arXiv:2303.03488v1 [cs.LG] 6 Mar 2023

# 神经网络聚合方法比较

John Pomerat Aviv Segev *

**摘要**

深度学习在理论层面已取得成功。要使深度学习在行业中取得成功，我们需要具备能够处理真实数据中出现的诸多不一致性的算法。这些不一致性可能会对深度学习算法的实施产生重大影响。人工智能目前正在改变医疗行业。然而，获得使用医疗数据来训练机器学习算法的授权是一个巨大的障碍。一种可能的解决方案是在不共享患者信息的情况下共享数据。我们针对深度学习算法提出了一种多方计算协议。该协议能够同时保护训练数据的隐私和安全。本文分析了三种神经网络组合方法：迁移学习、平均集成学习和串联网络学习。通过不同实验，将结果与基于数据共享的方法进行了比较。我们还分析了所提出协议的安全问题。尽管该分析基于医疗数据，但机器学习训练的多方计算结果具有理论意义，并可在多个研究领域中实施。

## 1 引言

近年来，机器学习的理论进展有望彻底改变从制造业 [1], 、医疗保健 [2] 和交通运输 [19], 到教育 [20]等众多行业领域。尽管学习算法已有许多成功的应用，但机器学习的大部分进展仍停留在理论层面 [3]。缺乏实际应用的一个原因是

*计算机科学系，南阿拉巴马大学，美国阿拉巴马州莫比尔 36688电子邮件：segev@southalabama.edu。

，尤其是在医疗保健领域，学习算法的实用性、韧性和安全性 [2, 4]。数据是机器学习的基础；因此，对于许多现实世界中的应用，必须仔细考虑数据的形态、组织方式、数量和质量 [5]。随着对医疗保健数据集的需求不断增加，数据共享 [5] 被提议作为获取医疗保健模型数据的一种策略。在数据共享中，医院会将数据重新格式化为约定的结构，并对内容进行匿名化，以免泄露患者机密数据。我们提出了一种使用安全多方计算（MPC）的数据共享替代方案。多方计算是密码学的一个分支，致力于对私有的、由用户持有的数据进行函数计算。一个具有代表性的例子是：两个人希望确定谁的薪水更高，同时双方都不向对方透露自己的薪水。借助安全多方计算，有一种算法能够解决这一问题以及其他类似问题。我们提出了一种协议，用于在私有数据集上训练神经网络，然后对这些神经网络进行组合，从而在不暴露私有数据的情况下，使模型的最终性能与基于合并后的私有数据集训练的模型相当。本文研究了三种神经网络聚合方法，用于组合在不同数据集上训练、但共享底层函数的网络。对于这三种方法，底层网络架构和数据集均保持不变。此外，包括激活函数、优化器和批量大小在内的超参数也保持不变。我们通过均方误差衡量性能，并记录该指标以比较这三种方法。这三种方法分别是迁移学习、平均集成学习和串联网络学习。

本文将深入探讨这些神经网络聚合方法。

# 2 相关工作

## 2.1 安全

在医疗保健领域，维护数据隐私的重要性显而易见。因此，应尽可能对敏感数据进行匿名化，以防止任何形式的数据泄露。针对学习算法存在多种攻击 [6–9]。2015年，Goodfellow 等人提出，对抗攻击是神经网络的一种安全漏洞 [10]。此后，针对神经网络安全的研究不断增多，也发现了更多攻击向量 [6–9]。按照我们的定义，我们的问题不易受到黑盒对抗攻击的影响。我们的系统易受攻击的一种途径是恶意对手提供的训

练代码 [11]。为防范这种攻击，特定参与方所采用的代码应开源并接受独立审查。此外，生成模型还存在一种攻击途径 [12] ，采用生成模型的某些实现应考虑这一点，但这超出了本文的讨论范围。我们主要关注的攻击途径是成员推断攻击 [18]。成员推断攻击是针对训练好的神经网络分类器的一种黑盒攻击途径。这种攻击是一种算法，可通过统计方法判断某个输入元组是否属于私有训练集 [18]。为防范成员推断攻击，模型应避免过拟合。此外，加入正则化、限制预测向量以及提高预测向量的熵，都有助于防止成员推断攻击 [18]。

## 2.2 迁移学习

神经网络聚合的一种广为人知的方法是迁移学习。最近的研究表明，迁移学习很有用且极具适应性，尤其适用于强化学习和深度神经网络模型 [16, 22–24]。此外，与本文探讨的一些方法相比，迁移学习的适用范围更广，因为它能够与种类更多的学习算法配合使用，包括卷积神经网络 [22]。此外，研究表明，迁移学习能够很好地应用于时间序列预测和循环神经网络 [25]。另外，在医疗保健领域背景下对迁移学习的研究已初显成效 [26–28]。Gupta 等人 [26] 的一篇论文利用迁移学习，将医疗保健领域中的模型推广到同一领域中的类似任务。类似地，Chen 等人 [27] 的一项研究利用可穿戴技术，采用迁移学习处理时间序列健康数据，从而提升 FedHealth 模型的性能并增强个性化程度。迁移学习的研究结果表明，神经网络聚合在医疗保健领域深度学习中的应用前景良好。

迁移学习方面的既有研究表明，在大数据医疗保健应用中，神经网络聚合作为数据共享的替代方案，前景广阔。这些成果与安全和多方计算领域的一些研究相结合，构成了本文的研究动机。

# 3 神经网络聚合

## 3.1 问题陈述

设定如下。令 $D_1, D_2, \cdots, D_k$ 为 $\mathbb{R}^n$ 的子集，并以数据集表示。接着，令 $G: \mathbb{R}^n \to \mathbb{R}^m$ 为一个可微函数，以参数为 $\theta_g$ 的多层感知机表示。我们关注的是如何从 $D_i$ 中生成 $\theta_g$，使得 $G$ 的损失与从 $\bigcap_{i=1}^{k} D_i$ 获得 $\theta_g$ 的损失相当，同时在计算上无法从 $G$中提取有关 $D_i$ 成员的信息。这个从多个互不相交的数据集中训练神经网络的过程称为神经网络聚合。神经网络聚合有三种方法：串联网络学习、平均集成学习和迁移学习。

## 3.2 串联网络学习

第一种方法称为串联网络学习，其运作方式是将一个预训练的“专家”神经网络作为额外输入，训练另一个神经网络。在我们的实验中，我们考虑一个在第一个数据集上训练的单一神经网络，并记录该神经网络在测试集上的性能。随后，该网络会为第二个数据集中的每个条目生成预测；接着，为第二个数据集创建一个新的神经网络，并将预测数组作为额外输入。然后在第二个数据集上训练该神经网络，并记录均方误差。直观而言，第二个神经网络的均方误差很可能会有所改善，因为它会“学会”何时信任第一个网络的预测，何时改用自己的计算结果，见图 1。

**算法 1：** 串联网络

**对于**除最后一个参与方之外的所有参与方**执行**
  在参与方的数据上训练网络；
**end**
取该网络，并将每个已训练网络的输出追加为一个新的输入神经元，然后在最后一组参与方的数据上训练得到的网络；

## 3.3 平均集成学习

第二种方法考虑两个具有完全相同架构的神经网络，$N_1$，$N_2$ 它们具有相同的激活函数、优化器、隐藏层数量和神经元数量。随后，分别使用结构相同的不同数据集训练这两个网络，并记录测试

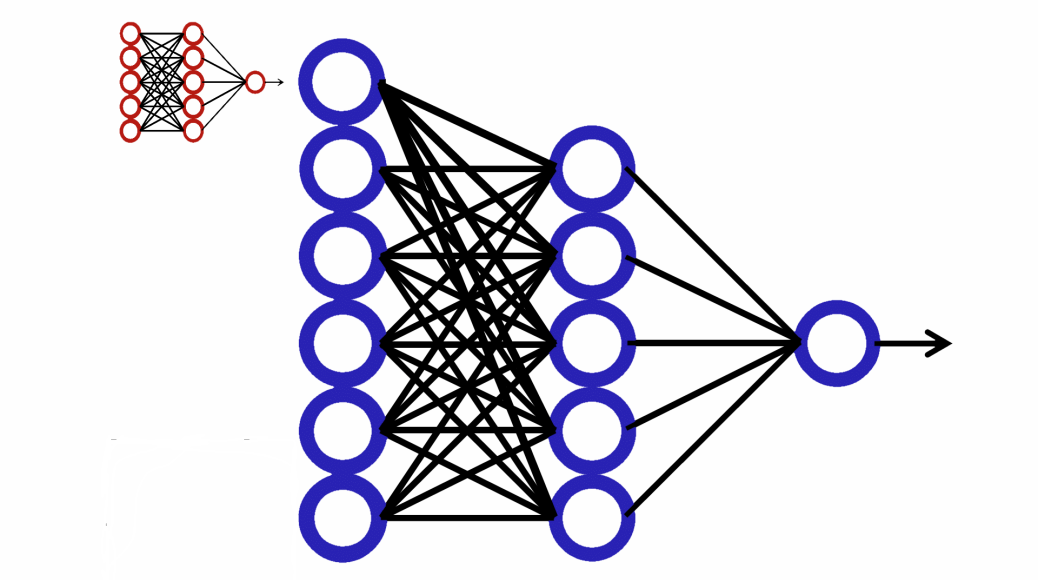

图1：将神经网络作为输入，以辅助训练第二个网络。

集上的均方误差。然后，将这两个神经网络组合成第三个具有完全相同结构的网络 $N_3$ （图 2）。$N_3$ 的权重和偏置分别是 $N_1$ 和 $N_2$中对应权重和偏置的平均值。更具体地说，如果 $N_3$中权重和偏置的总数为 $n$ ，且 $N_3$中的 $i$个权重或偏置由$i$（$N_3$）表示，则对于所有 $0 < i \leq n$,

$$N_3(i) = \frac{N_1(i) + N_2(i)}{2}$$

$N_3$ 随后在测试集上进行测量，其性能与 $N_1$ 和 $N_2$的性能进行比较。除了纯平均外，还会考虑其他策略。首先，可以根据数据集的大小进行加权平均，以确保在显著更多数据上训练的模型不会与在数据量小得多的数据集上训练的模型被同等对待。另一种方法是进行加权平均时，不仅考虑数据集的大小，还考虑疾病预测案例中阳性和阴性样本的比例。这样做是为了确保信息量不大的较大数据集不会压过包含更多信息的较小数据集。

## 3.4 迁移学习

第三种方法是迁移学习。与其合并两个神经网络，迁移学习通过在额外数据集上训练一个

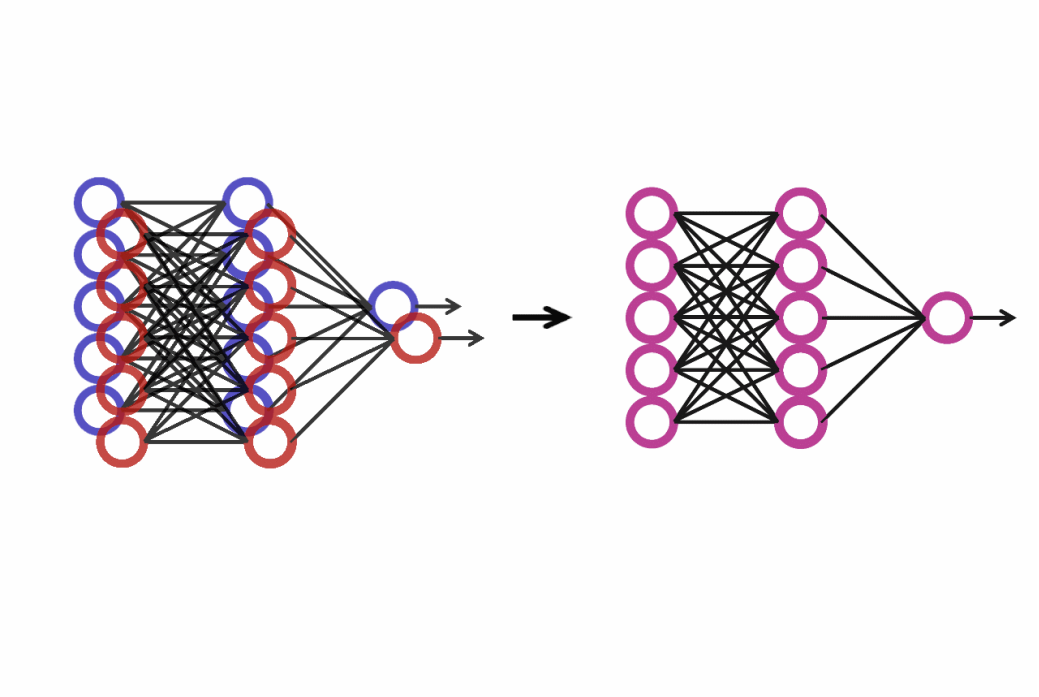

图2：对神经网络对应的权重和偏置取平均

不进行权重重新初始化的单一神经网络来实现 [16,17]。我们的实验考虑了一个具有随机初始化的权重、在第一个数据集上训练的单一神经网络。记录其在测试集上的均方误差。随后，在不重新初始化权重的情况下，在第二个数据集上训练该神经网络，并再次记录均方误差。然后，通过先在第二个数据集上训练，再在第一个数据集上训练来重复这一过程。全程记录均方误差。

# 4 实验

为比较所提出的神经网络聚合方法，我们进行了两项实验：一项使用人工生成的多项式数据，另一项使用威斯康星大学麦迪逊医院的乳腺癌数据集 [21]。开展这些测试的目的是初步比较所提出的方法与在全部数据上同时训练、代表数据共享的神经网络之间的性能。

**算法 2:** 平均集成学习

```
对于所有参与方执行
    在参与方的数据上训练相同的网络;
end
初始化一个与其他模型相同的新模型;
对于网络中的每个权重和偏置执行
    对于每个已训练的网络执行
        对相应的权重或偏置值求和;
    end
    新网络中的权重或偏置等于该总和除以参与方的数量
     ;
end
```

## 4.1 数据

本文中的神经网络使用真实数据和人工生成的数据进行训练。人工生成的数据创建过程如下：采用正态分布生成二维数组，其中填充了指定范围内的随机有理数。数组的每一行由 7 个随机有理数组成，代表数据特征。实验创建了多个数据集，分别生成了大小为 3200、1600、800 和 400 的数组。生成数组后，按照字典序项序创建了一个次数为 $n$ 的多元多项式，其系数随机选自正态分布。

$$f(x_1, x_2 ... x_7) \tag{1}$$

接下来，对于生成数据中的每组 7 个值，将生成数据中的值代入多项式，计算了 $\gamma$, $f(\gamma)$。图 3 以 2 维而非 7 维的形式展示了这一过程。

在计算出每个数组中所有元组的 $f(\gamma)$ 后，将这些值与生成的数据合并，形成一个数据集，每行包含 8 个值：7 个随机有理数，以及根据生成函数计算出的 y 值。因此，实验中的网络将使用这 7 个有理数进行训练，以学习底层的多项式函数。随后，这些数据集被划分为两个训练集和一个测试集，分别包含 80% 和 20% 的条目。之后，训练集又被划分为两个大小相等的训练集。

本文使用的真实数据来自威斯康星大学麦迪逊医院的乳腺癌数据集 [21]。该数据集也被划分为两个大小相同的训练集和一个测试集。乳腺癌

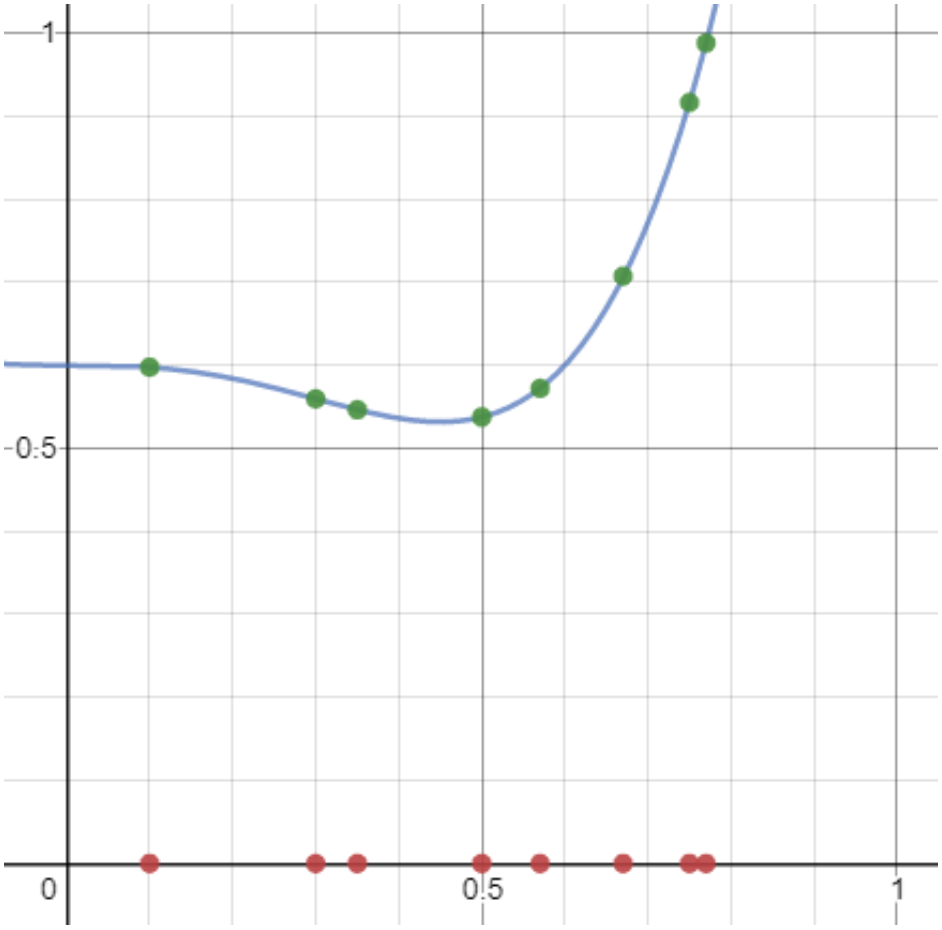


图 3：生成的多项式函数（蓝色）上的随机 x 值（红色）和计算得到的 y 值（绿色）。

数据集包含 569 行和 32 个特征。数据被划分为两个训练集，每个训练集包含 256 个训练样本；测试集包含 57 个样本。乳腺癌数据中的特征描述了肿瘤。部分特征包括肿块厚度、细胞大小均匀性和形状、边缘黏附性等。此外，数据中的每个特征都以三种不同的方式记录。每个特征都有平均值、低值和高值。所采用的数据预处理包括归一化和少量特征调整，以使数据格式适合构建规范的训练集和测试集。

## 4.2 回归

在本实验中，我们取用上述定义的回归数据，并将其拆分为训练集和测试集，其中 80% 的样本用于训练，20% 的样本用于测试。我们使用训练集训练神经网络，然后记录其在测试集上的损失。由于所有数据都集中在一处，因此得到的模型代表了使用通过数据共享创建的数据集训练的网络。随后，我们将训练集拆分为两个大小相等的较小训练集。接着，我们分别采用三种方法，使用拆分后的数据集训练神经网络，并记录每种方法的损失。所选神经网络架构力求与数据最佳匹配，并在不同条件下多次重复测试时保持不变。测试条件包括：训练轮数从 10 到 200 不等；回归数据中的噪声通过对数偏移加入，系数在 1 到 3 之间变化；数据集大小从 400 到 32,000 不等；底层数据集的多项式次数从 2 到 5 不等。所有测试中各方法的平均损失（包括以“None”表示的数据共享模型的损失）见下方图4。此外，不同噪声添加程度下测试的平均损失见表1。数据中添加的噪声由以下内容给出

$$y = f(x_1, x_2, ...x_7) + ndr$$

其中 $x_1$，$x_2...x_7$ 是一个数据点，$n$ 是一个选定的噪声值，$r$ 是从 -2 到 2 之间的正态分布中随机选取的实数，且$f$ 是一个次数为 $d$ 的多项式函数。

| 方法 | 平均均方误差 | 噪声 n |
|---|---|---|
| 平均集成 | 0.015 | 0 |
| | 0.011 | 1 |
| | 0.011 | 2 |
| 串联网络 | 0.013 | 0 |
| | 0.010 | 1 |
| | 0.010 | 2 |
| 迁移学习 | 0.011 | 0 |
| | 0.007 | 1 |
| | 0.009 | 2 |
| None | 0.006 | 0 |
| | 0.008 | 1 |
| | 0.008 | 2 |

表1：添加噪声的方法的损失比较

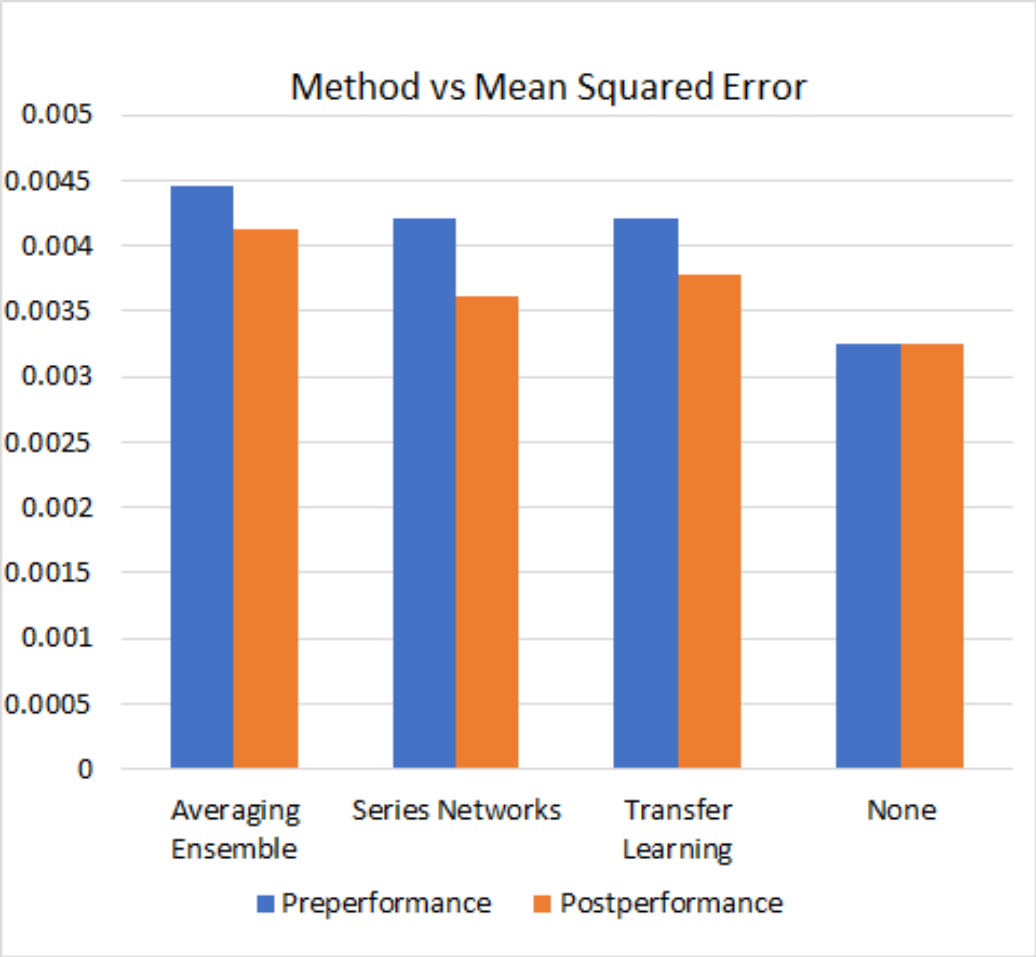


图4：回归数据上各方法的损失比较

聚合前性能是指模型在第一个数据集上完成学习后，在测试集上测得的损失。同样，聚合后性能是指聚合第二个数据集后测得的损失。这样做的目的是观察通过聚合多个数据集进行数据共享时，该方法是否会收敛到通过数据共享获得的模型性能。训练完成后，这三种方法的聚合性能均与基于合并后的“共享”数据训练的模型（None）相当。在这三种方法中，串联网络的性能最佳，并且聚合后的性能提升幅度也最大。

## 4.3 乳腺癌分类

本实验旨在训练一个分类器，以判断肿瘤是良性还是恶性。乳腺癌数据集包含569行和32个特征。与回归实验类似，我们对数据进行了划分，根据数据配置了神经网络架构，随后计算了三种方法的准确率。此外，在将训练集一分为二之前，我们先用数据训练了数据共享对应的模型，并记录了准确率。我们改变了多种超参数进行测试，包括批量大小、训练轮数和神经元数量。测试的准确率见图5，测试的ROC曲线见图6，精确率、召回率和F1分数见表2。

在此例中，所有方法的表现都优于通过数据共享获得的对应模型。此外，这个例子也证明了我们的方法在医疗保健领域的可行性。

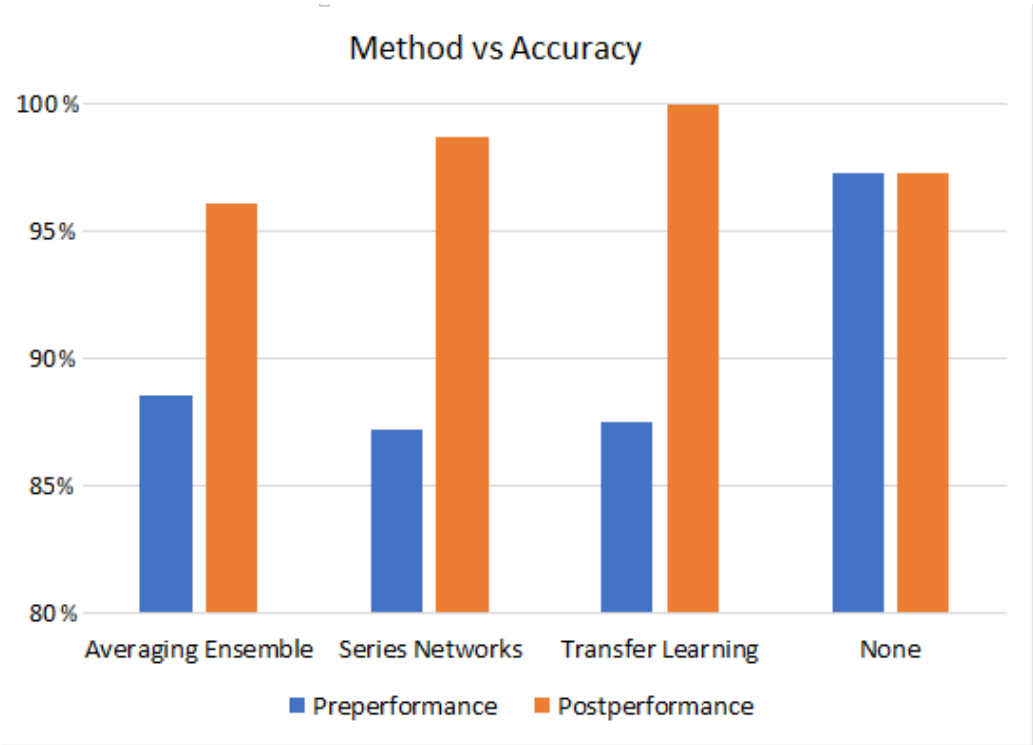


图5：乳腺癌数据上各方法的准确率比较

从准确率图（图 5）和 ROC曲线（图 6）可以看出，迁移学习和串联网络表现最佳，优于在合并数据集上进行训练。这很可能是因为数据集规模较小时，在更小的数据子集上进行训练能够带来更好的泛化能力。

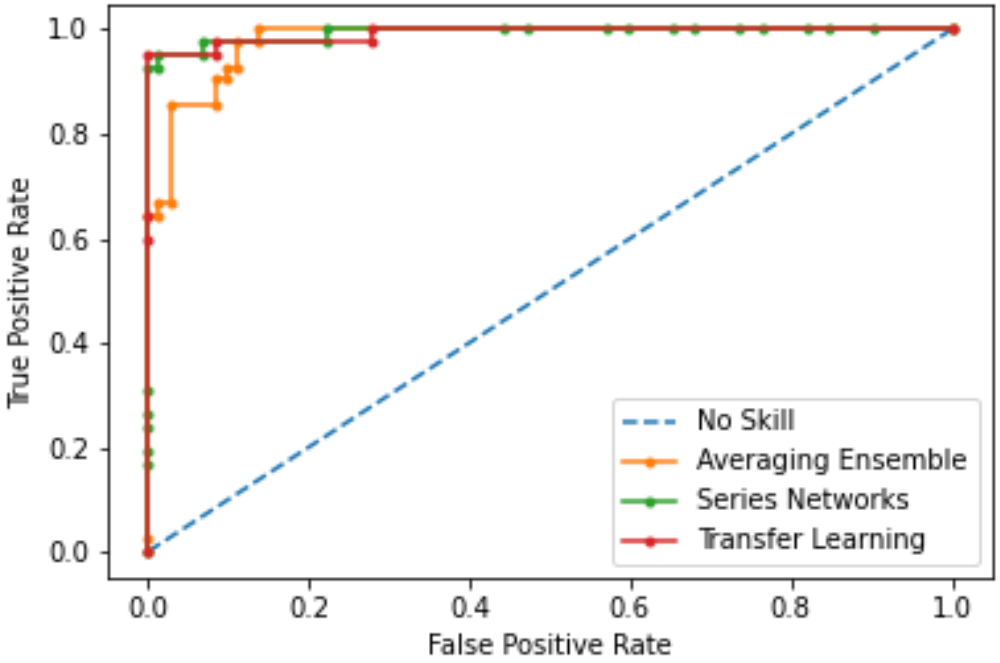


图 6：乳腺癌数据的 ROC 曲线

| 方法 | 精确率 | 召回率 | F1分数 |
|---|---|---|---|
| 平均集成 | 0.76 | 1.00 | 0.87 |
| 串联网络 | 1.00 | 0.93 | 0.96 |
| 迁移学习 | 1.00 | 0.93 | 0.96 |

表 2：所比较方法的指标

# 5 结论

要让神经网络聚合被充分认可为数据共享的更优替代方案，还需要进行更多测试。此外，未来的工作应研究所提出模型的可扩展性，并考察随着不相交数据集数量增加，模型是否能够收敛。如果迁移学习或串联网络学习能够通过将训练分散到多个数据集上，收敛到与通过数据共享获得的相同模型，那么这种方法就是可行的。此外，还需要开展更多关于成员推断攻击的研究，以降低安全方面的顾虑。由于成员推断攻击对过拟合模型具有很强的攻击效果，因此它将

有趣的是，可以观察串联网络或迁移学习在多个数据集上训练后的最终表现。尽管如此，迁移学习和串联网络学习似乎都是在私有数据集上分散训练的可行方法。因此，虽然还需要进行更多测试来证明神经网络聚合的可行性，但本文初步提出，神经网络聚合可以作为数据共享的一种有效替代方案。

# 参考文献

[1] L. Columbus。2019年机器学习革新制造业的10种方式。《福布斯杂志》，2019年。[2] T. Davenport 和 R. Kalakota。人工智能在医疗保健领域的潜力。《未来医疗保健杂志》第*6*卷第*2*期，2019年。[3] B. Bergstein。这就是人工智能尚未重塑大多数企业的原因。《麻省理工科技评论》，2019年。[4] J. He 等。人工智能技术在医学领域的实际应用。《自然医学》第*25*卷第*1*期，2019年。[5] K. Benke 和 G. Benke。人工智能与公共卫生大数据。《国际环境研究与公共卫生杂志》第*15*卷第*12*期，*2796*。*10*，2018年。[6] N. Papernot 等。对抗环境下深度学习的局限性。*arXiv*预印本 *arXiv:1511.07528v1*，2015年。[7] N. Carlini 和 D. Wagner。迈向神经网络鲁棒性的评估。*arXiv*预印本 *arXiv:1608.04644v2*，2017年。[8] A. Kurakin、I. Goodfellow 和 S. Bengio。大规模对抗机器学习。*arXiv*预印本 *arXiv:1611.01236v2*，2017年。[9] F. Tramer 等。集成对抗训练、攻击与防御。*arXiv*预印本 *arXiv:1705.07204v5*，2020年。[10] I. Goodfellow、J. Shlens 和 C. Szegedy。解释并利用对抗样本。*arXiv*预印本 *arXiv:1412.6572v3*，2015年。[11] C. Song、T. Ristenpart 和 V. Shmatikov。记忆过多的机器学习模型。*arXiv*预印本 *arXiv:1709.07886v1*，2017年。

[12] N. Carlini、C. Liu、 ´U. Erlingsson、J. Kos 和 D. Song。秘密共享者：评估和测试神经网络中的非预期记忆。*arXiv*预印本 *arXiv:1802.08232v3*，2019年。[13] C. Clifton、M. Kantarcioglu、J. Vaidya、X. Lin 和 M. Y. Zhu。隐私保护分布式数据挖掘工具。*ACM SIGKDD*知识发现与数据挖掘特别兴趣小组探索通讯，第*4*期，2002年。[14] D. Bogdanov、M. Niitsoo、T. Toft 和 J. Willemson。用于数据挖掘应用的高性能安全多方计算。*IJIS*《国际信息安全杂志》，第*11*卷，*403–418*，2012年。[15] W. Du、Y. S. Han 和 S. Chen。隐私保护多变量统计分析：线性回归与分类。*SIAM*国际数据挖掘会议，会议论文集，2004年。[16] C. Tan、F. Sun、T. Kong、W. Zhang、C. Yang 和 C. Liu。深度迁移学习综述，*arXiv*预印本 *arXiv:1808.01974*，2018年。[17] K. Bonawitz 等。面向用户持有的数据开展联邦学习的实用安全聚合，*arXiv*预印本 *arXiv:1611.04482*，2016年。[18]R. Shokri、M. Stronati、C. Song 和 V. Shmatikov。针对机器学习模型的成员推断攻击，*arXiv*预印本 *arXiv:1610.05820v2*，2017年。[19] B. Keshav。自动驾驶汽车：过去、现在与未来——对上个世纪的发展、当前状况以及自动驾驶汽车技术预期未来的综述，*ICINCO*——第*12*届控制、自动化与机器人信息学国际会议，会议论文集，2015年。[20] O. Zawacki-Richter、V. Marin、M. Bond 和 F. Gouverneur。高等教育人工智能应用研究的系统综述 —— 教育工作者在哪里？《高等教育教育技术国际期刊》*16，39*，2019年。[21] D. Dheeru 和 G. Casey。UCI机器学习资料库 加利福尼亚大学尔湾分校信息与计算机科学学院，2017年。[22] H. Mahbub、B. Jordan 和 F. Diego。用于图像分类的卷积神经网络迁移学习研究 第*18*届英国计算智能年度研讨会，诺丁汉，会议论文集，2018年。

[23] K. Weiss、T. Khoshgoftaar 和 D. Wang。《迁移学习综述》《大数据期刊》*3, 9*，2016年。[24] M. Taylor 和 P. Stone。《强化学习领域的迁移学习：综述》《机器学习研究期刊》*10, 1633- 1685*，2009年。[25] A. Giel 和 R. Diaz。《用于动作识别的循环神经网络与迁移学习》，2015年。[26] P. Gupta、P. Malhotra、J. Narwariya、L. Vig 和 G. Shroff。《使用深度神经网络进行临床时间序列分析的迁移学习》*arXiv* 预印本 *arXiv:1904.00655*，2019年。[27] Y. Chen、J. Wang、C. Yu、W. Gao 和 X. Qin。《FedHealth：面向可穿戴医疗保健的联邦迁移学习框架》*arXiv* 预印本 *arXiv:1907.09173*，2019年。[28] M.、J. Kleinberg、C. Zhang 和 S. Bengio。《Transfusion：理解医学影像中的迁移学习》*arXiv* 预印本 *arXiv:1902.07208*，2019年。