelerinin atanması

- $m \leftarrow$ ortalama (vektör)
- $C \leftarrow$ Kovaryans matris (matris)
- $\sigma \leftarrow$ adım sayısı (skaler)

- $\lambda \leftarrow$ popülasyon büyüklüğü

- $\mu \leftarrow$ seçilen bireyler sayısı

- max_iter $\leftarrow$ maksimum iterasyon sayısı

2.Adım: Her iterasyon için 1'den max_iter'e kadar

a: Yeni popülasyon oluştur

for i = 1 to λ :

- $x_i \leftarrow m + \sigma * \text{sqrt}(C) * z_i$

- z_i : normal dağılımdan örneklenmiş bir vektör

b: Her birey için fitness fonksiyonunu değerlendir.

$f(x_i) \leftarrow \text{evaluate_fitness}(x_i)$

c: En iyi bireyleri seç

- fitness değerlerine göre sırala

- elit grup olarak en iyi μ bireylerini seç.

d: Ortalamayı güncelle

- $m \leftarrow$ en üst μ bireylerin ağırlıklı ortalaması

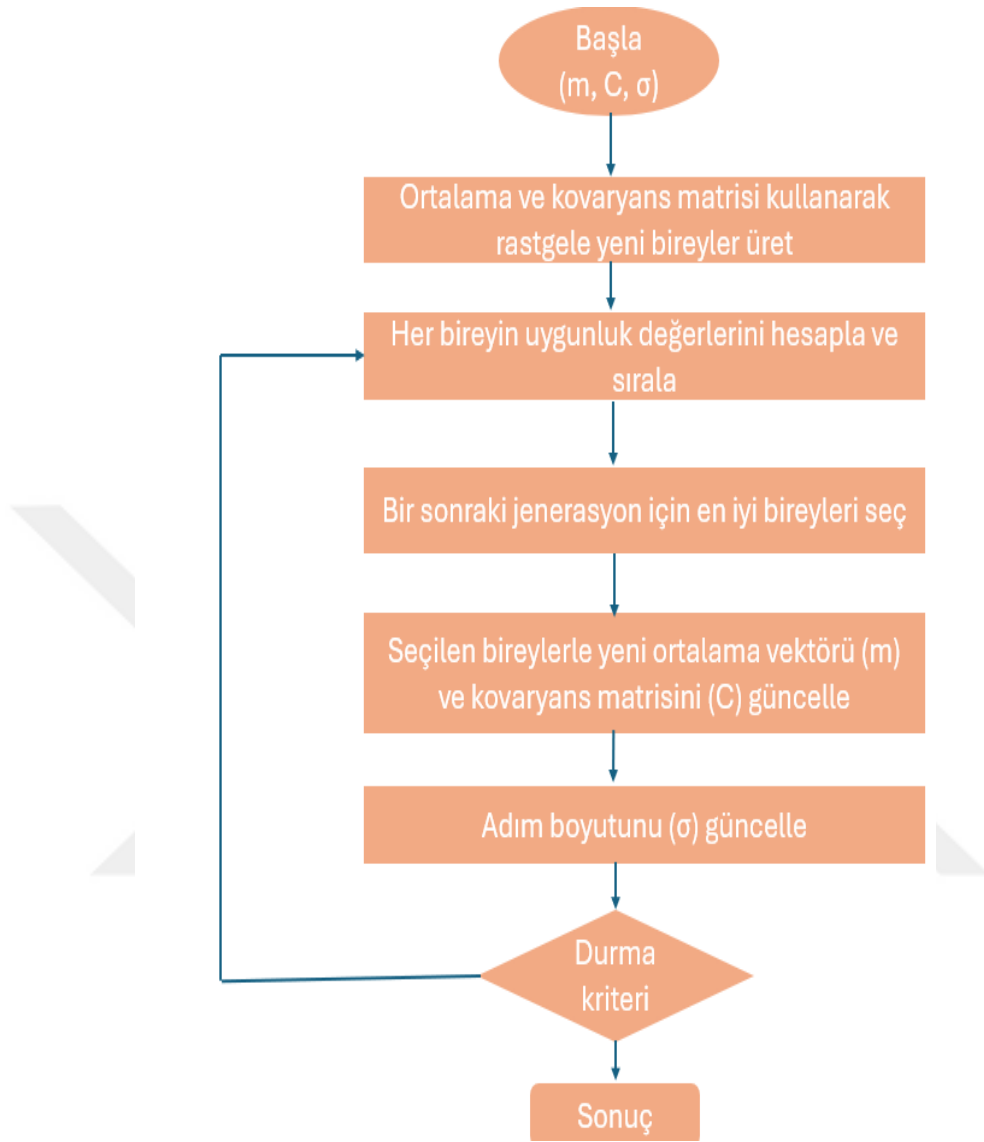
e: Kovaryans matrisini güncelle

- En üst μ bireyler ile yeni ortalama arasındaki farkları kullanarak ağırlıklı kovaryans matrisini hesapla.

f: Adım sayısını güncelle

- Uygunluğu iyileştirmedeki ilerlemeye bağlı olarak σ 'yı ayarla.

3.Adım: Bulunan en iyi değeri geri döndür.



Şekil 1. 4 CMA-ES algoritmasının akış diyagramı

CMA-ES, birçok mühendislik ve bilimsel alanda yaygın olarak kullanılmaktadır. Özellikle, geleneksel yöntemlerin zorlandığı yüksek boyutlu ve çok modlu optimizasyon problemlerinde etkili bir alternatif olarak öne çıkmaktadır. Yapay zeka, makine öğrenmesi, robotik ve kontrol sistemleri gibi alanlarda da başarıyla uygulanmaktadır.

1.4.2. Particle Swarm Optimization (PSO)

PSO, ilk olarak Kennedy ve Eberhart tarafından 1995 yılında geliştirilen bir optimizasyon algoritmasıdır. PSO, doğal sistemlerdeki kuş sürüsü veya balık sürülerinin hareketlerinden esinlenerek tasarlanmıştır (Kennedy,1995).

PSO, çözüm uzayında bir dizi aday çözümü "parçacıklar" olarak modelleyerek çalışan bir optimizasyon yöntemidir. Her parçacık, çözüm uzayında belirli bir pozisyona sahiptir ve bu pozisyon zaman içinde güncellenir. Parçacıkların hız ve pozisyon güncellemeleri, her bir parçacığın kendi en iyi bulgularını ve tüm parçacıkların en iyi bulgularını dikkate alarak yapılır. Bu yöntem, parçacıkların daha iyi çözümler bulmasını teşvik eder (Shi, 1998).

PSO'nun hız ve pozisyon güncellemeleri matematiksel olarak denklem (1.27) ve (1.28)'deki gibi tanımlanmaktadır:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_{best,i} - x_i(t)) + c_2 \cdot r_2 \cdot (g_{best,i} - x_i(t)) \quad (1.27)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (1.28)$$

Denklemlerdeki semboller aşağıda maddelenmiştir.

$v_i(t)$:Parçacığın mevcut durumundaki hız vektörü.

w :Parçacığın kendi yönündeki hareket katsayısı.

$x_i(t)$:Parçacığın mevcut durumundaki konum vektörü.

$p_{best,i}(t)$:Parçacığın geçmiş bilgilerinden elde ettiği en iyi konum.

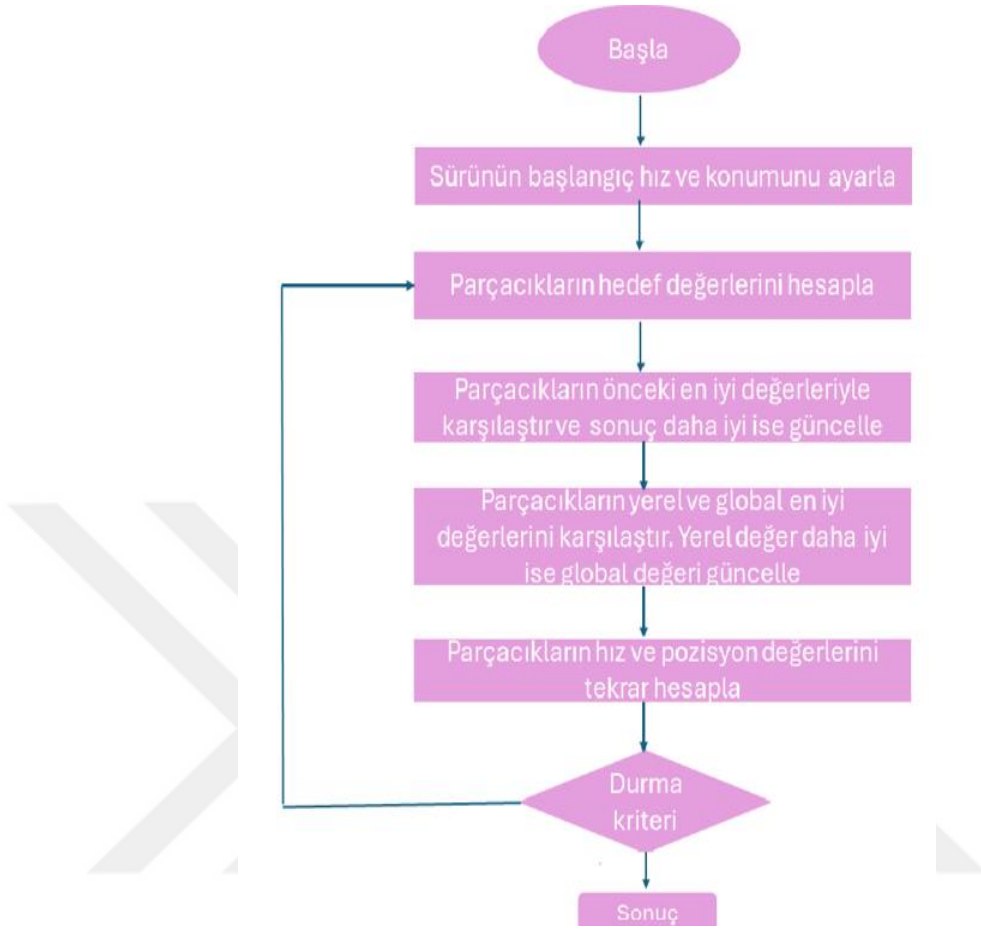
$g_{best,i}(t)$:Tüm parçacıklar kapsamında elde edilen en iyi konum.

c_1 :Parçacığın kendi geçmişinden gelen öğrenme katsayısı.

c_2 :Parçacığın toplumun geçmişinden gelen öğrenme katsayısı.

r_1 ve r_2 : $(0,1)$ aralığında atanan rastgele değerler olarak tanımlanır.

Şekil 1.7'de PSO algoritmasına ait akış diyagramı yer almaktadır.



Şekil 1. 5 PSO akış diyagramı

PSO basitliği ve esnekliği açısından geniş bir uygulama yelpazesine sahiptir. Bunların yanında PSO hem sürekli hem de kesikli optimizasyon problemlerinde etkili bir şekilde çalışabilir. Optimizasyon problemlerinde PSO'nun kullanımı, birçok mühendislik ve bilimsel alanda yaygındır. Örneğin, yapay sinir ağlarının hiper-parametre ayarlarında, kontrol sistemleri tasarımında ve makine öğrenme modellerinin performansını artırmada sıkça tercih edilir.

1.5. DERİN ÖĞRENME VE AĞAÇ TEMELLİ ALGORİTMALAR İLE İLGİLİ LİTERATÜR ÇALIŞMASI

Kan hücrelerinin analizi, birçok hastalığın teşhisinde kilit bir rol oynamaktadır. Kan hücrelerinin sınıflandırılmasında derin öğrenme yöntemleri, bu yöntemlerin birleştirilmesi veya melez yaklaşımlar son yıllarda sıklıkla kullanılmaktadır. Bu yaklaşımların performansı arttırdığı gözlemlenmektedir.

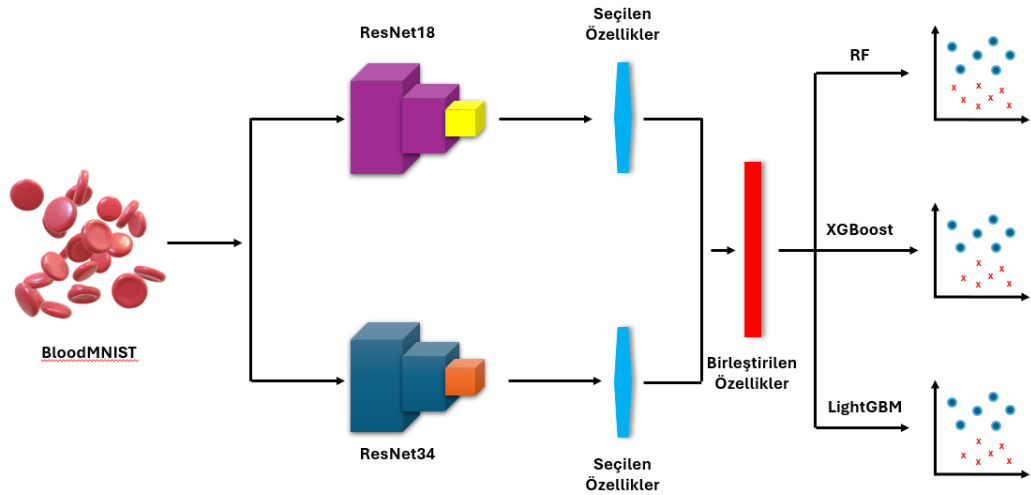
Kan hücrelerinin sınıflandırılması ile ilgili literatürde yapılmış başarılı derin öğrenme ve melez yaklaşımlar yer almaktadır (Cengil, 2021). Çalışmasında, 4 sınıflı kan hücresi veri setinde ResNet18, GoogleNet ve AlexNet modelleri birleştirilmiş buradan elde edilen vektör destek vektör makineleri (ing: Support vector machine, SVM) algoritması ile sınıflandırılmıştır. Bu melez yaklaşım ile yüzde %99,83 doğruluk değerine ulaşılmıştır. Çalışmanın ikinci aşamasında ise özneliklerin birleştirme aşamasından sonra boyut azaltma yöntemi olarak PCA algoritması uygulanmıştır. (Toğaçar,2020) çalışmasında ise yine benzer bir yaklaşım söz konusudur. Bu çalışma, MRI görüntülerinde, AlexNet, VGG-16 ve VGG-19 modellerini birleştirerek SVM algoritması ile sınıflandırma yapmaktadır. Bu çalışma %99,41 doğruluk değeri ile doğru sınıflandırma yapmaktadır. (Sengür,2019) çalışmasında ise, beyaz kan hücrelerini sınıflandırmak için önerilen üç katmanlı bir yapı bulunmaktadır. Bu hibrit yapı, ilk katmanlarda görüntü işleme ve iyileştirme algoritmalarını kullanarak bölütleyici ve öznelik çıkarımı işlemleri gerçekleştirirken, son katmanda Uzun-Kısa Vadeli Bellek (LSTM) ağını kullanarak sınıflandırma yapmaktadır. Önerilen model, 349 kan yayma görüntüsü veri kümesi üzerinde şekil tabanlı özelliklerde %80.0 doğruluk elde ederken, ince ayrıntılar üzerinde ise %82.9 doğruluk sağlamıştır. (Malkawi,2020) çalışmasında ise beyaz kan hücrelerinin sınıflandırılması için öznelik çıkarıcı olarak evrişimli sinir ağları (ing: convolutional neural networks, CNN) kullanılırken sınıflandırma için SVM, K-en yakın komşu (KNN) ve RF algoritmaları kullanılmıştır. En yüksek doğruluk değeri RF algoritması ile %98.7 olarak bulunmuştur. (Fathy, 2023) çalışmasında beyaz kan hücrelerini sınıflandırmak için melez bir yapı önerilmiştir. ResNet101 ile öznelik çıkarımı, BBCO algoritması ile öznelik azaltma ve SVM algoritması ile sınıflandırma yapmaktadır. ResNet101-BBCO-SVM hibrit yapı ile beyaz kan hücreleri %99.21 ile doğru sınıflandırma yapmaktadır.

İKİNCİ BÖLÜM

2. ÖNERİLEN YÖNTEMLER

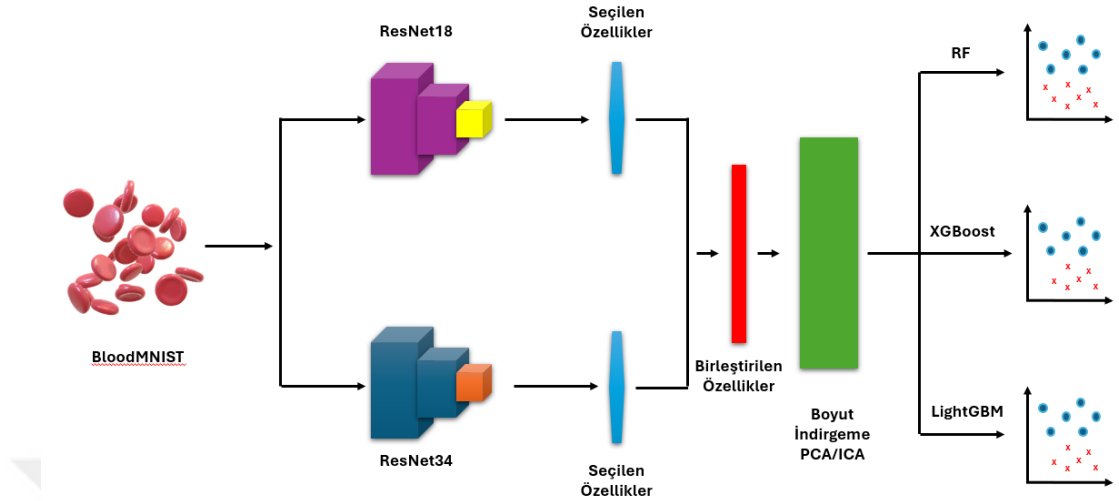
2.1. DERİN ÖĞRENME-AĞAÇ TEMELLİ ALGORİTMA MELEZ ÇERÇEVESİ

Melez derin öğrenme-ağaç tabanlı algoritma çerçevesi kan hücrelerinin sınıflandırılması için tasarlanmıştır. Bu model çalışmanın ilk melez çerçevesi 2 aşamadan oluşmaktadır. Derin öğrenme algoritmaları ile öznitelik çıkarma ve ağaç tabanlı algoritmalar ile sınıflandırma. Çalışmaya ait akış diyagramı Şekil 2.1’de yer almaktadır.



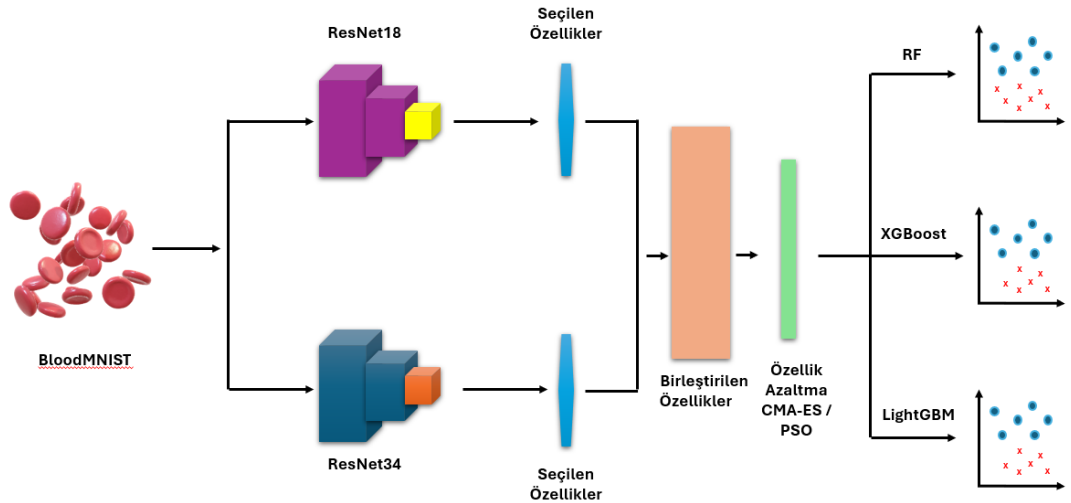
Şekil 2. 1 Birinci melez çerçevenin akış diyagramı

Çalışmanın ikinci melez çerçevesinde, elde edilen ResNet- ağaç tabanlı algoritma melez çerçevesinin ara katmanına karmaşıklığı azaltmak amacıyla boyut indirgeme algoritmalarından PCA, ICA, UMAP ve t-SNE algoritmaları uygulanarak sınıflandırma performansı üzerindeki etkisi incelenmiştir. İkinci melez çerçeveye ait akış diyagramı Şekil 2.2’de yer almaktadır.



Şekil 2. 2 İkinci melez çerçeve için akış diyagramı

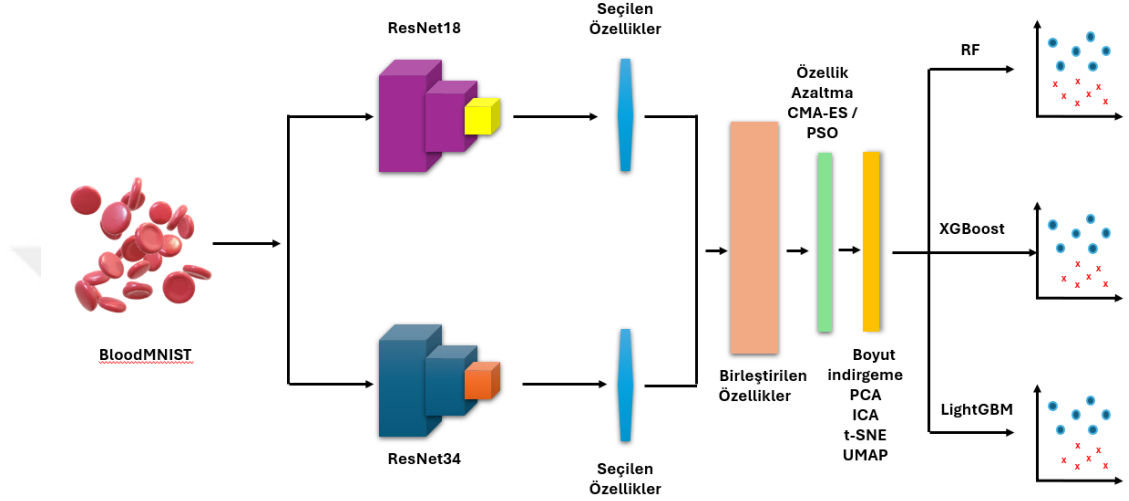
Çalışmanın üçüncü melez çerçevesinde elde edilen ResNet- ağaç tabanlı algoritma melez çerçevesinin ara katmanına karmaşıklığı azaltmak ve aşırı öğrenmeyi önleyerek daha seçkin öznitelikleri seçebilmek adına meta-sezgisel algoritmalarından CMA-ES ve PSO algoritmaları kullanılarak sınıflandırma performansına olan etkisi incelenmiştir. Üçüncü melez çerçeveye ait akış diyagramı Şekil 2.3'te yer almaktadır.



Şekil 2. 3 Üçüncü melez çerçeve için akış diyagramı

Çalışmanın dördüncü melez çerçevesinde ise üçüncü bölümdeki meta-sezgisel algoritmalarla özniteliklerin azaltılmasından sonraki katmana boyut azaltma

yöntemlerinden PCA, ICA, t-SNE ve UMAP uygulanarak sınıflandırma performansına olan etkisi incelenmiştir. Dördüncü melez çerçeveye ait akış diyagramı Şekil 2.4'te yer almaktadır.



Şekil 2. 4 Dördüncü melez çerçeve için akış diyagramı

ÜÇÜNCÜ BÖLÜM

3. DENEYSEL ÇALIŞMALAR

Bu bölüm altında tez çalışmasında kullanılan derin öğrenme modelleri ve ağaç tabanlı algoritmaların melez çerçevesi gerçekleştirirken, öğrenme ve test aşamalarında kullanılmasına karar verilen deneysel tasarım kararları ve veri setleri hakkında genel bilgilere yer verilmiştir.

3.1. DENEYSEL TASARIM

Çalışmada kullanılan yaklaşımlar, Python 3.10 versiyonu programlama dili kullanılarak geliştirilmiştir. Veri setinin okunması için Numpy, Pandas, Math, görselleştirme için Matplotlib, MÖ algoritmaları ve hiper-parametre optimizasyonu için scikit-learn, derin öğrenme mimarilerinin geliştirilmesi için Pytorch kütüphanelerinden yararlanılmıştır.

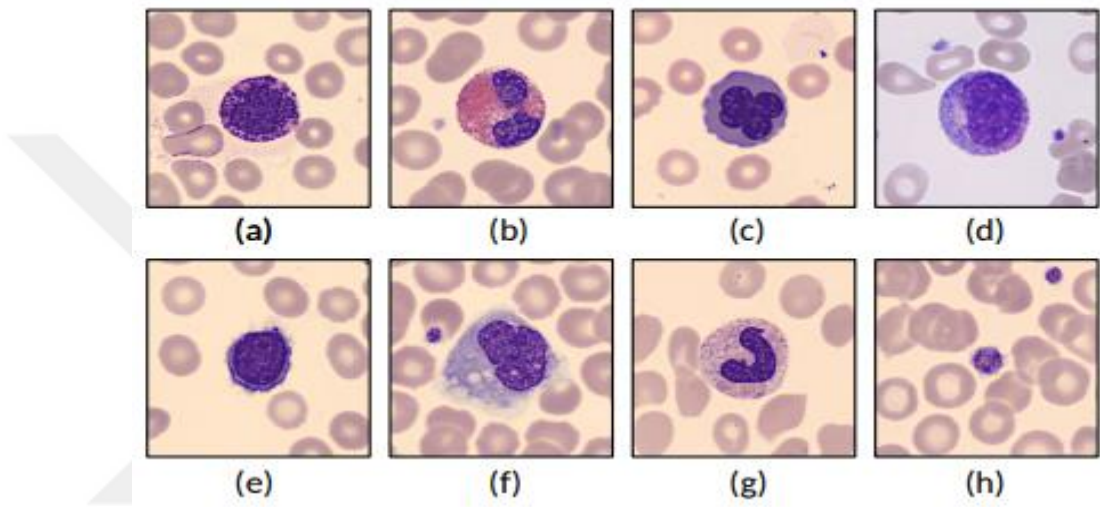
Ayrıca daha özelleşmiş tasarım kararları ve seçilen parametre değerleri “Bulgular ve Tartışmalar” bölümü altında ilgili altbaşlıklarda açıklanacaktır.

3.2. VERİ SETİ

Bu çalışmada, MedMNIST’ te bulunan BloodMNIST veri kümesi kullanılmıştır. BloodMNIST, tıbbi görüntü analizinde kullanılan bir veri kümesidir ve beyaz kan hücrelerinin sınıflandırılması gibi hematolojik araştırmalarda önemli bir rol oynar. Bu veri kümesi, çeşitli kan hücrelerinin görüntülerini içerir ve derin öğrenme modellerinin eğitiminde yaygın olarak kullanılır. BloodMNIST, Medical MNIST koleksiyonunun bir parçasıdır ve diğer tıbbi görüntüleme veri kümeleriyle birlikte sağlık ve biyomedikal alanında derin öğrenme uygulamaları için değerli bir kaynaktır.

BloodMNIST, toplamda 17,092 adet kan hücresi görüntüsü içermektedir. Bu görüntüler, farklı kan hücresi türlerini temsil eden sekiz sınıfa ayrılmıştır: Bazofil, Eozinofil, Eritrosit, Lenfosit, Monosit, Nötrofil, Normoblast ve Trombosit. Her bir görüntü 28x28 piksel boyutundadır. Bu boyut, veri kümesinin derin öğrenme modellerine uygun bir şekilde işlenmesini kolaylaştırır. Ancak, görüntülerin 28x28

piksel boyutunda olması, bazı ayrıntıların kaybolmasına neden olabilir. Bu nedenden dolayı, özellikle ince detayların ayırt edilmesi kaçırılabilir. BloodMNIST, eğitim, doğrulama ve test olmak üzere üç ayrı alt kümeye bölünmüştür. Bu bölünme 7,1,2 oranında bölünmüştür. Eğitim kümesi 11,959 görüntü, doğrulama kümesi 1,321 görüntü ve test kümesi 3,812 görüntü içerir. Bu yapı, modellerin eğitim ve değerlendirilmesi için dengeli bir veri dağılımı sağlar.



Şekil 3. 1 BloodMNIST veri kümesindeki örnek kan hücresi görüntüleri: (a) bazofil, (b) eozinofil, (c) eritrosit, (d) normoblast, (e) lenfosit, (f) monosit, (g) nötrofil, (h) trombosit

3.3. DEĞERLENDİRME METRİKLERİ

Tez kapsamında sınıflandırma algoritmalarının başarımını ölçmek amacıyla doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skor (F1 score) performans metrikleri kullanılmıştır.

3.4. HİPER-PARAMETRE DEĞERLERİ

Sınıflandırma ve meta-sezgisel algoritmalarında hiper-parametrelerin optimize edilmesi performansı oldukça etkilemektedir. Bu çalışmada hiper-parametre optimizasyonu Grid search ve Random search olmak üzere iki farklı şekilde yapılmıştır. Sınıflandırma algoritmalarından, RF algoritması için optimize edilen hiper-parametreler *n_estimators*, *max_features*, *max_depth*, *criterion* ve

min_sample_leaf olarak seçilmiştir. XGBoost algoritması için seçilen parametreler, *n_estimators*, *learning_rate*, *objective*, *max_depth*, *subsample*, *feature_fraction* ve *min_child_samples* olarak seçilmiştir. LightGBM algoritması için ise *n_estimators*, *objective*, *max_depth*, *feature_fraction*, *min_child_samples* ve *num_leaves* olarak seçilmiştir. Meta-sezgisel algoritmalarından PSO algoritması için *swarm_size*, *max_iter*, ve *inertia_weight* parametreleri seçilmiştir. CMA-ES algoritması için ise *pop_size*, *step_size* (*sigma*), ve *max_iter* parametreleri seçilmiştir. Parametrelerin değer aralıkları Tablo 3.1’de yer almaktadır. Bulunan en iyi değerler Tablo 3.2’de yer almaktadır.

Tablo 3. 1: Algoritmalar için bulunan hiper-parametre aralıkları

Algoritma	Parametre	Bulunan Değer
Rastgele Orman	<i>n_estimators</i>	[50, 75, 100, 150, 200]
	<i>max_features</i>	[auto, sqrt]
	<i>max_depth</i>	[5, 6, 7, 8, 9, 10, 11]
	<i>criterion</i>	[gini, entropy]
	<i>min_sample_leaf</i>	[5, 6, 7, 8, 9, 10, 11]
XGBoost	<i>n_estimators</i>	[50, 75, 100, 150, 200]
	<i>learning_rate</i>	[0.01, 0.05, 0.1, 0.2, 0.3]
	<i>objective</i>	binary: logistic
	<i>max_depth</i>	[5, 6, 7, 8, 9, 10, 11]
	<i>subsample</i>	[0.5, 0.6, 0.8]
	<i>feature_fraction</i>	[0.5, 0.6, 0.7, 0.8, 0.9]
	<i>min_child_samples</i>	[5, 6, 7, 8, 9, 10]

LightGBM	<i>n_estimators</i>	[50, 75, 100, 150, 200]
	<i>objective</i>	binary: logistic
	<i>max_depth</i>	[5, 6, 7, 8, 9, 10, 11]
	<i>feature_fraction</i>	[0.5, 0.6, 0.7, 0.8, 0.9]
	<i>min_child_samples</i>	[5, 6, 7, 8, 9, 10]
	<i>num_leaves</i>	[35, 40, 50, 60]
CMA-ES	<i>pop_size</i>	
	<i>max_iter</i>	[20, 35, 45, 55]
	<i>sigma</i>	[30, 50, 75, 100]
PSO		[0.3, 0.5, 0.7, 0.9]
	<i>swarm_size</i>	[20, 35, 45, 55]
	<i>max_iter</i>	[50, 65, 75, 100]
	<i>weight</i>	[0.4, 0.5, 0.6, 0.7, 0.8]

Tablo 3. 2: Algoritmalar için bulunan hiper-parametre değerleri

Algoritma	Parametre	Bulunan Değer
Rastgele Orman	<i>n_estimators</i>	100
	<i>max_features</i>	Auto
	<i>max_depth</i>	6
	<i>criterion</i>	Gini
	<i>min_sample_leaf</i>	6

XGBoost	<i>n_estimators</i>	200
	<i>learning_rate</i>	0.1
	<i>objective</i>	binary: logistic
	<i>max_depth</i>	8
	<i>subsample</i>	0.8
	<i>feature_fraction</i>	0.6
LightGBM	<i>min_child_samples</i>	8
	<i>n_estimators</i>	200
	<i>objective</i>	binary: logistic
	<i>max_depth</i>	6
	<i>feature_fraction</i>	0.8
	<i>min_child_samples</i>	11
CMA-ES	<i>num_leaves</i>	50
	<i>pop_size</i>	45
	<i>max_iter</i>	75
PSO	<i>sigma</i>	0.5
	<i>swarm_size</i>	45
	<i>max_iter</i>	65
weight	<i>weight</i>	0.6

DÖRDÜNCÜ BÖLÜM

4. BULGULAR VE TARTIŞMALAR

Bu bölümde, tez çalışmasında bir önceki bölümde anlatılmış olan deneysel tasarımlar baz alınarak veri setleri üzerinde yapılmış olan deneysel bulgular ve ilgili değerlendirmeler yer almaktadır.

Tez kapsamında tasarlanan deneysel çalışmalar dört melez yöntemden oluşmaktadır. İlk melez yöntemde, ResNet18 ve ResNet34 modellerinden ön-eğitim edilmiş ağırlıklar ile gelen öznitelikler birleştirilerek ağaç tabanlı algoritmalar ile sınıflandırma yapılmıştır. İkinci melez yöntemde, modellerin birleştirilmesinden sonraki aşamada ise performansı arttırmak amacıyla öznitelik seçme algoritmalarından PCA, ICA, t-SNE ve UMAP algortimaları uygulanmıştır. Üçüncü emelez yöntemde ResNet modellerinin birleştirilmesinden sonra daha seçkin özniteliklerin seçilebilmesi için sonraki katmana meta-sezgisel algoritmalarından CMA-ES ve PSO algoritmaları uygulanmıştır. Dördüncü melez yöntemde ise, üçüncü bölümde yapılan öznitelik azaltma algoritmaları aşamasından sonra boyut indirgeme algoritmalarından PCA, ICA, t-SNE ve UMAP uygulanarak performanslar değerlendirilmiştir. Ayrıca meta-sezgisel ve ağaç tabanlı algoritmaların başarımı hiper-parametre başarımına bağlı olduğu için her bölümde bütün algoritmalar için grid search yöntemi ile hiper-parametre optimizasyonu da yapılmıştır. Her sınıflandırma algoritması için çapraz

doğrulama yapılmış ve 10 fold'a bölünmüştür. Her algoritma 10 kez çalıştırılıp bulunan değerlerin ortalaması alınmış ve standart sapmaları hesaplanmıştır. Aşağıda çalışmaya ait sonuçlar yer almaktadır.

4.1. BİRİNCİ MELEZ ÇERÇEVE İÇİN SONUÇLAR

Çalışmadaki ilk melez çerçeve, ResNet18 ve ResNet34 modelleri birleştirilmiş, ağaç tabanlı algoritmalar ile sınıflandırılmıştır.

Tablo 4. 1: ResNet-Ağaç tabanlı sınıflandırıcı melez çerçevesinin varsayılan değerleri ile elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%73,21	%73,40	%70,42	%71,42
XGBoost	%74,69	%74,25	%71,05	%72,10
LightGBM	%76,24	%76,25	%73,05	%75,64

Tablo 4. 2: ResNet-Ağaç tabanlı algoritma melez çerçevesinde ağaç tabanlı algoritmaların hiper-parametre optimizasyonu sonucunda bulunan en iyi değerler ile elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%74,47	%74,00	%70,42	%72,58
XGBoost	%75,98	%75,91	%71,20	%72,48
LightGBM	%77,12	%77,30	%74,96	%74,48

Tablo 4.1 ve Tablo 4.2 sonuçları incelendiğinde, en iyi hiper-parametre seçiminin başarıyı arttırdığı gözlemlenmiştir.

4.2. İKİNCİ MELEZ ÇERÇEVE İÇİN SONUÇLAR

Çalışmanın ikinci melez çerçevesinde, ResNet18 ve ResNet34 modelleri ön-eğitim edilmiş ağırlıklar ile birleştirildikten sonraki katmana boyut indirgeme algoritmaları uygulanmıştır. Buradan elde edilen gömme vektörü, ağaç tabanlı algoritmalara girdi olarak verilip sınıflandırma başarısına olan etkisi incelenmiştir.

Tablo 4. 3: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PCA uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%74,98	%74,20	%72,52	%73,12
XGBoost	%76,12	%75,72	%73,58	%73,16
LightGBM	%77,82	%76,72	%74,42	%75,32

Tablo 4. 4: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına ICA uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%73,52	%73,48	%70,52	%70,40
XGBoost	%74,52	%74,48	%71,58	%71,62
LightGBM	%76,42	%77,05	%74,50	%74,35

Tablo 4. 5: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına UMAP uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%43,96	%42,96	%43,26	%42,27
XGBoost	%43,53	%43,28	%43,12	%42,96
LightGBM	%44,56	%44,68	%43,94	%43,56

Tablo 4. 6: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına t-SNE uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%39,41	%39,24	%38,56	%39,30
XGBoost	%40,51	%40,12	%40,18	%40,30
LightGBM	%42,58	%42,16	%42,56	%42,32

Tablolar incelendiğinde ikinci melez çerçevede UMAP ve t-SNE boyut indirgeme yöntemlerinin kullanılmasının başarımı oldukça düşürdüğü, buna ek olarak PCA ve ICA yaklaşımlarının kullanılmasının birinci melez çerçeveye kıyasla performans değerlerine etkisinin olmadığı gözlenmiştir.

4.3. ÜÇÜNCÜ MELEZ ÇERÇEVE İÇİN SONUÇLAR

Çalışmanın üçüncü melez çerçevesinde, ResNet18 ve ResNet34 modelleri ön-eğitim edilmiş ağırlıklar ile birleştirildikten sonraki katmana daha seçkin öznitelikleri seçmek amacıyla meta-sezgisel algoritmalar uygulanmıştır. Buradan elde edilen gömme vektörü, ağaç tabanlı algoritmalara girdi olarak verilerek sınıflandırma başarısı incelenmiştir.

Tablo 4. 7: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına CMA-ES uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%75,60	%76,00	%75,40	%75,30
XGBoost	%76,90	%76,20	%76,12	%76,20
LightGBM	%79,60	%78,09	%78,12	%76,09

Tablo 4. 8: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına hiper-parametre optimizasyonu yapılan CMA-ES uygulanması sonucunda bulunan en iyi değerler ile elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%77,80	%77,69	%76,82	%77,90
XGBoost	%79,32	%79,20	%79,12	%78,82
LightGBM	%81,70	%80,60	%80,32	%80,62

Tablo 4. 9: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PSO uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%72,46	%72,66	%72,18	%71,98
XGBoost	%74,52	%74,14	%74,47	%74,08
LightGBM	%78,79	%78,10	%78,32	%78,46

Tablo 4. 10: ResNet-Ağaç tabanlı algoritma melez çerçevesi ara katmanına hiper-parametre optimizasyonu yapılan PSO uygulanması sonucunda bulunan en iyi değerler ile elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%77,80	%77,50	%76,76	%76,08
XGBoost	%78,87	%78,47	%77,72	%77,78
LightGBM	%81,90	%81,66	%80,92	%80,52

Üçüncü melez çerçeve ile ilgili yapılan analizler, PSO ve CMA-ES gibi meta-sezgisellerin en-iyi hiper-parametre değerleri kullanılarak (Tablo 4.8 ve Tablo 4.10) öz niteliklerin seçilmesi sonucunda yapılan sınıflandırmada başarımların arttığı gözlemlenmiştir. PSO ve CMA-ES kıyaslandığında aralarında başarımların çok önemli bir farkın olmadığı görülmüştür.

4.4. DÖRDÜNCÜ MELEZ ÇERÇEVE İÇİN SONUÇLAR

Çalışmanın dördüncü melez çerçevesinde, ResNet-ağaç tabanlı algoritma melez çerçevesinde, ara katmanda meta-sezgisel algoritmalarından CMA-ES uygulandıktan sonraki katmana boyut indirgeme algoritmalarından PCA, ICA, UMAP ve t-SNE algoritmaları uygulanmıştır. Buradan elde edilen gömme vektörü ağaç tabanlı algoritmalara girdi olarak verilip sınıflandırma üzerindeki başarısı incelenmiştir.

Tablo 4. 11: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına PCA uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%70,80	%70,20	%70,24	%70,36
XGBoost	%69,64	%69,50	%69,36	%69,34
LightGBM	%72,80	%72,26	%72,46	%71,96

Tablo 4. 12: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına ICA uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%68,36	%68,12	%67,96	%67,90
XGBoost	%66,77	%66,24	%66,47	%66,72
LightGBM	%70,80	%70,14	%70,02	%70,12

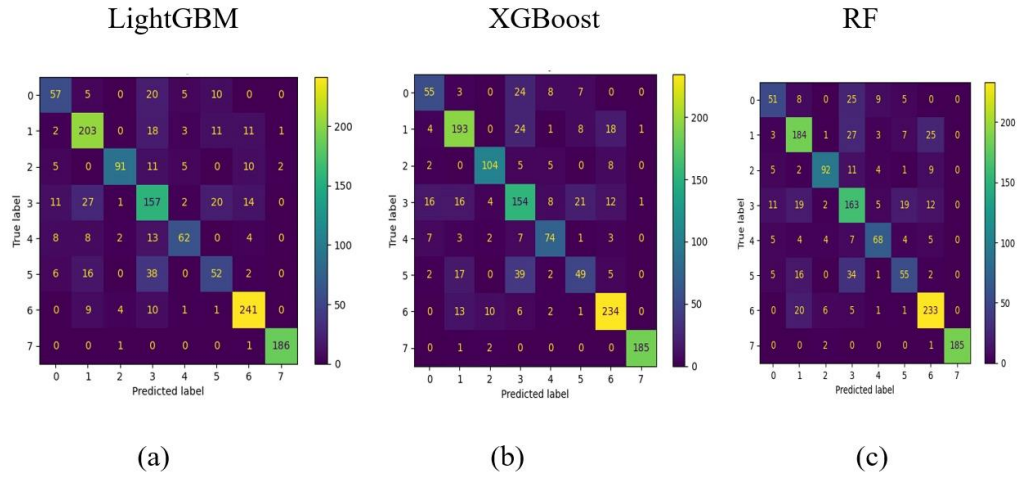
Tablo 4. 13: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına UMAP uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%57,12	%57,02	%57,18	%57,14
XGBoost	%56,30	%56,28	%56,20	%56,42
LightGBM	%59,12	%59,08	%59,10	%59,04

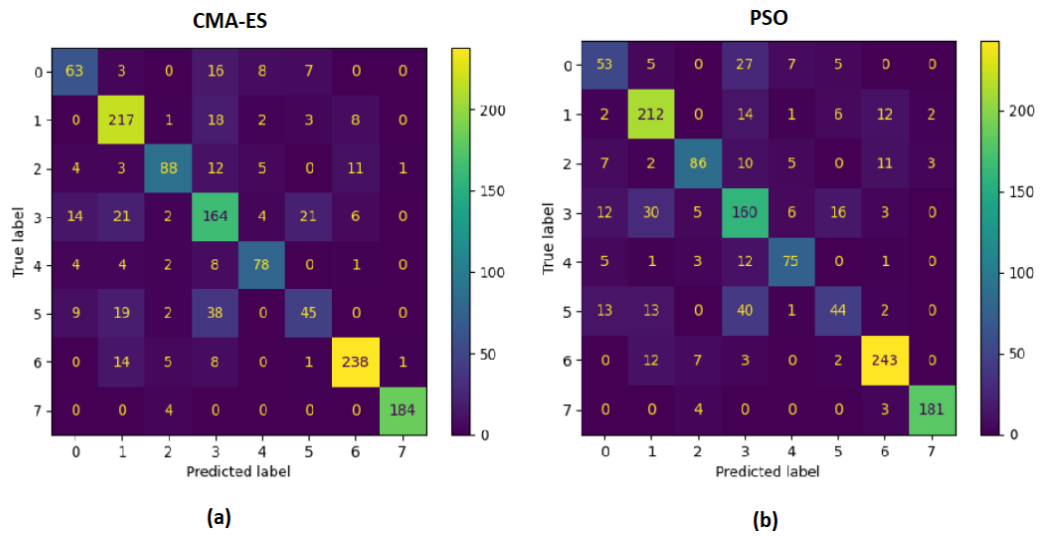
Tablo 4. 14: ResNet-CMA-ES-Ağaç tabanlı algoritma melez çerçevesi ara katmanına t-SNE uygulanması sonucunda elde edilen sonuçlar

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RF	%43,78	%43,16	%43,14	%43,38
XGBoost	%42,92	%42,68	%42,40	%42,80
LightGBM	%45,72	%43,36	%45,48	%45,12

Dördüncü melez çerçeve analiz sonuçları göstermiştir ki öz-nitelik seçiminin ardından uygulanan boyut indirgeme yaklaşımlarının sınıflandırma başarımlarını düşürmektedir. Bu da aslında öz-niteliklerin meta-sezgisel yaklaşımlar ile seçilmesinin yeterli olduğunu, genelleştirme performansını arttırdığını göstermektedir. CMA-ES ve PSO ile en etkili öz-nitelikler seçilmiş olup, ardından yapılan boyut azaltma yaklaşımları olası gerekli öz-niteliklerin etkilerini azalttığı için başarımları düşürmüştür.



Şekil 4. 1 İkinci melez çerçevede PCA algoritması ile (a) LightGBM algoritması için hata matrisi, (b) XGBoost algoritması için hata matrisi, (c) RF algoritması için hata matrisi



Şekil 4. 2 Üçüncü melez çerçevede LightGBM sınıflandırması ile (a) CMA-ES algoritması uygulandığında elde edilen hata matrisi (b) PSO algoritması uygulandığında elde edilen hata matrisi

Elde edilen sonuçlara göre LightGBM algoritması bütün algoritmalar içinde ve bütün performans değerlerinde en yüksek doğruluk değerini vermiştir. Ayrıca algoritmalar süre bakımından da kıyaslandığında yine en yüksek performans LightGBM ile oluşturulurken, en uzun süre XGBoost algoritması tarafından elde edilmiştir. Bu beklenen bir durumdur, XGBoost algoritması ağacı oluştururken sürekli yenileme yaptığı için ağacın oluşması uzun sürmektedir. Meta-sezgisel algoritmaların dahil edilmesi çalışma süresini arttırmıştır ancak özneliliklerin azaltılıp daha seçkin bir kümeye odaklanması performansı iyileştirmiştir. Hiperparametrelerin optimize edilip en iyilerinin seçilmesi bütün yöntemlerde performansı artırırken optimizasyon ile en yüksek performans artışı PSO algoritmasında görülmüştür. Boyut azaltma algoritmaları veriyi daha az boyut ile temsil ederek performansı artırır. Bu tez çalışmasında ICA ve PCA algoritmaları iyi performans sergilerken UMAP ve t-SNE algoritmaları bu veri seti için oldukça düşük performans sergilemiştir. Meta-sezgisel algoritmalarından sonraki aşamada tekrar boyut indirgeme algoritmalarını uygulamak hem zaman hem de sınıflandırma performansında düşüşe neden olmuştur.

SONUÇ

Bu çalışmada, BloodMNIST veri kümesi kullanılarak derin öğrenme, meta-sezgisel ve ağaç tabanlı algoritmaların melez bir çerçevesi oluşturuldu.

Çalışmadaki melez çerçeve öznitelik çıkarmayı ve sınıflandırmayı farklı algoritmalar ile yapmaktadır. Çalışmanın ilk melez çerçevesinde ResNet18 ve ResNet34 modelleri ön-eğitim edilmiş ağırlıklar ile öznitelik çıkarmak için kullanılmıştır. İki modelden gelen öznitelikler birleştirilmiş ve sınıflandırma algoritmalarına girdi olarak verilebilmesi için gömme vektörü elde edilmiştir. Bu vektör üç farklı ağaç tabanlı sınıflandırma algoritmasına verilerek sınıflandırma başarımı değerlendirilmiştir. İkinci melez çerçevede, ResNet-ağaç temelli algoritma çerçevesinin ara katmanına boyut indirgeme algoritmalarından PCA, ICA, UMAP ve t-SNE algoritmaları eklenerek verinin boyutu azlatılarak sınıflandırma başarısına olan etkisi incelenmiştir. Üçüncü melez çerçevede, ResNet-ağaç temelli algoritma çerçevesinin ara katmanına öznitelikleri azaltmak ve seçkin öznitelikleri seçebilmek için meta-sezgisel algoritmalarından CMA-ES ve PSO uygulanarak sınıflandırma başarımı üzerindeki etkisi incelenmiştir. Çalışmanın son melez çerçevesinde ise üçüncü çerçevedeki CMA-ES algoritmasının uygulanmasından sonraki katmana boyut indirgeme algoritmaları uygulanarak sınıflandırma başarımı üzerindeki etkisi incelenmiştir.

Sonuçlar incelendiğinde ilk melez çerçevede ResNet18 ve ResNet34 modellerinin ön-eğitilmiş ağırlıklarla özniteliklerinin çıkarılması oldukça hızlı bir işlemdir ve sınıflandırma aşamasında en yüksek doğruluk değerini LightGBM algoritması ile %76,24 olarak elde etmiştir. Sınıflandırma algoritmalarına hiper-parametre optimizasyonu uygulamanın doğruluk değerini arttırdığı görülmüştür.

İkinci melez çerçevede ara katmana eklenen boyut indirgeme algoritmaları PCA ve ICA için performansı etkilemezken, UMAP ve t-SNE algoritmalarının bu veri seti için performansı düşürdüğü gözlemlenmiştir.

Üçüncü melez çerçevede meta-sezgisel algoritmaların dahil edilmesi ile daha seçkin özniteliklerin seçilmesi hedeflenmiştir. Bu katmanda ara katmana eklenen CMA-ES ve PSO algoritmaları özellikle optimizasyon sonrasında bütün melez çerçevedeler arasındaki en yüksek doğruluk değerlerini vermiştir. Bu doğruluk

değerleri CMA-ES algoritması için %81,70 iken PSO algoritması için %81,90 olarak ölçülmüştür. İki değer birbirine oldukça yakın olmasına rağmen optimize edildikten sonra başarı artımı PSO için daha fazla olmuştur. İki algoritma arasında süre bakımından oldukça büyük fark vardır. CMA-ES algoritması PSO algoritmasına göre yaklaşık 5 kat daha hızlı çalışmaktadır.

Çalışmanın son melez çerçevesinde meta-sezgisel algoritmalarından sonraki katmana boyut indirgeme algoritmalarının uygulanması bütün boyut indirgeme algoritmalarında sınıflandırma başarımını düşürmüştür. CMA-ES ve PSO ile en etkili öz nitelikler seçilmiş olup, ardından yapılan boyut azaltma yaklaşımları olası gerekli öz niteliklerin etkilerini azalttığı için başarımı düşürmüştür.

Gelecekte, farklı veri setleri ve derin öğrenme yapıları üzerinde daha geniş kapsamlı deneyler yapılabilir. Ayrıca, farklı optimizasyon yöntemleri ve hiperparametre kombinasyonları da değerlendirilebilir. Bu çalışma kapsamında ResNet50 ve MobileNet algoritması birleştirilmiş ancak doğruluk değeri %45 olarak bulunmuştur. Gelecek çalışmalarda daha iyi kombinasyonlara odaklanabilir.

KAYNAKÇA

- LeCun, Y., Bengio, Y., & Hinton, G.** (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Şeker, A., Diri, B., & Balık, H. H.** (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Mühendislik Bilimleri Dergisi*, 3(3), 47-64.
- Anderson, D., & McNeill, G.** (1992). Artificial neural networks technology. Kaman Science Corporation, New York, USA.
- He, K., Zhang, X., Ren, S., & Sun, J.** (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Breiman, L.** (2001). Random forests. *Machine learning*, 45, 5-32.
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C.** (2010). Variable selection using random forests. *Pattern recognition letters*, 31(14), 2225-2236.
- Cutler, A., Cutler, D. R., & Stevens, J. R.** (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157-175.
- Chen, T., & Guestrin, C.** (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Li, M., Fu, X., & Li, D.** (2020, March). Diabetes prediction based on XGBoost algorithm. In *IOP conference series: materials science and engineering* (Vol. 768, No. 7, p. 072093). IOP Publishing.
- Kaya, Z., Kuş, Z., & Kiraz, B.** (2023). Hücre Canlılığı Tespitinde Yapay Öğrenme Yaklaşımları.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y.** (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Pearson, K.** (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11), 559-572.
- Sewell, M.** (2008). Principal component analysis. *University College London: London, UK*.

- McInnes, L., Healy, J., & Melville, J. (2018).** *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.*
- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W., Ng, L. G., ... & Newell, E. W. (2019).** Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*, 37(1), 38-44.
- McInnes, L., & Healy, J. (2020).** *UMAP: A Novel Dimension Reduction Technique.* Journal of Open Source Software, 3(29), 861
- Van der Maaten, L., & Hinton, G. (2008).** Visualizing data using t-SNE. *Journal of machine learning research*, 9(11)
- Linderman, G. C., & Steinerberger, S. (2019).** Clustering with t-SNE, provably. *SIAM journal on mathematics of data science*, 1(2), 313-332
- Van Der Maaten, L. (2014).** Accelerating t-SNE using tree-based algorithms. *The journal of machine learning research*, 15(1), 3221-3245.
- Hansen, N., & Ostermeier, A. (2001).** Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2), 159-195
- Kennedy, J., & Eberhart, R. (1995, November).** Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942-1948). Ieee
- Shi, Y., & Eberhart, R. (1998, May).** A modified particle swarm optimizer. In *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)* (pp. 69-73). IEEE
- Cengil, E., Çınar, A., & Yıldırım, M. (2022).** A hybrid approach for efficient multi-classification of white blood cells based on transfer learning techniques and traditional machine learning methods. *Concurrency and Computation: Practice and Experience*, 34(6), e6756.
- Toğaçar, M., Ergen, B., Cömert, Z., & Özyurt, F. (2020).** A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *Irbm*, 41(4), 212-222.
- Şengür, A., Akbulut, Y., Budak, Ü., & Cömert, Z. (2019, September).** White blood cell classification based on shape and deep features. In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-4). Ieee.
- Malkawi, A., Al-Assi, R., Salameh, T., Alquran, H., & Alqudah, A. M. (2020, October).** White blood cells classification using convolutional neural

network hybrid system. In *2020 IEEE 5th middle east and Africa conference on biomedical engineering (MECBME)* (pp. 1-5). IEEE.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4), 197-387.

Fathy, K. A., Yaseen, H. K., Abou-Kreisha, M. T., & ElDahshan, K. A. (2023). A novel meta-heuristic optimization algorithm in white blood cells classification. *Computers, Materials and Continua*, 75(1), 1527-1545.

