

Medikal Verilerin Sınıflandırılmasında Federe Öğrenme

Beyza Nur AKŞİT^{*1,2} Bahriye AKAY³ Adam SLOWIK⁴

^{*1}Aksaray Üniversitesi Mühendislik Fakültesi Yazılım Mühendisliği, AKSARAY

²Erciyes Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, KAYSERİ

³Erciyes Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, KAYSERİ

⁴Electronics and Computer Science, Koszalin University of Technology, Koszalin, Poland

(Alınış / Received: 01.08.2023, Kabul / Accepted: 20.11.2023, Online Yayınlanma / Published Online: 30.12.2023)

Anahtar Kelimeler
Federe Öğrenme,
İşbirlikçi Öğrenme,
Medikal Veri,
Veri Gizliliği,
Derin Öğrenme

Öz: Medikal verilerin sınıflandırılması ve analizi, çeşitli sağlık sorunlarının tanısında ve tedavisinde önemli bir rol oynar. Medikal veriler, içerdikleri hassas bilgiler nedeniyle özel güvenlik önlemlerine gereksinim duyarlar. Bu nedenle, veri paylaşımı olmadan model eğitimi işlemlerinin uç noktalarda gerçekleştirilmesini, veri paylaşımı yerine model parametrelerinin paylaşılmasını ve güncellenmesini sağlayan federe öğrenme yaklaşımları kullanılır. Bu sayede, veriler merkezi bir sunucuda toplanmadığından ve paylaşılmadığından, gizlilik riskleri azalır. Bir federe öğrenme yaklaşımı olan FedAvg, katılan tüm uç noktaların yerel model parametrelerinin ortalamasını alarak küresel modeli günceller. Ancak bu yöntemde, uç noktalar farklı model performanslarına sahip olduğunda yakınsama süresi ve performansı etkileyen sınırlamalar ortaya çıkar. Bu çalışmada bu sınırlamaları ortadan kaldırmak için en yüksek yerel model test doğruluğuna sahip uç noktanın yerel model parametrelerini kullanarak küresel modeli güncelleyen FedBest isimli bir yaklaşım önerilmiştir. Önerilen FedBest yaklaşımı ile FedAvg yaklaşımının performansları BloodMNIST, PathMNIST ve DermaMNIST veri setleri üzerinde kıyaslanmıştır. Deneylemlerden elde edilen sonuçlara göre, FedAvg'nin medikal sınıflandırmada başarılı olduğu ancak FedBest algoritmasının, daha yüksek doğruluk oranlarına ve daha hızlı bir yakınsamaya sahip olduğu görülmüştür.

Federated Learning for Medical Data Classification

Keywords
Federated Learning,
Collaborative Learning,
Medical Data,
Data Privacy,
Deep Learning

Abstract: The classification and analysis of medical data play a crucial role in the diagnosis and treatment of various health issues. Due to containing sensitive information, medical data require special security measures. Therefore, federated learning approaches are used to perform model training operations at the endpoints without data sharing, allowing the sharing and updating of model parameters instead. This way, privacy risks are reduced as data is not collected and shared on a central server. One federated learning approach called FedAvg updates the global model by taking the average of local model parameters from all participating endpoints. However, this method has limitations on convergence time and performance when endpoints have varying model performances. To address these limitations, this study proposes an approach called FedBest, which updates the global model using the local model parameters of the endpoint with the highest local model test accuracy. The performance of the proposed FedBest approach is compared to the FedAvg approach on the BloodMNIST, PathMNIST, and DermaMNIST datasets. According to the results obtained from the experiments, FedAvg demonstrates success in medical classification, but the FedBest algorithm shows higher accuracy rates and faster convergence.

*İlgili Yazar, email: aksitbeyza2@gmail.com

1. Giriş

Medikal verilerin işlenmesi ve analizi, sağlık alanında önemli bir rol oynamaktadır. Bu verilerin doğru şekilde analizi, hastalıkların erken teşhisi, tedavi planlaması ve hastaların sağlık durumlarının takibi gibi birçok klinik uygulama için hayati önem taşır. Ancak, medikal verilerin gizliliği ve güvenliği konuları, bu alanda çalışan araştırmacılar ve sağlık kuruluşları için önem teşkil etmektedir. Geleneksel olarak, medikal veriler merkezi sunucularda depolanır ve analiz edilir. Ancak, bu yaklaşım güvenlik açıklarına yol açabilir ve hasta gizliliği riskini artırabilir. Özellikle hassas kişisel sağlık bilgileri içeren verilerin merkezi sunucularda toplanması, yetkisiz erişim, veri ihlalleri ve kötü niyetli saldırılar gibi güvenlik risklerini beraberinde getirebilir.

Bu sorunları çözmek ve medikal verilerin gizliliğini korumak için, federe öğrenme yaklaşımı son yıllarda büyük ilgi görmektedir. Federe öğrenme, verilerin merkezi bir sunucuda bir araya getirilmesi yerine, veri sahipleri arasında dağıtılmış bir öğrenme sürecini kullanır [1]. Bu yöntemde, model eğitimi uç noktalarda gerçekleştirilir ve sadece güncellenmiş model parametreleri merkezi sunucuya iletilir. Bu sayede, medikal veriler uç noktalarda kalır ve gizlilik riski azalırken, aynı zamanda modelin toplam performansı da artırılabilir.

Federe öğrenme, medikal alanda özellikle medikal veri analizinde çeşitli çalışmalarda incelenmiştir. Zhang ve ark. izole medikal kurumlardaki verilerin gizliliğini dikkate alan, COVID19 tespiti yapabilen dinamik füzyon tabanlı bir federe öğrenme yaklaşımı önermişlerdir [2]. Odera, sağlık alanında federe öğrenme uygulaması için yapılan literatür taramasını kapsayan kapsamlı bir inceleme sunmuş, federe öğrenmede karşılaşılan zorluklar ve uygulama çerçevelerini ele almış ve aynı zamanda HIPPA ve GDPR gibi düzenlemelere uygun olarak gizlilik zorluklarına yönelik olası çözümleri tartışmıştır [3]. Sheller ve ark. çok sayıda uç noktada bulunan hasta verilerini paylaşmadan derin öğrenme ile beyin tümörü segmentasyonunu mümkün kılan bir federe öğrenme yaklaşımı önermişlerdir [4]. Yoo ve ark. medikal alanda federe öğrenme ortamı oluşturmak için yönergeler oluşturmuş ve bu yönergelerle birlikte federe öğrenimindeki kritik konuları ortaya koymayı ve uygulanabilir çözümler sunmayı amaçlamışlardır [5]. Wibawa ve ark. medikal verilerin gizliliğini ve güvenliğini artırmak için federe öğrenmede homomorfik şifreleme yaklaşımı önermişlerdir [6]. Farhad ve ark. Medikal Nesnelerin İnterneti'nde (IoMT) veri yönetimi ve mahremiyet sorununu ele almak için federe öğrenmenin potansiyelini tartışmışlardır [7]. Zhao ve Huang, bağımsız ve özdeş dağılımlı (Independent and Identically Distributed, IID) olmayan İveri sorunlarının yaygın olduğu medikal kurumları arasındaki her uç noktada yerel olarak veri üretirek uç noktalar arasında IID olmama durumunu azaltan, medikal görüntü sınıflandırmasına yönelik bir yenilikçi dağıtım bilgi paylaşımı federe öğrenme yöntemi olan FedDIS'i önermiştir. Deneysel sonuçlar, FedDIS'in IID olmayan veri senaryolarında Alzheimer hastalığı MRI ve MNIST veri kümeleri üzerindeki federe öğrenme performansının önemli ölçüde artırabileceğini göstermiştir [8]. Sohan ve Basalamah, medikal görüntü analizinde federe öğrenme üzerine sistematik bir inceleme yürütmüşlerdir [9]. Yi ve ark. hasta verilerinin gizliliğini sağlamak, büyük miktarda etiketli veri toplama ve üretme maliyetini ortadan kaldırmak için federe öğrenme tabanlı bir medikal teşhis yöntemi önermişlerdir [10]. Nazir ve Kaleem, derin sinir ağları ile medikal görüntü analizinde federe öğrenme uygulamalarını incelemiş, federe öğrenme model performansını artırmak için yapılan bazı çalışmalara değinmiş ve federe öğrenmenin gelecekteki araştırma yönlerini vurgulamışlardır [11]. Genel olarak, bu çalışmalar federe öğrenmenin medikal veri analizinde yüksek doğruluk elde ederken mahremiyeti koruyabilme potansiyelini göstermiştir.

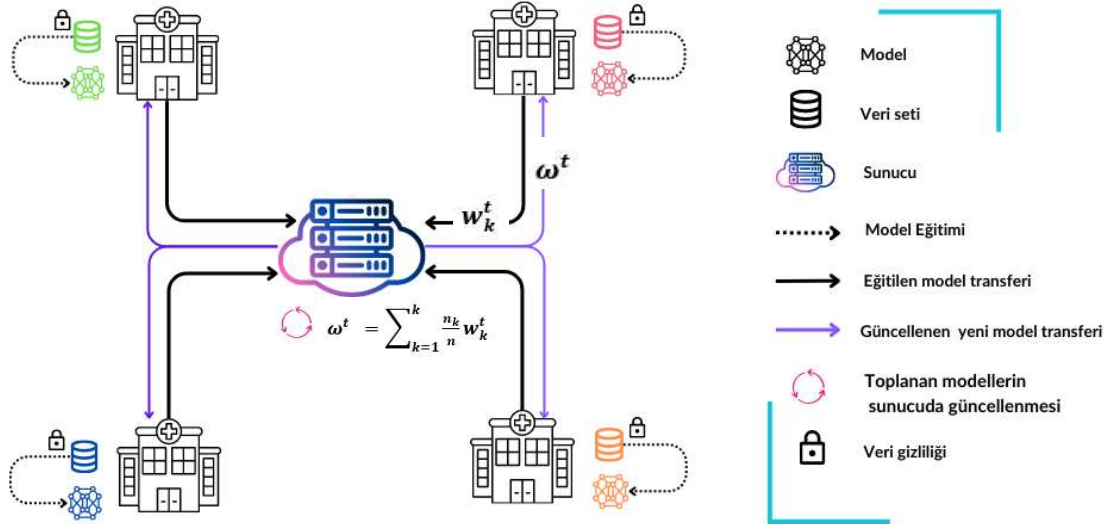
Bu çalışmanın amacı, medikal verilerin sınıflandırılmasında daha yüksek doğrulukta ve hızlı yakınsayabilen bir federe öğrenme yaklaşımı sunmaktır. Yeni bir federe öğrenme yaklaşımı olarak FedBest yaklaşımı önerilmiş ve literatürde yaygın kullanılan FedAvg yöntemi ile BloodMNIST, PathMNIST ve DermaMNIST veri setleri üzerinde kıyaslanmıştır. Sonuçlar incelendiğinde, FedBest yaklaşımı, küresel model test doğruluğu bakımından FedAvg yöntemine göre daha hızlı yakınsama elde ederek daha yüksek performans göstermiştir. Özellikle DermaMNIST ve BloodMNIST veri setlerinde, FedBest yaklaşımının daha yüksek başarı elde ettiği gözlenmiştir. DermaMNIST veri seti üzerinde, FedAvg yaklaşımında üç eğitim turu sonunda küresel model performansı %67,07'den %71,86'ya yükselirken, FedBest yaklaşımında ise %68,26'dan %73,05'e yükselmiştir. BloodMNIST veri seti için ise, FedAvg yaklaşımında üç eğitim turu sonunda küresel model performansı %37,68'den %84,51'e yükselirken FedBest yaklaşımında ise %73,94'ten %84,86'ya yükselmiştir. Bu sonuçlar, medikal veri sınıflandırma görevlerinde FedBest yaklaşımının Federe öğrenme süreci için etkili bir alternatif olabileceğini vurgulamaktadır.

Makalenin organizasyonu şu şekildedir: Birinci bölümde konuyla ilgili literatüre sunulan çalışmalar verilmiştir. İkinci bölümde, federe öğrenme yöntemleri ilgili bilgiler verilerle önerilen yöntem sunulmuştur. Üçüncü bölümde elde edilen sonuçlar verilmiştir. Dördüncü bölümde ise tartışma ve sonuca yer verilmiştir.

2. Materyal ve Metot

2.1. Federe Öğrenme Mimarisi

McMahan ve ark. [14] tarafından 2016'da önerilen Federe öğrenme [12,13], birden çok uç noktanın, verilerini merkezi bir sunucuya paylaşmadan küresel bir modelin eğitilmesine olanak tanıyan dağıtılmış bir makine öğrenimi tekniğidir. Federe öğrenmenin temel süreci Şekil 1'de gösterildiği üzere şu şekilde ilerler: İlk olarak, bir başlangıç modeli sunucu tarafından seçilen uç noktalara gönderilir. Daha sonra seçilen uç noktalar, kendi verileriyle yerel modeli eğitirler. Ardından, bu yerel modeller, güncellenmiş yerel model parametrelerini merkezi sunucuya gönderir. Sunucu, aldığı güncellenmiş parametreleri bir araya getirir ve küresel modeli günceller. Bu güncellenmiş küresel model tekrar uç noktalara gönderilir ve bu süreç tekrarlanır. Her iterasyonda, uç noktalar kendi verileri üzerinde öğrenme yapar, güncellemeleri paylaşır ve merkezi sunucu küresel modeli günceller. Bu şekilde, verilerin merkezi bir sunucuya gönderilmesi yerine, sadece güncellenmiş parametrelerin paylaşılmasıyla öğrenme gerçekleştirilir. Böylece, kullanıcıların veya veri sahiplerinin gizliliği korunurken, küresel model de sürekli olarak iyileştirilir.



Şekil 1. Federe öğrenme süreç şeması

2.2.1. Federated Averaging (FedAvg)

Google tarafından 2016 yılında önerilen FedAvg [15] yaklaşımında küresel modelin güncellenmesinde her bir uç noktadaki yerel model parametrelerinin ortalama bilgisinden faydalanılır. Her uç nokta, ham verilerini merkezi sunucuya göndermeden ve diğer uç noktalarla paylaşmadan model eğitim sürecine katkıda bulunur ve bu da Federe öğrenmeyi merkezi olmayan ortamlarda gizliliği koruyan değerli bir yaklaşım haline getirir. Algoritma 1’de FedAvg Algoritması sunulmuştur.

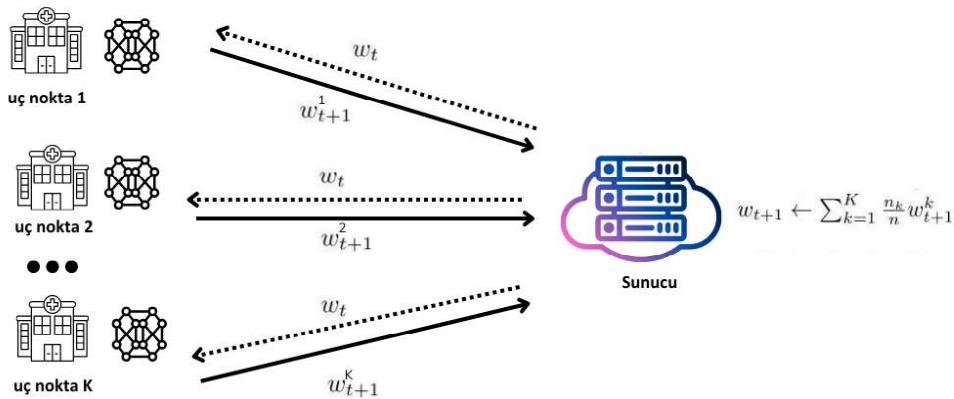
Algoritma 1. FedAvg algoritması [16]. K , k değişkeni tarafından endeksli uç nokta sayısı; n bütün katılımcılarda bulunan toplam örnek sayısı; n_k k ’nci katılımcının örnek sayısı, B , yerel minibatch boyutu; E , yerel epoch sayısı ve η , öğrenme oranıdır.

FedAvg Algoritması

[Sunucu]:
 Başla w_0
for her $t=1,2,\dots$ **do**
 $m \leftarrow \max([C.K], 1)$
 $S_t = \text{rastgele } m \text{ katılımcı ayarla}$
 for her katılımcı için $k \in S_t$ **in parallel do**
 $w_{t+1}^k = \text{YerelGuncelle}(k, w_k)$
 $w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k$

[Yerel]:
YerelGuncelle(k, w_k):
for her yerel epoch $i \in [1, E]$ **do**
 Yerel verileri böl, B ($\frac{B}{n_k}$ B yığınları)
 for yığın $b \in B$ **do**
 $w_{t,k} \leftarrow w_{t,k} - \eta_{\text{yerel}} \Delta l(w_{t,k}; b)$
return w to sunucu

Merkezi sunucuda, k değişkeni tarafından endekslenen K sayıda uç nokta seçilir. Seçilen uç noktalarda eğitilen yerel modellerin parametreleri merkezi sunucuya $\text{YerelGuncelle}()$ işlevi ile gönderilir. Son olarak, sunucu, K sayıda uç noktadan alınan tüm yerel model parametrelerinin (w_{t+1}^k) ortalamasını alarak küresel modeli günceller. Şekil 2’de FedAvg yaklaşımı sürecinin şeması sunulmuştur.



Şekil 2. FedAvg yaklaşımı süreç şeması.

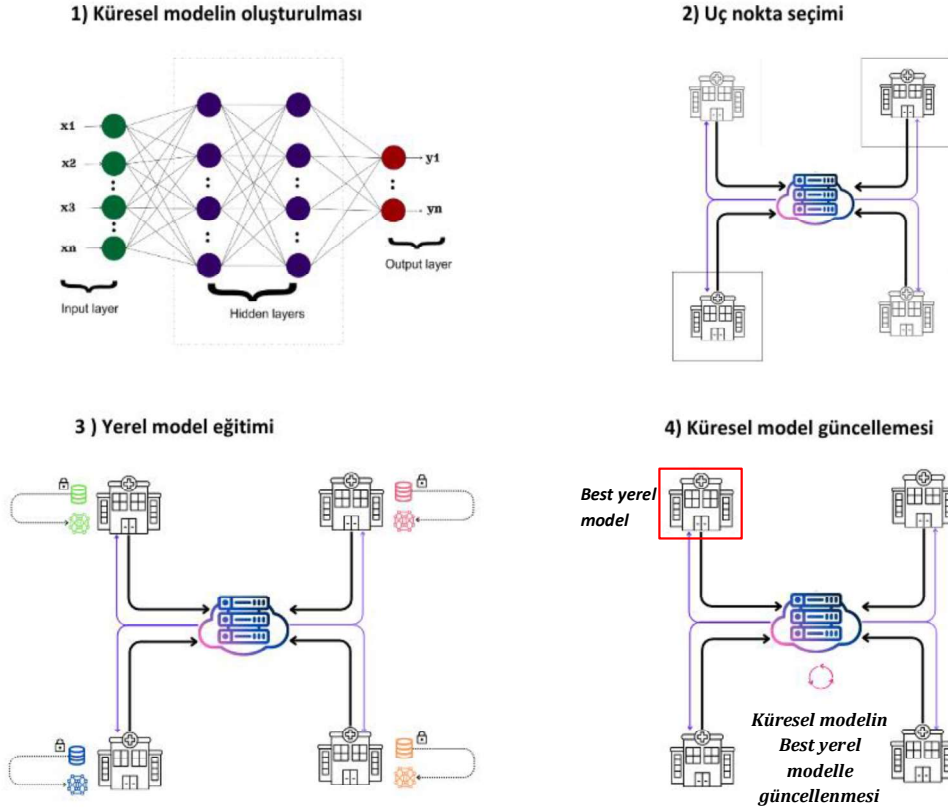
2.2.2 FedBest

FedBest, federe öğrenme sürecindeki en iyi performans gösteren yerel model parametrelerini kullanarak küresel modeli güncellemeyi amaçlayan bir yaklaşımdır. FedBest algoritmasının temel adımları Algoritma 2'de verilmiştir:

Algoritma 2. FedBest algoritmasının temel adımları

FedBest Algoritması
1. Başlangıç küresel modelin oluşturulması.
2. Rasgele uç noktaların seçilmesi.
repeat
3. Seçilen uç noktanın yerel model güncellemesi yapması:
a. Seçilen uç nokta, yerel verilerini kullanarak yerel modeli eğitir.
b. Güncellenen yerel model, merkezi sunucuya iletilir.
4. Küresel model güncelleme işlemi:
a. Merkezi sunucu, tüm uç noktalardan gelen yerel modelleri alır.
b. Küresel model, en iyi performans gösteren yerel modelin parametreleri ile güncellenir.
until Belirli bir durum sağlanana kadar (örneğin, maksimum iterasyon sayısına ulaşılanaya kadar veya belirli bir durma kriterine kadar).

1. **Başlangıç küresel modelin oluşturulması:** Oluşturulan başlangıç küresel model, verilerin özelliklerini yakalamak ve sınıflar arasında ayırım yapmak için etkili bir yapıda olmalıdır. Bu çalışmada CNN (Convolutional Neural Network) tabanlı bir başlangıç modeli oluşturulmuştur.
2. **Rastgele uç noktaların seçilmesi:** Rastgele uç nokta seçimi, federe öğrenme sürecinin dağıtılmış ve işbirlikçi yapısını sağlar. Farklı uç noktaların katılımı, daha çeşitli ve kapsayıcı bir veri seti temsili anlamına gelir. Bu da küresel modelin daha genel ve güncel bir şekilde güncellenmesini sağlar.
3. **Seçilen uç noktaların yerel model güncellemesi yapması:**
 - a. **Seçilen uç nokta, yerel verilerini kullanarak yerel modeli eğitir:** Her seçilen uç nokta, kendi yerel verilerini kullanarak eğitim işlemi gerçekleştirir. Yerel model, uç noktanın sahip olduğu veriye özgü bilgileri içerir ve yerel eğitim süreci boyunca güncellenir.
 - b. **Güncellenen yerel model, merkezi sunucuya iletilir:** Uç noktalardaki yerel eğitim tamamlandıktan sonra güncellenen yerel model, merkezi sunucuya iletilir.
4. **Küresel model güncelleme işlemi:**
 - a. **Merkezi sunucu, tüm uç noktalardan gelen yerel modelleri alır.**
 - b. **Küresel model en iyi performans gösteren yerel modelin parametreleri ile güncellenir:** Merkezi sunucu, alınan yerel modeller arasından en iyi performans gösteren modeli seçer. Bu yaklaşım, küresel modelin daha iyi performans gösteren yerel modelin bilgisini küresel modele entegre etmesini sağlar. En iyi performans gösteren yerel model parametreleri, küresel modelin performansını artırmak için kullanılır. Şekil 3'te FedBest yaklaşımı süreç şeması sunulmuştur.



Şekil 3. FedBest yaklaşımı süreç şeması.

3. Deneysel Çalışmalar ve Bulgular

Bu bölümde, FedAvg ve bu çalışmada önerilen FedBest yaklaşımlarının medikal veri sınıflandırmada etkinliklerini değerlendirilen deneysel çalışmalara yer verilmiştir. Bu çalışmada sınıflandırma algoritması olarak CNN (Convolutional Neural Network) mimarisi kullanılmıştır. Çalışma kapsamında uç nokta sayısı dört olarak belirlenmiştir yani, federe öğrenme, CNN modelinin dört uç noktadan oluşan merkezi olmayan bir ağ üzerinde eğitilmesini içeriyor. Medikal veriler, dört uç noktaya eşit şekilde dağıtıldıktan sonra veriler uç noktalarda kalmış ve bu şekilde veri gizliliği ve güvenliği sağlanmıştır. CNN modeli, her uç noktanın kendi verileri üzerinde yerel olarak eğitilmiş ve böylece her uç noktanın veri kümesinin belirli özelliklerinden öğrenmesine olanak sağlanmıştır. Yerel eğitimden sonra, küresel bir model oluşturmak için dört uç noktadan alınan yerel modeller sunucuda bir araya getirilmiştir. Federe öğrenme algoritmalarından FedAvg yaklaşımının performansı ve önerilen FedBest yaklaşımının performansı BloodMNIST, PathMNIST ve MedMNIST medikal veriler ile yapılan deneyler üzerinden değerlendirilmiştir.

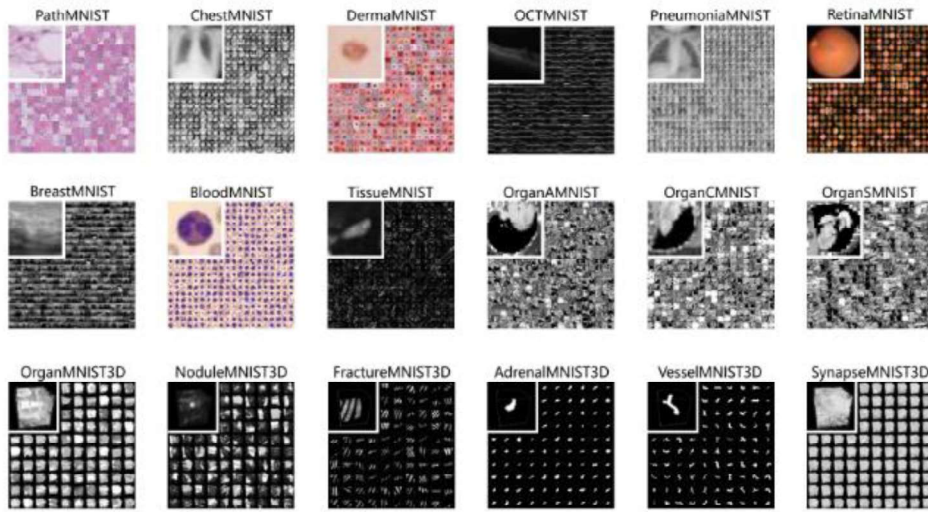
Modeller Google Colab'da Python programlama dili ile kodlanmıştır. Deneyler Intel (R) Core (TM) i7-10750H 2.6 GHz işlemci, 16 GB RAM ve Windows 11 işletim sistemi bulunan bilgisayar ile yapılmıştır. Bu çalışmada, Keras kütüphanesi kullanılarak bir CNN modeli oluşturulmuştur. Oluşturulan CNN modelin mimarisine ait parametreler Tablo 1'de verilmektedir.

Tablo1. CNN Modelinin yapısı ve parametre sayısı

Layer(type)	Output Shape	#parameters
conv2d(CONV2D)	(None,26,26,28)	784
Max_pooling2d(MaxPooling2D)	(None,13,13,28)	0
Conv2d_1(CONV2D)	(None,11,11,56)	14168
Max_pooling2d_1(MaxPooling2D)	(None,5,5,56)	0
Conv2d_2(CONV2D)	(None,3,3,56)	28280
flatten (Flatten)	(None,504)	0
dense (Dense)	(None,56)	28280
dense_1(Dense)	(None,10)	570
Droupout (Droupout)	(None,10)	0
Total parameters	-	72,082
Trainable parameters	-	72,082
Non-trainable parameters	-	0

3.1. Veri Seti

Şekil 4'te bazı örnekleri verilen MedMNIST, standartlaştırılmış biyomedikal görüntüler içeren geniş ölçekli bir veri kümesidir. 2-Boyutlu (2B) 12 veri kümesi ve 3-Boyutlu (3B) 6 veri kümesi içermektedir. MedMNIST, farklı veri ölçeklerinde (100'den 100.000'e kadar) ve farklı görevlerde (ikili / çok sınıflı, ordinal regresyon ve çok etiketli) 2B ve 3B görüntüler üzerinde sınıflandırma yapmak için tasarlanmıştır. Tüm görüntüler, ilgili sınıflandırma etiketleriyle birlikte 28 x 28 (2B) veya 28 x 28 x 28 (3B) boyutlarına önceden boyutlandırılmıştır. Biyomedikal görüntülerdeki farklı veri türlerini kapsayan toplamda 708.069 adet 2B görüntü ve 9.998 adet 3B görüntü içermektedir [18].

**Şekil 4.** MedMNIST 'e genel bakış

Bu çalışmada, medikal görüntülerden oluşan MEDMNIST veri setinden BloodMNIST, PathMNIST, DermaMNIST kullanılarak deneyler yapılmıştır.

3.1.1 BloodMNIST Veri Seti

BloodMNIST veri kümesi, enfeksiyon, hematolojik ya da onkolojik hastalığı olmayan ve kan toplama anında herhangi bir farmakolojik tedavi görmeyen bireylerin normal hücrelerinden elde edilen bir veri kümesidir. Toplam 17.092 görüntü içermektedir ve 8 sınıfa ayrılmıştır. Kaynak veri kümesi eğitim, doğrulama ve test seti olmak üzere 7:1:2 oranında bölünmüştür. Kaynak görüntülerin çözünürlüğü $3 \times 360 \times 363$ pikseldir ve merkezden kırılarak $3 \times 200 \times 200$ boyutlarına indirgenmiş, ardından $3 \times 28 \times 28$ boyutlarına yeniden boyutlandırılmıştır [18].

3.1.2. PathMNIST Veri Seti

PathMNIST, kolorektal kanser histoloji slaytlarından sağkalım tahminine yönelik 100.000 adet görüntü yaması (NCT-CRC-HE-100K) ve farklı bir klinik merkezden sağlanan 7.180 adet görüntü yaması (CRC-VAL-HE-7K) içeren bir veri setidir. Veri seti, 9 farklı doku türünden oluşur ve çok sınıflı sınıflandırma görevini içerir. Kaynak görüntüleri $3 \times 224 \times 224$ piksel boyutundan $3 \times 28 \times 28$ piksele yeniden boyutlandırılmıştır. NCT-CRC-HE-100K 9:1 oranında eğitim ve doğrulama setlerine ayrılmıştır. CRC-VAL-HE-7K test seti olarak kullanılmaktadır [18].

3.1.3. DermaMNIST Veri Seti

DermaMNIST, yaygın pigmentte cilt lezyonlarının çok kaynaklı dermatoskopik görüntülerinin büyük bir koleksiyonu olan HAM1000019,20,21 temel alınarak oluşturulmuştur. Veri kümesi, 7 farklı hastalığa ait sınıflandırılmış 10.015 dermatoskopik görüntü içermektedir ve çoklu sınıf sınıflandırma görevi olarak tanımlanmıştır. Görüntüler eğitim, doğrulama ve test setlerine 7:1:2 oranında bölünmüştür. Kaynak görüntülerin boyutu $3 \times 600 \times 450$ 'dir ve $3 \times 28 \times 28$ boyutuna yeniden boyutlandırılmıştır [18].

3.2. MedMNIST Veri Seti Deneysel Sonuçları

Kullanılan MedMNIST veri setlerinin adil bir şekilde dağıtılmasını sağlamak için eğitim veri seti eğitim öncesi karıştırılmış ve dört uç nokta arasında eşit olarak paylaştırılmıştır. Eğitim sürecinde her turda ve her uç noktada eğitilen model için aynı sayıda epoch sayısı kullanılmıştır ve modelin aşırı öğrenmesini önlemek için federe öğrenme eğitim turu sayısı üç olarak belirlenmiştir. Tablo 2-4'te, BloodMNIST, PathMNIST ve DermaMNIST için her bir uç noktanın yerel olarak eğitilmiş modellerin performansını yansıtan eğitim doğruluğu ve test doğruluğu metrikleri sunulmaktadır. Yerel modellerin bir araya getirilmesiyle elde edilen küresel modelin test sonucu da Tablo 2-4'ün son satırlarında gösterilmektedir.

3.2.1. BloodMNIST Veri Seti Deneysel Sonuçları

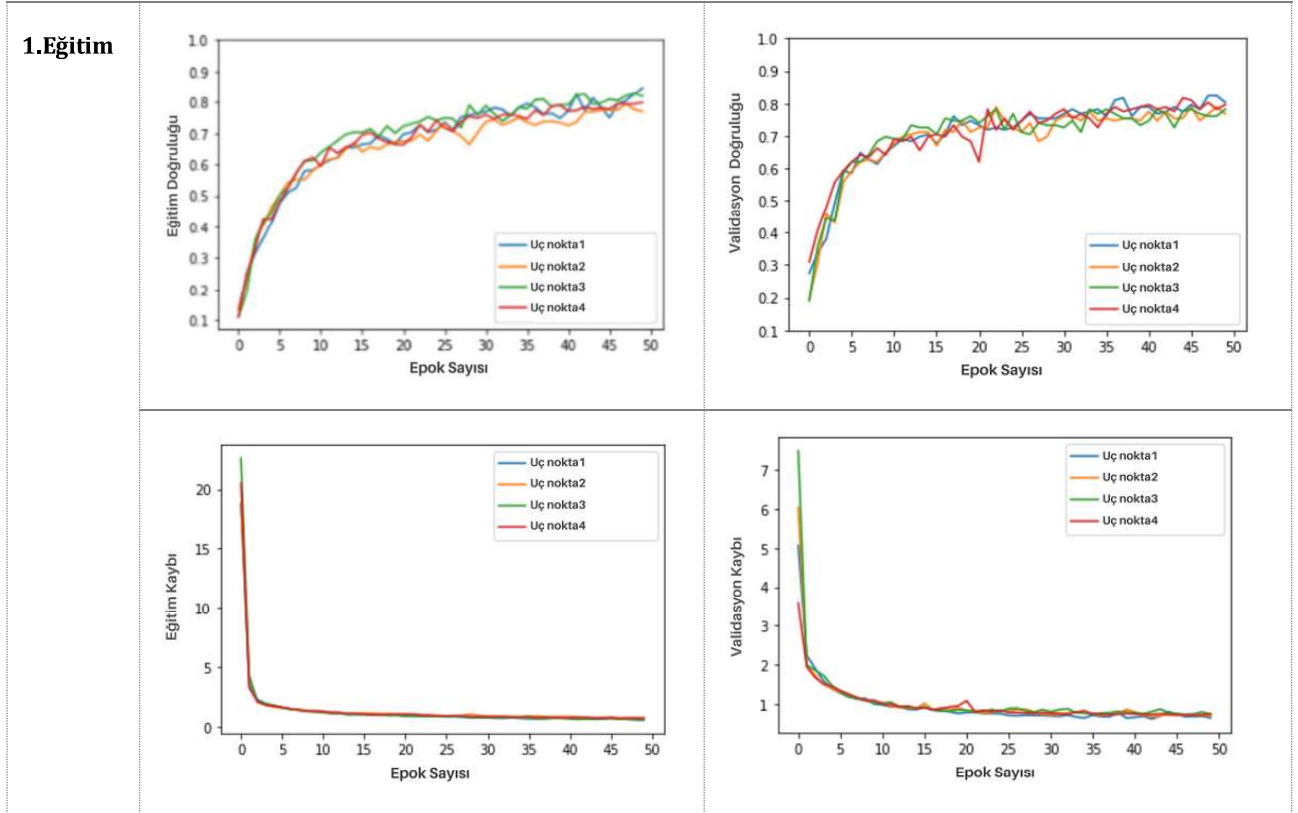
Tablo 2, BloodMNIST veri seti için FedAvg ve FedBest yaklaşımlarına ait üç eğitim turundaki sınıflandırma sonuçlarını sunmaktadır. Tablo 2'deki sonuçlar, FedAvg ve FedBest yaklaşımları için üç eğitim turunun sonunda, tüm uç noktadaki yerel model doğruluklarının ilk turdaki eğitim sonuçlarına göre nispeten yüksek doğruluk değerlerine ulaştığını göstermektedir. FedAvg yaklaşımı kullanılarak eğitilen küresel model, üçüncü eğitim turundan sonra %84,51'lik bir doğruluk elde ederken, FedBest yaklaşımında %84,86'lık bir doğruluk elde etmiştir. Bu değerler, modelin doğruluğunu geliştirmede her iki yaklaşımın etkinliğini göstermektedir. FedAvg ve FedBest yaklaşımlarında, uç noktaların yerel model doğruluk değerleri üç eğitim turu gerçekleştikten sonra incelendiğinde, önemli performans farklılıkları ortaya çıkmaktadır. Üç eğitim turu sonrası FedAvg yaklaşımında, uç nokta 1 yerel modeli %82,39, uç nokta 2 yerel modeli %82,04, uç nokta 3 yerel modeli %83,45 ve uç nokta 4 yerel modeli %83,10 doğruluk elde etmiştir. FedBest yaklaşımında ise uç nokta 1 yerel modeli %81,69, uç nokta 2 yerel modeli %83,45, uç nokta 3 yerel modeli %82,75 ve uç nokta 4 yerel modeli %82,75 doğruluk elde etmiştir. FedAvg yaklaşımında üç eğitim turu sonunda küresel model performansı %37,68'den %84,51'e yükselirken FedBest yaklaşımında ise %73,94'den %84,86'ya yükselmiştir.

FedBest yaklaşımı, BloodMNIST veri setinin sınıflandırılmasında yakınsama hızı ve genel performans açısından FedAvg'ye göre daha iyi performans göstermiştir. Her eğitim turu öncesi küresel model güncellemesi için sadece en iyi test doğruluğuna sahip uç noktanın yerel model parametreleri kullanıldığında, bir sonraki turda uç noktaların yerel model performansları test doğruluğu bakımından bir önceki tura göre genellikle daha iyi sonuç elde etmiştir.

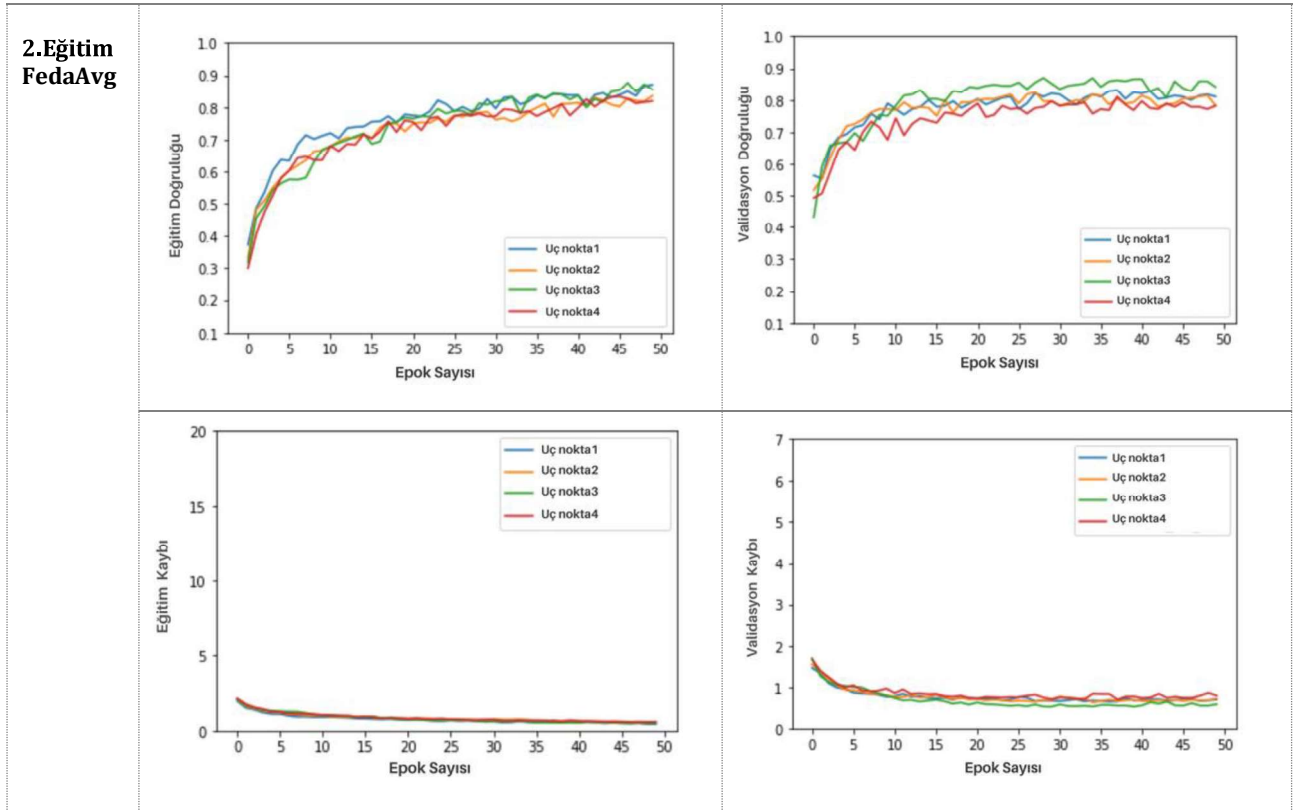
Tablo 2. BloodMNIST Veri Seti Deneysel Sonuçları

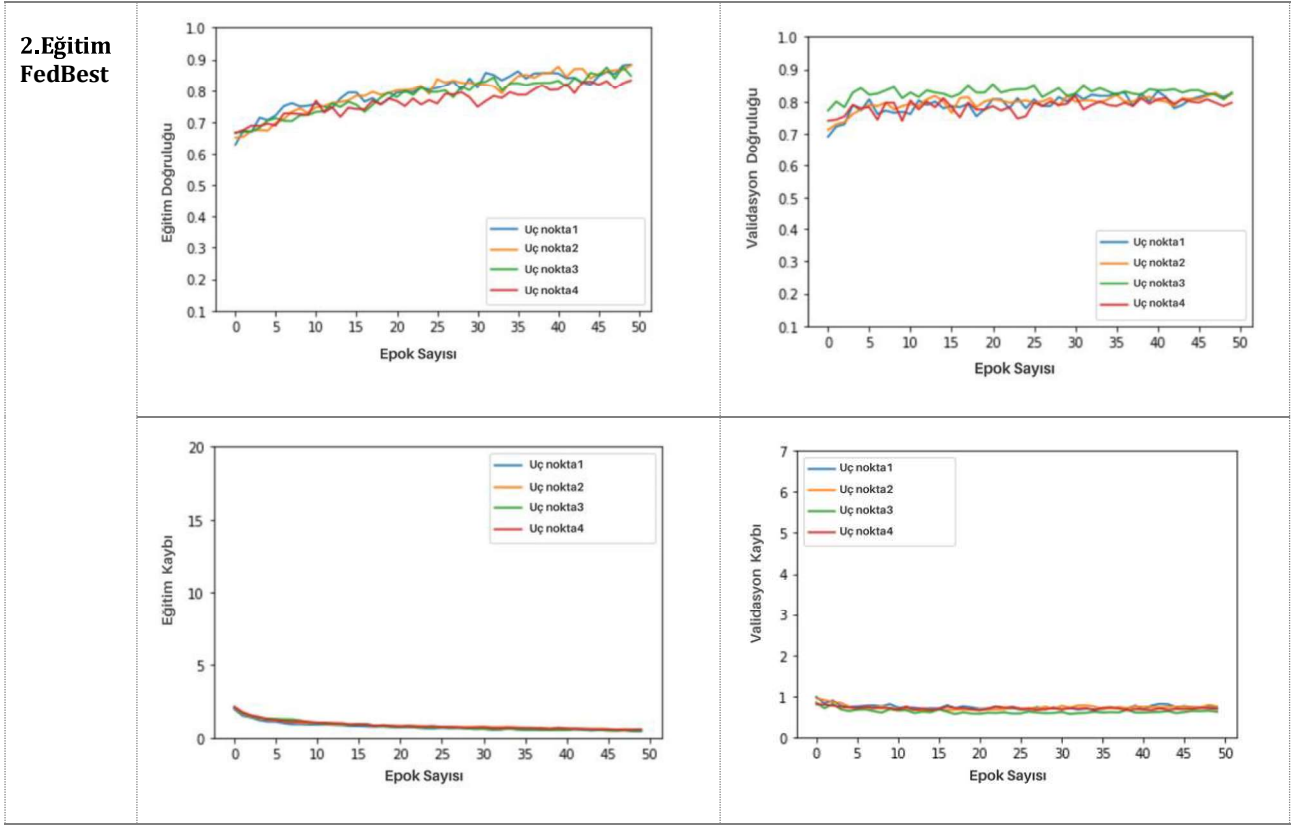
Uç noktalar		CNN Model	2. Eğitim FedAvg	2. Eğitim FedBest	3. Eğitim FedAvg	3. Eğitim FedBest
1	Train acc:	0.8464	0.8705	0.8835	0.8665	0.9127
	Test acc:	0.7711	0.8451	0.8275	0.8239	0.8169
2	Train acc:	0.7691	0.8343	0.8815	0.8685	0.8906
	Test acc:	0.7465	0.7711	0.8063	0.8204	0.8345
3	Train acc:	0.8183	0.8544	0.8484	0.8785	0.9016
	Test acc:	0.8134	0.8169	0.8134	0.8345	0.8275
4	Train acc:	0.7982	0.8183	0.8333	0.8765	0.8835
	Test acc:	0.7606	0.7923	0.8169	0.8310	0.8275
Küresel Model	Test acc:	FedAvg: 0.3768 FedBest: 0.7394	0.7570	0.8451	0.8451	0.8486

Şekil 5-7'de üç eğitim boyunca alınan uç noktalardaki yerel model performans sonuçlarının grafikleri sunulmaktadır. Grafikler, 50 epoch sonunda her uç noktanın yerel modelinin eğitim doğruluğu (train accuracy) ve kayıp değerlerini (train loss), ayrıca validasyon doğruluğu (validation accuracy) ve kayıp değerlerini (validation loss) göstermektedir. Grafikler, her uç noktanın eğitim sürecindeki performansını ve yerel modelin doğruluğu ile kayıp üzerindeki değişimleri takip etmemize olanak sağlamaktadır. Şekil 5'te ilk eğitim turu için merkezi sunucudan her uç noktaya gönderilen başlangıç CNN modelinin, uç noktalardaki yerel verilerle eğitim performans sonuçları grafikleri bulunmaktadır. Şekil 5'teki grafikler incelendiğinde, her bir uç noktanın yerel model performansı eğitim sürecinin başlangıcında eğitim ve validasyon doğruluğu için düşük bir değerde başladığını ancak yerel modellerin 50 epok sonunda eğitim ve validasyon performanslarının arttığını göstermiştir. Eğitim ve validasyon kayıp grafiklerinde ise başlangıçta yüksek bir kayıp değeri vardır, ancak eğitim turu sonuna doğru kayıp değeri düşmüştür. Bu kayıp grafiğindeki düşüş, modelin genel performansını iyileştirdiğini göstermektedir. Şekil 6'da ilk eğitim turunda eğitilen yerel model parametreleri merkezi sunucuda FedAvg ve FedBest yaklaşımları ile güncellenen yeni küresel modelin, uç noktalardaki yerel verilerle ikinci eğitim turu performans sonuçlarının grafikleri bulunmaktadır. İkinci eğitim turu öncesi merkezi sunucuda güncellenen küresel model, ilk eğitim turundaki başlangıç CNN modelin, eğitim ve validasyon başlangıç doğruluk değerlerine göre her iki yaklaşımda da daha yüksek değer ile başladığı görülmüştür. Özellikle FedBest yaklaşımı, başlangıç değerlerinin FedAvg'ye göre daha yüksek değerde başladığı grafiklerde gösterilmiştir. Bu durum, ikinci eğitim turu için uç noktalara gönderilen küresel modelin, ilk eğitim turundaki başlangıç CNN modeline kıyasla, federe öğrenme yaklaşımlarıyla güncellenerek küresel model performansının geliştiğini göstermektedir. Şekil 7'de ise ikinci eğitim turunda eğitilen yerel modellerin merkezi sunucuda FedAvg ve FedBest yaklaşımları ile küresel modeli güncellenmesi sonucunda elde edilen yeni küresel modelin, her bir uç noktaya tekrar gönderilip eğitime başlandığında, yerel model başlangıç performanslarının bir önceki tura göre daha yüksek başlangıç değerle eğitime başladığı görülmüştür. Üçüncü eğitim turu sonunda, model performansının belirli bir seviyeye ulaştığı ve değişiminin sabitlendiği gözlenmiş ve üçüncü eğitim turunda federe öğrenme süreci tamamlanmıştır.

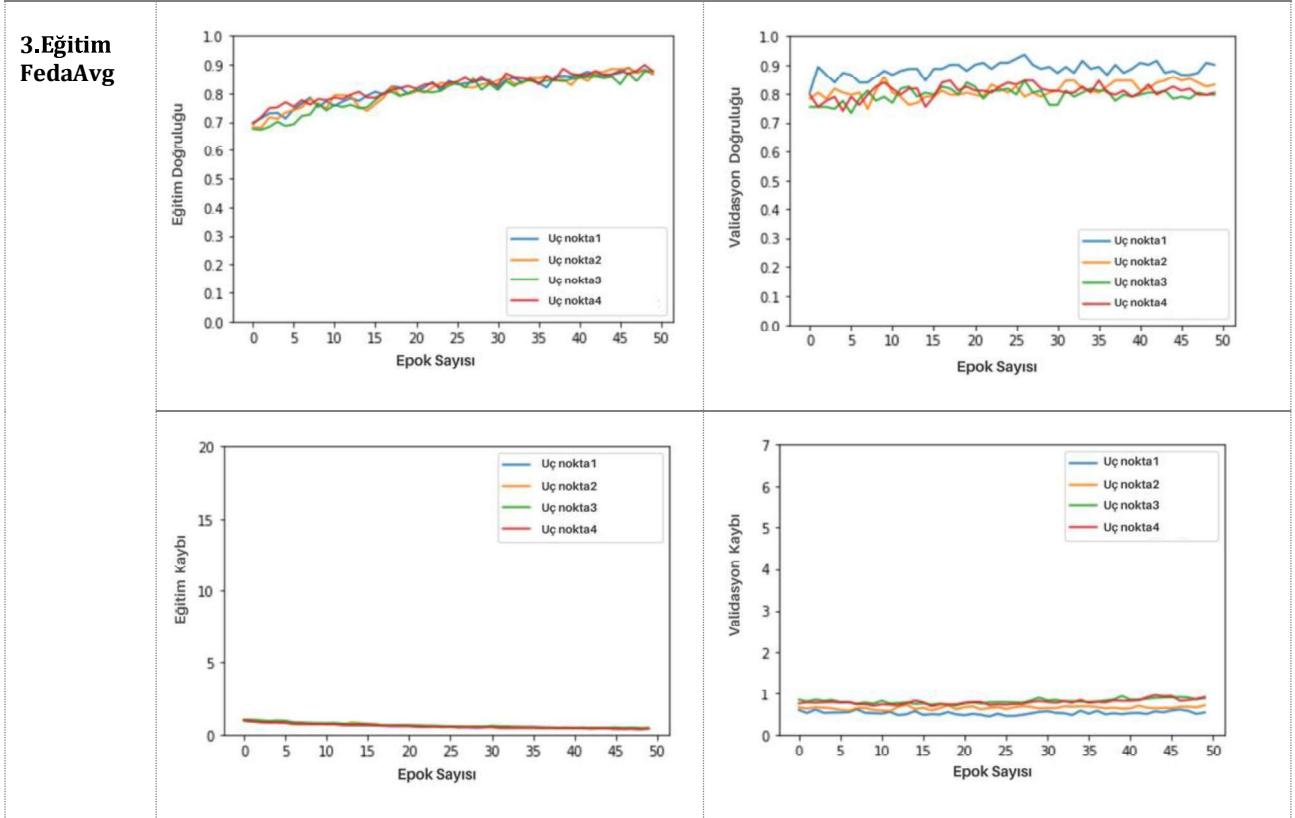


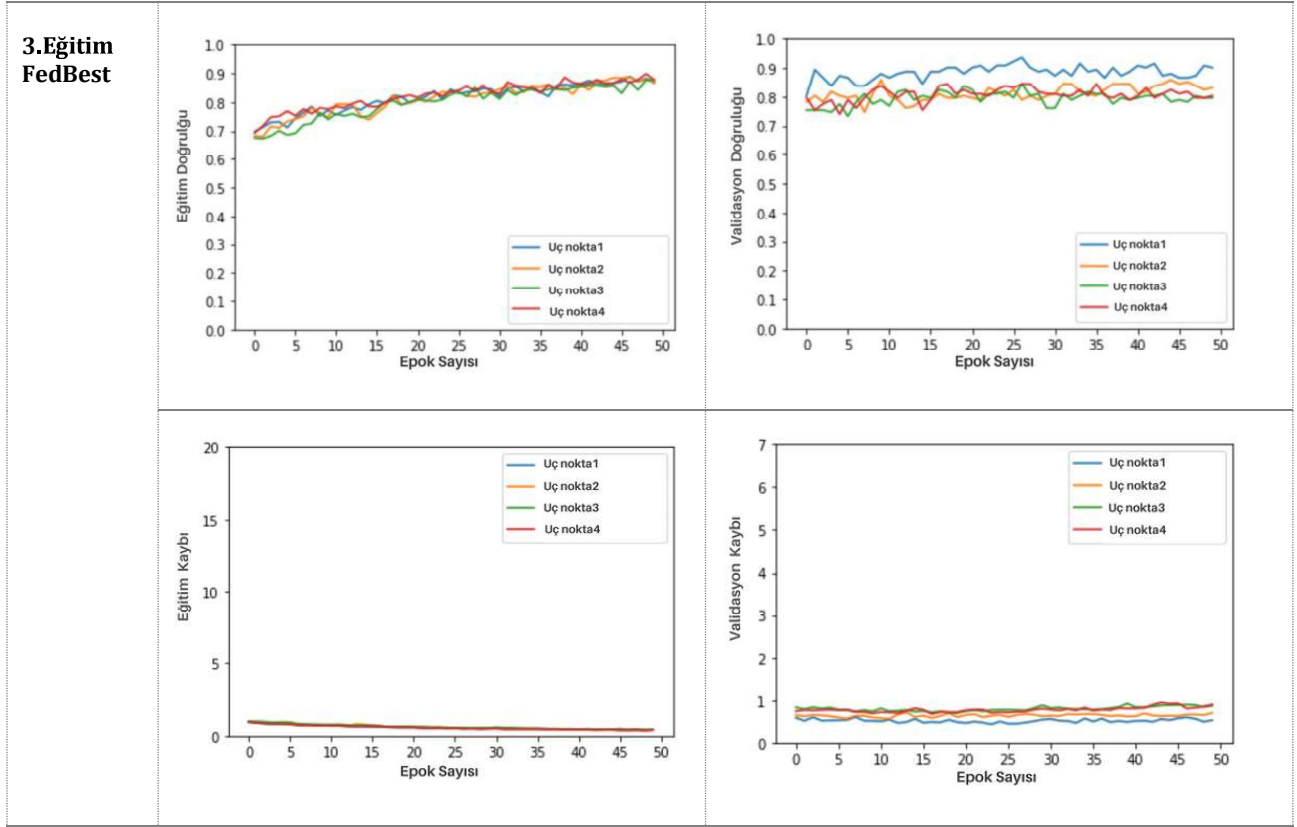
Şekil 5. Birinci eğitim sonunda BloodMNIST veri seti için bulunan sonuçların grafikleri; x eksenini, epok sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/Validation accuracy) ve Eğitim/Validasyon kaybı (Train/Validation loss) değerlerini göstermektedir.





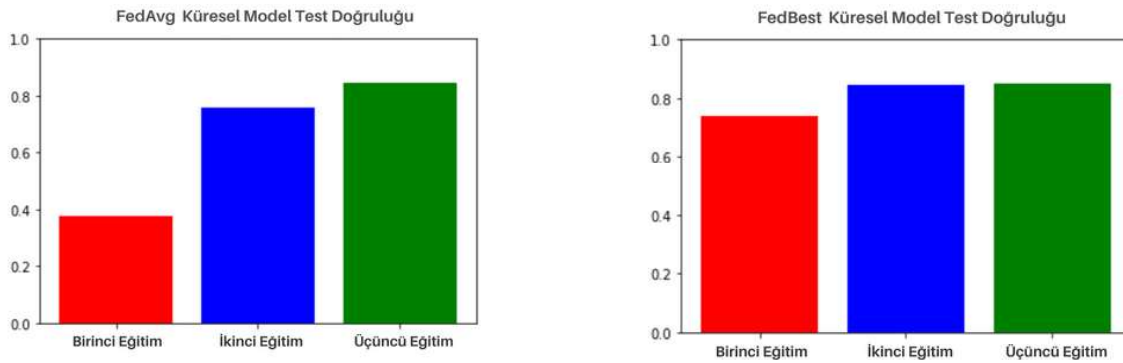
Şekil 6. İkinci eğitim sonunda BloodMNIST veri seti için bulunan sonuçların grafikleri; x eksenini, epok sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/ Validation accuracy) ve Eğitim/Validasyon kaybı (Train/ Validation loss) değerlerini göstermektedir.





Şekil 7. Üçüncü eğitim sonunda BloodMNIST veri seti için bulunan sonuçların grafikleri; x eksenini, epoch sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/Validation accuracy) ve Eğitim/Validasyon kayıp (Train/Validation loss) değerlerini göstermektedir.

Şekil 8'de Tablo 1'de son satırda belirtilen her üç eğitim turu sonunda merkezi sunucuda FedAvg ve FedBest yaklaşımlarıyla güncellenen küresel modelin test doğruluk sonuçlarının çubuk grafikleri sunulmuştur. FedAvg yaklaşımı kullanarak güncellenen küresel model performansı ilk eğitim turunda %37,68, ikinci eğitim turunda %75,70, üçüncü eğitim turunda %84,51 değerinde model test doğruluğu performansı göstermiştir. FedBest yaklaşımı kullanılarak güncellenen küresel model performansı ilk eğitim turunda %73,94, ikinci eğitim turunda %84,51, üçüncü eğitim turunda %84,86 değerlerinde model test doğruluğu performansı göstermiştir. Şekil 8'de görüldüğü üzere FedBest yaklaşımı FedAvg'ye göre küresel model test doğruluğu bakımından daha hızlı yakınsama elde etmiştir.



Şekil 8. BloodMNIST veri seti her eğitim turu sonu FedAvg ve FedBest yaklaşımlarıyla elde edilen küresel model performansları.

3.2.2. PathMNIST Veri Seti Deney Sonuçları

PathMNIST veri seti için FedAvg ve FedBest yaklaşımlarına ait üç eğitim turundaki sınıflandırma sonuçları Tablo 3'te sunulmaktadır. Buna göre, tüm uç noktaların yerel model doğrulukları üç eğitim turu sonunda daha yüksek doğruluk değerlerine ulaşmıştır. FedAvg yaklaşımı kullanılarak eğitilen küresel modelin üç eğitim turu sonunda küresel model test doğruluğunun %79,93 doğruluk verdiğini, FedBest yaklaşımında ise küresel modelin %74,08 test doğruluğu verdiği görülmektedir. Bu değerler, küresel modelin doğruluğunu geliştirmede her iki yaklaşımda etkinliğini göstermektedir. FedAvg ve FedBest yaklaşımlarında, uç noktaların yerel model doğruluk değerleri üç eğitim turu gerçekleşikten sonra incelendiğinde, FedAvg yaklaşımında, uç nokta 1 yerel modeli %69,40, uç nokta 2 yerel modeli %71,74, uç nokta 3 yerel modeli %70,23 ve uç nokta 4 yerel modeli %72,07 doğruluk elde etmiştir. FedBest yaklaşımında ise uç nokta 1 yerel modeli %81,69, uç nokta 2 yerel modeli %83,45, uç nokta 3 yerel modeli %82,75 ve uç nokta 4 yerel modeli %82,75 doğruluk elde etmiştir. FedAvg yaklaşımında üç eğitim turu sonunda küresel model performansı %14,05'ten %79,93'e yükselirken, FedBest yaklaşımında ise %68,06'dan %74,08'e yükselmiştir.

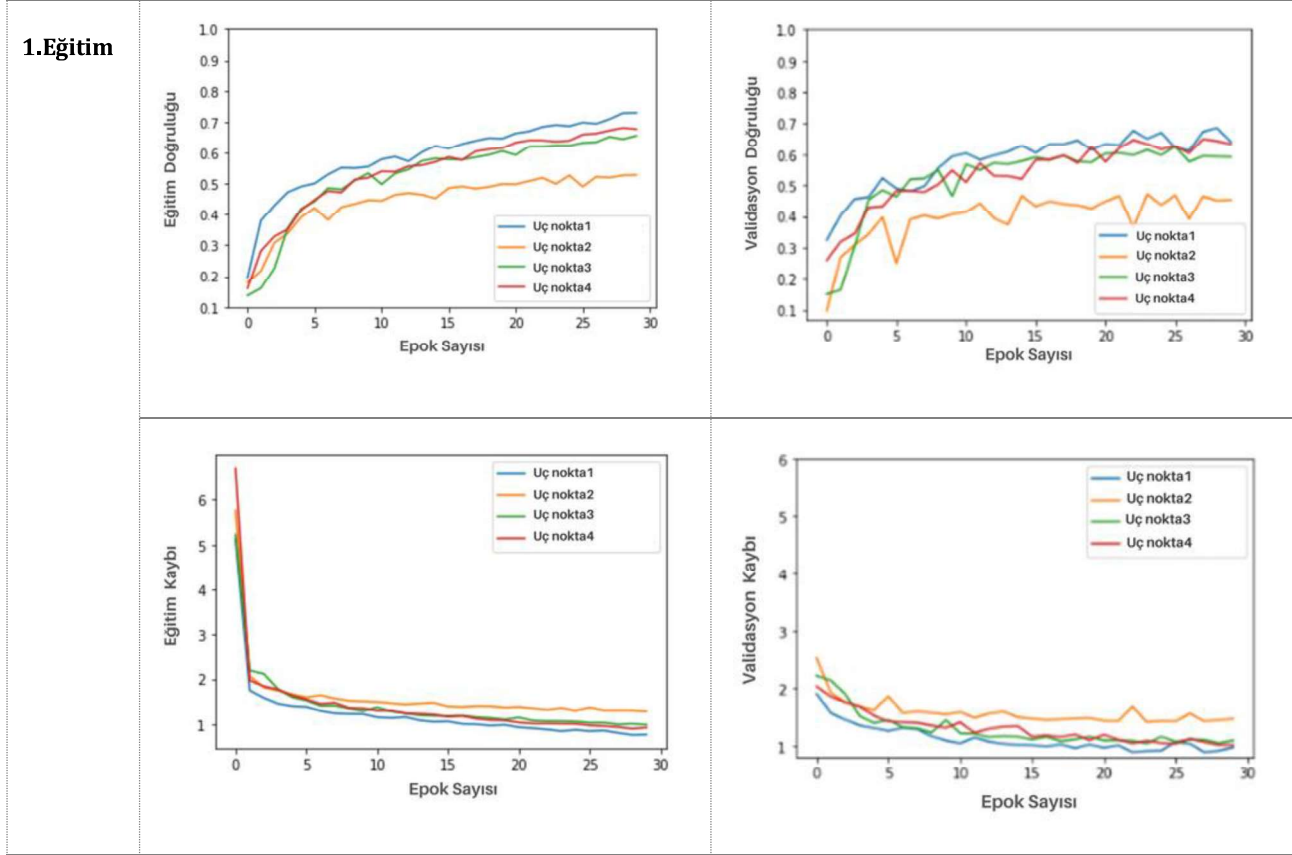
FedBest yaklaşımı PathMNIST veri setinin sınıflandırılmasında FedAvg'ye göre daha hızlı bir yakınsama elde etmiş olsa da üçüncü eğitim sonunda FedAvg yaklaşımı kullanılarak eğitilen küresel model test doğruluğu daha iyi sonuç vermiştir. Küresel model güncellemesinde sadece en iyi test doğruluğuna sahip yerel modelin parametreleri kullanıldığında bir sonraki turlarda uç noktaların yerel model test doğruluğu performanslarında genel anlamda bir önceki tura göre daha iyi sonuçlar elde edilmiştir.

Tablo 3. PathMNIST Veri Seti Deney Sonuçları

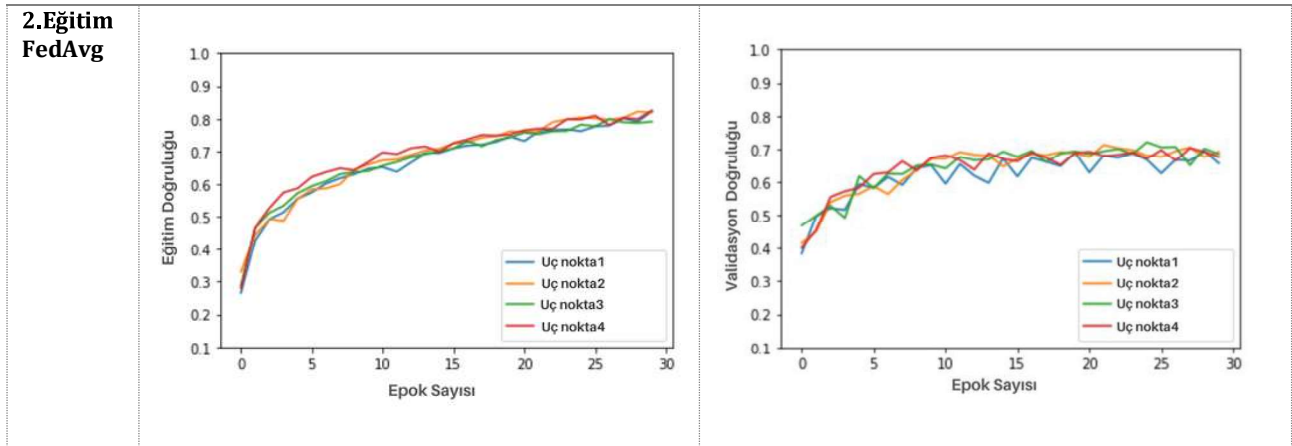
Uç noktalar	CNN Model	2. Eğitim FedAvg	2. Eğitim FedBest	3. Eğitim FedAvg	3. Eğitim FedBest
1	Train acc:	0.7289	0.8222	0.8065	0.9075
	Test acc:	0.6388	0.6656	0.7993	0.6940
2	Train acc:	0.5279	0.8182	0.8121	0.8995
	Test acc:	0.5635	0.6756	0.7609	0.7174
3	Train acc:	0.6554	0.7906	0.8092	0.8903
	Test acc:	0.6739	0.7375	0.6973	0.7023
4	Train acc:	0.6765	0.8252	0.7825	0.8848
	Test acc:	0.6756	0.6823	0.6890	0.7207
Küresel Model	Test acc:	FedAvg : 0.1405 FedBest: 0.6806	0.6221	0.7074	0.7993

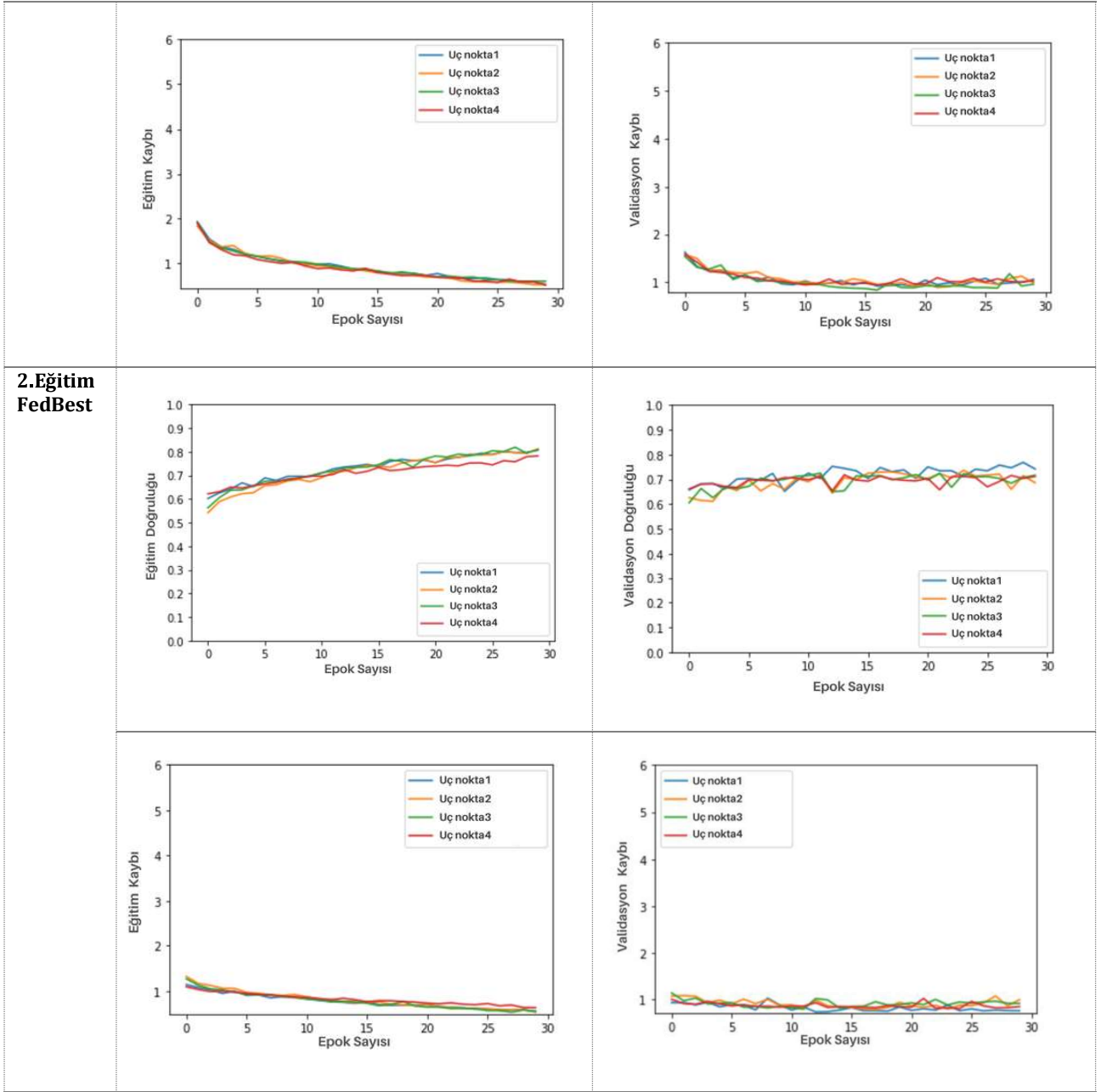
Şekil 9-11'de üç eğitim boyunca alınan uç noktalardaki yerel model performans sonuçlarının grafikleri sunulmaktadır. Grafikler, gösterilen 30 epoch sonunda her uç noktanın yerel modelinin eğitim doğruluğu (train accuracy) ve kayıp değerlerini (train loss), ayrıca validasyon doğruluğu (validation accuracy) ve kayıp değerlerini (validation loss) göstermektedir. Grafikler, her uç noktanın eğitim sürecindeki performansını ve yerel modelin doğruluğu ile kayıp üzerindeki değişimleri takip etmemize olanak sağlamaktadır. Şekil 5'te ilk eğitim turu için merkezi sunucudan her uç noktaya gönderilen başlangıç CNN modelinin, uç noktalardaki yerel verilerle eğitim performans sonuçları grafikleri bulunmaktadır. Şekil 9'daki grafikler incelendiğinde, her bir uç noktanın yerel model performansı eğitim sürecinin başlangıcında eğitim ve validasyon doğruluğu için düşük bir değerde başladığını ancak yerel modellerin 30 epok sonunda eğitim ve validasyon performanslarının arttığını göstermiştir. Eğitim ve validasyon kayıp grafiklerinde ise başlangıçta yüksek bir kayıp değeri vardır, ancak eğitim turu sonuna doğru kayıp değeri düşmüştür. Bu kayıp grafiğindeki düşüş, modelin genel performansını iyileştirdiğini göstermektedir. Şekil 10'da ilk eğitim turunda eğitilen yerel model parametreleri merkezi sunucuda FedAvg ve FedBest yaklaşımları ile güncellenen yeni küresel modelin, uç noktalardaki yerel verilerle ikinci eğitim turu performans sonuçları grafikleri bulunmaktadır. İkinci eğitim turu öncesi merkezi sunucuda güncellenen küresel model, ilk eğitim turundaki başlangıç CNN modelin, eğitim ve validasyon başlangıç doğruluk değerlerine göre her iki yaklaşımda da daha yüksek değer ile başladığı görülmüştür. Özellikle FedBest yaklaşımı, başlangıç değerlerinin FedAvg'ye göre daha yüksek değerde başladığı grafiklerde gösterilmiştir. Bu durum, ikinci eğitim turu için uç

noktalara gönderilen küresel modelin, ilk eğitim turundaki başlangıç CNN modeline kıyasla, federe öğrenme yaklaşımlarıyla güncellenerek küresel model performansının geliştiğini göstermektedir. Şekil 11'de ise ikinci eğitim turunda eğitilen yerel modellerin merkezi sunucuda FedAvg ve FedBest yaklaşımları ile küresel modeli güncellenmesi sonucunda elde edilen yeni küresel modelin her bir uç noktaya tekrar gönderilip eğitime başlandığında, yerel model başlangıç performanslarının bir önceki tura göre daha yüksek başlangıç değerle eğitime başladığı görülmüştür. Üçüncü eğitim turu sonunda, model performansının belirli bir seviyeye ulaştığı ve değişiminin sabitlendiği gözlenmiş ve üçüncü eğitim turunda federe öğrenme süreci tamamlanmıştır.

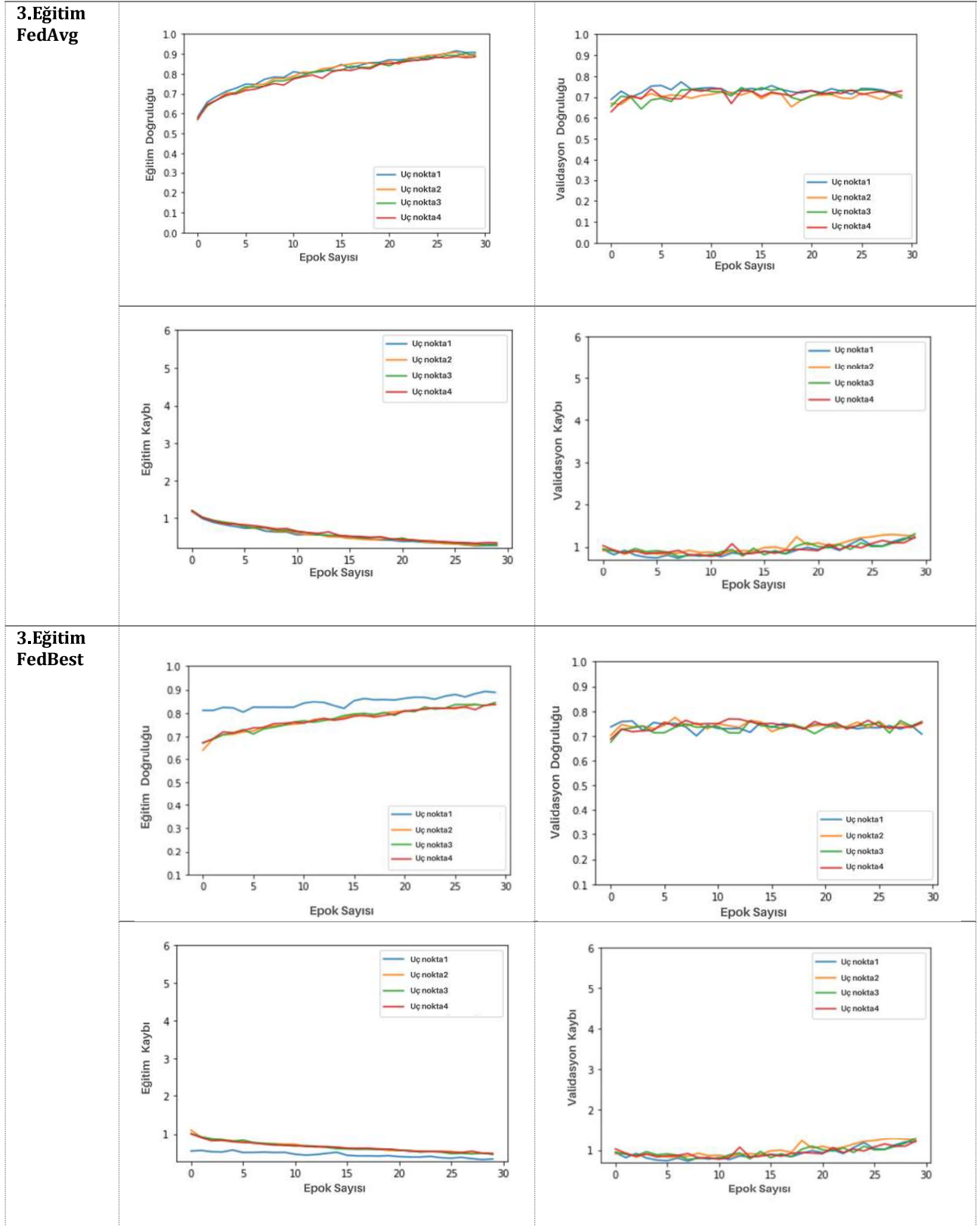


Şekil 9. Birinci eğitim sonu PathMNIST veri seti için bulunan sonuçların grafikleri; x eksenini, epoch sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/Validation accuracy) ve Eğitim/Validasyon kaybı (Train/Validation loss) değerlerini göstermektedir.





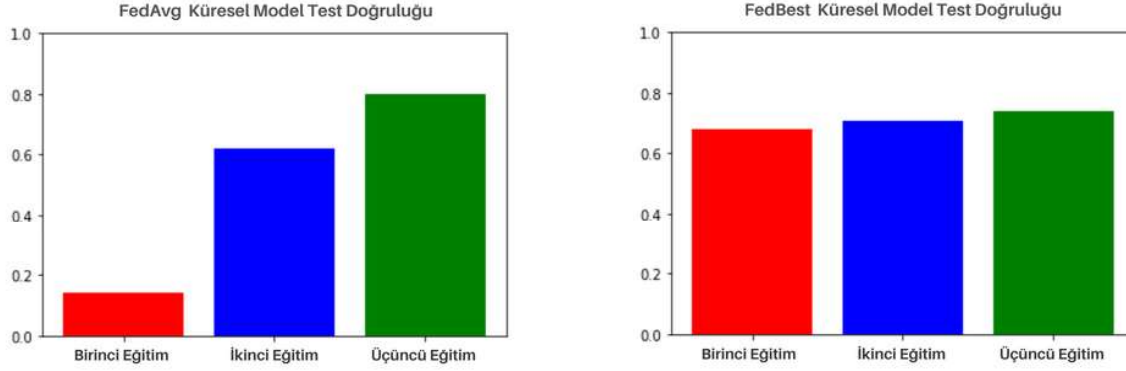
Şekil 10. İkinci eğitim sonu PathMNIST veri seti için bulunan sonuçların grafikleri;x eksenı, epoch sayısını, y eksenı ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/Validation accuracy) ve Eğitim/Validasyon kaybı (Train/Validation loss) değerlerini göstermektedir.



Şekil 11. Üçüncü eğitim sonu PathMNIST veri seti için bulunan sonuçların grafikleri;x eksenini, epoch sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/Validation accuracy) ve Eğitim/Validasyon kayıp (Train/Validation loss) değerlerini göstermektedir.

Şekil 12'de üç eğitim turu sonunda alınan FedAvg ve FedBest küresel model test doğruluğunun sonuçlarının bar grafikleri sunulmuştur. FedAvg yaklaşımı kullanarak güncellenen küresel model performansı ilk eğitim turunda %14,05, ikinci eğitim turunda %62,21, üçüncü eğitim turunda %79,93 değerinde model test doğruluğu performansı göstermiştir. FedBest yaklaşımı kullanılarak güncellenen küresel model performansı ilk eğitim turunda %68,06, ikinci eğitim turunda %70,74, üçüncü eğitim turunda %74,08 değerlerinde model test doğruluğu

performansı göstermiştir. Şekil 12’de görüldüğü üzere FedBest yaklaşımı FedAvg’ye göre model doğruluğu bakımından daha hızlı yakınsama elde etsede FedAvg yaklaşımı ile güncellenen küresel model üçüncü eğitim sonunda daha iyi test doğruluğuna ulaşmıştır.



Şekil 12. PathMNIST veri seti her eğitim turu sonu FedAvg ve FedBest yaklaşımıyla elde edilen küresel model performansları.

3.2.3. DermaMNIST Veri Seti Denev Sonuçları

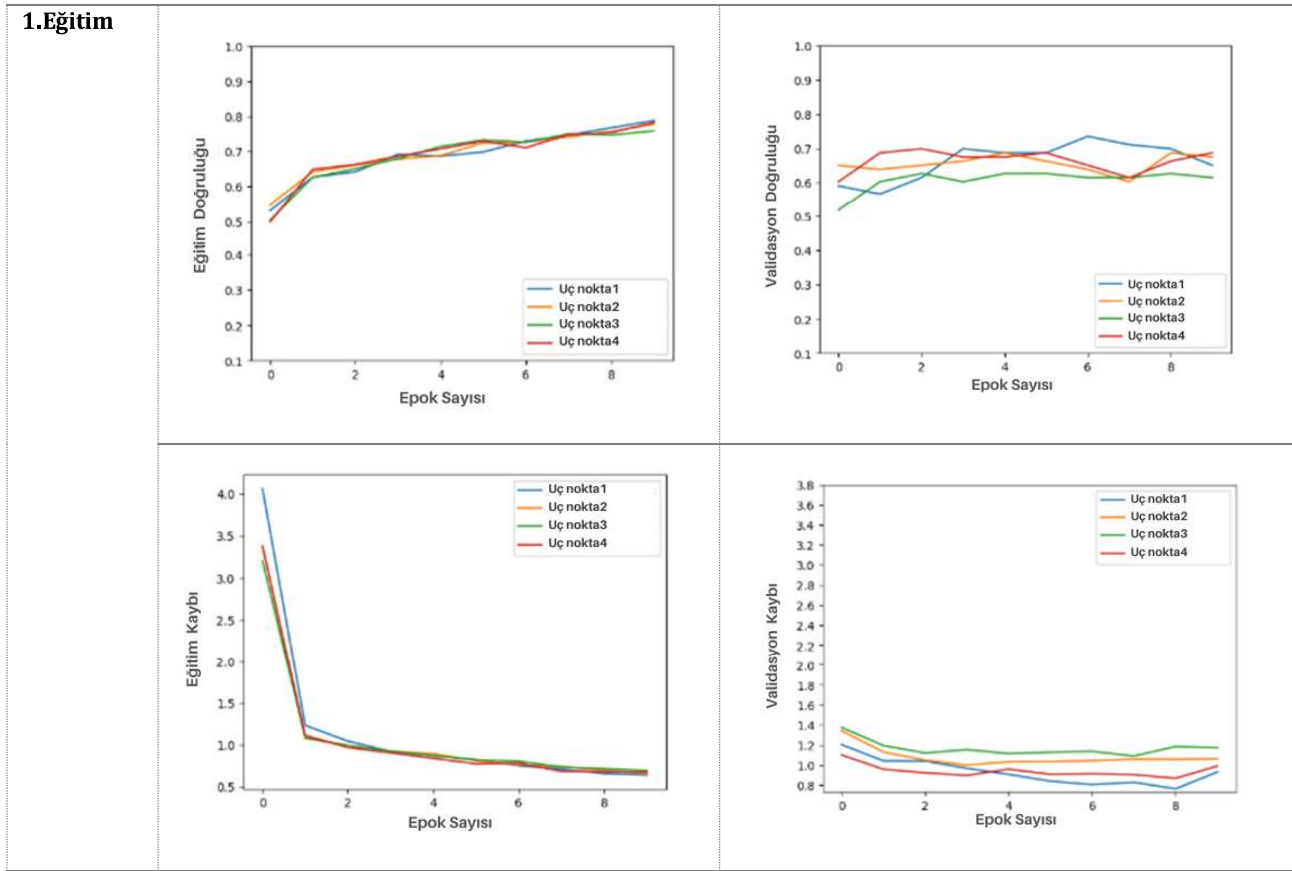
Tablo 4, DermaMNIST veri seti için FedAvg ve FedBest yaklaşımlarına ait üç eğitim turundaki sınıflandırma sonuçlarını sunmaktadır. Tablo 4’teki sonuçlar, tüm uç noktaların yerel model üç eğitim turu sonunda ilk tura göre daha yüksek doğruluk değerlerine ulaştığını göstermektedir. FedAvg yaklaşımı kullanılarak eğitilen küresel modelin test doğruluğunun %71,86 doğruluk verdiğini, FedBest yaklaşımında ise küresel modelin %73,05 test doğruluğu verdiği görülmektedir. Bu değerler, modelin doğruluğunu geliştirmede her iki yaklaşımın etkinliğini göstermektedir. FedAvg ve FedBest yaklaşımlarında, uç noktaların doğruluk değerleri üç eğitim turu gerçekleştikten sonra incelendiğinde; FedAvg yaklaşımında, uç nokta 1 yerel modeli %67,66, uç nokta 2 yerel modeli %69,46, uç nokta 3 yerel modeli %61,08 ve uç nokta 4 yerel modeli %64,67 doğruluk elde etmiştir. FedBest yaklaşımında ise uç nokta 1 yerel modeli %72,46, uç nokta 2 yerel modeli %68,86, uç nokta 3 yerel modeli %59,28 ve uç nokta 4 yerel modeli %70,66 doğruluk elde etmiştir. FedAvg yaklaşımında üç eğitim turu sonunda küresel model performansı %67,07’den %71,86’ya yükselirken, FedBest yaklaşımında ise %68,26’dan %73,05’e yükselmiştir.

FedBest yaklaşımı ile DermaMNIST veri setinin sınıflandırmasında FedAvg’ye göre daha hızlı bir yakınsama elde etmiş ve DermaMNIST veri seti sınıflandırmasında küresel model test doğruluğu için daha iyi sonuç vermiştir. Küresel model güncellemesinde sadece en iyi test doğruluğuna sahip yerel model kullanıldığında bir sonraki turlarda yerel modellerin test doğruluğunda genelde bir önceki tura göre daha iyi sonuçlar elde edilmiştir.

Tablo 4. DermaMNIST Veri Seti Denev Sonuçları

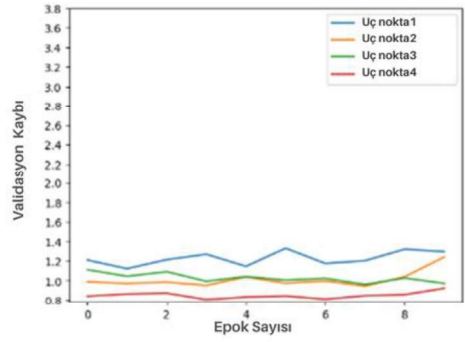
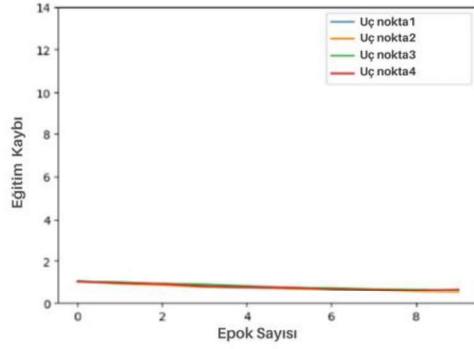
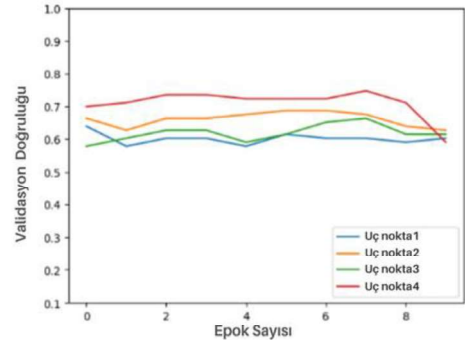
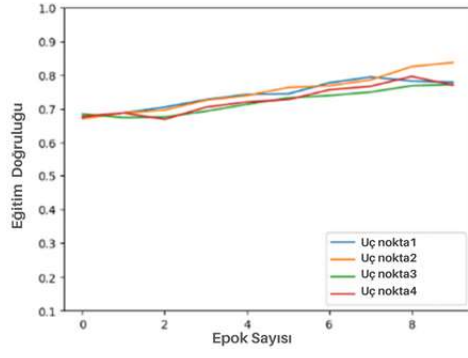
Uç noktalar	CNN Model	2. Eğitim FedAvg	2. Eğitim FedBest	3. Eğitim FedAvg	3. Eğitim FedBest
1	Train acc:	0.7873	0.7787	0.7993	0.7513
	Test acc:	0.6766	0.6766	0.6168	0.6766
2	Train acc:	0.7770	0.8370	0.8268	0.7788
	Test acc:	0.6766	0.6766	0.6407	0.6946
3	Train acc:	0.7581	0.7719	0.8062	0.8316
	Test acc:	0.6766	0.6527	0.6826	0.6108
4	Train acc:	0.7822	0.7702	0.7890	0.7959
	Test acc:	0.7186	0.6228	0.6647	0.6467
Küresel Model	Test acc:	FedAvg: 0.6707 FedBest: 0.6826	0.6946	0.7186	0.7305

Şekil 13-15'te üç eğitim turu boyunca alınan uç noktadaki yerel model performans sonuçlarının grafikleri sunulmaktadır. Grafikler, gösterilen 10 epoch sonunda her uç noktanın yerel modelinin eğitim doğruluğu (train accuracy) ve kayıp değerlerini (train loss), ayrıca validasyon doğruluğu (validation accuracy) ve kayıp değerlerini (validation loss) göstermektedir. Şekil 13'te ilk eğitim turu için merkezi sunucudan her uç noktaya gönderilen başlangıç CNN modelinin, uç noktalardaki yerel verilerle eğitim performans sonuçları grafikleri bulunmaktadır. Şekil 14'te ilk eğitim turunda eğitilen yerel model parametreleri merkezi sunucuda FedAvg ve FedBest yaklaşımları ile güncellenen yeni küresel modelin, uç noktalardaki yerel verilerle ikinci eğitim turu performans sonuçları grafikleri bulunmaktadır. Şekil 15'te ise ikinci eğitim turunda eğitilen yerel modellerin merkezi sunucuda FedAvg ve FedBest yaklaşımları ile küresel modeli güncellenmesi sonucunda elde edilen yeni küresel modelin, uç noktalardaki yerel verilerle üçüncü eğitim turu performans sonuçları grafikleri bulunmaktadır. Üçüncü eğitim turu sonunda, model performansının belirli bir seviyeye ulaştığı ve değişiminin sabitlendiği gözlenmiş ve üçüncü eğitim turunda federe öğrenme süreci tamamlanmıştır.

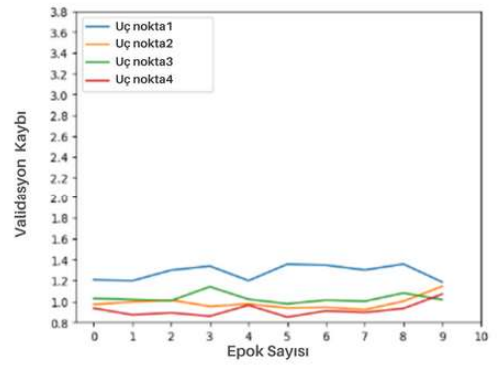
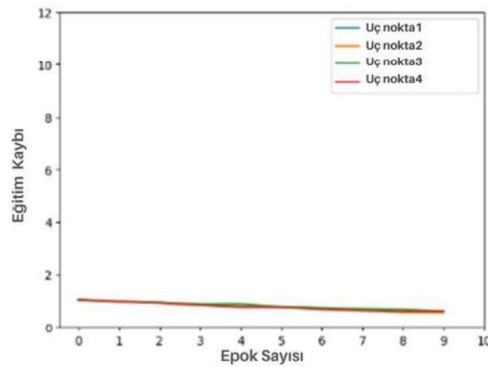
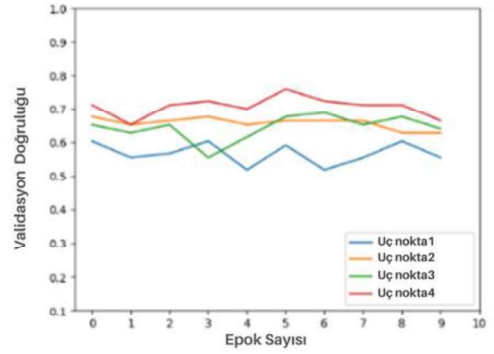
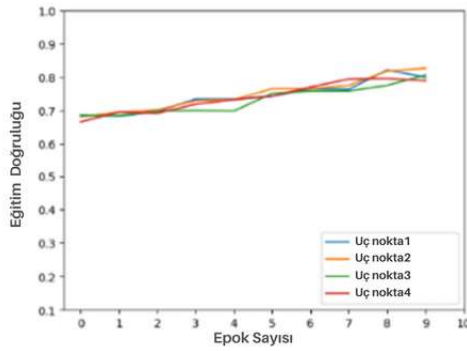


Şekil 13. Birinci eğitim sonu DermaMNIST veri seti deney sonuçları grafikleri; x eksenini, epoch sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/ Validation accuracy) ve Eğitim/Validasyon kayıp (Train/ Validation loss) değerlerini göstermektedir.

2.Eğitim
FedAvg

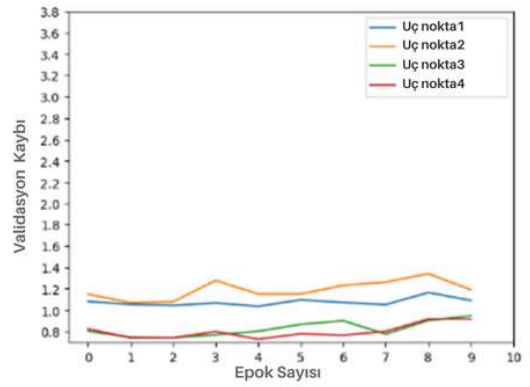
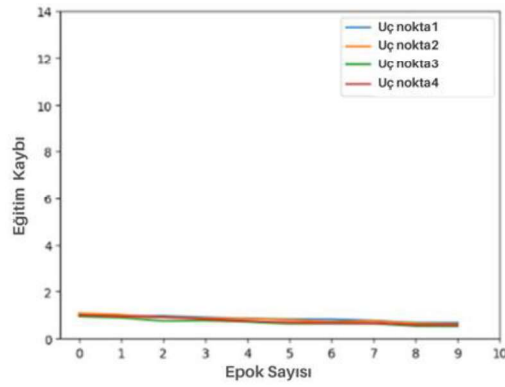
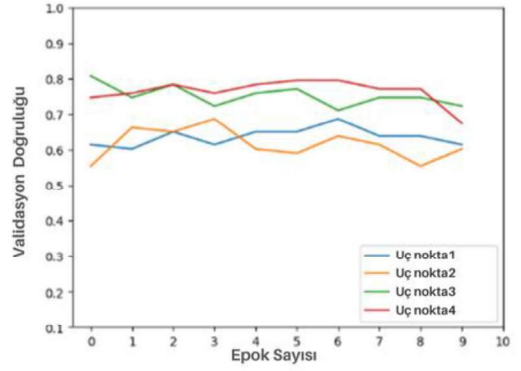
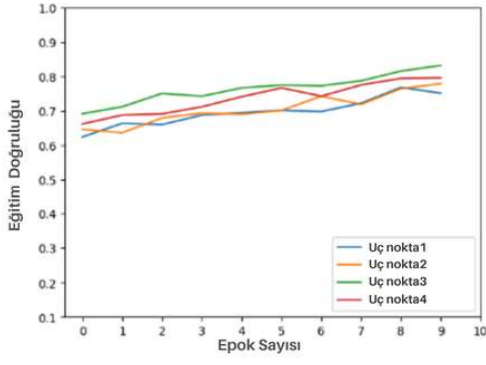


2.Eğitim
FedBest

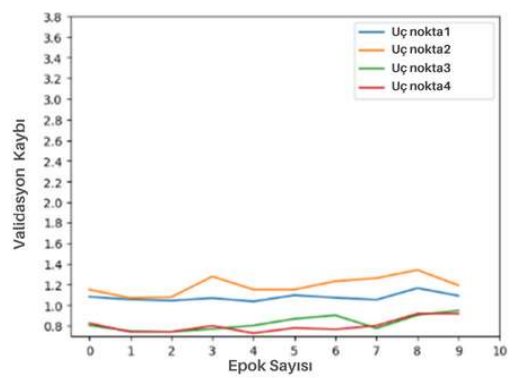
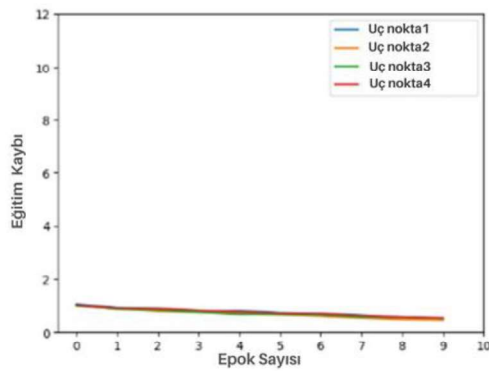
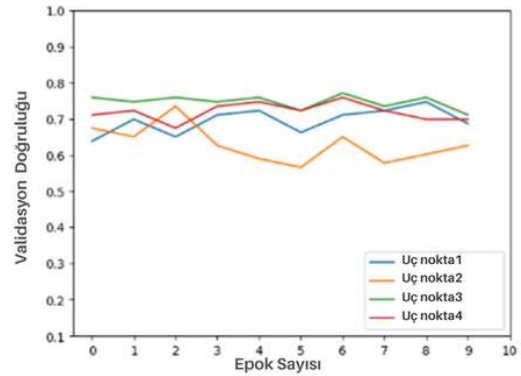
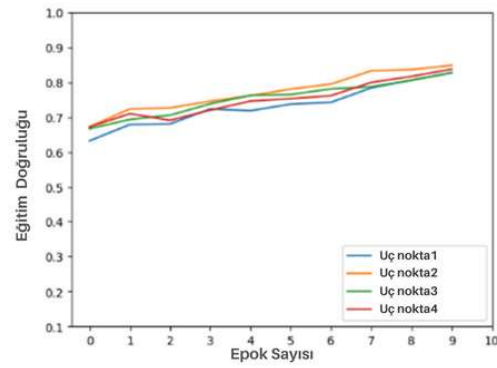


Şekil 14. İkinci eğitim sonu DermaMNIST veri seti deney sonuçları grafikleri; x eksenini, epoch sayısını, y eksenini ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/ Validation accuracy) ve Eğitim/Validasyon kaybı (Train/ Validation loss) değerlerini göstermektedir.

3. Eğitim FedAvg

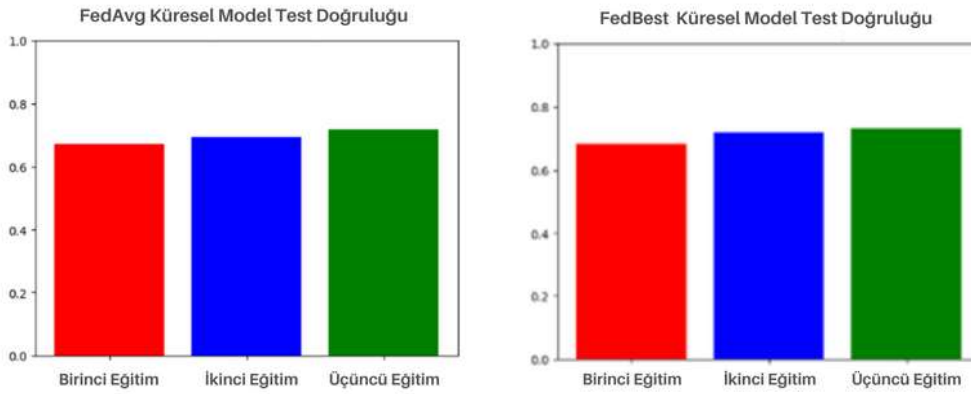


3. Eğitim FedBest



Şekil 15. Üçüncü eğitim sonu DermaMNIST veri seti deney sonuçları grafikleri; x eksen, epoch sayısını, y eksen ise eğitim sırasında elde edilen Eğitim/Validasyon doğruluk (Train/ Validation accuracy) ve Eğitim/Validasyon kayıp (Train/ Validation loss) değerlerini göstermektedir.

Şekil 16'da üç eğitim turu sonunda alınan FedAvg ve FedBest sonuçların grafikleri sunulmuştur. FedAvg yaklaşımı kullanarak güncellenen küresel model performansı ilk eğitim turunda %67,07, ikinci eğitim turunda %69,46, üçüncü eğitim turunda %71,86 değerinde model test doğruluğu performansı göstermiştir. FedBest yaklaşımı kullanılarak güncellenen küresel model performansı ilk eğitim turunda %68,26, ikinci eğitim turunda %71,86, üçüncü eğitim turunda %73,05 değerlerinde model test doğruluğu performansı göstermiştir. Şekil 11'de görüldüğü üzere FedBest yaklaşımı kullanılarak güncellenen küresel model test doğruluğunun FedAvg yaklaşımına göre model test doğruluğu daha hızlı yakınsama elde etmiş ve daha iyi performans göstermiştir.



Şekil 16. DermaMNIST veri seti her eğitim turu sonu FedAvg ve FedBest yaklaşımıyla elde edilen küresel model performansları.

Tablo 5'te, BloodMNIST, PathMNIST ve DermaMNIST tüm veri setleri için merkezi öğrenme yöntemiyle eğitilen CNN modelinin sonuçları, bu üç veri seti için uç noktalara dağıtılan verilerle eğitilen yerel modellerin ilk eğitim turunda elde edilen sonuçlarının ortalaması ve üç eğitim turunun sonunda alınan FedAvg ve FedBest sonuçları özetlenmiştir.

Tablo 5. Kullanılan Veri Setleri ve Deney Sonuçları

Veri Setleri	CNN (Tüm veri)	CNN (Ortalama)	FedAvg	FedBest
BloodMNIST	Test acc: 0.9211	Test acc: 0.7729	Test acc: 0.8451	Test acc: 0.8486
PathMNIST	Test acc: 0.7926	Test acc: 0.6379	Test acc: 0.7993	Test acc: 0.7408
DermaMNIST	Test acc: 0.7365	Test acc: 0.6871	Test acc: 0.7186	Test acc: 0.7305

Tablo 5'e göre üç farklı veri seti üzerinde yapılan deneyler sonucunda, BloodMNIST veri setinin uç noktalara dağıtılmadan önceki tüm eğitim veri seti kullanılarak eğitilen CNN model için test doğruluğu yaklaşık %92,11 olarak hesaplanmıştır. Federe Öğrenme süreci öncesi tüm eğitim veri seti dört uç noktaya rastgele dağıtılmış ve rastgele dağıtılan eğitim verileri ile eğitilen yerel modellerin test doğrulukları hesaplanmış ve bu dört uç nokta yerel model test doğruluklarının sonuçlarının ortalaması alındığında CNN model test doğruluğu ortalaması %77,29 olarak hesaplanmıştır. Federe öğrenme süreci sonunda FedAvg yaklaşımı kullanılarak elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %77,29'dan başlayıp %84,51'e yükselmiştir. FedBest yaklaşımı ile elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %77,29'dan başlayıp %84,86'ya yükselmiştir.

PathMNIST veri setinin uç noktalara dağıtılmadan önceki tüm eğitim veri seti kullanılarak eğitilen CNN model için test doğruluğu yaklaşık %79,26 olarak hesaplanmıştır. Tüm eğitim veri seti dört uç noktaya rastgele dağıtılmış ve rastgele dağıtılan eğitim verileri ile eğitilen yerel modellerin test doğrulukları hesaplanmış ve bu dört uç nokta yerel model test doğruluklarının sonuçlarının ortalaması alındığında CNN model test doğruluğu ortalama %63,79 olarak hesaplanmıştır. Federe öğrenme süreci sonunda FedAvg yaklaşımı kullanılarak elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %63,79'dan başlayıp %79,93'e yükselmiştir. FedBest yaklaşımı ile elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %63,79'dan başlayıp %74,08'e yükselmiştir.

DermaMNIST veri setinin uç noktalara dağıtılmadan önceki tüm eğitim veri seti kullanılarak eğitilen CNN model için test doğruluğu yaklaşık %73,65 olarak hesaplanmıştır. Tüm eğitim veri seti dört uç noktaya rastgele dağıtılmış ve rastgele dağıtılan eğitim verileri ile eğitilen yerel modellerin test doğrulukları hesaplanmış ve bu dört uç nokta yerel model test doğruluklarının sonuçlarının ortalaması alındığında CNN model test doğruluğu ortalama %68,71 olarak hesaplanmıştır. Federe öğrenme süreci sonunda FedAvg yaklaşımı kullanılarak elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %68,71'den başlayıp %71,86'ya yükselmiştir. FedBest yaklaşımı ile elde edilen küresel modelin test doğruluğu CNN ortalama değer olan %68,71'den başlayıp %73,05'e yükselmiştir.

4. Tartışma ve Sonuç

Bu çalışmada, CNN derin öğrenme algoritması kullanarak medikal veri sınıflandırması için federe öğrenme yaklaşımlarından FedAvg ve önerilen FedBest algoritmalarının uygulaması araştırılmıştır. FedAvg algoritması, uç noktalardan gelen yerel modellerin parametrelerinin ortalamalarını alarak küresel modeli güncellemeyi içeren bir temel yaklaşım olarak kullanılmıştır. Bu yaklaşım, verileri uç noktalarda tutarak her uç noktadan öğrenilen bilgileri birleştirmemize olanak sağlamıştır. Federe öğrenme sürecini daha da geliştirmek için FedBest algoritması kullanıma sunulmuştur. Bu algoritma, model güncellemesi için en yüksek model test doğruluğuna sahip uç noktanın yerel model parametrelerini kullanan bir yaklaşımdır. Sonuçlar, FedBest'in FedAvg'ye kıyasla BloodMNIST ve DermaMNIST için daha hızlı yakınsama ve yüksek doğruluk elde ettiğini ve önerilen algoritmanın medikal veri sınıflandırma için etkinliğini göstermiştir. PathMNIST veri seti için ise FedBest daha hızlı yakınsama sağlasada FedAvg yaklaşımı üç eğitim sonunda FedBest 'e göre küresel model test doğruluğunda daha iyi sonuç vermiştir.

Sonuçlar, en yüksek test doğruluğuna sahip yerel modelin katkıları daha fazla dikkate alındığında, küresel modelin daha hızlı yakınsayacağını ve daha iyi sınıflandırma yapabileceğini göstermiştir. FedBest'te her eğitim turu öncesi en iyi performansa sahip olan uç noktanın bilgisinin yanı sıra çeşitliliğin sağlanması bakımından bazı turlarda diğer uç noktalardan da faydalanılabilir. Bu nedenle, FedAvg ve FedBest gibi algoritmaların performansı, kullanıldıkları senaryoya, veri dağılımına, cihaz sayısına ve diğer faktörlere bağlı olarak değişebilir. Genel olarak, bu çalışma, medikal veri sınıflandırması için federe öğrenmenin potansiyelini ortaya koymuş ve sağlık hizmetleri uygulamalarında mahremiyeti koruma tekniklerinin önemini vurgulamıştır. Gelecekteki araştırmalar, federe öğrenme algoritmalarını daha da optimize etmeye ve bunların diğer medikal veri kümelerine ve sınıflandırma görevlerine uygulanabilirliğini keşfetmeye odaklanabilir.

Kaynakça

- [1] Büyüknacar, Yavuz Canbay-Yaşar. "Federe öğrenme ve veri mahremiyeti." Akademik Bilişim Konferansı. 2019.
- [2] Zhang, W., Zhou, T., Lu, Q., Wang, X., Zhu, C., Sun, H., ... & Wang, F. Y. (2021). Dynamic-fusion-based federated learning for COVID-19 detection. *IEEE Internet of Things Journal*, 8(21), 15884-15891.
- [3] Odera, David. "Federated learning and differential privacy in clinical health: Extensive survey." (2023).
- [4] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2019). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4* (pp. 92-104). Springer International Publishing.

- [5] Yoo, J. H., Jeong, H., Lee, J., & Chung, T. M. (2022). Open problems in medical federated learning. *International Journal of Web Information Systems*, 18(2/3), 77-99.
- [6] Wibawa, F., Catak, F. O., Sarp, S., & Kuzlu, M. (2022). BFV-Based Homomorphic Encryption for Privacy-Preserving CNN Models. *Cryptography*, 6(3), 34.
- [7] Farhad, Arshad, Sandra I. Woolley, and Peter Andras. "A Preliminary Scoping Study of Federated Learning for the Internet of Medical Things." *Studies in Health Technology and Informatics* 281 (2021): 504-505.
- [8] Zhao, Leiyang, and Jianjun Huang. "A distribution information sharing federated learning approach for medical image data." *Complex & Intelligent Systems* (2023): 1-12.
- [9] Sohan, Md Fahimuzzman, and Anas Basalamah. "A Systematic Review on Federated Learning in Medical Image Analysis." *IEEE Access* (2023).
- [10] Yi, Hang, Tongxuan Bie, and Tongjiang Yan. "Framework Construction of an Adversarial Federated Transfer Learning Classifier." *arXiv preprint arXiv:2211.04734* (2022).
- [11] Nazir, Sajid, and Mohammad Kaleem. "Federated Learning for Medical Image Analysis with Deep Neural Networks." *Diagnostics* 13.9 (2023): 1532.
- [12] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA, Aarti Singh and Xiaojin (Jerry) Zhu, Eds. 2017, vol. 54 of Proceedings of Machine Learning Research*, pp. 1273–1282, PMLR.
- [13] Reza Shokri and Vitaly Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA, 2015, CCS '15*, pp. 1310–1321, ACM.
- [14] Martin Abadi, Andy Chu, Ian Goodfellow, Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *23rd ACM Conference on Computer and Communications Security (ACM CCS)*, 2016.
- [15] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.
- [16] Behavioral Informatics & Interaction Computation Lab (BIIC) <https://biic.ee.nthu.edu.tw/blog-detail.php?id=2> (Erişim Tarihi: 14.06.2023).
- [17] Lee, Hagyeong, and Jongwoo Song. "Introduction to convolutional neural network using Keras; an understanding from a statistician." *Communications for Statistical Applications and Methods* 26.6 (2019): 591-610.
- [18] Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., ... & Ni, B. (2023). MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10(1), 41.