**基于联邦学习的电子病历共享系统**

作品名称： 基于联邦学习的电子病历共享系统

学 校： 桂 林 电 子 科 技 大 学

队 员：李平、袁恩昌、蓝颖聪、李新阳、文能翔

目 录

[摘要 3](#_Toc119413073)

[背景介绍 4](#_Toc119413072)

[系统概要设计 9](#_Toc119413073)

[总结 12](#_Toc119413075)

# 摘要

随着大数据和人工智能的发展，越来越多的新型技术应用到了医疗领域，用来建立以电子病历为核心的医院信息系统。这些新技术的应用不仅提高了医疗服务的质量和效率，也推动了医疗行业的数字化转型和升级。

传统的病历客观记录了临床医护人员的工作内容和过程，其效率较低且数据不安全、不精确。现有的电子病历也存在着三点问题：①数据安全，电子病历中存储的数据很敏感，如果被非法获取，会对患者的隐私产生严重的影响；②数据准确性：电子病历中的数据是由医生和护士等人员输入的，很容易发生输入错误，导致数据不准确；③数据兼容性：不同医院之间电子病历系统通常不相互兼容，很难将患者的病历数据共享和对接。

为解决上述问题，拟提出基于联邦学习的电子病历共享系统。它的工作流程如下：首先，各个医院需要在本地建立电子病历系统，并对本地数据进行清洗和标准化处理，确保数据的质量和可用性。其次，通过联邦学习的方式，各个医院可以将本地数据源共同参与到模型训练中来，从而实现数据的共享和融合。具体来说，各个医院会将本地模型的中间参数上传到中央服务器，中央服务器将收集到的所有参数进行聚合和加密，生成新的全局模型参数，并将更新后的全局模型参数重新分发给各个医院。这个过程中，原始数据不需要出本地，因此原始数据不需要流转，实现了数据的隐私保护。然后，各个医院可以基于新的全局模型参数，重新训练本地模型，并在本地应用新的模型参数进行预测和诊断。最后，通过联邦学习后的结果对当前病例进行判断，为诊断提供参考。

综上所述，该项目可以实现数据的共享和融合，提高模型的预测准确性和鲁棒性，同时也保护了患者的隐私和医院的商业利益。这种共享模式有助于建立更加完整和全面的电子病历数据库，为医学研究和分析提供更加可靠和丰富的数据源。同时，也可以帮助医生更加准确地了解患者的病情和病史，制定更加精准和有效的治疗方案，提高医疗服务的质量和效率。

关键词：联邦学习；隐私保护；电子病历；大数据

# 1.背景介绍

## 1.1.电子病历的介绍

近年来，我国医疗信息化政策频出，医疗信息化方面取得了长足的进步，并大力支持电子病历系统的建设，医院整体的电子病历应用水平逐步提高。根据中国医院协会信息专业委员会的调查，我国大部分医院对电子病历高度重视，并将其作为未来的重要工作。

电子病历（Electronic Medical Record, EMR）是指在电子计算机系统中存储、管理、更新和使用患者的医疗信息的系统，通常包含患者的个人信息、诊断信息、医疗服务记录、处方、检查报告、影像学图像等。它可以帮助医生和医疗团队更准确、有效和实时地了解患者的健康状况，并保证患者的医疗信息不丢失或损坏，促进信息共享和协作，并且可以帮助医生更好地监测患者的健康状况。此外，电子化方式管理的个人健康状态和医疗保健的信息也可以提高患者的医疗体验，因为它可以让患者随时了解自己的健康状况。

电子病历作为重要的医疗数据，是极具价值的医学研究资料。医学研究者可以对一定数量的电子病历开展研究，结合人工智能、大数据技术，为患者提供更高效、精确、人性化的诊疗方案。电子病历中的详细记录也可在医生接诊、医保核验、过度诊疗检测等方面起到重要作用。国家在政策上大力支持电子病历发展，不断出台电子病历规范类和支持性文件，推动医院的电子病历系统建设，行业迎来巨大发展。

## 1.2.电子病历发展历史

EMR的发展于20世纪80年代，但当时它的使用仍然有限，主要是因为计算机技术的局限性和高昂的成本。随着计算机的飞速发展，在21世纪初，国家陆续发布电子病历建设、评估等相关政策规划，政府也开始大力支持电子病历的建设，并制定了相关的政策。历经初步试点、推广普及、规范建设，电子病历逐渐向高质量管理方向迈进。

### 1.2.1试点阶段

**2002年，我国电子病历进入试点阶段，**原国家卫生部开始推进建设电子病历医院试点的相关工作。2002年10月，原国家卫生部发布《全国卫生信息化发展规划 纲要（2003—2009 年）》，要求三级医院创造条件，建设电子病历等信息化应用。2010年3月，原国家卫生部发布《电子病历试点工作方案》，要求在全国范围内至少遴选50家试点医院和3个试点区域承担电子病历试点工作，探索建立适合我国国情的电子病历系统。

试点阶段的目的是测试和评估电子病历系统的功能、效率、稳定性、安全性和接受度。试点阶段的内容通常包括系统的技术测试、用户的接受性评估、数据的准确性评估以及系统的改进。系统的技术测试是指确保电子病历系统的功能、效率、稳定性和安全性达到要求；用户的接受性评估是指评估医生、护士和其他医疗工作者对电子病历系统的接受程度；数据的准确性评估是指确保电子病历系统存储的病人信息准确可靠；系统的改进是指根据试点结果对电子病历系统进行改进，以确保系统能够适应更广泛的应用。

试点阶段的结果将为电子病历的普及提供重要的信息，并为电子病历系统的未来发展提供基础。在试点阶段结束后，电子病历系统可以扩展到更多的医院和医疗卫生机构，以改善病历记录质量和医疗信息共享的效率。

### 1.2.2推广阶段

**2011年，我国电子病历的发展进入了推广阶段。**国家先后发布电子病历推广应用的相关政策，鼓励支持在各级医院普及电子病历。2012年6月，国务院首次要求建设以电子病历和医院管理为重点的县级医院信息系统。 2015年3 月，国务院办公厅发布《全国医疗卫生服务体系规划纲要（2015—2020年）》， 要求到2020年，实现全员人口信息、电子健康档案和电子病历三大数据库基本覆 盖，并实现信息动态更新。

2017年，我国电子病历发展进入了规范化阶段。原国家卫计委发布了《电子病历应用管理规范》，明确了电子病历系统和电子病历的概念，对电子病历信息系统技术管理和电子病历质量管理提出具体要求。

在这一阶段，电子病历系统的建设和应用将进一步加强和普及。主要包括四个方面的工作。第一，加强技术支持是为了加强电子病历系统的技术支持，确保系统稳定、安全、高效运行；第二，完善管理制度是为了要完善电子病历系统的管理制度，确保系统的正确使用和维护；第三，普及应用是为了将电子病历系统普及应用到全国各地的医疗卫生机构中，让更多的患者受益于电子病历的优势；第四，加强培训是为了加强电子病历系统的培训工作，让医务人员更好地掌握和使用电子病历系统。

在电子病历推广阶段，我国的医疗信息化政策也随之不断完善，并大力推广电子病历的应用。通过大力投入，电子病历的研发、建设和推广得到了极大的推动，并逐渐成为了医疗信息化的重要组成部分。在推广过程中，各级政府不断加大投入，社会各界也不断加大关注，并对电子病历的安全性和数据的保密性提出了更高的要求，这也促进了电子病历的技术更新和完善。总的来说，电子病历在这一阶段得到了长足的发展，为我国的医疗信息化做出了积极的贡献

### 1.2.3规范建设

在电子病历的规范建设阶段，我国医疗信息化的政策更加明朗，电子病历的应用也更加普及。在这一阶段，各级政府颁布了一系列关于电子病历的法律法规和技术标准，为电子病历的发展和应用创造了良好的法律环境。同时，国家卫生健康委员会等部门也不断加强对电子病历的管理和监督，以确保电子病历的安全性和数据的真实性。

除此之外，在这一阶段，我国也开展了大规模的电子病历建设试点，不断提高电子病历的应用水平。各级医院也积极开展电子病历的建设和推广工作，并在实践中不断完善电子病历的技术和管理模式。这有助于提高电子病历的安全性、可靠性和效率。总的来说，电子病历在规范建设阶段得到了更加全面和深入的发展，为我国的医疗信息化做出了更加积极和显著的贡献。

## 1.3.现存的问题

信息技术在医疗上的应用主要包括疾病预测、医学影像识别、药物发现等场景。而基于医疗的算法或者统计往往需要大量的样本、单一的数据源很难满足数据的需求。这就需要有一个能够共享的数据医疗平台来实现数据的不断的拟合、打破数据的孤岛。目前电子病历虽然在改善医疗服务和信息管理方面取得了很大进展，但仍然存在一些问题需要解决。这些问题包括：

（1）数据安全：电子病历中存储的数据很敏感，如果被非法获取，会对患者的隐私产生严重的影响。

（2）数据准确性：电子病历中的数据是由医生和护士等人员输入的，很容易发生输入错误，导致数据不准确。

（3）数据兼容性：不同医院之间电子病历系统通常不相互兼容，很难将患者的病历数据共享和对接。

（4）技术应用难度：对于一些没有专业技术人员的医院，推广电子病历很难，也很难维护。

（5）接受程度：许多医生和护士仍然对电子病历系统的使用持有怀疑态度，电子病历的接受度和使用效果仍然需要提高。

## 1.4.拟解决问题

为解决上诉问题，本项目设计了一种基于联邦学习的电子病历共享系统来实现数据共享。它的核心思想是：不同的医院作为参与方，拥有自己的本地数据源，通过使用联邦学习技术，这些参与方可以将各自的数据进行联合建模和训练，从而得到一个全局的模型。这个过程中，原始数据不需要出本地也不需要流转，只需要通过交互模型中间参数进行模型联合训练，实现了数据的隐私保护，达到“数据可用不可见”的数据应用模式。

然而模型的精度也至关重要，只有经过大量数据的训练，才能获得好的模型，因此数据的准确性也至关重要。由于法律监管、数据隐私安全等方面的顾虑，个数据拥有者也不愿意直接交换原始数据，导致数据无法有效汇聚，从而影响机器学习的效果，因此我门的数据基于联邦学习来解决该问题。

基于这种模式该项目有以下三个优点：首先，它可以避免敏感数据泄露的风险，因为原始数据不需要流转。其次，它可以提高数据的利用价值，因为多个参与方的数据可以被联合起来训练模型，从而得到更加准确和鲁棒的模型。最后，它可以促进医学研究和临床实践的合作，从而推动医学领域的进步和发展。

## 1.5.应用前景分析

除了达到国家卫健委要求的电子病历评级要求外，医院的电子病历建设是长期化需求，也是建立智慧医院的必然要求。电子病历系统的发展满足了现代医院发展需求、医疗信息化自身发展要求以及病人患者诉求的需要，具有良好的需求前景。下面将从五个方面分析它的前景应用。

（1）提升医疗质量和效率：电子病历可以记录患者的全面健康信息、疾病诊断、治疗方案、药物使用等内容，这些信息可以方便医生查阅和使用。医生可以根据患者的历史病历和治疗情况制定更加准确和个性化的治疗方案，从而提高医疗效果。此外，电子病历还能够避免重复检查和治疗，提高医疗效率。

（2）促进医疗协同和共享：通过电子病历，多个医疗机构之间可以实现患者信息的共享，包括病历、检查结果、医嘱等信息。这些信息的共享可以促进医疗协同和合作，提高疾病的诊断和治疗水平。特别是在紧急情况下，通过电子病历可以快速获取患者信息，提高医疗决策的准确性和速度。

（3）改善患者体验和参与度：通过电子病历，患者可以更好地了解自己的病情和治疗情况。患者可以随时查看自己的病历和医嘱，了解自己的病情和治疗进程。患者也可以参与治疗方案的决策，通过电子病历与医生进行在线沟通和交流，提高患者的治疗满意度和参与度。

（4）促进健康管理和预防：通过电子病历，可以实现患者健康信息的实时跟踪和管理，包括个人基本信息、生活习惯、体征指标等内容。通过对患者的健康信息进行分析和挖掘，可以为健康管理和疾病预防工作提供数据支持，提高疾病预防和管理的效果。

（5）促进医学研究和创新：电子病历可以收集、分析和挖掘大量的医疗数据，包括患者的病历、诊疗数据、医学影像等内容。通过对这些数据进行分析和挖掘，可以发现疾病的规律和趋势，为医学研究和创新提供更加丰富和准确的数据支持。同时，医生和研究人员可以利用电子病历中的数据，对疾病的发病机制、治疗方案、药物研发等进行探索和研究，为医学研究和创新提供支持和基础。

（6）加强医疗安全和隐私保护：电子病历可以通过权限设置、加密保护等措施保护患者的隐私，避免患者信息被非法获取和使用。此外，通过电子病历，医生可以快速了解患者的病历和治疗情况，避免了因为病历遗漏或不完整等原因导致的医疗事故。

# 2.系统概要设计

本项目提出了安全高效的基于联邦学习的病历共享及预测系统如图1所示，解决了医疗数据孤岛问题，允许医疗机构在保证隐私的情况下分享经验。能够使得不同医院的医疗情况（包括患者病情，治疗方案等）在保证安全性的情况下进行数据流转，能够保护患者隐私，而且能够在患者到其他医院或者同其他医院协助治疗时，不泄露数据的情况下依旧能够对医疗数据进行使用，安全高效。联邦学习可以绕过医疗机构之间的信息壁垒，不考虑将各自数据做合并，而是通过协议在其间传递加密之后的信息，该加密过程具有一定的隐私保护机制，保证加密后的信息不会产生数据泄露。各个医疗机构通过使用这些加密的信息更新模型参数，从而实现在不暴露原始数据的条件下使用全部患者数据的训练过程。

该系统能够实现如下功能：

（1）医院之间对患者的病历信息不能直接进行数据交换通过密码学协议在数据交互过程传递加密信息，数据共享过程中患者隐私受保护。

（2）能够在使用对方数据时候，通过加密算法对双方医疗样本数据进行对齐求交，并共享双方特征指标。

实现的安全性目标如下：

（1）对患者个人以及对医疗机构在共享经验过程中必须保证数据隐私。

（2）除了病患和获得授权的申请者没有人可以对应二者的联系，也无法获得解密后的数据信息。

## 2.1.训练模型过程

(1）联邦学习问题建模:首先，针对本文的研究方向，本文要完成的是电子病历交互任务，参与方就是拥有电子病历的各医疗机构。对于该任务而言，各个医疗机构的用户群体可能分别来自不同的城市或地区，相互的交集很小，但是他们的数据极度相似,都属于同一类医疗数据。所以根据这样的数据分布特征，我们选用横向联邦学习框架来完成医疗图像分类任务。其次，对于具体框架类型的选择，考虑到在实际应用中，如果选择对等模型架构，各医疗机构并不能同时训练模型，而是需要收到上一个医疗机构训练好的模型之后，然后再用自己本地的数据训练，并不能同步进行，这使得等待时间过长。尤其是当医疗机构数量比较多时，这种架构会耗费非常多的时间。除此之外，具体来看，由于没有中心服务器，分类模型的参数并不能分批更新，而是连续更新，这也将导致训练模型的时间更长。所以本文最终选择使用客户-服务器架构。

(2）参与方初始化:首先，各医疗机构与中心服务器建立不同的TLS/SSL通道，以保证后续的通信安全;其次，各医疗机构在中心服务器的协调下共同研究并确定一个深度学习电子病历交互模型;然后将所有的电子病历数据按比例分为训练集、验证集与测试集;最后，为了保证整个联邦学习系统中数据的隐私安全。

(3）联邦学习模型训练:在启动整个模型的训练流程后，中心服务器和各医疗机构按照既定的横向联邦学习流程训练模型，简单来说就是每个医疗机构通过本地的分类模型对电子病历数据进行特征提取和向量化处理，得到诊疗结果向量，然后将训练后得到的模型参数提交给中心服务器。中心服务器则不间断地收集和聚合来自每个医疗机构的参数，进而构建一个全局模型，然后发送给所有的医疗机构。在这个过程中，训练数据对于每个医疗机构来说都是私有的，在学习过程中不会被共享，共享的仅仅是模型的可训练权重，从而保持了患者数据的私密性。

(4）联邦学习模型评估:为了更好地保证各医疗机构数据的隐私性，在每一个全局轮次结束后，对该轮得到的交互模型用测试集测试，保存记录下该模型的参数以及性能指标。然后，当所有的全局轮次都结束后，在其中选择性能最好的模型最为最后的部署模型。

(5）联邦模型部署:在训练过程结束以后，各医疗机构就可以将上个步骤中得到的最好的模型部署到自己的平台上。因为模型部署阶段和联邦学习模型训练与构建阶段是相对独立的过程，所以模型部署中可以使用传统的部署流程和技术。通过以上具体详细的构建流程，我们就可以得到一个基于横向联邦学习的电子病历交互模型。通过此模型，各医疗机构实现了模型参数信息的流通和共享，从而打破医疗“数据孤岛”，做到“数据可用不可见”，在有效保护病人隐私的同时获得更多更精准的诊断模型。

## 2.2.功能模型

（1）疾病风险预测：根据从不同医疗机构获得的针对同一种疾病的不同诊断结果、处方药以及身体状况等情况，在中心服务器训练后分发到各医疗机构，可以对新患者疾病情况进行比较或者排除一下错误的诊断，可以减少患者去不同的医疗机构进行重复检测的情况，提高对患者的治疗效果，减少疾病恶化的时间。

（2）临床诊断：运用联邦学习可以用于预测新病例的诊断结果，当判断类似病情时，可以将中心拟合后的结果呈现到当前的判断，增加诊断的精确性。另外，该模型能更好地预测少数人群中的疾病发病率和诊断准确性，从而减少由于人口结构差异而导致的误诊和漏诊。

（3）医学影像识别：医学影像数据量庞大且多变，联邦学习可以实现增量学习，即在模型不断学习的过程中逐渐适应新的医学影像数据，从而保证对该种影像的判断，对当前影像的判断提供了参考增加了普适性。

传回机构

医疗机构A

医疗机构B

各方数据不出域

模型A

模型B

加密样本对齐

加密模型训练

中央服务器

新的全局模型参数

传回机构

图1.系统模型图

总之，电子病历的应用场景非常广泛，可以提高医疗质量和效率，促进医疗协同和共享，改善患者体验和参与度，促进健康管理和预防，加强医疗安全和隐私保护，以及促进医学研究和创新等。

# 3.总结

电子病历是信息技术和网络技术在医疗领域的必然产物，是医院病历现代化管理的必然趋势，其在临床的初步应用，极大地提高了医院的工作效率和医疗质量。但目前大部分医院面临着电子病历数据不开放共享的现实问题，因此医生临床诊断的准确率和效率无法得到质的提升。通过利用隐私计算的技术，可以实现区域内医疗信息的安全共享（如：远程病患信息的传输和共享），在隐私计算平台之中可以给医生提供患者各种疾病发生的概率(如传染病预警信息、危重症提示信息等），从而帮助医生更好的对患者进行临床诊疗和指导。通过联邦学习将所有孤立的医疗机构等连接起来，使得他们分享有隐私保障的数据，实现了将以上各维度医疗数据进行有效融合与分析，破解了医疗领域数据安全与隐私保护的难题，并生成更精准的病例预测模型，助力医疗机构进行更好的医疗服务。