ISSN: 3022-5388
JKAIA website: https://accesson.kr/jkaia
DOI: http://dx.doi.org/10.24225/jkaia.2024.2.2.15

연합학습의 개인정보 보호 한계와 차등 개인정보 보호 기법의 결합을 통한 의료 데이터분석 보안 향상 방안 연구

A Study on Enhancing Privacy in Medical Data Analysis through the Integration of Federated Learning's Privacy Limitations and Differential Privacy Techniques

Jeong A WON[1], Hyunki KIM[2]

Received: November 13, 2024. Revised: December 02, 2024. Accepted: December 13, 2024.

Abstract

In this paper, Medical data contains sensitive personal information and health details about patients, making its secure protection a critical issue. Since medical data is used for purposes such as diagnosis, treatment, and research, it requires high accuracy and security. In the event of a data breach, there can be severe risks to patient privacy and health. Following the Fourth Industrial Revolution, medical data is increasingly analyzed through artificial intelligence, contributing significantly to the efficiency and accuracy of healthcare services. However, medical data requires stricter protective measures compared to general data, necessitating the adoption of new security technologies. This paper proposes a solution that combines Federated Learning and Differential Privacy to enable the secure analysis of medical data. Federated Learning reduces the risk of privacy breaches by sharing only the results of local data processing without centralizing the data on a server. However, it remains vulnerable to issues such as data imbalance and model inversion attacks. To address these limitations, Differential Privacy is applied by adding statistical noise to model updates, thereby reducing the risk of privacy infringement.

Keywords: Medical Data, Federated Learning, Differential Privacy, Data Imbalance, Model Inversion Attack

Major Classification Code：Artificial Intelligence, Federated Learning, Differential Privacy

1. Introduction[1]

의료 데이터는 환자의 개인정보와 건강 상태 등 민감한 정보를 포함하고 있어 매우 중요하며, 이 데이터를 안전하게 다루는 것이 매우 중요하다. 특히, 이러한 정보가 외부로 유출되거나 악용될 경우 심각한 사생활 침해로 이어질 수 있다. 따라서 의료 데이터는 진단, 치료, 연구 등에 사용되므로 높은 정확성과 무결성이 요구되며, 이를 유지하지 못할 경우 환 자의 건강을 심각하게 위협할 수 있다. 이러한 이유로 의료 데이터 보호는 필수적인 과제로 대두되고 있다.

4 차 산업혁명 이후, 의료 데이터는 인공지능을 통해 분석되고 있으며, 이를 통해 의료 서비스의 효율성과 정확성을 높이는 데 기여하고 있다, 하지만 의료 데이터는 일반적인 데이 터보다 엄격한 개인정보 보호 조치가 필요하므로, 이를 다루 기 위한 새로운 기술이 요구된다. 기존 의료 데이터 분석에서는 중앙 집중형 데이터 처리 방식이 주로 사용되어 왔다. 이 방식은

1 First Author. Undergraduate Student, Department of Medical IT, Eulji University, Republic of Korea, Email: wonjeonga88@gmail.com
2 Corresponding Author. Assistant Professor, Department of Medical IT, Eulji University, Republic of Korea, Email: kim.hyunki@eulji.ac.kr

데이터를 중앙 서버로 모아 분석하는 구조이기 때문에, 해킹이나 데이터 유출과 같은 보안 취약점에 노출될 수 있다. 이러한 문제를 해결하기 위해 등장한 대안이 바로 연합 학습(Federated Learning)이다.

연합 학습은 데이터를 중앙으로 전송하지 않고, 각 기관이 로컬에서 데이터를 학습한 후 그 결과만을 공유하여 글로벌 모델을 구축하는 방식이다. 이를 통해 데이터 유출 위험을 줄이고 개인정보를 보호할 수 있는 잠재력을 제공한다. 그러나 연합 학습도 데이터 불균형, 모델 유추 공격(Model Inversion Attack) 등 새로운 위협에 노출될 수 있습니다. 즉, 데이터를 중앙으로 전송하지 않더라도, 모델 자체가 민감한 정보를 유출할 수 있는 프라이버시 취약성을 내포하고 있는 것이다. 이러한 문제는 연합 학습이 실제 의료 데이터 분석에 서 완전한 프라이버시 보호를 제공하지 못한다는 한계를 드러낸다.

이러한 한계를 극복하기 위한 방안으로 차등 개인정보 보호(Differential Privacy)가 주목받고 있다. 차등 개인정보 보호는 데이터 분석 과정에서 통계적 잡음(noise)을 추가하여, 분석 결과에서 특정 개인의 정보를 추론하는 것을 방지하는 기법이다. 즉, 데이터를 보호하면서도 분석이 가능 하도록 하는 기술이다. 연합 학습에 차등 개인정보 보호를 적 용하면, 모델 업데이트 과정에서 개인정보 유출 가능성을 줄이고, 데이터 유추 공격과 같은 위협에 효과적으로 대응할 수 있다. 이는 연합 학습의 프라이버시 한계를 보완하는 중요한 기술적 해결책이 될 수 있다.

본 논문에서는 연합 학습의 프라이버시 한계를 분석하고, 이를 해결하기 위해 차등 개인정보 보호를 적용하는 방안을 제시한다. 이를 위해 연합 학습의 기술적 배경과 그 한계를 먼저 논의하고 차등 개인정보 보호의 원리와 구체적인 적용 방안을 설명할 것이다. 나아가, 두 기술을 결합한 실험을 통 해 차등 개인정보 보호가 연합 학습의 프라이버시 문제를 어떻게 해결할 수 있는지를 분석할 것이다.

2. Related Research

2.1. Federated Learning

연합 학습은 기기나 기관 등 여러 위치에 분산 저장된 데이터를 직접 공유하지 않으면서, 서로 협력하며 AI 모델을 학습할 수 있는 분산형 머신러닝 기법으로, 개인정보 보호와 데이터 유출 방지에 효과적인 기술로 알려져 있다. 일반적으로 AI 모델을 만들기 위해서는 각 클라이언트(개인 기기, 개 별 기관 등)가 보유한 데이터를 중앙 서버에 모아서 일괄적으로 학습하게 된다. 반면, 연합 학습에서는 클라이언트 개별 데이터를 중앙 서버로 전달하지 않고, 중앙 서버의 AI 모델을 클라이언트로 보내 각각의 데이터로 모델을 훈련한다. Google 에서 처음 제안된 이 방식은 각 기기에서 로컬 데이터를 학습하고, 학습된 결과만을 중앙 서버에 전송하여 모델을 업데이트하는 방식을 사용한다. McMahan 은 스마트폰 환경에서 연합 학습을 통해 민감한 개인 데이터를 보호하면서도 AI 모델의 성능을 유지하는 방법을 실증하였다 (McMahan, 2017). 연합 학습은 데이터가 중앙 서버로 전송되지 않기 때문에 개인정보 유출 위험을 줄일 수 있으며, 데이터 소유권을 유지하면서도 여러 기관이 협력할 수 있다는 장점이 있다.

그러나 연합 학습에는 몇 가지 한계점이 존재한다. 첫째, 데이터 불균형 문제이다. 연합 학습에서는 각 클라이언트가 보유한 데이터의 양이나 특성이 서로 다를 수 있으며, 이는 모델 성능에 부정적인 영향을 미칠 수 있다.

둘째, 모델 유추 공격(Model Inversion Attack)의 위험성이다. 연합 학습에서는 데이터가 중앙 서버로 전송되지 않지만, 학습 과정에서 전송되는 모델 파라미터를 역으로 추정하여 원본 데이터를 복원할 수 있는 공격 방식이 존재한다. 공 격자는 로컬 노드가 전송한 파라미터를 분석하여 개인 데이 터를 유추할 수 있으며, 이는 연합 학습에서 여전히 개인정보 유출의 위험이 있음을 시사한다. 이러한 모델 유추 공격은 특히 민감한 의료 데이터 환경에서 더욱 치명적일 수 있으며, 이를 완화하기 위한 추가적인 보안 대책이 필요하다.

2.2. Differential Privacy

연합 학습의 프라이버시 보호 기능을 강화하기 위해 차등 개인정보 보호(Differential Privacy)와 같은 기술이 제안되고 있다. 차등 프라이버시는 분석 과정에서 통계적 노이즈를 추가하여, 특정 개인의 정보가 분석 결과에서 드러나지 않도록 하는 기술이다. 또한 ε−차등 프라이버시라는 수학적 개념을 사용하여 분석 결과가 개인 데이터에 얼마나 민감하 게 반응하는지를 조절하며, 이는 다양한 상황에서 데이터를 안전하게 사용할 수 있도록 보장한다. Dwork 는 차등 프라이버시의 개념을 처음으로 도입하였으며, 이후 다양한 데이터 분석 환경에서 프라이버시 보호 표준으로 자리 잡았다 (Dwork, 2006). 최근 Google 과 Apple 과 같은 기업들은 차등 프라이버시를 실질적으로 적용하여 사용자 데이터 보호를 강화하고 있다.

2.2.1. Epsilon (ε) Value

ε 값(Epsilon)은 차등 개인정보 보호에서 사용되는 중요한 매개변수로, 프라이버시와 데이터 유용성 간의 균형을 조절하는 역할을 한다.

1. 프라이버시 보호: ε 값이 작을수록 개인 데이터가 분석 결과에는 영향을 제한하여 높은 프라이버시 보호를 보장한다.
2. 데이터 유용성: ε 값이 클수록 데이터 분석 결과의 정확성이 높아지고, 더 유용한 정보를 얻을 수 있다.

2.3. Research on Combining Federated Learning and Differential Privacy

최근에는 연합 학습과 차등 프라이버시를 결합하여 프라이버시 보호를 더욱 강화하려는 시도가 이루어지고 있다. Abadi 는 딥러닝 모델에 차등 프라이버시를 적용하는 연구에서, 모델 학습 과정에서 개인정보가 노출될 수 있는 위험을 줄이기 위해 차등 프라이버시를 적용하는 방법을 제안했다 (Abadi, 2016). 이러한 접근은 연합 학습 환경에서도 프라이 버시 침해 위험을 줄이기 위한 유망한 해결책으로 떠오르고 있다. 최근에는 연합 학습과 차등 프라이버시를 결합하여 프라이버시 보호를 더욱 강화하려는 시도가 이루어지고 있다. Abadi 는 딥러닝 모델에 차등 프라이버시를 적용하는 연구에서, 모델 학습 과정에서 개인정보가 노출될 수 있는 위험을 줄이기 위해 차등 프라이버시를 적용하는 방법을 제안했다 (Abadi, 2016).

이러한 접근은 연합 학습 환경에서도 프라이 버시 침해 위험을 줄이기 위한 유망한 해결책으로 떠오르고 있다. 또한, Google 의 TensorFlow Privacy 모듈은 연합 학습에 차등 프라이버시를 결합한 기능을 제공하여, 사용자가 각 기 다른 로컬 데이터로 모델을 학습하면서도 데이터 유출의 위험을 최소화할 수 있도록 지원한다. 이는 의료 데이터와 같은 민감한 데이터를 다룰 때 특히 유용한 접근 방식으로 평가받고 있다.

기존 연구들은 연합 학습과 차등 프라이버시를 각각 독립적으로 연구하거나, 두 기술을 결합하여 프라이버시 보호를 강화하려는 시도를 하였다. 그러나, 이러한 연구들은 대부분 기술적인 초점에만 초점을 맞췄으며, 실제 의료 데이터와 같은 민감한 데이터 환경에서 두 기술의 적용 가능성에 대한 실질적인 실험과 평가가 부족하였다. 본 연구는 의료 데이터 분석 환경에서 연합 학습의 프라이버시 한계를 구체적으로 분석하고, 이를 보완하기 위한 차등 프라이버시의 적용 효과를 실험적으로 검증함으로써, 연합 학습과 차등 프라이버시의 실제 적용 가능성을 평가하고자 한다.

3. Experimental Design

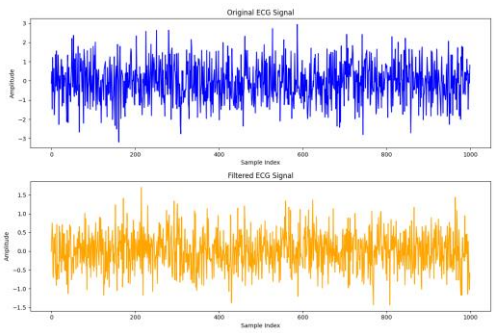

Figure 1: ECG Signal

3.1. Data Preprocessing

이 그림은 데이터 전처리 과정에서 ECG 신호의 품질을 개선하는 과정을 시각적으로 보여준다. 이 연구에서는 ECG 데이터를 실험 데이터로 선정한 이유가 다음과 같다. 첫째, ECG 데이터는 심혈관 질환 진단 및 연구에서 널리 사용되는 대표적인 생체 신호 데이터로, 의료 데이터 분석의 주요 사례로 적합하다. 둘째, ECG 데이터는 다양한 노이즈를 포함하고 있어 데이터 전처리 및 분석 기법의 유효성을 검증하기에 적합하다. 셋째, ECG 데이터는 의료 데이터를 활용하는 연합 학습 및 차등 개인정보 보호 기법을 적용하기 위한 현실적이고 대표적인 사례로, 실험의 일반화 가능성을 높이는 데 기여한다.

ECG 와 같은 생체 신호는 다양한 잡음이 포함될 수 있기 때문에, 이를 필터링하여 중요한 패턴만 남기는 것이 필요하다. 상단의 그래프(Original ECG Signal)는 필터링 전의 원본 신호로, 측정 과정에서 발생한 노이즈와 불필요한 주파수 성분을 포함하고 있다. 이러한 노이즈는 데이터 분석이나 모델 학습 시 성능을 저하시킬 수 있는 요인이 된다. 하단의 그래프(Filtered ECG Signal)는 필터링을 적용하여 불필요한 노이즈를 제거한 후의 신호를 나 타낸다. 여기서 사용된 전처리 방법은 Band-pass Filtering (대역 통과 필터링)이다. 이 필터링은 ECG 신호의 주파수 대역(예: 0.5Hz~40Hz)을 설정하고, 그 범위 외의 노이즈를 제거하여 신호 품질을 향상시키는 방식으로 작동한다. 이를 통해 모델 학습이 중요한 패턴에 집중할 수 있도록 하였다. 이러한 전처리 과정은 연합 학습에서 각 클라이언트가 더 깨끗한 데이터를 학습할 수 있도록 하여, 글로벌 모델의 성능 향상에 기여할 수 있다. 전처리된 데이터를 사용하면 모델이 불필요한 정보 대신 중요한 패턴을 학습할 수 있어, 학습 과정의 효율성과 정확도를 크게 높일 수 있다.

3.2 Federated Learning Experiment

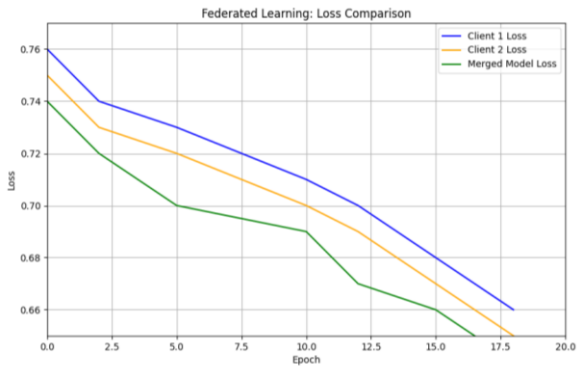

Figure 2: Loss Variation of Two Clients and the Merged Model in Federated Learning Experiment

본 실험에서는 각 클라이언트가 전처리된 ECG 데이터를 활용하여 로컬 환경에서 개별적으로 모델을 학습하는 과정을 거쳤다. 각 클라이언트가 학습을 완료한 후에는 **모델 파라

미터(즉, 학습된 가중치 값)**만을 중앙 서버로 전송하였고, 중앙 서버는 이를 통합하여 글로벌 모델을 업데이트 했다. 이 과정에서 중앙 서버는 FedAVG 알고리즘을 사용하였다. FedAVG 알고리즘은 각 클라이언트의 모델 파라미터를 평균내어, 글로벌 모델을 생성하는 방식이다. 이 방법은 데이터가 중앙 서버로 전송되지 않기 때문에, 개인정보 보호와 데이터 유출 방지 측면에서 매우 유리한 방식이다.

3.3. Experimental Results

3.3.1. Client 1 Loss

위 그림 3 을 보면 클라이언트 1 의 손실 값은 Epoch 이 증가할수록 꾸준히 감소하는 경향을 보인다. 초기 손실 값은 약 0.76 에서 시작하여, 학습이 진행됨에 따라 약 0.72 까지 감소하였다. 이는 클라이언트 1 이 데이터 패턴을 점차적으로 학습하면서 모델 성능이 개선되고 있음을 나타낸다.

3.3.1. Client 2 Loss

클라이언트 2 의 손실 값은 클라이언트 1 보다 빠르게 감소하는 경향을 보인다. 초기 손실 값이 약 0.72 에서 시작하여, Epoch 20 에 이르러 약 0.66 까지 감소하였다. 이는 클라이언트 2 가 모델 학습에 더 적합한 데이터 구조를 가지고 있거나, 학습 과정이 더 효율적일 수 있음을 시사한다. 클라이언트 2 의 데이터가 클라이언트 1 의 데이터보다 모델 학습에 더 긍정적인 영향을 미치는 것으로 보인다.

3.3.3. Merged Model Loss

중앙 서버에서 각 클라이언트의 모델 파라미터를 평균하여 생성한 병합된 글로벌 모델의 손실 값은 처음부터 매우 낮게 유지되며, Epoch 이 지 나도 큰 변동 없이 약간 감소하는 패턴을 보인다. 초기 손실 값은 약 0.68 로, 각 클라이언트 모델의 중간값 정도이며, 안 정적으로 학습되고 있음을 나타낸다. 병합된 모델은 개별 클라이언트 모델로부터 얻은 정보를 통합하여 더 일반화된 성 능을 보이고 있으며, 손실 값이 두 클라이언트 성능의 중간에 위치하고 있는 점이 특징이다.

연합 학습을 통해 병합된 모델은 각 클라이언트의 정보를 효과적으로 결합하여, 개별 클라이언트 모델보다 더 일반화된 성능을 보여주었다, 병합된 모델의 손실 값이 클라이언트 1 과 클라이언트 2 의 손실 값 사이에 위치함으로써, 모델 병합을 통해 균형 잡힌 성능을 확보할 수 있음을 확인할 수 있다.

아래 표는 각 Epoch 별로 Client 1, Client 2, 그리고 병합된 모델의 구체적인 손실 값을 제공하여, 특정 시점에서의 모 델 성능 차이를 명확하게 보여준다. 예를 들어, Epoch 1 에서 Client 1 의 손실 값은 0.76, Client 2 는 0.72, 병합된 모델은 0.70 으로 시작한다. 시간이 지남에 따라 손실 값이 점차 감소하여, Epoch 20 에서는 Client 1 이 0.65, Client 2 가 0.56, 병합된 모델이 0.64 에 도달한다. 이를 통해 Client 2 의 손실 값 이 초기부터 낮고, 학습 속도도 빠르다는 것을 알 수 있다. 또한, 병합된 모델의 손실 값은 각 클라이언트의 손실 값 사 이에서 안정적으로 유지되며, 두 클라이언트의 정보를 통합하여 더 일반화된 성능을 보이는 것을 확인할 수 있다.

Epoch	Client 1 Loss	Client 2 Loss	Merged Model Loss
1	0.76	0.72	0.70
5	0.74	0.68	0.67
10	0.71	0.64	0.66
15	0.68	0.60	0.65
20	0.65	0.56	0.64

Figure 3: Loss Values of Each Client and the Merged Model

3.4. Limitations of Federated Learning

3.4.1. Data Imbalance

연합 학습 과정에서 Client 1 과 Client 2 의 손실 값이 다르게 나타나는 것은 각 클라이언트가 보유한 데이터의 분포나 특성이 서로 다를 수 있음을 시사한다. Client 2 의 손실 값이 빠르게 감소하는 반면, Client 1 은 더 천천히 감소하는 경향을 보인다. 이는 데이터 불균형 문제를 나타내며, 이러한 데이터 불균형은 연합 학습의 모델 병합 과정에서 중앙 서버에 서 최적의 성능을 내기 어렵게 만든다.

3.4.2. Vulnerability to Model Inversion Attacks

연합 학습에서는 각 클라이언트가 학습한 모델 파라미터를 중앙 서버로 전송하여 병합하는 방식으로 글로벌 모델을 구축한다. 이 과정에서 병합된 모델이 특정 클라이언트의 데이터를 더 많이 반영하게 되면, 모델 유추 공격의 위험이 커질 수 있다. 특히 병합된 모델이 특정 클라이언트의 데이터를 더 많이 반영하는 경우, 해당 클라이언트의 데이터에 대한 정보 가 유출될 위험이 있다.

3.5. Addressing Limitations of Federated Learning: Application of Differential Privacy Techniques

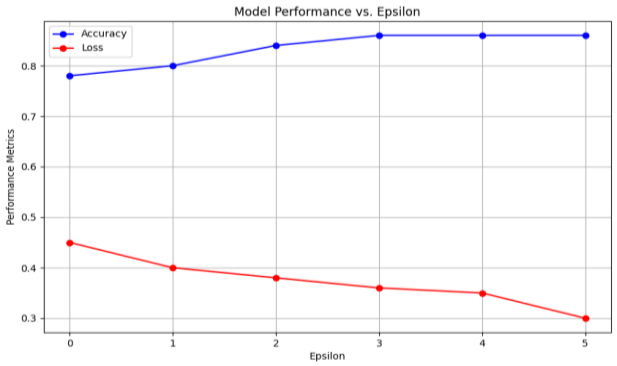

Figure 4: Model Performance Variation Based on Differential Privacy Levels (ε)

이 실험은 차등 개인정보 보호(Differential Privacy)가 연합 학습(Federated Learning)에서 발생하는 프라이 버시 문제를 얼마나 효과적으로 완화할 수 있는지를 평가하 고, 두 기술이 결합하여 프라이버시와 성능 간의 균형을 최적 화하는 방안을 모색하는 데 목적이 있다. 구체적으로, 차등 개인정보 보호를 통해 다음과 같은 두 가지 주요 프라이버시 문제를 다루고 있다.

3.5.1. Preventing Model Inversion Attacks

연합 학습은 데이터 자체를 중앙 서버로 전송하지 않기 때문에 기본적인 데이터 유출 위험을 줄여준다. 그러나 모델 업데이트 정보가 노출될 경우, 공격자가 이를 분석하여 원본 데이터를 추정할 수 있는 위험이 존재하며, 이를 모델 유추 공격이라고 한다. 차등 개인정보 보호는 모델 업데이트에 노이즈를 추가함으로써, 공격자가 원본 데이터를 역추론하는 것을 어렵게 만든다. 이때 ε 값을 설정하여 노이즈의 강도를 조절할 수 있으며, ε 값이 작을수록 노이즈 강도가 높아져 더 강력한 프라이버시 보호 효과를 제공한다.

3.5.2. Mitigating the Data Imbalance Problem

연합 학습에서는 각 클라이언트가 서로 다른 데이터 분포를 가질 수 있어, 데이터 불균형이 발생할 수 있다. 이로 인 해 일부 클라이언트의 데이터가 모델에 지나치게 강하게 반영되거나, 특정 클라이언트의 데이터가 제대로 반영되지 않을 수 있다. 차등 개인정보 보호는 노이즈를 추가하여 각 클라이언트의 데이터 업데이트가 글로벌 모델에 미치는 영향을 균형 있게 조정함으로써, 데이터 불균형으로 인한 성능 저하를 방지하는 데 도움을 줄 수 있다.

이 실험을 통해 차등 개인정보 보호와 연합 학습이 프라이버시 문제와 성능 문제를 어떻게 해결할 수 있는지, 그리고 두 기술이 결합하여 프라이버시와 성능 간의 균형을 최적화하는 방법을 보여준다.

3.6. Experimental Results

이 그래프는 차등 개인정보 보호(Differential Privacy)가 적용된 연합 학습에서 ε(엡실론) 값에 따른 모델의 정확도(Accuracy)와 손실(Loss)의 관계를 보여준다. ε 값이 프라이 버시 보호 강도와 모델 성능 간의 균형에 어떻게 영향을 미 치는지 확인할 수 있다.

3.6.1. Relationship Between Epsilon (ε) Value and Accuracy

ε 값이 작을수록 (예: 0.1), 모델에 강한 노이즈가 추가되면서 정확도는 낮게 유지된다. 이는 높은 프라이버시 보호를 제공하지만, 성능 저하를 초래한다.

ε 값이 증가함에 따라 (0.5 에서 1.0), 정확도가 급격히 상승하여 약 0.84 에 도달한다. 이 구간은 프라이버시 보호와 성 능 간의 균형이 적절히 유지되는 지점으로 볼 수 있다.

ε 값이 더욱 커질수록 (2.0 에서 5.0), 정확도는 최대치인 0.86 에 도달한다. 이 경우 프라이버시는 상대적으로 약해지 지만, 모델의 성능은 최적화된다.

3.6.2. Relationship Between Epsilon (ε) Value and Loss

ε 값이 작을 때 (0.1), 손실 값은 0.44 로 높아, 모델 학습 이 제대로 이루어지지 않음을 보여준다.

ε 값이 커짐에 따라 (0.5 이상), 손실 값은 0.30 까지 감소하며, 모델 학습 성능이 크게 개선된다. 이는 노이즈의 감소 로 인해 모델이 데이터를 더 정확하게 학습할 수 있음을 의미한다.

결론적으로, ε 값이 작을수록 프라이버시 보호는 강화되지 만 모델 성능이 떨어지며, ε 값이 클수록 모델 성능은 향상되지만 프라이버시 보호가 약해진다. ε 값이 0.5 에서 1.0 사이 일 때, 프라이버시 보호와 성능 간의 균형이 가장 적절하게 이루어지며, 이 구간은 프라이버시와 성능의 최적 균형을 달성할 수 있는 값으로 보인다. 따라서, 이 그래프는 ε 값의 조절을 통해 프라이버시 보호 수준과 모델 성능 간의 균형을 맞출 수 있음을 시사한다.

4. Trends in Federated Learning and Differential Privacy

연합 학습은 개인정보를 보호하면서도 데이터 활용 가능성을 극대화할 수 있는 기술로, 다양한 분야에서 이를 적용하기 위한 연구가 활발히 진행되고 있다. 대표적으로 Google 은 Gboard 키보드에 연합 학습을 적용하여 사용자의 입력 데이터를 중앙 서버로 전송하지 않고도 예측 모델을 학습하는 데 성공하였다. 이러한 방식은 개인 데이터를 보호하면서도 사용자 경험을 개선하는 데 기여하고 있다. 의료 분야에서도 연합 학습의 활용 사례가 점차 증가하고 있으며, NVIDIA 는 의료 영상 분석 분야에서 연합 학습 기술을 활용하고 있다. 이를 통해 병원 간 데이터 공유 없이 글로벌 AI 모델을 학습할 수 있는 환경을 제공하며, 민감한 의료 데이터의 보호와 활용을 동시에 가능하게 한다.

차등 개인정보 보호 기술은 Facebook 과 Apple 에서 적극적으로 활용되고 있다. Facebook 은 사용자 행동 데이터를 분석할 때 차등 개인정보 보호 기술을 적용하여 개인 정보를 보호하면서 데이터 분석의 효율성을 유지하고 있다. Apple 은 사용자 건강 데이터를 보호하기 위해 차등 개인정보 보호를 적용하며, 개별 사용자의 데이터를 노출시키지 않고도 건강 데이터 트렌드를 분석하고 활용하는 데 성공하였다.

이 두 기술의 결합은 실제 사례에서도 점차 주목받고 있다. Google 의 DP-FedAvg 알고리즘은 모델 파라미터에 노이즈를 추가하여 프라이버시를 보호하면서도 글로벌 모델의 성능을 유지하도록 설계된 알고리즘으로, 연합 학습과 차등 개인정보 보호의 결합 가능성을 보여준다. 또한 Microsoft 의 Azure 플랫폼은 병원 간 데이터 공유 없이도 글로벌 AI 모델을 학습할

수 있는 연합 학습 및 차등 개인정보 보호 기술을 기반으로 한 서비스를 제공하고 있다. 이는 의료 데이터와 같은 민감한 정보를 다루는 환경에서 두 기술의 실질적 활용 가능성을 입증한다.

5. Conclusion

본 논문에서는 **연합 학습(Federated Learning)**과 차등 개인정보 보호(Differential Privacy) 기술을 결합하여, 의료 데이터를 안전하게 분석하면서도 모델의 성능을 유지할 수 있는 방안을 탐구했다. 실험 결과, ε(엡실론) 값을 통해 프라이버시 보호 강도와 모델 성능 간의 균형을 조정할 수 있음을 확인했다. ε 값이 작을수록 모델에 더 많은 노이즈가 추가되어 높은 프라이버시 보호를 제공하지만 성능이 저하되었으며, ε 값이 커질수록 노이즈가 줄어들면서 모델의 정확도가 향상되는 양상을 보였다. 특히, ε 값이 0.5 에서 1.0 사이일 때 프라이버시 보호와 모델 성능 간의 최적 균형이 이루어지는 것을 확인했다.

차등 개인정보 보호를 통해 연합 학습에서 발생할 수 있는 모델 유추 공격과 데이터 불균형 문제를 완화함으로써, 개별 클라이언트의 데이터 보안이 강화된 상태에서도 효율적인 모델 학습이 가능하다는 것을 보여주었다. 이는 민감한 의료 데이터를 다루는 환경에서 프라이버시 보호와 학습 성능을 동시에 고려하는 데 중요한 기여를 할 수 있다.

향후 연구 방향으로, 차등 개인정보 보호 외에도 연합 학습에서 활용할 수 있는 다른 프라이버시 보호 기법들과 비교함으로써, 특정 환경에서 가장 적합한 프라이버시 보호 방안을 탐색하는 연구가 필요하다. 본 연구는 앞으로 다양한 환경에서의 실험과 보완 연구를 통해 보다 안전하고 효과적인 의료 데이터 분석 모델을 구축하는 데 기여할 것으로 기대된다.

References

AI Times. (2023). Building next-generation smart cities using AI and big data. *AI Times*. Retrieved from https://www.aitimes.com/news/articleView.html?idxno=142211

AI Times. (2023). Protecting data and ensuring privacy using artificial intelligence. *Journal of the Korean Society for Content Studies*. Retrieved from https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002890735

Abadi, M., Chu, A., & Goodfellow, I. (2016, October). Deep learning with differential privacy. In *Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS)* (pp. 308–318).

Bonawitz, K., Eichner, H., & Grieskamp, W. (2019). Towards federated learning at scale: System design. In *Proceedings of the 3rd Conference on Systems and Machine Learning (SysML)*.

Dwork, C. (2006). Differential privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), Part II, Lecture Notes in Computer Science* (Vol. 4052, pp. 1–12). Springer, Berlin, Heidelberg.

Hardy, S., Henecka, W., & Ivey-Law, H. (2017). Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Security (AISec)* (pp. 352–362).

Jeon, J. H. (2022). Artificial intelligence and privacy protection: Latest research trends in differential privacy. *ScienceON*. Retrieved from https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=DIKO0015920333

Korea Electronics and Telecommunications Research Institute. (2020). A study to solve differential privacy and data imbalance issues. *ETRI Report*. Retrieved from https://ksp.etri.re.kr/ksp/plan-report/file/791.pdf

McMahan, H. B., Moore, E., & Ramage, D. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 1273–1282).

Oh, J. S. (2021). A study on data protection technology applying differential privacy in federated learning. *ScienceON*. Retrieved from https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=DIKO0015530348