联邦学习中的差分隐私：使用随机响应缓解推理攻击

Ozer Ozturk,
Department of Computer Engineering
Marmara University
Istanbul, Turkiye
ozer.ozturk@marun.edu.tr

Busra Buyuktanir
Department of Computer Engineering
Marmara University
Istanbul, Turkiye
busra.buyuktanir@marmara.edu.tr

Gozde Karatas Baydogmus
Department of Computer Science
Loyola University of Chicago
Chicago, USA
gkaratasbaydogmus@luc.edu

Kazim Yildiz
Department of Computer Engineering
Marmara University
Istanbul, Turkiye
kazim.yildiz@marmara.edu.tr

ABSTRACT

用于由服务器和客户端组成的分布式架构的机器学习模型需要大量数据以实现高精度。从客户端获得的数据被收集在中央服务器上进行模型训练。然而，在中央服务器上存储数据引发了安全性和隐私性的担忧。为了解决这个问题，提出了联邦学习架构。在联邦学习中，每个客户端使用自己的数据训练一个本地模型。这些训练好的模型定期传输到中央服务器。然后，服务器通过使用联邦聚合算法结合接收到的模型以获得全局模型。该全局模型被分发回客户端，并且这个过程以循环的方式继续进行。尽管防止数据离开客户端增强了安全性，但仍存在某些担忧。攻击者可以对获取的模型执行推理攻击来近似训练数据集，可能导致数据泄露。在本研究中，应用了差分隐私来解决上述安全漏洞，并进行了性能分析。使用基于关联的数据无意识分类（duCBA）算法作为联邦聚合方法。通过随机响应技术在数据上实现了差分隐私，并且检查了不同 epsilon 值下的安全性和性能之间的权衡。随着 epsilon 值的减小，模型精度下降，并观察到类别预测不平衡。这表明更高的隐私级别并不总是导致实用结果，并且必须仔细考虑安全性和性能之间的平衡。

Keywords 联邦学习 · 差分隐私 · 随机响应 · 数据隐私 · duCBA

1 介绍

互联网连接设备数量的迅速增加导致数据产量显著提升 [1]。物联网 (IoT)、移动应用和社会媒体平台等来源在任何时候都会生成大量的数据。解释这些大量数据正变得越来越困难，需要使用基于机器学习的方法 [2]。机器学习模型需要大量的数据来达到足够的准确性和性能。处理大数据不仅需要强大的处理能力和存储容量，还会带来诸如数据完整性、访问速度、安全和隐私等各种挑战。

在由客户端和服务器组成的分布式系统中，传统的机器学习方法涉及从客户端收集数据并将其传输到中央服务器，在这些数据上执行模型训练 [3]。在这种结构中，需要将大量数据传输、存储和处理在中央服务器上的要求带来了显著的成本。然而，大数据结构在隐私保护和数据安全方面存在重大风险 [3]。由于隐私和安全漏洞，组织不仅遭受财务损失，还面临声誉和信任的丧失。个人对个人信息隐私日益敏感以及由此产生的伦理问题使得这种方法难以持续并增加实施成本。

作为解决上述问题的方案，联邦学习（FL）架构于 2016 年被开发出来 [4]。联邦学习是一种机器学习方法，它消除了集中收集数据的需求，并能够利用分布式系统中客户端本地存储的数据进行模型训练。该方法基于定期将客户端上本地训练的模型发送到中央服务器的原则，在那里使用适当的联邦聚合算法创建全局模型。然后将全局模型重新分发给客户端，整个过程迭代继续。因此，消除在中央服务器上收集和存储数据的需求，保护隐私并增强数据安全性。此架构的示意图如图 1 所示。

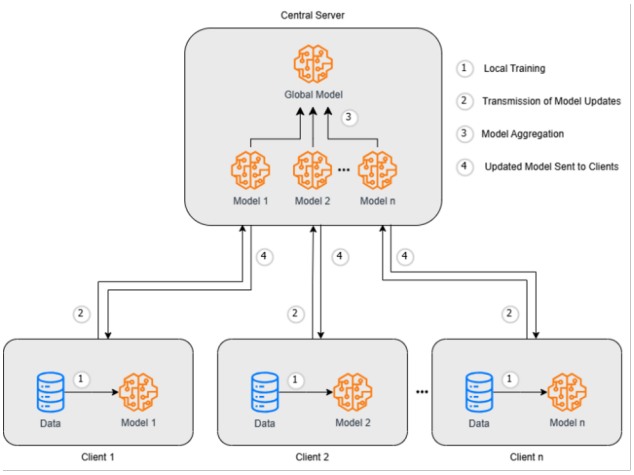

图 1: 联邦学习架构的工作原理

尽管 FL 架构消除了数据共享并创建了一个更安全的系统，但安全问题仍然存在。对数据没有完全控制权，并且由于分布式架构，控制机制被削弱了。这种情况催生了一种新的攻击方式，称为推理攻击 [5]。在推理攻击中，攻击者使用他们获得的模型和特定输入数据进行测试，并可以根据模型输出 [6] 对训练数据做出推断。此外，最近的文献还研究了机器遗忘技术作为解决联邦学习系统中的隐私和合规性挑战的一种方法 [7]。这使他们能够以较高的准确率预测模型是用什么数据进行训练的，甚至可以推测出一些敏感数据样本的内容。这种情况削弱了联邦学习保护数据隐私的说法，并提出了需要额外安全措施的需求。

本研究旨在通过整合安全机制来评估联邦学习架构的性能。为此，选择了基于关联的数据不可知分类（duCBA）[8] 方法作为聚合算法，并开发了一个原型以证明所提出结构的可行性。研究所使用的数据集被构造成客户端数据；应用了随机响应（RR）方法以确保客户端的隐私。因此，在保持客户端级别隐私的同时，分析了对模型准确率的影响，并详细考察了隐私与性能之间的权衡。

本研究的其余部分结构如下：第二部分回顾了相关文献。第三部分详细介绍了研究所使用的方法和数据集。第四部分分享了研究结果，而第五部分则讨论了这些结果，包括研究的贡献、局限性以及对未来研究的建议。

2 相关工作

与集中式学习相关的高通信成本、数据隐私问题以及可持续性挑战，已经促使机器学习领域开发新的方法。在此背景下，谷歌提出的联合学习架构能够在不进行集中式数据收集的情况下，在分布式客户端上训练本地模型，从而确保隐私保护和降低通信负载 [4, 9]。在联合学习过程中，客户端上训练的模型参数会在特定时间

间隔传输到中央服务器；通过将这些模型合并以创建更新后的全局模型，并重新分发给客户端。因此，该过程会迭代进行。

在联邦学习系统中，模型性能很大程度上取决于联邦聚合算法的有效性 [10]。这些算法根据特定策略合并从客户端接收的更新后的模型参数以生成新的全局模型。文献中突出的主要算法如下：

- **FedAvg（联邦平均）：** 这是最常用的聚合方法。在客户端训练的模型权重根据每个客户端的数据量加权平均，以创建一个全局模型。然而，在异构（非 IID）数据分布情况下，这种方法可能导致性能波动 [4, 11]。
- **FedPer（带有个性化层的联邦学习）：** 它只允许模型的下层（基础层）共享，而上层（决策层）保留在客户端。这实现了个性化，但全局模型的泛化能力可能受限 [12]。
- **联邦匹配平均（FedMA）：** 基于层次的神经元匹配方法将相似的神经元分组在一起。这提供了一种更有结构和意义的组合；然而，该方法以其高计算成本著称 [12]。
- **FedDist（联邦距离）：** 通过使用欧氏距离分析客户端中神经元之间的差异，将显示出显著偏差的神经元添加到全局模型中。这保留了特定于客户端的信息，同时也提高了全局模型的泛化能力。随着动态结构更新，在通信轮次 [12] 中，模型架构可以进化。
- **联邦支持向量机（FedSVM）：** 这是一种联邦学习算法，可以在分布式的客户端上使用本地数据训练 SVM 模型，并在中央服务器上合并这些模型的权重。该方法允许在无需共享数据 [13] 的情况下更新全局分类模型。
- **杜 CBA：** 本研究中使用的 duCBA 聚合算法是对传统基于关联的分类（CBA）[14, 15] 算法在联邦学习环境中的适应。duCBA 允许客户端根据其本地数据生成带有类别标签的关联规则，然后这些规则在一个中央服务器上融合以创建全局分类模型 [8]。

尽管联邦学习系统在数据隐私方面提供了比传统机器学习方法更安全的结构，但并不能说安全性完全得到保证 [16, 17]。这些系统仍然容易受到各种类型的攻击。在文献中，针对联邦学习的攻击主要分为两大类：模型中毒攻击 [18] 和模型推断攻击 [19]。

模型中毒攻击是指攻击者通过将故意操纵的模型注入联邦学习过程中，以降低全局模型的准确性或使其朝着有利于自己的方向发展的攻击。在这样的攻击中，恶意客户端会故意生成有害更新，并将这些更新纳入中心聚合过程 [18]。

相比之下，模型推理攻击旨在获取有关模型训练数据的信息 [19]。在此背景下，两种主要类型的攻击尤为突出：

1. *成员推断攻击*：在这种类型的攻击中，攻击者试图推断特定的数据样本是否包含在目标机器学习模型的训练集中。这通常是通过分析模型输出中的不确定性水平（例如，类别概率或熵）来实现的。模型对训练数据的过拟合增加了其对此类攻击的脆弱性 [20]。
2. *模型反转攻击*：此攻击旨在根据模型对特定类别的输出重构属于某类的典型输入数据。通常通过迭代优化输入数据以最大化类别概率来实现。该方法主要由基于梯度的优化技术或潜在空间反投影方法支持 [21]。

对抗这些攻击的主要防御方法之一是差分隐私 [22] 原理。差分隐私是一种隐私定义，它在数学上保证第三方无法可靠地判断某个个体是否包含在一个数据集中。这一原理可以通过拉普拉斯 [23] 和高斯 [24] 机制来实现。这两种机制都是为数值数据设计的，难以直接应用于分类数据。随机响应方法 [25] 是一种适用于分类数据的合适替代方案，最早由沃纳于 1965 年提出，并被用于保护调查研究中个体的隐私。通过今天开发的方法，随机响应方法可以在差分隐私框架内广泛应用于确保分类数据的隐私。

3 方法论

本节详细阐述了研究中采用的方法和方法论。

3.1 使用的技术和开发环境

本研究使用 Python 编程语言开发，并且优先使用 Anaconda 分发包中包含的 Spyder 集成开发环境 (IDE)。应用程序实现了 pyarc、numpy、pandas、scikit-learn 和 pyfim 库。这些库在数据预处理、关联规则推理以及分类过程中得到了有效利用。

3.2 数据集

本研究使用了由 Prosper Chuks 编译并在 Kaggle 数据共享平台上提供的“糖尿病、高血压和中风预测”数据集中的高血压子集，用于高血压预测 [26]。该数据集基于美国疾病控制与预防中心（CDC）进行的 2015 年行为风险因素监测系统（BRFSS）调查，并已通过预处理步骤提供给研究人员。

所使用的子集包含大约 26,000 个样本和 14 个属性。目标变量高血压的类别分布被观察到为阳性占 45.3%（n=11,809），阴性占 54.7%（n=14,274）。根据这种分布计算出的不平衡比率（IR≈1.21）低于文献中普遍接受的 1.5 的阈值，因此数据集被认为是平衡 [27] 的。此外，该数据集中没有缺失观测值，并且其结构可以直接应用于机器学习模型。表 1 提供了研究中使用的数据集的汇总信息。

表 1: 关于数据集的特征信息

列名	数据类型	解释
age	Numerical	Age
sex	Categorical	Gender
cp	Categorical	Type of chest pain
trestbps	Numerical	Resting blood pressure
chol	Numerical	Serum cholesterol
fbs	Categorical	A condition where fasting blood sugar is greater than 120 mg/dl
restecg	Categorical	Resting ECG results
thalach	Numerical	Maximum heart rate
exang	Categorical	Exercise-induced chest pain
oldpeak	Numerical	Exercise-induced ST segment depression
slope	Categorical	The slope of the ST segment during exercise
ca	Categorical	Number of large coronary arteries visualized by fluoroscopy
thal	Categorical	Myocardial perfusion status determined by thallium stress test
target	Categorical	The presence or absence of hypertension

3.3 数据预处理

为了确保模型训练得以正确进行，在本研究范围内应用了多种数据预处理步骤。首先，从数据集中移除了含有缺失值的观测值，从而防止因缺失数据而产生的偏差。由于 duCBA 算法只能处理分类数据，因此将数值变量适当地转换为分类形式。在此背景下，对于不适合直接使用的 thalach（最大心率）变量，通过利用该变量与年龄之间的相关性，创建了一个名为 thalach_ 比率的新派生变量。该变量被划分成特定区间并转换成了分类格式。

卡方独立性检验被应用于确定数据集中的自变量与目标变量 [28] 之间的关系。通过这一检验，未显示出与目标变量有统计显著关系的年龄和性别变量从数据集中移除。得益于这些预处理步骤，数据集符合了 duCBA 算法的要求，并在模型训练前确保了结构完整性。

3.4 联邦学习架构和 duCBA 合并算法

在本研究中，设计了一种基于联合学习架构的方法论，适用于边缘设备如移动设备、物联网系统和客户端-服务器基础的协作结构，并在此范围内应用了 duCBA 联合聚合算法。为了使 duCBA 算法能够运行，客户使用其本地数据的 CBA 算法生成基于分类规则的模型。CBA 算法在监督式数据挖掘的范围内运作，根据特定的支持度和置信度阈值产生标记关联规则。一旦训练完成，客户端仅将由标记规则组成的结构传输给中央服务器；数据本身并未发送到服务器。

这些在服务器上收集的模型是通过特定于 duCBA 算法 [8] 的合并模块集成的。该模块根据内容和标签比较来自客户端的规则，更新相同规则的支持度和置信度值，并且当匹配规则具有不同标签时优先选择支持度值较高的规则。在相等的情况下，则优先考虑先到达的规则。所有规则都按照其置信度值进行排序；如果置信度值相等，则根据支持度值和顺序应用次级排序。生成的规则列表代表服务器上创建的最终模型。在此过程中，开发了特殊公式来根据加权平均值更新支持度和置信度值。

在本研究范围内，创建了一个仿真环境来从隐私和安全的角度评估 duCBA 算法。所使用的数据集被分为两部分：80%用于训练，20%用于测试。训练数据被等分，并假设来自不同的客户端，进行了端点的模拟。例如，在有两个客户端的场景中，训练集被随机分成两个大致相等的部分。每一部分分别使用 CBA 算法进行训练，只将生成的模型（即标记规则）发送到服务器。最终模型通过使用 duCBA 算法合并这些规则获得。

在模型训练期间，支持阈值和置信度阈值分别设置为 0.02 和 0.5。最终的模型在测试数据上进行了评估，并执行了性能分析。

3.5 差分隐私

在 duCBA 架构中，客户端的每个单元都使用其本地数据基于分类规则通过 CBA 算法生成模型。然而，在此过程中，客户端数据的匿名性面临来自模型推理攻击等隐私威胁的风险。作为对抗这种威胁的对策，已将差分隐私机制集成到客户端。

在这项研究中，选择了特别适用于分类数据的差分隐私应用的随机响应方法。该方法在模型训练前应用于客户端数据，以实现敏感信息的隐藏。其目的是通过基于随机性原则随机扭曲每个数据样本的准确性来防止识别个别参与者。

差分隐私机制的实现可能会通过在一定程度上损害数据完整性而负面影响模型的分类性能。在这种情况下，隐私级别和模型性能之间存在权衡。这种权衡是通过差分隐私参数 epsilon (ϵ) 值来控制的。随着 ϵ 值的减小，系统中的隐私级别增加；然而，这种增加可能会导致准确性降低，因为增加了集成到模型中的随机性数量。

在研究范围内，创建了一个由三个客户端组成的仿真环境。对每个客户端的数据应用了隐私机制，并进行了本地模型训练。然后在服务器端将这些本地模型合并以获得全局模型。该过程重复使用不同的 ϵ 值来分析隐私与性能权衡对系统的影响。研究结果表明，必须谨慎管理这种权衡。

4 实验结果

本节介绍了获得的实验结果。开发的模型分类性能通过准确率、精度、召回率和 F1 分数等基本指标进行了评估，包括有隐私增强和无隐私增强两种情况。

该模型在一个模拟 FL 架构的结构中进行了评估，工作是在单台计算机上本地完成的。在此背景下，应用了一个联邦学习场景，仿佛存在一个去中心化的结构。首先，数据集被分成两部分：80%用于训练，20%用于测

试。训练数据被随机分为三个等份，每部分配置为表示一个独立客户端。在每个客户端上，首先使用 CBA 算法进行了局部模型训练。然后，利用 duCBA 聚合算法将客户端上的这些局部训练模型合并以获得集中式全局模型。没有增强隐私的合并模型性能评估结果见表 2，混淆矩阵见图 2，ROC 曲线见图 3。

表 2: 所提模型不带 RR 的性能指标

Class	Precision	Recall	F1-Score	Accuracy
No Hypertension	99.00	95.00	97.00	97.00
Hypertension	96.00	99.00	98.00	97.00
Macro Average	98.00	97.00	98.00	97.00

合并后的最终模型的准确率计算为 97%。这一高准确率表明该模型总体上展示了成功的分类性能。在查看图 2 中的混淆矩阵时，可以看到 2284 个样本被正确分类为真正例，而 2799 个样本被正确分类为假反例；另一方面，只有 114 个样本被分类为假正例，15 个样本被分类为真反例。图 3 展示了 ROC 曲线的一般形式和 AUC 值 87%，这表明该模型在类间具有高分辨能力，并且在泛化性能方面表现出可靠性。根据这些结果，可以说该模型特别成功地区分了高血压类别。此外，计算出的 F1 分数基于每个类别的数值都较高且彼此接近，表明该模型能够以平衡的方式学习两类并进行分类而不偏向任何一类。这也证明了该模型在类间分辨能力上的强度和可靠性。

随后，通过将隐私机制整合到相同的结构中来评估模型性能。首先，将 ϵ 值设置为 1，并检查所获模型的性能结果。使用 CBA 算法进行本地模型训练，在此过程中对客户端数据应用了差分隐私。这些本地训练的模型在中央服务器上通过 duCBA 合并算法被整合以获得全局模型。应用隐私机制后，该合并模型的性能评估结果见表 3。

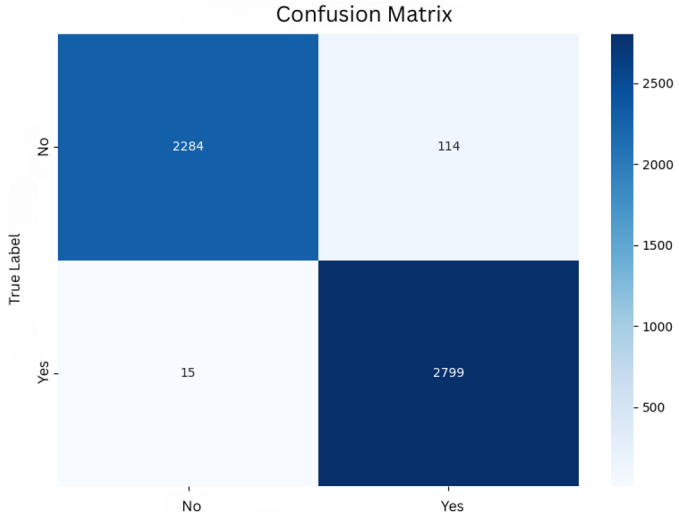

图 2: 最终合并的 duCBA 模型的混淆矩阵

表 3: RR 模型的性能指标

Class	Precision	Recall	F1-Score	Accuracy
No Hypertension	94.00	65.00	77.00	83.00
Hypertension	78.00	97.00	86.00	83.00
Macro Average	86.00	81.00	82.00	83.00

差分隐私在联邦学习中的应用：如何减轻推理攻击

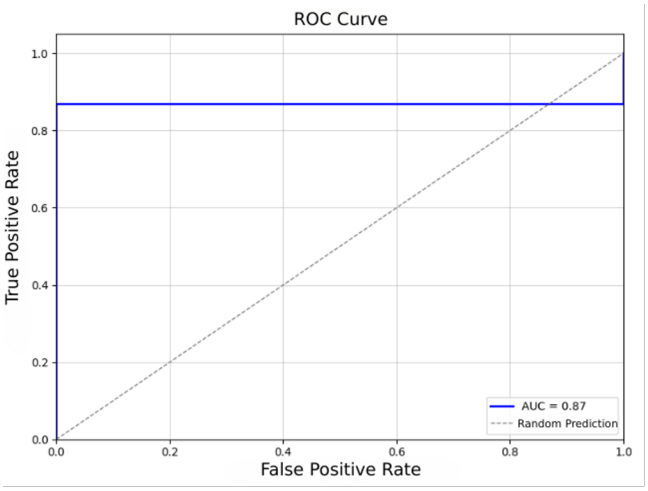

图 3: ROC 曲线的 duCBA 合并最终模型

表 2 显示，未应用隐私的模型在两类中都表现出高 F1 分数和平衡性能。然而，在表 3 中，当加入差分隐私（RR）时，敏感度和 F1 分数显著下降，特别是在“无高血压”类别中。这表明隐私机制对各类别具有不对称的影响。

模型性能的变化是通过在 0 到 5 之间取不同值的 ϵ 重复该过程来分析的。图 4 图形展示了随着 ϵ 值增加，在模型准确率和基于类别的 F1 分数中观察到的变化。当 ϵ 值减少，即差分隐私级别提高时，总体准确率以及两类的 F1 分数都出现了下降。

具体来说，在低 ϵ 值下，“无高血压”类别的 F1 分数受到的影响更大，并且产生的结果较低，与“存在高血压”类别相比。这种情况表明，在差分隐私范围内应用的随机化过程可能对类别模式产生不对称影响。

在文献中，根据 NIST SP 800-226 指南，ϵ<= 1 被认为提供了强保密性。另一方面，声明称随着 ϵ 值的增加，保密性保证会减弱。然而，一般建议是在性能损失可接受的范围内选择尽可能低的 ϵ 值。

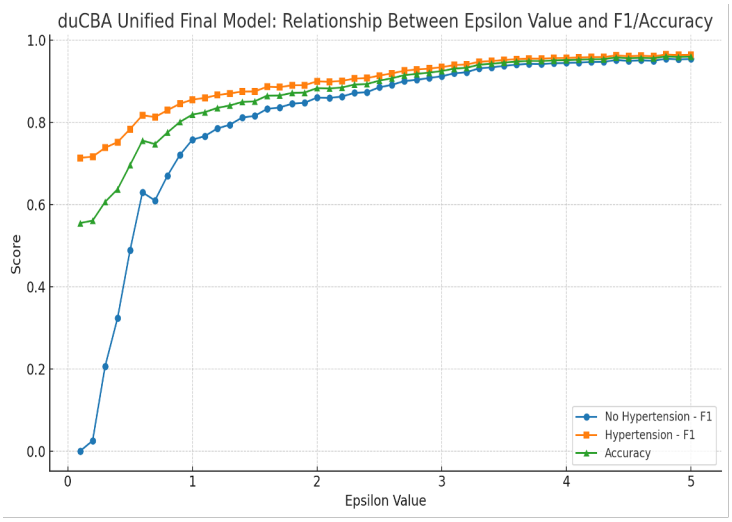

图 4: 模型准确性及基于类别的 F1 分数根据 Epsilon 值

图 4 展示了模型准确率和基于类别的 F1 分数随 ϵ 值增加而发生的变化。随着 ϵ 值的减少等，以及差分隐私级别提高时，整体准确率及两类的 F1 分数均有所下降。这表明为了保护隐私而实施的噪声操作显著降低了模

型的分类性能，尤其是在低 ϵ 值的情况下。"无高血压"类别的 F1 分数在低 ϵ 值下受到更大影响，并且得分低于"存在高血压"类别。这表明噪声过程对类模式的影响可能是不对称的。文献中认为根据 NIST SP 800−226 指南，当 ϵ<= 1 时能够提供较强的隐私保护。人们接受隐私保证会随着 ϵ 值的增加而减弱。然而，作为一般建议，在性能损失可以容忍的情况下，应使用尽可能低的 ϵ 值。

5 讨论与结论

在本研究的范围内，基于联邦学习架构开发的 duCBA 聚合算法从隐私和安全的角度进行了评估。duCBA 算法是传统 CBA 方法在联邦环境下的适应性改进，能够将客户端上本地训练的分类规则整合到中心化的结构中。然而，仅仅阻止联邦学习结构中的集中式数据共享并不能完全保证防止隐私侵犯。因此，利用了差分隐私机制，特别是为了加强个体数据隐私以对抗推断攻击。

在该研究中，随机响应方法被应用于差分隐私的范围内，并通过向客户端生成的数据添加受控噪声来匿名化模型更新。在进行的实验评估中，观察到应用噪声水平与模型准确性之间存在逆相关关系；这揭示了 ϵ 参数在系统性能中的关键作用。随着 ϵ 值的减小，隐私级别增加，但这导致了模型分类性能的显著下降。获得的结果表明，在联邦系统的设计中必须仔细分析隐私与性能之间的权衡。

未来的工作将重点放在开发针对模型中毒攻击的安全措施上，这是联邦学习系统面临的另一个威胁，并提高 duCBA 算法对此类攻击的韧性。

参考文献

[1] MG Sarwar Murshed, Christopher Murphy, Daqing Hou, Nazar Khan, Ganesh Ananthanarayanan, and Faraz Hussain. Machine learning at the network edge: A survey. *ACM Computing Surveys (CSUR)*, 54(8):1–37, 2021.

[2] Massimo Merenda, Carlo Porcaro, and Demetrio Iero. Edge machine learning for ai-enabled iot devices: A review. *Sensors*, 20(9):2533, 2020.

[3] Yi Liu, Li Zhang, Ning Ge, and Guanghao Li. A systematic literature review on federated learning: From a model quality perspective. *arXiv preprint arXiv:2012.01973*, 2020.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282, Apr 2017.

[5] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov. How to backdoor federated learning. In *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, pages 2938–2948, Jun 2020.

[6] M. Fredrikson, S. Jha, and T. Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proc. 22nd ACM SIGSAC Conf. Computer and Communications Security (CCS)*, pages 1322–1333, Oct 2015.

[7] B. Buyuktanir, K. Yildiz, and G. K. Baydogmus. A systematic mapping study on machine unlearning in federated learning. In *Proc. 7th Int. Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*, pages 1–10. IEEE, May 2025.

[8] B. Büyüktanir, K. Yildiz, E. Ülkü, and T. Bütüktanir. du-cba: Data-agnostic and incremental classification-based association rules extraction architecture. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(3), 2023.

[9] B. Buyuktanir, . Altinkaya, G. Karatas Baydogmus, and K. Yildiz. Federated learning in intrusion detection: advancements, applications, and future directions. *Cluster Computing*, 28(7):1–25, 2025.

[10] A. Ö. Önlü, B. Akca, B. Buyuktanir, K. Yildiz, and G. K. Baydogmus. An investigation and performance evaluation of aggregation algorithms in federated learning architecture. In *Proc. Innovations and Applications in Smart Systems Conf. (ASYU)*, Istanbul, Turkey, 2025.

[11] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konený, and H. B. McMahan. Adaptive federated optimization. arXiv preprint arXiv:2003.00295, Mar 2020.

[12] E. K. Sannara, F. Portet, P. Lalanda, and V. E. G. A. German. A federated learning aggregation algorithm for pervasive computing: Evaluation and comparison. In *Proc. IEEE Int. Conf. Pervasive Computing and Communications (PerCom)*, pages 1–10, Mar 2021.

[13] D. G. Nair, C. V. A. Narayana, K. J. Reddy, and J. J. Nair. Exploring svm for federated machine learning applications. In *Advances in Distributed Computing and Machine Learning, Proc. ICADCML*, pages 295–305, Singapore, 2022. Springer.

[14] G. Chen, H. Liu, L. Yu, Q. Wei, and X. Zhang. A new approach to classification based on association rule mining. *Decision Support Systems*, 42(2):674–689, 2006.

[15] B. E. Tosun, E. Yorulmaz, F. E. Ergul, B. Buyuktanir, G. K. Baydogmus, and K. Yildiz. Comparative analysis of relational classification algorithms on benchmark datasets. In *Proc. 10th Int. Conf. Computer Science and Engineering (UBMK)*, Istanbul, Turkey, 2025.

[16] Z. Çıplak, K. Yıldız, and . Altınkaya. Fedetect: a federated learning-based malware detection and classification using deep neural network algorithms. *Arabian Journal for Science and Engineering*, pages 1–28, 2025.

[17] B. Buyuktanir, Z. Ciplak, A. E. Cil, O. Yakar, M. B. Adoum, and K. Yildiz. Ddos_fl: Federated learning architecture approach against ddos attack. *Pamukkale University Journal of Engineering Sciences*, 31(6), 2025.

[18] E. S. Erdöl, H. Erdöl, B. Üstübiolu, F. Z. Solak, and G. Uluta. Impactful neuron-based secure federated learning. In *Proc. 32nd Signal Processing and Communications Applications Conf. (SIU)*, pages 1–4, May 2024.

[19] L. Bai, H. Hu, Q. Ye, H. Li, L. Wang, and J. Xu. Membership inference attacks and defenses in federated learning: A survey. *ACM Computing Surveys*, 57(4):1–35, 2024.

[20] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang. Membership inference attacks on machine learning: A survey. *ACM Computing Surveys*, 54(11s):1–37, 2022.

[21] S. V. Dibbo. Sok: Model inversion attack landscape: Taxonomy, challenges, and future roadmap. In *Proc. 36th IEEE Computer Security Foundations Symposium (CSF)*, pages 439–456, Jul 2023.

[22] X. Xiong, S. Liu, D. Li, Z. Cai, and X. Niu. A comprehensive survey on local differential privacy. *Security and Communication Networks*, 2020:8829523, 2020.

[23] N. Phan, X. Wu, H. Hu, and D. Dou. Adaptive laplace mechanism: Differential privacy preservation in deep learning. In *Proc. IEEE Int. Conf. Data Mining (ICDM)*, pages 385–394, Nov 2017.

[24] F. Liu. Generalized gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):747–756, 2018.

[25] Q. Ye, L. Yu, K. Huang, X. Xiao, W. Liu, and H. Hu. From randomized response to randomized index: Answering subset counting queries with local differential privacy. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 3877–3891, May 2025.

[26] P. Chuks. Diabetes, hypertension and stroke prediction [dataset]. `https://www.kaggle.com/datasets/prosperchuks/health-dataset`, 2025. Accessed: Jul. 18, 2025.

[27] M. S. Kraiem, F. Sánchez-Hernández, and M. N. Moreno-García. Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties: An approach based on association models. *Applied Sciences*, 11(18):8546, 2021.

[28] O. Yakar, B. Buyuktanir, A. E. Cil, and A. B. A. Girgin. Performance comparison of different classification algorithms and feature selection methods in turkish hate speech problem analysis. *European Journal of Science and Technology*, (53):97–111, 2024.