人体跌倒检测——多模态方法

Xi Wang阿尔伯塔大学多媒体学院埃德蒙顿，加拿大
xi29@ualberta.ca

Ramya Penta阿尔伯塔大学多媒体学院埃德蒙顿，加拿大
penta@ualberta.ca

Bhavya Sehgal阿尔伯塔大学多媒体学院埃德蒙顿，加拿大
bse
hgal1@ualberta.ca

Dale Chen-Song阿尔伯塔大学多媒体学院埃德蒙顿，加拿大
c
hensong@ualberta.ca

arXiv:2302.00224v1 [cs.CV] 1 Feb 2023

摘要——近年来，跌倒发生得越来越频繁，对老年人造成了危害。因此，跌倒检测变得十分重要，已有多个相关数据集和机器学习模型被提出。在本项目报告中，我们提出了一种采用多模态方法的人体跌倒检测方法。我们使用了由数十名志愿者通过不同传感器和两台摄像机采集的 UP-FALL 跌倒检测数据集。我们使用腕部传感器的加速度计数据，并将数据集标签设为二分类，即跌倒和未跌倒。我们融合摄像头数据和传感器数据，以提升性能。实验结果表明，与使用多传感器进行二分类相比，仅使用腕部数据并未影响跌倒检测模型的预测性能。

索引词—**跌倒检测、多模态、加速度计**

I. 引言

跌倒是老年人可能遭遇的严重问题，会造成严重伤害，例如髋部骨折、创伤性脑损伤以及其他伤害 [1]。如果老年人独自居住且无人照料，这些伤害可能会进一步恶化，因为他们可能长时间倒在地上而无法起身，失去意识、体温过低，甚至可能导致死亡 [2]。因此，有必要使用跌倒检测设备及时监测老年人是否跌倒，以防止此类伤害。

目前有多种系统可用于日常监测和检测人体跌倒。这些系统可分为两类：可穿戴设备和非可穿戴设备。可穿戴设备包括可附着在人体上的设备，例如加速度计 [3], 陀螺仪[4],和可监测跌倒动作的可穿戴摄像机 [5] 。这类设备的一些缺点是，持续佩戴可能笨重且不舒适，而且可能需要经常充电。此外，如果受试者需要佩戴不止一种设备，佩戴起来可能会更加不舒适。非可穿戴设备包括光学摄像机 [6], 麦克风 [7],微波雷达 [8], 或深度摄像机 [9]。然而，这些设备也存在局限性，例如摄像机容易受光照强度影响，麦克风需要安静的环境才能准确检测声音，而受雷达系统 [10]的探测范围和多普勒分辨率限制，人体动作的微波特征不够可靠。因此，减轻这些缺点的最佳方法是结合不同的系统采用多模态方法。然而，另一个挑战是确定哪些系统应结合起来实现多模态方法，因为某些系统可能与其他系统配合得更好，也可能存在冗余。

此外，跌倒检测还面临另一项挑战：需要确认对象确实发生了跌倒。设备可能会检测到某些并非真正跌倒的动作或行为。因此，设备需要区分行走、躺下和站起等常见动作，以及下蹲或从地上捡东西等常见的跌倒误报情形与真正的跌倒。

在本报告中，我们将提出一种使用特征融合和多层感知机、基于摄像头和加速度计数据检测人体跌倒的多模态方法。

II. 背景

A. 多模态方法

如上所述，有许多设备可用于检测跌倒。因此，可以通过多种方式组合这些系统，以实现多模态方法。 [11]提出了一种将深度网络应用于多个模态特征学习的新方法。他们为多模态学习提出了一系列任务，并展示了如何训练深度网络来学习用于解决这些任务的特征。他们展示了跨模态特征学习：如果在特征学习时存在多个模态（例如音频和视频），就可以学习到某一模态（例如视频）更优的特征。此外，他们还展示了如何学习模态之间的共享表示，并在一项独特的任务上对其进行评估：分类器使用纯音频数据进行训练，但使用纯视频数据进行测试，反之亦然。

整个任务分为三个阶段：特征学习、监督训练和测试。监督训练和测试采用简单的线性分类器，以利用多模态数据考察不同的特征学习模型。具体而言，他们考虑了三种学习设置：多模态融合、跨模态学习和共享表示学习。

特征学习最直接的方法之一，是分别为音频和视频训练受限玻尔兹曼机（RBMs）模型，但这种方法存在两个问题。首先，模型没有明确的目标来发现不同模态之间的相关性；模型可能会找到这样的表示：某些隐藏单元只对音频进行调谐，而另一些只对视频进行调谐。

其次，在监督训练和测试期间只存在一种模态的跨模态学习场景中，这些模型使用起来并不方便。

因此，他们提出了一种深度自编码器来解决这两个问题。他们采用了跨模态学习设置：在特征学习期间同时提供两种模态，但在监督训练和测试时仅使用单一模态。他们在监督训练和测试时仅提供单一模态的设置中使用深度自编码器模型。他们提出使用经过扩充但含有噪声的数据集训练双模态深度自编码器，其中加入了仅以单一模态作为输入的额外样本。两种模型均使用稀疏受限玻尔兹曼机进行预训练。

[12]在地球观测任务中，研究人员使用高光谱数据或激光雷达数据等单一遥感器数据时，由于难以挖掘并联合利用多模态异构数据的信息潜力，在识别和辨认感兴趣目标方面不可避免地会遇到性能瓶颈。尽管一些新近开发的方法已被证明能够有效融合多个遥感器数据源，但这些方法的能力仍然有限。为了解决这一问题，作者提出了一种名为跨通道重构模块的组件，以获得更稳健、紧凑的联合特征表示，并将其存储在编码潜向量中。

对于以往的传统模态融合方法，最直接的做法是将不同特征拼接成融合嵌入。根据特征拼接的阶段，可分为早期融合、中期融合和后期融合。然而，这些朴素方法的性能有限。为了解决这一问题，他们提出了 CCRfusion，该方法采用类似编码器-解码器的结构来获得稳健的特征表示。如详细示意图所示，卷积神经网络提取的特征主要按照预先设定的顺序简单拼接。拼接后的特征被输入编码器，以生成更紧凑的特征编码。随后，压缩后的特征编码将被进一步解码，还原为交换后的原始提取特征。对于目标函数，整体损失被分解为交叉熵损失项（用于评估分类准确率）和重建损失项（用于反映解码效果）。

来自 [13],一种能够通过多模态深度玻尔兹曼机对不同类型信息进行分类的生成模型。它提取统一表示，将各个模态融合在一起，并学习多模态输入空间上的概率密度。即使某些模态缺失，它也能通过对这些模态的条件分布进行采样并补全缺失内容，提取这种表示。针对多模态数据的学习已有多种方法。Huiskes 证明，除了标准的低级图像特征外，使用图像说明或标签，能够显著提高支持向量机（SVM）和线性判别分析（LDA）模型的分类准确率。多核 Guillaumin 提出的学习框架进一步证明，额外加入文本模态能够提高支持向量机在各种目标识别任务中的准确率。

B. 跌倒检测

一种用于跌倒检测的多模态方法是结合雷达和光学摄像机来检测跌倒。 [10] 提出采用卷积神经网络进行多传感器融合，通过结合雷达和光学摄像机来检测跌倒。对于雷达回波信号，通常使用时频（TF）域。跌倒以及三种常见动作——行走、下蹲和站起——的时频信号。作者提出采用基于 Alex 的卷积神经网络（CNN）和基于单次多边界框检测器（SSD）的卷积神经网络，分别对时频图像进行分类，并将两个卷积神经网络的结果合并，以得到最终检测结果。另一方面，光学摄像机拍摄的图像序列也由另一个基于 SSD 的卷积神经网络处理，并提供人体边界框的纵横比序列，以帮助确认是否发生跌倒事件。逐帧变化的边界框纵横比（高度除以宽度）用于帮助区分跌倒事件与非跌倒事件。

[14]旨在利用 MobiFall 数据集，并借助机器学习算法识别和区分跌倒活动与日常生活活动（ADL）。他们使用加速度计、陀螺仪和磁力计采集身体运动数据。所实施的五种分类方法是：朴素贝叶斯、k 近邻、人工神经网络和最小二乘法

跌倒检测分类系统分为五个阶段：预处理、特征提取、特征选择、模型训练和分类。数据经过中值滤波器和低通滤波器预处理。在特征提取阶段，共提取了 38 个特征，并使用基于滤波器排序的系统剔除不含信息的特征，提取排名前五的特征，以优化算法的维度。（z 中值和 x 中值为时域特征……x 均值和 y 均值为频域特征……信号幅度向量的偏度）见图 1。受试者工作特征曲线表明



图 1：跌倒检测方法

在 5 种分类器中，k 近邻算法的总体准确率为 87.5%，灵敏度为 90.70%，特异度为 83.78%。灵敏度表示检测跌倒的能力，特异度表示仅检测跌倒并忽略非跌倒事件的能力，而准确率表示总体中真实结果所占的比例。

本文的局限性在于，“MobiFall”数据集是在用户口袋中采集的，位置低于腰部；但研究发现，将设备放置在身体躯干上部、颈部以下的位置时效果最佳。这是因为其他位置的运动频率较高，且运动情况更为复杂。

[15]重点研究基于深度学习方法的跌倒检测。该方法使用的数据是从可穿戴设备采集的加速度计数据。模型部署在雾计算设备上；在本例中，作者使用树莓派进行实验。作者还提出了一种名为 CNN-3B3Conv 的卷积神经网络架构，由 3 个模块组成。最终，该方法的准确率达到 99.86%，而纯 LSTM 模型的准确率仅约为 95%。该架构使用 3 个模块。模块 1 和模块 2 的结构相同，各包含 3 个卷积层。模块 3 则包含 3 个全连接层。最终，输出为一个二元标签，用于表示用户是否正在跌倒。

III. 提出的方法

A. 数据集

我们使用泛美大学发布的 UP-FALL跌倒检测数据集 [16] 。该数据集用于人类活动识别，主要旨在检测老年人的跌倒。数据集包含 17 名受试者的数据；每名受试者进行 11 项活动，每项活动试验三次。这些活动包括六种简单的人类日常活动（行走、站立、捡起物体、坐下、跳跃和躺下）以及五种人体跌倒（用手向前跌倒、用膝盖向前跌倒、向后跌倒、向后跌坐在空椅子上以及侧向跌倒）。用于采集数据的设备包括两台摄像机（一台安装在受试者前方，另一台安装在受试者侧面）；12 个红外传感器，分布在测试场景周围，如图 2 所示；一个脑电波传感器（脑电图头盔）；以及六个佩戴在身体不同部位、用于检测加速度、角速度和亮度的可穿戴设备。该数据集包含两类数据：一类是摄像机拍摄的图像数据，另一类是存储在 CSV 文件中的传感器数据。该数据集收集的用于检测人类行为的数据非常全面。原始标签不仅包括跌倒或未跌倒，还包括站立、躺下等其他静态行为标签。

B. 预处理

本项目的主要目标是进行二元分类，判断是否发生跌倒。数据集的原始标签除了标记是否跌倒外，还包含其他行为标签，这些标签在当前阶段似乎是多余的。在预处理过程中，我们保留跌倒标签不变，并将其余所有标签设为未跌倒。此操作旨在匹配实验目标，即逐帧判断是否发生跌倒。此外，我们只使用了17名受试者中的3名。另外，考虑到真实世界使用的情况，尽管原始数据集中收集的数据很全面，但让老年人在全身佩戴多达六个可穿戴传感器来检测跌倒并不现实。

因此，我们将传感器数量从六个减少到一个，只保留手腕上的加速度计。这是因为将传感器嵌入智能手表的技术已相当成熟并实现了全面商业化，而且佩戴手表式设备对佩戴者日常生活的影响极小，如图 3 所示。我们还决定不使用该传感器的角速度和光度数据，只保留加速度计数据。我们采集了 x、y 和 z 方向上的原始加速度计数据，构成一个三维向量。我们通过以下公式计算原始加速度计数据的幅值：

$$SV_{total} = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{1}$$

其中， A_x、 A_y 和 A_z 表示三个轴向上的原始加速度计数据， SV_{total} 表示计算所得的幅值。其余传感器（脑电波传感器和红外传感器）也被移除，因为在实际场景中部署专用设备并不可行。最后，我们获得了预处理后的传感器数据，即由佩戴在手腕上的加速度计采集的 x、y 和 z 方向加速度及其幅值。

图像（帧）被转换为灰度图，并调整大小为 32×32。图4展示了带有修正后标注的预处理后的图像的可视化效果。

C. 多模态方法

我们选择了一种直观而简洁的多模态特征融合方法：直接拼接两组不同的特征。首先，我们使用卷积神经网络提取图像数据和传感器数据的特征图，然后将不同的模态特征展平并拼接起来。拼接后的特征嵌入被输入由全连接层构成的多层感知机（MLP）

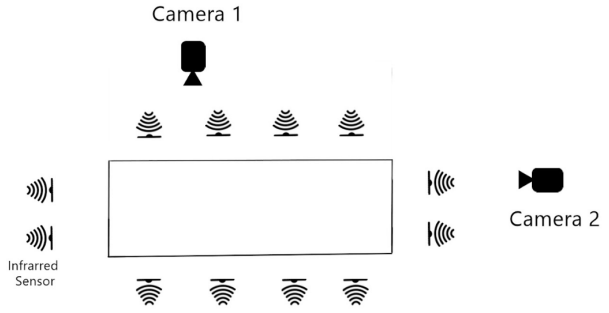


图2：UP-FALL数据集实验中的传感器布局

表 I：评估　　传感器、摄像头 1、摄像头 2、摄像头 1 和 2 的融合以及多模态融合的评估指标　　n

	传感器	摄像头 1	摄像头 2	摄像头 1 和 2	多模态融合
准确率	96.47	99.26	99.45	99.60	**99.72**
精确率	96.47	99.26	99.45	99.60	**99.72**
召回率	96.47	99.26	99.45	99.60	**99.72**
F1 分数	95.62	99.24	99.45	99.59	**99.72**

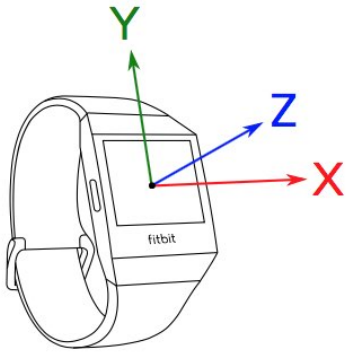


图 3：加速度计示例，显示 FitBit 的 x、y、z 轴



图 4：经预处理的灰度 32×32 图像，来自摄像机 1 和 2，并以二元标签标注是否发生跌倒。

层。最后一层通过 Softmax 函数激活，以输出相应的预测结果。

1）传感器数据：对于传感器数据，腕部加速度计虽然在每个样本（时间戳）中包含四个数据，即 A_x、A_y、A_z 和 SV_{total}，但信息的维度是单一的。我们首先尝试使用多层感知机进行特征提取，以单模态方式预测是否发生跌倒。此外，卷积神经网络具有良好的特征提取能力。一维卷积神经网络可用于一维数据。因此，我们也将传感器数据输入一维卷积神经网络，以进行单模态预测。结果表明，一维卷积神经网络的表现更好。由于我们将对图像数据使用二维卷积神经网络，因此对传感器数据采用相同的运算准则，显然更有利于特征融合。

2) 图像数据：多项研究和应用已证明，使用卷积网络提取图像数据的特征效果更佳。因此，我们采用二维卷积神经网络提取图像数据的特征。对于视频流中的每一帧，我们将 RGB 图像转换为单通道灰度图，然后将其调整为 32×32 尺寸，以便输入卷积神经网络。

3）特征融合：对于传感器数据，我们将其输入一个卷积块，该卷积块由一维卷积层、一维最大池化层和批归一化层组成，然后将其展平为一维嵌入。对于两组图像数据，我们首先将其输入一个由二维卷积层、二维最大池化层和批归一化层组成的卷积块，然后将其展平为一维嵌入。对于三组展平后的特征嵌入，我们将它们拼接在一起，并输入一个由两个全连接层和一个用于避免过拟合的丢弃层组成的多层感知机。最后一个全连接层将嵌入维度缩减为二，随后使用 Softmax 函数激活这个二维嵌入。得到的概率分数将转换为输出预测结果。

IV. 评估

A. 分类能力

我们分别评估了单模态传感器数据、单模态图像数据和多模态融合方法的结果。单模态图像数据已经融合了从摄像头 1 和摄像头 2 的视频流中提取的特征。作为评估指标，我们选择了准确率和 F1 分数。准确率是分类任务中最基本的指标，而 F1 分数结合了召回率和精确率，能够对分类模型进行更全面、更直观的评估。准确率、精确率、召回率和 F1 分数可通过以下公式计算：

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{2}$$

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

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

$$F1score = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

其中，TP、FP、TN 和 FN 分别表示真正例、假正例、真负例和假负例。评估指标的可视化结果见图7，详细的数值比较见表I。

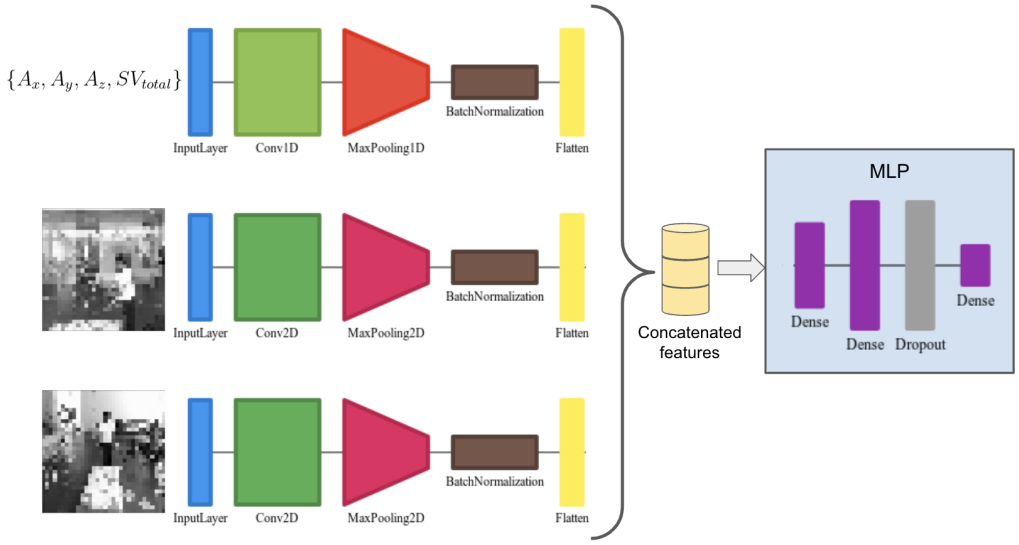
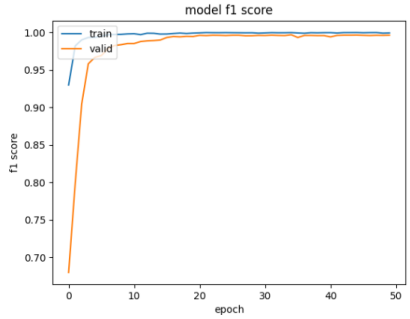
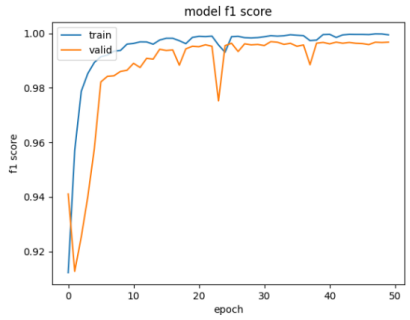


图5：多模态方法。



(a) 单传感器腕部加速度计在 50 个训练轮次中的 F1性能评估



(b) 多传感器 F1 分数评估，在 50 个训练轮次中的表现

图6：单传感器和多传感器的 F1分数可视化

B. 数据清洗的可行性

由于我们在数据预处理期间减少了传感器的数量，因此我们尝试在实施我们的方法后，使用原始传感器数据训练二元预测模型，以确保清洗操作确实不会影响跌倒检测模型的二元预测性能。两种方法的 F1 分数如图6所示，可以看出，两种方法在验证集上的最佳F1分数几乎没有差异。尽管从图中可以看出，使用多个传感器的模型表现更稳定，且在训练期间收敛更快，但单传感器模型的波动更大。为进一步验证数据细化操作的可行性，我们还在测试集上对其进行了评估。数值结果见表II。在测试集上，我们的模型甚至略占优势。我们可以得出结论，传感器数据的清理并未影响模型的分类性能。

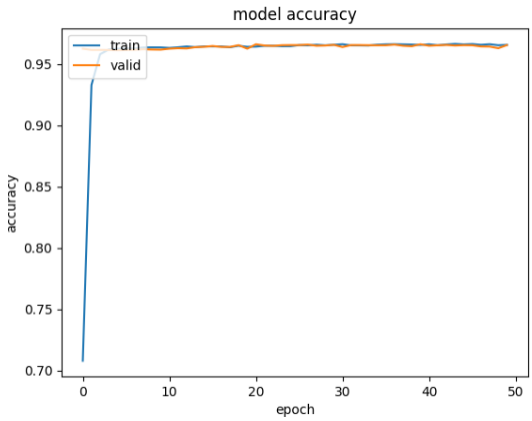
V. 讨论与结论

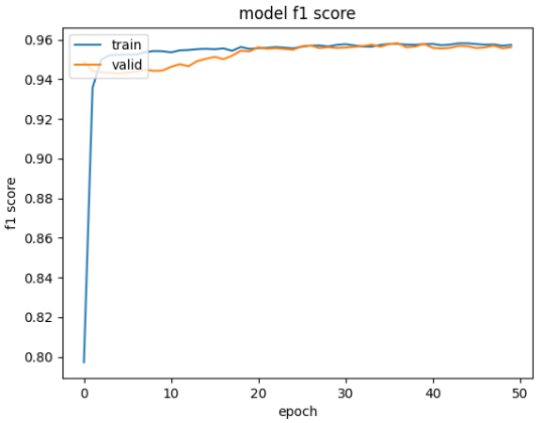
A. 结论

总之，我们清理了数据集，将受试者人数从17人减少到3人，并保留了11种活动和3次试验。传感器数量从6个减少到仅使用腕部传感器；同时，我们剔除了角速度和亮度，仅使用加速度计进行预测。所有标签均简化为两个，即跌倒和未跌倒。对于多模态方法，我们拼接了两组不同的特征用于跌倒检测。此外，我们融合了传感器数据和图像数据以提升性能。我们将这些数据输入具有全连接层的多层感知器，并使用 Softmax 函数激活最后一层。通过这些实验，我们取得了99.72%的准确率、精确率、召回率和 F1 分数，相比之下，多传感器数据的结果为99.68%。

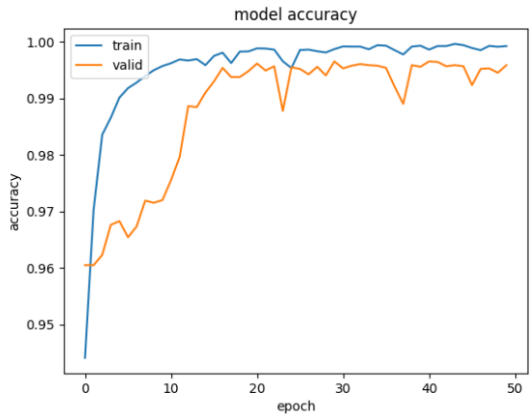
表II：单传感器（手腕）与多传感器的评估指标

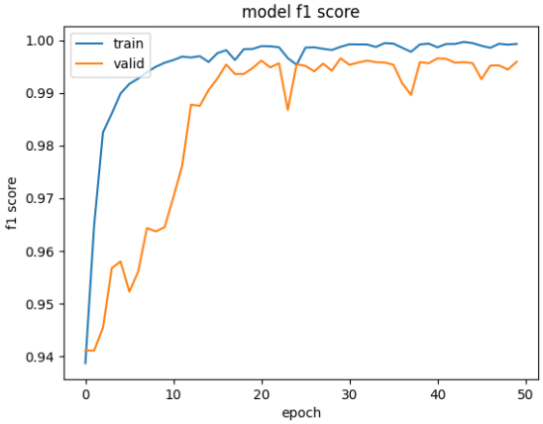
	准确率	精确率	召回率	F1 分数
单传感器（手腕）	**99.72**	**99.72**	**99.72**	**99.72**
多传感器	99.68	99.68	99.68	99.68



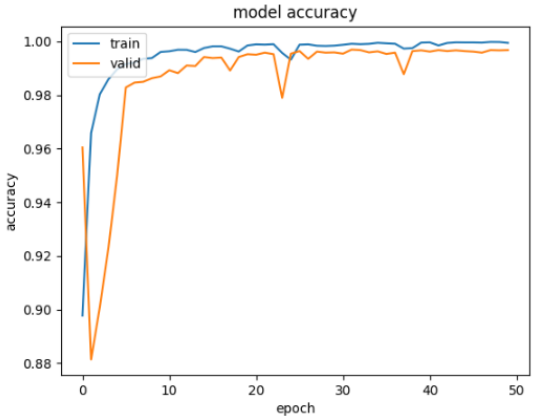


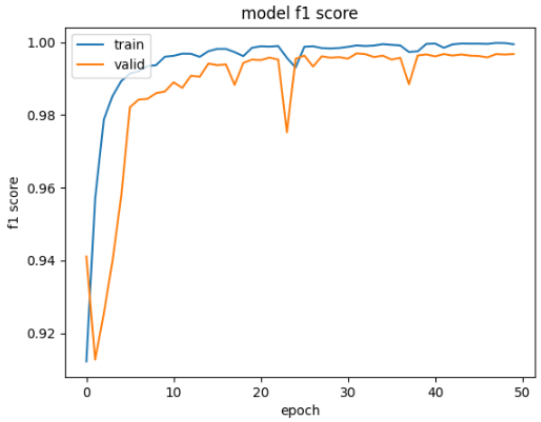
(a) 腕部传感器加速度计数据单模态准确率表现（50 个训练轮次） (b) 腕部传感器加速度计数据单模态 F1 分数表现（50 个训练轮次）





(c) 摄像头1和2数据单模态准确率表现（50 个训练轮次） (d) 摄像头1和2数据单模态 F1 分数表现（50 个训练轮次）





(e) 多模态融合准确率表现（50 个
训练轮次

(f) 多模态融合 F1 分数表现（50 个
训练轮次

图7：单模态和多模态的选定指标可视化。

B. 未来工作

未来我们可以通过多种方式改进这个项目。例如，未来我们将更加关注特征提取，以进一步了解与跌倒相关的情况。我们将重点尝试使用循环神经网络或长短期记忆神经网络，以提升跌倒检测问题的性能。为此，我们计划将帧划分为时间窗口，并进行时间序列预测。此外，在获得时间窗口后，我们还将扩展特征维度。我们发现，与结合所有数据相比，仅使用腕部数据并未改变准确率；不过，我们计划在未来尝试不同的类别组合，以提高模型性能。

参考文献

[1] P. Kannus、M. Palvanen、S. Niemi 和 J. Parkkari，“老年人跌倒所致颈椎损伤的数量和发生率惊人上升”，*老年学杂志：A辑*，第 62 卷，第 2 期，第 180–183 页，2007 年 2 月。 [在线]。可访问：https://doi.org/10 1093 62 2 180 。/gerona/。。M. E. Tinetti、[2] W.-L. Liu 和 E. B. Claus，“老年人跌倒后无法起身的预测因素及预后”，《美国医学会杂志》，第 269 卷，第 1 期，第 65–70 页，1993 年。[3] F. Bagala、C. Becker、A. Cappello、L. Chiari、K. Aminian、J. M. Hausdorff、W. Zijlstra 和 J. Klenk，“基于加速度计的跌倒检测算法在真实场景中的跌倒事件上的评估”，《公共科学图书馆·综合》，第 7 卷，第 5 期，文章编号 e37062，2012 年。[4] Q. T. Huynh、U. D. Nguyen、L. B. Irazabal、N. Ghassemian 和 B. Q. Tran，“基于加速度计和陀螺仪的跌倒检测算法优化”，《传感器杂志》，2015 年，第 2015 卷，2015 年。[5] K. Ozcan、A. K. Mahabalagiri、M. Casares 和 S. Velipasalar，“利用可穿戴嵌入式智能摄像机自动检测跌倒并进行活动分类”，《*IEEE*电路与系统新兴及重点主题期刊》，第 3 卷，第 2 期，第 125–136 页，2013 年。[6] E. Auvinet、L. Reveret、A. St-Arnaud、J. Rousseau 和 J. Meunier，“使用多摄像机检测跌倒”，载于*2008*年第*30*届*IEEE*医学与生物学工程学会年度国际会议。IEEE，2008 年，第 2554–2557 页。[7] Y. Li、K. Ho 和 M. Popescu，“用于自动跌倒检测的麦克风阵列系统”，《*IEEE*生物医学工程汇刊》，第 59 卷，第 5 期，第 1291–1301 页，2012 年。[8] L. Liu、M. Popescu、M. Skubic、M. Rantz、T. Yardibi 和 P. Cuddihy，“基于多普勒雷达运动特征的自动跌倒检测”，载于*2011*年第五届普适计算技术与医疗保健国际会议（*PervasiveHealth*）及研讨会。IEEE，2011 年，第 222–225 页。[9] Z.-P. Bian、J. Hou、L.-P. Chau 和 N. Magnenat-Thalmann，“基于深度摄像机身体部位跟踪的跌倒检测”，《*IEEE*生物医学与健康信息学期刊》，第 19 卷，第 2 期，第 430–439 页，2014 年。[10] X. Zhou、L.-C. Qian、P.-J. You、Z.-G. Ding 和 Y.-Q. Han，“利用融合多传感器的卷积神经网络进行跌倒检测”，载于*2018*年*IEEE*多媒体与博览会国际会议研讨会（*ICMEW*）。IEEE，2018 年，第 1–5 页。[11] J. Ngiam、A. Khosla、M. Kim、J. Nam、H. Lee 和 A. Y. Ng，“多模态深度学习”，第 689–696 页，2011 年。[12] X. Wu、D. Hong 和 J. Chanussot，“用于多模态遥感数据分类的卷积神经网络”，《*IEEE*地球科学与遥感汇刊》，第 60 卷，第 1–10 页，2022 年。[13] N. Srivastava 和 R. R. Salakhutdinov，“利用深度玻尔兹曼机进行多模态学习”，载于神经信息处理系统进展，F. Pereira、C. Burges、L. Bottou 和 K. Weinberger 编，第 25 卷。Curran Associates, Inc.，2012 年。 [在线]。可访问：https://proceedings.neurips.cc/paper/2012/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf[14] P. Vallabh、R. Malekian、N. Ye 和 D. C. Bogatinoska，“使用机器学习算法进行跌倒检测”，第 1–9 页，2016 年。

[15] G. L. Santos、P. T. Endo、K. H. d. C. Monteiro、E. d. S. Rocha、I. Silva 和 T. Lynn，“使用卷积神经网络进行基于加速度计的人体跌倒检测”，*Sensors*，第 19 卷，第 7 期，2019 年。 [在线]。可访问：https://www.mdpi.com/1424 8220 19 7 1644 - / / / L. Martınez-Villasenor、H. [16] ´ ˜ Ponce、J. Brieva、E. Moya-Albor、J. N´u˜nez- Mart´ınez 和 C. Pe˜nafort-Asturiano，“UP-FALL跌倒检测数据集：一种多模态方法”，*Sensors*，第 19 卷，第 9 期，文章编号 1988，2019 年。