REPUBLIC of TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
SOFTWARE ENGINEERING DEPARTMENT

THE UTILISATION of FEDERATED LEARNING IN THE CLASSIFICATION of MAGNETIC RESONANCE IMAGES

BURAK UZDUR

M.Sc.

ADANA 2025

REPUBLIC of TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
SOFTWARE ENGINEERING DEPARTMENT

THE UTILISATION of FEDERATED LEARNING IN THE CLASSIFICATION of MAGNETIC RESONANCE IMAGES

BURAK UZDUR

M.Sc.

THESIS ADVISOR
ASST.PROF.DR. ERKUT TEKELİ

SECOND ADVISOR
PROF.DR. TURGAY İBRİKÇİ

ADANA, 2025

DECLARATION OF CONFORMITY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

..../..../20..

[Signature]

Burak UZDUR

ÖZET

MANYETİK REZONANS GÖRÜNTÜLERİNİN SINIFLANDIRILMASINDA FEDERASYONLU ÖĞRENMENİN KULLANIMI

Burak UZDUR
Yüksek Lisans, Yazılım MühendisliğiAnabilim Dalı
Danışman: Dr.Öğr.Üyesi Erkut TEKELİ
II. Danışman: Prof.Dr. Turgay İBRİKÇİ
Ağustos 2025, 70 sayfa

Karaciğer tümörlerinin erken ve doğru tanısı, hasta sonuçlarını iyileştirmek ve uygun tedavi stratejilerini belirlemek açısından çok önemli bir rol oynamaktadır. Son yıllarda, makine öğrenimi teknikleri, özellikle derin öğrenme, tıbbi görüntü analizinde tanı doğruluğunu artırmada önemli bir potansiyel göstermektedir. Ancak, veri gizliliği ve güvenliği gibi zorluklar, merkezi tıbbi veri setlerini kullanarak derin öğrenme modellerinin eğitilmesinde önemli engeller oluşturmuştur. Federated Learning (FL), hasta verilerini paylaşmadan, farklı kurumlar arasında ortak model eğitimi yapılmasına olanak sağlayan umut verici bir çözüm olarak ortaya çıkmıştır.

Bu çalışma, Karaciğer tümörü sınıflandırması için Magnetic Resonance Imaging taramaları kullanarak Federated Learning'in yeni bir uygulamasını sunmaktadır. Önerilen model, FL çerçevesi içinde EfficientNetB0 mimarisini kullanmaktadır. FL modelinin performansı, CNN, EfficientNet, MobileNetV3, ResNet50 ve VGG16 gibi geleneksel derin öğrenme mimarileriyle karşılaştırılmıştır. Deneysel sonuçlar, FL tabanlı EfficientNetB0 modelinin, %93,75 doğruluk, %99,71 hassasiyet, %87,79 duyarlılık, %93,37 F1 skoru ve %99,19 ROC-AUC ile üstün performans gösterdiğini ortaya koymaktadır. Bu sonuçlar, FL'nin veri gizliliğini korurken yüksek sınıflandırma performansı sağlama potansiyelini vurgulamaktadır.

Sonuç olarak, bu çalışma, sağlık sektöründe gizlilik koruyan yapay zeka çözümlerinin giderek daha kritik hale geldiği bir dönemde, Federated Learning'in artan rolünü ortaya koymaktadır. Gelecekteki çalışmalar, FL'nin daha ileri düzey mimariler ve daha büyük, daha çeşitli veri setleriyle entegrasyonunu keşfederek modelin genelleştirilebilirliğini ve dayanıklılığını daha da artırmayı amaçlayabilir.

Anahtar Kelimeler: Federated Learning, FedAvg, Karaciğer Tümörünün Sınıflandırılması, Yapay Sinir Ağları, MRI

ABSTRACT

THE UTILISATION OF FEDERATED LEARNING IN THE CLASSIFICATION OF MAGNETIC RESONANCE IMAGES

Burak UZDUR
M.Sc., Department of Software Engineering

Supervisor: Asst. Prof. Dr. Erkut TEKELİ
Co-Supervisor: Prof.Dr. Turgay İBRİKÇİ
August 2025, 70 pages

The early and accurate diagnosis of liver tumors plays a crucial role in improving patient outcomes and determining appropriate treatment strategies. In recent years, machine learning techniques, particularly deep learning, have shown significant potential in enhancing diagnostic accuracy through medical image analysis. However, challenges such as data privacy and security concerns pose significant barriers to training deep learning models using centralized medical datasets. Federated Learning (FL) has emerged as a promising solution, enabling collaborative model training across multiple institutions without the need for sharing sensitive patient data.

This study presents a novel application of Federated Learning for liver tumor classification using Magnetic Resonance Imaging scans. The proposed model leverages the EfficientNetB0 architecture within the FL framework. The performance of the FL model is compared to conventional deep learning architectures, including CNN, EfficientNet, MobileNetV3, ResNet50, and VGG16. Experimental results demonstrate that the FL-based EfficientNetB0 model achieves superior performance, with an accuracy of 93.75%, precision of 99.71%, recall of 87.79%, F1-score of 93.37%, and ROC-AUC of 99.19%. These results highlight the potential of FL to provide high classification performance while preserving data privacy.

In conclusion, this study underscores the growing role of FL in the healthcare sector, where privacy-preserving AI solutions are becoming increasingly critical. Future work can explore the integration of FL with more advanced architectures and larger, more diverse datasets to further enhance model generalizability and robustness.

Keywords: Federated Learning, FedAvg, Liver Tumor Classification, Convolutional Neural Networks, MRI

Dedication

Throughout this research process, I would like to express my deepest gratitude to my first supervisor, Asst. Prof. Dr. Erkut TEKELİ, and my second supervisor, Prof. Dr. Turgay İBRİKÇİ, who guided me with their valuable knowledge and experience from the very beginning to the final stage of my thesis work, and who always provided me with their support and guidance. I would also like to thank Harun Ur Rashid and Prof. Dr. R. Geetha for their invaluable contributions to the data analysis process of my thesis. I would like to express my infinite gratitude to Bilge KIRAN, who has been by my side throughout my academic journey, motivating me with her patience and support, and to my family, who have empowered me at every stage of my educational life with their unconditional love and presence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
CFL	: Clustered Federated Learning
CNN	: Convolutional Neural Network
CT	: Computed Tomography
DL	: Deep Learning
FedAvg	: Federated Averaging
FedProx	: Federated Proximal
FL	: Federated Learning
FTL	: Federated Transfer Learning
HCC	: Hepatocellular Carcinoma
IID	: Independent and Identically Distributed
KD	: Knowledge Distillation
ML	: Machine Learning
MRI	: Magnetic Resonance Imaging
VFL	: Vertical Federated Learning

1. INTRODUCTION

The liver is the largest internal organ in the human body and plays a vital role in maintaining essential bodily functions. It is involved in crucial metabolic activities such as producing digestive enzymes, storing vitamins and iron, synthesizing cholesterol and triglycerides, generating blood clotting factors, regenerating liver tissues, and most importantly, detoxifying harmful substances from the body (Singh et al., 2016).

Today, the early diagnosis and accurate evaluation of various diseases play a critical role in their treatment process. Therefore, leveraging current technological advancements and medical data in the diagnostic and classification stages of diseases offers significant benefits in the field of healthcare (Bucak et al., 2010). In this context, due to the liver's vital functions, the early detection and accurate classification of liver-related diseases necessitate the effective use of medical technologies.

Magnetic Resonance Imaging (MRI) is one of the most commonly used imaging techniques in the diagnosis of liver diseases. Its widespread use stems from its ability to provide high-resolution and detailed information about both pathological and physiological structures. With the integration of advanced MRI techniques into clinical practice in recent years, the early and accurate detection of liver-related disorders has become more feasible, further elevating the significance of MRI in the diagnosis of hepatic conditions (Çaviş et al., 2024). The analysis of such detailed imaging data has recently found great potential in Deep Learning (DL) methods, which have emerged prominently in the field of medical imaging.

DL represents the most advanced approach in Machine Learning (ML). Its remarkable success in various pattern recognition applications has generated significant enthusiasm and high expectations regarding its potential to drive transformative advancements in healthcare (Chan et al., 2020). DL continues to play an increasingly vital role in radiology. Currently, it assists in the detection, classification, segmentation, and monitoring of diseases in medical imaging. Beyond automating time-consuming tasks for radiologists, it has also been demonstrated to reduce interobserver variability in diagnostic interpretation (Kwak et al., 2023). These

advancements have further emphasized the importance of DL in the early diagnosis of serious and widespread liver diseases, such as Hepatocellular Carcinoma (HCC).

HCC represents a significant public health concern and is projected to impact over one million individuals worldwide by 2025. HCC and colorectal metastases are the most common forms of primary and metastatic liver cancer (Llovet et al., 2021). Sub-Saharan Africa and Eastern Asia exhibit a higher incidence of HCC compared to other regions, primarily due to the prevalence of hepatitis B and C virus infections. According to numerous clinical guidelines, dynamic Computed Tomography (CT) is one of the primary modalities for HCC diagnosis (Heimbach et al., 2018), (Shao et al., 2021). However, the interpretation of CT images on a slice-by-slice basis is a complex and time-intensive process that requires the specialized expertise of hepatologists, surgeons, oncologists, or radiologists (Chen et al., 2022). Physicians can alleviate their workload and minimize subjective errors by leveraging computer vision techniques for feature extraction and pattern recognition (Song et al., 2021). Automating liver tumor identification through computer vision techniques improves the efficiency and accuracy of liver cancer diagnosis. This approach enables early detection, optimizes treatment strategies, and enhances patient survival rates (Song et al., 2020).

In recent years, Artificial Intelligence (AI) has emerged as a transformative tool across various domains, including healthcare. In the medical field, AI-driven systems have been employed to support clinical decision-making, predict disease outcomes, and assist in diagnostic imaging. Particularly within computer vision, AI has enabled automated analysis of medical images, facilitating faster and more consistent interpretation. These advancements have paved the way for the integration of more sophisticated techniques, such as DL, in medical imaging applications.

DL based models have significantly improved disease diagnosis in medical imaging by achieving high accuracy rates. Pre-trained models such as MobileNetV3, EfficientNet, VGG16, and ResNet50 offer powerful feature extraction and transfer learning capabilities, as they are trained on large-scale datasets. However, achieving high accuracy with these models requires access to large and diverse datasets. In medical imaging, this poses a significant

challenge, as data sharing is restricted due to patient privacy concerns, legal regulations, and data security issues**.**

Table 1. Overview of commonly used models in the literature, including their functional characteristics, advantages, and limitations.

Model	Functional Characteristics	Advantages	Limitations
CNN	Learns spatial hierarchies from input images through convolutional and pooling layers	High accuracy in image classification tasks, easy to implement	Requires large labeled datasets, prone to overfitting
EfficientNet	Compound scaling of depth, width, and resolution for efficient architecture	High accuracy with fewer parameters and lower computational cost	Performance can degrade with very small or imbalanced datasets
MobileNetV3	Lightweight model optimized for mobile and embedded devices	Fast inference, low memory usage, suitable for real-time applications	May sacrifice accuracy for efficiency
ResNet50	Deep residual learning with skip connections to avoid vanishing gradients	Very effective for deep image recognition tasks, avoids degradation problem in deep networks	Higher computational requirements, slower training
VGG16	Uses multiple 3x3 convolutional layers followed by max pooling	Simplicity, strong baseline for transfer learning	Large model size, computationally expensive, no residual connections

Today, data privacy and security have become major challenges in the development of AI-based models, particularly in the healthcare domain. Data sharing across hospitals and medical institutions is often limited due to ethical and legal constraints. At this point, Federated Learning (FL) emerges as an innovative and privacy-preserving learning approach that enables local training of models across decentralized sources without transferring the data to a central server. FL constructs a global model by aggregating only the model weights centrally, thus ensuring patient data confidentiality while promoting collaborative learning across institutions. In this study, to highlight the potential of FL in the healthcare field, a performance comparison was conducted between the FL model and widely-used DL models such as Convolutional Neural Network (CNN), EfficientNet, MobileNetV3, ResNet50, and VGG16 in the classification of liver tumors using MRI images. This allows for a

comprehensive evaluation of the advantages FL offers over traditional methods, both in terms of classification accuracy and data security.

The use of centralized datasets in medical imaging is significantly limited due to concerns related to privacy and regulatory compliance. FL offers an effective solution by enabling institutions to collaboratively train models without sharing raw data. This decentralized framework ensures the protection of patient confidentiality and facilitates adherence to legal requirements, making FL particularly advantageous in the healthcare domain. Figure 1 demonstrates the implementation of the FL approach within a medical context. This study focuses on the application of global FL algorithms to enhance diagnostic accuracy while preserving data privacy. Furthermore, it emphasizes the clinical relevance of MRI imaging in the diagnosis of HCCs, thereby highlighting both the scientific and practical value of the proposed approach.

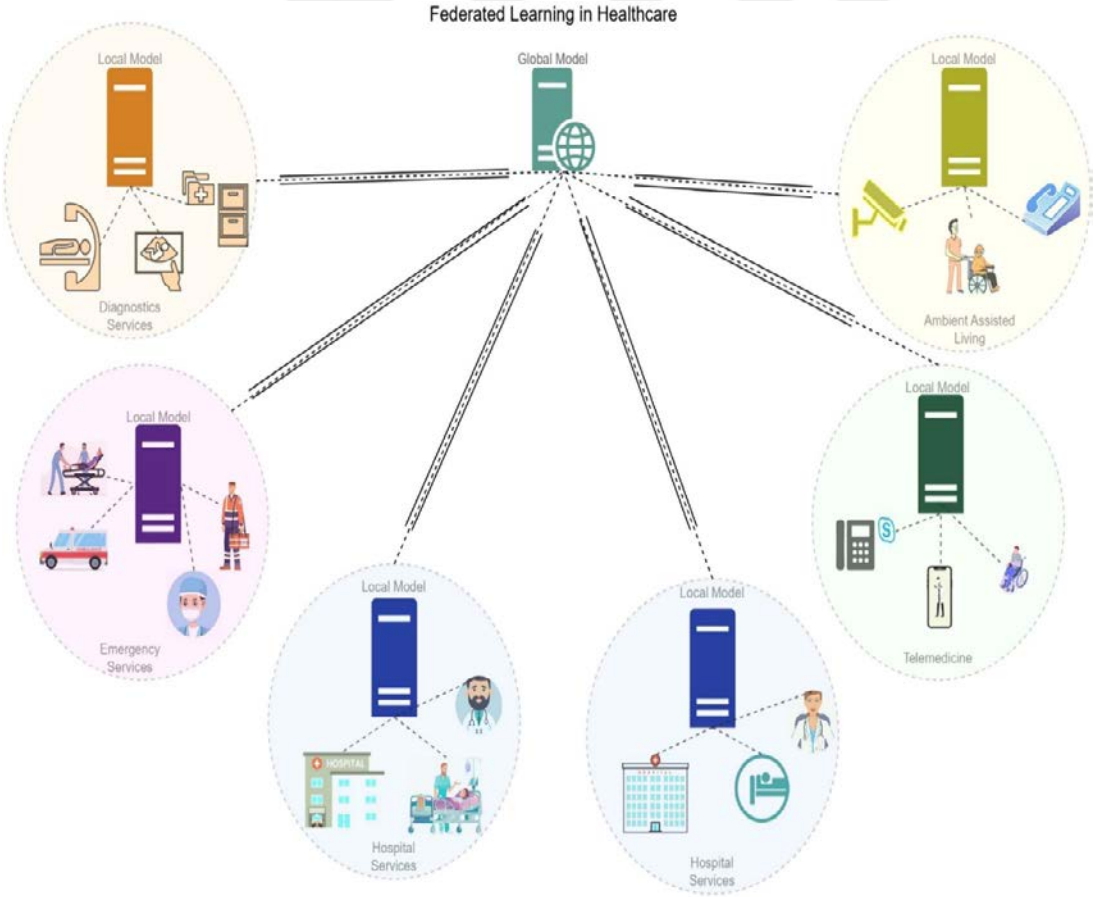

Figure 1. FL technology in the healthcare domain. (Srinivasu et al., 2024)

In summary, the limitations of centralized datasets in medical imaging, particularly due to privacy and regulatory concerns, have prompted the need for alternative approaches. FL addresses this challenge by enabling decentralized training of models, allowing institutions to collaborate while ensuring that raw patient data remains confidential. This approach not only complies with legal and ethical requirements but also enhances the overall utility of medical data. By utilizing FL, this study aims to improve diagnostic accuracy, specifically in the detection of HCCs, through the use of MRI images. The proposed solution underscores the clinical importance of MRI in liver disease diagnosis while demonstrating the potential of FL to offer both scientific advancements and practical solutions to pressing healthcare challenges.

2. THEORETICAL FOUNDATIONS AND SOURCE RESEARCH

In the realm of medical imaging, the application of ML and DL techniques has led to significant progress, particularly in the classification and diagnosis of liver tumors using MRI images. This section examines relevant studies in the field, emphasizing key approaches, findings, and their connection to the present research.

Table 2. Summarization for Related Works

Authors	Year	Number of patients/images	Data set description	Accuracy/Dice	Methodology
McMahan et al.	2017		Mobile device datasets; includes 4 datasets with 5 NN models		FL using the FedAvg algorithm
Roth et al.	2020	715.000	Mammography (BIRADS, multi-institutional)	6.3 (avg.), 45.8 (gen.)	FL for breast density classification across 7 clinical sites
Liu et al.	2020	15.282	Chest X-ray images for COVID-19 detection	ResNet18: 96.15% (train), 91.26% (test)	FL with MobileNet, ResNet18, MobileNetV2, COVID-Net over 100 rounds
Feki et al.	2021	108	Chest X-ray images from multiple institutions for COVID-19	Acc(%): 93,57	FL framework addressing non-IID and imbalanced data across 2 architectures
Kumar et al.	2021	34.006	CT images from multiple hospitals (COVID-19 detection)	Acc(%): 98,68	FL with Capsule Network and Blockchain integration; data normalization for heterogeneity
Atef et al.	2022	585	BraTS2021 dataset; MRI data (FLAIR, T1w, T1wCE, T2w) for brain tumor patients	Acc(%): 96,71	FL using EfficientNet-B3 via OpenFL across 10 institutions
Bernecker et al.	2022	428	CT and MRI images from 6 different public datasets	Dice: 96.1	FedNorm and FedNorm+ algorithms for FL-based segmentation

Table 3. Summarization for Related Works (continue)

Mahlool et al.	2022	253 (BT-small-2c), 3, 264 (BT-large-3c)	BT-small-2c: MRI.BT-large-3c: MRI (3 tumor types, 500 healthy).	BT-small-2c Acc(%): 82, BT-large-3c Acc(%): 96	FL for brain tumor diagnosis
Qayyum et al.	2022	361	X-ray and ultrasound datasets for COVID-19 diagnosis	F1-score improvement: +16% (X-ray), +11% (ultrasound)	Clustered Federated Learning (CFL) on edge devices
Islam et al.	2023	2.309	MRI datasets for brain tumor classification	Acc(%): 91.05	FL with CNN models and ensemble learning; validated on a larger dataset
Trivedi et al.	2023	576	MRI	Acc(%): 99.59	AlexNet-based FL tested on IID Liver dataset with lightweight CNNs
Chai et al.	2024	733	Gene expression data from TCGA and GEO (cross-institutional, anonymized)	Acc(%): 54.2	FL with AdFed, DeepSurv-based survival prediction
Lusnig et al.	2024	41 patients / 4,400 images	41 whole-slide images (JPEG2000), divided into 1024×1024-pixel patches; balanced dataset with 1100 images per stage	Acc(%): ~90	FL applied to high-resolution histopathology images; images categorized into transplant-suitable vs. unsuitable; HQNN was used classification
Zhou et al.	2024	3.260	Brain tumor detection using MRI	Acc(%): 83.12	EfficientNet-B0 + FedAvg; comparative model analysis, emphasis on data heterogeneity
Albalawi et al.	2024	7.023	Brain tumor classification using MRI	Acc(%): 98.00	FL + CNN (modified VGG16) + Transfer Learning; automatic classification with privacy protection
Gohari et al.	2024	3.064	MRI-based brain tumour classification	Acc(%): 94.38	FL + Knowledge Distillation (FedBrain-Distill); architecture independent, communication friendly solution

Table 4. Summarization for Related Works (continue)

Palas et al.	2024		CT scans; 4 classes: adenocarcinoma, large cell carcinoma, squamous cell carcinoma, normal tissues	Acc**(%)**: 93.92	FL + MobileNetV2; privacy-preserving distributed learning, multiple CNN architectures compared
Selvakanmani et al.	2024	15.500	Mammography and MRO images obtained from 3 different institutions	Acc(%): 98.8	FL + Transfer Learning + Domain Adversarial Training; ResNet-based, domain-invariant feature extraction provided
Shankar et al.	2025		Multi-modal dataset (CT, MRI, ultrasound + lab values)	Acc(%): 79.05	Prediction of liver disease using FL from imaging and clinical data
Sivakumar et al.	2025	7.023	Kaggle Brain Tumor MRI; 4 sınıf: glioma, meningioma, pituitary tumor, no tumor	Acc(%): 97.19	Hybrid FL (FedAvg + FedProx); CNN-based, local training for data privacy and heterogeneity
Tölle et al.	2025	8,104 (cardiac CT images)	Partially labeled cardiac CT datasets from 8 hospitals in a federated setting		Two-step semi-supervised learning: task-specific CNNs generate predictions on unlabeled data; a transformer model trained with label-specific attention heads; learns from partial labels across federated sites

McMahan et al. (2017) highlighted the limitations of conventional centralized data collection methods, particularly in environments involving sensitive and decentralized data such as mobile devices. They proposed a novel FL approach that performs model training locally on devices while only sharing model updates with a central server, thus addressing privacy and communication efficiency challenges.

A key contribution of the study is the implementation of the FedAvg (Federated Averaging) algorithm, tested across five neural network architectures and four datasets. The results demonstrated that FL remains effective even under non-IID (Independent and Identically Distributed) and unbalanced data distributions, conditions commonly encountered in real-

world applications. Furthermore, the study reported a 10 to 100-fold reduction in communication cost compared to traditional synchronized SGD, underscoring FL's scalability and efficiency.

While foundational, the study mainly focuses on mobile environments and does not explore high-resolution or domain-specific data such as medical imaging. Additionally, it does not offer a detailed analysis of how data heterogeneity affects model performance.

Roth et al. (2020) explore the potential of FL to create robust medical imaging classification models within a collaborative framework involving multiple clinical institutions. The study focuses specifically on breast density classification using the Breast Imaging, Reporting & Data System (BIRADS). The collaboration includes seven clinical institutions worldwide, working together on an FL-based initiative to train a model capable of classifying breast density without centralizing data, thus maintaining patient privacy. Despite the significant variations in the datasets—such as differences in mammography systems, class distributions, and dataset sizes—the authors demonstrate that FL can effectively train models across multiple institutions. The results indicate that the FL-trained models outperform those trained on data from a single institution, with a notable 6.3% average improvement in performance. Furthermore, the study highlights a significant 45.8% improvement in model generalizability when tested on data from other participating sites, demonstrating the power of FL in enhancing model performance across diverse datasets. However, the study could benefit from further analysis of the computational cost and the time required for training models in a federated setting. Moreover, while the collaborative nature of FL is a significant strength, there is a need for further discussion regarding the potential challenges of implementing FL across a large number of institutions, such as varying technical capabilities and data standardization issues. Despite these considerations, the study successfully demonstrates the effectiveness of FL in creating scalable, privacy-preserving AI models in medical imaging.

Liu et al. (2020) explored the use of FL to address the data privacy and scarcity issues in COVID-19 detection using chest X-ray images. With DL and computer vision showing significant potential in identifying COVID-19 infections, the challenge remains in accessing large datasets due to the sensitive nature of patient data. Hospitals are often unable to share

patient data directly, which hinders the creation of extensive datasets for training. To overcome this issue, the authors proposed a FL-based approach, which enables training a shared global model while maintaining local data privacy. The study compared four popular DL architectures—MobileNet, ResNet18, MobileNetV2, and COVID-Net—with and without the FL framework. After conducting experiments with identical training parameters over 100 communication rounds, the authors found that ResNet18 achieved the fastest convergence and highest accuracy, reaching 96.15% on training and 91.26% on testing. ResNeXt showed similar convergence speed but underperformed in accuracy, while COVID-Net ranked second in terms of accuracy. Although MobileNetV2 exhibited a comparable loss value to COVID-Net, its test accuracy was notably lower. The study successfully demonstrates that FL can be effectively utilized in medical imaging applications, particularly for COVID-19 detection. However, the results suggest that further refinement in model architecture and training protocols may be necessary to improve the performance of certain models, particularly in terms of test accuracy and consistency across datasets.

Feki et al. (2021) propose a FL framework designed to support the collaborative development of DL models for COVID-19 diagnosis from chest X-ray images, without requiring centralized data sharing. Given the urgency and importance of rapid and accurate diagnosis during the COVID-19 pandemic, DL presents a promising solution by automating the analysis of large-scale medical images, which can aid radiologists in diagnosing cases quickly and efficiently. However, privacy regulations surrounding patient data present a significant barrier to the centralization of such datasets, which is a prerequisite for traditional model training. To overcome this challenge, the authors introduce a decentralized approach that allows multiple medical institutions to collaboratively train a shared model while maintaining control over their local data. The study focuses on the common challenges of FL, such as non-IID and imbalanced data distributions, which are often encountered in real-world medical datasets. Experimental results across two different model architectures demonstrate that the FL framework performs similarly to conventional data-sharing-based models. These findings highlight the practical feasibility of FL for building privacy-preserving, robust diagnostic systems and suggest that FL could play a significant role in enabling widespread adoption of AI-based diagnostic tools in healthcare, particularly in pandemic scenarios. However, while

the study demonstrates strong performance, the authors do not address the potential technical challenges associated with implementing FL in diverse healthcare environments with varying levels of technological infrastructure, which could affect the scalability and generalization of the proposed approach. Furthermore, the impact of the data imbalance issue on diagnostic accuracy remains an area that requires further exploration.

Kumar et al. (2021) propose a FL-based approach to address two major challenges faced by the medical field during the COVID-19 pandemic: the shortage and limited reliability of testing kits, and the difficulty of global data sharing due to privacy concerns. Their study presents an innovative framework that enables the collaborative training of a global DL model using small, decentralized datasets from various hospitals, while maintaining data privacy through blockchain integration. The framework also incorporates a data normalization method to reduce heterogeneity caused by different CT scanners used at multiple institutions. A Capsule Network is employed to handle both segmentation and classification tasks, specifically targeting COVID-19 detection from CT images. Blockchain technology ensures data integrity and authenticity, while FL enables decentralized model training. The experimental results, obtained from real-world datasets, demonstrate that the proposed method achieves high accuracy in identifying COVID-19 cases, highlighting its potential as a reliable diagnostic tool that preserves both patient confidentiality and institutional privacy. However, while the proposed framework shows promise, the study does not sufficiently address the computational costs of integrating blockchain with FL, which may pose challenges in large-scale implementation. Furthermore, the scalability of this approach could be tested more thoroughly across diverse healthcare settings with different data infrastructure.

Atef et al. (2022) investigate the feasibility of applying FL to predict the methylation status of the MGMT (O6-Methylguanine-DNA Methyltransferase) gene in brain tumor patients using MRI data. The ability to predict MGMT promoter methylation is clinically significant, as it directly informs chemotherapy responsiveness and helps reduce the need for invasive surgeries. The study utilizes the BraTS2021 dataset, which includes four MRI modalities: FLAIR, T1w, T1wCE (T1Gd), and T2w. By leveraging the OpenFL framework along with an enhanced EfficientNet-B3 architecture, the FL model was trained across ten institutions. After

300 communication rounds, it achieved 99.99% of the quality of a centralized model, despite the challenges posed by imbalanced and heterogeneous data. Remarkably, the final accuracy reached 96% in both FL and traditional DL configurations, demonstrating that FL can not only match but also outperform traditional methods in environments where strict data privacy is paramount. While this study highlights the promising potential of FL for preserving patient data privacy while maintaining model performance, there are certain limitations that should be considered. The reliance on a specific dataset (BraTS2021) may limit the generalizability of the model to other datasets or populations with different MRI protocols or tumor types. Additionally, the high performance achieved in this study may not fully account for challenges that arise in more heterogeneous and non-IID data distributions in real-world clinical settings. The study also emphasizes the quality of the model after 300 communication rounds, but the computational cost and time required for such extensive training rounds could be a significant hurdle in clinical practice. Future studies should explore ways to balance the performance with the computational efficiency to make this approach more practical for widespread adoption in healthcare systems.

Qayyum et al. (2022) emphasize that despite significant advancements in recent years, cloud-based healthcare systems still face adoption challenges, primarily due to their inability to meet essential security, privacy, and quality of service requirements, particularly in terms of low latency. In response, the integration of edge computing and distributed ML techniques, such as FL, has emerged as a promising solution. In their study, the authors investigate the potential of intelligent clinical data processing at the edge, employing a Clustered Federated Learning (CFL) framework for automatic COVID-19 diagnosis. The approach is evaluated in various experimental settings using two benchmark datasets, showing that the CFL-based model achieves competitive performance when compared to centralized, specialized models trained on specific modalities. Specifically, the CFL model shows improvements in overall F1-scores, with 16% and 11% increases in performance on X-ray and ultrasound datasets, respectively. Moreover, the study addresses several technical challenges, tools, and techniques involved in deploying ML models at the edge, especially in privacy- and latency-sensitive medical applications. While the results are promising, the authors do not explore the scalability and computational complexity of deploying such systems in large-scale healthcare

environments, which could be a potential area for further research. Additionally, the integration of FL in edge computing is still in its early stages, and more evidence is needed to assess its long-term feasibility in real-world clinical settings.

Islam et al. (2023) propose a FL model for brain tumor classification, aiming to overcome the limitations of centralized data collection and address privacy concerns in medical imaging. By training different CNN models on MRI datasets and combining the top three performing models into an ensemble, the authors demonstrate the effectiveness of decentralized learning. Their results show that the FL-based model achieved 91.05% accuracy, with only a minor performance drop compared to the traditional ensemble model (96.68%). Furthermore, the authors validated the scalability of their approach on a larger dataset, reinforcing its potential for real-world application. While the study successfully highlights the privacy preserving and scalable nature of FL, it falls short in providing a detailed analysis of the trade-offs between model complexity, training efficiency, and communication costs—factors that are crucial for broader implementation in clinical environments. Nonetheless, the research contributes meaningfully by showing that FL can maintain high diagnostic accuracy even in a decentralized setting.

Trivedi et al. (2023) explored the application of lightweight FL techniques to classify liver cancer, specifically HCC, while prioritizing data privacy. The study tested a variety of pre-trained CNN, including MobileNet, SqueezeNet, VGG11, ResNet50, ResNet18, ShuffleNet, and AlexNet. Among these, AlexNet achieved the highest accuracy (99.59%) in a traditional centralized learning setting. The authors then adapted AlexNet into an FL framework to evaluate its effectiveness in a decentralized context.

The federated system, tested under IID conditions with approximately 576 liver images, demonstrated classification performance comparable to the baseline centralized system. While the study emphasizes that FL enhances resource efficiency and patient data privacy, it does not explore the impact of non-IID data distributions, which are commonly encountered in real-world multi-institutional healthcare settings. Moreover, the relatively small dataset—entirely composed of diseased livers—limits the model's ability to generalize across a more

varied clinical population, especially for early detection scenarios where distinguishing healthy from diseased tissue is critical.

Despite these limitations, the paper successfully demonstrates that a lightweight FL architecture using AlexNet can preserve diagnostic accuracy while significantly improving data security and system efficiency, making it a promising approach for privacy-conscious liver cancer diagnostics. However, further research is needed to validate the framework's scalability and robustness across larger, more diverse datasets and under realistic, heterogeneous conditions.

Chai et al. (2024) proposed AdFed, a FL framework for survival prediction in multiple cancers including liver cancer. Using 733 genetic profiles, the model achieved an AUC of 0.605 for liver cancer, surpassing comparable FL approaches. A major strength is its biological interpretability, with half of the top genes already known to be associated with liver cancer. However, the limited sample size and lack of discussion on real-world FL challenges, such as scalability and communication overhead, weaken its practical applicability.

Zhou et al. (2024) investigate the potential of FL in overcoming privacy-related challenges in medical image analysis, with a specific emphasis on brain tumor (BT) detection using MRI data. By combining EfficientNet-B0 with the FedAvg algorithm, the study introduces an optimized FL framework that balances diagnostic performance with data privacy. Their comparative analysis shows that EfficientNet-B0 outperforms other models, such as ResNet, in accuracy, loss minimization, and handling data heterogeneity. This not only demonstrates the model’s robustness but also its suitability for decentralized healthcare applications. However, while the authors conduct a detailed technical evaluation of preprocessing methods and hyperparameters, the study lacks a discussion on communication overhead and real-world deployment challenges—critical aspects in translating FL into clinical practice. Nevertheless, the paper contributes valuable insights by showcasing FL’s capability to preserve patient privacy without compromising diagnostic accuracy and by outlining directions for future research on data heterogeneity in medical imaging.

Lusnig et al. (2024) introduced a FL framework using hybrid quantum neural networks (HQNNs) for classifying Non-Alcoholic Fatty Liver Disease (NAFLD) from histopathological biopsy images. The model achieved 97% accuracy under centralized training and around 90% under FL, demonstrating strong performance while preserving data privacy. Despite the promising results, the study is limited by a small dataset (41 patients) and potential scalability issues due to quantum infrastructure requirements.

Albalawi et al. (2024) address the challenge of brain tumor classification using MRI images, proposing a FL-based approach to overcome the limitations of traditional methods. The authors argue that manual interpretation of MRI images faces scalability issues and is prone to human error, thus necessitating automation in tumor classification. Their proposed model, which uses a modified VGG16 architecture tailored for brain MRI images, integrates CNNs and FL to achieve decentralized, privacy-preserving model training. Additionally, the incorporation of transfer learning enhances the model's ability to generalize across different tumor types. The model shows promising performance, with high precision and accuracy rates across various tumor classifications. However, while the study demonstrates strong quantitative results, there is little discussion of potential implementation challenges, such as communication overhead, the complexity of deploying the model in real-world clinical environments, and its adaptability to different healthcare infrastructures. Despite these gaps, the study highlights the transformative potential of FL and Transfer Learning in automating brain tumor classification, offering a scalable solution for privacy-sensitive medical applications.

Gohari et al. (2024) introduce an innovative approach that combines Knowledge Distillation (KD) with FL to tackle the challenges of brain tumor classification using MRI. The study addresses a critical issue in medical imaging: the need to maintain patient privacy while utilizing distributed learning models. Although FL offers a solution for privacy preservation, it faces challenges such as high communication costs and the need for uniformity in model architecture across clients. To overcome these, the authors propose FedBrain-Distill, which enables independent model architectures on each client while ensuring privacy preservation by using knowledge distillation. This approach reduces communication costs and maintains

high accuracy, which is a significant improvement over traditional FL methods. The experimental results, evaluated on both IID and non-IID data, show promising outcomes, but the study could benefit from further exploration of the method's scalability when applied to larger and more complex datasets. Additionally, while the model shows promising accuracy, the research could benefit from an in-depth discussion of its potential limitations in clinical settings, such as computational complexity or real-time application challenges. Despite these minor gaps, the study demonstrates that the combination of KD and FL provides a scalable and efficient solution to brain tumor classification, marking a significant step toward privacy-preserving, distributed ML in medical imaging.

Palas et al. (2024) introduce a privacy-preserving FL framework aimed at the early detection of lung cancer using CT scan images. Lung cancer remains one of the leading causes of cancer-related mortality, and the study underscores the importance of early and accurate diagnosis in improving survival rates. The proposed method involves creating a labeled dataset with four distinct classes: adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and normal lung tissues. Several preprocessing techniques were applied to enhance image quality and consistency before evaluating various CNN architectures, including MobileNet, MobileNetV2, ResNet50V2, VGG16, and InceptionV3. MobileNetV2 emerged as the most effective model, achieving a classification accuracy of 92.27%. Building upon this result, the authors developed a federated version of MobileNetV2, allowing decentralized training across multiple clients without centralizing sensitive data, thus addressing privacy concerns effectively. The federated model outperformed conventional models, achieving a higher accuracy 93.92%, precision 93.50%, recall 93.50%, and F1-score 93.25%. Importantly, the FL approach ensures data privacy, as it eliminates the need for hospitals and clinics to share sensitive patient data directly, making it a highly effective and secure solution for lung cancer detection. While the approach demonstrated robust results, one area that could be further investigated is the model's performance on more diverse datasets to assess its generalizability to different populations and imaging conditions. Additionally, the challenges of handling heterogeneous data across multiple clients in real-world settings may require further exploration to ensure scalability and efficiency in large-scale deployments.

Selvakanmani et al. (2024) address the critical challenge of breast cancer classification by proposing a novel FL framework combined with transfer learning and domain adversarial training. Given the immense importance of early and accurate breast cancer detection and the challenges posed by data privacy and decentralized medical datasets, the authors introduce a privacy-preserving model utilizing mammography and MRO images collected from three different medical institutions. The proposed approach employs a pre-trained ResNet model as a feature extractor, which is then fine-tuned on institution-specific datasets to optimize classification performance. To mitigate domain shift across healthcare centers, the model incorporates domain adversarial training, which helps the system learn domain-invariant features. Experimental results show that this approach outperforms both traditional standalone methods and FL models without adaptation, achieving an impressive classification accuracy of 98.8%, precision of 98.9%, recall of 98.5%, specificity of 97.4%, and an F1-score of 98.2%. These results underscore the potential of combining FL and transfer learning to develop privacy-preserving, accurate breast cancer detection models, which could be implemented across multiple healthcare institutions.

Shankar et al. (2025) developed a FL framework that combines CT, MRI, ultrasound medical imaging data with clinical test data (e.g. bilirubin, ALT, AST) for liver disease prediction. The approach is to achieve an acceptable accuracy of 79.05% while maintaining data confidentiality. However, this study lacks transparency regarding the dataset size and patient numbers. The objective of this study is to identify cases of general liver disease as opposed to tumor-specific classification.

Sivakumar et al. (2025) addressed two critical challenges in brain tumor classification (BTC) using ML: data privacy and heterogeneity across healthcare settings. To create a secure and accurate diagnostic system, the authors proposed a distributed learning framework combining the FedAvg and Federated Proximal (FedProx) algorithms. This hybrid approach allows decentralized training of CNNs across multiple clients with their local datasets, without requiring the sharing of sensitive patient data. FedAvg aggregates locally trained models into a robust global model, while FedProx introduces a proximal regularization term to stabilize the training process and manage data non-IID (Independent and Identically Distributed) conditions. The framework was evaluated using the Brain Tumor MRI dataset from Kaggle,

with the task of classifying four tumor categories: glioma, meningioma, pituitary tumor, and no tumor. After ten communication rounds, where each client trained independently on their local data, the proposed method achieved impressive performance metrics, with a global model accuracy of 97.19%, precision of 97.25%, recall of 97.19%, and F1-score of 97.18%. These results demonstrate that the framework not only ensures data privacy but also effectively addresses distribution disparities, outperforming models trained in traditional centralized settings. This study highlights the potential of FL in collaborative diagnostic applications, especially in privacy-sensitive medical domains.

Tölle et al. (2025) tackle one of the ongoing challenges in FL applications—working with partially labeled medical datasets. In real-world settings, not all data across decentralized sources are fully annotated, leading to significant portions of potentially valuable data remaining unused. Their study introduces a novel two-step semi-supervised learning approach to address this issue, with a specific focus on cardiac CT analysis. Initially, task-specific CNN models are deployed across eight hospitals to generate predictions for unlabeled data pertaining to specific labels. These predictions are then utilized to train a transformer model using label-specific attention heads. This architecture enables the system to learn efficiently from all available partial labels simultaneously across the federation. The results reveal that this strategy not only enhances predictive accuracy but also improves generalizability in downstream tasks, surpassing the performance of conventional UNet-based models. A notable contribution of this study is that it represents the largest federated cardiac CT study conducted to date (n = 8,104), and the authors openly share their code and model weights, encouraging further research in transformer-based FL for cardiac imaging. However, while the study demonstrates promising results, the authors do not explore the computational overhead or potential challenges associated with the deployment of such a semi-supervised learning approach in real-world clinical environments, which could be a limitation for wider adoption. Additionally, the reliance on partially labeled data raises concerns about the quality and consistency of the labels used for training, which could affect the model’s overall performance in certain clinical scenarios.

While the study demonstrates the success of FL in preserving privacy and fostering collaboration across multiple centers, further in-depth exploration of the effects of data heterogeneity and model aggregation processes could provide valuable insights for future studies. Additionally, while the study's sample size and testing datasets offer promising findings, it would be beneficial to test FL's performance with more diverse datasets from different healthcare institutions in future work.

Overall, the study proves that FL offers a strong alternative for tasks like digital histopathology, where privacy is paramount. It provides a solid foundation for further research on how FL can be optimized for model aggregation strategies and applied to even larger datasets.

3. MATERIALS AND METHODS

3.1. Image Processing

Image processing serves as a fundamental component in the analysis of digital images, incorporating a range of techniques aimed at enhancing, interpreting, and transforming visual data to derive meaningful insights. Initially, the focus was on improving image quality to support human evaluation through methods such as contrast adjustment, noise suppression, and edge enhancement. As the field progressed, more complex strategies like histogram equalization, spatial domain filtering, and frequency-based transformations were adopted, facilitating a shift toward automated image interpretation.

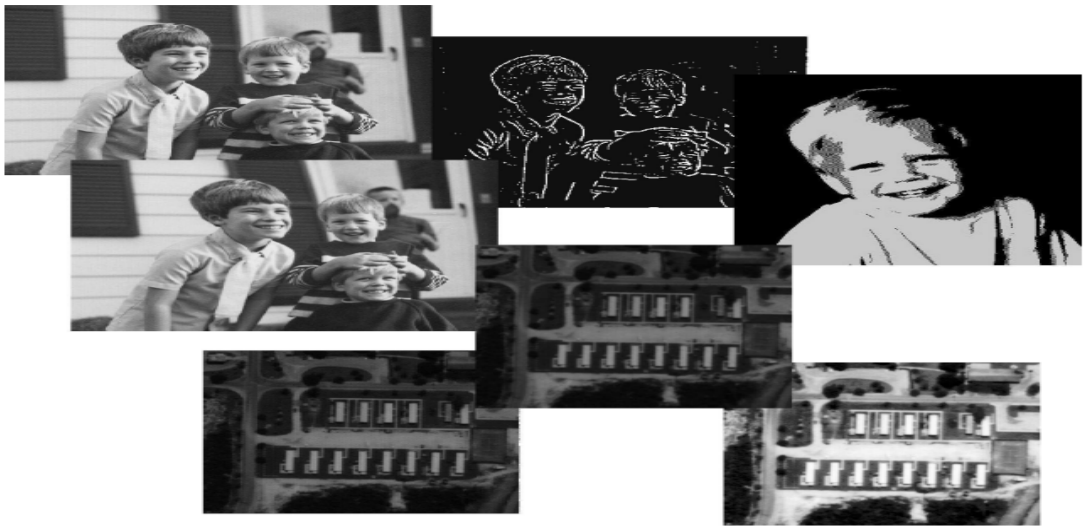

Figure 2. Image Processing (Hu et al. 2004)

The evolution of computational power enabled the transition from manual enhancement to automated image processing tasks such as segmentation, registration, and feature extraction. Segmentation divides an image into coherent regions to isolate significant structures, which is particularly important in areas like medical diagnostics and visual surveillance. Registration ensures spatial alignment between images obtained from different sensors or time points, allowing for consistent longitudinal analysis. Feature extraction involves identifying and quantifying critical image attributes, which are essential for subsequent classification or decision-making tasks.

Modern image processing has been significantly transformed by the integration of ML, enabling adaptive systems that refine their performance by learning from data. This synergy has improved both the accuracy and efficiency of image analysis, especially in handling complex and high-dimensional data. The advent of DL has further advanced the field by introducing models capable of learning multi-level features directly from raw images, minimizing the dependency on handcrafted features and paving the way for end-to-end learning in medical image interpretation and beyond.

3.2. Artificial Intelligence

AI is a multidisciplinary field that focuses on developing systems capable of mimicking human intelligence, including the abilities to learn, reason, and solve problems. Initially conceptualized in the mid-20th century, AI remained largely theoretical due to the limited computational resources of the time. However, advancements in hardware technology and the emergence of big data have significantly accelerated the practical deployment of AI systems. In particular, the rise of ML and artificial neural networks during the 1980s marked a turning point, transforming AI into a powerful tool now widely used across domains such as healthcare, finance, industry, and education.

The fundamental goal of AI is to design intelligent systems that replicate various aspects of human cognition. These systems are capable of performing complex tasks such as learning from data, making logical inferences, processing natural language, and recognizing images. In recent years, AI has shown great promise in healthcare, where it is employed in medical image analysis, early disease detection, and the development of personalized treatment plans. These applications contribute to improved patient outcomes and enhance the efficiency of healthcare services.

AI encompasses several subfields, including ML, DL, and artificial neural networks. These approaches enable machines to extract meaning from data, recognize patterns, and make autonomous decisions—functions traditionally associated with human intelligence. Figure 3. presents a visual overview of the conceptual relationship among these domains.

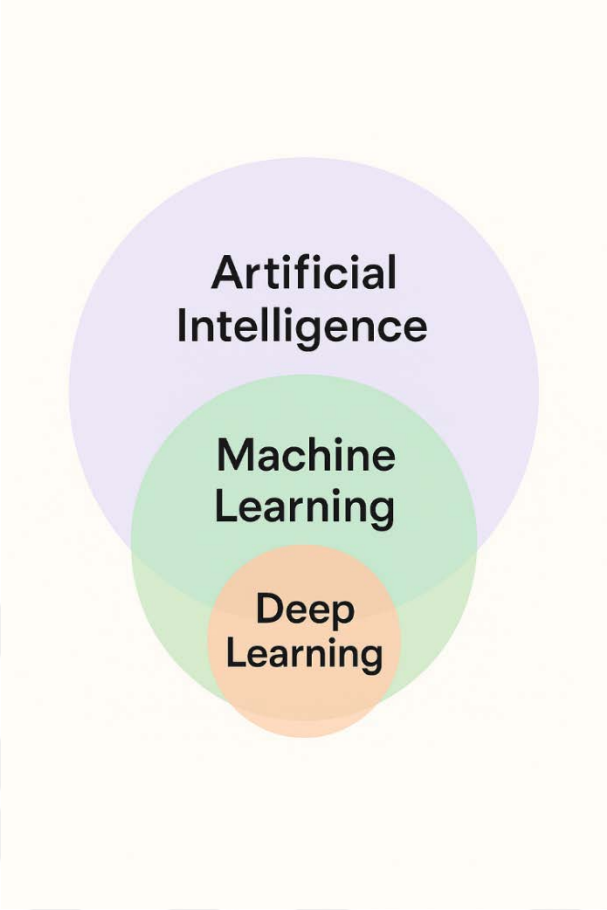

Figure 3. Drawing the Relationship between Artificial Intelligence, Machine Learning, and Deep Learning

3.3. Machine Learning

ML is a subfield of AI focused on the development of intelligent systems that can learn from data and improve their performance without being explicitly programmed. ML has found broad applications in domains such as image recognition, natural language processing, prediction modeling, and robotics. In the healthcare field, it plays a pivotal role in tasks like classifying medical images, extracting diagnostic insights from patient records, generating automated reports, and detecting fraudulent activities.

ML algorithms operate by analyzing historical data to extract patterns, which can then be applied to predict future outcomes or assist decision-making. The field primarily includes three major types of learning: supervised learning, unsupervised learning, and reinforcement learning—each with distinct methods tailored to various data types and objectives.

Among its key advantages are the automation of decision-making processes, the discovery of hidden patterns within large datasets, and the reduction of human error. Nevertheless, ML also faces limitations such as the need for vast quantities of labeled data, risks of algorithmic bias, and challenges related to model interpretability—especially critical in sensitive areas like healthcare.

Currently, ML is bringing transformative impacts in medical diagnosis, treatment planning, and patient monitoring. Ongoing research and development continue to foster the creation of more precise and efficient systems, facilitating the broader integration of AI-driven solutions into healthcare infrastructures.

3.4. Deep Learning

DL is a methodology distinct from traditional AI and ML approaches, capable of autonomously learning intricate patterns from large datasets through multilayered artificial neural networks. Inspired by the structure of the human brain, DL systems process data through successive layers, enabling them to extract complex and abstract relationships inherent in the input data. This allows for high performance particularly in handling high-dimensional, unstructured, or heterogeneous data.

Key advantages of DL include higher accuracy compared to conventional models, reduced need for manual feature engineering, and adaptability to a wide range of data types and problem domains. However, it also presents several limitations: it demands substantial data volumes for effective training, requires significant computational resources, and suffers from limited model interpretability—posing challenges in critical domains such as healthcare where explainability is vital.

DL is actively employed across diverse fields including image processing, natural language understanding, robotics, financial forecasting, and medical diagnostics. As the technology continues to evolve, it is expected to play an increasingly pivotal role in the development of intelligent systems. This advancement holds transformative potential across various sectors,

laying the groundwork for more autonomous and accurate decision-support technologies in the near future. Figure 4 shows the general scheme of DL (Marigliano, 2024).

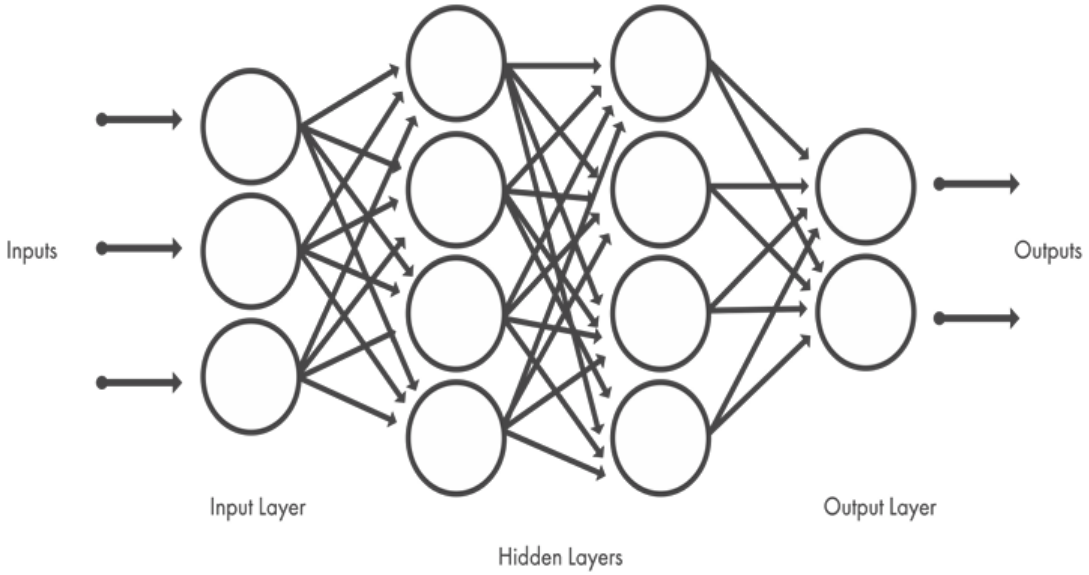

Figure 4. Deep learning

3.5. Common Computer Vision Tasks

In the field of computer vision, three primary tasks—classification, detection, and segmentation—constitute the foundation of most image analysis systems. Each of these tasks serves a unique function in extracting structured information from visual data and plays a crucial role in the automation of image-based decision-making processes.

Image classification refers to the process of assigning a category or label to an image based on its visual features. This can take the form of binary classification, where images are divided into two groups (e.g., benign vs. malignant tumors), or multiclass classification, which involves multiple categories (e.g., identifying different types of medical abnormalities or road signs). Classification techniques are widely used across domains such as industrial automation and medical diagnostics.

Object detection builds upon classification by not only recognizing objects within an image but also pinpointing their locations using bounding boxes. This capability is essential in applications that require spatial awareness, such as surveillance systems monitoring human activity, or autonomous vehicles that must detect and localize pedestrians, obstacles, and other vehicles in real-time environments.

Segmentation further refines image interpretation by dividing the image into meaningful regions. It includes semantic segmentation, where each pixel is labeled according to a class, and instance segmentation, which also differentiates individual instances of the same class. For example, distinguishing between separate tumors in a radiological image or identifying individual people in a crowded scene. These tasks collectively enable more precise and context-aware analysis in computer vision systems, significantly enhancing AI applications in fields such as healthcare, robotics, and smart surveillance. Figure 5 illustrates comparative overview of semantic segmentation, image classification with localization, object detection, and instance-level segmentation tasks.

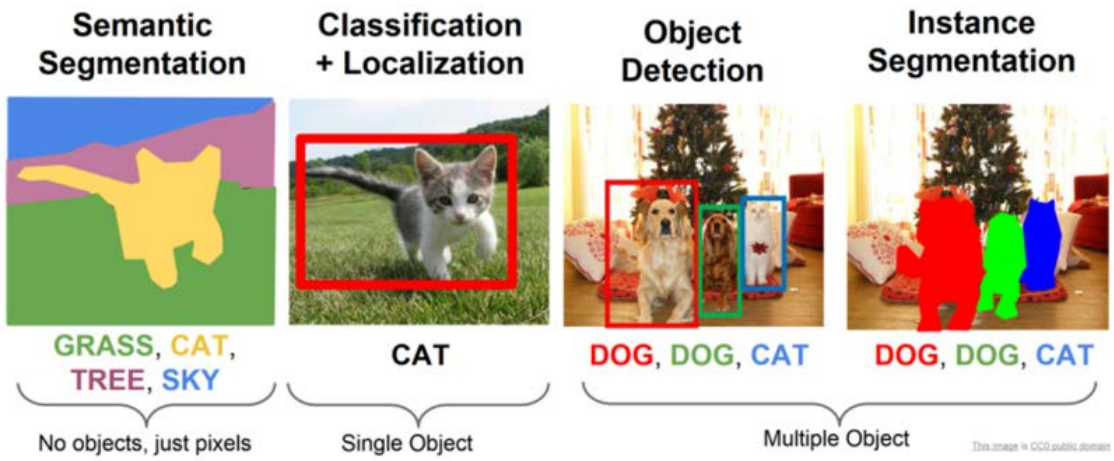

Figure 5. A comparative overview of semantic segmentation, image classification with localization, object detection, and instance-level segmentation tasks (Li, Johnson and Yeung, 2017)

3.6. Data Description

The dataset used in this study was obtained from a publicly available dataset from the ATLAS (A Tumour and Liver Automatic Segmentation) competition. The dataset includes contrast-enhanced T1-weighted magnetic resonance images (CE-MRI) obtained from 90 patients diagnosed with inoperable HCC at the University Hospital of Dijon between 2012 and 2023. All images were anonymized in accordance with French data protection laws, and no ethics committee approval was required according to local regulations. The dataset includes segmentation masks for both liver and tumor regions, allowing for detailed analysis.

For the purposes of this study, the dataset was divided into two classes: “normal” and “tumor”. The training dataset consists of 1,808 normal and 1,815 tumor images, while the test dataset contains 775 normal and 778 tumor images. All images were resized to 224×224 pixels and preprocessed to ensure compatibility with DL models. The aim was to increase the classification performance by maintaining a balanced distribution between classes. Figure 6 visually reveals the structure of the dataset by showing the distribution of training and test images belonging to both categories.

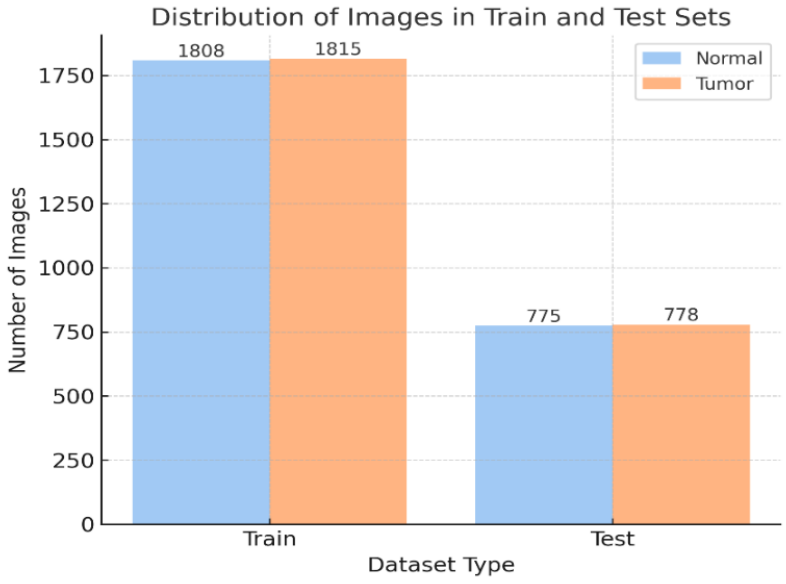

Figure 6. The numerical distribution of normal and tumor images in the training and test datasets

The dataset used in this study was split into training and test sets to ensure a robust model evaluation. As shown in Figure 7, 70% of the dataset was allocated for training, while 30% was reserved for testing. This split was chosen to maintain a balance between model optimization and reliable evaluation. The partitioning was performed randomly to prevent any bias and to ensure a representative distribution of the data across both subsets. This structured division supports the model's ability to generalize effectively while maintaining a sufficient amount of data for performance assessment.

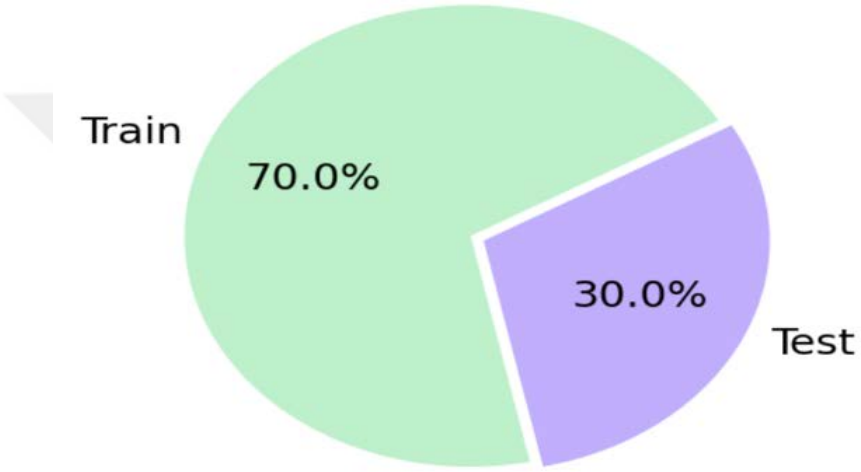

Figure 7. Train and test dataset distribution.

To further illustrate the dataset’s composition, Figure 8 presents sample MRI images from both the training and test sets, categorized as normal and tumor. These images provide a visual representation of the dataset’s diversity, highlighting variations in tumor appearance and normal anatomical structures. The inclusion of these examples aids in understanding the dataset’s characteristics and the challenges associated with tumor classification.

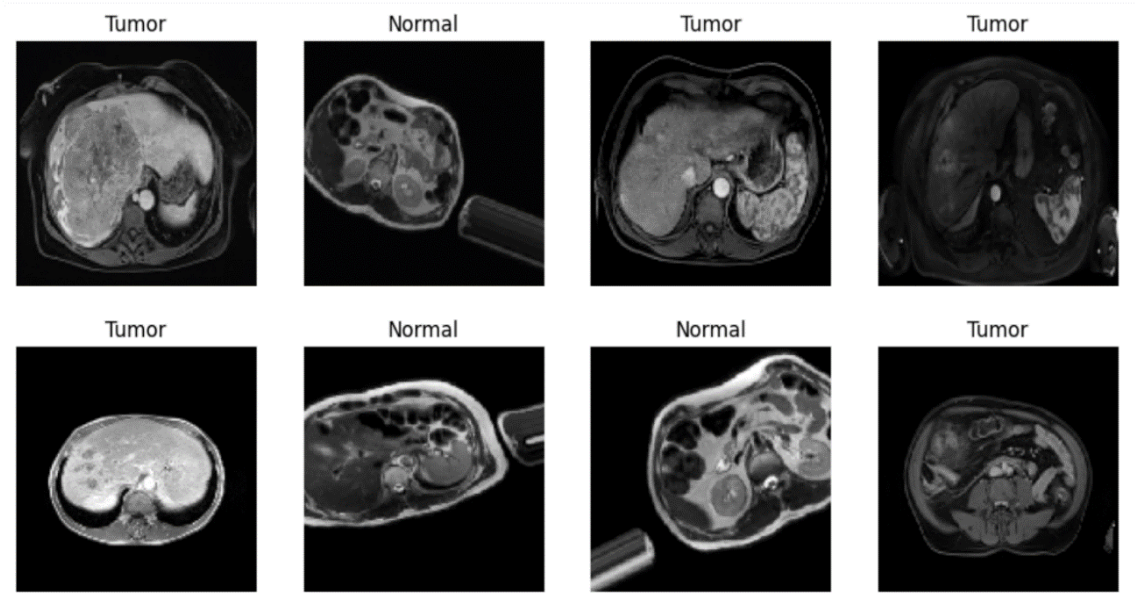

Figure 8. Sample images from the dataset.

3.7. Data Preprocessing

In this study, a rigorous data preprocessing pipeline was applied to enhance model performance and ensure data consistency. Initially, all MRI scan images were resized to 224 × 224 pixels to maintain uniformity across different models. To improve model generalization and reduce overfitting, data augmentation techniques were implemented, including rotation (±40°), horizontal flipping, zooming (±30%), brightness adjustment (range: 0.8–1.2), width and height shifting (±30%), and shear transformations (±30%)**.** These augmentations increased dataset diversity and improved model robustness. Furthermore, all images were normalized by rescaling pixel values to the [0,1] range, ensuring numerical stability during training. Since DL models require well-structured datasets, a class-balancing strategy was applied to prevent bias, ensuring an equal representation of tumor and non-tumor images.

3.8. Comparison Models for Liver Tumor Classification

To evaluate the effectiveness of the proposed FL model in liver tumor classification, five widely used DL architectures—CNN, EfficientNet, MobileNetV3, ResNet50, and VGG16—were implemented as baseline models. These models served as comparison baselines and were trained using identical hyperparameter configurations to ensure a fair evaluation.

Specifically, each model was trained for 10 epochs using a batch size of 32, the Adam optimizer, and binary cross-entropy as the loss function. ReLU was used as the activation function in hidden layers, while the output layer utilized the sigmoid function, appropriate for binary classification. Table 2 summarizes the training configurations.

Table 5. Summarization of the training configurations

	Number of Epochs	**Batch Size**	**Optimizer**	**Loss Function**	**Activation Functions**	**Output Layer**
CNN	10	32	Adam	Binary Crossentropy	Relu	Sigmoid
EfficientNet	10	32	Adam	Binary Crossentropy	Relu	Sigmoid
MobileNetV3	10	32	Adam	Binary Crossentropy	Relu	Sigmoid
ResNet50	10	32	Adam	Binary Crossentropy	Relu	Sigmoid
VGG16	10	32	Adam	Binary Crossentropy	Relu	Sigmoid

3.8.1. Convolutional Neural Networks (CNN)

In this study, a custom CNN was implemented as a baseline model for liver tumor classification. The architecture consists of three convolutional layers with increasing filter sizes (32, 64, and 128), each followed by a MaxPooling2D layer to reduce spatial dimensions and control overfitting. The feature maps are then flattened and passed through a dense layer with 128 neurons activated by ReLU. Finally, a sigmoid-activated output layer is used to perform binary classification. This relatively simple yet effective architecture enables a foundational comparison with more advanced pre-trained models.

3.8.2. EfficientNet

As part of the comparative analysis, EfficientNetB0 was utilized to assess liver tumor classification performance. The model was fine-tuned by unfreezing the last 50 layers of the

pre-trained base and retraining them on the target dataset. To adapt it for the binary classification task, several layers were added on top of the base model, including a GlobalAveragePooling2D layer, a dropout layer with a 0.3 rate to prevent overfitting, and two dense layers—one with 128 neurons and ReLU activation, and a final output layer with a sigmoid activation function. This configuration enabled the model to learn domain-specific features while benefiting from transfer learning.

3.8.3. MobileNetV3

MobileNetV3 Small was adapted in this study using a transfer learning approach tailored for medical image classification tasks. The model utilized pre-trained weights, with its initial layers frozen to preserve low-level feature extraction. Specifically, the MobileNetV3Small architecture was selected, followed by the addition of a Global Average Pooling layer, a fully connected Dense layer with 128 neurons and ReLU activation, a Dropout layer with a rate of 0.5, and a final output layer with a single neuron using sigmoid activation. During training, the Adam optimizer and binary cross-entropy loss function were employed. Furthermore, the learning rate was dynamically adjusted using the ReduceLROnPlateau technique, which monitored validation loss and reduced the learning rate when improvements plateaued.

3.8.4. ResNet50

ResNet50, known for its powerful residual blocks, was selected as one of the benchmark models in this study. The model architecture included the pre-trained ResNet50 as the base, with its weights frozen to retain previously learned features. Following the base model, a GlobalAveragePooling2D layer was applied to reduce the spatial dimensions, which was then followed by a fully connected Dense layer with 128 neurons and ReLU activation. A Dropout layer with a rate of 0.5 was added to prevent overfitting. Finally, the output layer consisted of a single neuron with a sigmoid activation function for binary classification. The model was compiled using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy as the loss function. Additionally, a learning rate scheduler (ReduceLROnPlateau) was implemented to reduce the learning rate by a factor of 0.1 if the validation loss plateaued for 3 consecutive epochs.

3.8.5. VGG16

One of the comparison models utilized was the VGG16 architecture. The base of the model consisted of the VGG16 network pre-trained on ImageNet, with its top layers removed. To prevent overfitting and focus training on newly added layers, the weights of the pre-trained layers were frozen. A Global Average Pooling layer was added to flatten the feature maps, followed by a fully connected dense layer with 128 neurons and ReLU activation, and a Dropout layer with a rate of 0.5. The final output layer employed a sigmoid activation function for binary classification. The model was compiled using the Adam optimizer with a learning rate of 0.001 and binary crossentropy as the loss function.

3.9. Federated Learning Approach

AI has revolutionized the field of medical imaging by offering the potential to automate diagnostics and improve accuracy across various healthcare environments. Nonetheless, the development of robust AI models necessitates access to extensive and heterogeneous labeled datasets—a requirement that remains difficult to fulfill in clinical settings due to strict data privacy regulations, inter-institutional variability, and the scarcity of annotated medical data. As a response to these limitations, collaborative and privacy-conscious methodologies have been introduced. Among these, FL stands out as a prominent decentralized approach. FL allows multiple medical institutions to locally train models on their private data and contribute to a shared global model by transmitting only model updates. This ensures that sensitive patient information remains within the source institution while facilitating the integration of knowledge from diverse data distributions (Lotfinia et al. 2025). To maximize the collaborative benefits of FL, a strategy involving full client participation was adopted in this study.

In this study, all three clients participated in the model training process in each federated round. Each client trained its local model for 3 epochs on its respective data, and this process was repeated over 10 federated rounds. This full participation strategy ensures that updates from every client are incorporated into the global model in each round, promoting more efficient and balanced learning.

FedAvg is one of the most commonly used optimization algorithms in horizontal FL, where client data remains decentralized and is not shared with a central server. This method has been widely studied in terms of convergence properties under different conditions such as data heterogeneity and loss function variations, demonstrating its robustness in collaborative learning environments (Overman et al. 2025). In this study, FedAvg was employed as the aggregation mechanism for synchronizing the local models of the three clients into a single global model. After each client completed training its local model over three epochs, the learned weights were transmitted to the central server, where layer-wise averaging was performed. The resulting global weights were then redistributed to all clients for the next training round. This iterative process was repeated for a total of 10 federated rounds, enabling the global model to progressively improve by integrating knowledge from all clients while preserving data privacy.

For the training process of the federated model, the binary cross-entropy loss function was employed during the training of the FL model. This choice is appropriate given the binary nature of the liver tumor classification task, where the model is expected to distinguish between two classes. Binary cross-entropy effectively measures the dissimilarity between predicted probabilities and true class labels in binary classification problems, making it a standard and reliable loss function for such applications. The optimization of the model was carried out using the Adam optimizer, known for its adaptive learning rate and robust performance in DL tasks. In addition to minimizing the loss function, several performance metrics were tracked during training and evaluation, including accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide a comprehensive assessment of the model's classification capability, especially in a clinical context where sensitivity and specificity are equally important.

Figure 9 summarizes the training and testing process applied in the study. This process is structured according to the FL approach. Within this structure, independent models are trained using local data on different clients. After each training round, the local model weights are transferred to the central server, where the FedAvg algorithm is applied to update the global

model. This cycle is repeated until the model gains accuracy. After the final model is created, its performance is evaluated using the central test data.

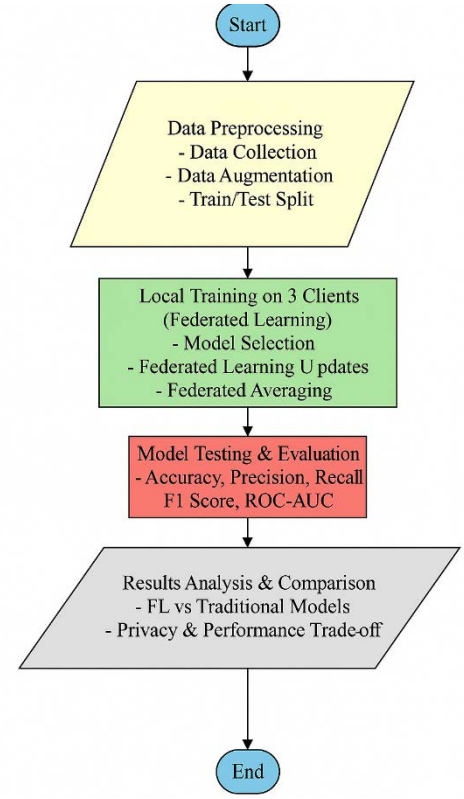

Figure 9. Sample images from the dataset

3.9.1. Types of Federated Learning

FL fundamentally enables the collaborative training of ML models across multiple devices or servers without centralizing sensitive data. By keeping data localized within each institution it ensures privacy while still contributing to a global model. Within this framework, FL is generally categorized into three primary types: horizontal FL, vertical FL, and federated transfer learning.

3.9.1.1. Horizontal Federated Learning

Horizontal FL, also referred to as sample-based FL, is applied in scenarios where data sets share the same feature space but differ in data samples Figure 10. For instance, two

regional banks may serve significantly different customer bases with little overlap. However, since they operate in the same business domain, the feature space of their data remains similar. In such cases, a collaborative DL framework was proposed where participants train models locally and share only subsets of their parameter updates (Shokri and Shmatikov, 2015).

In 2017, Google introduced a horizontal FL solution for updating Android device models (McMahan et al., 2016). In this framework, each Android user locally updates the model parameters and uploads them to the Android cloud, thereby contributing to training a centralized global model. To ensure user privacy, a secure aggregation scheme was introduced (Bonawitz et al., 2017). Additionally, homomorphic encryption was proposed to securely aggregate model parameters without exposing sensitive information (Phong et al., 2018).

A secure client-server architecture was designed in which local models trained on client devices collaborate at the server side to construct a global federated model, while preventing data leakage (McMahan et al., 2016). Similarly, methods were developed to reduce communication overhead when training centralized models using data distributed across mobile clients (Konecný et al., 2016). To address communication bottlenecks in large-scale distributed training, a compression strategy called Deep Gradient Compression was introduced (Lin et al., 2017).

In horizontal FL systems, each participant (i and j) possesses distinct input (Xi, Xj) and output (Yi, Yj) data, uniquely identified by their indices ($i \neq j$). This relationship is formally represented by the following equation:

$$Xi = Xj \ , Yi = Yj \ , Ii \neq Ij \ , \forall Di \ , Dj \ , i \neq j \tag{1}$$

Such systems typically assume honest participants and an honest-but-curious server, meaning the server is the only entity that could potentially compromise data privacy (Bonawitz et al., 2017; Phong et al., 2018). More recent work considered adversarial participants and outlined additional privacy concerns in such environments (Hitaj et al., 2017). At the end of training, the final global model and its parameters are shared with all participating clients. Figure 10 illustrates horizontal FL (Yang et al., 2019).

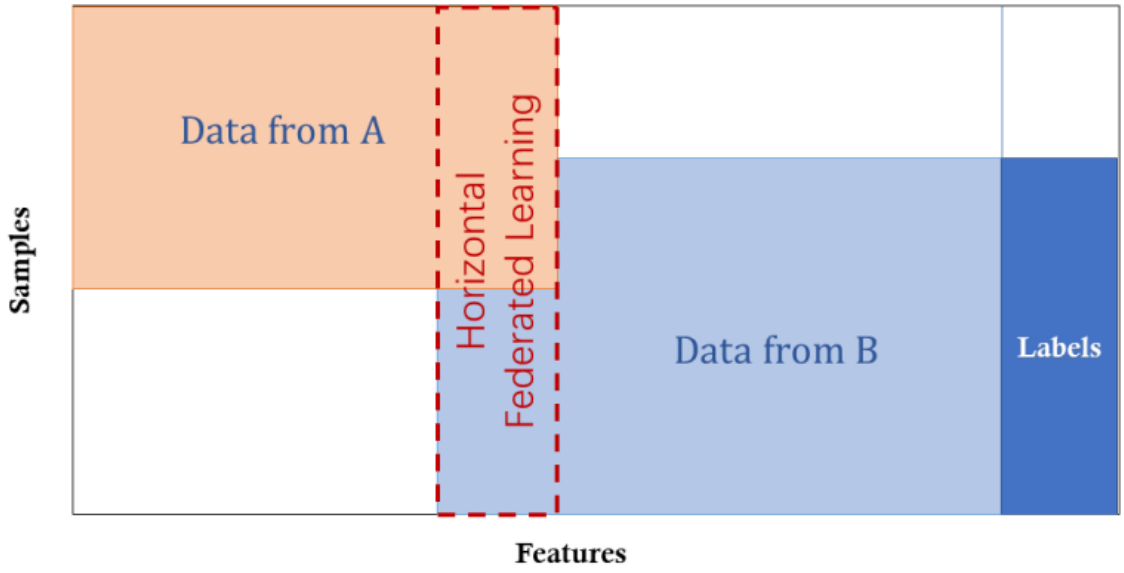

Figure 10. Horizontal Federated Learning (Yang et al., 2019)

3.9.1.2. Vertical Federated Learning

Vertical Federated Learning (VFL) is a FL paradigm designed for scenarios where multiple parties possess information about the same data instances but with different feature sets. In such settings, the data from each party aligns at the sample level—such as patients, customers, or users—but the attributes describing those samples vary. For example, a hospital may hold demographic and clinical information about a patient, while a genomics lab may hold DNA sequence data for the same individual. These complementary data views necessitate collaborative learning without centralized data aggregation due to privacy and regulatory constraints.

VFL enables collaboration through secure exchange of intermediate information, such as encrypted gradients or statistical representations, without exposing raw data. To preserve privacy, advanced cryptographic techniques such as Secure Multi-Party Computation (SMPC), homomorphic encryption, and differential privacy are commonly integrated into the VFL pipeline. These mechanisms ensure that while collaboration occurs, no party gains access to another's sensitive data.

This approach plays a vital role in sectors like healthcare, finance, and government, where data is distributed vertically across institutions and strict privacy regulations are in place. For instance, a bank and an insurance company may wish to collaborate on fraud detection or credit scoring, using different sets of features for the same set of clients. VFL allows them to jointly train a predictive model without compromising data confidentiality.

In conclusion, VFL is a powerful method that facilitates privacy-preserving ML on samples with distributed but complementary attributes. It supports ethical, lawful, and high-performance collaboration by allowing institutions to learn from structurally diverse yet semantically related data sources. Figure 11 illustrates VFL (Yang et al., 2019).

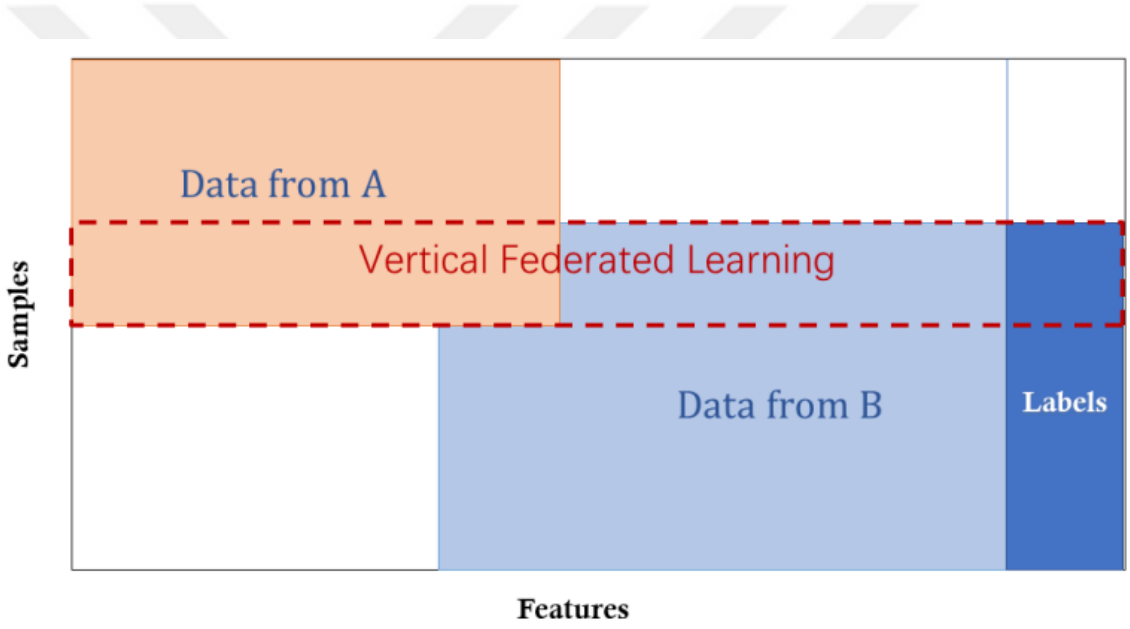

Figure 11. Vertical Federated Learning (Yang et al., 2019)

3.9.1.3. Federated Transfer Learning

Federated Transfer Learning (FTL) emerges as an effective solution in scenarios where traditional FL methods fall short—particularly when there is minimal overlap in both data instances and feature spaces among participating parties. While horizontal FL handles the same features across different samples and VFL addresses the same samples with different features, FTL is designed for environments where neither the sample sets nor the feature sets

are substantially shared. As such, FTL enables collaboration between organizations from distinct domains that operate on structurally dissimilar data sources.

The primary objective of FTL is to facilitate knowledge transfer across parties without the need to exchange raw data. This is achieved by leveraging transfer learning techniques, which allow pre-trained knowledge from one party's model to be reused or adapted by another party. For example, shared model parameters, feature embeddings, or network layers can be transferred and fine-tuned for the target task. This process allows collaborative learning to take place even in the absence of substantial data alignment, ultimately enhancing model performance.

Privacy preservation is a core component of the FTL framework. To ensure that no sensitive information is exposed during the training process, privacy-preserving mechanisms such as secure multi-party computation (SMPC), differentially private stochastic gradient descent (DP-SGD), and homomorphic encryption are integrated. These cryptographic and algorithmic techniques make FTL particularly suitable for sensitive sectors such as healthcare, finance, and legal industries, where regulatory compliance is crucial.

In summary, FTL offers a strategic approach for applications where data sharing is restricted or infeasible but where the value of knowledge transfer is significant. By enabling collaborative learning across structurally heterogeneous and privacy-sensitive environments, FTL promotes the development of robust, generalizable, and ethically sound ML models. It plays a critical role in advancing data-driven solutions in domains where data silos and privacy regulations pose substantial challenges. Figure 12 illustrates FTL (Yang et al., 2019).

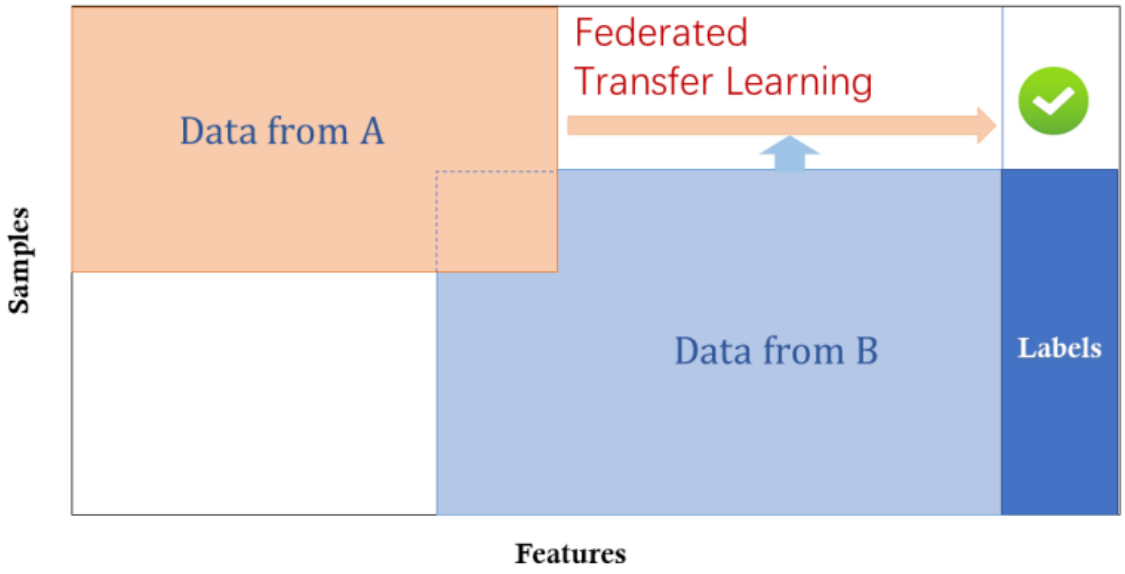

Figure 12. Federated Transfer Learning (Yang et al., 2019)

3.9.2. Advantages of Federated Learning

FL offers significant advantages over traditional centralized learning approaches, especially in domains where data privacy is of critical importance. Its primary benefit lies in enabling the training of ML models without compromising data security and privacy. Participating clients or organizations retain data locally rather than transferring it to a central server; only model parameters or gradients are shared. This makes collaborative learning feasible even in highly regulated domains such as healthcare, finance, and law.

Another key advantage is reduced communication and infrastructure costs by preserving data locality. In scenarios involving numerous low-power edge devices, transferring all data to a centralized system is neither practical nor cost-effective. FL allows local computation, eliminating the need for centralized data storage and minimizing network traffic.

Additionally, FL leverages data diversity to improve the generalizability of models. By incorporating data from various geographical, demographic, or institutional sources, the trained models avoid overfitting to a single dataset and learn more universal patterns. This enhances robustness and reliability in real-world applications.

Finally, FL provides a framework that facilitates cross-institutional collaboration. Competing or independent entities can work together toward a common learning goal without sharing sensitive data, promoting innovation while staying within ethical and legal boundaries. Furthermore, the ability to tailor models to local conditions enables the development of more personalized services.

3.9.3. Challenges of Federated Learning

FL, while providing a privacy-preserving collaborative model training framework, faces a range of technical, operational, and domain-specific challenges that impact its feasibility and effectiveness. One primary issue is data heterogeneity (non-IID data distribution), where data held by different institutions or devices vary significantly. This complicates achieving consistent learning performance across all parties and can degrade overall model accuracy.

Communication overhead and latency present additional challenges, especially in settings involving numerous edge devices or diverse healthcare institutions. Frequent parameter exchanges require substantial bandwidth, and network instability can cause training delays, making efficient communication protocols essential.

Adversarial clients pose a further threat by potentially sending malicious updates that degrade model performance or compromise data privacy. Security measures such as secure aggregation and anomaly detection are crucial in mitigating such risks.

In medical contexts, particularly for diseases like HCC diagnosis, FL introduces further challenges. Data heterogeneity is accentuated due to differences in MR imaging devices, protocols, and patient demographics across hospitals, complicating model generalization and necessitating extensive preprocessing.

Moreover, the sensitive nature of healthcare data invokes strict ethical and legal regulations (e.g., GDPR, HIPAA), making data handling and sharing complex. Institutions must enforce stringent privacy and security policies throughout federated training to comply.

Infrastructure variability among hospitals is another obstacle, where disparities in computational resources lead to performance inconsistency and synchronization issues during training.

Data quality and annotation in medical imaging are critical for diagnosis accuracy. Variability in image quality and the costly, expert-dependent labeling process complicate producing consistent and reliable models in federated environments.

Finally, the risk of indirect privacy leakage via model updates is particularly critical in medical applications, necessitating advanced privacy-preserving techniques like differential privacy and homomorphic encryption. However, these methods often increase computational demands and may reduce system efficiency.

3.9.4. Summary of Model Comparisons

To benchmark the performance of the proposed FL model, five different DL architectures were employed for liver tumor classification. The custom CNN model featured a straightforward structure with three convolutional layers, max pooling, and fully connected layers. EfficientNetB0, used with fine-tuning of the last 50 layers, incorporated a Global Average Pooling layer, Dropout, and dense layers. MobileNetV3 and ResNet50, both initialized with pre-trained ImageNet weights, followed a similar architectural pattern—each with a Global Average Pooling layer, dense ReLU-activated layers, and Dropout regularization. VGG16 was also utilized with frozen convolutional layers and additional dense layers for classification. Across all models, binary crossentropy was used as the loss function and Adam was chosen as the optimizer. The consistent design in compilation and training settings ensured a fair and robust comparison among the models.

Table 6. Summary of architectural and training parameters of the comparison models

Model	Pretrained Weights	Trainable Layers	Pooling Type	Dropout Rate	Dense Layer (Units)
CNN	No	All	MaxPooling2D	None	128
EfficientNet	ImageNet	Last 50 layers	GlobalAveragePooling2D	0.3	128
MobileNetV3	ImageNet	Frozen (feature extractor)	GlobalAveragePooling2D	0.5	128
ResNet50	ImageNet	Frozen (feature extractor)	GlobalAveragePooling2D	0.5	128
VGG16	ImageNet	Frozen (feature extractor)	GlobalAveragePooling2D	0.5	128

4. FINDINGS AND DISCUSSION

A variety of metrics are employed to comprehensively evaluate the performance of a liver tumor classification model. These metrics serve as fundamental evaluation criteria to assess the model's accuracy, its ability to distinguish between different tumor types, and its overall effectiveness in the classification task.

4.1. Accuracy

Accuracy is a fundamental metric used to evaluate the overall performance of a model and is defined as the ratio of all correctly classified instances (both positive and negative) to the total number of predictions. Mathematically, accuracy is calculated in Equation 2. While accuracy is a reliable performance indicator in balanced datasets, it can be misleading when there is an imbalance between classes. For instance, in datasets where one class dominates the other, the model may achieve a high accuracy by predominantly predicting the majority class, which does not accurately reflect the model's classification ability.

1. TP = True Positives.
2. TN = True Negatives.
3. FP = False Positives.
4. FN = False Negatives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

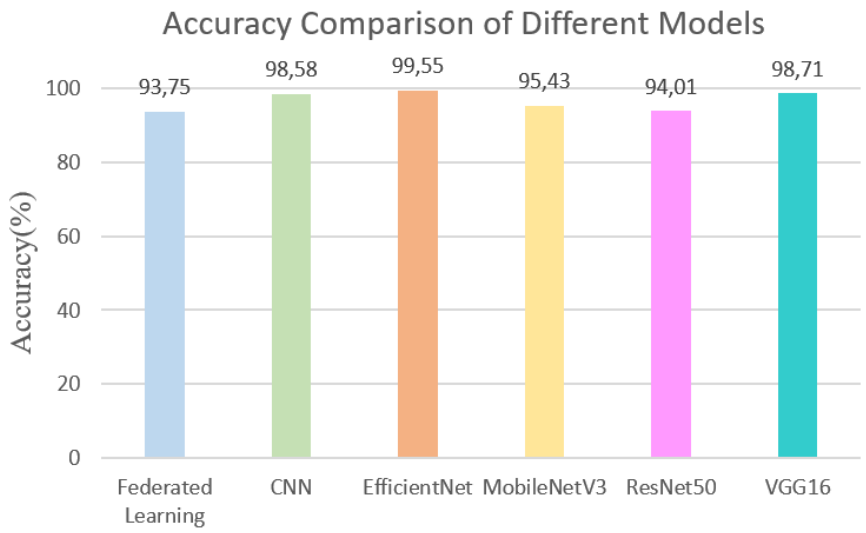

Figure 13. Accuracy Comparison of Different Models

4.2. Precision

Precision measures the proportion of instances classified as positive that are truly positive, and it is calculated in Equation 3. A high precision value indicates that the model minimizes false positives and makes more reliable positive classifications. This metric is particularly important in cases where false positives carry a significant cost, such as in medical diagnosis systems. In liver cancer classification, for example, a low precision score may lead to healthy individuals being incorrectly diagnosed as having cancer, resulting in unnecessary medical interventions and psychological distress.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{3}$$

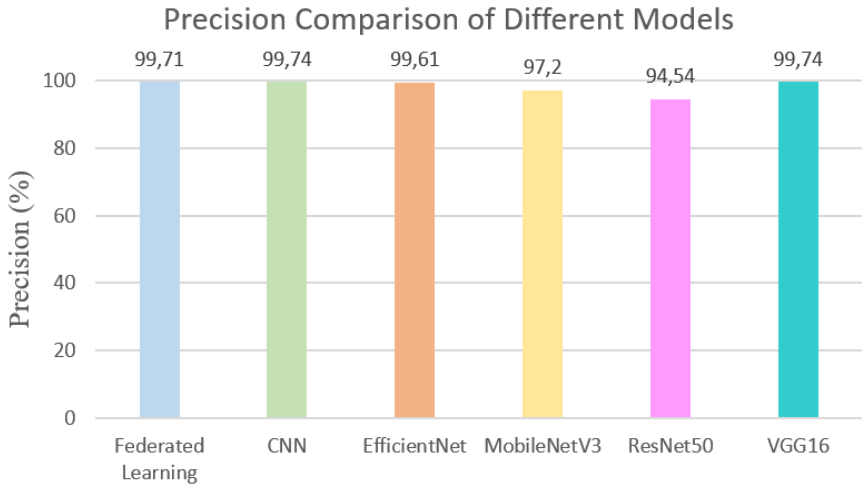

Figure 14. Precision Comparison of Different Models

4.3. Recall

Recall measures how effectively a model identifies positive instances and is calculated Equation 4. A high recall value indicates that the model minimizes false negatives and correctly captures most actual positive cases. This metric is particularly crucial in scenarios where missing a positive instance carries significant consequences, such as in medical diagnosis applications. In liver cancer detection, for example, a low recall score means that some patients may be incorrectly classified as healthy, leading to missed diagnoses and delayed medical interventions. Therefore, in healthcare-related classification problems, achieving a high recall score is essential for ensuring reliable and comprehensive disease detection.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{4}$$

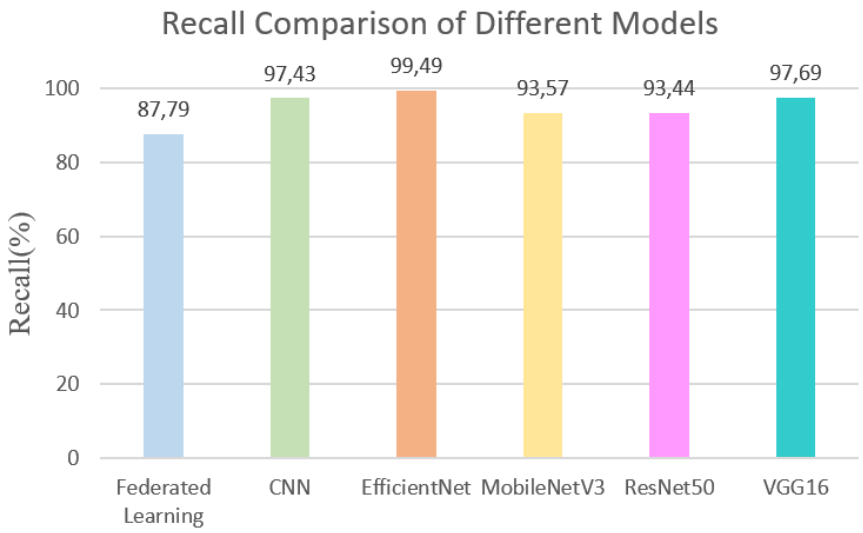

Figure 15. Recall Comparison of Different Models

4.4. F1 - Score

F1-Score is the harmonic mean of precision and recall (sensitivity), calculated Equation 5. This metric is particularly useful when dealing with imbalanced datasets, as it considers both false positives and false negatives rather than focusing on a single performance measure. A high F1-Score indicates that the model effectively minimizes both false positives and false negatives. In critical classification problems such as cancer diagnosis, F1-Score serves as a valuable metric for evaluating the model's overall effectiveness, ensuring that both missed diagnoses and incorrect positive classifications are minimized.

$$F1\_score = 2 \text{ X } \frac{\text{Precision X Recall}}{Precision+Recall} \quad (5)$$

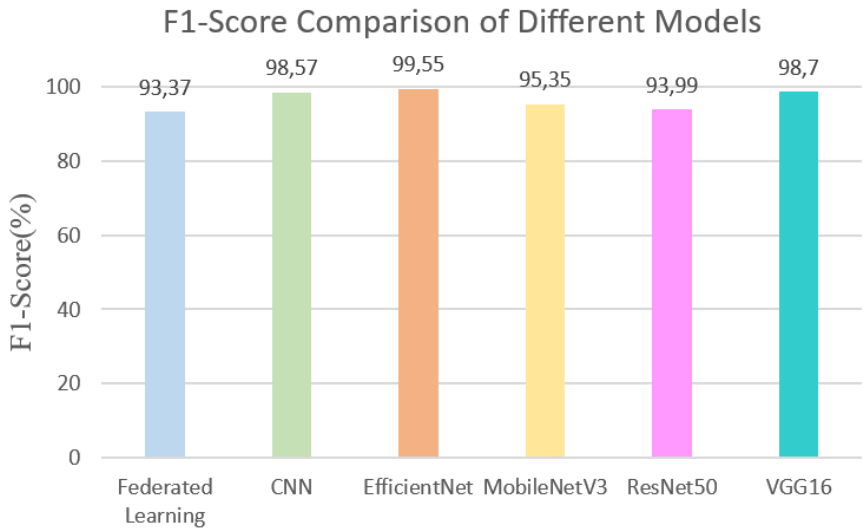

Figure 16. F1_Score Comparison of Different Models

4.5. ROC – AUC Score

The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of a classification model at various threshold values. The ROC curve is generated by plotting the True Positive Rate (TPR) (also known as Recall) against the False Positive Rate (FPR) at different threshold levels.

The Area Under the Curve (AUC) represents the area beneath the ROC curve and measures the overall discriminative ability of the model. The AUC value ranges between 0.5 and 1, where values close to 1 indicate a high discriminative capability, while values near 0.5 suggest that the model performs similarly to random guessing.

The AUC-ROC metric is particularly important for evaluating classification models on imbalanced datasets. A higher AUC value signifies that the model can effectively distinguish between positive and negative classes.

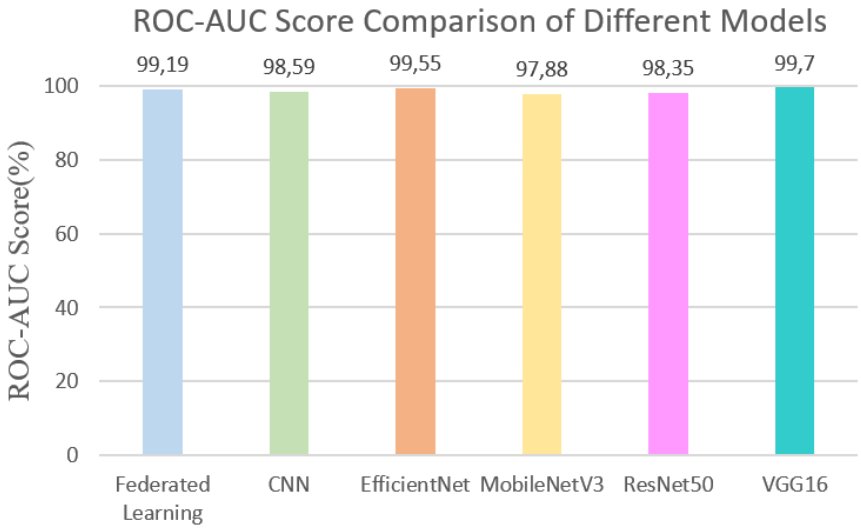

Figure 17. AUC Score Comparison of Different Models

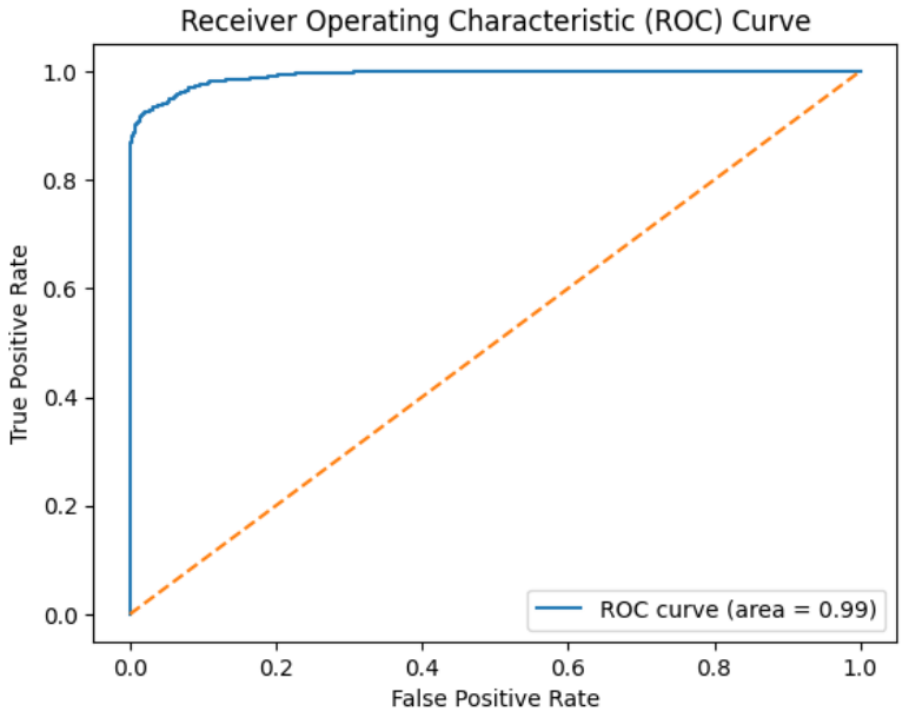

Figure 18. FL ROC Curve

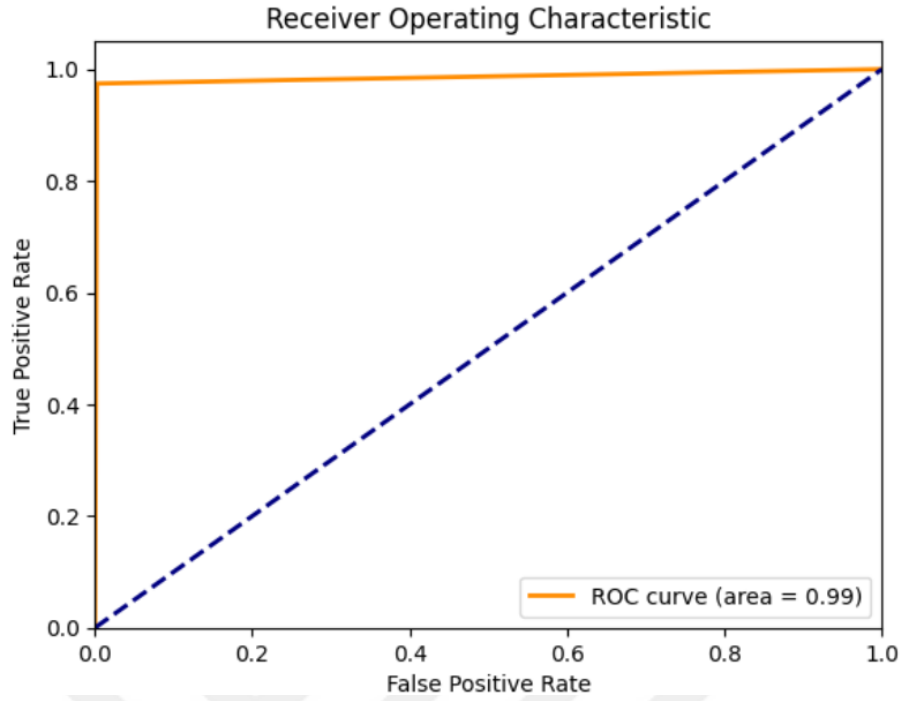

Figure 19. CNN ROC Curve

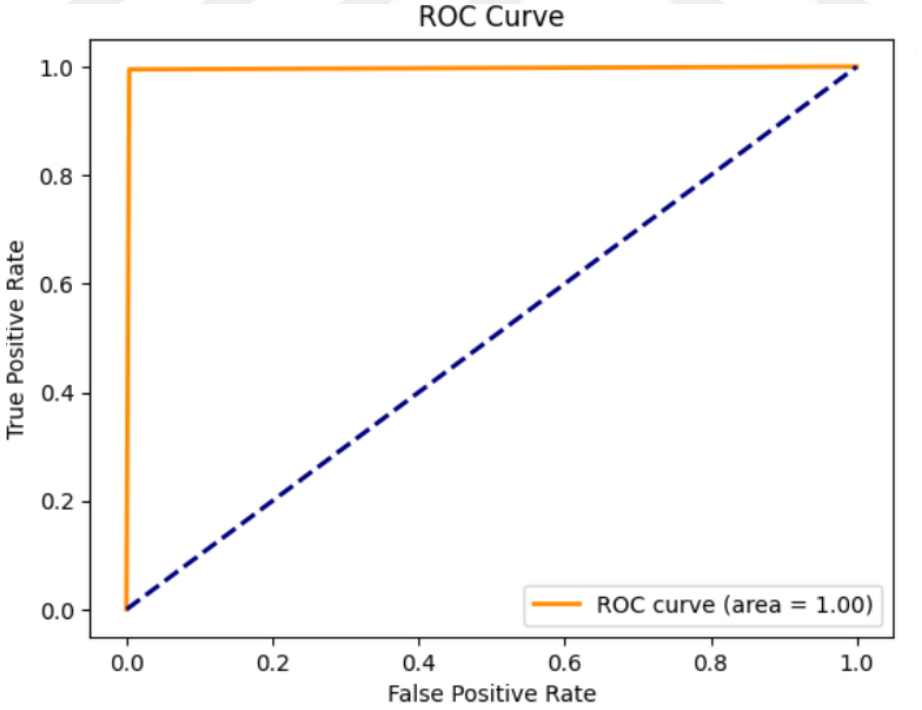

Figure 20. EfficientNet ROC Curve

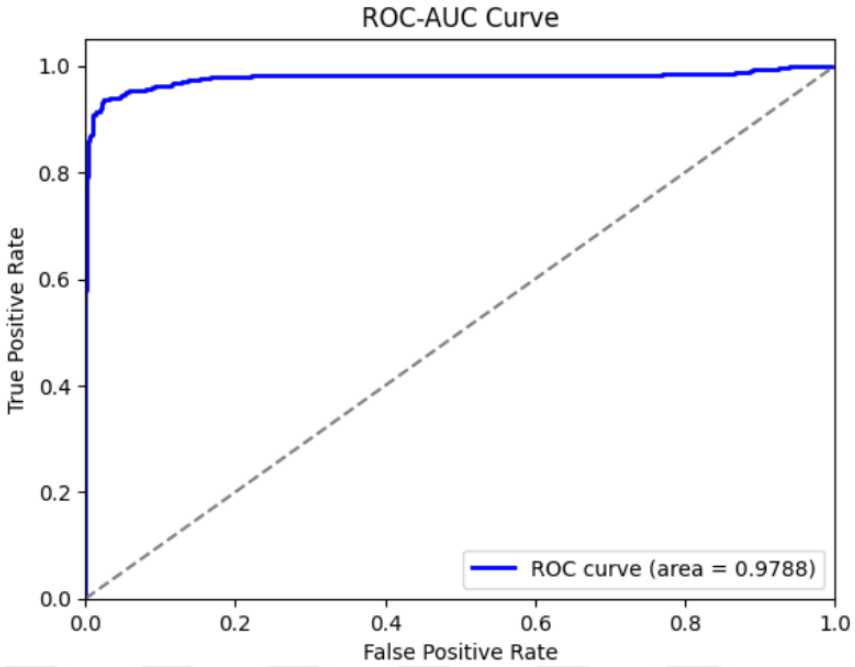

Figure 21. MobileNetV3 ROC Curve

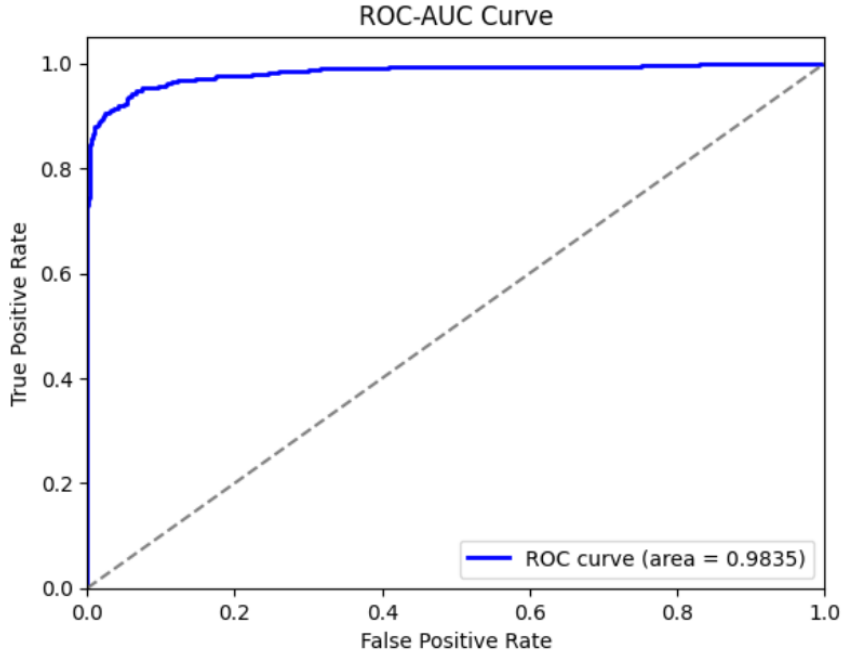

Figure 22. ResNet50 ROC Curve

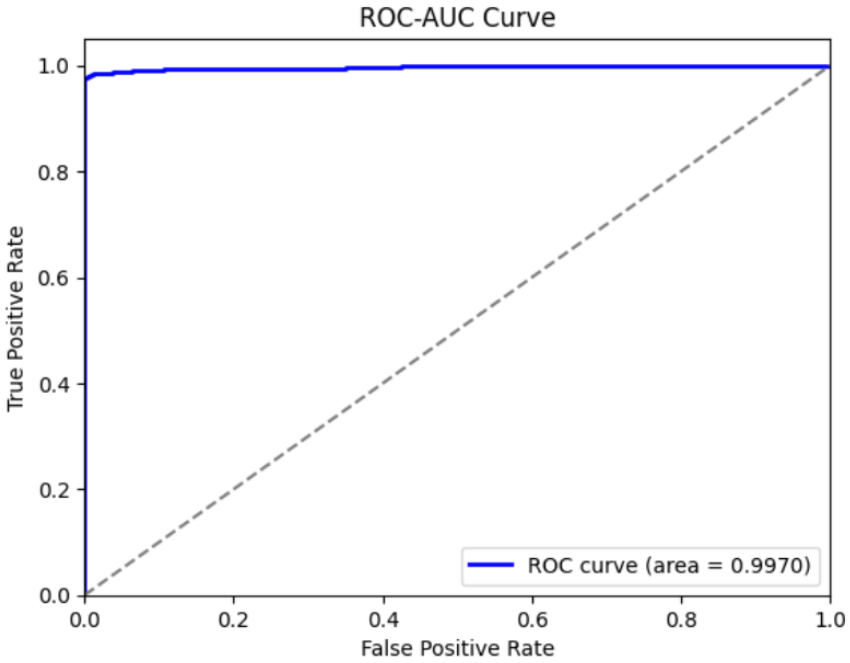

Figure 23. *VGG16 ROC Curve*

Table 7. Classification report

	Accuracy(%)	**Precision(%)**	**Recall(%)**	**F1-Score(%)**	**ROC-AUC (%)**
Federated Learning	93.75	99.71	87.79	93.37	99.19
CNN	98.58	99.74	97.43	98.57	98.59
EfficientNet	99.55	99.61	99.49	99.55	99.55
MobileNetV3	95.43	97.20	93.57	95.35	97.88
ResNet50	94.01	94.54	93.44	93.99	98.35
VGG16	98.71	99.74	97.69	98.70	99.70

5. CONCLUSIONS

The FL-based model has achieved particularly noteworthy results, especially when considering its applications in the healthcare domain. FL provides a significant advantage by enabling secure learning across different datasets without the need to centralize data on a single server, thus ensuring the privacy of healthcare data. In this context, the FL -based model achieved an accuracy of 93.75%, a precision value of 99.71%, and a recall of 87.79%. These results demonstrate that the model provides highly reliable and accurate positive results in the diagnosis of liver tumors, while also effectively differentiating between classes. The F1-Score was 93.37%, and the ROC-AUC value was measured at 99.19%, both of which indicate the high overall performance of the model and its potential for use in critical decision support systems in healthcare.

In comparison, EfficientNet exhibited the highest performance with an accuracy of 99.55%. Its Precision and Recall values are very close, reflecting the model's success in accurately diagnosing tumors. Both VGG16 and CNN models also showed high accuracy rates 98.71% and 98.58% and excellent Precision values 99.74%. However, when compared to the FL model, these models exhibited slightly lower Recall values 97.69% and 97.43%, which is a crucial factor in healthcare applications, where false negatives can have significant consequences. This highlights that FL offers a more balanced and reliable solution.

MobileNetV3 and ResNet50 models, with accuracy rates of 95.43% and 94.01%, respectively, presented the lowest performance. Nevertheless, their Precision and F1-Score values are still at levels suitable for healthcare applications, providing useful results.

In conclusion, FL has great potential, especially in the healthcare sector. While offering significant advantages in terms of data privacy and security, its performance is comparable to traditional DL models. In the diagnosis of liver tumors, the FL -based model provides an effective alternative compared to other models, with high accuracy, Precision, and ROC-AUC values, making it a promising solution for broader use in healthcare applications.

Data Availability This study uses a publicly available dataset from the ATLAS (A Tumour and Liver Automatic Segmentation) challenge. The dataset can be accessed at the URL https://atlas-challenge.u-bourgogne.fr/dataset.

REFERENCES

Albalawi E., Mahesh T.R., Thakur A., Kumar V. V., Gupta, M., Khan S. B., Almusharraf A. (2024). Integrated Approach of Federated Learning With Transfer Learning for Classification and Diagnosis of Brain Tumor. Albalawi et al. BMC Medical Imaging (2024) 24:110.

Atef O., Salam M. A., Abdelsalam H. (2022). Federated Learning Approach for Measuring the Response of Brain Tumors to Chemotherapy. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 10, 2022.

Bernecker T., Peters A., Schlett C. L., Bamberg F., Theis F., Rueckert D., Weib J., Albarqouni S. (2022). FedNorm: Modality-Based Normalization in Federated Learning for Multi-Modal Liver Segmentation. ArXiv preprint arXiv:2205.11096.

Bonawitz K., Ivanov V., Kreuter B., Marcedone A., McMahan H. B., Patel S., Ramage D., Segal A., Seth K. (2017). Practical Secure Aggregation for PrivacyPreserving Machine Learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS ’17). ACM, New York, NY, USA, 1175–1191. https://doi.org/10.1145/3133956.3133982 .

Bucak İ. Ö., Baki S. (2010). Diagnosis of Liver Disease by Using CMAC Neural Network Approach. Volume 37, Issue 9, September 2010, Pages 6157-6164.

Chai H., Huang Y., Xu L., Song X., He M., Wang Q. (2024). A decentralized federated learning-based cancer survival prediction method with privacy protection. Heliyon 10 (2024) e31873.

Chan H. P., Samala R. K., Hajiiski L. M., Zhou C. (2020). Deep Learning in Medical Image Analysis. Doi.org/10.1016/B978-0-12-810408-8.00032-8.

Chen H., Gomez C., Huang C. M., Unberath M. (2022). Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence from a Systematic Review. npj Digit. Med., vol. 5, no. 1, pp. 156-171, 2022.

Çaviş T., Arda K. N. (2024). Advanced Magnetic Resonance Imaging Techniques in the Diagnosis of Liver Diseases. The Turkish Journal of Current Gastroenterology 2024;26:130-141.

Feki I., Ammar S., Kessentini Y., Muhammad K. (2021). Federated learning for COVID-19 screening from Chest X-ray images. Applied Soft Computing 106 (2021) 107330.

Gohari R.J. (2024). FedBrain-Distill: Communication-Efficient Federated Brain Tumor Classification Using Ensemble Knowledge Distillationon Non-IID Data. 10.1109/ICCKE65377.2024.10874662.

Heimbach J. K., Kulik L. M., Finn R. S., Sirlin C. B., Abecassis M. M., Roberts L. R., Zhu A. X., Murad M. H., Marrero J. A. (2018). AASLD Guidelines for the Treatment of Hepatocellular Carcinoma. Hepatology 67(1):p 358-380, January 2018.

Hitaj B., Ateniese G., Cruz F. P. (2017). Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning. CoRR abs/1702.07464 (2017).

Hu Q., He X., Zhou J. (2004). Multi-scale edge detection with bilateral filtering in spiral architecture. Proceedings of the Pan-Sydney area workshop on Visual information processing. 10.5555/1082121.1082126(29-32). Online publication date: 1-Jun-2004.

Islam M., Reza Md. T., Kaosar M., Parvez M. Z. (2023). Effectiveness of Federated Learning and CNN Ensemble Architectures for Identifying Brain Tumors Using MRI Images. Neural Processing Letters (2023) 55:3779–3809.

Konecný J., McMahan H. B., Ramage D., Richtárik P. (2016). Federated Optimization: Distributed Machine Learning for On-Device Intelligence. CoRR abs/1610.02527 (2016). arXiv:1610.02527.

Kumar R., Khan A. A., Kumar J., Zakria, Golilarz N. A., Zhang S., Ting Y., Zheng C., Wang W. (2021). Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging. IEEE SENSORS JOURNAL, VOL. 21, NO. 14, JULY 15, 2021.

Kwak L., Bai H. (2023). The Role of Federated Learning Models in Medical Imaging. Radiology: Artificial Intelligence 2023; 5(3):e230136.

Li F. F., Johnson J., Yeung S. (2017). Detection and Segmentation, CS231n: Convolutional Neural Networks for Visual Recognition,Stanford University, [online] Available at http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf [Accessed 10 June 2019].

Lin Y., Han S., Mao H., Wang Y., Dally W. J. (2017). Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training. CoRR abs/1712.01887 (2017). arXiv:1712.01887 http: //arxiv.org/abs/1712.01887.

Liu B., Yan B., Zhou Y., Yang Y., Zhang Y. (2020). Experiments of Federated Learning for COVID-19 Chest X-ray Images. Doi.org/10.48550/arXiv.2007.05592.

Llovet J. M., Kelley R. K., Villanueva A., Singal A. G., Pikarsky E., Roayaie S., Lencioni R., Koike K., Rossi J. Z., Finn R. S. (2021). Hepatocellular Carcinoma, Nature Rev. Dis. Primers, vol. 7, no. 1, pp. 6-34, 2021 Jan 21;7(1):6.

Lotfinia M., Tayebiarasteh A., Samiei S., Joodaki M., Arasteh S. T. (2025). Boosting multi-demographic federated learning for chest x-ray analysis using general-purpose selfsupervised representations. doi.org/10.48550/arXiv.2504.08584.

Lusnig L., Sagingalieva A., Surmach M., Protasevich T., Michiu O., McLoughlin J., Mansell C., Petris G. D., Bonazza D., Zanconati F., Melnikov A., Cavalli F. (2024). Hybrid quantum image classification and federated learning for hepatic steatosis diagnosis. 14(5):558, 2024, DOI: 10.3390/diagnostics14050558.

Mahlool D. H., Abed M. H. (2022). Distributed Brain Tumor Diagnosis Using a Federated Learning Environment. Vol. 11, No. 6, December 2022, pp. 3313-3321.

Marigliano P. (2024). Analyzing tourism reviews using Deep Learning and AI to predict sentiments. Quality & Quantity. 10.1007/s11135-024-01840-x. McMahan B., Moore E., Ramage D., Hampson S., Arcas B. A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In Artificial Intelligence and Statistics (pp. 1273-1282). PMLR.

McMahan H. B., Moore E., Ramage D., Arcas B. A. Y. (2016). Federated Learning of Deep Networks using Model Averaging. CoRR abs/1602.05629 (2016). ArXiv:1602.05629.

Overman T., Klabjan D. (2025). Continuous-Time Analysis of Federated Averaging. Doi.org/10.48550/arXiv.2501.18870.

Palas I., A., Yousuf M., A. (2024). A Federated Learning-based Model for the Detection of Lung Cancer from CT Scan Images. 10.1109/ICEEICT62016.2024.10534496.

Phong L. T., Aono Y., Hayashi T., Wang L., Moriai S. (2018). Privacy Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Trans. Information Forensics and Security 13, 5 (2018), 1333–1345.

Qayyum A., Ahmad K., Ahsan M. A., Al-Fuqaha A., Qadir J. (2022). Collaborative Federated Learning for Healthcare: Multi-Modal COVID-19 Diagnosis at the Edge. Digital Object Identifier 10.1109/OJCS.2022.3206407.

Roth H.R., Chang K., Singh P., Neumark N., Gupta V., Gupta. S., Qu L., Ihsani A., Bizzo B.C., Wen Y., Buch V., Shah M., Kitamura F., Mendonça M., Lavor V., Harouni A., Compas C., Tetreault J., Dogra P., Cheng Y., Erdal S., White R., Hashemian B., Schultz T., Zhang M., McCarthy A., Yun B.M., Sharaf E., Hoebel K.V., Patel J.B., Chen B., Ko S., Leibovitz E., Pisano E.D., Coombs L., Xu D., Dreyer K.J., Dayan I., Naidu R.C., Flores M., Rubin D., Cramer J.K. (2020). Federated Learning for Breast Density Classification: A Real-World Implementation. Conference paper First Online: 26 September 2020.

Selvakanmani S., Devi G.D., Rekha V., Jeyalakshmi J. (2024). Privacy-Preserving Breast Cancer Classification: A Federated Transfer Learning Approach. Journal of Imaging Informatics in Medicine (2024) 37:1488–1504.

Shankar P.U., Rahul E. S., Rao K.D., Satish K., Kumar U.D., Ravindra D., Subbarao G. (2025). Liver Disease Prediction using Federated Learning. International Journal of Innovative Science and Research Technology, 10(4), 880-887. DOI.org/10.38124/ijisrt/25apr626.

Shao Y.Y., Wang S.Y., Lin S.M., Diagnosis Group, Systemic Therapy Group (2021). Management Consensus Guideline for Hepatocellular Carcinoma: 2020 Update on Surveillance, Diagnosis, and Systemic Treatment by the Taiwan Liver Cancer Association and the Gastroenterological Society of Taiwan. J Formos Med Assoc. 2021 Apr;120(4):1051-1060.

Shokri R., Shmatikov V. (2015). Privacy-Preserving Deep Learning. In Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). ACM, New York, NY, USA, 1310–1321. https: //doi.org/10.1145/2810103.2813687.

Singh A., Pandey B. (2016). Diagnosis of Liver Disease by Using Least Squares Support Vector Machine Approach. International Journal of Healthcare Information Systems and Informatics Volume 11 Issue 2 April-June 2016.

Sivakumar N., Khan A.R., Umar S., Ravikumar R.N., Bremnavas I., Lunagaria M., Vaghela K., Tejani G.G., Sharma S.K. (2025). A Hybrid Brain Tumor Classification Using FL With FedAvg and FedProx for Privacy and Robustness Across Heterogeneous Data Sources. Digital Object Identifier 10.1109/ACCESS.2025.3549440.

Song D., Wang Y., Wang W., Wang Y. (2021). Using Deep Learning to Predict Microvascular Invasion in Hepatocellular Carcinoma Based on Dynamic Contrast-Enhanced MRI Combined with Clinical Parameters.J. Cancer Res. Clin. Oncol., vol. 147, no. 12, pp. 3757-3767, 2021.

Song L., Geoffrey K., Kaijian H. (2020). Bottleneck Feature Supervised U-net for Pixel-Wise Liver and Tumor Segmentation. Expert Syst. Appl., vol. 145, 2020.

Srinivasu P.N., Lakshmi G.J., Narahari S.C., Shafi J., Choi J., Ijaz M.F. (2024). Enhancing Medical Image Classification Via Federated Learning and Pre-trained Model. Egyptian Informatics Journal 27 (2024) 100530.

Tölle M., Garthe P., Scherer C., Seliger J. M., Leha A., Krüger N., Simm S., Martin S., Eble S., Kelm H., Bednorz M., Andre F., Bannas P., Diller G., Frey N., Grob S., Hennemuth A., Kaderali L., Meyer A., Nagel E., Orwat S., Seiffert M., Friede T., Seidler T., Engelhardt S. (2025). Real world federated learning with a knowledge distilled transformer for cardiac CT imaging. Doi.org/10.1038/s41746-025-01434-3.

Triverdi N.K., Shukla S., Tiwari R.G., Agarwal A.K., Gautam V. (2023). Liver Cancer Diagnosis with Lightweight Federated Learning Using Identically Distributed Images. 12th International Conference on System Modeling & Advancement in Research Trends, 22nd–23rd, December, 2023.

Yang Q., Liu Y., Chen T., Tong Y. (2019). Federated Machine Learning: Concept and Applications. Doi.org/10.48550/arXiv.1902.04885.

Zhou L., Wang M., Zhou N. (2024). Distributed Federated Learning-Based Deep Learning Model for Privacy MRI Brain Tumor Detection. Journal of Information, Technology and Policy (2023): 1-12.