储藏马铃薯质量检测：基于迁移学习的卷积神经网络与视觉Transformer方法

作者

Shrikant Kapse*、Priyankkumar Dhrangdhariya、Priya Kedia、Manasi Patwardhan、Shankar Kausley、Soumyadipta Maiti、Beena Rai、Shirish Karande

TCS Research，欣贾瓦迪拉吉夫·甘地信息技术园三期，欣贾瓦迪，浦那，皮姆普里-钦奇瓦德，马哈拉施特拉邦，印度-*411057*

*通讯作者：TCS Research，第三期，欣贾瓦迪拉吉夫·甘地信息技术园，欣贾瓦迪，浦那，皮姆普里-钦奇瓦德，马哈拉施特拉邦，印度－411057 电子邮箱：shrikant.kapse@tcs.com

摘要

基于图像的深度学习为储藏期间监测马铃薯品质提供了一种非侵入式、可扩展的解决方案，可应对芽体检测、失重估计和货架期预测等关键挑战。本研究在受控的温度和湿度条件下，历时200天采集图像及相应的重量数据。借助ResNet、VGG、DenseNet和视觉Transformer（ViT）等强大的预训练架构，我们设计了两种专用模型：（1）用于芽体检测的高精度二分类器；（2）先进的多分类预测器，用于估计失重并预测剩余货架期，且准确率出色。DenseNet表现卓越，芽体检测准确率达到98.03%。货架期预测模型在采用粗粒度类别划分（2–5类）时表现最佳，准确率超过89.83%；而采用更细的类别划分（6–8类）时，由于视觉差异细微且每个类别的数据有限，准确率有所下降。这些发现表明，将基于图像的模型集成到自动分选和库存系统中切实可行，可实现对发芽土豆的早期识别，并依据储藏阶段进行动态分类。其实际意义包括改善库存管理、实施差异化定价策略，以及减少供应链中的食物浪费。尽管精确预测货架期区间仍具挑战，但采用较宽泛的类别划分有助于确保稳健的性能。未来研究应致力于开发基于多种马铃薯品种和储藏条件训练的通用模型，以增强适应性和可扩展性。总体而言，这种方法经济高效且无损，可用于质量评估，助力提升马铃薯储藏与配送的效率和可持续性。

关键词

芽体分类、货架期预测、失重、卷积神经网络、视觉Transformer、迁移学习、多类别图像分类

1. 引言

马铃薯是继水稻和小麦之后全球种植面积排名第三的粮食作物，为数十亿人口提供主食，在全球粮食安全中发挥着至关重要的作用。然而，在采后贮藏和配送过程中，马铃薯会受到温度、湿度、周围气体浓度和储藏时间等环境条件的影响，发生显著的生理和生化变化。这些变化会体现在内部和外部品质参数上，包括糖含量、失重、pH值、可溶性固形物总量，以及发芽、皱缩、真菌生长和表皮变黑等可见的表面变化。其中，失重和发芽是品质劣变的重要指标。失重主要由蒸腾作用和呼吸作用引起，会导致明显的皱缩和商品质量下降。失重超过10%的马铃薯通常被认为不适合销售（Pinhero等，2009年）。发芽由涉及赤霉素、细胞分裂素和脱落酸（ABA）的激素变化驱动，不仅会加速失重，还会降低消费者接受度（Sonnewald与Sonnewald，2014）。芽体的表皮相比马铃薯表面具有明显更高的透水性，进一步提高了蒸腾速率。

鉴于每年只有3至4个月能买到新鲜马铃薯，其余时间市场主要供应储藏马铃薯。这些储藏块茎的储藏时间和品质各不相同，因此需要采用有效的基于品质的分级方法，以支持差异化定价、库存管理以及及时的买卖决策。传统的人工检查方法劳动强度大、主观性强，且难以高效应用于大规模作业。近年来，基于图像的深度学习取得了进展，在自动化采后品质评估方面展现出良好前景。与依赖手工设计特征和受控成像条件的传统机器学习方法不同，深度学习，尤其是卷积神经网络（CNN）和视觉Transformer（ViT），能够自动学习复杂的视觉模式，例如芽形态、表面皱纹和变色（Wang等，2025 & Gülmez等，2025）。这些模型已成功应用于多种作物的分级（成熟阶段）和缺陷检测，包括枣、番茄、香蕉和马铃薯（Piedad等，2018；Surya Prabha与Satheesh Kumar，2015；Nasiri等，2019）。例如，基于卷积神经网络的系统利用深度成像进行马铃薯分级时，准确率超过86%；而基于YOLO的模型已用于实时检测绿化和腐烂等外部缺陷（Su 等人，2020；Wu等，2025）。

尽管取得了这些进展，关于基于图像预测马铃薯货架期和成熟度的研究仍存在明显空白。虽然一些研究探讨了水分含量估算、病害检测和缺陷分类，但很少有研究将视觉线索与失重和发芽数据相结合，以估算剩余货架期（Kamilaris & Prenafeta-Boldú，2018；Le & Lin，2019）。以往研究采用了多种机器学习和图像处理技术，利用视觉特征评估马铃薯品质。基于颜色和纹理的水分含量估算采用了多层感知机（MLP）模型（Ebrahimi 等人，2012）。研究人员使用标记控制分水岭分割分析长宽比（Si 等人，2017），并将基于主成分分析的特征选择与最近邻和遗传算法相结合，以检测腐烂和绿化等外部缺陷（Dacal-Nieto 等人，2009）。研究人员还采用主成分分析、判别分析和人工神经网络模型，利用颜色、纹理和形状进行品种识别（Azizi 等人，2016）。Wei 等人提出了一种跨模态检测策略，将机器视觉与可见光/近红外光谱相结合，实现对马铃薯外部和内部缺陷的实时智能分类（Wei 等人，2024）。成像光谱与机器学习相结合，已被证明是表征和预测块茎作物品质属性的一种有效无损方法（Su 和 Xue，2021）。近期将深度相机与卷积神经网络相结合的研究，在大小和外观分类方面展现出较高准确率，凸显了三维成像在采后品质评估中的潜力（Su 等人，2020）。自动化视觉系统通过利用从数字图像中提取的颜色、大小、纹理和硬度特征，在马铃薯无损品质检测方面展现出巨大潜力（Ibrahim 等人，2020）。近期成像光谱、机器视觉、深度成像和自动化视觉系统与机器学习及人工智能的结合，为马铃薯品质评估提供了快速、无损的解决方案，可依据物理、化学和表面属性进行准确分级（Danielak 等人，2023）。可解释人工智能与迁移学习在深度学习框架中的结合，能够准确且可解释地对马铃薯叶部病害进行分类，准确率超过97%，同时提高了其在实际农业应用中的透明度（Alhammad 等人，2025）。基于视觉的深度学习技术可实现工业涂料和涂层缺陷的自动化、准确分类，减少对低效人工检查的依赖（Dhrangdhariya 等人，2025）。这些研究表明，基于图像的方法在马铃薯无损品质评估方面具有应用潜力。

本研究旨在通过开发全面的深度学习框架来弥补这一空白，利用图像数据进行马铃薯品质检测。具体目标如下：

1. 对马铃薯图像中是否存在芽进行分类。2. 预测失重百分比，并使用多类别分类估算剩余货架期。

本研究还考察了输出类别数量的变化对货架期预测准确率的影响，旨在为供应链管理中的实际部署确定最佳分类粒度。

2. 材料与方法

2.1 数据收集

‘FC3’栽培品种的马铃薯采自印度浦那的一家当地农场。清洗以去除泥土和杂物后，进行为期两周的愈伤处理。愈伤处理结束后，将18个马铃薯分装到托盘中，每个托盘放置6个。储藏条件设定为温度21 ± 2°C、相对湿度（RH）70 ± 10%。

根据可见变化的程度，每隔 1 至 5 天拍摄一次马铃薯图像。拍摄时，将 Nikon Coolpix P100 数码相机固定在支架上，以确保成像距离和角度一致。每张图像拍摄一个装有六个马铃薯的托盘，随后使用截图工具手动裁剪图像，以分离出单个马铃薯图像（如图 1 所示）。裁剪后的图像尺寸约为 250× 250 像素，分辨率为 96 每英寸点数。同时，在为期 200 天的实验中，每隔相同时间使用 METTLER TOLEDO PB3002-S 精密天平记录托盘重量。根据公式（1），计算每个马铃薯在任一时间点的累计失重率：

$$Cumulative\ weight\ loss\ \% = \frac{W_0 - W_t}{W_0} \times 100 \quad (1)$$

其中，

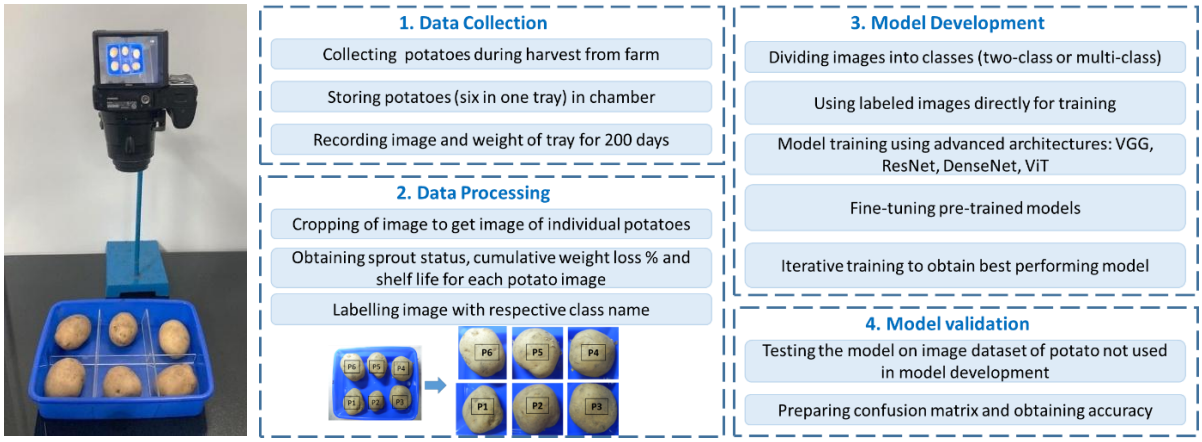
W_0 为马铃薯开始储藏时的重量

W_t 表示马铃薯在任意时刻 t 的重量

货架期定义为马铃薯达到 10% 累计失重所需的天数。在任意给定时刻，每个马铃薯的剩余货架期均使用以下公式（2）计算：

$$Remaining\ shelf\ life\ (days) = Shelf\ life\ (days) - Current\ day \quad (2)$$

定期收集图像和重量数据后，为每张图像标注相应的输出类别，以便对相应的分类模型进行监督训练



a) b)

图 1：a) 用于数据收集的实验装置 b) 马铃薯基于图像的模型开发总体架构

2.2 芽体分类

芽检测模型采用二分类设计，包含两个类别："有芽"和"无芽"。数据集由单个马铃薯的图像组成。为确保模型评估稳健，将 80% 的图像分配用于训练，其余 20% 留作测试。通过目视检查对每张图像进行人工标注，以指明是否存在芽。最终得到一个标注数据集，其中包含 255 张训练图像和 51 张测试图像。在对基于 CNN 和 ViT 的模型进行微调时，采用五折交叉验证并结合 GridSearchCV 策略进行模型评估。

2.3 多分类货架期预测

蒸腾作用导致的失重是影响马铃薯在储藏期间品质（货架期）的关键因素。预测累计失重率以及相应的剩余货架

期需要基于马铃薯图像进行预测，这是一项复杂的任务，主要是因为马铃薯的视觉变化是逐渐发生的，而且在储藏初期往往并不明显。起初，马铃薯可能几乎没有可见变化，随后会长出小芽。随着储藏时间延长，更多芽会出现，已有的芽会长得更大，表面也开始形成皱纹，表明老化程度加深。与香蕉或番茄等水果不同，水果与年龄相关的变化通常体现在颜色或大小等简单的二维特征上，而马铃薯则呈现出更复杂的三维视觉线索，例如发芽和皱缩。这些特征使得使用传统的基于图像的方法估算失重和货架期颇具挑战性。

为解决这一问题，我们开发了多类分类模型，根据马铃薯的累计失重率对其进行分类；该指标与剩余货架期相关。我们探索了多种类别配置（从 2 类到 8 类），详见表1；图 2 展示了各类别的样本图像。通过这些配置，我们旨在确定类别数量的最佳设置，以兼顾预测准确率和实际可用性。

为开发货架期预测模型，我们为马铃薯的逐日图像和重量数据标注了相应的累计失重率和剩余货架期，这些数值根据公式（1）和（2）计算得出。货架期预测被设定为一个多类分类问题，并针对 2 至 8 种类别配置训练模型。在每种配置中，各类别代表特定范围的失重百分比。每种配置中的最后一个类别均包含失重大于 10% 的所有样本，其余类别则根据类别数减一，将 0–10% 的范围等分（详见表1）。

针对每种分类配置，均对数据集进行了相应标注，并按照指定的类别标签将图像整理到不同目录中。每个目录代表一个独立类别，用于训练分类模型。该数据集包含六个单独的马铃薯的图像。模型评估采用五折交叉验证，并结合 GridSearchCV 策略进行。在此设置下，255 张图像用于训练，另有 51 张图像留作测试。

T表1：马铃薯多类别货架期预测模型中的类别分布

类别数量	类别	累计失重 (%)	剩余货架期 (天)
2	I级	0 – 10	1 – 121
	II类	>10	0
3	I级	0 – 5	37 – 121
	II类	5 – 10	1 – 36
	III级	>10	0
4	I级	0 – 3.3	52 – 121
	II类	3.3 – 6.6	20 – 51
	III级	6.6 – 9.9	1 – 19
	IV级	>10	0
5	I级	0 – 2.5	62 – 121
	II类	2.5 – 5	38 – 61
	III级	5 – 7.5	14 – 37
	IV级	7.5 – 10	1 – 13
	V级	>10	0
6	I级	0 – 2	74 – 121
	II类	2 – 4	48 – 73
	III级	4 – 6	28 – 47
	IV级	6 – 8	12 – 27
	V级	8 – 10	1 – 11
	VI级	>10	0
7	I级	0 – 1.66	82 – 121
	II类	1.66 – 3.32	52 – 81
	III级	3.32 – 4.98	38 – 51
	IV级	4.98 – 6.64	21 – 37
	V级	6.64 – 8.3	11 – 20
	VI级	8.3 – 9.96	1 – 10
	VII级	>10	0
8	I级	0 – 1.43	89 – 121
	II类	1.43 – 2.86	56 – 88
	III级	2.86 – 4.29	46 – 55
	IV级	4.29 – 5.72	29 – 45
	V级	5.72 – 7.15	17 – 28
	VI级	7.15 – 8.58	8 – 16
	VII级	8.58 – 10	1 – 7
	八年级	>10	0

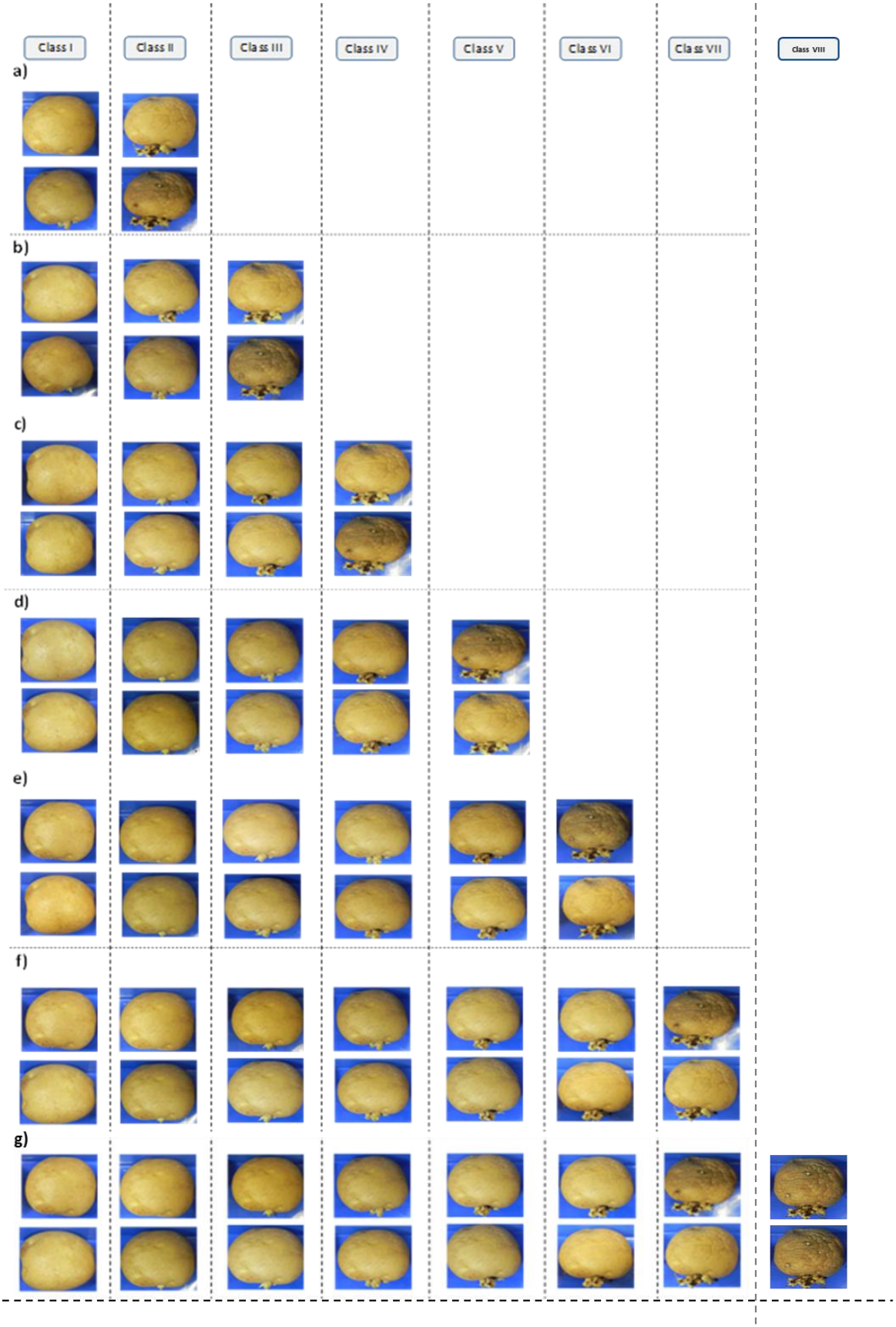


图 2：展示多类别货架期预测模型中不同马铃薯类别视觉差异的样本图像（2–8 个类别）

2.4 模型架构

本研究评估了四种先进的深度学习架构：VGG、ResNet、DenseNet 和视觉 Transformer（ViT），用于对马铃薯失重类别进行多类分类。每种架构都代表了计算机视觉领域中一个独特的架构系列，涵盖传统卷积神经网络（CNN）和现代基于Transformer的视觉模型。所有模型均采用预训练的 ImageNet 权重（Russakovsky 等人，2015年）进行初始化，以利用迁移学习，随后针对马铃薯多类失重预测任务进行微调。以下小节将讨论各架构的结构细节及其所做的修改。

VGG

VGG 由牛津大学视觉几何组提出，是图像识别领域中历史最悠久、应用最广泛的卷积神经网络架构之一（Simonyan 和 Zisserman，2014）。它采用简单而有效的设计，由多个卷积层深度堆叠而成，并使用尺寸较小的 3×3 滤波器，且步幅和填充保持一致。通过定期插入最大池化层，降低空间分辨率并提取更高层次的特征。

在各种变体中，VGG-16 和 VGG-19 应用最为广泛，分别包含 16 个和 19 个权重层，涵盖卷积层和全连接层。本研究选择 VGG-16 模型，以在准确率与计算成本之间取得平衡。为使模型适用于马铃薯失重分类任务，原有的 1000 类输出层被替换为两个包含 1024 个神经元的全连接层，后接一个 Softmax 层，用于生成各输出类别的概率（n），如图3所示。修改后的 VGG-16 模型包含约 1489 万个可训练参数。

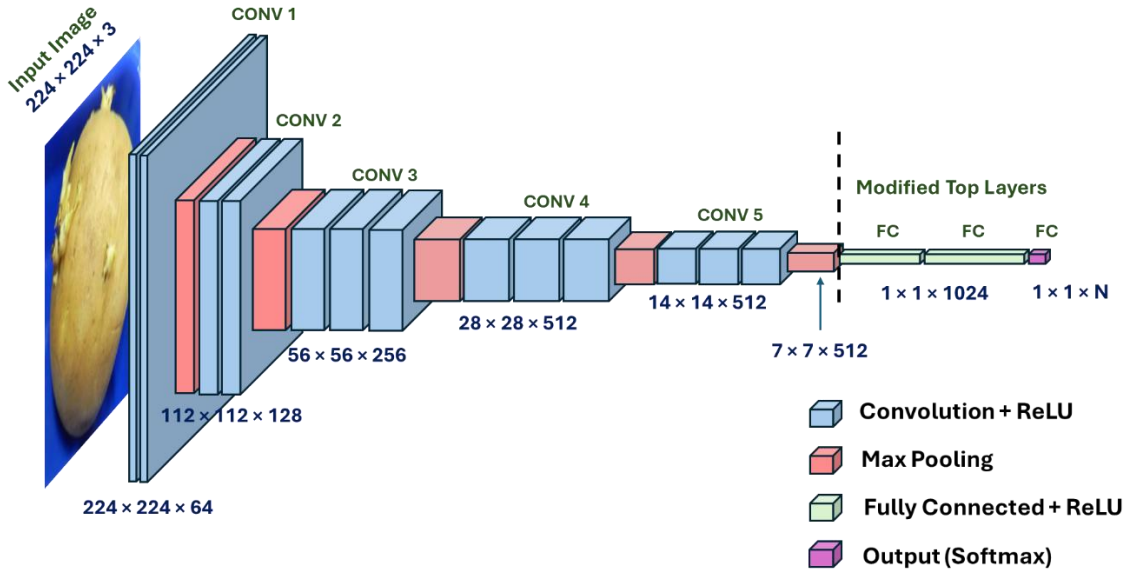


图3：采用修改后顶层的VGG-16架构

ResNet

残差网络（ResNet）由 He 等人于 2016年提出，旨在解决阻碍超深层卷积神经网络训练的梯度消失问题。ResNet 的关键创新是残差连接，也称为“快捷连接”，它允许某一层的输入绕过一个或多个中间层，并直接与输出相加。这使得包含数百层的网络能够稳定训练，同时提高特征复用率，且不增加计算复杂度。

每个残差块在数学上定义为：

$$y = F(x, \{W_i\}) + x \tag{1}$$

其中，x是输入，$F(x, \{W_i\})$表示需要学习的残差映射，y是块的输出。加上x可确保恒等映射，并在反向传播过程中改善梯度流。

本研究采用 ResNet-50 作为深度残差架构的代表。该模型由一个包含 7×7 个滤波器的初始卷积层和后续的最大池化组成，并设有4个阶段的瓶颈残差块，每个阶段分别包含3、4、6和3个残差块，因此总计有50个权重层。各阶段的输出逐步下采样，同时特征通道数量不断增加，使网络能够学习丰富的层次化表示。顶层分类层经过修改，以匹配n类输出。ResNet-50 模型

包含约2352万个可训练参数。图4展示了本研究使用的改进版 ResNet-50 架构，其中显示了卷积层、各阶段的瓶颈残差块、下采样操作以及最终的全连接分类层。

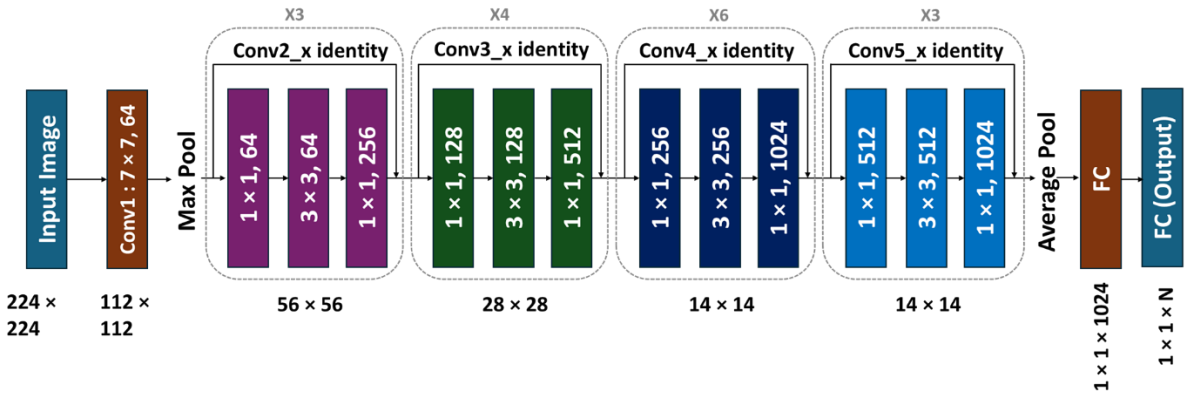


图4：ResNet-50 详细架构及修改后的顶层

DenseNet

密集卷积网络（DenseNet）由 Huang 等人（2017年）提出，引入了一种密集连接模式，其中每一层都会接收所有前序层的特征图。形式上，l^{th} 层的输入是此前所有层特征图的拼接：

$$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}]) \quad (2)$$

其中，$H_l()$表示由批归一化、ReLU 激活和卷积操作组成的复合函数。这种密集连接增强了信息流，缓解了梯度消失问题，并通过促进特征复用大幅减少了参数数量。

本研究所使用的 DenseNet-121 模型由四个密集块组成，各密集块之间由执行卷积和池化操作的过渡层分隔。最后的全局平均池化层之后连接全连接层（每层包含1024个神经元）和用于分类的 Softmax 层，如图5所示。DenseNet-121 模型包含约696万个可训练参数，比 VGG-16 少近17倍，因此在大规模分类任务中具有较高的计算效率。

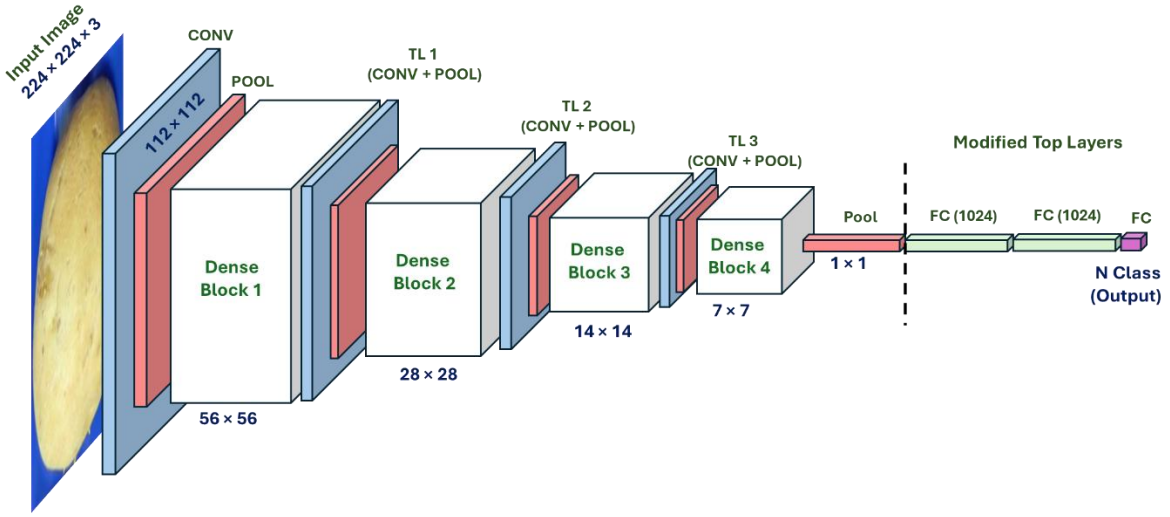


图5： 修改顶层后的 DenseNet-121 架构

视觉Transformer

受 Transformer 在自然语言处理领域卓越性能的启发，Dosovitskiy 等人提出的视觉 Transformer（ViT）（2020年）将自注意力机制应用于视觉识别任务。不同于基于卷积的架构，ViT 通过将224 × 224 输入划分为一组固定大小且互不重叠的图像块来处理图像（通常为16 × 16）。每个图像块都会被展平，并通过线性投影映射为向量表示，从而形成图像块嵌入序列。为保留空间上下文，模型会向这些嵌入中添加位置编码。此外，序列开头还会添加一个可学习的类别标记，使变换器编码器能够聚合全局图像信息，如图6所示。

ViT 的关键组件是多头自注意力机制（Vaswani 等，2017年），它使模型能够捕捉不同图像区域之间的长距离依赖。注意力运算在数学上表示为：

$$Attention\,(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

其中，Q、K 和 V 分别表示查询矩阵、键矩阵和值矩阵，d_k 表示键向量的维度。多头结构会执行多个并行的注意力计算（h 个头），并将其输出拼接后进行线性投影。每个

变换器编码器块还包含残差连接（He 等人，2016年）和层归一化，以稳定学习过程

本研究采用 ViT-B/16 配置，包括 12 个变换器编码器层、12 个注意力头（h = 12）以及 768 的嵌入维度。分类头经过定制，包含两个全连接层（每层含 1024 个神经元），后接用于 10 类预测的 Softmax 层。ViT-B/16 模型包含约 8660 万个可训练参数。本研究采用的整体网络结构如图6所示。

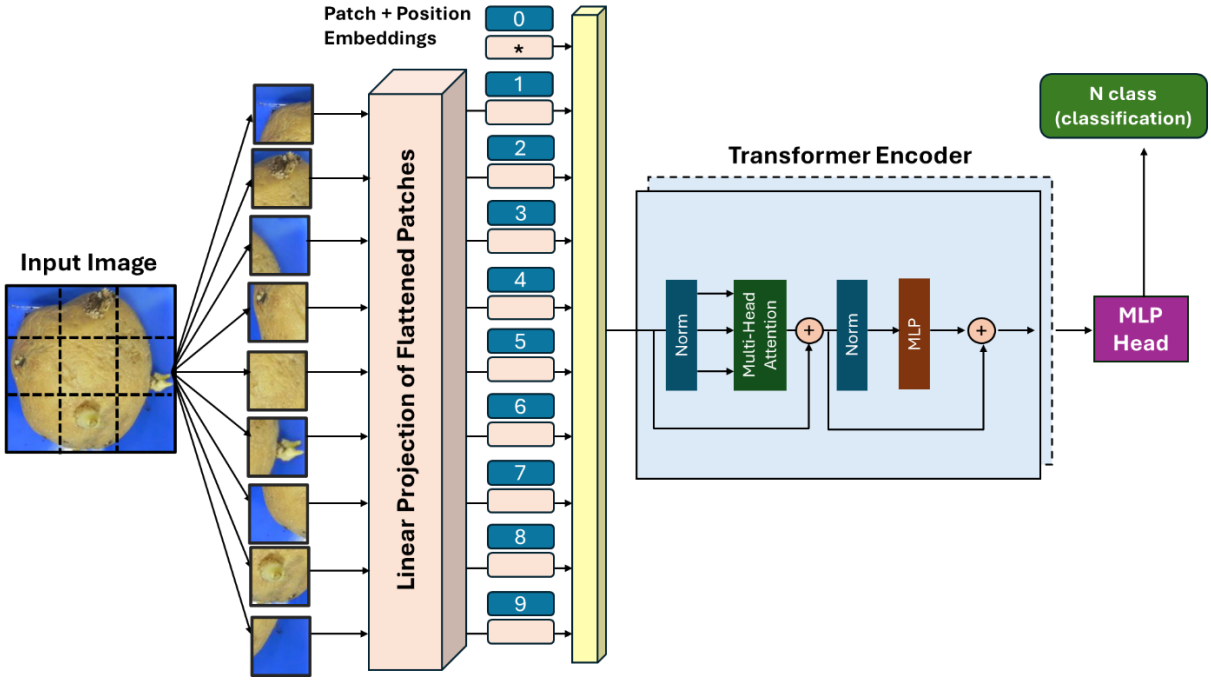


图6：视觉Transformer（ViT）详细架构

2.5 模型训练

训练过程采用了全面的策略，以确保模型具有稳健的性能和泛化能力。所有模型最多训练 500 轮，批次大小为 16，输入图像分辨率为 224 × 224 像素。CNN 架构使用 Adam 优化器，学习率为 0.001；视觉Transformer (ViT) 的学习率则为 0.0001。采用带标签平滑（ε = 0.1）的交叉熵损失，以减轻预测过度自信的问题。使用学习率调度器（ReduceLROnPlateau），将衰减因子设为 0.5、耐心值设为 30 轮；同时采用耐心值为 100 轮的提前停止，以防止过拟合。使用 GridSearchCV 进行超参数调优，并对模型

评估采用 5折分层交叉验证，计算各折的平均准确率和标准差。每一折均初始化并训练一个全新的模型，并记录准确率、F1 分数、精确率和召回率等性能指标。使用 PyTorch 的默认变换进行数据增强，包括随机缩放裁剪、水平和垂直翻转、旋转（±20°）、颜色抖动和仿射变换，同时避免使用强度过大的增强，以保持收敛稳定性。对 VGG-16、ResNet-50、DenseNet-121 和 ViT-B/16 等预训练骨干网络进行微调，并配备适配目标类别数量的自定义分类器。最终评估在每一折的测试子集上进行，并生成混淆矩阵以开展详细的误差分析。

2.6 模型性能评估

开发的基于图像的模型使用未见测试数据集进行了验证。对于所有模型，测试数据约占总数据的 20%。模型性能通过混淆矩阵进行评估，该矩阵比较模型中每个类别的正确预测数和错误预测数。记录混淆矩阵中的真正例（TP）、真负例（TN）、假正例（FP）和假负例（FN）的值，并根据下文所示的公式（公式 (3) 至 (6)）进一步估算精确率、召回率、F分数和分类准确率等统计参数:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

$$Classification\ accuracy = \frac{Number\ of\ correctly\ classified\ potatoes}{Total\ number\ of\ samples} \tag{6}$$

还使用热图评估了模型对芽分类和非芽分类的性能。热图显示模型在判定类别时关注了输入图像的哪些部分。我们从测试数据集中为两个二分类模型分别随机选取样本，并生成相应的热图。这些热图使用梯度加权类激活映射（Grad-CAM）算法生成，该算法利用预测类别得分的梯度来获取神经元重要性权重。将这些权重进行加权组合后，再经过修正线性单元（ReLU）处理，即可得到类激活图。将类激活图以热图形式呈现，可以帮助我们了解模型在分配重要性时所关注的图像区域，

以此预测图像的分类。这可能与我们所关注的马铃薯特定类型损伤所在位置相对应，例如芽或真菌斑块。

3. 结果与讨论

3.1 顶层微调

采用 VGG-16、ResNet-50、DenseNet-121 和视觉Transformer（ViT-B/16）构建分类模型。图7展示了不同顶层架构对分类准确率的影响。评估首先采用不含额外顶层的配置，记为 NoTop-4；随后测试了包含一个由 1024 个神经元组成的顶层的架构（1024-4），以及包含两层和三层顶层的架构（1024-1024-4 和 1024-1024-1024-4）。在大多数模型中，1024-4 和 1024-1024-4 等配置取得了最高准确率，而 NoTop-4 配置则使 VGG-16 的准确率大幅下降（低于 0.3），凸显了顶层微调对于有效适配特征的重要性。DenseNet-121 和 ViT-B/16 在所有配置下均表现出色，与 VGG-16 和 ResNet-50 相比，展现出更强的泛化能力和鲁棒性。

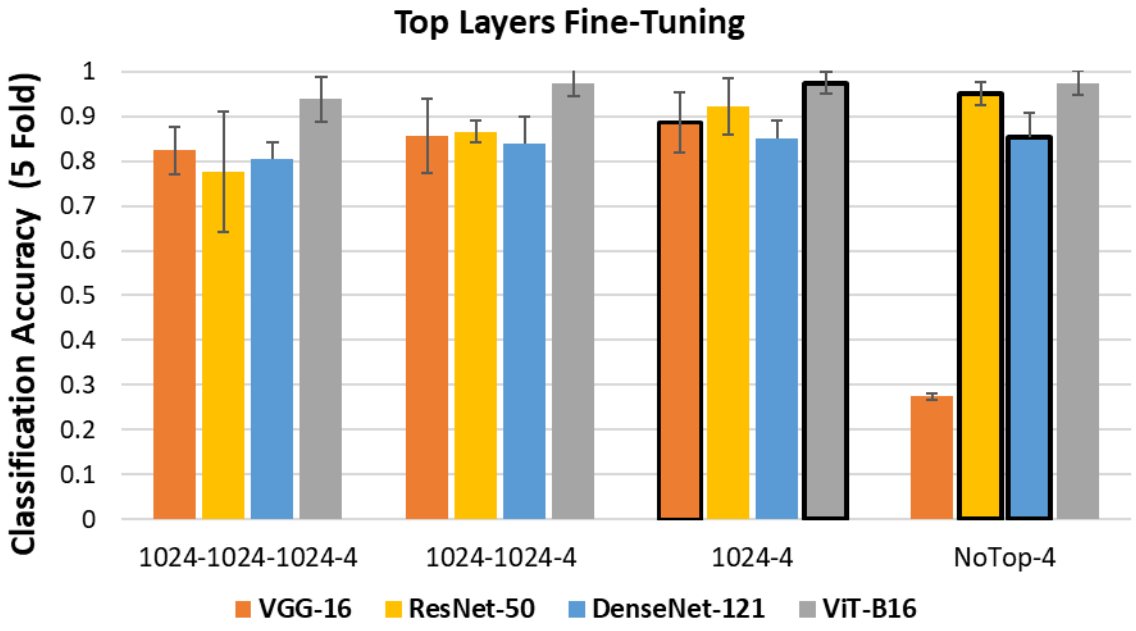
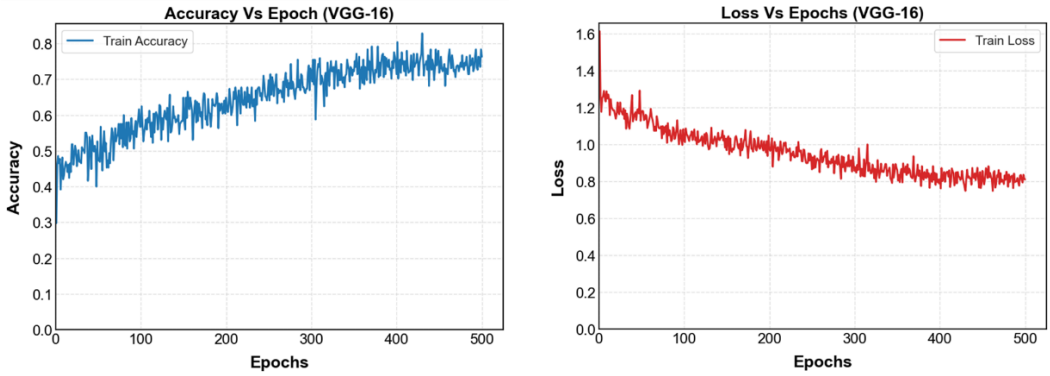
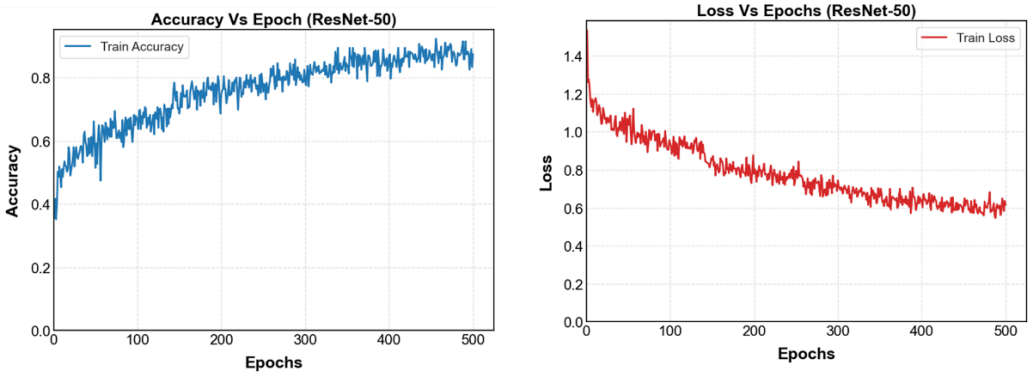


图7：顶层微调对分类准确率的影响

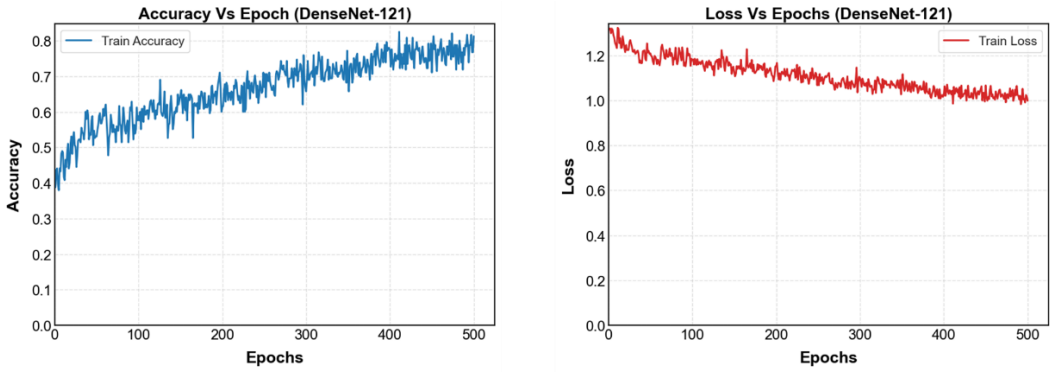
图 8 展示了模型训练和损失曲线。由于采用了学习率衰减，随着模型逐渐接近收敛，损失波动逐步减小。训练过程中采用了耐心值为 100 个训练轮次的提前停止机制，所有模型均训练了约 500 个训练轮次，以确保充分学习并减少偏差。此外，还在顶层应用了丢弃率为 0.5 的丢弃法和批归一化，以防止过拟合并促使模型正则化。还可以观察到，与其他模型相比，ViT 的收敛速度显著更快，表明其学习表征的效率更高（图 8d）。



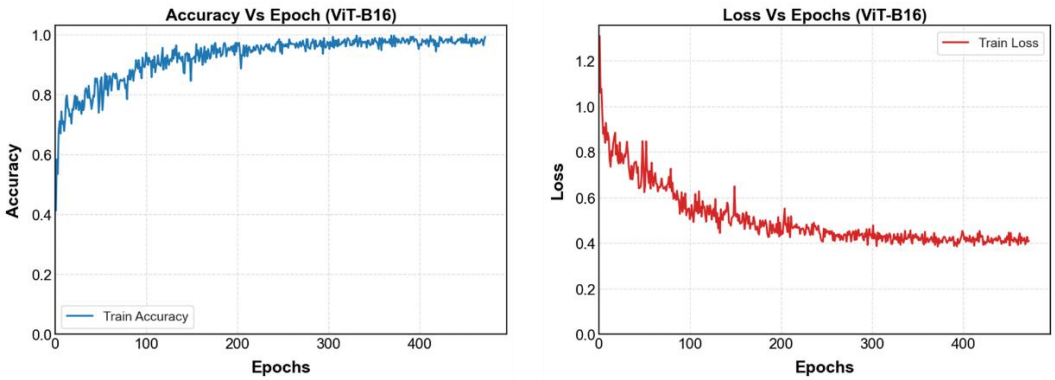
a) VGG-16



b) ResNet-50



c) DenseNet-121



d) ViT-B/16

图 8：训练历史图 a) VGG-16 b) ResNet-50 c) DenseNet-121 d) ViT-B/16

3.2 马铃薯芽分类

我们开发了一种芽分类模型，利用标注图像检测马铃薯是否发芽，图像标签为“有芽”或“无芽”。数据集中的马铃薯图像涵盖了不同的发芽阶段，从看不到芽，到局部出现少量芽，再到马铃薯表面大范围发芽。数据集包含 255 张训练图像和 51 张测试图像。我们使用单独的测试集对模型进行评估，并通过混淆矩阵评估其性能。训练过程中，我们通过各训练轮次的损失曲线和准确率曲线监控训练情况。

芽分类任务被定义为一个二分类问题，并采用四种不同的架构评估芽体检测准确率。表 2 汇总了四种模型的准确率、精确率、召回率和 F1 分数等性能指标，这些指标以五折结果的均值及其标准差表示。此外，图 9 以带有误差线的直方图展示了芽/非芽检测的总体分类准确率，直观对比各模型的性能。所有模型都取得了近乎完美的准确率，其中 DenseNet-121、ResNet-50 和 ViT-B/16 均达到 98%，VGG-16 略低，为 92%。这些结果证实，基于卷积神经网络和 ViT 的深度学习模型能够可靠地识别马铃薯是否发芽，不受芽的大小或分布影响。

DenseNet-121 和 ResNet-50 表现出色，这可归因于它们通过密集连接和残差学习捕捉细粒度局部特征的能力。ViT 的竞争力表明，基于 Transformer 的架构在建模全局依赖关系方面具有优势，这有利于检测马铃薯表面分布广泛的发芽模式。在微调受限的情况下，VGG-16 表现较差，这表明与现代深度网络相比，较早的卷积神经网络架构可能难以应对复杂的视觉变化

。微调策略对性能有显著影响，其中部分层适配取得了最佳结果。DenseNet-121 和 ResNet-50 在局部特征检测方面仍然非常有效，而 ViT 在农业图像分析中展现出巨大潜力，尤其是在全局上下文至关重要时。

表 2: 芽体分类模型架构的比较分析

评估矩阵	VGG-16		ResNet-50		DenseNet-121		ViT-B/16	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
准确率	0.9181	0.0544	0.9771	0.0187	0.9803	0.0137	0.9804	0.0179
F1 分数	0.9118	0.0624	0.9773	0.0182	0.9804	0.0137	0.9803	0.0181
精确率	0.9291	0.0432	0.9793	0.0159	0.9816	0.0126	0.9806	0.0181
召回率	0.91813	0.0545	0.9771	0.0018	0.9803	0.0137	0.9804	0.0179

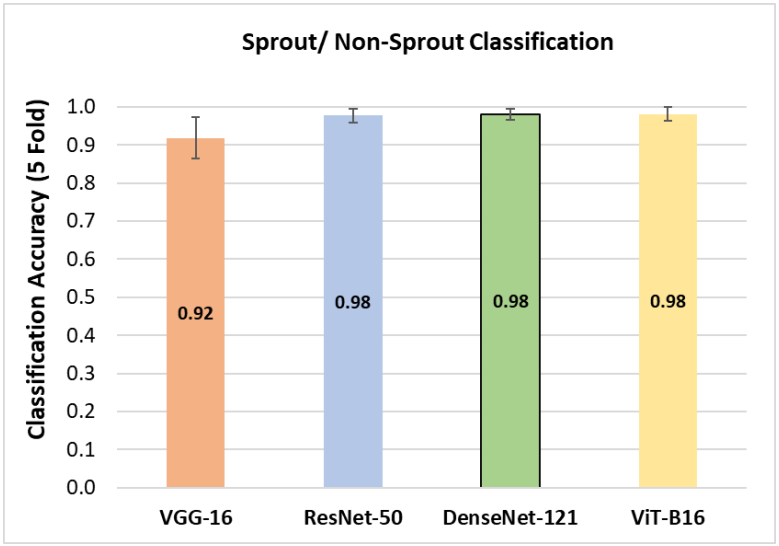


图 9：芽/非芽检测的总体分类准确率对比

混淆矩阵显示类别之间几乎实现了完美分离：模型正确识别出18/18个无芽样本和42/43个芽样本（TN=18、FP=0、FN=1、TP=42），准确率达到98.36%，平衡准确率达到98.84%（图10）。没有出现假阳性（FP=0）表明特异度达到完美，而漏检一个芽使芽的召回率为97.67%（精确率100.00%，F1为98.82%）。对于无芽类别，94.74%的精确率和100.00%的召回率（F1为97.30%）表明标签泄漏极少。马修斯相关系数为0.962，证实尽管该折存在类别不平衡（43个芽样本对18个无芽样本），模型仍具有很强的一致性。从实际应用看，错误分布主要由一个假阴性芽构成，这表明决策边界较为保守；如果应用优先考虑检出每一个芽，则可以适度调整阈值，或

成本敏感训练可以用少量假阳性的增加，换取消除这些罕见漏检。

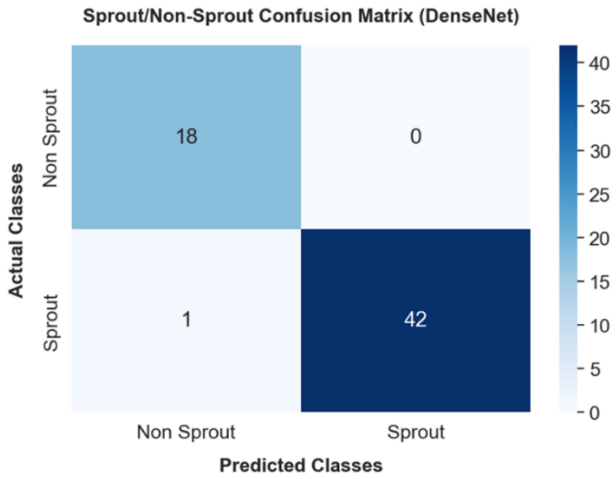


图10： 使用 DenseNet-121 进行芽/非芽分类的混淆矩阵

使用 Grad-CAM（梯度类激活图）对基于卷积神经网络的架构进行热图可视化：

Grad-CAM 热图突出显示了影响不同卷积神经网络架构对发芽马铃薯进行分类决策的图像区域。图10展示了所有训练模型对四张样本图像生成的类激活图。

在图11 (a) 中，三个卷积神经网络都准确地突出显示了芽所在区域，并据此将图像分类为“发芽”。然而，ResNet 有时会突出显示与芽无关的区域，而 VGG 生成的注意力图则更加分散。在这种情况下，DenseNet 的定位最为精准。在图11 (b) 中也观察到了类似的模式：DenseNet 将注意力集中在较大的芽上，而 VGG 和 ResNet 的注意范围更广。在图11 (c) 中，VGG 将图像误分类为“非芽”，因此生成的热图几乎一片平坦。芽的大小较小可能导致了这一错误。相比之下，DenseNet 准确地突出显示了发芽像素。在图11 (d) 中，VGG 突出显示了三个区域，其中两个对应芽，另一个对应表面不规则区域，而 ResNet 漏掉了一个芽。DenseNet 再次展现出最精准的激活效果。

总体而言，与 VGG 和 ResNet 相比，DenseNet 能够始终更准确地定位芽特征。VGG 倾向于将注意力分散到马铃薯表面，表明其特征关注不够精准。这些观察结果证实，像 ResNet 和 DenseNet 这样的深层架构

提供更出色的可解释性和特征敏感性，这对于农业质量评估中实现可靠分类至关重要。

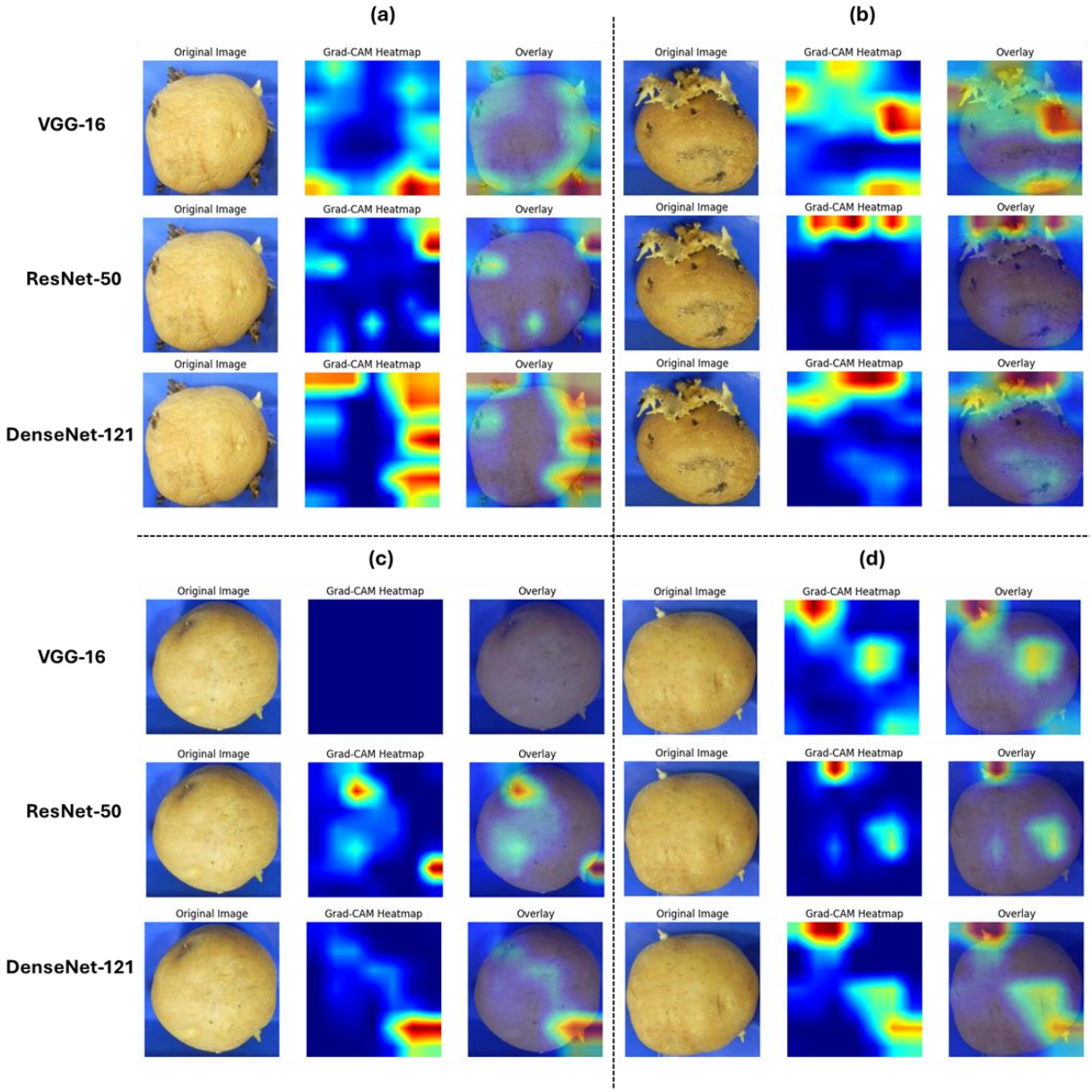


图11： 热图可视化

3.3 货架期预测（多分类模型）

在为期 200 天的储藏期内，每天都会采集每个马铃薯的图像，并为每张图像标注相应的失重百分比和剩余货架期。训练后的模型会预测给定马铃薯当前的失重和剩余货架期所属的类别（即范围）。该数据集由单个马铃薯的图像构成。5 折

采用了交叉验证，其中使用 255 张图像进行训练，51 张图像进行测试。

图 12 展示了类别数量增加对模型准确率的影响。准确率从二分类模型的 ~99% 几乎线性下降至七分类模型的 ~82%。这一降幅主要由两个因素造成：

(a) 随着类别数量增加，相邻类别之间的视觉区分度降低，使模型更难学习判别特征。

(b) 类别数量越多，每类样本越少，因为数据集总大小保持不变。

在这些架构中，ViT-B/16 在大多数类别配置下的准确率最高，其次是 DenseNet-121 和 ResNet-50；VGG-16 的表现始终落后，尤其是在类别数量较多时。这些结果表明，基于 Transformer 的模型更适合捕捉全局上下文特征，而基于 CNN 的模型在提取局部视觉线索方面仍具竞争力（例如表 3（五分类））。

多类别货架期预测在粗粒度类别划分（2–5 类）下可实现较高准确率，但随着类别划分更加精细，性能会下降。对于实际应用，类别较少的模型可能在准确率和可用性之间实现更好的平衡。

显然，随着类别数量从 2 增加到 7，准确率几乎呈线性下降，从 99.01% 降至 82.68%（图 12）。这主要有两个原因：a) 类别数量越多，相邻类别之间的视觉差异就越不明显，模型也就更难学到有意义的模式；b) 由于各模型的数据集大小基本保持不变，类别数量增加后，每类训练样本数量会随之减少。

表 3： 五分类模型架构比较分析

	五分类							
	DenseNet-121		**VGG-16**		**ResNet-50**		**ViT-B/16**	
评估矩阵	**Mean**	**STD**	**Mean**	**STD**	**Mean**	**STD**	**Mean**	**STD**
准确率	0.8918	0.0298	0.7213	0.0418	0.8131	0.0377	0.8984	0.0618
F1 分数	0.8908	0.0308	0.6911	0.0522	0.8043	0.0462	0.8975	0.0627
精确率	0.9015	0.0236	0.7074	0.0811	0.8545	0.0287	0.9045	0.0618
召回率	0.8918	0.0298	0.7213	0.0418	0.8131	0.0377	0.8984	0.0618

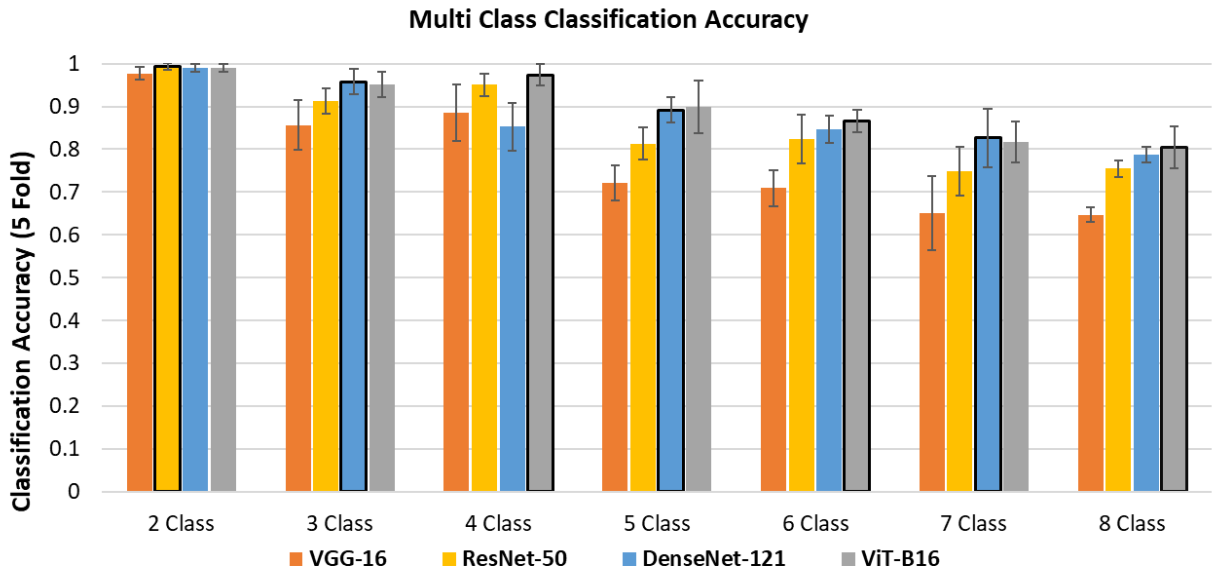
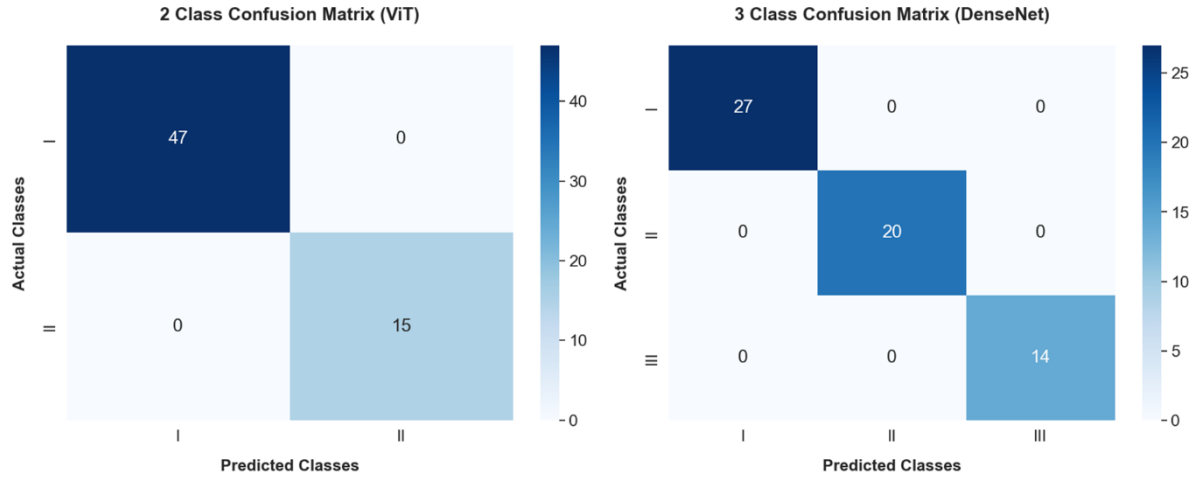


图 12：多类分类总体分类准确率对比

混淆矩阵表明，所提出的模型在各类别区间上均表现出较强的分类性能，预测结果集中在对角线上，非对角线误差极少（图13）。大多数样本都被正确分类到各自的百分比范围，反映出模型对每个类别均具有较高的精确率和召回率。若出现误分类，也仅限于相邻类别，这表明模型能够有效捕捉潜在模式，但在区分密切相关的区间时仍面临些许挑战。总体而言，结果证实了该方法在实现准确的基于百分比的分类方面具有鲁棒性和可靠性。



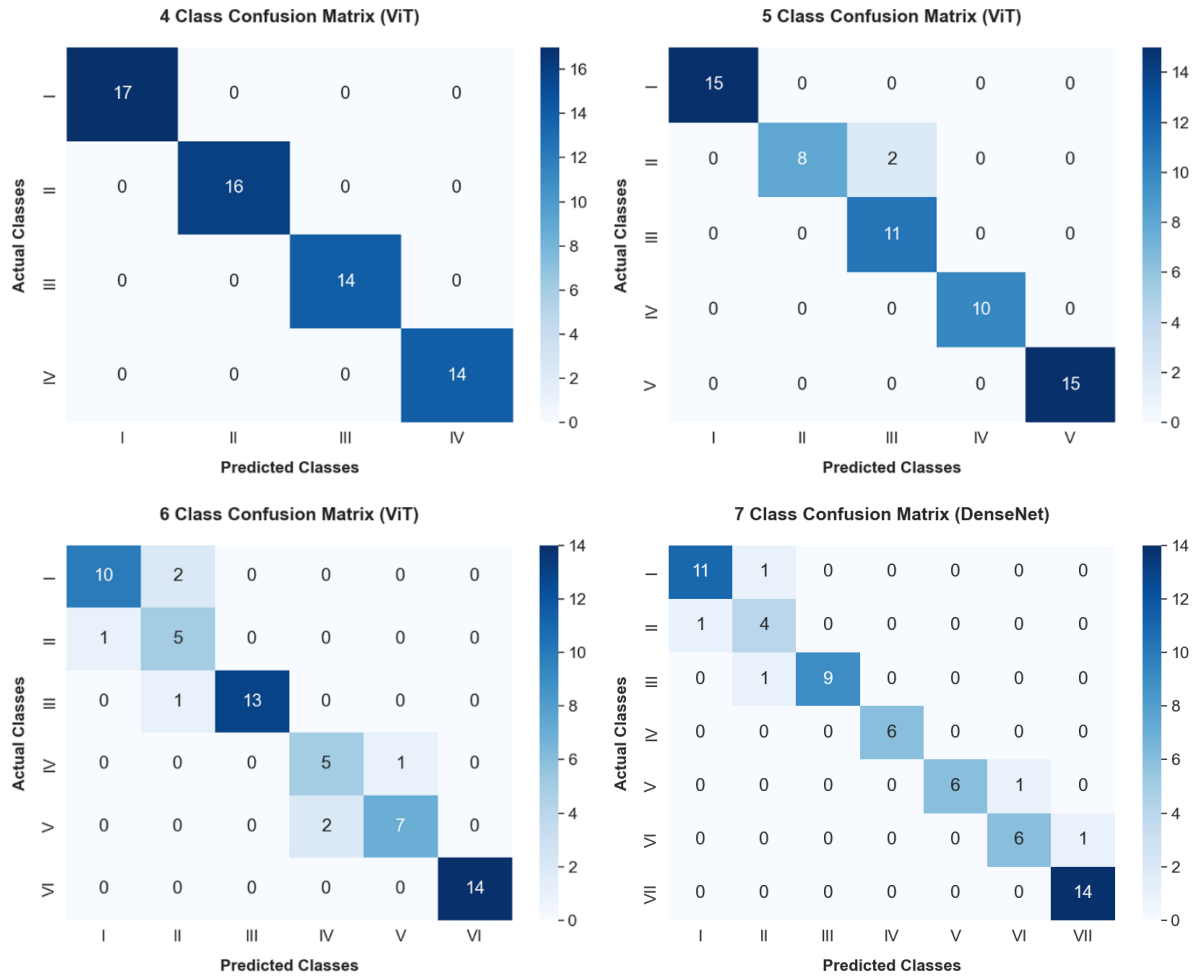


图13：2至7类分类中表现最佳模型的混淆矩阵

全球种植着众多马铃薯品种，它们在营养成分、干物质含量、形状、大小、颜色和遗传性状等方面各不相同。这些差异会影响各品种在储藏期间的表现。多项研究报告指出，不同栽培品种的失重、休眠期和发芽等参数存在显著差异（Gupta等人，2015年；Pande等人，2007年）。例如，Gupta等人（2015年）观察到，44个马铃薯品种储藏75天后的休眠期为45至75天，失重率为7.5%至17.5%。

这种差异表明，马铃薯的实际货架期可能不同于基于特定品种训练的模型所预测的货架期。因此，将本研究开发的模型应用于其他马铃薯品种时应保持谨慎。为了扩大适用范围，建议使用包含多个马铃薯品种的多样化数据集训练更加通用的模型。

3.4 计算复杂度与效率分析

表4总结了所评估模型的架构复杂度和计算性能。DenseNet-121 是参数效率最高的网络，仅包含 6.96M 个参数和 2.9 GFLOPs，但其推理时间（0.01219 秒/图像）略高于 ResNet-50 和 VGG-16。尽管 ResNet-50 包含 23.52M 个参数，但其训练时间最快（13.16 分钟），且推理延迟具有竞争力（0.00486 秒/图像），表明其在网络深度和效率之间实现了良好权衡。VGG-16 包含 14.89M 个参数，却产生了显著更高的计算成本（15.4 GFLOPs）和更长的训练时间（16.95 分钟），反映出其架构优化程度较低。ViT-B/16 的资源需求最高（86.6M 个参数、17.61 GFLOPs），训练时间也最长（19.45 分钟），这与基于 Transformer 的模型的复杂性相一致。这些结果表明，ResNet-50 和 DenseNet-121 是用于实时马铃薯芽分类的实用选择，能够在准确率、可解释性和计算效率之间实现平衡（表4）。

表4：基于计算复杂度的模型架构比较

模型	**参数量**	**GFLOPs**	**平均训练时间** **（分钟）（1 折）**	**平均推理时间** **（秒/张图像）**
DenseNet-121	6.96 M	2.9 GMac	15.72	0.01219
ResNet-50	23.52 M	4.13 GMac	13.16	0.00486
VGG-16	14.89 M	15.4 GMac	16.95	0.00191
ViT-B/16	86.6 M	17.61 GMac	19.45	0.00422

3.5 人工准确率评估

为建立模型性能基准并评估改进空间，我们在用于多类分类的同一数据集上开展了人工准确率研究。五位领域专家针对四分类场景独立标注了图像。汇总后的人类表现指标如下：准确率 = 0.9020、精确率 =0.9139、召回率 = 0.9020，F1 分数 = 0.9079。这些结果表明，尽管我们的模型 DenseNet-121、VGG-16、ResNet-50 和 ViT-B/16 在二分类中达到了很高的准确率（最高可达 99%），但在七分类中的表现降至约 82%。人工基准凸显了这项任务固有的复杂性，并为该数据集上可达到的准确率提供了现实的上限。这项比较对于解读模型表现和指导未来改进至关重要。

3.6 实际应用意义

基于图像的深度学习模型在马铃薯自动芽体检测和多类别货架期预测中取得了很高的准确率，展现出其在革新储藏和供应链管理方面的巨大潜力。将这些模型集成到分拣系统中，可以及早识别发芽土豆，减少人工检查工作，并及时决定将其出售、加工或处置，最终最大限度地减少经济损失和食物浪费。同样，采用粗粒度类别划分（2–5个类别）的货架期预测模型，可以可靠地将马铃薯划分为储藏早期、中期或晚期，从而支持动态库存管理和差异化定价策略，在为消费者优化价值的同时，降低供应商的腐损成本。基于图像的自动分类能够提升大型设施的效率和可扩展性，取代劳动密集型的人工检查。然而，类别划分更精细（6–8 个类别）时，准确率有所下降，这凸显了由于视觉差异细微且每类数据有限，精确预测时间区间所面临的挑战。因此，实际部署应着重采用较少且范围较广的类别，以确保稳健表现；未来研究则应致力于开发基于多样化马铃薯品种和储藏条件训练的通用模型，从而提升其在不同供应链中的适应性。

4. 结论

本研究证明，基于图像的模型可有效用于供应链中马铃薯的筛选和货架期预测。研究开发了两种不同的模型：芽检测模型和货架期预测模型。芽检测模型根据马铃薯是否有芽进行分类，准确率超过98%。货架期预测模型将马铃薯划分为2、3、4、5、6或7个类别，平均准确率分别达到99.34%、95.76%、97.38%、89.18%、86.60%和82.68%。结果表明，将采后生命周期划分为2、3或5个类别时，均可获得超过89%的可靠准确率。研究还发现，模型性能取决于训练所用马铃薯的差异程度以及数据集的大小，这凸显了使用大型、多样化数据集来提升泛化能力和预测准确率的必要性。

未来研究应着重将失重的视觉指标与维生素含量、pH值、干物质和酚类化合物等理化特性相结合，开发能够通过图像预测多种品质属性的综合模型。此外，该方法还可拓展至其他根茎类作物，包括

甘薯、胡萝卜、萝卜和甜菜，这些作物在储藏和运输过程中也会出现类似的品质劣变。通过针对整个根茎类作物类别开发通用模型，并考虑不同环境条件下的呼吸作用、蒸腾作用、休眠解除和机械损伤等生理过程，即可实现对多种作物的品质预测。

CRediT 作者贡献声明

Shrikant Kapse： 概念构思、数据整理、方法设计、形式分析、调查研究、验证、撰写初稿、审阅与编辑。**Priyankkumar Dhrangdhariya：**形式分析、方法设计、调查研究、撰写初稿、审阅与编辑。**Priya Kedia：**数据整理、撰写初稿、审阅与编辑。**Manasi Patwardhan：**数据整理、形式分析、方法设计。**Shankar Kausley：**调查研究、方法设计、撰写初稿、审阅与编辑。**Beena Rai：**概念构思、方法设计、监督。 **Shirish Karande：**概念构思、方法设计。

利益冲突声明

作者声明，他们不存在已知的可能影响本文所报告研究工作的财务利益冲突或个人关系。

致谢

作者衷心感谢 TCS Research 负责人 Sachin Lodha博士和塔塔咨询服务公司首席技术官 Harrick Vin博士在本项目中提供的支持。作者还感谢实验室助理 Sudam Konelu先生、Rupesh Shinde先生和Yaseen Shaikh先生协助收集内部实验数据。

参考文献

Alhammad, S. M., Khafaga, D. S., El-Hady, W. M., Samy, F. M., & Hosny, K. M. (2025). 利用深度学习和可解释人工智能对马铃薯叶部病害进行分类。*Frontiers inArtificial Intelligence, 7,* 1449329. https://doi.org/10.3389/frai.2024.1449329

Azizi, A., Abbaspour-Gilandeh, Y., Nooshyar, M., & Afkari-Sayah, A. (2016年). 使用机器视觉和人工神经网络识别马铃薯品种。*International Journal of Food Properties*，*19*(3), 618–635。https://doi.org/10.1080/10942912.2015.1038834

Dacal-Nieto, A., Vázquez-Fernández, E., Formella, A., Martin, F., Torres-Guijarro, S., & González-Jorge, H. (2009年，十一月). 一种通过计算机视觉对马铃薯进行分类的特征选择遗传算法方法。载于*2009*年 *IEEE*工业电子学第*35*届年会（第 1955–1960 页）。IEEE。https://doi.org/10.1109/IECON.2009.5414871

Danielak, M., Przybył, K., & Koszela, K. (2023年)。采用人工智能方法和成像技术对马铃薯进行无损质量评估的机器需求。*Sensors*，*23*(4), 1787。https://doi.org/10.3390/s23041787

Dhrangdhariya, P., Saini, P., Maiti, S., & Rai, B.（2025）。利用迁移学习对油漆/涂层缺陷进行多类分类。*Engineering Applications of ArtificialIntelligence, 156,* 111320。https://doi.org/10.1016/j.engappai.2025.111320

Dosovitskiy, A.（2020）。一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer。*arXiv* 预印本 *arXiv:2010.11929*。https://arxiv.org/abs/2010.11929

Ebrahimi, E., Mollazade, K., & Arefi, A.（2012年）。一种利用图像处理和人工神经网络对马铃薯块茎进行分类的专家系统。*International Journal of FoodEngineering, 8*(4)。https://doi.org/10.1515/1556-3758.2656

Gülmez, B.（2025）。马铃薯种植中基于卷积神经网络的病害检测策略综合综述。*Potato Research, 68*（2），1295–1329。https://doi.org/10.1007/s11540-024-09786-1

Gupta, V. K., Luthra, S. K., & Singh, B. P.（2015年）。印度马铃薯品种的储藏特性和烹饪品质。《食品科学与技术杂志》，*52*（8），4863–4873。https://doi.org/10.1007/s13197-014-1608-z

He, K., Zhang, X., Ren, S., & Sun, J.（2016年）。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集（第770–778页）。https://doi.org/10.1109/CVPR.2016.90

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q.（2017年）。密集连接卷积网络。载于*IEEE*计算机视觉与模式识别会议论文集（第4700–4708页）。https://doi.org/10.1109/CVPR.2017.243

Ibrahim, A., El-Bialee, N., Saad, M., & Romano, E.（2020年）。使用自动化视觉系统对马铃薯块茎进行无损质量检测。《国际先进科学、工程与信息技术期刊》，10(6), 2419–2428。

Kamilaris, A. 和 Prenafeta-Boldú, F. X.（2018年）。农业中的深度学习：综述。《农业中的计算机与电子学》，*147*，70–90。https://doi.org/10.1016/j.compag.2018.02.016

Le, T. T. 和 Lin, C. Y.（2019年）。用于成簇园艺作物无创分类的深度学习–以香蕉果实等级为例。《采后生物学与技术》，*156*，110922。https://doi.org/10.1016/j.postharvbio.2019.05.023

Nasiri, A.、Taheri-Garavand, A. 和 Zhang, Y. D.（2019年）。基于图像的深度学习在枣果自动分选中的应用。《采后生物学与技术》，*153*，133–141。https://doi.org/10.1016/j.postharvbio.2019.04.003

Pande, P. C., Singh, S. V., Pandey, S. K., & Singh, B. (2007年). 印度马铃薯（马铃薯）品种的休眠、发芽行为和失重。《印度农业科学杂志》, *77*(11)。

Pinhero, R. G., Coffin, R., & Yada, R. Y.（2009年）。马铃薯的采后贮藏。载于《马铃薯化学与技术进展》（第339–370页）。学术出版社。https://doi.org/10.1016/B978-0-12-374349-7.00012-X

Piedad, E. J., Larada, J. I., Pojas, G. J., & Ferrer, L. V. V.（2018年）。使用基于层级的机器学习对香蕉（小果野蕉）进行采后分类。《采后生物学与技术》，*145*卷，第93–100页。https://doi.org/10.1016/j.postharvbio.2018.06.004

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L.（2015年）。ImageNet大规模视觉识别挑战赛。《国际计算机视觉杂志》，*115*(3), 211–252。https://doi.org/10.1007/s11263-015-0816-y

Simonyan, K., & Zisserman, A.（2014）。用于大规模图像识别的超深卷积网络。*arXiv*预印本 *arXiv:1409.1556*。 https://arxiv.org/abs/1409.1556

Si, Y., Sankaran, S., Knowles, N. R., & Pavek, M. J.（2017年）。使用图像分析评估马铃薯块茎长宽比。《美国马铃薯研究杂志》，*94*(1), 88–93。https://doi.org/10.1007/s12230-016-9545-1

Sonnewald, S., & Sonnewald, U. (2014). 马铃薯块茎发芽的调控。*Planta,239*(1), 27–38. https://doi.org/10.1007/s00425-013-1968-z

Su, Q., Kondo, N., Al Riza, D. F., & Habaragamuwa, H. (2020). 基于深度成像和卷积神经网络的马铃薯品质分级。 《食品质量杂志》, *2020*年,8815896. https://doi.org/10.1155/2020/8815896

Su, W. H., & Xue, H. (2021). 利用成像光谱和机器学习智能判定马铃薯和甘薯品质。 《*Foods*》, *10*(9), 2146.https://doi.org/10.3390/foods10092146

Surya Prabha, D., & Satheesh Kumar, J. (2015). 通过图像处理技术评估香蕉果实成熟度。 《食品科学与技术杂志》, *52*(3), 1316–1327.https://doi.org/10.1007/s13197-013-1188-3

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I.（2017年）。《神经信息处理系统进展》，*30*，5998–6008。

Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., & Wang, L.（2025）。用于图像分类的视觉Transformer：比较综述。*Technologies*，*13*（1），32。https://doi.org/10.3390/technologies13010032

Wei, Q., Zheng, Y., Chen, Z., Huang, Y., Chen, C., Wei, Z., ... & Chen, F. (2024). 基于跨模态技术的实际在线生产中马铃薯品质无损感知。《国际农业与生物工程学报》，*16*（6），280–290。DOI：https://doi.org/10.25165/j.ijabe.20231606.8076

Wu, H., Zhu, R., Wang, H., Wang, X., Huang, J., & Liu, S. (2025). Flaw-YOLOv5s：一种基于多尺度特征融合的轻量级马铃薯表面缺陷检测算法。*Agronomy*，*15*（4），875。https://sdoi.org/10.3390/agronomy15040875