

一种用于猴痘诊断的可解释自然启发框架：结合 NGBoost 和非洲秃鹫优化算法的 Xception 特征

Ahmadreza Shateri^a, Negar Nourani^a, Morteza Dorrigiv^a, Hamid Nasiri^{a,b,*}

^a*Electrical and Computer Engineering Department, Semnan University, Semnan, Iran*

^b*School of Computing and Communications, Lancaster University, Lancaster, UK*

Abstract

猴痘在全球的近期传播，尤其是在历史上不常见的地区，引起了重大的公共卫生关注。早期和准确的诊断对于有效的疾病管理和控制至关重要。对此，本研究提出了一种基于深度学习的新框架，用于从皮肤病变图像中自动检测猴痘，利用迁移学习、降维和先进的机器学习技术的力量。我们使用新开发的猴痘皮肤病变数据集（MSLD），该数据集包括猴痘、水痘和麻疹的图像，以训练和评估我们的模型。所提出的框架采用 Xception 架构进行深度特征提取，随后进行主成分分析（PCA）以降维，并使用自然梯度提升（NGBoost）算法进行分类。为了优化模型的性能和泛化能力，我们引入了非洲秃鹫优化算法（AVOA）进行超参数调优，确保有效地探索参数空间。我们的结果表明，所提出的 AVOA-NGBoost 模型达到最先进的性能，准确率为 97.53%，F1 分数为 97.72%，AUC 为 97.47%。此外，我们使用 Grad-CAM 和 LIME 技术增强模型的可解释性，提供对决策过程的见解，并突出影响分类的关键特征。该框架提供了一种高度精确和有效的诊断工具，有可能在资源有限的环境中帮助医疗服务提供者进行早期检测和诊断。

Keywords: Monkeypox Disease, Transfer Learning, African Vultures Optimization Algorithm (AVOA), Natural Gradient Boosting (NGBoost), Grad-CAM

1. 介绍

自 2022 年 5 月以来，在此前未观察到该疾病的国家中出现了猴痘病例。截至本文撰写时，115 个非流行国家报告了 95,196 例确诊病例，7 个流行国家报告了 4,322 例病例

*Corresponding author

Email address: h.nasiri@lancaster.ac.uk (Hamid Nasiri)

(Centers for Disease Control and Prevention, 2022)。根据世界卫生组织 (WHO) 的报告, 这是首次在不同地理区域的流行和非流行国家同时报告猴痘病例。在 COVID-19 疫情继续对世界构成挑战的同时, 由猴痘病毒引发的新疫情引起了公共卫生当局的担忧, 担心这是否可能构成新的威胁 (Yang, 2022)。

猴痘的病原体, 称为猴痘病毒 (MPXV), 与天花、牛痘和痘苗病毒一样, 属于痘病毒科的正痘病毒属, 它可以导致多种疾病 (Realegeno et al., 2017)。尽管 MPXV 传染给人类的确切方式仍然很大程度上未知, 但大多数感染被认为是人畜共患病, 通过直接或间接接触从动物传播给人类。在某些情况下, 人际间传播发生, 可能是通过大的呼吸道飞沫或通过直接或间接接触体液、病变物质、被污染的物体或其他物质 (Alakunle et al., 2020)。该疾病首次在 1958 年被检测到, 当时在用于研究目的的生物群中观察到了两例类似痘病的病例。随后, 1970 年在刚果民主共和国报告了首例人类病例 (Kumar et al., 2022)。

暴露后, 症状可能在几天甚至几周内都不会出现。猴痘的早期症状类似于流感, 包括发热、寒战、头痛、疼痛、肌肉痛、疲劳和淋巴结肿大 (Fatima and Mandava, 2022)。通常几天后出现皮疹, 最初表现为痛苦、红色、扁平的丘疹。这些隆起区域最终会发展成充满液体的水疱, 在两到四周内变硬并脱落。需要注意的是, 不是所有患者都会出现每一种症状 (Adalja and Inglesby, 2022)。猴痘的临床表现包括症状和病变, 通常难以与天花区分 (Rizk et al., 2022)。虽然猴痘通常表现为比历史上更严重的天花轻微的症状, 但仍然可能致命, 死亡率为 1% 到 10%, 在儿童和青少年中的死亡率更高, 而且在免疫缺陷的个体中病程更为严重。淋巴结病, 或淋巴结的肿大, 观察到在超过 90% 的患者中, 并作为与天花区分的特征。由于其症状和病变的非特异性性质, 必须考虑广泛的鉴别诊断, 包括水痘、传染性软疣、麻疹、立克次氏体感染、细菌性皮肤感染 (例如由金黄色葡萄球菌引起的感染)、炭疽、疥疮、梅毒和药物反应, 以及其他非感染性皮疹原因 (Petersen et al., 2019)。

在没有诊断测试的情况下, 临床上区分皮疹疾病是具有挑战性的。这些评估在识别猴痘方面起着关键作用, 并且可以使用临床样本对疾病进行分类。这些测试在与患者临床信息结合时最为有效, 如疫苗接种历史 (Koenig et al., 2022)。常见的方法, 如从临床样本中分离病毒、电子显微镜和免疫组织化学, 仍然有效, 但需要先进的技术技能和培训。可以使用实时聚合酶链反应 (PCR) 检测样本中的牛痘病毒或猴痘病毒的存在, 尽管 PCR 在检测病毒 DNA 方面有效, 但它的使用通常仅限于大型实验室, 这使得在农村或资源匮乏

乏地区使用较不实际。

四芯体正痘类病毒生物威胁警报提供了一种不需要先进设备或专业技术人员的替代方案，因此适合资源有限的地区。然而，由于该测试旨在诊断正痘病毒属而非专门针对猴痘，其使用理想限于疾病流行的地区 (MacNeil et al., 2009)。鉴于这些潜力，得出了一个重要结论：需要一种新方法来弥补这一空白。

近年来，诸如人工智能和机器学习等领域的进步使它们成为临床医生最有用的工具之一 (Bohr and Memarzadeh, 2020)。深度学习是人工智能的一个子领域，它帮助创建模型，无需人工参与即可自动提取特征，训练模型并生成结果 (Myszczyńska et al., 2020)。各种成像技术已经在医学领域中使用，帮助医疗专业人员诊断各种疾病，包括乳腺癌 (Abbasniya et al., 2022) 和其他呼吸系统疾病，如肺炎和结核病 (Haloi et al., 2018)，以及 COVID 和其他症状。最近对医学图像通过深度学习进行分析进行了广泛研究。由于猴痘病例数量众多以及测试套件的不足，这一领域存在需求。由于专业临床医生的稀缺，很难为每个医院提供一位。此外，深度学习模型可以协助解决诸如 RT-PCR 套件供应不足、高昂成本以及诊断结果等待时间漫长等挑战，这在专业临床医生缺席时可能补充传统测试方法 (Ozturk et al., 2020)。还研究了深度学习策略，以确定机器学习是否能够制定有效的猴痘诊断分诊策略的问题提供可行的解决方案。在这项研究中，部署了深度学习模型以提高猴痘诊断的准确性。

本分析侧重于使用新开发的猴痘图像数据集，即猴痘皮肤病变数据集 (MSLD)，对猴痘检测的性能进行研究。为了应对猴痘诊断中的挑战，我们采用了基于深度学习的卷积神经网络 (CNN) 模型和提升算法。CNN 因其独立学习特征的能力而广受认可，已成为医学影像分类中最受欢迎的方法。具体来说，我们利用了 Xception 架构，这是一种以减少梯度消失问题、增强特征复用并最小化参数数量而闻名的尖端技术。Xception 的设计包含线性残差连接，这在训练期间改善了梯度流动并使信息能够在层之间有效传播。这些残差连接将位于每个中间流动模块内的层连接在一起，而不是将所有层连接在一起。与其他图像处理技术相比，由于具有深度可分离卷积，Xception 的架构更加复杂，使其能够捕获更多视觉信息，使其特别适合此任务 (Chollet, 2016)。本研究的主要目标是通过分析皮肤病变图像开发一种可靠的框架，以检测猴痘患者。我们提出了一种具有高精度的方法，以提高检测速度和准确性，以区分非猴痘（如水痘和麻疹）。为了验证所提出的方法，进行了一系列实验，并将结果与其他竞争特征提取和参数优化算法进行比较。统计测试进一步证

实了我们方法的稳定性和有效性。本工作的贡献可以总结如下：

1. 我们利用基于深度学习的模型，包括 ResNet50V2、ResNet101V2、ResNet152V2、InceptionV3、InceptionResNetV2、MobileNet、MobileNetV2、DenseNet121、DenseNet169、DenseNet201、NASNetMobile、NASNetLarge 和 Xception 架构，研究从皮肤病变图像中进行猴痘早期诊断和筛查的潜力。
2. 我们实现了 NGBoost 模型，这是一种相对较新的概率方法，用于猴痘分类。该模型利用其概率预测能力和处理复杂高维数据的能力，同时提供校准良好的不确定性估计。
3. 非洲秃鹫优化算法 (AVOA) 是一种新颖的自然启发的元启发式优化技术，被用来自动调整 NGBoost 模型的超参数。该优化策略通过有效地探索超参数空间和识别最佳配置来增强模型性能，从而提高分类准确性和泛化能力。
4. 我们通过 Grad-Cam 和 LIME 技术增强模型的可解释性，这为决策过程提供了宝贵的见解，并识别出影响猴痘病变分类的关键特征。
5. 本研究通过引入一种基于迁移学习的改进且高效的方法，解决了猴痘早期检测的挑战，这对于增强疾病监测和控制特别是对于像猴痘这样的稀有疾病具有重要价值。

本文的其余部分组织如下：

第 2 节回顾了医学图像分析中深度学习模型的相关工作。第 3 节详细介绍了猴痘皮肤病灶数据集 (MSLD) 和数据增强技术。第 4 节解释了方法论，包括使用预训练 CNN 进行数据预处理和特征提取。还概述了使用分类器如 LightGBM、SVM、XGBoost 和 NGBoost，并通过非洲秃鹫优化算法 (AVOA) 进行超参数优化。第 6 节基于准确性、F1-分数和 AUC 评估模型，并通过 Grad-CAM 讨论模型可解释性。最后，第 7 节总结了结果和未来的研究方向。

2. 相关工作

在本节中，我们重新审视了最近一些与深度学习相关的方法及其在医学图像分析中的应用。

由于高级硬件的广泛可用性，近年来使用深度学习技术进行医学图像分析经历了显著的增长。通常，对于基于分类的问题，人们更加重视特征提取，这可以说是任何表示学习任务中最重要方面。

(Pramanik et al., 2023) 开发了一个用于检测皮肤病变图像中猴痘病毒的集成学习框架。他们的方法涉及在 MSLD 上对三个预训练的深度学习模型——Inception V3、Xception 和 DenseNet169 进行微调。研究人员从这些基础学习器中提取预测概率，并将其纳入到一个集成框架中。他们提出了一种 Beta 函数归一化方案，以有效聚合来自基础学习器的互补信息，随后使用和规则集成方法结合结果。该框架在一个公开的猴痘病变数据集上使用五折交叉验证进行了评估，取得了出色的性能，准确率为 93.39 %，精确率为 88.91 %，召回率为 96.78 % 和 F1 分数为 92.35 %。

(Almufareh et al., 2023) 提出了一种无创、非接触的计算机视觉方法，通过分析皮肤病变图像来进行猴痘的诊断。他们采用深度学习技术将病变分类为猴痘阳性或阴性。该方法在两个公共数据集上进行了评估——Kaggle 猴痘皮肤图像数据集 (MSID) 和 MSLD。通过敏感性、特异性和平衡准确性指标对多个深度学习模型进行了基准测试。所提出的方法取得了令人期待的结果，其中 MobileNet 在 MSID 数据集上实现了 96.55 % 的平衡准确率、0.93 的特异性和最大的敏感性。相比之下，InceptionV3 在 MSLD 数据集上展示出最佳指标，具有 94 % 的平衡准确率、最大特异性和 0.88 的敏感性。

(Saha et al., 2023) 的研究重点是对多个预训练的 CNN 模型进行比较分析。他们的研究使用了各种架构，包括 VGG-16、VGG-19、Resnet50、Inception-V3、Densnet、Xception、MobileNetV2、Alexnet、Lenet，以及用于分类的多数投票法。他们结合了多个数据集，如猴痘对水痘、猴痘对麻疹、猴痘对正常皮肤，以及猴痘对所有疾病。在他们的实验中，多数投票法在识别猴痘与水痘时获得了 97 % 的准确率，Xception 在猴痘与麻疹时达到了 79 % 的准确率，MobileNetV2 在猴痘与正常皮肤时取得了 96 % 的准确率，LeNet 在猴痘对所有疾病时表现为 80 % 的准确率。他们的目标是比较不同的 CNN 架构用于猴痘与其他类似疾病和正常皮肤的分类。

为了应对区分猴痘和麻疹症状的挑战，(Ariansyah et al., 2023) 开发了一种使用 CNN 架构和 VGG-16 迁移学习的图像分类方法。他们在第 15 轮次中，成功实现了 83.333 % 的高准确度，用于区分猴痘和麻疹症状。其目标是开发一种基于视觉表现自动区分这两种疾病的图像分类器。

(Sitaula and Shahi, 2022) 比较了 13 种不同的预训练深度学习模型用于检测猴痘病毒。他们通过添加通用自定义层对所有模型进行了微调，并使用精确度、召回率、F1-得分和准确率分析了结果。在确定了表现最佳的模型后，他们通过对概率输出进行多数投票对这些模型进行集成，以提高整体性能。实验在一个公开数据集上进行，并通过他们提出的集成方法实现了 85.44 % 的精确度、85.47 % 的召回率、85.40 % 的 F1-得分和 87.13 % 的准确率。这表明集成多个深度学习模型能够增强猴痘检测能力的强大之处。

(Irmak et al., 2022) 使用开源的 2022 猴痘皮肤图像数据集研究了三种预训练的 CNN 模型——MobileNetV2、VGG16 和 VGG19 的性能。在这些架构中，MobileNetV2 表现最佳，其准确率为 91.38 %，精确度为 90.5 %，召回率为 86.75 %，F1 分数为 88.25 %，而 VGG16 和 VGG19 的准确率分别为 83.62 % 和 78.45 %。

采取一种实际的猴痘诊断方法，(Sahin et al., 2022) 开发了一款移动应用程序，利用深度学习模型从相机图像进行初步的猴痘诊断。该 CNN 模型通过迁移学习在公共病变图像上进行了训练。在模型选择并转换为 TensorFlow Lite 后，他们在具有快速推理时间的设备上测试了该应用程序。该系统允许感染者在寻求专家确认之前获得初步诊断。测试显示准确率为 91.11 %。

(Islam et al., 2022) 引入了 2022 年猴痘皮肤图像数据集，这是迄今为止最大规模的同类公共数据集。他们利用这个数据集研究基于皮肤图像的猴痘检测深度学习模型。他们的研究发现，先进的深度 AI 模型在检测猴痘方面具有巨大潜力，准确率达到 83 %。

(Haque et al., 2022) 研究了卷积块注意模块 (CBAM) 与深度迁移学习相结合在基于图像的猴痘病分类中的应用。他们实施了五种深度学习模型 - VGG19、Xception、DenseNet121、EfficientNetB3 和 MobileNetV2，这些模型集成了通道和空间注意力。对比分析显示，结合了 Dense 层的 Xception-CBAM 模型性能最佳，在猴痘与其他疾病的分类中实现了 83.89 % 的验证准确性。

不同于以往利用视觉数据集并专注于从皮肤病图像中提取特征来诊断猴痘的研究，(Farzipour et al., 2023) 创建了一个全新的纯文本数据集，基于由全球健康组织收集和公布的数据，并被 WHO 用于展示症状与猴痘疾病之间的关系。利用该数据集，他们分析并比较了几种机器学习方法，包括极限梯度提升 (XGBoost)、CatBoost、LightGBM、支持向量机 (SVM) 和随机森林。他们的目标是提供一个基于症状诊断猴痘的机器学习模型。在他们的实验中，XGBoost 取得了最佳性能。为了评估模型的稳健性，他们使用了 k 折交

叉验证，在五个不同的测试集划分中达到了平均 90 % 的准确率。此外，他们利用 Shapley 加法解释 (SHAP) 来检查和解释 XGBoost 模型的输出。

尽管在使用深度学习进行猴痘检测方面取得了显著进展，但现有方法仍面临关键限制。许多方法依赖于集成技术或预训练的架构，而未能有效优化超参数，这可能会影响性能。此外，尽管有时使用了可解释性技术，但它们在工作流程中的集成仍不一致，降低了模型预测的实际可信度。

我们提出的框架通过利用 AVOA 进行自动化和高效的超参数调整来解决这些挑战。这使得模型在准确性和预测速度方面表现极其优秀。此外，我们加入了解释性技术，例如 Grad-CAM 和 LIME，以确保决策过程透明且与临床相关。这些增强使得与之前的研究相比，该框架在猴痘检测中更加稳健和实用。

3. 数据集描述

一项在 Kaggle (一个面向数据科学家和机器学习从业人员的在线社区) 上公开的数据集，用作本研究的一部分实验的基础。

3.1. 数据收集

用于训练迁移学习模型的数据集名为“猴痘皮肤病变数据集”(MSLD) (Ali et al., 2022)。这是一个关于猴痘与非猴痘图像的二分类数据集，发布在 Kaggle 上。MSLD 是通过收集和分析从各种网页抓取来源获得的图像生成的，包括新闻门户网站、网站和公开病例报告。

该数据集包含 228 张图像，包括 102 例猴痘病例和 126 例非猴痘病例 (包括水痘和麻疹)。这些图像的分辨率为 $224 \times 224 \times 3$ 。Figure 1 展示了一些来自两个类别的图像示例。

3.2. 数据增强

获得的数据集包括一个由其创建者扩充的原始版本。在训练深度神经网络时，数据集增强用于增加数据集的规模以提高网络性能 (Lewy and Mańdziuk, 2023)。创建者应用了多种增强技术，包括旋转、平移、反射、剪切、色调、饱和度、对比度和亮度抖动、噪声和缩放。这些方法旨在使数据集多样化，并提高模型的泛化能力 (Nayak et al., 2023)。通过使用这些方法，模型接触到更广泛的变化范围，并获得对图像基础特征的更深理解。

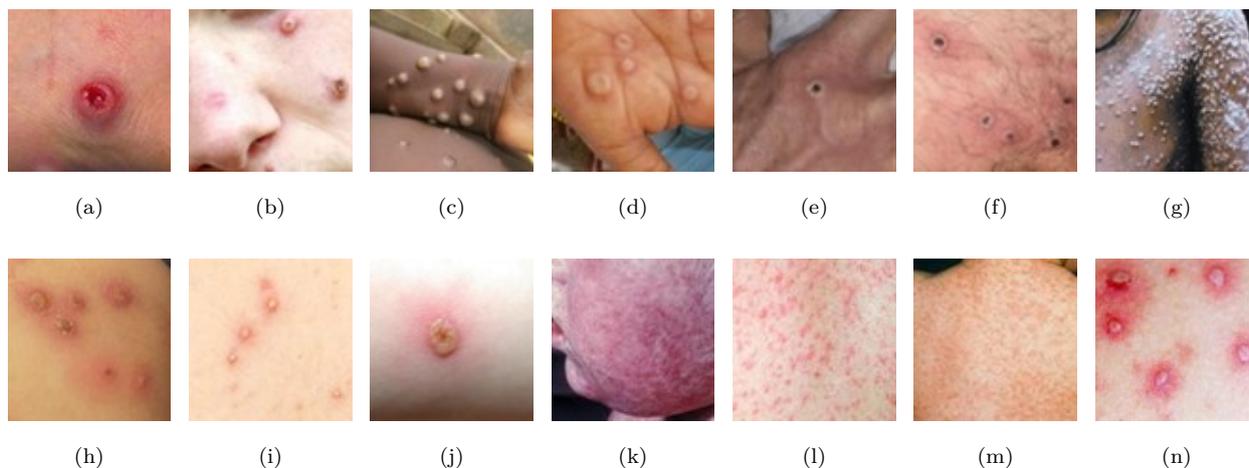


Figure 1: 数据集的示例。图像 a-g: 类别 “MonkeyPox”。图像 h-n: 类别 “Others”。

总共对原始数据集进行了十四倍增强，使得每个类别中的图像数量增加到“猴痘”类别中的 1,428 张图像和“其他”类别中的 1,764 张图像。Figure 2 展示了一些增强后的图像，这些图像被存储在增强图像文件夹中。

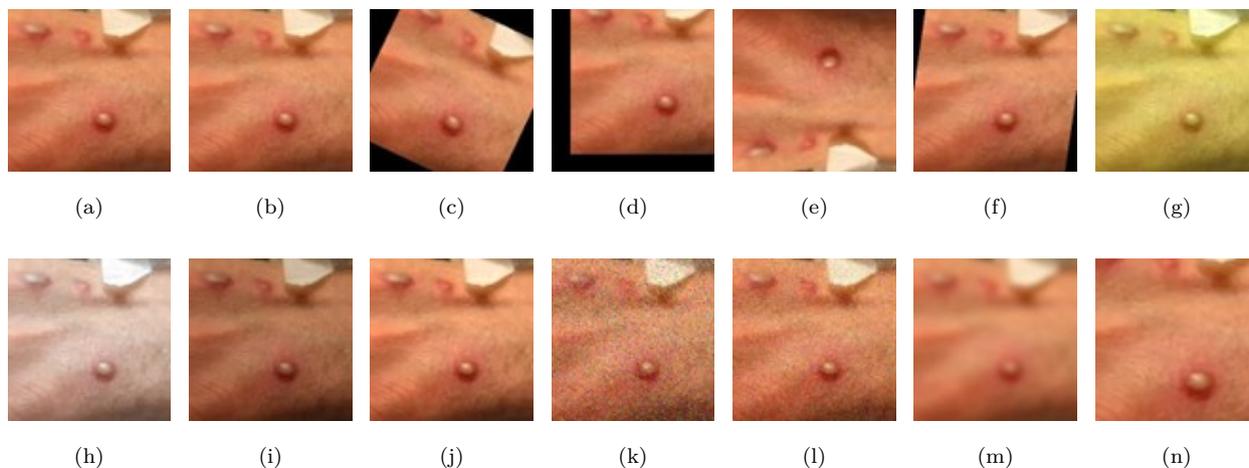


Figure 2: 从“猴痘”类别中一张图像的十四倍增强的示例。图像 a: 原始图像。图像 b-n: 增强后的图像。

数据集包括原始和增强图像（训练、测试、验证），比例为 70:10:20。我们在训练和验证阶段使用了来自增强训练文件夹的数据，其中包含 2,142 张图像。数据集的详细分布在 Table 1 中提供。

Table 1: MSLD 的分布

Class Label	No. of Original Images	No. of Unique Patients	No. of Augmented Images
Monkeypox	102	55	1428
Others	126	107	1764
Total	228	162	3192

4. 方法论

用于区分猴痘和非猴痘病例的所提模型的主要思想是迁移学习，最初由 Stevo Bozinovski 和 Ante Fulgosi 引入用于神经网络 (Bozinovski, 2020)。

该过程从将猴痘图像数据集加载到工作环境中开始。接下来，执行了数据预处理步骤，包括使用标签编码器对类别标签进行编码，并对图像 (Li et al., 2020) 进行归一化处理。结果是，提供给网络的输入数据冗余较少。

在下一步中，为了进行特征提取，图像被输入到通过 Keras 开源库实现的预训练特征提取模型中。深度特征迁移学习涉及使用在大规模数据集（如包含数百类的 ImageNet）上训练的神经网络来预测其他数据集 (Deng et al., 2009) 的标签。所提出的方法通过移除通常基于卷积提取特征对图像进行分类的全连接层，并用一个备用分类器替换它们来实现。

提取的特征随后被转移到另一个分类器进行训练。这个附加的分类器可以是一个全连接层或者一个外部的分类算法。

所提出的方法包含几个关键步骤：

- 数据预处理
- 模型开发
- 训练和测试模型（使用交叉验证）
- 超参数设置

4.1. 数据预处理

数据预处理步骤包括特征缩放和数据调整大小，如下所示：

第一步涉及导出数据集以便操作和准备数据。这一步确保了数据能够被各种机器学习模型有效利用，以准确检测猴痘病例。数据预处理具体包括调整图像大小、执行特征缩放以及从图像中提取标签等操作。

首先使用 OpenCV 方法从图像中提取标签，然后将图像缩小到 285×285 像素。随后，对提取的标签进行编码以使其与分类器兼容，并将图像转换为 NumPy 数组以便后续使用。在此过程中，特征缩放对于提高模型性能至关重要。

特征缩放，特别是归一化，通过处理自变量将输入数据标准化到特定范围。各种归一化技术的存在，包括最小-最大、Z-分数和小数缩放。本研究采用最小-最大归一化，这是一种通过线性变换来保持原始数据关系的方法。归一化过程包括将每个像素值除以最大可能值 (255)，以将数据缩放到 $[0,1]$ 的范围 (Pei and Lin, 1995)。

因此，归一化后的值用符号 x' 表示。

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

，其中 x 是图像 (Bala et al., 2023) 的原始强度。

4.2. 模型开发

模型开发阶段包括预训练的深度学习模型、降维方法、一些机器学习分类器，以及所提出的模型。

4.2.1. 深度学习模型

特征提取是将初始数据转换为可用信息的过程。特征提取的主要目标是减少需要处理的数据量，同时有效地表征初始数据。在图像处理过程中，特征提取器有助于发现形状、边缘和运动等特性。特征提取在不丢失关键信息的情况下减少处理资源需求时，具有显著优势。模型或模型的一个子集可以对输入进行预处理，为每个输入图像生成一个输出（即矢量），此输出可以随后用作新模型 (Khalid et al., 2014) 的训练数据。

在本研究中，采用了十三种深度学习预训练模型进行特征提取：Xception、ResNet50V2、ResNet101V2、ResNet152V2、InceptionV3、InceptionResNetV2、MobileNet、MobileNetV2、DenseNet121、DenseNet169、DenseNet201、NASNetMobile 和 NASNetLarge。此过程中配置了两个重要参数：池化类型（是否包含或排除最后一层）和输入形状，所有模型的输入形状均标准化为 285×285 。下面简要介绍 Xception 模型。

4.2.1.1. Xception

Xception 是一种深度卷积神经网络架构，由 (Chollet, 2016) 提出，作为“深度可分离的 Inception 变体”。在该架构中，逐通道应用的 $n \times n$ 空间卷积被称为“深度卷积”。

Xception 背后的核心思想是假设卷积神经网络中的跨通道相关性和空间相关性可以完全解耦。为了探讨这一假设，Xception 采用了深度可分离卷积，这涉及两个关键步骤：首先，对每个输入通道独立应用深度空间卷积，然后点卷积将输出通道投射到一个新的通道空间。

Xception 架构总共包括 71 层，其中 36 层卷积层被组织成 14 个模块。最初和最后的模块被设计用于逐渐调整维度，而中间的 12 个模块形成了核心特征提取的“中间流”。

具体而言，这种架构以一个包含三个模块的“入口流”开始，这些模块利用深度可分离卷积层来提取特征并降低空间分辨率。接下来是八个重复的中间流模块，每个模块配备有围绕深度可分离卷积的线性残差连接，以确保有效的信息传播。这些中间流模块逐步提取越来越复杂的特征。最后，该架构以一个包含可分离卷积和全局平均池化用于分类的“出口流”模块结束。

在 14 个模块中，除第一个和最后一个之外的所有模块都包含残差连接。这些线性残差连接将深度可分离卷积封装在每个中间流模块中，以增强训练期间的梯度流动。

总共来看，14 个模块中的 36 个卷积层包含了 2290 万个可学习的参数。在 ImageNet (一组包含超过 1400 万张图片 and 1000 类的数据库) 上的评测中，Xception 取得了 79 % 的 top-1 准确率以及 94.5 % 的 top-5 准确率。这一表现略微超过了 Inception V3，其参数数量相当，但特意为 ImageNet 分类进行了优化。Xception 性能的提升表明，它在跨通道和空间相关性上的解耦能够更有效地利用参数 (Chollet, 2016)。

4.2.2. 机器学习分类器

在本研究中使用了四种不同的机器学习分类器，以各种模型结构来基于深度神经网络提取的深度特征对猴痘疾病进行分类。下面简要描述这些分类器。

4.2.2.1. LightGBM

梯度提升决策树 (GBDT) 是一种可靠的机器学习模型，通过迭代训练弱分类器 (决策树) 来优化性能并最小化过拟合 (Zhou and Feng, 2019)。在 2017 年，来自微软和北京大

学的研究人员推出了 LightGBM，以解决现有方法（如 GBDT 和 XGBoost）在效率和可扩展性方面的挑战，特别是针对大规模数据集和高维特征 (Li et al., 2023)。

LightGBM 是一个基于 GBDT 的时间高效、分布式提升框架，提供更快训练速度和更低的内存占用 (Sharma and Singh, 2020)。它采用了一种叶节点生长策略（见 Figure 3），其重点是对具有较高梯度的叶节点进行分裂，这与其他 GBDT 工具中对同一层级的所有叶节点一视同仁的层级增长策略相反。这种创新方法最大限度地减少了不必要的计算，并允许生成更复杂的树 (Fu et al., 2023; Li et al., 2022)。

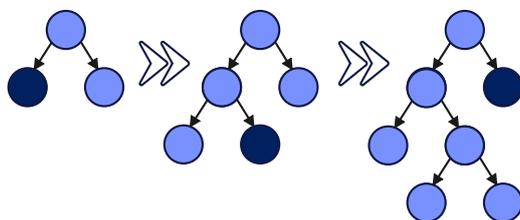


Figure 3: LightGBM 叶节点树增长 (Abbasniya et al., 2022)。

此外，LightGBM 采用了先进技术，如基于梯度的一侧采样 (GOSS) 和独占特征捆绑 (EFB)。GOSS 通过关注梯度更大的样本来减少样本数量，这些样本对信息增益的影响更显著。EFB 通过捆绑互相独占的特征进一步提高效率，有效处理高维数据 (Li et al., 2022)。这些特征与基于直方图的数据处理算法结合，使得在较短时间内实现平滑和准确的数据扫描、聚类 and 分类成为可能。

4.2.2.2. 支持向量机

支持向量机 (SVM) 由 Vapnik 于 1995 年提出，是一种监督学习算法，以其在处理稀疏、噪声和高维数据方面的鲁棒性而闻名 (Cortes and Vapnik, 1995; Furey et al., 2000)。这种方法基于 VC 维理论和结构风险最小化原则，关注于最小化实际风险而非经验误差 (Jing and Hou, 2015)。

SVM 的核心概念是识别一个最优的决策边界或超平面，该超平面在分隔两个数据类别的同时最大化它们之间的间隔 (Ahmadipour et al., 2022; Chiu et al., 2020)。

SVM 的核心概念是识别一个最优的决策边界，或称超平面，以分隔两个类别的数据，

并最大化它们之间的边距。对于一个数据集 $\{x_i, y_i \mid i = 1, 2, \dots, n\}$ ，其中 $x_i \in R^d$ 表示输入样本， $y_i \in \{-1, 1\}$ 是相应的标签，超平面由 $w^T x_i + b = 0$ 定义，其中 w 是权重向量， b 是偏置项。优化问题旨在最小化 $\{\frac{1}{2}\|w\|^2 + C \sum_{i=1}^n \xi_i\}$ ，其中 C 是惩罚参数， ξ_i 是松弛变量，测量错误分类点到超平面 (Chiu et al., 2020) 的距离。

SVM 最初是为线性可分问题设计的，但也可以使用核方法解决非线性情形。核将数据从低维非线性空间映射到高维特征空间，在那里可以实现线性分离。常见的核包括线性、多项式、高斯径向基 (RBF) 和 sigmoid 函数。本研究使用的 RBF 核对于许多实际应用是有效的。它定义为 $K(x, x_i) = e^{(-\gamma\|x-x_i\|^2)}$ ，其中 $\gamma = \frac{1}{2\sigma^2}$ ，并且 σ 是核宽度参数。

以下函数是此处使用的决策函数

$$f(x) = \text{sign} \left(\sum_{i=1}^n \lambda_i y_i K(x, x_i) + b \right) \quad (2)$$

。SVM 的决策函数使用核来计算 Equation 2。 $f(x)$ 在核函数 K 中通过新的查询向量 x 计算每个支持向量的 x 值。在方程中， b 表示沿法向量的超平面的偏移量，并通过 SVM 训练获得。 λ_i 表示拉格朗日乘子，其遵循 $0 \leq \lambda_i \leq C$ (Chiu et al., 2020)。通过结合核，SVM 有效地解决了“维度灾难”并实现了在复杂特征空间中的稳健分类。

4.2.2.3. XGBoost

极端梯度提升 (XGBoost) 由 Chen 和 Guestrin 于 2016 年提出，是梯度提升算法 (Chen and Guestrin, 2016) 的一个高度可扩展的变体。XGBoost 与其他提升决策树之间最显著的区别在于它的可扩展性。此外，尽管其快速学习能力增加了过拟合的风险，XGBoost 的正则化技术有效地防止了过拟合，使其与其他梯度提升算法 (Abbasniya et al., 2022) 区分开来。由于这些特性，XGBoost 最近广泛被数据科学家采用，并在机器学习竞赛中取得了出色的成绩。

XGBoost 背后的关键理念是通过顺序地添加新的弱学习器 (决策树) 来建立一个加法模型，以最大化目标函数 (Ghaehri et al., 2024) 的改进。在每一次迭代中，构造一个新的树来预测模型上一轮迭代预测的残差 (误差)。这类似于在函数空间中进行梯度下降，其中梯度由最新的决策树 (Patnaik et al., 2021) 近似。

XGBoost 的细节如下所述。

给定的样本集包含 n 个样本和 m 个特征，可以表达为 $D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}^n)$ ，其中 x 是特征值， y 是真值 (Jiao et al., 2021)。树提升模型输出 \hat{y}_i ，具有 K 棵树，定义如下：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (3)$$

，其中 F 表示决策树的空间，定义为：

$$F = \{f(x) = \omega_q(x)\} (q: \mathbb{R}^m \rightarrow T, \omega \in \mathbb{R}^T) \quad (4)$$

每个 $f(x)$ 将树分为结构部分 q 和叶权重部分 ω 。这里 T 表示叶节点的数量。需要最小化的目标函数由两项组成：训练损失 l ，用于衡量预测值 \hat{y}_i 和真值 y_i 之间的差异，以及用于惩罚模型复杂度的正则化项 Ω ：

$$O = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (5)$$

正则化项表示为：

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\| \quad (6)$$

其中 γ 惩罚叶子数量 (T)， λ 是叶权重 ω 的惩罚系数，通常为常数。在训练过程中，将添加新树以拟合上一轮的残差 (Jiao et al., 2021)。因此，当模型具有 t 棵树时，表示如下：

$$\hat{y}^t = \hat{y}^{t-1} + f_t(x) \quad (7)$$

将 Equation 7 代入目标函数 Equation 5 得到函数

$$O^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (8)$$

XGBoost 通过利用二阶泰勒展开来近似 Equation 8。对于每个实例，计算损失函数的一阶梯度 g_i 和二阶梯度 h_i ，从而能够精确地更新模型：

$$O^{(t)} \approx \tilde{O}^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (9)$$

为了确定树的最佳结构，算法评估每个叶子的权重。对于叶子 j ，最佳权重计算为：

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (10)$$

其中 I_j 表示叶子 j 中的实例集合。将其代入得分函数，树的最终目标变为：定义 ?? 为评分函数，以评估树结构 $q(x)$ ，并找到用于分类的最优树结构。

4.2.2.4. NGBoost

在提升算法领域，若干现有方法已被证明在提高预测准确性方面有效。然而，在处理概率预测和量化不确定性方面存在显著差距。像 AdaBoost、梯度提升和 XGBoost 等方法通常专注于减少预测误差，并常常忽视估计预测的不确定性。它们主要能够根据预测变量 (Duan et al., 2020) 估计响应变量的条件均值。这一疏漏可能会削弱模型量化不确定性以及在需要精确评估不确定性的情况下（例如医学诊断或金融风险分析中）实现准确概率预测的能力。因此，这一差距突出了需求，一种不仅提高预测准确性而且专注于提供良好校准的概率估计和可信的不确定性测度的复杂提升技术。

一个合适的解决方案候选是名为自然梯度提升 (NGBoost) 的一种创新方法。NGBoost 可以通过使用概率预测（包括实际的点预测）(Jabeur et al., 2022)，利用梯度提升技术来计算统计不确定性。此外，该算法的概率预测功能使得获取关于任何给定搜索区域的所有可能结果的概率成为可能，而不是确定一个特定的输出作为预测 (Kavzoglu and Teke, 2022)。

事实上，NGBoost 将提升过程转化为一个概率框架，使用自然梯度来弥补现有梯度提升方法 (Dutta, 2020) 的概率预测限制。

在点预测设置中，关注的对象是一个标量函数的估计值，比如 $E[y | x]$ ，其中 x 是特征向量， y 是预测目标。相反，NGBoost 算法使用条件概率分布函数 $P_\theta(y | x)$ 的概率预测，该函数是通过预测目标分布 θ 的参数生成的。NGBoost 假设 $P_\theta(y | x)$ 是一个指定的参数形式，然后在适当的评分规则 $S(\theta, y)$ 下，将分布的 p 参数 $\theta \in \mathbb{R}^p$ 估计为 x 的函数 (基学习器)，并拟合自然梯度 ∇_θ (Chen and Geyer, 2022; Duan et al., 2020)。

该模块化和可扩展算法由三个主要组件组成：(1) 基本学习器（例如，决策树等）；(2) 参数化概率分布（Bernoulli, 正态, 拉普拉斯分布等）；(3) 合适的评分规则（MLE, CRPS 等）(Peng and Cheng, 2020)。NGBoost 模型的概念表示在 Figure 4 中展示。

在 Figure 4 中， x_i 代表输入特征， $f^{(m)}$ 是基本学习器， y_i 是目标标签， θ 代表参数化概率密度函数， M 是 Boosting 迭代次数 (Hussain et al., 2021)。

NGBoost 算法的工作过程如下：

1. 计算初始的 $\theta_0 \leftarrow \arg \min_\theta \sum_{i=1}^n S(\theta, y_i)$ ，强调 θ 可以是参数的向量。
2. 执行 M 个梯度提升阶段 (Kavzoglu and Teke, 2022)。

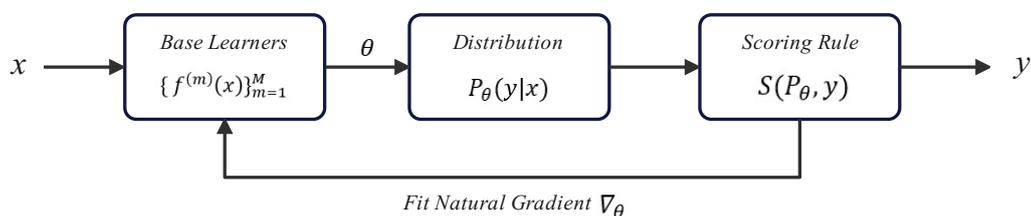


Figure 4: NGBoost 在基学习器、分布和评分规则 (Duan et al., 2020) 的选择上是模块化的。

与传统的提升算法相比，该算法可以扩展以处理大型变量或观测数据，并且具有合理的复杂性 (Jabeur et al., 2022)。此外，应注意，根据我们所知，这是首篇关于应用 NGBoost 算法诊断猴痘的论文。

NGBoost 提供了一个灵活的框架，可以在训练 (Chen and Geyer, 2022) 之前预先自定义基础学习器、参数化概率分布和评分规则作为超参数。为了获得满意的结果，除了这三个组件之外，我们还使用了一种元启发式算法，根据训练集调整超参数，该算法通过启发式搜索过程高效地探索参数空间，以找到最优的参数集。

4.2.3. 超参数优化

超参数被定义为模型中的一个更高层次的概念，例如其复杂性或可学习性。超参数是在学习过程之前设定的参数，而不是通过训练获得的参数。通常，在机器学习的过程中，需要对这些参数进行优化，并为学习者选择一组最优的超参数，以增强学习的性能和效果 (Dong et al., 2020)。然而，超参数调优是主观的，并依赖于试错方法和经验判断 (Bergstra et al., 2011)。超参数优化算法克服了对人工搜索、经验和试错的依赖 (Jiao et al., 2021)。

随着各种非线性问题的日益复杂化，迫切需要基于自然现象和生物行为进一步开发元启发式优化算法，以有效解决这些复杂的搜索和优化问题。

超参数优化过程由四个基本组成部分构成：具有目标函数的估计器、搜索空间（配置区域），以及用于定位超参数配置的搜索或优化方案。

超参数配置的主要步骤如下：

在模型的设计阶段，可以通过使用优化算法有效搜索超参数空间来确定模板的最佳超参数。提供更好的参数优化过程包括关键要素，例如预测器的目标和结果，需要一个空间

来理解数据并优化组合，以及应用各种评价函数以便对结果进行比较 (Jiao et al., 2021)。

本节概述了用于混合机器学习模型的元启发式方法。

4.2.3.1. 非洲秃鹰优化算法 (AVOA)

一种新的基于群体的元启发式优化算法 AVOA，其灵感来源于环境，模仿主要生活在非洲的一种秃鹰的觅食和捕食行为。此外，与其他元启发式算法相比，该算法更具适应性且计算复杂度更低 (Abdollahzadeh et al., 2021)。

AVOA 算法原理可以总结为以下四个主要阶段：

第一阶段：最佳秃鹰组的确定

AVOA 种群由 N 只秃鹰组成， N 根据问题的需求可调整。初始种群是随机生成的，并对所有个体进行适应度评估。在每次迭代中，解决方案被分组如下：表现最好的解决方案形成第一优秀秃鹰组，而次优秀的解决方案形成第二优秀秃鹰组。其余的秃鹰使用基于选择概率的轮盘法在这两个组之间分配，选择概率由以下公式计算：

$$p_i = \frac{F_i}{\sum_{i=1}^n F_i} \quad (11)$$

，其中 F_i 是第 i^{th} 只秃鹰的适应度值。然后，将种群中的最佳秃鹰分配为：

$$R(i) = \begin{cases} BestVulture_1, & \text{if } p_i = L_1 \\ BestVulture_2, & \text{if } p_i = L_2 \end{cases} \quad (12)$$

。参数 L_1 和 L_2 是预设的，范围从 0 到 1，其中 $L_1 + L_2 = 1$ 。更高的 L_1 值增加了集中的力度，而更高的 L_2 值增强了多样化。

第二阶段：秃鹰的饥饿率

这一阶段决定了秃鹰的饥饿率。当过度进食时，秃鹰会探索更远的地方，但当饥饿时，它们会留在附近寻找食物。这种行为被建模为：

$$F = (2 \cdot \text{rand}_1 + 1) \cdot z \cdot \left(1 - \frac{\text{iteration}_i}{\text{max iterations}}\right) + t \quad (13)$$

。Equation 13 计算饱和率，它用于把 AVOA 从探索阶段过渡到开发阶段。

$$t = h \left(\sin^w \left(\frac{\pi}{2} \times \frac{\text{iteration}_i}{\text{max iterations}} \right) + \cos \left(\frac{\pi}{2} \times \frac{\text{iteration}_i}{\text{max iterations}} \right) - 1 \right) \quad (14)$$

, 其中 $iteration_i$ 是当前迭代次数, z 是介于-1 和 1 之间的随机值, h 是介于-2 和 2 之间的随机数。如果 z 为正数, 秃鹫则是满足的; 如果为负数, 则是饥饿的。参数 w 影响探索和开发阶段之间的过渡。当 $|F| > 1$ 时, 秃鹫处于探索阶段。

第三阶段: 探索

在此阶段, 具备高视觉能力的秃鹫会寻找食物并预测即将死去的猎物。探索策略依赖于参数 P_1 , 该参数控制搜索方法。采用了两种技术:

1. 聚焦搜索: 围绕群体中最好的秃鹫, 鼓励局部探索。这一策略使得在当前最佳解决方案附近进行深入调查成为可能。
2. 全局搜索: 在预定义边界内的整个搜索空间中进行搜索。

秃鹫在下一次迭代中的位置使用以下公式更新:

$$P(i+1) = R(i) - D(i) \cdot F \quad \text{if } P_1 \geq \text{rand}_{P_1} \quad (15)$$

$$P(i+1) = R(i) - F + \text{rand}_2 \cdot ((ub - lb) \cdot \text{rand}_3 + lb) \quad \text{if } P_1 < \text{rand}_{P_1} \quad (16)$$

其中 P_1 是一个预设的搜索参数, 介于 0 和 1 之间, 用于控制搜索策略; $R(i)$ 是由 Equation 12 在当前迭代中获得的最佳秃鹫; $D(i) = |X \cdot R(i) - P(i)|$, X 是用于随机运动的系数向量; F 是使用 Equation 13 获得的饥饿率; $P(i+1)$ 是秃鹫在下一次迭代中的位置; ub 和 lb 是搜索空间的上限和下限。

第四阶段: 开发利用

如果 $|F| < 1$, AVOA 进入开发阶段, 该阶段有两个阶段, 每个阶段涉及两种策略。每种策略的选择取决于 P_2 和 P_3 , 范围均在 0 和 1 之间。

阶段 1: 旋转飞行与围攻战斗

当 $0.5 \leq |F| < 1$ 时, 秃鹫会执行两种战术: 旋转飞行和围攻。选择哪种战术是通过生成一个随机数 rand_{P_2} 并将其与 P_2 进行比较来确定的。以下方程表示这些策略:

- 围攻战斗:

$$\begin{aligned} P(i+1) &= D(i) \times (F + \text{rand}_4) - d(t) \quad P_2 \geq \text{rand}_{P_2} \\ d(t) &= R(i) - P(i) \end{aligned} \quad (17)$$

- 旋转飞行:

$$\begin{aligned}
P(i+1) &= R(i) - (S_1 + S_2) \\
S_1 &= R(i) \times \left(\frac{\text{rand}_5 \times P(i)}{2\pi} \right) \times \cos(P(i)) \quad P_2 < \text{rand}_{P_2} \\
S_2 &= R(i) \times \left(\frac{\text{rand}_6 \times P(i)}{2\pi} \right) \times \sin(P(i))
\end{aligned} \tag{18}$$

其中 rand_4 、 rand_5 和 rand_6 是介于 0 和 1 之间的随机数，用于增加随机性。 $\cos(P(i))$ 或 $\sin(P(i))$ 的值是大小为 $n \times 1$ 的数组向量，其中 n 表示发电单元的数量。

阶段 2: 激进围攻与徙动飞行

当 $|F| < 0.5$ 时，第二阶段的利用开始，涉及更具攻击性的围攻战术。策略的选择使用 P_3 ，如下所示：

- 激进围攻

$$\begin{aligned}
P(i+1) &= \frac{A_1 + A_2}{2}, \quad P_3 \geq \text{rand}_{P_3} \\
A_1 &= \text{BestVulture}_1(i) - \frac{\text{BestVulture}_1(i) \times P(i)}{\text{BestVulture}_1(i) \times P(i)^2} \times F \\
A_2 &= \text{BestVulture}_2(i) - \frac{\text{BestVulture}_2(i) \times P(i)}{\text{BestVulture}_2(i) \times P(i)^2} \times F
\end{aligned} \tag{19}$$

- 徙动飞行策略：

$$P(i+1) = R(i) - |d(t)| \times F \times \text{Levy}(d), \quad P_3 < \text{rand}_{P_3} \tag{20}$$

其中， $\text{BestVulture}_1(i)$ 和 $\text{BestVulture}_2(i)$ 分别是当前迭代中的最佳和次佳秃鹰， $P(i)$ 是当前秃鹰位置的向量。 $d(t)$ 、 $R(i)$ 和 F 如之前所定义，而 $\text{Levy}(d)$ 代表用于提高 AVOA 效率的 Le'vy 飞行，其数学表达式为：(LF)。

$$\begin{aligned}
LF(x) &= 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}} \\
\sigma &= \left(\frac{\Gamma(1 + \beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma(1 + \beta/2) \times \beta \times 2 \left(\frac{\beta-1}{2}\right)} \right)^{\frac{1}{\beta}}
\end{aligned} \tag{21}$$

其中 u 和 v 都是介于 0 和 1 之间的随机数， β 表示固定常数 1.5 (Zhou and Xu, 2023)。

4.3. 拟议模型

在本节中，我们提出了一种模型，用于从皮肤病变图像中分类猴痘病例。我们的模型无缝地集成了迁移学习、降维和先进的机器学习技术，以达到最先进的成果。以下步骤描述了我们提出的模型：

在我们模型的初始阶段，我们专注于准备数据集，以确保其适合训练。这涉及两个关键的预处理步骤：标签编码和归一化。标签编码用于将类别数值化，以确保与机器学习算法的兼容性。此外，我们应用了图像归一化，将像素值缩放到 $[0, 1]$ 的范围内。这些预处理技术不仅提高了计算效率，还改善了模型在数据中辨别有意义特征的能力。

我们的模型基础在于使用 Xception 深度神经网络，该网络以 ImageNet 数据集的预训练权重进行初始化。我们将 Xception 配置为输入形状为 (285×285) ，并排除顶层以用于我们的特定任务。关键是，我们将网络的层设置为不可训练，保留网络中编码的预训练知识。

经过 Xception 的特征提取后，我们获得了皮肤病变图像的高维特征表示。这一步使我们能够提取 2048 个特征，捕捉到复杂的模式和细节，这对于准确分类至关重要。

为了减轻过拟合的风险并提高计算效率，我们采用主成分分析 (PCA) 将特征空间缩减至更易管理的 530 个特征。PCA 是一种降维技术，用于解决高度相关特征和“维度灾难”的挑战，后者是指特征数量常常超过数据集中实例数量的问题。通过将原始高维数据转换为一组较少的无相关变量，PCA 在消除冗余的同时保留了最重要的信息。这样的降维提高了计算效率，减轻了过拟合的风险，并简化了特征空间，使其更易于分类器处理。作为我们方法的一部分，PCA 帮助聚焦于数据中最具信息量的方面，提高了模型性能并确保更快的预测。

在我们提出的模型的最后阶段，使用 NGBoost 算法对猴痘和非猴痘病例进行最终分类。我们的 NGBoost 模型通过非洲秃鹫优化算法进行了微调，以确保最佳性能。

Figure 5 说明了用于分类和预测猴痘病的提出的过程和方法的步骤。

5. 评估指标

整体实验结果采用广泛采用的统计方法进行测量和展示，包括准确率、精确率、召回率、特异性和 F1 分数。这些定量测量根据其对分类任务的有效性以及在密切相关研究中的频繁使用而被选择。

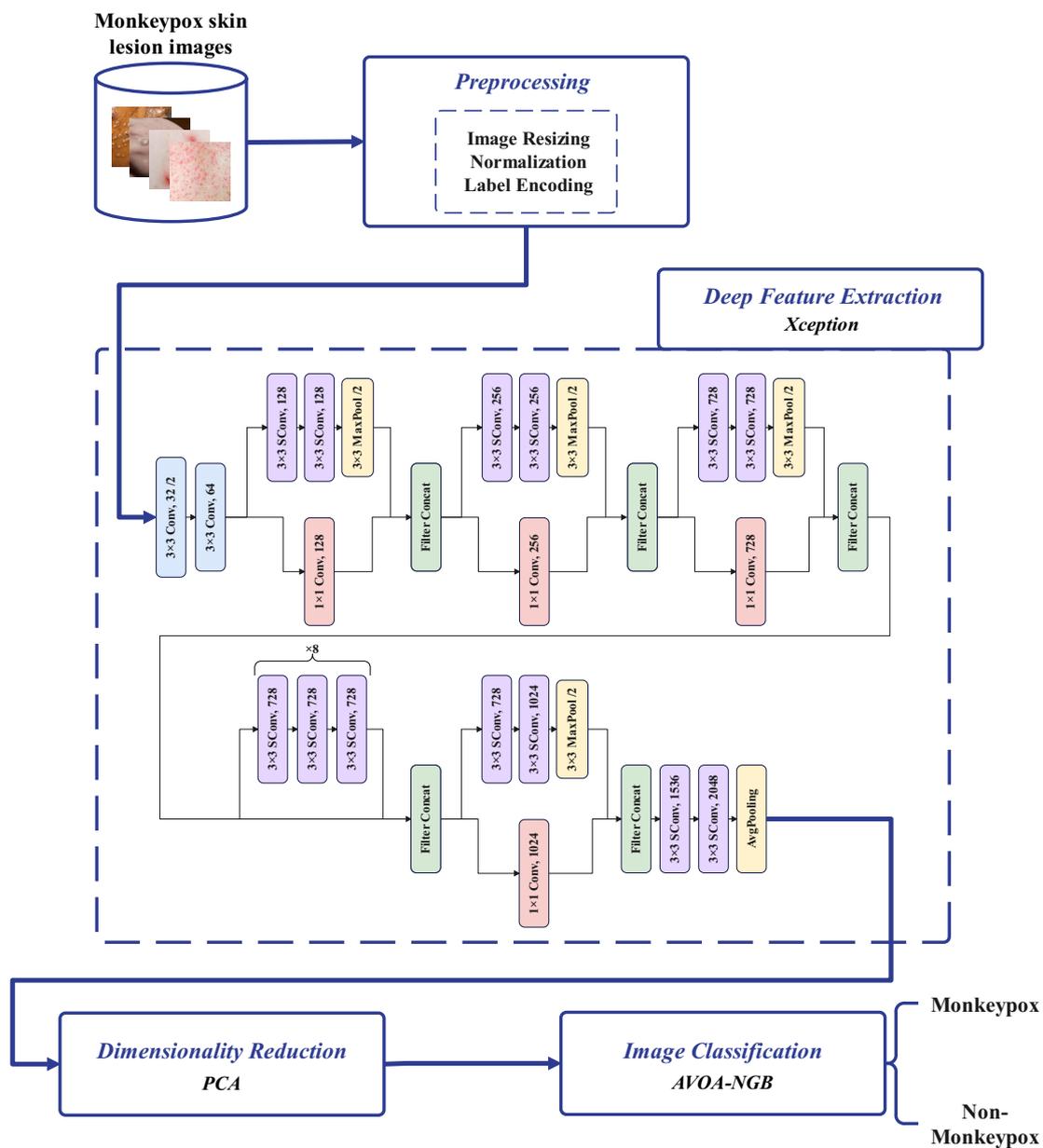


Figure 5: 提出的方法论架构。

模型准确分类测试数据集中的图像的程度称为其“准确率”。

“精准度”一词指的是能够准确地估计两种类别中的实际正向结果的能力。

“召回率”表示在两类中都为真正阳性的案例所占的比例。当假阴性与高成本相关联时，召回率是用于选择最佳模型的主要统计量。“特异性”识别并计算出正确识别的真正阴性的数量。

“F1-score”是一种用来展示精确度和召回率之间关系的指标。它在机器学习和信息检索中常用于评估分类模型的性能。F1-score 是精确度和召回率的调和平均数，对两个度量给予相同的权重。一个高的 F1-score 表明模型在获得高精度和高召回率两方面都表现良好，意味着它能够正确识别出大多数相关的案例，同时最大限度地减少误报。

性能比较如果仅基于准确性指标，可能会导致误导性的结果。这种评估还必须考虑错误的成本。在这种情况下，我们使用 Cohen 的 kappa 指标来评估分类器的可靠性，该指标考虑了实际和预测类别值之间的协议在多大程度上可以用随机性来解释。通常，kappa 值在-1 到 1 之间。当分类器的计算 kappa 值为 1 时，分类器的性能被认为与真实值完全一致。

在我们的数据集中，如果个体被正确诊断，猴痘可以被分类为真正例（TP）或真负例（TN）；如果个体被误诊，则分类为假阳性（FP）或假阴性（FN）。本研究使用的性能评估标准公式如下：

$$\text{Accuracy} = \frac{TP + TN}{TN + TP + FP + FN} \quad (22)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (23)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (24)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (25)$$

$$F_1\text{-score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (26)$$

$$\left\{ \begin{array}{l} P_o = AC = \frac{TP+TN}{TP+TN+FP+FN} \\ P_e = \frac{(TP+FP)(TP+FN)+(TN+FP)(TN+FN)}{(TP+TN+FP+FN)^2} \\ \kappa = \frac{P_o - P_e}{1 - P_e} \end{array} \right. \quad (27)$$

其中， P_o 表示总体一致概率， P_e 表示由于偶然的一致概率。

6. 实验结果

在本节中，我们展示并分析了我们提出的方法在使用猴痘皮肤病变数据集 (MSLD) 进行猴痘检测方面的性能。

由于收集的数据集规模较小，采用了交叉验证技术以确保所提方法的效能。作为 k 折交叉验证策略的一部分，整个数据集被随机分成 k 个大小相同的子文件夹。随后，使用 $(k-1)$ 个训练数据子文件夹对学习函数进行训练，而剩下的一个用于测试模型。每个独立的数据子集都被用于训练或验证。在依次遍历所有数据子集后，计算出了一个累积的平均验证得分。我们采用了 5 折交叉验证技术，其中四折 (80 % 的数据) 用于训练，剩下的一折 (20 % 的数据) 用于评估模型的测试准确性。

在初始阶段，使用了 13 个预训练的深度学习模型从输入图像中提取特征。这些特征随后被提供给四个不同的机器学习分类器：LightGBM、SVM、XGBoost 和 NGBoost，以预测最终标签。这些模型的性能结果展示在 Table 4 中，其中 L、S、X 和 N 分别代表 LightGBM、SVM、XGBoost 和 NGBoost。由 Table 4 中呈现的结果可见，使用 Xception 模型作为特征提取器在所有 52 个测试模型中达到了最高的准确率。因此，它被选为本研究的特征提取器。表中的粗体值表示分类器和特征提取器的最大指标。

进行了一项综合分析，以确定 PCA 的最佳方差比率。评估了四种分类器——LightGBM、SVM、XGBoost 和 NGBoost——在不同方差部分的性能。结果显示在 Figure 6 中，其中显示了分类性能分数随方差比率的变化。

如 Figure 6 所示，分类性能随着方差比率的变化而变化。值得注意的是，当方差比率设置为 0.95 (266 个特征) 和 0.97 (530 个特征) 时，SVM 和 NGBoost 均表现出优越的性能。这些最佳方差比率是基于达到最高分类准确度选择的。

Table 2 展示了有无降维情况下的分类结果。该表格清楚地表明，降维对于提高分类结果的准确性是必不可少的。

除了评估分类性能外，还分析了 PCA 对四个分类器训练时间的影响。Table 3 展示了应用 PCA 前后完整数据集的训练时间。

此外，如 ?? 所示，对不同阶段的数据分布进行了分析。原始数据集包含每个个体的 2048 个属性，代表 2048 个维度。为了可视化数据，采用 t 分布随机邻域嵌入 (t -SNE) 方法将这种高维数据投射到二维平面中。为了对两个类实现更好的聚类可视化，实验在数据

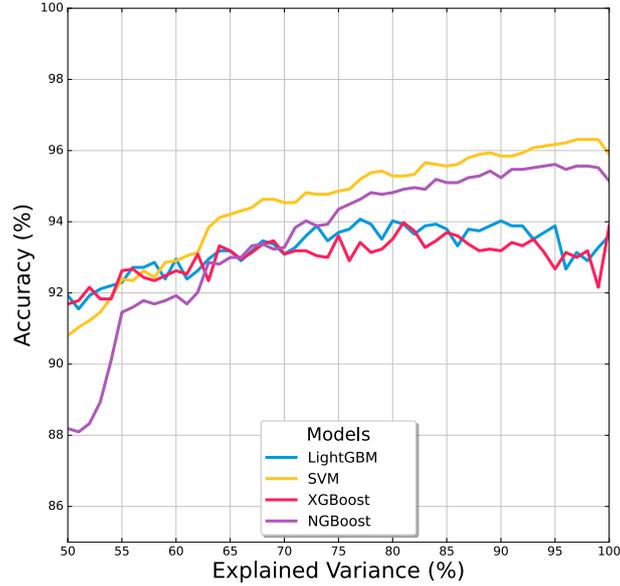


Figure 6: 分类性能作为 PCA 方差比的函数。

Table 2: 带有和不带有降维的分类结果比较。

Metric	Before Dimensionality Reduction				PCA (variance ratio = 0.95)				PCA (variance ratio = 0.97)			
	L	S	X	N	L	S	X	N	L	S	X	N
Accuracy (%)	93.60	95.89	93.51	95.15	93.00	96.17	93.60	95.61	93.18	96.31	93.42	95.57
Precision (%)	92.47	94.75	92.73	93.65	91.39	95.29	92.26	94.11	92.05	95.53	92.55	94.11
Recall (%)	96.11	97.88	95.60	97.70	96.17	97.78	96.34	98.04	95.73	97.78	95.54	97.95
Specificity (%)	90.79	93.63	91.20	92.20	89.35	94.35	90.53	92.83	90.26	94.67	90.99	92.82
Kappa	0.8705	0.9168	0.8688	0.9016	0.8581	0.9226	0.8706	0.9112	0.8620	0.9255	0.8668	0.9102
F1-Score (%)	94.22	96.27	94.10	95.61	93.70	96.50	94.21	96.02	93.83	96.62	94.01	95.67

Table 3: 分类器在进行主成分分析前后的训练时间。

Classifier	Training Time (sec.)		
	Before Dimensionality Reduction	PCA (number of features = 266)	PCA (number of features = 530)
LightGBM	16.36	3.31	6.28
SVM	3.36	0.64	0.89
XGBoost	28.37	7.06	9.57
NGBost	85.97	17.75	22.58

Table 4: 不同深度学习模型的分类结果。

Feature extractor	Classifier	Results					
		Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	Kappa	F1-Score (%)
Xception	L	93.60	92.47	96.11	90.79	87.05	94.22
	S	95.89	94.75	97.88	93.63	91.68	96.27
	X	93.51	92.73	95.60	91.20	86.88	94.10
	N	95.15	93.65	97.70	92.20	90.16	95.61
ResNet50V2	L	85.48	87.78	85.16	85.87	70.77	86.43
	S	87.63	93.85	82.74	93.55	75.34	87.92
	X	84.83	86.99	84.75	85.03	69.48	85.83
	N	83.94	84.08	87.41	80.31	67.61	85.52
ResNet101V2	L	94.91	94.26	96.49	93.11	89.71	95.35
	S	94.91	93.66	97.17	92.31	89.70	95.37
	X	94.68	94.08	96.24	92.96	89.25	95.12
	N	94.54	92.62	97.68	90.88	88.93	95.07
ResNet152V2	L	95.15	93.99	97.25	92.70	90.17	95.58
	S	95.33	94.09	97.51	92.78	90.54	95.76
	X	93.84	92.77	96.13	91.16	87.52	94.40
	N	93.93	91.45	97.93	89.23	87.68	94.57
InceptionV3	L	83.75	86.73	82.79	85.00	67.38	84.68
	S	86.88	93.50	81.57	93.22	73.85	87.11
	X	83.33	86.05	82.84	84.14	66.54	84.35
	N	81.23	80.74	86.52	75.47	62.08	83.34
InceptionResNetV2	L	84.03	86.32	83.98	84.16	67.86	85.10
	S	87.91	94.13	82.94	93.86	75.87	88.17
	X	84.45	86.93	84.07	84.90	68.71	85.46
	N	83.94	84.10	87.14	80.49	67.58	85.46
MobileNet	L	82.21	85.49	81.13	83.77	64.37	83.14
	S	84.36	91.72	78.38	91.73	68.94	84.41
	X	81.51	84.12	81.46	81.85	62.90	82.67
	N	79.83	80.04	84.62	74.75	59.32	82.00
MobileNetV2	L	82.12	84.87	81.65	82.74	64.06	83.21
	S	83.61	88.81	79.93	88.14	67.28	84.09
	X	81.79	83.68	82.65	80.88	63.32	83.13
	N	78.20	76.44	86.96	68.24	55.64	81.20
DenseNet121	L	93.70	92.43	96.18	90.77	87.25	94.25
	S	92.25	89.94	96.48	87.26	84.26	93.08

Continued on next page

集中的 300 个样本上进行，并选择欧几里得度量来衡量样本之间的相似性。

Table 4 – Continuation of Table 4

Feature extractor	Classifier	Results					
		Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	Kappa	F1-Score (%)
DenseNet121	X	93.37	92.14	95.95	90.34	86.58	93.98
	N	88.80	83.82	98.28	77.57	77.03	90.46
DenseNet169	L	84.73	88.93	82.15	87.90	69.43	85.37
	S	86.41	93.87	80.25	93.77	72.96	86.51
	X	83.99	87.44	82.43	85.99	67.89	84.80
	N	79.93	79.85	84.85	74.53	59.44	82.10
DenseNet201	L	93.84	93.28	95.39	92.06	87.57	94.31
	S	93.42	92.12	96.04	90.38	86.68	94.02
	X	93.51	93.13	94.95	91.80	86.89	94.01
	N	91.27	88.39	96.57	85.12	82.27	92.27
NASNetMobile	L	91.41	90.00	94.64	87.64	82.60	92.24
	S	91.88	89.86	95.76	87.38	83.55	92.68
	X	90.24	89.18	93.26	86.70	80.24	91.16
	N	90.71	87.11	97.24	83.11	81.10	91.86
NASNetLarge	L	83.99	86.70	83.26	84.93	67.82	84.93
	S	86.09	92.77	80.69	92.53	72.27	86.30
	X	83.99	86.19	83.95	84.12	67.79	85.03
	N	81.84	81.18	87.02	76.14	63.28	83.84

t-SNE 无法提供 Xception 网络 (??) 和 PCA 方法 (??) 之间的类特异簇比较，因为 PCA 是一种无监督的降维技术。然而，详细比较 ?? 和 ?? 可以发现 ?? 中的点更为集中。通过比较 ?? 和 ?? 的橙色和蓝色点，这些点代表数据中的两个类别，可以看出结合 PCA 和 Xception (??) 结果在组之间的分离更加明显。在 ?? 和 ?? 的坐标轴上也观察到显著差异。

除了通过观察蓝色和橙色点的位置来分析单个分布，还使用了标准差 (SD) 和 t 检验等统计方法来分析两组数据的值。t 检验用于确定两个独立组是否表现出具有统计显著性差异的现象，结果接近零表明两组之间没有差异。结果显示在 ?? 的每个面板的上角。?? 展示了具有 t 值为-8.806 的原始数据分析。两类别的 SD 值分别为 0.228 和 0.211，作为对照组。通过 Xception 模型处理的数据在 ?? 中展示，其中 t 值提高到了 1.991，两个类别的 SD 值减少到了 0.183 和 0.166。这些改进伴随着水平和垂直轴曲线的明显变化。在

?? 中最终处理的数据展示了 t 值为 2.126，且两类别的 SD 值分别为 2.059 和 2.082。尽管 SD 值没有显著改善，但 t 值显示出较大的进展，使得个别分布组变得大致分离。然后将从 Xception 提取的特征转移到分类器，以建立最终的决策边界，将数据分为两类。

使用标准差和 t 检验的统计分析突出了蓝色和橙色点的优缺点。t 检验结果确认 ?? 产生了最出色的结果。通过使用这些指标评估两组，实现了更清晰的区分，验证了组合方法的优越性。

AVOA-NGB 集成需要大量的超参数调优。为了解决这个问题，我们重点关注了最敏感的超参数。为了建立模型，应用了支持向量回归 (SVR) 作为基学习器。选择 SVR 是因为它具有优越的性能和较少需要调整的参数。只有两个敏感参数——学习率和估计器的数量——需要进行优化。使用 AVOA 优化的 NGBoost 超参数及其域范围详见 Table 5。

Table 5: 使用 AVOA 优化的 NGBoost 超参数及其域范围。

Hyperparameter	Description	Values	
		Domain Range	Optimal Value
learning_rate	Step size for shrinkage update.	$[1 \times 10^{-7}, \dots, 0.9]$	1.0921481×10^{-1}
n_estimator	Number of boosting iterations to perform.	$[3, \dots, 20]$	5
Base_learner	Base component of multiple classifier systems.	('SVR', 'DecisionTreeRegressor', 'Ridge')	SVR
Probability_distribution	Distribution for output type.	('k_categorical', 'Bernoulli')	Bernoulli

本文使用 AVOA 优化四种分类算法的超参数。Table 6 展示了所提算法的配置参数。这些参数包括种群大小、训练迭代次数、P1、P2、P3，最后是 alpha 和 gamma 的值，分别设置为 0.8 和 2.5。

如描述所述，NGBoost 分类器在数据集的 80 % 上进行了训练，并在剩下的 20 % 上进行了测试。训练阶段涉及对 SVR 基础学习器参数的调整和 NGBoost 参数的优化。

为了评估所提出技术的有效性，评估了准确性和计算时间。准确性比较是在三个算法变体之间进行的：(1) 基线 NGBoost 算法，(2) NGBoost 减少特征但没有超参数优化，(3) AVOA-NGB 减少特征并优化 l_rate 和 n_est。

如 Table 7 所示，所提出的 AVOA-NGB 方法结合降维技术和优化的超参数提供了最高的准确性。关于计算时间，基准算法由于处理整个特征集最慢。相比之下，AVOA-NGB 方法使用基于 PCA 的降维技术，消除了约 75 % 的冗余特征，将计算速度加快了近一个数量级。这种计算效率的显著提高表明在实时应用中实施该方法的可行性。

Table 6: 提出的控制参数值。

Parameter	Value
Population size	50
Number of iterations	40
γ	2.5
α	0.8
P_1	0.6
P_2	0.4
P_3	0.6

Table 7: 三种算法变体的准确性和计算时间比较。

Algorithm	Number of Features	Performance Metrics	
		Accuracy (%)	Computational Time (sec.)
Baseline NGBoost	Whole feature set (2048)	95.15	0.0054
NGBoost with PCA	530	95.57	0.0012
NGBoost with PCA and optimized NGBoost hyperparameters	530	97.53	0.0003

为了展示调优策略的性能，AVOA 被拿来与其他元启发式优化器进行比较，即哈里斯鹰优化（HHO）(Heidari et al., 2019)、混合改进鲸鱼优化算法（HI-WOA）(Tang et al., 2019)、改进的猩猩部队优化（mGTO）(You et al., 2023)、混合灰狼-鲸鱼优化算法（GWO-WOA）(Dhakhnamoorthy et al., 2023) 和嵌入 Q 学习的正弦余弦算法（QLESCA）(Hamad et al., 2022)。为了公平比较，所有竞争优化器使用相同的一组候选输入、预处理方法、训练和验证样本及预测样本。唯一的变量是优化器类型（例如，用 HHO-NGB 替换 AVOA-NGB 等）。基于测试结果，AVOA-NGB 表现出最佳性能。因此，可以得出结论，AVOA 具有优于其他方法的优化能力。Table 8 展示了实验中使用的其他竞争算法的设置。

Figure 8 展示了各种参数优化方法的平均适应度函数的收敛曲线。纵坐标代表达到的最大精度水平，横坐标表示迭代次数。图中显示，GWO-WOA 和 QLESCA 收敛较慢。HI-WOA 收敛速度更快，且优于 GWO-WOA 和 QLESCA。尽管 mGTO 比 HI-WOA 略高的精度，但 HI-WOA 更快达到收敛。AVOA 和 mGTO 在所有优化方法中表现出最快的收敛速度。GWO-WOA 提前进入收敛状态，这导致了其最终精度不理想，并未找到全

Table 8: 竞争算法的配置参数。

Algorithm	Parameter	Value
HHO	Population size	50
	Number of iterations	40
HIWOA	Population size	50
	Number of iterations	40
	Maximum iterations of each feedback	10
mGTO	Population size	50
	Number of iterations	40
	Probability of transition in exploration phase (P)	0.03
GWO_WOA	Population size	50
	Number of iterations	40
QLESCA	Population size	50
	Number of iterations	40
	Learning rate in Q-learning (α)	0.1
	Discount factor (γ)	0.9

局最优值。AVOA 在所有收敛性能中取得了最佳表现。此外，mGTO 在经过多次迭代后继续改善，这表明其优化机制促进了子群多样性，防止了局部粒子聚集，并增强了对局部最优解的搜索。

AVOA 与 NGBoost 的性能与其他分类器：LightGBM、SVM 和 XGBoost 进行了比较。每种模型都有影响其分类性能的参数。因此，使用 AVOA 来寻找分类器参数的最优值。使用 AVOA 优化后的超参数值及其域范围如 Table 9 所示。

如 Table 10 所示，所提出的方法在所有分类指标上表现优越，达到了准确率（97.53 %）、Kappa 值（0.95）、F1-Score（97.72 %）和 AUC（97.47 %）。AVOA-SVM 方法排名第二。相比之下，使用 PCA 的 LightGBM 的准确率（93.18 %）在本研究中所考察的分类器中是最低的。为了突出 AVOA 的重要性，展示了对同一输入数据而不使用 AVOA 的模型的结果。很明显，当不使用 AVOA 时，所有分类器的性能都有所下降。例如，未使用

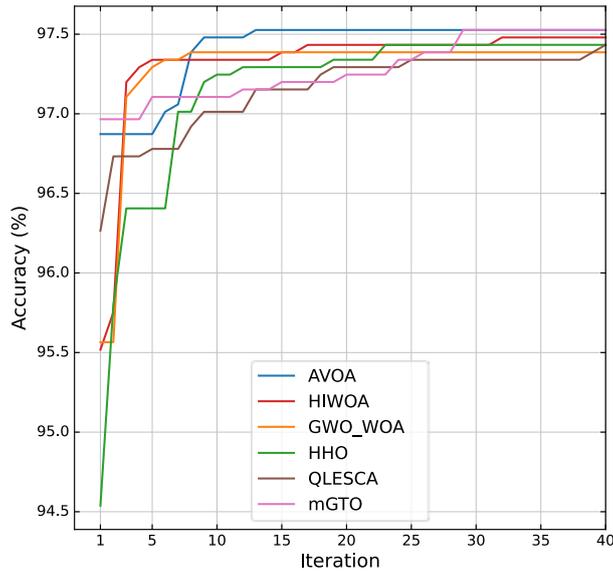


Figure 8: 参数优化方法的平均适应度函数的收敛曲线。

AVOA 的 SVM 的准确率 (95.89 %)、Cohen Kappa (0.9168) 和 F1-Score (96.27 %) 明显低于 AVOA 优化的 SVM。

由模型获得的不同折叠的结果，以及所有折叠的标准差和平均值，展示在 Table 11 中。

分析显示计算出的标准差较低，表明所提出的模型在所有折叠中表现一致，其中在第 3 和第 5 折叠中观察到优秀的结果。各折叠分数的低变动表明该模型具有鲁棒性，并且在任何单一折叠上没有表现出高方差或过拟合。该模型在数据的多个不同子集上的强劲表现，即不同折叠所代表的，支持其作为一个准确预测模型在该问题上的普适性和可靠性。

为了检查模型的模式，每个折叠都计算了一个混淆矩阵，然后叠加，如 Figure 9 所示。所有折叠的组合条目形成了重叠的混淆矩阵，这表明所提出的架构以 98.10 % 的准确率正确分类了猴痘病例，以 96.84 % 的准确率正确分类了非猴痘病例。这些结果突出了该方法在检测真实阳性样本方面的有效性。此外，第一个折叠表现出最弱的性能，在 429 个测试样本中有 15 个被误分类。

接收者操作特征 (ROC) 曲线及其曲线下面积 (AUC) 被用于将我们的模型与之前的最先进预测模型进行比较。ROC 曲线可视化显示在不同阈值下的真阳性率与假阳性率。这使得可以根据用户特定需求 (通常与可变的错误成本和效率假设相关) 评估和选择分类模型。该图展示了召回率和特异性之间的权衡，而 AUC 量化了模型的预测能力。较高的

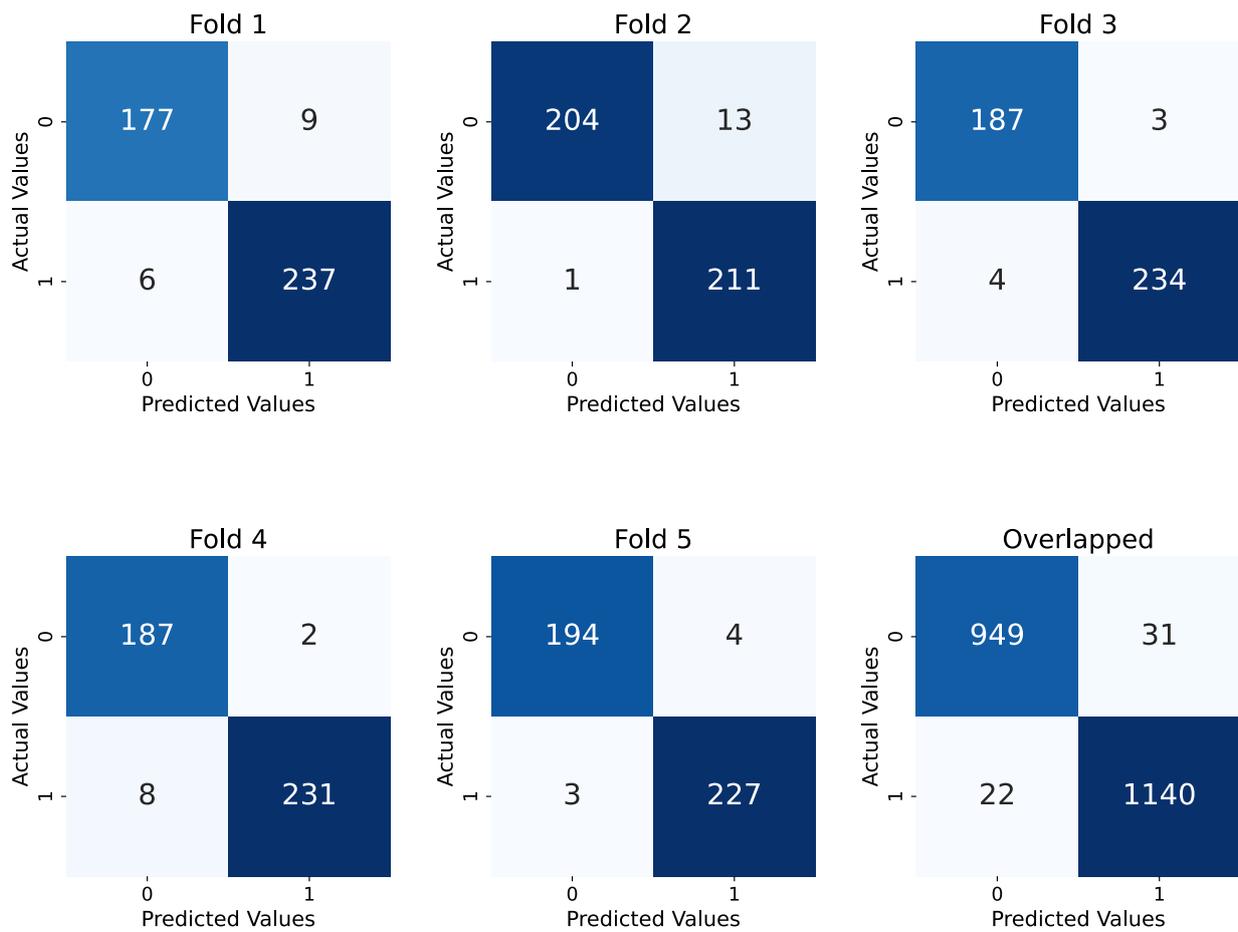


Figure 9: 对所提出模型的混淆矩阵进行可视化。

Table 9: 使用 AVOA 的分类器的优化超参数及其范围

Classifier	Hyper-parameters	Domain Range	Optimal Value
LightGBM	Number of decision trees	[15, ..., 200]	84
	Maximum depth of a tree	[5, ..., 300]	127
	Number of boosting iterations	[70, ..., 300]	100
	Learning rate	[1×10^{-7} , ..., 0.9]	2.405572×10^{-1}
	Maximum number of leaves on one tree	[5, ..., 400]	291
SVM	Kernel	('linear', 'rbf')	rbf
	Cost	[0.1, ..., 20]	14.681521
	Kernel coefficient (γ)	[1×10^{-4} , ..., 0.9]	6.535413×10^{-3}
	Probability estimates	-	True
XGBoost	Base Learner	('gblinear', 'gbtree')	Gradient boosted tree
	Learning rate (η)	[1×10^{-7} , ..., 0.9]	2.863612×10^{-2}
	Lagrange multiplier (γ)	[1×10^{-5} , ..., 0.9]	1.571300×10^{-3}
	Maximum depth of a tree	[5, ..., 200]	23
	Maximum number of leaves on one tree	[5, ..., 800]	720

AUC 表明更高的有用性，展示了分类器在不同类别之间进行区分的能力。明显向左上角倾斜的 ROC 曲线表明性能更佳，反映出高召回率和低假阳性率。Figure 10 展示了这些结果，以放大的格式代表了 AUC。ROC 曲线表明我们提议的模型优于其他分类器。

深度学习模型所做的决策可以通过可解释人工智能 (XAI) 技术来阐明。本研究使用的两种方法是局部解释模型无关解释 (LIME) 和 Grad-CAM。

LIME 是一种可视化工具，它突出显示卷积神经网络在训练期间用于进行预测的重要特征。它通过使用一个可解释的模型 (Nayak et al., 2023) 来局部近似地解释分类器的预测。LIME 对象是以一个特定的查询点和定义数量的显著预测因子生成的。此过程生成一个合成数据集，之后用于训练一个简单、可理解的模型，该模型解释围绕查询点的合成数据的预测。在 Figure 11e 和 Figure 11j 中为 “Monkeypox” 类展示 LIME，在 Figure 11o 和 Figure 11t 中为 “Normal” 类展示 LIME。这些图中的超像素是使用回归树生成的，其中较暖区域指示对做出预测更具责任的重要特征，而较冷区域则表示在决策过程中重要性较小的特征。

Table 10: 各种模型的分类指标。

Classifiers	Performance Metrics						
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	Kappa	F1-Score (%)	AUC (%)
LightGBM	93.60	92.47	96.11	90.79	0.8705	94.22	93.45
SVM	95.89	94.75	97.88	93.63	0.9168	96.27	95.75
XGBoost	93.51	92.73	95.60	91.20	0.8688	94.10	93.75
NGBoost	95.15	93.65	97.70	92.20	0.9016	95.61	94.95
AVOA-LGBM	94.12	93.67	95.68	92.40	0.8811	94.64	94.04
AVOA-SVM	96.87	96.42	97.90	95.78	0.9368	97.13	96.84
AVOA-XGB	94.44	93.65	96.38	92.30	0.8876	94.96	94.12
AVOA-NGB	97.20	97.08	97.80	96.60	0.9435	97.42	97.20
LightGBM + PCA	93.18	92.05	95.73	90.26	0.8620	93.83	92.92
SVM + PCA	96.31	95.53	97.78	94.67	0.9255	96.62	96.23
XGBoost + PCA	93.42	92.55	95.54	90.99	0.8668	94.01	92.86
NGBoost + PCA	95.57	94.11	97.95	92.82	0.9102	95.97	95.39
AVOA-LGBM + PCA	93.88	92.62	96.41	91.04	0.8763	94.45	92.88
AVOA-SVM + PCA	96.78	96.18	97.98	95.49	0.9349	97.05	96.67
AVOA-XGB + PCA	94.17	93.11	96.43	91.67	0.8821	94.69	94.01
Proposed Method	97.53	97.34	98.15	96.90	0.9500	97.72	97.47

这些发现得到了 Grad-CAM 的支持，如 Figure 11c 和 Figure 11h 中针对“猴痘”类别以及 Figure 11m 和 Figure 11r 中针对“正常”类别所示。Grad-CAM 是一种可视化技术，旨在展示 CNN 的“思考”过程，从而使用户能够更有效地理解模型的决策过程。它利用来自最后卷积层的梯度信息，识别输入图像中对特定类别影响最大的区域。通过为每个神经元分配重要性值并生成热图，Grad-CAM 突出显示模型认为对分类至关重要的区域。这种方法不仅提供了对 CNN 内部逻辑的宝贵见解，还通过强调与诊断相关的区域来支持临床应用。在我们的特定情境中，Grad-CAM 主要聚焦于各种疾病类别的污染区域，在这些区域可以发现疾病的主要迹象。

Figure 11 证明了 Xception 模型能够有效识别用于分类的差异区域。Grad-CAM 用黄色或深黄色突出显示病毒感染区域，而 LIME 能够在地图上用其超级像素圈出可能感染的区域。对于 Grad-CAM，使用了 Xception 模型的“block14 sepconv2 act”层。对于 LIME，参数设置如下：特征数量为 5，样本数量为 1,000，Xception 模型的前五个标签。结合起来，这些技术展示了模型准确学习和定位感染区域的能力。

Table 11: 5 折交叉验证结果。

Number of Fold	Evaluation Metrics					
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	Kappa	F1-Score (%)
Fold 1	96.503	96.34	97.53	95.16	0.9287	96.93
Fold 2	96.736	94.20	99.53	94.01	0.9384	96.79
Fold 3	98.364	98.73	98.32	98.42	0.9669	98.53
Fold 4	97.663	99.14	96.65	98.94	0.9528	97.88
Fold 5	98.364	98.27	98.70	97.98	0.9671	98.48
Mean	97.53	97.34	98.15	96.90	0.9500	97.72
Standard Deviation	0.78	1.84	0.99	1.95	0.015	0.74

7. 结论

这项工作在学习应用于传染病检测方面标志着重要的进展，特别是在资源匮乏的环境中。未来的研究可以探索集成到移动医疗系统中的实时诊断工具，从而在偏远地区提供及时的诊断支持。

References

- Mohammad Reza Abbasniya, Sayed Ali Sheikholeslamzadeh, Hamid Nasiri, and Samaneh Emami. Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods. *Computers and Electrical Engineering*, 103:108382, 10 2022. ISSN 00457906. doi: 10.1016/j.compeleceng.2022.108382.
- Benyamin Abdollahzadeh, Farhad Soleimani Gharehchopogh, and Seyedali Mirjalili. African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. *Computers & Industrial Engineering*, 158:107408, 8 2021. ISSN 03608352. doi: 10.1016/j.cie.2021.107408.
- Amesh Adalja and Tom Inglesby. A novel international monkeypox outbreak. *Annals of Internal Medicine*, 175:1175–1176, 8 2022. ISSN 0003-4819. doi: 10.7326/M22-1581.
- Masoud Ahmadipour, Muhammad Murtadha Othman, Rui Bo, Zainal Salam, Hussein Mohammed Ridha, and Kamrul Hasan. A novel microgrid fault detection and classification method using maximal overlap discrete wavelet packet transform and an augmented lagrangian particle swarm optimization-support vector machine. *Energy Reports*, 8:4854–4870, 11 2022. ISSN 23524847. doi: 10.1016/j.egy.2022.03.174.
- Emmanuel Alakunle, Ugo Moens, Godwin Nchinda, and Malachy Ifeanyi Okeke. Monkeypox virus in nigeria:

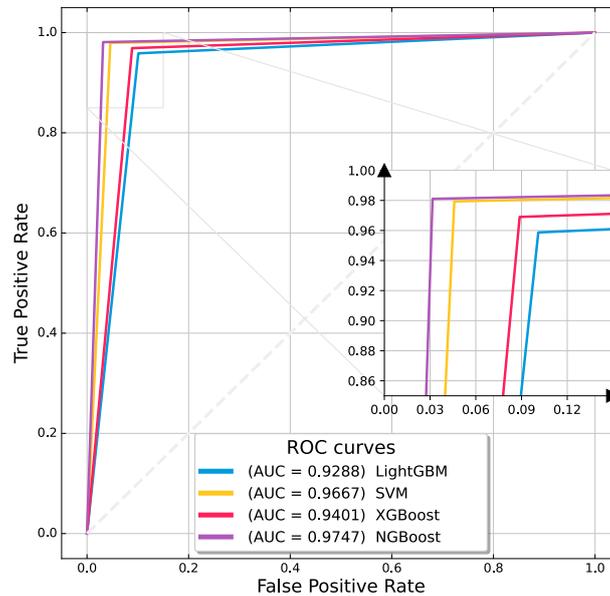


Figure 10: 不同分类器的 ROC 曲线。

Infection biology, epidemiology, and evolution. *Viruses*, 12:1257, 11 2020. ISSN 1999-4915. doi: 10.3390/v12111257.

Shams Nafisa Ali, Md. Tazuddin Ahmed, Joydip Paul, Tasnim Jahan, S M Sakeef Sani, Nawsabah Noor, and Taufiq Hidayat Hasan. Monkeypox skin lesion detection using deep learning models: A feasibility study. *arXiv preprint arXiv:2207.03342*, 2022.

Maram Fahaad Almufareh, Samabia Tehsin, Mamoona Humayun, and Sumaira Kausar. A transfer learning approach for clinical detection support of monkeypox skin lesions. *Diagnostics*, 13:1503, 4 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13081503.

M Hafidz Ariansyah, Sri Winarno, and Ramadhan Rakhmat Sani. Monkeypox and measles detection using cnn with vgg-16 transfer learning. *Journal of Computing Research and Innovation*, 8:32–44, 2 2023. ISSN 26008793. doi: 10.24191/jcrinn.v8i1.340.

Diponkor Bala, Md. Shamim Hossain, Mohammad Alamgir Hossain, Md. Ibrahim Abdullah, Md. Mizanur Rahman, Balachandran Manavalan, Naijie Gu, Mohammad S. Islam, and Zhangjin Huang. Monkeynet: A robust deep convolutional neural network for monkeypox disease detection and classification. *Neural Networks*, 161:757–775, 4 2023. ISSN 08936080. doi: 10.1016/j.neunet.2023.02.022.

James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In J Shawe-Taylor, R Zemel, P Bartlett, F Pereira, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Adam Bohr and Kaveh Memarzadeh. *The rise of artificial intelligence in healthcare applications*, pages

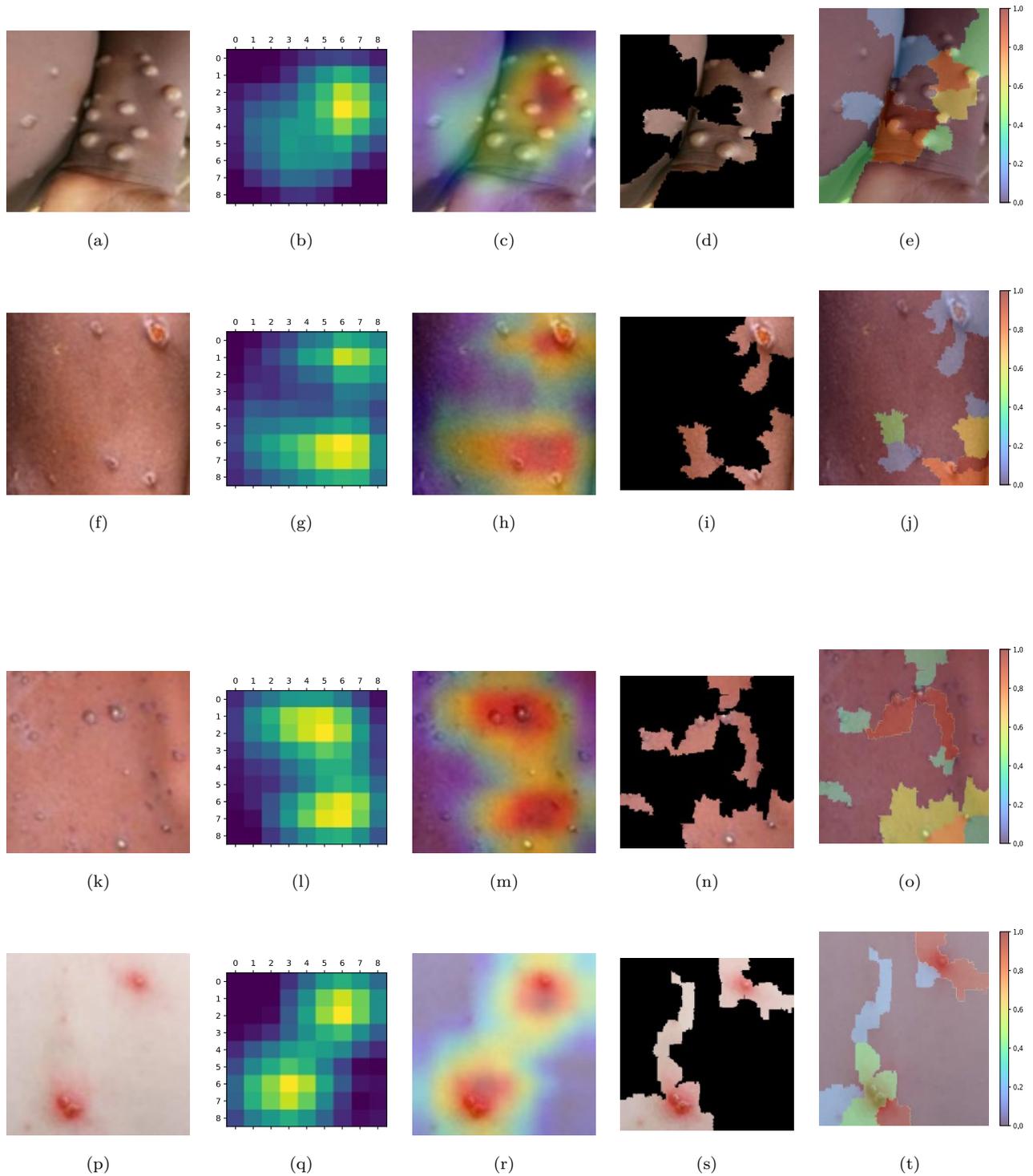


Figure 11: (a) 和 (f) 代表'猴痘'类, (k) 和 (p) 代表'正常'类。(b) 和 (g) 展示'猴痘'类的热图, 而 (l) 和 (q) 展示'正常'类的热图。(c) 和 (h) 提供'猴痘'类的 Grad-CAM 可视化, (m) 和 (r) 为'正常'类。(d)、(i)、(n) 和 (s) 显示来自 LIME 可视化的前 7 个特征, (e)、(j)、(o) 和 (t) 则展示它们各自的 LIME 图像。

- 25–60. Elsevier, 2020. doi: 10.1016/B978-0-12-818438-7.00002-2.
- Stevó Bozinovski. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica*, 44, 9 2020. ISSN 1854-3871. doi: 10.31449/inf.v44i3.2828.
- Centers for Disease Control and Prevention. 2022-2023 mpox outbreak global map, 2022. URL <https://archive.cdc.gov/#/details?url=https://www.cdc.gov/poxvirus/mpox/response/2022/world-map.html>.
- Tianqi Chen and Carlos Guestrin. Xgboost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 8 2016. ISBN 9781450342322. doi: 10.1145/2939672.2939785.
- Xia Chen and Philipp Geyer. Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty. *Applied Energy*, 307:118240, 2 2022. ISSN 03062619. doi: 10.1016/j.apenergy.2021.118240.
- Huan-Jung Chiu, Tzuu-Hseng S Li, and Ping-Huan Kuo. Breast cancer–detection system using pca, multilayer perceptron, transfer learning, and support vector machine. *IEEE Access*, 8:204309–204324, 2020. doi: 10.1109/ACCESS.2020.3036912.
- François Chollet. Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2016.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 9 1995. ISSN 0885-6125. doi: 10.1007/BF00994018.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 6 2009. ISBN 978-1-4244-3992-8. doi: 10.1109/CVPR.2009.5206848.
- Chitradevi Dhakhinamoorthy, Sathish Kumar Mani, Sandeep Kumar Mathivanan, Senthilkumar Mohan, Prabhu Jayagopal, Saurav Mallik, and Hong Qin. Hybrid whale and gray wolf deep learning optimization algorithm for prediction of alzheimer’ s disease. *Mathematics*, 11:1136, 2 2023. ISSN 2227-7390. doi: 10.3390/math11051136.
- Haipei Dong, Dakuo He, and Fuli Wang. Smote-xgboost using tree parzen estimator optimization for copper flotation method classification. *Powder Technology*, 375:174–181, 9 2020. ISSN 00325910. doi: 10.1016/j.powtec.2020.07.065.
- Tony Duan, Anand Avati, Daisy Yi Ding, Khanh K Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. Ngboost: natural gradient boosting for probabilistic prediction. In *Proceedings of the 37th International Conference on Machine Learning*. JMLR.org, 2020.
- Shawni Dutta. Revealing brain tumor using cross-validated ngboost classifier. *International Journal of Machine Learning and Networked Collaborative Engineering*, 4:12–20, 8 2020. ISSN 25813242. doi:

10.30991/IJMLNCE.2020v04i01.002.

- Alireza Farzipour, Roya Elmi, and Hamid Nasiri. Detection of monkeypox cases based on symptoms using xgboost and shapley additive explanations methods. *Diagnostics*, 13(14):2391, 2023.
- Nabeela Fatima and Kiranmai Mandava. Monkeypox- a menacing challenge or an endemic? *Annals of Medicine & Surgery*, 79, 7 2022. ISSN 2049-0801. doi: 10.1016/j.amsu.2022.103979.
- Xin-Yu Fu, Xin-Li Mao, Hao-Wen Wu, Jia-Ying Lin, Zong-Qing Ma, Zhi-Cheng Liu, Yue Cai, Ling-Ling Yan, Yi Sun, Li-Ping Ye, and Shao-Wei Li. Development and validation of lightgbm algorithm for optimizing of helicobacter pylori antibody during the minimum living guarantee crowd based gastric cancer screening program in taizhou, china. *Preventive Medicine*, 174:107605, 9 2023. ISSN 00917435. doi: 10.1016/j.ypmed.2023.107605.
- Terrence S. Furey, Nello Cristianini, Nigel Duffy, David W. Bednarski, Michèl Schummer, and David Hausler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16:906–914, 10 2000. ISSN 1367-4811. doi: 10.1093/bioinformatics/16.10.906.
- Paria Ghaheri, Hamid Nasiri, Ahmadreza Shateri, and Arman Homafar. Diagnosis of parkinson’ s disease based on voice signals using shap and hard voting ensemble method. *Computer Methods in Biomechanics and Biomedical Engineering*, 27:1858–1874, 10 2024. ISSN 1025-5842. doi: 10.1080/10255842.2023.2263125.
- Mrinal Haloi, K Rajalakshmi, and Pradeep Walia. Towards radiologist-level accurate deep learning system for pulmonary screening. *arXiv preprint arXiv:1807.03120*, 2018.
- Qusay Shihab Hamad, Hussein Samma, Shahrel Azmin Suandi, and Junita Mohamad-Saleh. Q-learning embedded sine cosine algorithm (qlsca). *Expert Systems with Applications*, 193:116417, 5 2022. ISSN 09574174. doi: 10.1016/j.eswa.2021.116417.
- Md. Enamul Haque, Md. Rayhan Ahmed, Razia Sultana Nila, and Salekul Islam. Classification of human monkeypox disease using deep learning models and attention mechanisms. *arXiv preprint arXiv:2211.15459*, 2022.
- Ali Asghar Heidari, Seyedali Mirjalili, Hossam Faris, Ibrahim Aljarah, Majdi Mafarja, and Huiling Chen. Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97:849–872, 8 2019. ISSN 0167739X. doi: 10.1016/j.future.2019.02.028.
- Saddam Hussain, Mohd Wazir Mustafa, Khalil Hamdi Ateyeh Al-Shqeerat, Faisal Saeed, and Bander Ali Saleh Al-rimy. A novel feature-engineered-ngboost machine-learning framework for fraud detection in electric power consumption data. *Sensors*, 21:8423, 12 2021. ISSN 1424-8220. doi: 10.3390/s21248423.
- Muhammed Coskun Irmak, Tolga Aydin, and Mete Yaganoglu. Monkeypox skin lesion detection with mobilenetv2 and vggnet models. In *2022 Medical Technologies Congress (TIPTEKNO)*, pages 1–4. IEEE, 10 2022. ISBN 978-1-6654-5432-2. doi: 10.1109/TIPTEKNO56568.2022.9960194.

- Towhidul Islam, M A Hussain, Forhad Uddin Hasan Chowdhury, and B M Riazul Islam. Can artificial intelligence detect monkeypox from digital skin images? *bioRxiv*, 2022. doi: 10.1101/2022.08.08.503193.
- Sami Ben Jabeur, Houssein Ballouk, Wissal Ben Arfi, and Rabeh Khalfaoui. Machine learning-based modeling of the environmental degradation, institutional quality, and economic growth. *Environmental Modelling & Assessment*, 27:953–966, 12 2022. ISSN 1420-2026. doi: 10.1007/s10666-021-09807-0.
- Wenjiang Jiao, Xingwei Hao, and Chao Qin. The image classification method with cnn-xgboost model based on adaptive particle swarm optimization. *Information*, 12:156, 4 2021. ISSN 2078-2489. doi: 10.3390/info12040156.
- Chen Jing and Jian Hou. Svm and pca based fault classification approaches for complicated industrial process. *Neurocomputing*, 167:636–642, 11 2015. ISSN 09252312. doi: 10.1016/j.neucom.2015.03.082.
- Taskin Kavzoglu and Alihan Teke. Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (xgboost). *Bulletin of Engineering Geology and the Environment*, 81:201, 5 2022. ISSN 1435-9529. doi: 10.1007/s10064-022-02708-w.
- Samina Khalid, Tehmina Khalil, and Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In *2014 Science and Information Conference*, pages 372–378. IEEE, 8 2014. ISBN 978-0-9893193-1-7. doi: 10.1109/SAI.2014.6918213.
- Kristi L. Koenig, Christian K. Bej, and Aileen M. Marty. Monkeypox 2022 identify-isolate-inform: A 3i tool for frontline clinicians for a zoonosis with escalating human community transmission. *One Health*, 15:100410, 12 2022. ISSN 23527714. doi: 10.1016/j.onehlt.2022.100410.
- Narendra Kumar, Arpan Acharya, Howard E. Gendelman, and Siddappa N. Byrareddy. The 2022 outbreak and the pathobiology of the monkeypox virus. *Journal of Autoimmunity*, 131:102855, 7 2022. ISSN 08968411. doi: 10.1016/j.jaut.2022.102855.
- Dominik Lewy and Jacek Mańdziuk. An overview of mixing augmentation methods and augmentation strategies. *Artificial Intelligence Review*, 56:2111–2169, 3 2023. ISSN 0269-2821. doi: 10.1007/s10462-022-10227-z.
- Bohao Li, Kai Liu, Ming Wang, Yanfang Wang, Qian He, Linmei Zhuang, and Weihua Zhu. High-spatiotemporal-resolution dynamic water monitoring using lightgbm model and sentinel-2 msi data. *International Journal of Applied Earth Observation and Geoinformation*, 118:103278, 4 2023. ISSN 15698432. doi: 10.1016/j.jag.2023.103278.
- Boyi Li, Felix Wu, Ser-Nam Lim, Serge J Belongie, and Kilian Q Weinberger. On feature normalization and data augmentation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12378–12387, 2020.
- Kun Li, Haocheng Xu, and Xiao Liu. Analysis and visualization of accidents severity based on lightgbm-tpe. *Chaos, Solitons & Fractals*, 157:111987, 4 2022. ISSN 09600779. doi: 10.1016/j.chaos.2022.111987.

- Adam MacNeil, Mary G. Reynolds, Zach Braden, Darin S. Carroll, Vanda Bostik, Kevin Karem, Scott K. Smith, Whitney Davidson, Yu Li, Amba Moundeli, Jean-Vivien Mombouli, Aisha O. Jumaan, D. Scott Schmid, Russell L. Regnery, and Inger K. Damon. Transmission of atypical varicella-zoster virus infections involving palm and sole manifestations in an area with monkeypox endemicity. *Clinical Infectious Diseases*, 48:e6–e8, 1 2009. ISSN 1058-4838. doi: 10.1086/595552.
- Monika A. Myszczyńska, Poojitha N. Ojamies, Alix M. B. Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M. Hautbergue, Joanna D. Holbrook, and Laura Ferraiuolo. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, 16:440–456, 8 2020. ISSN 1759-4758. doi: 10.1038/s41582-020-0377-8.
- Tushar Nayak, Krishnaraj Chadaga, Niranjana Sampathila, Hilda Mayrose, Nitila Gokulkrishnan, Muralidhar Bairy G, Srikanth Prabhu, Swathi K. S, and Shashikiran Umakanth. Deep learning based detection of monkeypox virus using skin lesion images. *Medicine in Novel Technology and Devices*, 18:100243, 6 2023. ISSN 25900935. doi: 10.1016/j.medntd.2023.100243.
- Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U. Rajendra Acharya. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, 121:103792, 6 2020. ISSN 00104825. doi: 10.1016/j.combiomed.2020.103792.
- Bhaskar Patnaik, Manohar Mishra, Ramesh C. Bansal, and Ranjan K. Jena. Modwt-xgboost based smart energy solution for fault detection and classification in a smart microgrid. *Applied Energy*, 285:116457, 3 2021. ISSN 03062619. doi: 10.1016/j.apenergy.2021.116457.
- Soo-Chang Pei and Chao-Nan Lin. Image normalization for pattern recognition. *Image and Vision Computing*, 13:711–723, 12 1995. ISSN 02628856. doi: 10.1016/0262-8856(95)98753-G.
- Chong Peng and Qiang Cheng. Discriminative ridge machine: A classifier for high-dimensional data or imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, PP:1–15, 7 2020. doi: 10.1109/TNNLS.2020.3006877.
- Eskild Petersen, Anu Kantele, Marion Koopmans, Danny Asogun, Adesola Yinka-Ogunleye, Chikwe Ihekweazu, and Alimuddin Zumla. Human monkeypox. *Infectious Disease Clinics of North America*, 33: 1027–1043, 12 2019. ISSN 08915520. doi: 10.1016/j.idc.2019.03.001.
- Rishav Pramanik, Bihan Banerjee, George Efimenko, Dmitrii Kaplun, and Ram Sarkar. Monkeypox detection from skin lesion images using an amalgamation of cnn models aided with beta function-based normalization scheme. *PLOS ONE*, 18:e0281815, 4 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0281815.
- Susan Realegeno, Andreas S. Puschnik, Amrita Kumar, Cynthia Goldsmith, Jillybeth Burgado, Suryaprakash Sambhara, Victoria A. Olson, Darin Carroll, Inger Damon, Tetsuya Hirata, Taroh Kinoshita, Jan E. Carette, and Panayampalli Subbian Satheshkumar. Monkeypox virus host factor screen using haploid cells identifies essential role of garp complex in extracellular virus formation. *Journal of*

- Virology*, 91, 6 2017. ISSN 0022-538X. doi: 10.1128/JVI.00011-17.
- John G. Rizk, Giuseppe Lippi, Brandon M. Henry, Donald N. Forthal, and Youssef Rizk. Prevention and treatment of monkeypox. *Drugs*, 82:957–963, 6 2022. ISSN 0012-6667. doi: 10.1007/s40265-022-01742-y.
- Sourav Saha, Trina Chakraborty, Rejwan Bin Sulaiman, and Tithi Paul. A comparative analysis of cnn-based pretrained models for the detection and prediction of monkeypox. *arXiv preprint arXiv:2302.10277*, 2023.
- Veysel Harun Sahin, Ismail Oztel, and Gozde Yolcu Oztel. Human monkeypox classification from skin lesion images with deep pre-trained network using mobile application. *Journal of Medical Systems*, 46:79, 10 2022. ISSN 1573-689X. doi: 10.1007/s10916-022-01863-7.
- Abhibhav Sharma and Buddha Singh. Ae-lgbm: Sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and lightgbm. *Computers in Biology and Medicine*, 125:103964, 10 2020. ISSN 00104825. doi: 10.1016/j.combiomed.2020.103964.
- Chiranjibi Sitaula and Tej Bahadur Shahi. Monkeypox virus detection using pre-trained deep learning-based approaches. *Journal of Medical Systems*, 46:78, 10 2022. ISSN 1573-689X. doi: 10.1007/s10916-022-01868-2.
- Chenjun Tang, Wei Sun, Wei Wu, and Min Xue. A hybrid improved whale optimization algorithm. In *2019 IEEE 15th International Conference on Control and Automation (ICCA)*, pages 362–367. IEEE, 7 2019. ISBN 978-1-7281-1164-3. doi: 10.1109/ICCA.2019.8900003.
- Zhilong Yang. Monkeypox: A potential global threat? *Journal of Medical Virology*, 94:4034–4036, 9 2022. ISSN 0146-6615. doi: 10.1002/jmv.27884.
- Jinhua You, Heming Jia, Di Wu, Honghua Rao, Changsheng Wen, Qingxin Liu, and Laith Abualigah. Modified artificial gorilla troop optimization algorithm for solving constrained engineering optimization problems. *Mathematics*, 11:1256, 3 2023. ISSN 2227-7390. doi: 10.3390/math11051256.
- Jianguo Zhou and Zhongtian Xu. A novel three-stage hybrid learning paradigm based on a multi-decomposition strategy, optimized relevance vector machine, and error correction for multi-step forecasting of precious metal prices. *Resources Policy*, 80:103148, 1 2023. ISSN 03014207. doi: 10.1016/j.resourpol.2022.103148.
- Zhi-Hua Zhou and Ji Feng. Deep forest. *National Science Review*, 6:74–86, 1 2019. ISSN 2095-5138. doi: 10.1093/nsr/nwy108.