# Abstract

深度学习已经成为医学图像分析中的重要工具，在病理图像分析领域也不例外。病理图像是临床医生用于诊断疾病的重要工具，然而，由于病理图像的复杂性和多样性，人类的识别和分析难度很大。深度学习在这个领域中的应用使得病理图像的自动化分析和诊断成为可能。癌症是一种高度异质性的疾病，是导致死亡的主要原因之一，利用深度学习可以为该类病病理分析的有价值的指南，组织病理学图像的准确分类可用于开发自动化和精确的癌症早期检测分级系统以及准确的预后，在本综述中，我们将讨论深度学习在病理图像分析中的方法。

# Introduction

机器学习(ML)可以通过从数据中学习而不显式编码来开发和应用算法来解决模式识别问题。卷积神经网络(Convolutional neural networks, CNNs)是一种人工神经网络，具有两层以上的神经元，并添加了用于自动特征提取的卷积层。DNs由深度学习(DL)算法训练，该算法自动识别数据中的一致表示(表示学习)。DN模型学习直接从数据中提取“高级”特征，并从以前的“经验”中提高准确性(即广泛的数据可视化)，这是对人脑学习功能的终极模拟。几十年前，机器学习(ML)方法，开始用于乳腺癌图像分类。深度学习(DL)被引入来克服ML模型的准确性问题。DL允许大量的数据学习，也提高了声音和图像识别能力。本文探讨了利用DL方法进行组织病理学乳腺癌图像分类的最新技术进展。

# 方法

①卷积神经网络

卷积神经网络（CNN）是深度学习中最为常见的一种网络结构，它在病理图像的分割、分类和检测等任务中得到了广泛应用。例如，在肺癌病理图像中，CNN可以识别细胞的形态和组织结构，并将其分为恶性和良性细胞。在结直肠癌的病理图像分析中，CNN可以通过识别不同的组织类型来分割出肿瘤区域。

②生成对抗网络

生成对抗网络（GAN）是另一种深度学习技术，它可以用于生成具有高度相似性的图像，这些图像可以用于增加病理图像的数据量。例如，在乳腺癌的病理图像分析中，GAN可以生成具有类似于真实乳腺癌图像的图像，从而提高算法的准确性。

③迁移学习

由于深度学习需要大量的标记数据，因此在病理图像分析中，标记数据的获取成为了一个瓶颈。迁移学习可以通过将已经训练好的模型应用于新的领域，从而减少标记数据的需求。例如，在结直肠癌的病理图像分析中，迁移学习可以将已经训练好的神经网络应用于肝癌的分析中。

⑤MixPatch

一种用于组织病理学图像分类的新方法，基于卷积神经网络的图像处理已被积极应用于组织病理学分析，实现癌变肿瘤的自动检测和分类。然而，基于卷积神经网络的分类器对标签的预测往往过于自信，这在医学领域成为一个严重的问题。MixPatch是一种新的训练方法，它为每个小批生成并使用一个新的子训练数据集，该数据集由混合补丁及其预定义的真实值标签组成，混合斑块是使用病理学家确认的小尺寸干净斑块生成的，而它们的真实标签是使用基于比例的软标记方法定义的。这种方法能够通过专门解决预测不确定性问题来改进基于卷积神经网络的分类器，并检查其在组织病理学图像分析背景下提高诊断性能的有效性。

⑥带有自我注意的弱监督组织病理学图像分割

组织病理学图像像素级的精确分割在数字病理工作流程中起着至关重要的作用。弱监督组织病理学图像分割方法的发展将病理学家从耗时和劳动密集型的工作中解放出来，为全玻片组织病理学图像的进一步自动化定量分析提供了可能。多实例学习(MIL)作为一种有效的弱监督方法，在组织病理图像处理中取得了巨大的成功。

通过将像素作为实例，可以将组织病理学图像分割任务转化为MIL中的实例预测任务，但MIL中实例之间缺乏关系限制了分割性能的进一步提高。因此，一种新的弱监督方法，称为SA-MIL应运而生，用于组织病理学图像的像素级分割。SA-MIL在MIL框架中引入了一种自我关注机制，它可以捕获所有实例之间的全局相关性。该方法通过自注意模块(SAMs)将自注意机制整合到MIL中，该模块自适应地从全图像依赖中捕获上下文信息。MIL中独立实例之间的关系，具有更好的可解释性，极大地促进了分割结果。

挑战问题及可能研究的方向

①病理医学图像的准确和自动分类一直是DL中最重要的挑战之一。

可能的研究方向：集成学习和嵌入式融合模型比其他集成方法表现出更好的性能。此外，具有模型融合的CNN是精确的特征提取和组织病理学图像分类的强大工具。采用在线相互知识转移策略作为嵌入CNN的融合策略的建议可能对其他类型的乳腺癌检测有希望。

②模型可解释性

深度学习模型通常被视为黑盒子，难以解释其决策过程。在病理图像分析中，模型的可解释性非常重要，因为医生需要了解模型是如何做出决策的。因此，研究人员需要探索如何提高深度学习模型的可解释性，例如使用可视化技术和解释性机器学习方法。

③数据收集和管理

虽然越来越多的病理图像数据被数字化，但这些数据通常来自不同的医疗机构和国家，因此数据的质量和一致性可能存在差异。收集和管理大量的病理图像数据需要投入大量的人力和资源，因此研究人员需要寻找更加高效和经济的数据收集和管理方法。

④数据隐私和安全性问题

由于病理图像包含患者的个人身体信息，因此数据隐私和安全性一直是病理图像分析中的重要问题。在未来，随着越来越多的病理图像数据被数字化并在互联网上共享，保护患者数据的隐私和安全性将成为更加紧迫的问题。研究人员需要采取措施来保护数据隐私，例如采用匿名化技术和数据加密技术等。

# 结论

这篇综述论文旨在帮助创建更好的癌症分类设计和方法，以帮助癌症的识别过程。此外，通过综合比较融合各种方案，可以针对组织病理学图像中癌细胞的检测和分类找到更精确、更完善的方案，可能会导致癌症的早期检测和人类生存率的提高。