随机蜂群：集成深度学习器在太阳高能粒子预测问题中的应用

Patrick M. O'Keefe[a]、Viacheslav Sadykov[b]、Alexander Kosovichev[c, d]、Irina N. Kitiashvili[d]、Vincent Oria[a]、Gelu M. Nita[c]、Fraila Francis[a]、Chun-Jie Chong[a]、Paul Kosovich[c]、Aatiya Ali[b]、Russell D. Marroquin[e, b]

[a]新泽西理工学院计算机科学系，*323 Dr Martin Luther King Jr Blvd*，纽瓦克 *07102*，新泽西州，美国[b]佐治亚州立大学物理和天文学系，*25 Park Place NE*，亚特兰大 *30303*，佐治亚州，美国[c]新泽西理工学院物理系，*323 Dr Martin Luther King Jr Blvd*，纽瓦克 *07102*，新泽西州，美国[d]美国国家航空航天局艾姆斯研究中心，*NASA*研究园区，*Moffett Field 94035*，加利福尼亚州，美国[e]加州大学圣迭戈分校物理系，*9500 Gilman Dr*，拉霍亚 *92093*，加利福尼亚州，美国

arXiv:2303.08092v2 [astro-ph.SR] 7 May 2024

摘要

近年来，机器学习和深度学习技术已广泛应用于各种现象的预测，其中包括大量使用非集成的传统神经网络（CoNN）。这些技术效率高，且能够在无需人工干预的情况下从数据中发现关系，因此广受欢迎。然而，某些 CoNN 配置可能不适用于某些数据集，尤其是传入的参数（包括模型参数和超参数）可能带有较强的任意性，需要不断更新并重新训练模型；特别是新增特征可能使旧模型过时。采用神经网络委员会可以在一定程度上缓解这一问题：委员会成员的输入特征相同，且以随机方式初始化，并共同"投票"作出决策。然而，委员会成员的架构相似，接收的特征也相同，因此可能会对所有节点和边采用完全相同的权重和偏置。委员会也无法扩充以纳入新特征，因此必须重新训练整个委员会。我们提出随机蜂群（RH）方法，以缓解这一问题：该方法使用多个神经网络估计器，根据特征的随机排列作出决策，并规定了一种确定各个估计器决策权重的方法。我们通过实验展示了 RH 的有效性：在预测具有危害性的太阳高能粒子（SEP）事件时，将其与 CoNN 以及上述输入特征相同的委员会配置进行比较。结果表明，RH 的表现可与 CoNN 和基于委员会的方法相当或更好；此外，各项实验的得分分布范围更小，并且在捕捉几乎所有会引发太阳高能粒子事件的耀斑实例方面展现出良好前景。

关键词： 太阳：活动，太阳：粒子发射，日地关系，计算方法：提升，机器学习，神经网络

目录

1 引言 2 数据准备 3 机器学习方法 **2**

4 结果与讨论 5 结论 **3**

4

8

12

1. 引言

太阳高能粒子（SEP）事件的预测及其前兆的理解，无论从业务应用还是研究角度来看，都是日球物理学和空间天气领域面临的重大挑战。太阳高能粒子事件通量的增加受到各类用户的关注，用户涵盖政府和私人空间天气机构、航空公司以及电网运营商。美国国家海洋和大气管理局（NOAA）下属的空间天气预报中心（SWPC）已建立针对太阳质子事件（SPE）的常规每日预报、短期预警和警报系统（Balch, 1999，2008）。业务预报系统的表现仍远不足以预测每一次太阳高能粒子事件（Bain 等，2021年）。

太阳高能粒子事件由太阳耀斑和日冕物质抛射（CMEs，Reames，2021年）引发。长期以来，人们已知耀斑软X射线特性（例如，在单温近似下计算的耀斑温度与1-8Å和0.5-4Å通量峰值之比成正比，Ryan等人，2012年；Sadykov等人，2019年）与随后发生的日冕物质抛射和太阳高能粒子事件之间存在统计关系。太阳耀斑根据美国国家海洋和大气管理局地球静止环境业务卫星（GOES）网络在1-8Å波长范围内观测到的软X射线（SXR）通量最大值进行分类。在此分类中，A级耀斑的软X射线通量最大值大于 10^{-8} W/m^2；B、C、M和X级耀斑的SXR通量则分别大于 10^{-7}、 10^{-6}、 10^{-5}和 10^{-4}W/m^2 。具体而言，研究发现，软X射线类别越低，与太阳高能粒子事件相关的耀斑和非SEP耀斑之间的峰值温度差异越大；与太阳高能粒子事件相关的耀斑温度较低（Garcia, 1994年）。利用更多耀斑事件对这些关系进行了量化，并将其用于太阳高能粒子预测（Garcia，2004年）。后来，相关结果也得到复现（Kahler和Ling，2018年）；作者尝试分别针对太阳西半球和东半球，采用k近邻机器学习算法和神经网络预测与太阳高能粒子事件相关的耀斑。尽管观测到的关系很明确，但低温耀斑为何更易与太阳高能粒子事件相关，其确切原因仍大多未知（Kahler和Ling，2018年）。此外，研究发现，耀斑持续时间和温度在统计上与日冕物质抛射的特性相关（Ling和Kahler，2020年；Kahler和Ling，2022年）。因此，耀斑持续时间和温度可用于约束SEP参数（Kahler和Vourlidas，2013年），也可作为建立太阳高能粒子预测经验模型的基础（Richardson等人，2018年）。太阳质子事件实时警报经验模型（ESPERTA）预报工具（Laurenza等人，2009年、2018年）还利用积分SXR强度和1 MHz处的积分无线电强度，对质子能量高于10兆电子伏特和100兆电子伏特的太阳质子事件进行短期预测。

这些研究的延伸是采用机器学习（ML）和深度学习技术，尤其是根据前序（母体）太阳耀斑的特性来预报太阳高能粒子事件。例如，Aminalragia-Giamini 等人（2021）使用在太阳耀斑期间直接观测到的软X射线通量时间序列上训练的神经网络。作者发现，在 1988—2013 年期间，该模型能够预测绝大多数与太阳高能粒子事件相关的耀斑（超过 85%），同时保持较低的假阳性率。Boubrahimi 等（2017）分析了地球静止环境业务卫星的软X射线和质子通量时间序列之间的相关性，并采用分类决策树模型来预报 100 MeV太阳质子事件。Lavasa 等人（2021）分析了多种机器学习算法，例如随机森林、神经网络、极端随机树和极端梯度提升。他们得出结论，在软X射线参数中，注量对于预报太阳高能粒子事件最为重要。

在文献中，使用传统神经网络（下文简称 CoNN；本文将其定义为任意架构的单个神经网络，即非集成神经网络）进行空间天气预报十分常见（例如 Torres 等，2022年；Nishizuka 等，2020年等）。CoNN 在如何利用新数据进行训练方面具有很强的灵活性和可塑性，但对于模型大小和形状、超参数及模型选择等方面所提供的参数，可能需要调整后才能用于其他数据集。因此，随着数据逐渐过时，或需要添加新的特征，就必须不断重新训练模型。理想情况下，还应根据输入模型的特征调整模型的训练轮数和学习率。尤其是，较少的训练轮数和较高的学习率有助于避免模型对噪声数据过拟合（Afaq 和 Rao，2020年；You 等，2019年），但同时也需要确保模型经过充分训练，以便从总体上学习数据中的模式。这种矛盾可能导致为整个模型设定的训练轮数与学习率组合并非最优，因为同一组合未必适用于所有特征，尤其是当某些特征比其他特征更能决定给定数据集的真实标签时。可以通过调整这些参数来折中，但这种折中也可能再次导致模型在某些参数上欠拟合，

而在其他参数上过拟合。通常，防止过拟合的一种策略是采用集成学习（Cunningham，2000年）。

与SEP预测相关的另一项挑战是，这类事件十分罕见，在分类问题中属于少数类事件。例如，在第22—24太阳活动周中，增强质子通量能量大于 10 兆电子伏特且粒子通量大于 10 质子通量单位（pfu）（一个粒子通量单位 pfu 等于每平方厘米每秒每球面度一个粒子）的天数，与没有增强通量的天数之比为 ~ 1/23（Ali 等人，2023）。仅考察第24太阳活动周时，这一比值甚至更低（~ 1/34）（Sadykov等，2021）。针对 ESPERTA模型（Stumpo 等人，2021）的研究得出结论：算法的性能（具体而言，虚警率 FAR）取决于训练数据集中的类别不平衡比。可采用多种技术来处理类别不平衡数据，例如过采样、欠采样和误分类权重（Ahmadzadeh 等人，2021），以及使用合成数据（Chen 等人，2021）。除传统的以数据为中心的类别不平衡处理方法外，还可在此类问题中采用集成分类器（Galar 等人，2012）。对于太阳高能粒子事件的预测问题，先前采用基于神经网络的委员会集成（Aminalragia-Giamini等人，2021）和随机森林集成算法（Lavasa 等人，2021）已取得了令人鼓舞的结果。

在我们此前的工作中，我们介绍了随机森林机器学习算法在预测太阳高能粒子事件方面的应用，并测试了各种类别不平衡处理技术（O'Keefe等，2022）。在本研究中，我们将考察范围扩展至新型机器学习算法，包括常规神经网络（CoNN）、采用投票方法的CoNN集成（委员会，Aminalragia-Giamini等人，2021），并提出一种称为随机蜂群（RH）的加权共识方法。当各集成成员彼此独立，并以确定性方式参与分类决策时，我们所研究的这两种集成方法均属于所谓的"装袋法"集成分类器。本文的重点是考察这些算法在给定的、与太阳高能粒子事件相关的耀斑数据集上的相对表现。本文结构如下：第2节介绍本文采用的数据准备方法，包括软X射线数据的处理、耀斑与太阳高能粒子事件的关联，以及为机器学习（ML）分析准备数据集。第3节介绍本研究测试的机器学习算法。第4节呈现结果并进行讨论，第5节则给出结论。

2. 数据准备

地球静止环境业务卫星在0.5-4 Å和1-8 Å波长通道观测到的太阳软X射线（SXR）辐射，可在单温度等离子体近似下用两个参数表示，即等离子体温度T及其发射量EM。我们采用2002—2017年间的历史数据集，其中T 和EM 值是使用基于温度和发射量的背景扣除算法（TEBBS，Ryan等人，2012；Sadykov等人，2017）估算得出，并收录于太阳耀斑交互式多仪器数据库（IMIDSF[1], Sadykov等人，2017）。除温度和发射量的峰值T_{max} 和EM_{max}外，我们还采用扣除背景后的耀斑级别（SXR_{max}）、耀斑持续时间、T_{max}、EM_{max}和SXR_{max} 峰值相对于耀斑起止时间的时刻，以及宿主耀斑在太阳圆面上观测到的X、Y坐标。总计而言，这一数据处理为每个太阳耀斑提供了12个参数。对于部分耀斑，由于X射线辐射变化复杂且背景噪声较强，该算法未能奏效（产生了不合理的$T_{max} \geq 100$兆开尔文值和负时间差），因此将这些耀斑从数据集中剔除。纳入分析的耀斑总数为18311个。其中包括：按SXR分类，5个A级、6919个B级、10074个C级、1207个M级和106个X级耀斑。这些耀斑级别是在扣除SXR背景后计算得出的。Sadykov等人（2019）的图4c显示，背景扣除主要影响较弱的BC级耀斑的耀斑等级，但能使T曲线和EM曲线的变化更可靠。

为将耀斑与SEP记录相关联，我们使用了美国国家海洋和大气管理局空间环境服务中心[2] 提供的《影响地球环境的太阳质子事件》列表。该数据集代表峰值通量高于10质子通量单位的 >10 兆电子伏特质子的SEP事件。我们列表中的64个耀斑与SEP事件相关联，另有18247个耀斑不含SEP（非SEP耀斑），因此类别不平衡比极为悬殊，为

[1] https://data.nas.nasa.gov/helio/portals/solarflares/
[2] https://umbra.nascom.nasa.gov/SEP/

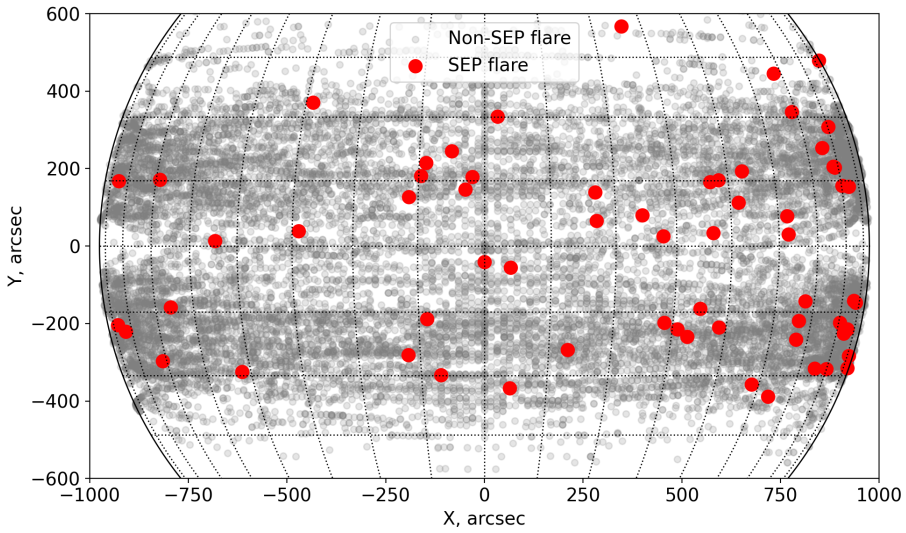

图 1：位置数据集中太阳圆面上太阳高能粒子耀斑与非SEP耀斑的位置。灰点标示未产生太阳高能粒子事件，红点标示产生了太阳高能粒子事件的耀斑。

1/285。就SXR_{max}参数而言，SEP事件中有8个对应C级耀斑、36个对应M级耀斑，20个对应X级耀斑。所研究耀斑的列表可在太阳高能粒子预测门户（SEP[3]）网站[3]公开获取。我们在此指出，其他一些关于SEP预测问题的研究（例如，Papaioannou 等人，2016年；Lavasa 等人，2021年）所考虑的SEP事件数量显著多于当前工作（分别为314个和257个）。这两项研究涵盖了更长的时间跨度（检测于1984年至2013年以及1988年至2013年的事件），因此SEP事件的统计数量更大。此外，在当前工作中，我们未考虑以下SEP事件：（1）未与耀斑事件相关联的事件；（2）根据GOES耀斑目录无法确定宿主耀斑事件坐标的事件。这两项限制导致损失了15条SEP事件记录，使事件总数剩余64个。图1展示了SEP耀斑和非SEP耀斑在太阳圆面上的位置。显然，SEP耀斑更倾向于起源于西半球，该区域与地球具有更直接的磁连接，不过也有一些SEP耀斑起源于靠近东边缘的位置。

图 2 展示了与太阳高能粒子事件相关的耀斑和非SEP太阳耀斑在耀斑温度与软X射线峰值通量、耀斑持续时间与发射量，以及耀斑上升时间与耀斑衰减时间关系图中的分布。这里，耀斑上升时间定义为SXR_{max}减去耀斑开始时间，耀斑衰减时间定义为耀斑结束时间减去SXR_{max}时刻。可以看出，即使在软X射线峰值通量相同的耀斑中，产生太阳高能粒子事件的耀斑也并非随机分布。具体而言，图 2a 表明，在软X射线峰值通量（或耀斑等级）相同的耀斑中，与太阳高能粒子事件相关的耀斑平均温度较低。Garcia（1994，2004）也发现了类似的相关性。图 2b 和图 2c 所示参数未呈现明显规律，但仍需进一步研究，以了解它们与太阳高能粒子事件相关耀斑之间的关系。在机器学习分析中，我们将数据集划分为训练子集和测试子集；后者占完整数据集的 30%。本文所述的每项机器学习实验均随机重复进行十次训练集与测试集的划分。

3. 机器学习方法

本文针对太阳高能粒子事件预测问题，探讨了三种基于神经网络的方法。第一种是传统神经网络（下文简称CoNN），它采用单一的全连接神经网络架构。对于集成深度学习器，还构建了另外两种基于神经网络的方法。第一种

[3] https://sun.njit.edu/SEP3/datasets.html

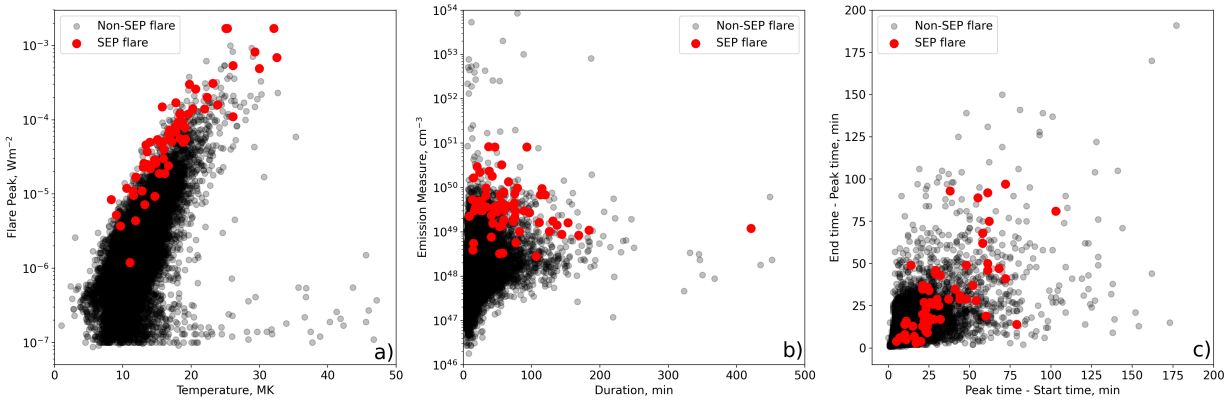

图 2：（a）太阳高能粒子耀斑和非SEP耀斑在 1–8 Å 软X射线通量与耀斑峰值温度图中的分布；（b）耀斑持续时间与峰值发射量图；（c）耀斑上升时间与耀斑衰减时间图。黑点表示未产生太阳高能粒子事件的太阳耀斑，红点表示产生了太阳高能粒子事件的耀斑。

集成方法是 Aminalragia-Giamini 等人（2021）采用的委员会方案，其中包含多个具有相同输入特征和输入层形状的神经网络估计器。第二种集成方法是随机蜂群（RH），通过从训练数据中随机筛选耀斑特征，将其用作输入特征来构建。

考虑开发 RH 方法有几个原因。首先，随机蜂群允许单个估计器分别扩展，以纳入更多特征，而无需重新训练整个集成。只需重新训练部分而非全部估计器，便可能缩短深度学习模型的重新训练时间。其次，随机蜂群不仅利用集成学习通常能够抑制过拟合这一优势（Cunningham，2000），还会根据各估计器所用特征的重要性，调整其训练轮数、学习率和投票权重。这样，随机蜂群就能通过降低不重要特征被纳入或影响预测的可能性，来应对过拟合风险。

对于本文介绍的一系列测试，我们考虑了两种存在若干差异的随机蜂群（RH）实现。首先，两种实现使用的输入特征数量不同。第一种 RH 实现（下文称为 RH.v1）将特征总数的平方根向上取整，作为 RH 中每个神经网络的输入特征数量。第二种实现（下文称为 RH v2）使用一半的特征作为输入。每个委员会内部各估计器的布局保持不变。从某种意义上说，特征的半随机选择（其被选中的概率仍与其权重成正比；见下文）受到了随机森林集成学习算法（Breiman，2001年）的启发。相应地，有 12 个特征输入 CoNN 或每个委员会成员，而 4 个或 6 个特征则通过下文介绍的流程选出，并输入 RH分类器。每种集成配置（即委员会、RH v1 和 RH v2）均包含 10 个神经网络估计器。所用机器学习方法的架构示意见图3。

特征的随机排列通过首先计算特征与 SEP 事件的发生之间的 χ 平方统计量和互信息增益统计量来选择，并根据每个特征在判断给定耀斑是否引发太阳高能粒子事件时的重要程度为其评分。每个特征的得分按以下公式计算：

$$s_i = \sqrt{\chi_i^2} + \kappa_i \quad (1)$$

这里，s_i 是给定特征的总得分，$\chi^2{}_i$ 是给定特征与 SEP事件的发生之间的 χ-平方统计量，κ_i 是给定特征与 SEP事件的发生之间的互信息增益统计量。根据此公式为每个特征评分后，会对这些分数进行归一化，使其总和（“特征权重”）为一。随后将特征权重用作概率，以此按各自的权重选取给定随机蜂群估计器所使用的特征。

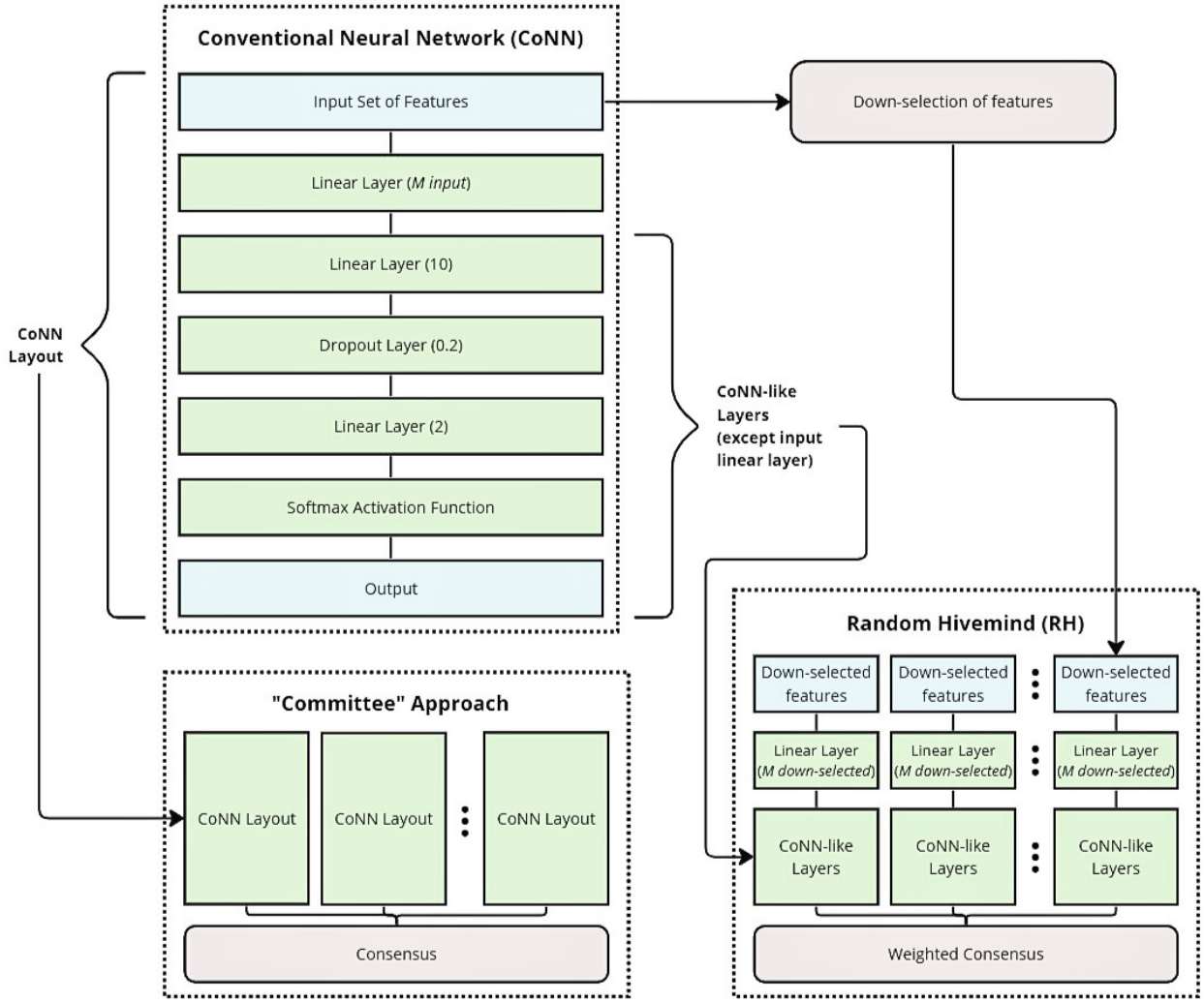

图3：传统神经网络（CoNN）、委员会网络和随机群体网络（RH）的布局示意图。线性层后括号中的数字表示该层的神经元数量，丢弃层后括号中的数字表示训练过程中每个连接/权重被丢弃的概率。

每个神经网络，包括 CoNN、委员会估计器和 RH 估计器，其输入层的大小等于该估计器所测试的特征数量；网络还包含一个输入和输出形状均为 10 的全连接层、一个丢弃概率为 0.2 的丢弃层，以及一个输出形状等于预测特征数量的输出层。这些网络使用 Skorch 库实现（Tietz等人，2017年）。根据后文所述的超参数优化，所有 CoNN 和委员会配置的训练轮数和学习率分别为$n_{\text{epochs}} = 500$ 和 $\alpha = 0.001$。此外，随机蜂群会使用以下公式增加其训练轮数和学习率：

$$n_{\text{epochs}} = 500 \times (2 - x) \tag{2}$$

$$\alpha_{\text{RH v1}} = 0.001 \times \left[1 + \ln\left(\frac{e^x - x}{\eta} + 1\right)\right] \tag{3}$$

$$\alpha_{\text{RH v2}} = 0.001 \times (0.5 + x) \tag{4}$$

其中，x 表示给定估计器的特征权重总和，n^{epochs} 表示训练过程中的训练轮数，α 表示学习率，而e 是自然对数的底数。参数 $\eta = e^{x}\Sigma x\, n_{\text{features}}$中，$n_{\text{features}}$ 表示

为给定估计器选取的特征数量，而 Σx 表示集成中所有估计器所含全部特征的权重总和。RH分类器经过两次测试，第一次中 α 等于 α_1，第二次中等于 α_2。将预测数据输入训练阶段构建的每个估计器，并查看各估计器选出的预测结果，即可得出预测结果。每个委员会将所有估计器给出的结果视为同等重要，并通过简单的相对多数投票确定给定数据点所属的类别。每个随机蜂群都将分类投票中各估计器的值视为相等，该值为该估计器输入特征的特征权重总和。对于所有神经网络，包括 CoNN、委员会估计器和 RH 估计器，均使用 Adam 优化器（Kingma 和 Ba，2015年），采用带平衡类别权重的交叉熵损失函数，并使用包括丢弃概率为 0.2 的 Dropout 层和数据打乱在内的过拟合防止措施。

让我们更详细地看一个特征权重的例子。如果某个耀斑的软 X 射线峰值通量的特征权重为 0.25，发射量峰值的特征权重为 0.1，温度峰值的特征权重为 0.05，持续时间的特征权重为 0.01，且所有特征权重之和为 1，那么峰值软 X 射线通量被选入 RH估计器的概率为 0.25，发射量峰值被选入的概率为 0.1，依此类推。然而，CoNN 和委员会会将所有可用特征一视同仁地作为输入特征。在训练期间，使用这四个参数的 RH估计器将在第一次测试中经历 15 个训练轮数，学习率约为$\alpha_{\mathrm{RH\ v1}} = 0.00162$；在第二次测试中，学习率约为$\alpha_{\mathrm{RH\ v2}} = 0.00186$。然而，在此例中，CoNN 和委员会估计器各自只会经历 10 个训练轮数，学习率为 $\alpha = 0.001$，因为它们无法根据特征选择自动计算这些参数。在作出判断时，每个 RH估计器会将其特征权重之和作为数值，因此，使用这四个参数的估计器在投票时的数值为 0.5。每个委员会估计器的数值则为 1，因为同样没有机制可以根据特征选择来确定如何计算这些数值。

超参数调优对于机器学习的重要性早已得到认可。对于基于神经网络的方法，超参数可能包括与网络架构相关的参数（隐藏层数量、每层的神经元数量、激活函数等），以及与训练过程相关的参数（训练轮数、学习率、优化器和正则化参数等）。若考虑集成方法，参数空间还会进一步扩大。探索整个参数空间的成本非常高。因此，在本研究中，我们将模型架构限制为图3所示的架构，仅对与训练过程相关的参数进行优化。我们还注意到，本研究采用的集成方法均以相同的 CoNN 结构为基础；因此，优化 CoNN 的训练过程也应能为集成方法提供最优训练方案。我们对学习率 $\alpha = \{0.01、0.0025、0.001、0.0005\}$ 以及最多达到$n_{\mathrm{epochs}} = 2000$轮的训练轮数构成的网格进行探索，每组参数组合进行 10 次实验，以找到最优且成本相对不高的 CoNN 训练流程。

我们使用 $\alpha = 0.001$ 优化 CoNN 的结果如图4所示。这里需要指出，将学习率进一步降至 $\alpha = 0.0005$ 并未提高 HSS 和 TSS 评分。可以看出，随着训练轮数增加，CoNN 的性能总体上有所提升。我们决定在预测实验中选用 $n_{\mathrm{epochs}} = 500$：显然，在训练轮数达到这一数值时，分类器的 TSS 评分首次开始显著下降。此外，训练轮数相对较少，因此可以在评估 CoNN 和集成方法性能的同时，开展大量运行。不过，我们也指出，针对这一问题仍可进一步微调超参数。

总体而言，RH 算法的超参数包括单个估计器的架构、具体加权方案的实现、特征重要性估计、每个估计器使用的特征数量，以及 RH 训练参数的调整等。此处考察的 RH 实现仅用于展示 RH 的概念验证。具体而言，我们进一步表明，在 SEP预测问题上，RH v2 的表现优于 RH v1。这说明 RH 算法的不同实现可能在同一数据集上产生不同效果，因此需要仔细调整 RH参数和选项，才能针对特定任务构建最优模型。

CoNN 和每个集成成员都有两个输出通道，用于预测输入属于正类（与 SEP 相关）或负类（非 SEP）的可能性。在二分类中，选择可能性较大的类别作为预测结果。要得出集成预测结果，RH分类器会选择获得最高加权票数的结果。我们展示了每种方法的混淆矩阵各项（真阳性预测，*TP*；真阴性预测，*TN*；假阳性预测，*FP*；以及假阴性预测，*FN*），这些结果是对 50 次随机训练—测试划分取平均后得到的。我们还使用多种指标来比较二分类预测结果，包括准确率、真实技巧评分（TSS）、Heidke技巧评分2（HSS）、精确率和召回率。

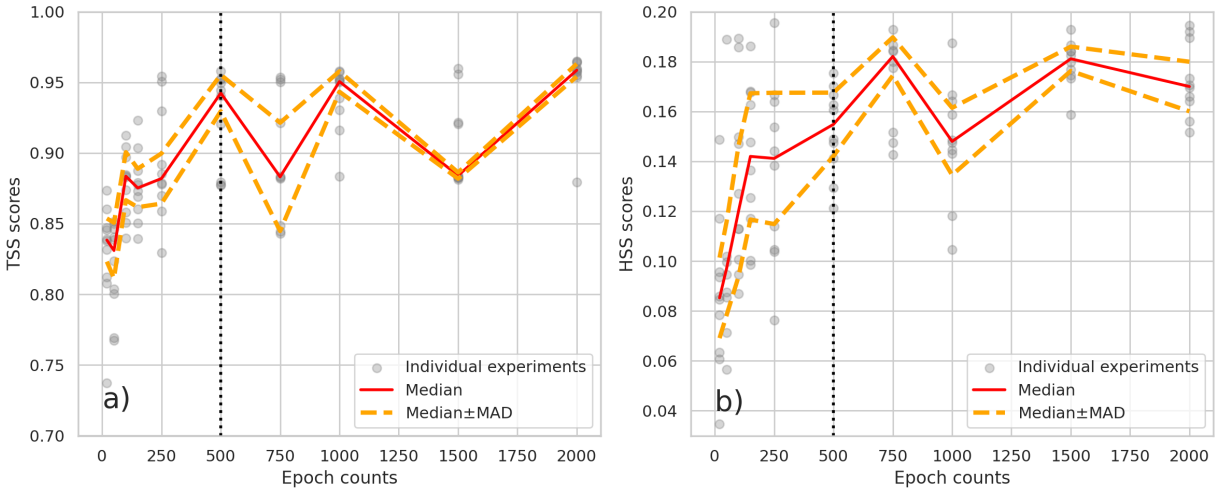

图4：CoNN在不同训练轮次和学习率为 $\alpha = 0.001$时的性能中位数（红色）和中位数绝对偏差（橙色）。面板(a)对应TSS评分，面板(b)对应HSS评分。灰色点表示单次实验的结果。黑色虚线表示被视为最优并用于本研究的训练轮次数。

有关这些指标的定义，请参阅例如 Bobra 和 Couvidat（2015年）及其中的参考文献。第 4 节也更详细地讨论了 TSS 和 HSS 分数。为减轻准确率指标易受数据集类别高度不平衡影响的问题，我们还考虑平衡准确率，其定义如下：

$$BA = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right). \tag{5}$$

我们还为每个分类器绘制受试者工作特征曲线，并计算曲线下面积（ROC曲线下面积）作为预测指标。单个估计器的类别概率由似然通过softmax函数计算得出。RH测试的正类概率（用于绘制ROC曲线）通过对每个估计器的概率预测进行加权平均计算得出，权重为各估计器的权重总和，并对所有权重总和进行归一化。

4. 结果与讨论

本研究采用的分类算法在混淆矩阵元素和各项预测评分方面的结果列于表1（所有分类器的汇总结果，以平均评分和标准差表示）和表2（所有分类器的汇总结果，以中位数评分和中位数绝对偏差表示）。结果的统计特性还通过图5中的箱线图以及图6中基于50次实验取平均的受试者工作特征（ROC）曲线加以展示。50次实验足以保证结论具有统计稳健性的一个指标是：对于所考虑的几乎所有指标和分类器，均值和中位数在最后一位有效数字上相一致。

表1和表2表明，总体而言，在空间天气预报中通常使用的指标——HSS和TSS——方面，集成方法的表现优于 CoNN 分类器，无论是从均值还是中位数来看。例如，CoNN 分类器的TSS评分中位数为TSS =0.906±0.042，而委员会方法和两种 RH 集成分类器的评分则分别升至TSS =0.926±0.023、TSS =0.915±0.010 和TSS =0.944±0.005（RH v2）。尽管HSS 评分相对较低，但从 CoNN 转向集成方法时，评分并未下降，甚至有所提高，从HSS =0.163±0.026 变为HSS =0.168±0.005、HSS =0.163±0.010 和HSS =0.168±0.008。这表明，与 CoNN 分类器相比，集成方法的平均表现更好。另一方面，我们在此指出，所有分类器的 TSS 和 HSS 评分几乎总是在不确定性范围内（标准差或中位数绝对偏差）相互重叠，而这些不确定性源于各项实验结果。

算法 / 指标	CoNN	委员会	RH v1	RH v2
TN	5233.0±74.4	5250.9±18.9	5244.2±25.5	5245.2±22.8
FP	237.0±74.4	219.0±18.9	225.7±25.5	224.8±22.8
FN	1.2±1.3	0.8±0.9	1.1±0.7	0.4±0.6
TP	22.8±1.3	23.2±0.9	22.9±0.7	23.6±0.6
精确率	0.094±0.022	0.096±0.005	0.093±0.009	0.096±0.008
召回率	0.945±0.053	0.966±0.038	0.956±0.031	0.985±0.025
准确率	0.957±0.013	0.960±0.003	0.959±0.005	0.959±0.004
平衡准确率	0.953±0.021	0.963±0.018	0.957±0.014	0.972±0.011
TSS	0.906±0.043	0.926±0.035	0.915±0.029	0.944±0.023
HSS	0.163±0.036	0.168±0.009	0.163±0.014	0.168±0.013
ROC曲线下面积	0.9903±0.0005	0.9907±0.0001	0.9901±0.0005	0.9906±0.0003

表1：本文所考虑的分类器性能的平均值和标准差。

算法 / 指标	CoNN	委员会	RH v1	RH v2
TN	5233.0±46.5	5250.9±10.0	5244.2±20.0	5245.2±15.5
FP	237.0±46.5	219.0±10.0	225.7±20.0	224.8±15.5
FN	1.2±1.5	0.8±0.5	1.1±0.0	0.4±0.0
TP	22.8±1.5	23.2±0.5	22.9±0.0	23.6±0.0
精确率	0.094±0.016	0.096±0.003	0.093±0.006	0.096±0.005
召回率	0.949±0.063	0.966±0.021	0.956±0.0	0.985±0.0
准确率	0.957±0.008	0.960±0.002	0.959±0.004	0.959±0.003
平衡准确率	0.953±0.021	0.963±0.011	0.957±0.005	0.972±0.002
TSS	0.906±0.042	0.926±0.023	0.915±0.010	0.944±0.005
HSS	0.163±0.026	0.168±0.005	0.163±0.010	0.168±0.008
ROC曲线下面积	0.9903±0.0002	0.9907±0.0001	0.9901±0.0003	0.9906±0.0001

表 2：本文所考虑的分类器性能的中位数和中位数绝对偏差（计算方式为各个得分与中位数之差的绝对值的中位数）。

分类器表现整体相近，这一点也从图6所示的 ROC 曲线中显而易见：这些曲线有显著重叠。Aminalragia-Giamini 等人（2021）在委员会方法的研究中，以及 Lavasa 等人（2021）在随机森林分类器的研究中，都曾注意到集成分类器的良好表现。有趣的是，CoNN 分类器在个别案例中的表现甚至可能优于单个集成分类器测试（图5中 CoNN 四分位距的上界可证实这一点），这说明了在多个训练—测试划分上评估方法并展示其对随机划分的稳健性的重要性。

从表1和表2中还可以看出另一种模式，即 CoNN 与集成分类器的标准差或中位数绝对偏差之间存在显著差异。例如，CoNN 的平均TSS得分及其标准差为TSS =0.906±0.043，而委员会为TSS =0.926±0.035，RH v1 为TSS =0.915±0.029，RH v2 为TSS =0.944±0.023。对于中位数和中位数绝对偏差，这种趋势更为明显（将 CoNN 的TSS =0.906±0.042与上述集成方法的TSS =0.926±0.023、TSS =0.915±0.010和TSS =0.944±0.005进行比较）。对于表中所考虑的所有其他得分，较小离散度的趋势同样存在。图5还表明，CoNN 的四分位距和须线跨度（表示四分位距之外各个实验结果的离散程度）通常更大，尤其是在HSS 技能得分的情况下。总体而言，这种行为表明，集成方法相对于数据集的随机训练—测试划分和训练过程具有较强的稳健性，而单个分类器的训练则可能失败。因此，这些集成算法复杂度的提高，因其在不平衡数据集上的稳健表现而是合理的（Galar 等，2012）。

表1和表2还表明：（1）RH 分类器未必优于委员会方法，

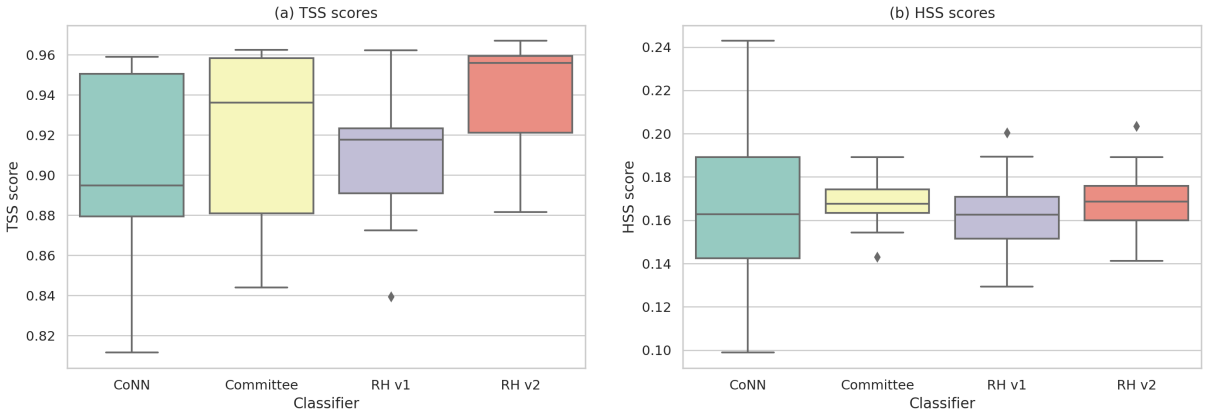

图5：箱线图总结了 CoNN、委员会以及所考虑的两个随机蜂群（RH）版本在50次实验中的表现，评估指标为 TSS 评分和 HSS 评分。每个彩色矩形覆盖评分的第二和第三四分位数，水平线表示中位数。须线表示箱体四分位距范围内最后一个单独测试结果的位置。菱形标记表示位于箱体四分位距范围之外的离群值。

（2）其表现取决于参数的选择。这一结论是基于以下事实得出的：与委员会或 RH v1 相比，RH v2 在几乎所有指标上的平均表现都相同或更好。委员会分类器与 RH 分类器之间另一个显著差异是，委员会的 ROC 曲线呈阶梯状，而 RH 的曲线更为平滑，这一点在图 6 中显而易见。RH 与委员会分类器的关键差异在于为每个集成成员选择特征的方式。委员会成员使用所有可用特征，而 RH 成员使用经过筛选的特征（在本例中为 12 个特征中的 4 个或 6 个），并采用确定性算法来计算每个委员会成员对最终结果的贡献。我们可以推测，为每个成员筛选特征能够提高预测评分，因为这有助于避免单个委员会成员关注噪声或无关特征，并防止其成员对相同的模型参数（权重和偏置）达成“一致”。这证实了特征选择过程的重要性；在太阳瞬变事件预测领域，特征选择仍是一个活跃的研究课题（Bobra和Couvidat，2015；Sadykov和Kosovichev，2017；Yeolekar等，2021）。此外，尽管委员会方法（Aminalragia-Giamini等，2021）有助于降低网络参数（权重和偏置）收敛到局部或全局极小值时对“偶然性”的依赖，但其集成成员仍是结构相似的 CoNN。RH 则引入了更多样化的集成成员种群，并将经过筛选、各不相同的特征集作为输入；与采用规模完整但几乎相同的学习器相比，这种做法可能更有益。

如前所述，表1和表2中的结果表明，所有接受测试的分类器（包括 RH 分类器）的精确率和 HSS 评分都相对较低。同时，相应的 TSS 评分较高。为了解模型出现这种表现的原因，我们来看看 SEP 预测中的一些模式。观察混淆矩阵元素的中位数，可以发现，按从大到小排列的 RH 分类器满足真阴性 $\approx 5244 \gg$ 假阳性 $\approx 226 \gg$ 真阳性 $\approx 22.9 \gg$ 假阴性 ≈ 1.1，以及真阴性 $\approx 5245 \gg$ 假阳性 $\approx 225 \gg$ 真阳性 $\approx 23.6 \gg$ 假阴性 ≈ 0.4。假设可以忽略高一阶的小量项，则可以将关注的指标改写为：

$$Precision = \frac{TP}{TP+FP} \approx \frac{TP}{FP} \tag{6}$$

$$HSS = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP+FN)(FN+TN)+(TP+FP)(FP+TN)} \approx$$

$$\approx \frac{2 \cdot TP \cdot TN}{TP \cdot TN + FP \cdot TN} \approx \frac{2 \cdot TP}{TP+FP} = 2 \cdot Precision \tag{7}$$

$$TSS = \frac{TP}{TP+FN} - \frac{FP}{FP+TN} = 1 - \frac{FN}{TP+FN} - \frac{FP}{FP+TN} \approx 1 - \frac{FN}{TP} - \frac{FP}{TN} \to 1 \tag{8}$$

在上述混淆矩阵元素条件下，精确率和 HSS 分数均

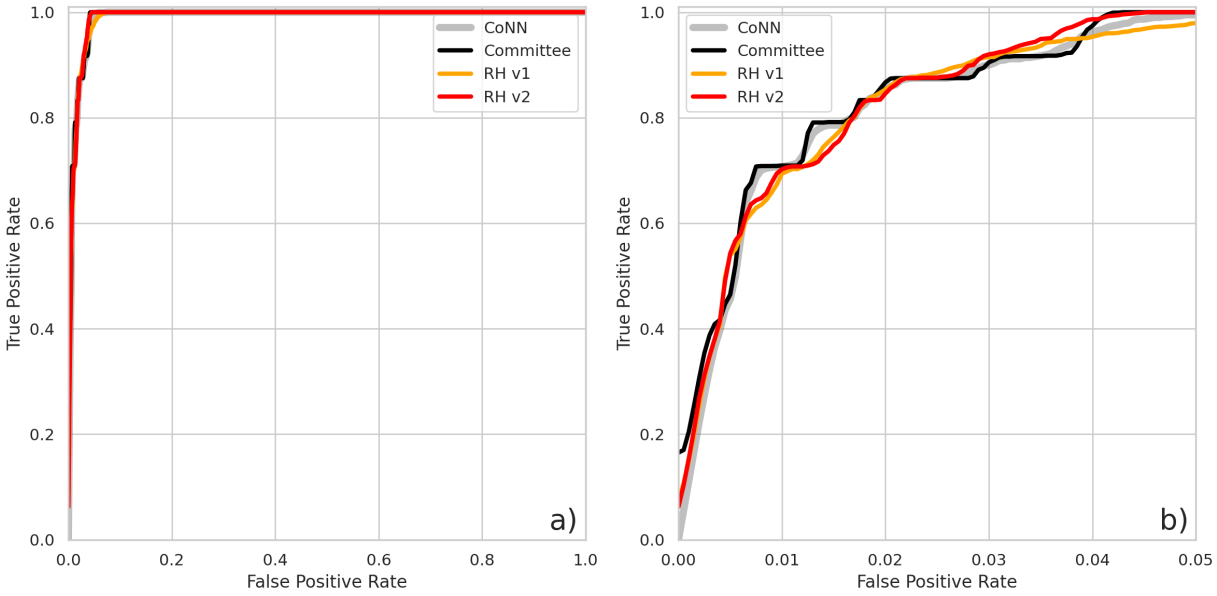

图 6：(a) 接收　　者操作特征（ROC）曲线：对本研究考虑的分类器进行 50 次实验后取平均。（b　　）相同曲线仅在较低 假阳性率范围内绘制。

由TP/FP 比值决定。对于本研究中实现的模型，该比值约为 $\sim 1/10$。同时，由于两个减数都很小，TSS评分非常接近 1：$FN/TP \sim 1/20$ 或更小，且$FP/TN \sim 1/23$。在测试数据集中，模型几乎捕获了每一次太阳高能粒子耀斑（平均有 24 次事件），但产生的虚警数量几乎是太阳高能粒子耀斑数量的十倍（RH v1 和 RH v2 的FP 中位数分别为 $= 225.7 \pm 20.0$ 和$FP = 224.8 \pm 15.5$）。此外，考虑到正样本与负样本的比值为 $\sim 1/285$，该数据集也高度不平衡。上述条件使 HSS评分受到$HSS \approx 2 \times TP/FP \leq 2 \times P/FP$的限制。它们还使 HSS 分数极易受到虚警数量变化的影响。例如，若能将虚警数量减半（RH v2 的$FP = 224.8 \rightarrow FP = 112.4$ 和$TN = 5245.2 \rightarrow TN = 5357.6$），相应的 HSS评分将几乎翻倍，达到$HSS \approx 0.290$。与此同时，TSS评分将从$TSS \approx 0.944$ 升至$TSS \approx 0.963$，相对变化幅度较小。由此可见，对于高度不平衡的数据集，较低的 HSS 分数与模型产生的大量虚警有关（相对于测试数据集中正样本的数量），但 TSS 分数仍可能非常接近 1。

此外，我们注意到，尽管 Heidke技巧评分2（HSS）通常被标注为衡量相对于随机机会预报的性能，但此处给出的预报显然并非随机：对于 RH v1，漏报率 假阴性/(假阴性 + 真阳性) ≈ 0.046 较低（每次试验几乎都能命中每个太阳高能粒子事件），且虚警率 假阳性/(假阳性 + 真阴性) ≈ 0.041 也较低；对于 RH v2，漏报率更为理想，为 假阴性/(假阴性 + 真阳性) ≈ 0.017，而虚警率仍约为 0.041。尽管如此，HSS 分数与 0 的偏差并不明显（平均而言，RH v1 的 $HSS = 0.163 \pm 0.014$，RH v2 的$HSS = 0.168 \pm 0.013$）。因此，我们认为，将较低的 HSS 分数与预报接近随机机会预报联系起来是不恰当的。此外，上一段中的示例表明，在预报用途上，将 HSS 分数与TSS 分数结合考虑非常有益。与 TSS 相比，HSS分数对虚警 假阳性的减少更为敏感。因此，依据模型产生的虚警数量，可以更好地区分漏报事件数量大致相同的模型。

虽然这里没有明确测试全清预报，但此处实现的分类方法表明，它们在全清场景中具有实用性。这里需要指出的是，全清太阳高能粒子预测通常被定义为全日面预测任务，并且还必须指定发布预报的时间窗口

。我们在此讨论一个相关但并不完全相同的问题：如何在保持低虚警率的同时，可靠地预测每一次太阳高能粒子活跃耀斑（漏报事件极不可取）。对于这类预报，通常希望漏报率极低（漏报事件的平均比率为 $FN/(FN+TP) \approx 0.05$，即便使用 CoNN 也是如此）（Sadykov等，2021）。对于 RH v2 分类器，漏报率几乎为零（$FN/(FN+TP) \approx 0.017$），因此 RH 适用于这项任务。尽管 CoNN 分类器的 FN 中位数仍相对较低（$FN = 1.2 \pm 1.5$，见表 2），委员会方法的中位数也较低（$FN = 0.8 \pm 0.5$），但 RH v2 的 FN 更低，为 $= 0.4 \pm 0.0$。中位绝对偏差为 0.0，表明 RH v2 的 50 次试验中，超过一半（即超过 25 次）的结果是假阴性预测为零。总体而言，我们还可以预期，RH 方法的某些配置或许能够在其他机器学习问题中实现漏报事件近乎为零且虚警率较低的预报，包括涉及高度不平衡类别的问题。

5. 结论

在本研究中，我们提出了一种集成算法——Random Hivemind（RH），并比较了它的两种选定实现与传统神经网络（CoNN）以及针对 CoNN 的委员会集成方法。比较针对的是根据宿主软 X 射线耀斑的特性预测太阳高能粒子（SEP）事件这一问题。本研究的主要结果如下：

- RH 算法的性能取决于其实现方式和训练参数（可能包括输入特征的数量、特征加权方案、学习率提升策略等）。RH v1 和 RH v2 实现对应的 TSS 和 HSS 分数分别为$TSS = 0.915 \pm 0.010$ 和$TSS = 0.944 \pm 0.005$，以及$HSS = 0.163 \pm 0.010$ 和 $HSS = 0.168 \pm 0.008$。

- 两种集成方法（委员会方法和随机蜂群）在数据集的随机训练—测试划分下均表现稳健，这一点反映在较低的标准差或中位数绝对偏差上。尽管 CoNN 在预报指标方面的表现通常与委员会方法相当，但其标准差要高得多（中位数绝对偏差也往往更高）。

- 与空间天气预报中通常使用的指标 HSS和 TSS相比，两种集成方法在均值和中位数方面的表现相近或更好。可以比较 CoNN 的中位数 $TSS = 0.906 \pm 0.042$ 与委员会方法、RH v1 和 RH v2 对应的 $TSS = 0.926 \pm 0.023$、$TSS = 0.915 \pm 0.010$ 和 $TSS =0.944 \pm 0.005$；还可以比较 $HSS = 0.163 \pm 0.026$ 与 $HSS =0.168 \pm 0.005$、$HSS = 0.163 \pm 0.010$ 和 $HSS = 0.168 \pm 0.008$。

- 平均而言，RH v2 集成分类器在几乎所有指标上的表现都优于委员会、CoNN 和 RH v1 方法，并且在十次随机训练-测试划分实验中结果稳定一致。

- 所有分类器（包括 RH）的表现都表明，在 SEP 预测问题上，精确率和HSS 评分相对较低。尽管如此，将HSS 作为预报的补充指标仍很有益，因为与TSS相比，它更容易受到虚警率下降的影响。这有助于根据模型产生的虚警数量，更好地区分漏报事件数量大致相同的模型。

- 所有分类器的假阴性预测数量都非常少。CoNN、委员会、RH v1 和 RH v2 分类器的假阴性 中位数分别为 $= 1.2 \pm 1.5$、 $=0.8 \pm 0.5$、 $= 1.1 \pm 0.0$ 和 $= 0.4 \pm 0.0$。不过，前文提到的 RH 分类器的稳健性，尤其是 RH v2 的稳健性，使其成为解决太阳高能粒子活跃耀斑类似“解除警报”预报问题的可行选择。

根据上述结果，我们可以得出结论：随机蜂群是一种有效的机器学习算法，即使存在类别不平衡，也能表现良好。总体而言，随机蜂群的表现与 CoNN 相当或更好，也优于无权重同构 CoNN 委员会机器。要全面了解随机蜂群方法的潜力，尤其是其在空间天气预报方面的用途（包括对太阳高能粒子事件进行“一切平静”预报），还需要进一步研究该方法（包括采用不同的特征权重实现与处理方式、调整学习率和迭代轮数，以及在随机蜂群训练中考虑的耀斑类别边界

训练）对于理解其总体潜力，尤其是其在空间天气预报中的用途（包括太阳高能粒子事件的“一切平静”预报）至关重要。

致谢

作者感谢匿名审稿人提出的建议，这些建议显著提升了稿件的质量。本研究得到 NASA早期阶段创新计划项目资助（编号 80NSSC20K0302 ）、 NASA 宜居世界计划资助（编号 80NSSC19K0068 ）、美国国家科学基金会 EarthCube 项目资助（编号 1639683、1743321 和 1927578），以及美国国家科学基金会项目资助（编号 1835958）。VMS 感谢 NSF FDSS 项目资助（编号 1936361）和美国国家科学基金会项目资助（编号 1835958）。

参考文献

C. C. Balch，《SEC质子预测模型：验证与分析》，《辐射测量》30（1999）231–250。C. C. Balch，《空间天气预报中心太阳高能粒子预测模型的更新验证》，《空间天气》6（2008）S01001。H. M. Bain，R. A. Steenburgh，T. G. Onsager，E. M. Stitely，《美国国家海洋和大气管理局空间天气预报中心质子事件预报性能与技巧概述》，《空间天气》19（2021年）e2020SW002670。D. V. Reames，《太阳高能粒子：理解源、加速与传播的现代入门》，第

978, 2021 10．卷。

1007/978-3-030-66402-2
doi：。D. F. Ryan，R. O. Milligan，P. T. Gallagher，B. R. Dennis，A. K. Tolbert，R. A. Schwartz，C. A. Young，《利用GOES X射线观测研究三个太阳活动周内太阳耀斑的热特性》，《天体物理学杂志增刊系列》202（2012）11。V. M. Sadykov，A. G. Kosovichev，I. N. Kitiashvili，A. Frolov，《太阳耀斑软X射线辐射的统计特性》，《天体物理学杂志》874（2019）19。H. A. Garcia，《高能质子事件的温度与硬X射线特征》，《天体物理学杂志》420（1994）422。H. A. Garcia，《利用太阳软X射线预报质子风暴的发生与强度》，《空间天气》2（2004）S02002。S. W. Kahler，A. G. Ling，《利用耀斑X射线峰值比预报太阳高能粒子（SEP）事件》，《空间天气与空间气候杂志》8（2018）A47。A. G. Ling，S. W. Kahler，《大型太阳X射线耀斑的峰值温度及相关日冕物质抛射速度和宽度》，《天体物理学杂志》891（2020）54。S. W. Kahler，A. G. Ling，《太阳X射线耀斑时间尺度和峰值温度与相关日冕物质抛射的比较》，《天体物理学杂志》934（2022）175。S. W. Kahler，A. Vourlidas，《渐进型太阳高能粒子事件的强度和能量与相关日冕物质抛射动力学特性的比较》，《天体物理学杂志》769（2013）143。I. G. Richardson，M. L. Mays，B. J. Thompson，《基于日冕物质抛射速度和方向以及相关太阳现象观测的简单算法预测太阳高能粒子事件质子峰值强度》，《空间天气》16（2018）1862–1881。M. Laurenza，E. W. Cliver，J. Hewitt，M. Storini，A. G. Ling，C. C. Balch，M. L. Kaiser，《一种基于耀斑位置、耀斑大小和粒子逃逸证据的太阳高能粒子事件短期预警技术》，《空间天气杂志》7（2009）S04008。M. Laurenza，T. Alberti，E. W. Cliver，《一种基于ESPERTA的中等至极端太阳质子事件短期预报工具》，《天体物理学杂志》857（2018）107。S. Aminalragia-Giamini，S. Raptis，A. Anastasiadis，A. Tsigkanos，I. Sandberg，A. Papaioannou，C. Papadimitriou，P. Jiggens，A. Aran，I. A. Daglis，《利用太阳耀斑软X射线测量和机器学习预测太阳高能粒子事件的发生》，《空间天气与空间气候杂志》11（2021年）59。S. F. Boubrahimi，B. Aydin，P. Martens，R. Angryk，《利用GOES卫星数据预测≥100兆电子伏特太阳高能粒子事件》，载于：2017年IEEE大数据国际会议（Big Data），2017，第2533–2542页。doi：10.1109/BigData.2017.8258212。E. Lavasa，G. Giannopoulos，A. Papaioannou，A. Anastasiadis，I. A. Daglis，A. Aran，D. Pacheco，B. Sanahuja，《利用机器学习技术评估太阳高能粒子的可预测性》，《太阳物理学》296（2021年）107。J. Torres，L. Zhao，P. K. Chan，M. Zhang，《利用日冕物质抛射特性预测SEP事件的机器学习方法》，《空间天气》20（2022年）e2021SW002797。N. Nishizuka，Y. Kubo，K. Sugiura，M. Den，M. Ishii，《可靠的太阳耀斑概率预报：Deep Flare Net-Reliable（DeFN-R）》，《天体物理学杂志》899（2020）150。S. Afaq，S. Rao，《训练神经网络时训练轮数的重要性》，《国际科学与技术研究杂志》9（2020）485–488。K. You，M. Long，J. Wang，M. I. Jordan，《学习率衰减如何助力现代神经网络？》，arXiv电子预印本（2019）arXiv:1908.01878。P. Cunningham，《基于特征选择的分类集成中的过拟合与多样性》，Trinity研究档案库（2000）。A. Ali，V. Sadykov，A. Kosovichev，I. N. Kitiashvili，V. Oria，G. M. Nita，E. Illarionov，P. M. O’Keefe，F. Francis，C.-J. Chong，P. Kosovich，R. D. Marroquin，《利用GOES质子与软X射线通量统计预测第22—24太阳活动周的太阳质子事件》，arXiv电子预印本（2023）arXiv:2303.05446。V. Sadykov，A. Kosovichev，I. Kitiashvili，V. Oria，G. M. Nita，E. Illarionov，P. O’Keefe，Y. Jiang，S. Fereira，A. Ali，《利用机器学习预测太阳质子事件：与业务预报的比较及“一切平静”展望》，arXiv电子预印本（2021年）arXiv:2107.03911。M. Stumpo，S. Benella，M. Laurenza，T. Alberti，G. Consolini，M. F. Marcucci，《太阳质子事件统计预报中的未决问题：机器学习视角》，《空间天气杂志》19（2021年）e2021SW002794。

A. Ahmadzadeh，B. Aydin，M. K. Georgoulis，D. J. Kempton，S. S. Mahajan，R. A. Angryk，《如何训练耀斑预测模型：重新审视稀有事件的鲁棒抽样》，《天体物理学杂志增刊系列》254（2021年）23。Y. Chen，D. J. Kempton，A. Ahmadzadeh，R. A. Angryk，《面向耀斑预报的合成多变量时间序列生成》，载于：L. Rutkowski，R. Scherer，M. Korytkowski，W. Pedrycz，R. Tadeusiewicz，J. M. Zurada（编），《人工智能与软计算》，施普林格国际出版，尚，2021年，第296–307页。M. Galar，A. Fernandez，E. Barrenechea，H. Bustince，F. Herrera，《类别不平衡问题的集成方法综述：装袋法、提升法和混合方法》，《IEEE系统、人和控制论汇刊C辑（应用与评论）》42（2012）463–484。P. M. O'Keefe，V. M. Sadykov，A. G. Kosovichev，G. M. Nita，V. Oria，S. Sharma，F. Francis，C. J. Chong，《机器学习应用中处理高度不平衡数据》，Zenodo，2022年。URL：https://doi.org/10.5281/zenodo.6780972。doi：10.5281/zenodo.6780972。V. M. Sadykov，A. G. Kosovichev，V. Oria，G. M. Nita，《太阳耀斑交互式多仪器数据库》，《天体物理学杂志增刊系列》231（2017年）6。A. Papaioannou，I. Sandberg，A. Anastasiadis，A. Kouloumvakos，M. K. Georgoulis，K. Tziotziou，G. Tsiropoula，P. Jiggens，A. Hilgers，《太阳耀斑、日冕物质抛射和太阳高能粒子事件特征》，《空间天气与空间气候杂志》6（2016）A42。L. Breiman，《随机森林》，《机器学习》45（2001）5–32。M. Tietz，T. J. Fan，D. Nouri，B. Bossan，skorch开发者，《skorch：一个兼容scikit-learn、封装PyTorch的神经网络库》，2017年。URL：https://skorch.readthedocs.io/en/stable/。D. P. Kingma，J. Ba，《Adam：一种随机优化方法》，CoRR abs/1412.6980（2015）。M. G. Bobra，S. Couvidat，《利用SDO/HMI矢量磁场数据和机器学习算法预测太阳耀斑》，《天体物理学杂志》798（2015）135。V. M. Sadykov，A. G. Kosovichev，《视线方向磁场特征与太阳耀斑预报之间的关系》，《天体物理学杂志》849（2017）148。A. Yeolekar，S. Patel，S. Talla，K. Puthucode，A. Ahmadzadeh，V. M. Sadykov，R. A. Angryk，《耀斑预报测试平台上的特征选择：24种方法的比较研究》，载于：2021年国际数据挖掘研讨会（ICDMW），IEEE计算机学会，洛斯阿拉米托斯，加利福尼亚州，美国，2021年，第1067–1076页。URL：https://doi.ieeecomputersociety.org/10.1109/ICDMW53433.2021.00138。doi：10.1109/ICDMW53433.2021.00138。