SSI-GAN：用于神经元尖峰分类的半监督式受 Swin 启发的生成对抗网络

Danial Sharifrazi[1*]，Nouman Javed[1,2], Mojtaba Mohammadi[1], Seyede Sana Salehi[3], Roohallah Alizadehsani[1], Prasad N. Paradkar[2], U. Rajendra Acharya[4, 5], Asim Bhatti[1]

[1]智能系统研究与创新研究所（*IISRI*），迪肯大学，吉朗，澳大利亚[2]澳大利亚联邦科学与工业研究组织卫生与生物安全部门，澳大利亚动物卫生实验室，吉朗，澳大利亚[3]计算机工程系，伊斯兰自由大学伊斯兰沙赫尔分校，伊斯兰沙赫尔，伊朗[4]数学、物理与计算学院，南昆士兰大学，斯普林菲尔德，澳大利亚[5]健康研究中心，南昆士兰大学，斯普林菲尔德，澳大利亚

摘要

蚊子是虫媒病毒性疾病的主要传播媒介。人工分类其神经元尖峰模式非常耗费人力且成本高昂。大多数现有的深度学习解决方案都需要完全标注的尖峰数据集和经过高度预处理的神经元信号，这降低了其在实际实地场景中大规模应用的可行性。为解决带标签数据稀缺的问题，我们提出了一种新的生成对抗网络（GAN）架构，称为**半**监督式**Swin-启**发**生成对抗网络**（SSI-GAN）。该网络采用受 Swin 启发的移位窗口判别器以及基于 Transformer 的生成器，用于分类神经元尖峰序列，进而检测病毒嗜神经性。我们在一种扁平的、基于窗口的 Transformer 判别器中采用多头自注意力模型，使其能够学习捕捉较稀疏的高频尖峰特征。仅使用 1% 至 3% 的有标签数据，SSI-GAN 即利用在感染后五个时间点采集的超过 1500 万个尖峰样本进行训练，并将记录分类为寨卡病毒感染、登革热病毒感染或未感染类别。我们使用贝叶斯Optuna框架优化超参数，并通过五折蒙特卡洛交叉验证验证模型的鲁棒性。感染后第三天，仅使用 3% 的有标签数据，SSI-GAN 的分类准确率便达到 99.93%。仅凭 1% 的监督数据，它在感染的各个阶段都保持了较高的准确率。与性能相当的标准监督方法相比，这使人工标注工作量减少了 97%–99%。本文提出的移位窗口Transformer设计大幅超越所有基线，在基于尖峰的神经元感染分类方面创下新的最佳成绩。本文提出了一种鲁棒且数据需求量大的神经病毒学方法，

* 通讯作者：D. Sharifrazi，智能系统研究与创新研究所（IISRI），迪肯大学，澳大利亚维多利亚州吉朗 3216。电子邮箱：d.sharifraz@deakin.edu.au（D. Sharifrazi），n.javed@deakin.edu.au（N. Javed），m.mohammadikarachi@deakin.edu.au（M. Mohammadi），sanaa.salehi@iau.ir（S. Salehi），r.alizadehsani@deakin.edu.au（R. Alizadehsani），prasad.paradkar@csiro.au（P. Paradkar），Rajendra.Acharya@unisq.edu.au（R. Acharya），asim.bhatti@deakin.edu.au（A. Bhatti）。

研究表明，将 Transformer 注意力与半监督对抗学习相结合，即使只有极少的专家标注，也能有效检测神经元尖峰数据中细微的病毒特征。
关键词：半监督学习、*Transformer*模型、生成对抗网络、神经尖峰分类。

1. 引言

寨卡病毒（ZIKV）和登革热病毒（DENV）属于黄病毒科和正黄病毒属[1]。该科的其他病毒包括西尼罗病毒（WNV）、日本脑炎病毒（JEV）和黄热病毒（YFV）[1]。ZIKV于1947 [2]在乌干达的一只恒河猴中被发现。自2007年以来，非洲、美洲、亚洲和太平洋地区均曾暴发疫情，随着时间推移已成为严重的全球健康问题[3]。截至目前，该病毒已在50多个国家被记录，全球约有3.9亿人感染[4]。在此期间，人们广泛观察到严重的神经系统疾病，如吉兰-巴雷综合征、脑炎、脊髓炎和神经痛，以及小头畸形和眼部畸形等发育问题[5]。

寨卡病毒主要由埃及伊蚊传播 [6]。它也可以通过母婴传播、性接触和输血在人与人之间直接传播 [7]。研究病毒如何影响蚊子的行为和神经系统（这是病毒生命周期的一部分），将有助于我们了解蚊媒疾病的传播 [8]。哺乳动物和昆虫的神经系统在进化上存在许多相似之处 [9]。两者使用相同的化学信使和受体 [10]。两类生物的受体基本结构相似，因此我们可以比较它们体内的功能 [11]。它们的神经系统在结构和功能上也有相似之处 [12]。这两类生物都具有中枢神经系统和周围神经系统 [13]。虽然存在一些差异，但这些相似之处表明，研究昆虫中的寨卡病毒和登革病毒神经活动，有助于了解它们在人类中的发病机制和感染机制 [14]。

寨卡病毒对神经系统的影响已通过多种模型得到广泛研究，包括使用人多能干细胞（hPSC）衍生的神经祖细胞和类器官开展的体外研究，以及使用不同病毒株开展的小鼠模型体内研究 [15]。此前一项聚焦于蚊子生命周期的研究，利用微电极阵列技术记录了未感染、感染寨卡病毒和感染登革病毒的埃及伊蚊产生的神经尖峰 [16]。对行为和神经尖峰的分析显示，病毒具有嗜神经性，会影响蚊神经元，并导致神经活动增强和行为变化 [16]。这些变化包括飞行运动发生改变、叮咬次数增加、取食增多，以及对繁殖性能的影响，例如生殖力和生育力 [16]。这些变化可通过增加蚊人与人相互作用，使病毒更容易传播 [16]。

在分析神经系统行为的各种方法中，神经尖峰分类是一种有效的方法，因为它能分离尖峰——大脑传递信息的信号 [17]。研究蚊虫神经元活动及神经元的尖峰分类，对于理解寨卡病毒和登革病毒的病毒感染机制至关重要 [18]。利用深度学习对这些昆虫的尖峰进行分类，有助于我们了解这些病毒如何影响蚊子和人类 [19]。深度学习能够实现数据处理自动化，并提高

通过减少分类任务中的人为错误来提升准确率[20]。此外，它还能识别神经元活动中的模式和相关性，而这些模式和相关性可能难以通过传统方法轻易发现 [21]目前几乎没有研究关注利用半监督学习对昆虫数据进行基于深度学习的尖峰分类，以检测寨卡病毒和登革病毒[18] 。

综述显示，已有多种传统方法用于尖峰分类 [22]。例如，Lewicki采用贝叶斯方法 [23], , 而Letelier和Weber采用小波分析，根据神经尖峰的时频特征对其进行分类，从噪声和重叠活动中分离信号 [24]。Haas、Cohen等人提出了一种基于方差的模板匹配模型，该模型使用CMOS方差估计器对数据进行实时分类 [25]。Takekawa、Isomura等人引入了一种混合算法，将多模态加权主成分分析与聚类技术相结合 [26]。传统模型在某些基准测试中表现良好，但通常准确率较低且泛化能力较差，而且大多数情况下只能在特定数据集上有效运行 [27]。另一方面，深度学习模型展现出令人印象深刻的性能，这得益于其过滤噪声、增强信号解码并产生更可靠结果的能力 [28]。例如，Meyer、Schanze等人 [28] 开发了一种基于尖峰波形的单层多层感知机网络。结果表明，该模型性能良好，其准确率根据噪声水平不同介于89%至95%之间 [29]。Liu、Feng等人提出了一种使用卷积神经网络（CNN）和长短期记忆网络进行尖峰分类的方法 [30]。Li、Tang等人提出了一种用于不平衡数据尖峰分类的混合深度强化学习方法 [31]。在他们的研究中，该问题被建模为马尔可夫序列决策，并在训练期间使用动态奖励函数来提高数据中少数类的影响 [31]。Saif-ur-Rehman、Ali等人推出了“SpikeDeep-Classifier”，这是一种将监督学习与无监督K均值聚类相结合的自动尖峰排序算法。其核心思想是对所有数据集采用相同的超参数设置 [32]。Park、Eom等人使用全连接神经网络对细胞外记录中的神经尖峰进行分类，该方法能够有效处理形状相似的尖峰和高噪声水平 [33]。他们利用主成分分析和聚类生成伪标签，研究表明这种方法能够有效处理未标记数据 [33]。García、Suárez等人提出了一种混合模块化计算系统，用于在不了解波形结构的情况下对多单元神经记录中的尖峰进行分类 [34]。所提出的系统由预处理模块和处理模块组成，并使用混合无监督多层人工神经网络来识别特征、聚类尖峰和过滤噪声 [34]。Bernert和Yvert开发了一种利用脉冲时序依赖可塑性的脉冲神经网络，用于实时无监督尖峰排序 [35]。结果显示，该方法在噪声数据上具有较高的准确率，能够适应不同的尖峰模式，并适用于多通道记录 [35]。

如上述研究所示，深度学习在尖峰分类方面非常有效 [36]。然而，处理寨卡病毒和登革病毒数据时面临的一项主要挑战是标注 [36]。大型数据集的标注过程需要大量实验室工作，并且在很大程度上依赖熟练技术人员对复杂的生物数据进行准确分类 [37, 38]。这种人工流程不仅耗时且成本高昂，而且随着数据量增加，也难以扩展 [37, 38]。解决这一问题的有效方法是采用半监督算法 [37, 38]。半监督学习技术使用少量有标签样本以及大量无标签数据，从而减轻实验室工作负担、降低成本并提高模型准确率 [37, 38]。半监督算法的成功应用案例包括文本情感分析 [39], 语音识别 [40], 和 DNA测序 [41]。鉴于此，本文采用半监督技术进行数据准备和标注。

鉴于分析蚊子神经活动对于研究病毒感染的重要性，本文提出一种混合方法，将半监督技术与深度学习相结合，用于对感染寨卡病毒和登革病毒的蚊子进行神经尖峰分类。值得注意的是，本文采用了最先进的深度学习方法，并使用实验室生成的实验数据训练模型；这些数据包含未感染、感染登革病毒以及感染寨卡病毒的埃及伊蚊的神经尖峰信息。据作者所知，这是首次运用半监督技术和先进深度学习模型，研究昆虫的尖峰分类问题。通过将半监督技术与深度学习相结合，本研究旨在弥补神经尖峰分类研究领域现有的不足。

2. 数据收集

2.1. 病毒株及制备

寨卡病毒于2010年从一名柬埔寨人身上分离（GenBank KU955593）。该病毒在白纹伊蚊 C6/36细胞中传代数次，并于感染后72小时收获。登革病毒2型分离自一名在东帝汶感染的澳大利亚士兵，并在C6/36细胞系中扩增了七次。使用L15培养基制备这些病毒，随后用于感染蚊子和原代神经元培养物。

2.2. 蚊子饲养与感染

埃及伊蚊在受控实验室条件下培养，温度为26°C、湿度为60–70%，并采用14小时光照、随后10小时黑暗的光周期模拟自然黎明与黄昏。在第7天，将感染性血液以37°C喂给雌性蚊子，持续一小时，并筛选出受感染的蚊子用于后续实验。

2.3. 免疫组织化学（IHC）

我们使用免疫组织化学检测了11只寨卡病毒感染组和4只对照组蚊子的头部切片中的病毒，使用的4G4抗体稀释比例为1:100。我们在97°C下进行20分钟抗原修复，并封闭内源性过氧化物酶。随后，使用EnVision FLEX+ Mouse Linker和HRP标记二抗进行信号放大，并以AEC显色底物显色。之后，使用徕卡显微镜在4× 和40× 倍放大下观察并拍摄切片。

2.4. 原代神经元分离

羽化后3至4天，取出成熟雌性埃及伊蚊的大脑，并转移至L15培养液中；培养液含有胰蛋白胨磷酸盐肉汤（10%）、胎牛血清（10%）、青霉素–链霉素（50 U/mL）和两性霉素B。收集15至20个大脑，在2 ml离心管中用轻柔吹打的方式机械性打散。离心两次后，将沉淀悬浮于50 µL新鲜培养基中，用于细胞培养。

2.5. 神经元培养与感染

按照标准流程，采用类似的方法制备原代神经元培养物，用于共聚焦显微镜观察、半数组织培养感染剂量测定和逆转录聚合酶链式反应。进行共聚焦检测时，将聚乙烯亚胺（PEI）涂布于盖玻片上，在 37oC 下处理 30 分钟，用无菌水清洗三次，然后在生物安全柜中晾干。用移液器将 25 µl 细胞混合液加入 24孔板的每个孔中。

将培养板置于 28oC 下孵育 30 分钟，以促进细胞黏附；随后加入 1 ml 含 10% 胎牛血清的 L15培养基。第 7 天，待神经元培养物成熟后，按感染复数（MOI）感染细胞。

微电极阵列（MEA）神经元培养方案与 Hales等人 [42]的方案相似，但作了以下修改：预包被时先使用聚乙烯亚胺（PEI），再使用层粘连蛋白；接种时采用不同的细胞密度；每隔3–4天更换一次培养基，但记录前一天除外。最后，感染时使用了不同体积的病毒。

2.6. 使用微电极阵列进行数据采集

A. 记录系统
使用多通道系统的MEA记录神经元网络活动。该阵列由60个平面电极组成，按方形网格排列，其中一个电极用作接地参考电极。使用MEA60inv系统在线记录神经元信号，并以30 kHz的采样率采集动作电位。

所有记录均在II级生物安全柜（BSCII）中进行，以避免病毒污染；记录温度为28°C，以优化神经元的存活和功能。放置MEA后等待5–10分钟，待神经元从机械应激中稳定下来后再进行记录。在感染后第0、2、3和7天记录自然神经活动，每次持续30分钟；在感染后第8天（dpi），进行加巴嗪诱导测试，记录15分钟。每次记录会话结束后，更换50%的培养基。对初始电压记录进行巴特沃斯高通滤波器处理，以检测低频噪声。

C. 记录时间线Af
ter 将神经元接种到多电极阵列上后，培养物在体外培养 7 天以形成网络。第 7 天，用寨卡病毒和登革病毒（作为对照组）感染神经元。每项实验均对神经信号进行了六次记录。首次记录于第 7 天感染前进行，并作为参考。感染后记录分别在第 2、3 和 7 天进行。

为以 30 kHz 的高采样率进行连续电记录，取出雌性蚊子的大脑，并在 MEAs 上培养神经元。随后，使用高通滤波器对记录进行滤波，以消除背景噪声。

在感染首日测定对照组、登革病毒2型和寨卡病毒各信号类型的标准差（std），分别为 6.1 μV、4.08 μV 和 5.44 μV。接下来，我们将标准差的平均值四舍五入后作为阈值。只有当信号超过设定为噪声水平两倍（10 μV）的阈值时出现的神经元尖峰才会被计数。

2.7. 最终数据集
从三个组别收集了 15,728,580 个数据样本，包括感染后不同天数（0、1、2、3 和 7 天）的寨卡病毒组、登革病毒组和对照组。每个组别均采集了 1,048,572 条记录。

3. 提出的方法

3.1 尖峰检测

第一步预处理旨在清理原始数据集并检测神经尖峰。尖峰检测首先对采集到的蚊子神经元记录应用巴特沃斯高通滤波器，截止频率为 [43]，采样率为 30 kHz [44, 45]。此举可抑制低频噪声，同时保留尖峰事件特有的高频成分。在滤波信号中，振幅大于 +10 µV 或小于 −10 µV 的信号被视为尖峰事件。所有其他数值均设为零，仅保留有意义的尖峰信息，同时去除基线噪声。由于所提出的模型使用尖峰序列而非原始信号进行训练，这一步可确保分类器专注于有信息量的神经活动。接下来，将清理后的尖峰信号分割为互不重叠的 100 个样本窗口，以构成用于分类的输入序列。该窗口长度通过实验确定：较短的序列无法捕获足够的时间结构，而较长的序列会增加计算量，却无法提升性能。由于各片段互不重叠，因此每个尖峰值只会出现在一个样本中。

3.2 数据标注

尖峰检测和分段后，数据集被划分为训练集和测试集，其中随机抽取 20% 的数据用于测试。对于半监督学习，仅从训练数据中随机选取并标注 3%，其余 97% 保持无标签数据状态。此外，从有标签数据子集中预留 1% 作为验证数据，以便在训练期间监测模型性能并降低过拟合风险。所有数据集的划分，包括训练集、验证集、有标签集和无标签集，均采用随机方式进行。

3.3 生成器模型

完成数据准备后，构建所提出的 **SSI-GAN** 中生成器和判别器的架构。尽管许多 GAN 研究采用基于卷积神经网络的生成器与判别器配对，但注意力机制在多个领域表现出色，因此本研究采用基于 Transformer 的组件。

研究了两种生成器变体：一种采用卷积层，另一种则融入注意力块。由于半监督生成对抗网络中的生成器主要用于支持判别器学习，而非生成高保真合成样本，因此两种设计下分类器的最终性能表现有所不同。我们将在消融研究部分进一步阐述核心架构设计假设。因此，生成器同时采用卷积层和 Transformer 风格层。

生成器接收一个 128 维的随机噪声向量，并通过稠密投影、重塑、Conv2DTranspose 层、批归一化、LeakyReLU 激活和最终的 tanh 输出激活，逐步将其转换为合成的 100×60 类尖峰序列。

3.4 判别器模型（受 Swin 启发的移位窗口设计）

与生成器不同，判别器充当**最终分类器**，因此是所提出模型性能的核心。由于输入尖峰序列较为稀疏，除尖峰位置外大部分都是零，因此模型必须能够有选择地关注包含信息的时间区域。这一需求与自注意力机制天然契合。

在本研究中，我们针对一维神经尖峰序列提出了一种受 Swin 启发的移位窗口 Transformer 判别器。与采用多阶段下采样的原始分层 Swin Transformer 不同，我们的判别器：

- 不进行分层合并，从而保留尖峰序列的完整100 × 60 分辨率，

- 将序列划分为固定大小的窗口，
- 在每个窗口内局部应用自注意力，并且
- 通过滚动操作使用移位窗口，从而在不同层之间交替改变注意力边界。

这种简化的移位窗口设计能够实现高效的局部建模，同时在交替层中支持跨窗口交互；对于可能出现在窗口边界附近的稀疏尖峰模式而言，这一点至关重要。由于没有采用分层下采样，该模型尤其适用于神经数据，因为保持时间粒度至关重要。最终形成的判别器架构包含：

- 一个输入投影层，
- 两个移位窗口的受 Swin 启发的Transformer块，
- 全局平均池化，
- 采用 LeakyReLU 和丢弃的全连接层，以及
- 生成对照组、登革热和寨卡分类结果的 Softmax层。

训练过程中还会加入生成器生成的虚假样本，以支持半监督学习。图1展示了完整的 SSI-GAN 架构。这种新颖的非层次式、受 Swin 启发并采用移位窗口的判别器，是 SSI-GAN 框架的一项关键贡献。

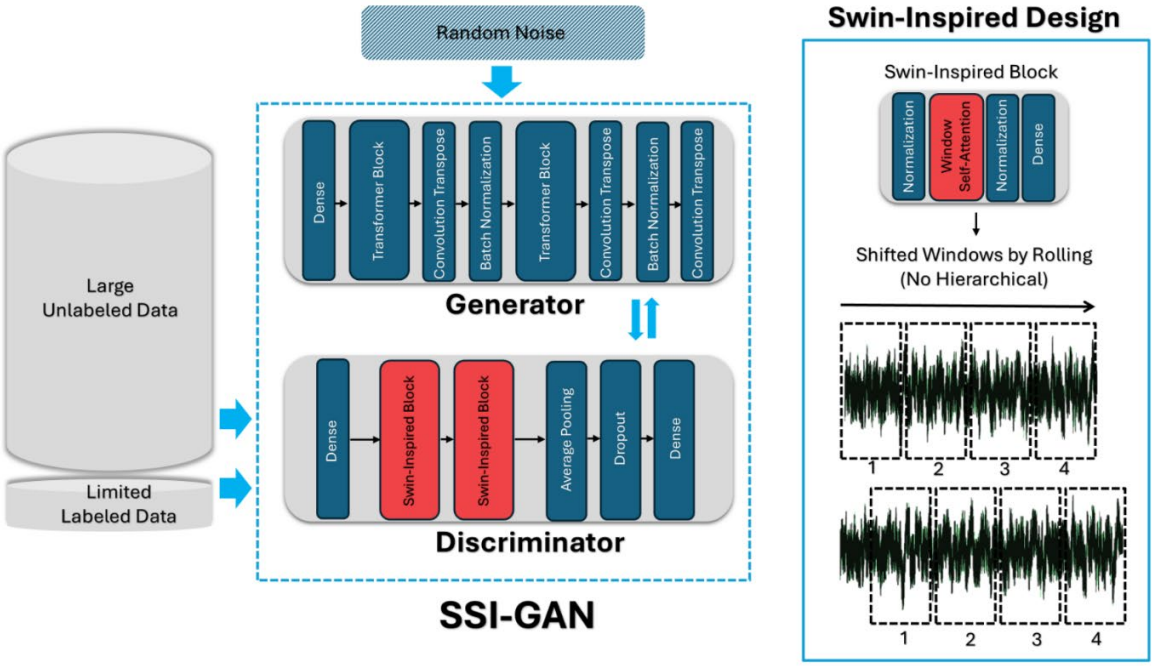

图1。基于受 Swin 启发的判别器所提出的 SSI-GAN 模型架构 .

3.5 超参数优化

鉴于生成器和判别器组件涉及大量超参数，我们采用了组合策略。关键超参数通过 Optuna 平台使用贝叶斯方法进行优化 [46], 而次要参数则根据已有文献手动调整 [47]。经过 20 次 Optuna 迭代，最佳配置达到了 99.48% 的准确率。表1总结了搜索范围和最优值。

表 1. 基于 Optuna 平台的模型超参数名称、范围和最佳建议汇总。

超参数名称	超参数范围	最佳结果
随机噪声的维度	[64, 100, 128]	128
Transformer头大小	[64, 128, 256]	64
头数	[2, 4, 8]	4
前馈维度	[32, 64, 128]	128
Transformer块数	[1, 2, 3, 4]	2
Dropout率	[0.1 – 0.5]	0.29
学习率	[1e-5 – 1e-3]	0.0009
批量大小	[64, 128]	128

3.6 模型训练与评估

使用选定的超参数，SSI-GAN 训练 500 次迭代。在整个训练过程中，生成器与提出的移位窗口Swin启发式判别器进行联合优化。对准确率曲线（图2）的监测证实，联合训练稳定且不存在过拟合。

为确保统计可靠性，采用蒙特卡洛 交叉验证 [48, 49], 将完整方法执行五次，每次运行都会重新随机划分所有数据（训练集、测试集、验证集、有标签数据和无标签数据）。最终结果以五次执行结果的均值和标准差报告。

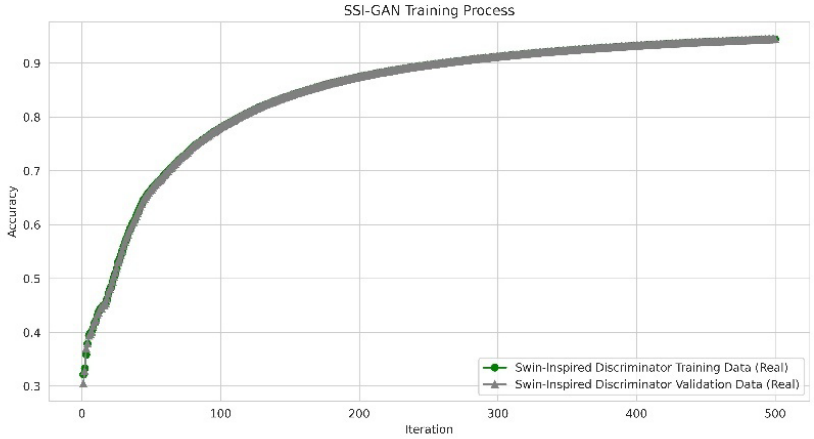

图2。生成器和判别器模型经过500次迭代后的训练过程。

3.7 方法总结

所提出方法的整体概览如图3所示，详细实施步骤以伪代码形式列于附录A。总而言之，该方法包括以下步骤:

1. 记录对照组、登革热组和寨卡组条件下的蚊虫神经信号。2. 应用高通滤波检测尖峰，并去除非尖峰值。3. 将尖峰信号分割为包含100个样本的非重叠窗口。4. 将数据集划分为测试集、有标签集（3%）、验证集（有标签数据的1%）和无标签集。5. 构建并优化 SSI-GAN，其中包含一种新颖的移位窗口、受 Swin 启发的判别器。6. 采用半监督学习训练模型，使有标签数据、无标签数据和合成数据协同提升分类效果。7. 使用蒙特卡洛交叉验证，将整个流程重复五次。图3总结了 SSI-GAN 的完整方法。

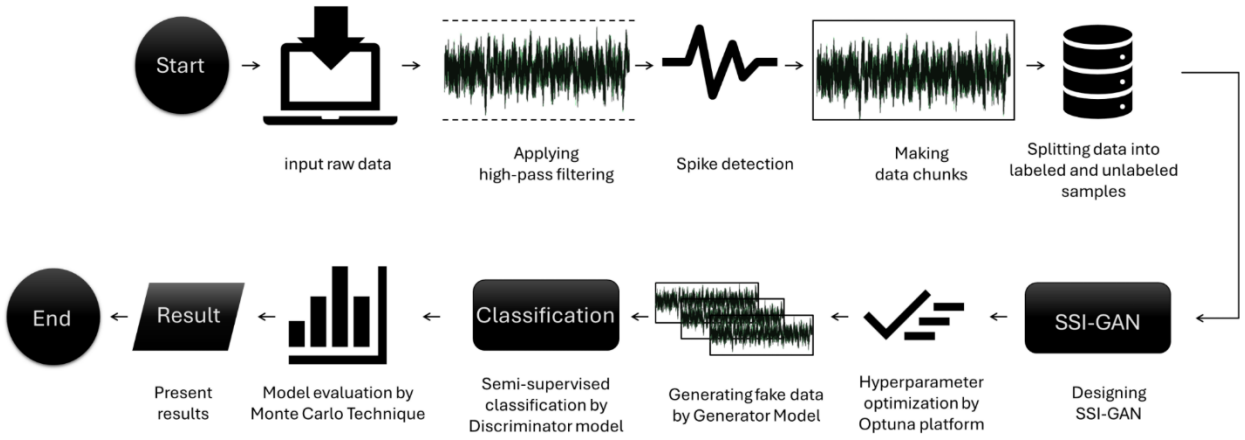

图3。所提出方法的简要步骤。

4. 实验结果

4.1 全监督架构结果

我们的实验数据集包含从埃及伊蚊采集的15,728,580个神经尖峰样本。数据集由三个实验组组成：寨卡病毒感染组、登革病毒感染组和未感染对照组。神经记录使用60通道微电极阵列采集，该阵列包含59个按方形网格排列的平面电极。动作电位的采样率设为10 kHz。在多个时间点采集数据样本，包括第0天（感染前）、感染后第1天、第2天、第3天和第7天（感染后天数）。每组在每个时间点贡献了1,048,572条记录。在预处理流程中，我们以30,000 Hz的采样率应用了截止频率为700 Hz的巴特沃斯高通滤波器，从而有助于从背景噪声中分离神经尖峰。随后，识别出振幅超过基线波动±10 个标准差的信号，并将所有非尖峰值设为零。接着，将连续信号分割为由100个时间单位组成的非重叠序列。此分割方式经经验确定，旨在平衡模式识别能力与计算效率。

按照半监督学习方法，随机选取总样本数的20%作为测试集，将数据集划分为不同部分。剩余的80%作为训练数据。 此外，从已标注子集中留出1%作为验证数据，用于评估训练进度并防止过拟合。所有数据划分均采用蒙特卡洛交叉验证，并进行5次随机重复。

我们使用四项标准分类指标评估模型：准确率、精确率、召回率和F1分数。结果报告为5次蒙特卡洛交叉验证的均值。对病毒感染五个时间阶段（感染后第0、1、2、3和7天）进行的全面评估显示，基线架构之间的性能存在显著差异。这凸显了所提出的 SSI-GAN 方法在神经元尖峰分类方面的优越性。表2展示了基线模型的定量结果，并将这些模型与我们提出的方法进行了比较。

表2。Optuna 平台上获得的各模型性能汇总（%）

dpi	*Evaluation Metric*	*LSTM*	*CNN-LSTM*	*RNN*	*CNN-RNN*	*CNN*	*GRU*	*CNN-GRU*	*SSI-GAN*
感染后第*0*天	***AC***	33.20	86.11	50.25	57.67	79.14	85.48	91.57	**99.45**
	PR	24.93	86.67	50.86	59.49	81.22	86.67	91.94	**99.46**
	RE	33.74	86.12	50.20	57.70	79.17	85.47	91.59	**99.46**
	FS	23.01	86.02	49.39	57.03	78.89	85.28	91.56	**99.45**
感染后第*1*天	***AC***	33.39	78.93	55.68	68.95	84.31	84.41	91.08	**99.61**
	PR	15.43	80.26	56.56	69.85	84.77	85.12	91.78	**99.62**
	RE	33.39	78.93	55.66	68.99	84.27	84.41	91.12	**99.61**
	FS	19.25	78.99	55.46	68.88	84.35	84.49	91.11	**99.62**
感染后第*2*天	***AC***	34.03	77.09	53.41	66.40	74.26	75.77	83.21	**97.09**
	PR	20.81	77.83	56.50	69.03	75.68	77.02	84.04	**97.34**
	RE	33.94	77.07	53.43	66.38	74.23	75.72	83.18	**97.07**
	FS	20.71	77.22	53.32	66.77	74.62	75.80	83.29	**97.10**
感染后第*3*天	***AC***	33.00	70.78	78.95	85.94	86.38	87.80	90.32	**99.82**
	PR	11.00	71.77	79.30	86.24	87.46	88.29	90.60	**99.82**
	RE	33.33	70.79	78.99	85.93	86.33	87.80	90.27	**99.82**
	FS	16.54	70.41	78.98	85.97	86.01	87.69	90.17	**99.82**
感染后第*7*天	***AC***	33.14	69.36	54.82	66.70	76.98	82.67	84.93	**99.93**
	PR	11.05	71.06	55.15	67.72	77.80	84.32	85.65	**99.93**
	RE	33.33	69.45	54.83	66.73	77.05	82.63	84.91	**99.93**
	FS	16.59	68.91	53.95	66.58	76.49	82.50	84.87	**99.93**

* 缩写：AC：准确率，PR：精确率，RE：召回率，FS：F1分数。

根据表2所示结果，简单的循环架构在不同实验时间点的表现始终不佳。长短期记忆网络（LSTM）的分类准确率最低，在所有时间点均介于33.00%至34.03%之间。这表明，在三分类任务中，它们仅略高于随机猜测水平。表现严重不佳的原因在于训练深度神经网络（如标准 LSTM 架构）时固有的梯度消失问题。循环神经网络（RNN）模型

准确率略有提升，在各感染阶段的准确率介于50.25%至78.95%之间。3 dpi（78.95%）的感染中期产生了包含足够事件起始间隔信息的尖峰模式，使基础循环处理能够有效检测感染。然而，RNNs在感染早期（1 dpi：55.68%）和晚期（7 dpi：54.82%）表现不佳，表明其对可变的事件起始间隔模式较为敏感。

混合卷积神经网络-循环神经网络架构的性能从0 dpi时的57.67%到3 dpi时的85.94%不等。这表明，将卷积特征提取与循环时序建模相结合，仅带来了中等程度的提升。在该方法中，卷积层能够有效提取局部尖峰波形特征，而循环组件则捕捉时间依赖关系。然而，该方法存在的问题是，RNN处理的序列特性限制了其实现并行化的能力，并降低了计算效率。此外，纯卷积神经网络架构表现出显著更强的性能，准确率介于74.26%至86.38%之间。卷积神经网络的最佳性能出现在3 dpi（86.38%），表明通过卷积滤波器进行空间特征学习，能够有效捕捉病毒嗜神经效应达到峰值期间具有可区分性的尖峰波形特征。然而，卷积神经网络在7 dpi时（76.98%）出现性能下降，这表明持续性感染会引入波形变异性，从而给纯粹的卷积方法带来挑战。

卷积神经网络-长短期记忆网络混合模型的结果不稳定，准确率介于 69.36% 至 86.11% 之间。这一发现凸显了尖峰分类任务中混合架构设计的复杂性。

门控循环单元（GRU）的性能始终优于循环神经网络，准确率介于 75.77% 至 87.80% 之间。门控循环单元在各感染阶段均保持稳定的性能，并在感染后第 3 天达到峰值准确率（87.80%）。此外，门控机制在保持计算效率的同时，也有效解决了长短期记忆网络中出现的梯度消失问题。

卷积神经网络-门控循环单元混合架构已成为表现最出色的传统监督基线，准确率介于 83.21% 至 91.57% 之间。该架构结合卷积神经网络空间特征提取、门控循环单元时序处理和门控机制，始终优于所有其他非生成对抗网络基线模型。在感染后第 1 天达到的 91.08% 峰值性能表明，该方法尤其适用于感染早期，此时尖峰模式开始呈现新兴的病毒特征。

大多数架构都取得了中等至较高的准确率，反映出实验组之间存在神经元变异性。性能通常在感染后1天（1 dpi）下降。这表明病毒感染已开始改变神经元动力学，但尚未形成充分的判别性特征。病毒嗜神经性的关键阶段发生在感染后2–3天（2–3 dpi），此时感染诱导的神经生理变化达到最显著的程度；因此，大多数架构在这一阶段的性能有所恢复。不过，许多架构在感染后7天（7 dpi）的性能略有下降，这表明慢性感染可能导致尖峰模式变异性增加，或使其趋于稳定。

我们的 SSI-GAN 在所有时间阶段和标注比例下始终优于所有基线架构。仅使用 3% 的带标签数据时，所提出的模型在 0、1、2、3 和 7 感染后天数分别达到了 99.45%、99.61%、97.09%、99.82% 和 99.93% 的准确率。与卷积方法或循环方法相比，Transformer架构的自注意力能力结合受 Swin Transformer 模型启发的移位窗口方法，能够捕捉尖峰序列中的长程时间依赖关系。多头注意力层处理尖峰模式的不同

方面，并识别出传统架构通常会忽略的细微判别特征。此外，半监督学习框架通过对抗训练有效利用了大量未标记尖峰数据。该特性在保持较高分类准确率的同时，降低了对昂贵、耗时且全面的标注阶段的依赖。下一节将展示所提出的 SSI-GAN 模型在不同带标签数据和无标签数据比例下的比较结果。

4.2 标记与未标记数据变化结果

图4比较了 SSI-GAN 模型在不同数量的带标签训练数据下，各项指标及五个感染后天数的平均值。评估了三种不同的标注情景：1%带标签数据（99%未标记数据）、2%带标签数据（98%未标记数据），以及3%带标签数据（97%未标记数据）。

带标签数据比例为2%和3%时，性能显著提升；但仅使用1%带标签数据时，所提出的模型与标准生成对抗网络表现相近。带标签数据比例为2%时，SSI-GAN 的准确率、精确率、召回率和 F1分数均高于97%。带标签数据比例为3%时模型性能最佳，四项指标均超过99%；与1%带标签数据相比，准确率约提升1.34%，其他指标也有相近幅度的提升。

即使监督程度较低（仅有 1% 的带标签数据），所提出的模型相较于第 4.1 节提到的全监督方法，仍保持了有竞争力的性能，在五个时间点的准确率均超过 96%。这些发现表明，即使带标签数据稀少，尤其是在采用先进的 SSI-GAN架构时，仍能实现鲁棒分类。对于电生理分析而言，带标签数据的采集成本高且耗时，这凸显了半监督学习框架的实用价值。

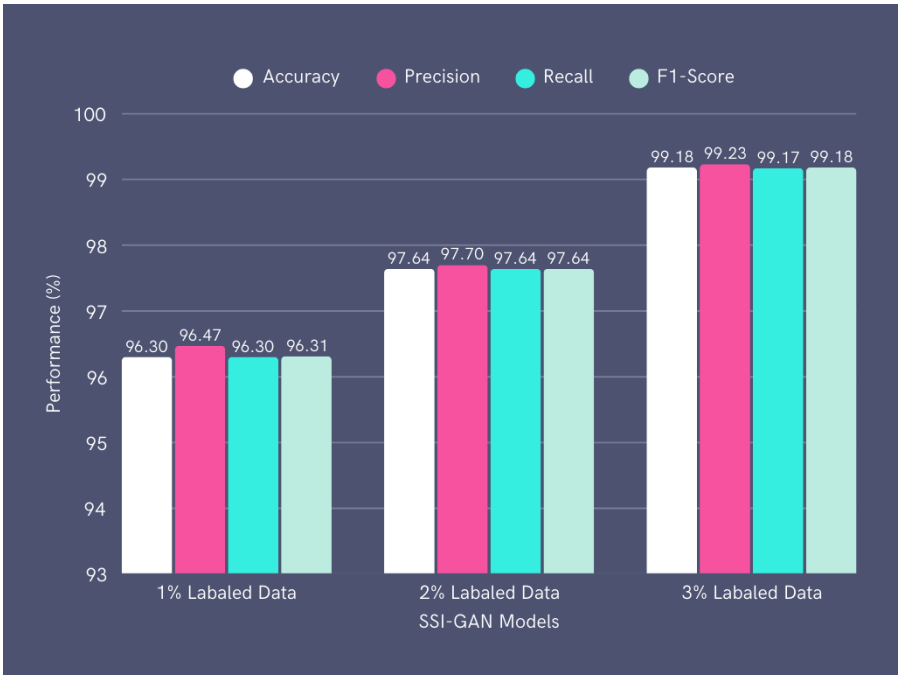

图4. 比较　　所提出的 SSI-GAN 模型在不同数量的有标签数据和无标签数据下的比较　　.

4.3 五折交叉验证的总体结果

表3展示了所提出的 SSI-GAN 模型在所有感染时间点（感染后第 0、1、2、3 和 7 天）上使用 3% 带标签数据时的总体分类性能。结果以均值和标准差表示，数据来自五次采用不同随机初始化的蒙特卡洛交叉验证。

报告标准差以体现统计稳健性，无论随机选择种子，还是分配特定的标注样本。所有配置下的标准差都较低（通常低于 0.05%），这证明模型在有标签监督较少的情况下，依然能够稳定地实现较高的分类准确率。这对于神经科学应用中的实际部署至关重要，因为人工进行尖峰标注需要投入大量资源。

表3。蒙特卡洛运行 5 次的总体性能（%）及标准差。

dpi	准确率	精确率	召回率	***F1*****分数**
0	99.06 ± 0.46	99.08 ± 0.44	99.06 ± 0.46	99.06 ± 0.46
1	97.49 ± 03.89	97.60 ± 03.68	97.51 ± 03.86	97.50 ± 03.87
2	97.86 ± 0.93	97.95 ± 0.85	97.86 ± 0.95	97.86 ± 0.94
3	99.20 ± 0.76	99.20 ± 0.76	99.20 ± 0.77	99.20 ± 0.76
7	99.84 ± 0.11	99.84 ± 0.11	99.84 ± 0.11	99.84 ± 0.11
总体性能	**98.69**	**98.73**	**98.69**	**98.69**

5. 消融研究

为了解所提出的 SSI-GAN 中各架构组件的作用，我们通过系统地调整生成器和判别器架构，设置了五种替代基线，开展了消融研究。这些配置在不同感染后天数（dpi）下进行了评估，并使用标准分类指标衡量其性能：准确率（AC）、精确率（PR）、召回率（RE）和 F1 分数（FS）。完整结果见表4。

首个关键观察结果是，基线1同时采用基于 Transformer 的生成器和 Swin Transformer 判别器，表现非常糟糕（例如，在所有感染后天数下，准确率均为 32–34%）。这表明，采用 Swin Transformer 判别器会导致过拟合。由于 Swin Transformer 的整体架构对于我们的案例研究而言过于复杂，判别器因此导致了过拟合。需要注意的是，我们仅将 Swin Transformer 架构用作

判别器，而非 Liu 等人提出的预训练 Swin Transformer。[50] 因为预训练模型是在与我们的案例研究无关的图像数据上训练的。

基线 2 将 CNN 生成器与普通 Transformer 判别器相结合，性能明显更好，在 0 dpi 时准确率超过 97%，在其他水平下甚至更高。这表明，判别器中的自注意力对分类的影响比生成器中的自注意力更大。

基线 3 的生成器和判别器均采用标准 Transformer模型，性能相近，但略逊于所提出的基线。这表明，在生成器中加入注意力机制并不一定能提升性能，还可能引入不必要的复杂性。

值得注意的是，基线 4 采用 CNN 生成器和我们提出的受 Swin 启发的判别器，其表现优于此前的基线，尤其是在中高 dpi 水平下（例如，1 dpi 时为 99.28%，3 dpi 时为 99.18%）。这表明，受 Swin 启发的架构显著增强了模型的鲁棒性，尤其是在使用真实世界噪声数据时。

基线 5 是一个完全基于 CNN 的生成对抗网络，表现逊于所有其他设置，仅优于基线1；在 3 dpi 和 7 dpi 时的得分尤其低。这证实，单靠卷积模型不足以完成此任务，这可能是因为它们无法聚焦于输入中的稀疏尖峰密集区域。

相比之下，所提出的 SSI-GAN 将基于 Transformer 的生成器与采用移位窗口机制的受 Swin 启发的判别器相结合（不进行分层下采样），在所有 DPI 上始终优于所有基线。该模型的分类性能近乎完美，各个感染后天数的准确率均超过 99%，在感染后 7 天达到 99.93%。消融实验要点：

- 判别器架构对性能的影响大于生成器。
- 我们采用移位窗口（非分层结构）的受 Swin 启发的判别器，能够有效学习局部特征和跨窗口特征，尤其适用于稀疏尖峰数据。
- 在生成器中加入注意力并非必不可少，但将其与我们受 Swin 启发的判别器结合，可获得最稳定、最准确的结果。
- 完整的 SSI-GAN 模型优于所有消融变体，证实最终架构中的两个组件均有助于在神经尖峰数据上实现最先进的分类性能。

表4。基于我们提出的算法进行不同核心设计的消融实验。基线1：生成器：Transformer，判别器：Swin Transformer。基线2：生成器：卷积神经网络，判别器：Transformer。基线 3：生成器：Transformer，判别器：Transformer。基线 4：生成器：卷积神经网络，判别器：受 Swin 启发。基线 5：生成器：卷积神经网络，判别器：卷积神经网络。提出的 SSI-GAN：生成器：Transformer，判别器：受 Swin 启发。

Dpi	*Evaluation Metric (%)*	*Baseline 1*	*Baseline 2*	*Baseline 3*	*Baseline 4*	*Baseline 5*	*SSI-GAN*
感染后第***0***天	***AC***	32.96	97.15	98.45	93.97	96.21	**99.45**
	PR	10.98	97.36	98.47	94.11	96.42	**99.46**
	RE	33.33	97.12	98.46	93.97	96.21	**99.46**
	FS	16.52	97.16	98.46	94.00	96.18	**99.45**
感染后第***1***天	***AC***	33.76	97.82	97.32	99.28	95.38	**99.61**
	PR	11.25	97.88	97.38	99.28	95.84	**99.62**
	RE	33.33	97.83	97.32	99.28	95.41	**99.61**
	FS	16.82	97.83	97.32	99.28	95.35	**99.62**
感染后第***2***天	***AC***	34.04	87.93	90.36	98.48	96.87	**97.09**
	PR	11.34	89.35	90.38	98.48	97.03	**97.34**
	RE	33.33	87.88	90.30	98.48	96.86	**97.07**
	FS	16.93	87.85	90.32	98.48	96.85	**97.10**
感染后第***3***天	***AC***	32.10	98.71	98.55	99.18	98.53	**99.82**
	PR	14.24	98.71	98.55	99.20	98.54	**99.82**
	RE	32.54	98.68	98.55	99.19	98.53	**99.82**
	FS	17.31	98.69	98.55	99.19	98.53	**99.82**
感染后第***7***天	***AC***	33.65	99.12	98.09	82.75	95.56	**99.93**
	PR	11.21	99.13	98.15	83.73	96.73	**99.93**
	RE	33.33	99.11	98.06	82.69	95.56	**99.93**
	FS	16.78	99.12	98.07	82.36	95.46	**99.93**

6. 讨论

本研究首次将半监督学习应用于作为病毒感染指标的昆虫神经尖峰分类，从而为计算神经科学、媒介生物学和机器学习交叉领域开辟了新的研究方向。其他研究中的全监督学习框架依赖完整标注的训练数据。因此，本研究针对现有神经元尖峰分类研究中发现的一项关键局限展开。对现有工作的全面回顾表明，先前所有关于尖峰分类的深度学习工作均采用全监督范式，而这需要对尖峰波形进行详尽的人工标注。新提出的 SSI-GAN 仅使用 3% 的带标签数据，便达到了最先进的性能（99.93% 的峰值准确率）。

与传统的全监督方法相比，我们的半监督框架将人工标注需求减少了 97–99%。神经科学专家进行的人工尖峰排序是电生理学研究中的一大障碍，因为每次记录会话都需要耗费数小时甚至数天进行目视检查、波形分组和分类验证。这种方法的成本日益增加，且无法实现规模化。

除了数据效率之外，Transformer架构结合移位窗口方法，还具备当前最先进方法所不具备的时间模式识别能力。事实上，正是自注意力机制能够对整个尖峰序列（100个时间单位）中的长程依赖进行建模，才解释了性能的大幅提升：准确率比 CNN模型 高出 12.27-24.40%，较 LSTM网络 提高 65.68-67.98%。

这对于稀疏且经过大量预处理的尖峰数据尤为重要，因为其中大多数值都被设为零，而相关的判别特征在时间上可能相距甚远。其他优势还包括采用多头注意力设计，从而能够并行处理多个时间关系，并并行捕捉区分 ZIKV、DENV 和对照组神经元活动的互补特征。在五个时间阶段（0、1、2、3、7 感染后天数）上的模型泛化能力保持了较高准确率，范围为 97.09%-99.93%。这种时间鲁棒性与基线模型形成鲜明对比，后者在感染过渡阶段表现出陡峭的性能下降：基于 CNN 的 GAN 在感染后 1 天的准确率下降了 2.83%，而传统模型则下降了 10-15%。

使用不同随机种子的蒙特卡洛交叉验证，是进行检查并确保可复现性的简便方法。在机器学习领域，大多数人并不会讨论这一点。所有指标的标准差都较低，这表明报告的性能提升反映的是强大的学习能力，而不是有利的数据划分或特定标注样本选择所造成的假象。

表5展示了与其他研究进行比较的性能分析。针对这一特定数据集（寨卡病毒/登革病毒）的机器学习研究较为匮乏，反映出在媒介传播疾病背景下，昆虫神经元电生理学的计算方法仍处于发展阶段。

表5。神经元数据上尖峰分类方法的性能比较。

研究	方法	学习范式	有标签数据必需	*Peak* 准确率（*%)*	数据集
Gaburro 等, 2018 [16]	NeuroSigX（手动分类工具）	手动	100%	100	寨卡病毒/登革病毒 埃及伊蚊 埃及伊蚊 神经元
Sharifrazi 等人, 2024 [18]	卷积神经网络 + XGBoost	完全监督式	100%	97.99	寨卡病毒/登革病毒 伊蚊 埃及伊蚊 神经元（60-通道多电极阵列）
Meyer 等，2019 [29]	MLP	完全监督式	100%	89-95	哺乳动物皮层尖峰
Liu 等人, 2020 [30]	卷积神经网络-长短期记忆网络	完全监督式	100%	91-93	基准尖峰数据集
Park 等人等，2020 [33]	FCN	完全监督式	100%	87-90	多电极阵列
拟议研究 2026	**SSI-GAN**	**半监督式**	**1-3%**	**98.69**	**寨卡病毒/登革病毒 伊蚊 埃及伊蚊 神经元**

如表5所示，所有以往的深度学习方法均采用完全监督式学习，因此在人工标注方面成本非常高，不适用于高通量媒介传播疾病研究；而所提出的半监督框架仅使用1-3%的有标签数据便达到了当前最先进的性能，这表明其不仅是准确率的渐进式提升，更代表了方法论上的根本进步，并且能够实际部署于资源有限的神经科学应用中。通过比较相同的数据，这种直接比较可以消除由数据集特征、记录质量或生物变异差异所引起的混杂变量，从而为所提出方法的优越性提供最有力的证据。与使用不同数据集、规模更大的尖峰分类研究相比，所提出的方法仍具有显著优势。其他完全监督式深度学习方法，例如Meyer的多层感知机网络（准确率为89-95%）和Liu的CNN-LSTM混合模型（准确率为91%-93%），在尖峰分类方面同样表现良好。

本研究 这代表着一项重大的方法学进展，但也存在一些局限性 s:

- 来源与环境控制：该数据集仅采集自体外培养的埃及伊蚊原代神经元，培养环境为受控的实验室条件（26 °C，湿度为 60–70 percent），因此可能无法充分反映野生蚊群体在不同生态与环境压力因素下呈现的生物多样性。

- 物种与病毒范围：本分析仅针对一种蚊子（埃及伊蚊）和两种黄病毒毒株（寨卡病毒和登革病毒2型），因此研究结果可能不适用于其他伊蚊属物种或其他嗜神经性黄病毒（例如西尼罗病毒、日本脑炎病毒或黄热病毒）。

- 计算需求：基于Transformer的模型能力出色，但与简单的基线架构相比，其训练需要更多的计算资源，这可能会造成实际应用上的限制。

- 静态评估设置：该模型仅在静态数据集上进行了评估，未在实时部署场景中进行测试；而在这些场景中，延迟、流式数据和连续推理至关重要。

- 可解释性挑战：尽管注意力机制有助于为特征赋予权重，但驱动分类决策的特征可能并不容易解释，这会使可解释性分析和生物学理解更加困难。

7. 结论

本研究首次将受Swin启发的半监督生成对抗网络（SSI-GAN）应用于感染病毒的蚊子的神经元尖峰分类。所提出的模型仅使用3%的带标签训练数据，在感染后3天达到了99.93%的峰值分类准确率。该方法优于本研究评估的其他生成对抗网络基线模型及所有传统深度学习架构。采用移位窗口方法的Transformer判别器所具备的自注意力机制，也展现出优于常见的基于注意力、卷积和循环方法的性能；在各时间阶段，其准确率较纯卷积神经网络模型提升了12.27%至24.40%，较LSTM网络提升了65.68%至67.98%。

该半监督学习框架通过将人工标注需求减少约 97–99%，同时保持与全监督方法相当或更高的准确率，解决了神经科学研究中的一个关键局限。即使仅有极少量监督（1% 的带标签数据），仍能在各感染时间点达到 95.04–97.83% 的准确率，这表明该框架适用于高通量电生理筛查。通过系统评估感染后第 0、1、2、3 和 7 天的时间性能模式，揭示了寨卡病毒和登革病毒嗜神经性的生物学进程。分类准确率峰值与感染后第 3 天（3 dpi）病毒诱导的神经生理变化达到最大值相吻合。

这项工作提供了一个定量框架，用于理解黄病毒如何改变蚊神经动力学，并对操控媒介行为、传播生态学和干预措施开发具有启示意义。

未来的工作可将该方法扩展至更多蚊媒物种和全面的黄病毒面板，以验证其普适性。通过主动学习或自监督预训练，进一步探索更低的有标签数据比例（例如零样本机制），或可进一步降低人工标注的要求。该方法实现了完全无监督的尖峰分类，并推动了可部署于活体神经元记录的实时分类系统的开发；这些系统可用于闭环实验，根据对感染状态的自动检测动态调整实验参数。与同步行为追踪相结合，可将神经元活动模式与蚊子运动、吸血倾向和寻主行为联系起来。最后，可解释人工智能方法（如注意力权重可视化和基于梯度的显著性映射）有助于识别对分类决策影响最大的时间位置和尖峰特征。

致谢
我们衷心感谢 Julie Gaburro 博士为收集和整理本研究所用数据集所付出的宝贵努力与奉献。她的贡献对本研究的成功至关重要。

数据和材料的可用性
本研究所用数据集可在联系通讯作者后获取，但须经澳大利亚疾病防范中心（澳大利亚联邦科学与工业研究组织，CSIRO）的伦理审批。该中心是澳大利亚政府的法定机构，拥有该数据的合法权利。所提出方法的代码可通过提供的链接在 GitHub 平台上公开获取。

利益冲突
作者　　声明本文的发表不存在利益冲突　　r.

资助声明
本文由迪肯大学资助，以供开放获取发表。

伦理声明
本研究不涉及人类或动物，因此动物伦理和参与同意声明不适用。

附录A

使用伪代码展示所提出的 SSI-GAN 架构用于检测登革热、寨卡和对照组类别的完整实施流程。

所提算法： 半监督Swin启发式生成对抗网络（SSI-GAN）
输入： 原始多电极阵列记录信号
输出： 使用 3% 已标注数据检测出的类别（对照组、登革热和寨卡）

开始

1. 尖峰数据预处理

1.1 读取 3 个类别的原始信号：
- CT（标签 = 0）、DN（标签 = 1）、ZK（标签 = 2）
- 来源：读取_each_类别_所有（感染后天数）

1.2 将原始信号分割为多个片段：
- 每个分段长度为 = 100 个采样点
- 使用滑动分段（不重叠）

1.3 拼接并分配类别标签

1.4 应用巴特沃斯高通滤波器：
- 截止频率 = 700 Hz，采样率 = 30 kHz
- Removes low-fr频率 noise

1.5 通过阈值提取尖峰：
- 将（−10，10）μV 范围内的值置零
- 保留高振幅事件作为尖峰

输出：尖峰_数据 ∈ R^{N × 100}，标签

2. 生成器 G(z)

输入：潜向量 z ∈ R128
- 投影并重塑为二维特征图
- 应用 2 个 Transformer块（在空间布局上进行注意力计算）
- 使用 Conv2DTranspose 层进行上采样
- 输出：合成尖峰信号 ∈ R^{100 × 60}

3. 受 Swin 启发的判别器 D(x)

输入：序列 x ∈ R^{100 × 60}
- 将项目输入映射到嵌入维度
- 应用堆叠的受 Swin 启发的 Transformer1D 模块（窗口注意力）
- 全局平均池化 → 全连接层 → Softmax（3 类）
- 对于无监督式任务：对 Softmax logits 应用 sigmoid（真/假）

4. GAN 架构
- G(z) → D(x)
- 用于监督式和无监督分支的判别器
- 无监督分支通过自定义 sigmoid 输出真/假概率

5. 训练流程

对于 t = 1 到 T：
- 抽取 3% 的带标签数据进行监督训练
- 使用真实（有标签数据）和伪造（生成）数据训练 D
- 训练 G 以最大化 D 的误分类（通过生成对抗网络损失）
- 每 N 次迭代记录损失/准确率

Use:
- 标签平滑（真实值 = 0.9）
- 丢弃 = 0.29

End

参考文献

1. Thomas, S.J., L.J. Martinez, 和 T.P. Endy，《黄病毒：黄热病、日本脑炎B型、西尼罗病毒及其他病毒》，载于《人类病毒感染：流行病学与防控》。2023，施普林格。第1–62页。 2. Dick, G.W.，乌干达分离出的一些病毒的流行病学札记（黄热病、裂谷热、邦巴热、西尼罗病毒、门戈病毒、塞姆利基森林病毒、班尤姆韦拉病毒、恩塔亚病毒、乌干达S病毒和寨卡病毒）。《热带医学与卫生学皇家学会会刊》，1953。47(1)：第13–48页。 3. Chang, C., 等，《21世纪的寨卡疫情》。《自身免疫杂志》，2016。68：第1–13页。 4. Khan, M., 等，《COVID-19：一项历史悠久的全球挑战——流行病学及迄今进展》。《分子》，2020。26(1)：第39页。 5. Muñoz, L.S., P. Barreras, 和 C.A. Pardo.《成人寨卡病毒相关神经系统疾病：吉兰–巴雷综合征、脑炎和脊髓炎》。载于《生殖医学研讨会》。2016。蒂梅医学出版社。 6. Thangamani, S., 等，寨卡病毒在埃及伊蚊中的垂直传播。《美国热带医学与卫生学杂志》，2016。95(5)：第1169页。 7. Villalobos-Sánchez, E., 等，《寨卡病毒在人类胎盘中垂直传播的可能途径：全面综述》。《病毒免疫学》，2022。35(6)：第392–403页。 8. Yu, X., 等，《深入了解蚊媒病毒生命周期的进展》。《寄生虫学趋势》，2019。35(12)：第1009–1017页。 9. Osorio, D., J.P. Bacon, 和 P.M. Whitington，节肢动物神经系统的演化：现代昆虫和甲壳类动物在五亿多年前从共同祖先分化而来，但其神经回路仍保留许多共同特征。《美国科学家》，1997。85(3)：第244–253页。 10. Lovinger, D.M.，大脑中的通信网络：神经元、受体、神经递质和酒精。《酒精研究与健康》，2008。31(3)：第196页。 11. Teleanu, R.I., 等，神经递质——中枢神经系统神经系统疾病和神经退行性疾病的关键因素。《国际分子科学杂志》，2022。23(11)：第5954页。 12. Miguel-Tomé, S. 和 R.R. Llinás，拓展神经系统的定义，以更好地理解动植物的演化。《植物信号与行为》，2021。16(10)：第1927562页。 13. Ghysen, A.，神经系统的起源与演化。《国际发育生物学杂志》，2003。47(7-8)：第555–562页。 14. Wu, S., 等，寨卡病毒、西尼罗病毒和登革病毒对中枢神经系统免疫反应的调节。《医学病毒学评论》，2024。34(3)：第e2535页。 15. Harschnitz, O. 和 L. Studer，利用人类干细胞模型研究中枢神经系统中的宿主–病毒相互作用。《自然评论：免疫学》，2021。21(7)：第441–453页。 16. Gaburro, J., 等，与媒介埃及伊蚊中寨卡病毒感染相关的嗜神经性和行为变化。《新发微生物与感染》，2018。7(1)：第1–11页。 17. Sharifrazi, D., 等，利用人工智能技术对脉冲信号数据进行功能分类：综述。arXiv预印本 arXiv:2409.17516，2024。 18. Sharifrazi, D., 等，利用神经脉冲分析通过机器学习方法自动检测埃及伊蚊中的寨卡病毒和登革病毒。《生物医学信号处理与控制》，2024。96：第106594页。 19. Flynn, M，利用机器学习对危险蚊种进行分类。2022，东英吉利大学。 20. Morais, C., 等，识别人为错误及其影响因素：一种机器学习方法。《安全科学》，2022。146：第105528页。 21. Koutsoukas, A., 等，深度学习：研究深度神经网络超参数，并比较其在生物活性数据建模中的性能与浅层方法。《化学信息学杂志》，2017。9：第1–13页。 22. Wilson, S.B. 和 R. Emerson，尖峰检测：算法综述与比较。《临床神经生理学》，2002。113(12)：第1873–1881页。 23. Lewicki, M.S.，神经信号的贝叶斯建模与分类。《神经计算》，1994。6(5)：第1005–1030页。 24. Letelier, J.C. 和 P.P. Weber，基于离散小波变换系数的尖峰排序。《神经科学方法杂志》，2000。101(2)：第93–106页。 25. Haas, A.M., M.H. Cohen, 和 P.A. Abshire.基于实时方差的模板匹配尖峰排序系统。载于2007年IEEE/NIH生命科学系统与应用研讨会。2007。IEEE。

26. Takekawa, T., Y. Isomura, 和 *T. Fukai*，通过多模态加权*PCA*和显式稳健变分贝叶斯对异质神经元类型进行尖峰排序。《*Frontiers in Neuroinformatics*》，*2012*。*6*：第*5*页。 *27. Coleman, C.,* 等，*DawnBench*分析：机器学习性能基准中的达到准确率所需时间。《*ACM SIGOPS Operating Systems Review*》，*2019*。*53(1)*：第*14–25*页。 *28. Yu, S., J. Ma,* 和 *W. Wang*，深度学习去噪。《*Geophysics*》，*2019*。*84(6)*：第*V333–V350*页。 *29. Meyer, L.M., T. Schanze,* 和 *F. Samann*，用于尖峰波形分类的单隐层神经网络。《*Automation in Medical Engineering*论文集》，*2023*。*2(1)*：第*747–747*页。 *30. Liu, M.,* 等，利用卷积神经网络和长短期记忆对重叠尖峰进行分类。《*Computers in Biology and Medicine*》，*2022*。*148*：第*105888*页。 *31. Li, S.,* 等，将深度强化学习应用于多类别不平衡条件下的尖峰排序。《*Computers in Biology and Medicine*》，*2023*。*164*：第*107253*页。 *32. Saif-ur-Rehman, M.,* 等，*SpikeDeep-Classifier*：一种基于深度学习的全自动离线尖峰排序算法。《*Journal of Neural Engineering*》，*2021*。*18(1)*：第*016009*页。 *33. Park, I.Y.,* 等，用于细胞外记录的深度学习模板匹配尖峰分类。《*Applied Sciences*》，*2020*。*10(1)*：第*301*页。 *34. García, P.,* 等，利用混合多层人工神经网络对神经尖峰进行无监督分类。《*Journal of Neuroscience Methods*》，*1998*。*82*：第*59–73*页。 *35. Bernert, M.* 和 *B. Yvert*，一种用于无监督尖峰排序的基于注意力的脉冲神经网络。《*International Journal of Neural Systems*》，*2019*。*29(08)*：第*1850059*页。 *36. Dipietro, L.,* 等，大数据在神经科学和神经病学中的演进。《*Journal of Big Data*》，*2023*。*10(1)*：第*116*页。 *37. Learning, S.-S.*，半监督学习。*CSZ2006.html*，*2006*。*5(2)*：第*1*页。 *38. Hady, M.F.A.* 和 *F. Schwenker*，半监督学习。《神经信息处理手册》，*2013*：第*215–239*页。 *39. Jafarlou, M.* 和 *M.M. Kubek.*通过半监督学习降低情感分析中的标注成本。载于《*2024*年第*8*届自然语言处理与信息检索国际会议论文集》。*2024*。 *40. Zhang, Y.,* 等，推动半监督学习在自动语音识别中的极限。*arXiv*预印本 *arXiv:2010.10504*，*2020*。 *41. Mourad, R*，利用伪标签进行调控序列预测的半监督学习。*bioRxiv*，*2023*：第*2023.12. 21.572780*页。 *42. Hales, C.M., J.D. Rolston,* 和 *S.M. Potter*，如何在微电极阵列（*MEAs*）上培养、记录和刺激神经元网络。*JoVE*（可视化实验杂志），*2010(39)*：第*e2056*页。 *43. Chen, Y.,* 等，优化神经记录前端，以提升高通道数系统中的尖峰排序准确率。《*IEEE Transactions on Neural Systems and Rehabilitation Engineering*》，*2025*。 *44. Liu, C.-W.,* 等，小脑以数值精度和跨个体一致性计算运动的频率动力学。*Research Square*，*2024*：第*rs. 3. rs-4615547*页。 *45. Li, R.,* 等，用于脑电监测和*SSVEP*响应检测的高采样率多通道无线记录仪的设计与实现。《*Frontiers in Neuroscience*》，*2023*。*17*：第*1193950*页。 *46. Akiba, T.,* 等，*Optuna*：新一代超参数优化框架。载于第*25*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集。*2019*。 *47. Sardari, S.,* 等，使用卷积自编码器进行基于音频的抑郁症检测。《*Expert Systems with Applications*》，*2022*。*189*：第*116076*页。 *48. Shan, G*，针对二元结局和有限样本量研究的蒙特卡洛交叉验证。《*BMC Medical Informatics and Decision Making*》，*2022*。*22(1)*：第*270*页。 *49. Xu, Q.-S.* 和 *Y.-Z. Liang*，蒙特卡洛交叉验证。《 *Chemometrics and Intelligent Laboratory Systems* 》，*2001*。*56(1)*：第*1–11*页。 *50. Liu, Z.,* 等，*Swin Transformer*：采用移位窗口的分层视觉变换器。载于*IEEE/CVF*国际计算机视觉会议论文集。*2021*。