《生物医学计算机方法与程序更新》9 (2026) 100227

目录 列表 可在 ScienceDirect 计算机 方法 与 程序 于 生物医学 更新 期刊 主页:
www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update

关于 使用 混合 LSTM-KAN 架构 对 不平衡 数据 集 进行 呼吸 声音 分类 的 研究

Nithinkumar K.V., Anand R. *
电气 与 电子 工程 系， *Amrita* 工程 学院， *Amrita Vishwa Vidyapeetham*，哥印拜陀，*641112*，印度

文 章 信 息

数据集 链接:
https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

关键词：呼吸 音 分类 深度 学习
长短期记忆网络
Kolmogorov-Arnold 网络 （
KAN）类别 不平衡 SMOTE 焦点
损失 听诊 ICBHI 数据集

摘要

通过听诊采集的呼吸音包含诊断肺部疾病的重要线索。对这些声音进行自动分类，既要区分细微的声学模式，也要应对临床数据集中固有的严重类别不平衡问题。本研究探讨了将呼吸音分为多个疾病类别的方法，重点关注如何缓解显著的类别不平衡。在本研究中，我们开发并评估了一种混合深度学习模型，该模型采用长短期记忆网络（LSTM）作为特征序列编码器，并使用 Kolmogorov–Arnold 网络（KAN）进行分类。该架构结合了全面的特征提取流程和针对性的不平衡缓解技术。我们使用一个公开的呼吸音数据库对模型进行评估，该数据库包含六个类别，且类别分布高度偏斜。研究采用焦点损失、针对特定类别的数据增强以及合成少数类过采样技术（SMOTE）等策略，以提升少数类识别能力。结果表明，所提出的 LSTM-KAN 混合模型总体准确率达到 94.6%，宏平均 F_1-分数为 0.703。考虑到占主导地位的类别（慢性阻塞性肺疾病）占数据总量的 86% 以上，这一表现尤为突出。尽管对最稀有类别（细支气管炎和上呼吸道感染）的识别仍面临挑战，其 F_1-分数分别约为 0.45 和 0.44，但与简单基线相比，该方法显著提升了对这些类别的检测效果，并且在其他少数类（如支气管扩张，F_1-分数为 ≈ 0.84）上也表现出色。本研究有助于开发用于早期检测呼吸系统疾病的智能听诊工具，并凸显了将循环神经网络与先进的 KAN 架构相结合、同时采用针对性类别不平衡处理方法的潜力。

1. 引言

呼吸系统疾病构成了全球重大的健康负担，在世界范围内对发病率和死亡率造成了显著影响 [1]。慢性阻塞性肺疾病（COPD）、哮喘、支气管扩张，以及肺炎和细支气管炎等各种感染性疾病影响着数百万人的健康 [2]。因此，早期且准确的诊断对于有效治疗和患者管理至关重要。在临床环境中，肺音听诊仍是基础性的诊断实践 [3]，，但其高度依赖临床医生的经验和专业知识，因而会引入主观性和局限性，尤其是在远程监测和远程医疗场景下 [4]。

人工智能 和 信号 处理 领域 的 最新 进展 推动 了 利用 声学 生物标志物 开发 自动化 诊断 系统 。基于 声音 的 诊断 方法 在 各 种 医疗 应用 中 展现 出巨大潜力 。例如，[5] 探讨 了 用于 诊断 流程 的 声音数据分析，证明了声学特征提取在医疗情境中的有效性。同样，[6] 介绍了处理和分析医疗声音数据的全面方法，突显了机器学习技术在呼吸声音分类中的潜力。这些研究凸显了基于声音的生物标志物在现代医疗诊断中的重要性日益提升。

尽管取得了这些进展，自动呼吸音分类系统仍面临若干严重挑战，限制了其在临床中的部署。首先，大规模且标注完善的数据集稀缺，制约了稳健深度学习模型的训练。其次，真实世界的数据集往往存在显著的类别不平衡 [7]，，多数样本属于常见疾病（如慢性阻塞性肺疾病），而罕见但具有重要临床意义的疾病（如细支气管炎、上呼吸道感染）则严重代表不足。这种不平衡会使模型偏向多数类，导致对较罕见且往往十分危重的疾病检出效果不佳 [8]。第三，传统机器学习方法难以捕捉呼吸音中固有的复杂时序与频谱模式。

* 通讯 作者：电子邮件 地址： r_anand2@cb.amrita.edu （Anand R.）。

https://doi.org/10.1016/j.cmpbup.2025.100227收稿日期 2025 年 6 月 16 日； 修回日期 2025 年 12 月 5 日； 接收日期 2025 年 12 月 20 日
在线发布于 2025 年 12 月 26 日 2666-9900/© 2025 作者。由爱思唯尔出版。本文为开放获取文章，遵循知识共享署名-非商业性使用-禁止演绎许可（http://creativecommons.org/licenses/by- nc-nd/4.0/）。

第四， 现有 深度 学习 架构 （如 卷积 神经 网络 （卷积神经网络） 和 循环 神经 网络 （RNN）），尽管 效果 良好， 却 往往 缺乏 可解释性 ——这 是 临床接受度的一项关键要求。

为应对这些挑战，本研究提出了一种新颖的混合深度学习架构，将长短期记忆网络（LSTM）与 Kolmogorov–Arnold 网络（KAN）相结合。与采用固定激活函数的传统多层感知机（MLP）不同，KAN网络在网络边上使用可学习的单变量样条函数，从而增强函数逼近能力并提高可解释性 [9]。长短期记忆网络组件捕捉提取出的声学特征中的时间依赖关系，而 KAN 后端则提供强大的非线性分类能力，并有望实现对已学习变换的可视化。该架构专门针对长短期记忆网络输出的潜在特征表示进行设计，并利用 KAN 基于样条函数的结构，更有效地区分细微的呼吸模式。

此外，我们采用了一套全面的不平衡缓解技术，包括使用焦点损失 [10] 降低分类正确样本的权重、通过类别专属的数据增强生成少数类的合成变体，以及对提取的特征向量应用合成少数类过采样技术（SMOTE）[11] 。将这些技术整合到两阶段训练策略中（先在平衡子集上预训练，再在完整数据集上微调），为应对严重的类别不平衡提供了一种系统化方法。

研究 问题： 本 研究 探讨 以下 关键 研究 问题 ：

1. 与传统深度学习方法相比，混合 LSTM-KAN 架构如何提升不平衡呼吸音数据集的分类准确率？2. 结合焦点损失、类别特定数据增强和 SMOTE，在缓解呼吸音分类中的严重类别不平衡方面效果如何？3. 基于 KAN 的分类器能否为所学习的特征变换提供可解释的见解，以辅助呼吸系统疾病诊断？4. 在真实临床环境中部署所提出的混合架构，其计算成本和实际可行性如何？

创新 贡献： 本 工作的 主要 贡献 如下 ：

1. 创新架构：首次将混合 LSTM-KAN 架构应用于呼吸音分类，利用 KAN 基于样条的可学习函数，提升对 LSTM 编码特征的判别能力。2. 全面的不平衡处理：在两阶段训练框架中系统整合焦点损失、类别特定数据增强和 SMOTE，专门针对高度不平衡的医学数据集进行设计。3. 提升可解释性：通过可视化学习到的样条函数，展示 KAN 在可解释性方面的优点，并揭示与呼吸系统疾病相关的声学特征变换。4. 临床验证框架：在 ICBHI 数据集上进行详细的性能分析，并专门选用适用于不平衡分类的评估指标，包括逐类别分析、混淆模式和校准评估。

本文其余部分安排如下：第 2 节综述呼吸音分类和类别不平衡技术的相关研究。第 3 节介绍数据集、预处理、特征提取、所提出的混合 LSTM-KAN 架构以及训练策略。第 4 节展示实验结果、消融研究，以及 计算 成本 分析。 第 5 节 讨论 研究 发现、 可解释性 方面、 临床 意义 以及 未来 工作， 包括 在 现实 场景 中的 部署 可能性。第6节总结了本文。

2. 相关 工作

肺音自动分类研究已从采用手工设计特征和传统分类器，逐步发展到深度学习[12]。应用于频谱图的卷积神经网络（CNN）[13] 以及用于建模时间动态的循环神经网络（RNN），例如长短期记忆网络（LSTM）单元，均展现出良好的应用前景[14]。基于 Transformer 的模型也已应用于呼吸音分析，并展现出具有竞争力的性能[15]。

基于声音的诊断系统已在多个医学领域得到广泛关注。[5] 针对诊断流程中的声音数据分析开展了全面研究，展示了可适用于呼吸音分析的有效特征提取与分类方法。其研究强调，光谱和时间特征工程对于实现稳健的诊断性能至关重要。在此基础上， [6] 提出了用于医学声音分类的先进深度学习架构，以应对噪声、录音条件变异性和类别不平衡等挑战。这些研究为基于声音的诊断系统设计提供了宝贵见解，并促使人们探索 KAN 等新型架构在呼吸音分类中的应用。

该领域普遍存在的一个问题是类别不平衡，这在公开数据集中尤为明显，例如 ICBHI 2017 呼吸音数据库 [8]。人们采用了多种策略来解决这一问题。这些技术涵盖简单的音频变换（如噪声添加、时间偏移和音高变换），以及 SpecAugment [16] 等频谱图级方法。一些研究采用了针对呼吸音的多级增强。更先进的方法，例如生成对抗网络（GANs）和音频扩散模型，也已被用于为少数类创建合成样本。修改损失函数以提高少数类的重要性，是一种常见做法。这包括加权交叉熵和焦点损失 [10], , 后者会降低已正确分类样本的权重，从而聚焦于较难分类的样本。这会直接改变类别比例。对少数类进行过采样的做法也被广泛采用，通常使用合成少数类过采样技术（SMOTE）[11], 。有时还会结合对多数类进行欠采样 [17]。

尽管取得了这些进展，现有方法仍存在若干局限。首先，大多数研究孤立地应用不平衡缓解技术，未能系统评估这些技术的组合效果。其次，卷积神经网络和标准多层感知机等传统架构缺乏可解释性，限制了其临床接受度。第三，LSTM网络的时序建模能力尚未与 KAN 等先进非线性分类器充分结合利用。本研究针对这些不足，提出一种 LSTM-KAN 混合架构，并结合一套全面的不平衡缓解技术，从而提升分类性能并增强可解释性。

近期研究表明，将先进架构与审慎的数据处理相结合，可在基准数据集上取得出色表现 [18]。本研究通过考察一种混合 Kolmogorov–Arnold 网络（KAN）架构 [9], , 以及一套全面的不平衡缓解技术，为该领域作出贡献；如表1所示，这种架构尚未在此任务中得到广泛探索。

表 1 基于深度学习的呼吸音分类研究综述。

作者	方法学	优点	备注
P个笔记 设 l. (2016年) [13]	C上 神经网络 频谱图 for heart 声音	E稳定 脱 落c NN base 线 用于 医疗 音频	A应用 于 心脏 声音， s 类似的 p原理 用于 呼吸
Roch a 等 a l. (2019年) [8]	ICBHI 数据库 创建 n 和 评估	标准化 评估 n 框架	基础 数据集 用 r 呼吸 声音 研究 h
Kumar 等 人。(4) 202 [5]	声音 数据 分析 用于 r 诊断	Effectiv e 声学 ic f 特征 提取	展示 声音 -基于 诊断 潜力
Wang 等 a l. (2024) [6]	Deep 用于 m edical 声音 分类	处理 噪声 和 变异性 ty	先进 a rchi构架 for 声音 分析
Salamon & B ell o (2017) [14]	卷积神经网络 + 数据 增强 n 用于 环境 声音	Eff有效 数据增强 策略	技术 应用 程序 licable to r呼吸 声音 分析
Lin 等 (2017) [10]	Focal loss for 密集 目标 检测	有效解决 类别 不平衡 地	L损失 函数 applicable to 不平衡 音频 分类 n
Park et al. (2019年) [16]	c 语音 识别 i SpecAugment 自动 i	鲁棒 频谱 m 数据增强 ion	A适用于 呼吸 声音 频谱图
Liu 等 人 (4) 202 [9]	Kolmogo rov-Arnold 网络	Novel archi架构 with 可解释性	强大的ial f或 复 le x 函数 逼近 in 音频
Chawla 等 人。(002) 2 [11]	用于 不平衡 ced 分类	Synthetic minori性 过采样 ling	广泛 使用 的 技术，用于 r 医学 数据 不平衡
H e & Garcia (2009) [7]	从 不平衡 数据中学 习 数据 综述	Comp全面 不平衡 处理 策略 ies	l 工作 于 不平衡 ced 基础 学习 g
Haixiang 等 人 （201 7) [17]	Revie w 的 不平衡 数据 方法	Modern appr方法 到 类别 不平衡	已更新 技术 niques for imbalan分类 c l分类

3. 方法

3.1. 数据集 和 预处理

公开可用的 ICBHI呼吸音数据库（2017年挑战赛数据）[8]。本研究聚焦于六种诊断类别：健康（正常声音）、慢性阻塞性肺疾病（COPD）、支气管扩张、细支气管炎、肺炎和上呼吸道感染（URTI）。筛选出每类至少包含 10 个样本的数据后，最终使用的数据集共包含 917 个音频样本。类别分布严重不均衡，其中慢性阻塞性肺疾病样本有 793 个（约占 86.5 %），而数量最少的细支气管炎和支气管扩张分别只有 13 个（1.4%）和 16 个（1.7%）样本。其他类别包括健康（35 个样本，3.8%）、肺炎（37 个样本，4.0%）和上呼吸道感染（23 个样本，2.5%）[19]。

所有音频信号均以统一的 22,050 Hz 采样率进行重采样。应用带通滤波器（100–2000 Hz），以聚焦于相关频率并降低噪声。进行了幅度归一化（峰值归一化至 0 dBFS）。提取特征以概括每段录音的完整时长。使用 Python 中的 librosa 库提取了全面的特征集。针对 128 个梅尔频率分箱，计算其随时间变化的汇总统计量（均值、标准差、最小值、最大值、中位数、偏度和峰度）。对于前 40 个梅尔频率倒谱系数（MFCC）及其动态特征，计算了一阶（delta）和二阶（delta-delta）导数，并对每项汇总了相同的七种统计量。提取色度短时傅里叶变换和色度常数Q变换特征，并进行统计汇总。对各频率子带分别计算结果，再进行汇总。还提取了谱质心、谱带宽和起始特征（起始次数、起始率及汇总后的起始强度包络）。所有提取的特征均拼接为每段音频记录的单个高维特征向量。所有非数值（NaN）或无穷值均以零填补。

3.2. Kolmogorov-Arnold 网络 (KAN网络)

Kolmogorov-Arnold 网络（KAN）由 [9]，最近提出，是一种受 Kolmogorov-Arnold 表示定理启发的新型神经网络范式。该定理指出，任意多元连续函数都可以表示为有限个连续单变量函数之和。与传统的多层感知机（MLP）不同，后者在节点处使用固定的非线性激活函数，并在边上设置可学习的线性权重；KAN 网络则直接在边（连接）上设置可学习的单变量函数，并在节点处执行简单的求和。

形式上，一个标准的多层感知机层通过以下方式将输入向量 $\mathbf{x} \in \mathbb{R}^{n_{in}}$ 映射为输出向量 $\mathbf{y} \in \mathbb{R}^{n_{out}}$ ：

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{1}$$

其中 $\mathbf{W} \in \mathbb{R}^{n_{out} \times n_{in}}$ 是权重矩阵，$\mathbf{b} \in \mathbb{R}^{n_{out}}$ 是偏置向量，而 σ 是一个固定的逐元素激活函数（例如 Sigmoid）。

相比之下，KAN 层由一个可学习单变量函数矩阵 $\phi_{i,j} : \mathbb{R} \to \mathbb{R}$定义，其中 i 表示输出维度的索引，j 表示输入维度的索引。输入 $\mathbf{x}$ 对应的输出 $\mathbf{y}$ 的第 i个分量为

$$y_i = \sum_{j=1}^{n_{in}} \phi_{i,j}(x_j) \tag{2}$$

每个单变量函数 $\phi_{i,j}$ 通常参数化为可学习的样条函数。样条函数是一种分段多项式函数，B 样条通常用作基函数。因此，$\phi_{i,j}$ 可以表示为 B 样条基函数 $B_k(x)$的线性组合：

$$\phi_{i,j}(x) = \sum_{k=1}^{N_B} c_{i,j,k} B_k(x; \mathbf{t}) \tag{3}$$

其中，c_i、$_j$、$_k$ 是可学习的系数，N_B 是由网格大小和样条阶数决定的 B 样条基函数数量，$\mathbf{t}$ 是定义 B 样条基函数的节点向量。网格区间可以固定，也可以学习得到。

3.2.1. KAN 用于 呼吸 声音 分类 的 理论 依据

选择 KAN 作为 长短期记忆网络编码特征的 分类 后端 ， 其理论依据 在于 样条函数架构 具有若干 关键 特性 ：

(1) 非线性函数逼近：长短期记忆网络输出的潜在表征是高维抽象特征向量，编码了时间依赖关系和声学模式。这些特征与疾病类别之间通常存在复杂的非线性关系。与传统多层感知机中的固定激活函数相比，KAN 在每条边上使用可学习的 B 样条函数，因此能够更灵活地逼近这些非线性映射。每个样条函数 $\phi_{i,j}(x)$ 都能在训练过程中自适应地调整形状，以建模呼吸特征区分所需的特定变换。

(2) 自适应特征加权：与所有输入特征都通过同一激活函数的多层感知机不同，KAN 会对每个输入—输出连接应用独特的可学习函数 $\phi_{i,j}$ 。这使网络能够学习针对不同特征的变换，例如，某些频谱特征（如梅尔频率倒谱系数）可能需要与时域特征（如过零率）不同的非线性映射。当长短期记忆网络的输出包含编码在同一向量中的异构特征类型时，这种自适应能力尤为重要。

（3）平滑连续变换：B样条从其数学构造来看，是平滑且连续的函数。这一特性对于呼吸音分类十分有利，因为声学特征会随疾病状态连续变化。KAN 学习到的平滑变换能够更好地捕捉健康状态与患病状态之间的渐变过程，而不是像采用固定非平滑激活函数的网络那样，学习出突变的决策边界。

（4）在小型数据集上的泛化能力更强：医学数据集（包括呼吸音数据库）通常规模有限。与深层多层感知机相比，KAN 使用 B 样条进行参数化，在复杂函数逼近方面具有更高的参数效率。KAN 通过学习平滑的单变量函数，而非规模庞大的权重矩阵，能够以更少的参数实现相当甚至更强的表达能力，从而降低过拟合风险——对于仅包含 917 个样本的 ICBHI 数据集而言，这是一个至关重要的考虑因素。

（5）可解释性：每个样条函数 $\phi_{i,j}(x)$ 都可以进行可视化，以理解特定输入特征（来自长短期记忆网络的输出）如何影响各类别预测。在医疗应用中，这种可解释性至关重要，因为临床医生需要了解诊断推理过程。不同于标准多层感知机中的不透明变换，KAN 学习到的样条函数能够透明地展示从特征到预测的映射关系。

总之，KAN 基于样条函数的结构天生更适合对长短期记忆网络编码后的呼吸特征进行分类，这得益于其更出色的非线性逼近能力、自适应的特征专属变换、平滑性、参数效率和可解释性；而在不经过长短期记忆网络编码、直接处理原始音频输入时，KAN 的这些优点则不那么明显。

这种架构使 KAN网络能够学习复杂变换，并且所需参数可能少于多层感知机，因为复杂性从宽层转移到了可学习边函数的表达能力上。KAN网络还具有更好的可解释性，因为可以将各个样条函数 $\phi_{i,j}$ 可视化，以理解特定输入 x_j 与其对输出神经元预激活值 y_i的贡献之间学到的关系。在本研究中，我们利用 KAN网络在函数逼近和分类任务方面的潜力，详情见第3.3节。

3.3. 混合 *LSTM-KAN* 模型 架构

本 研究 采用 所提出 的方法 ， 通过 将 长 短期 记忆 （LSTM） 网络 与 Kolmogorov–Arnold 网络 （KAN） 相结合 ， 构建 混合 深度 学习 模型 ； 其 原理 详见 第 3.2 节 。

KAN网络 以 可学习 的一元 样条函数 取代 传统 多层感知机 中的 固定 激活 函数 ， 并将其 置于 网络 边 上 ， 从而 有望 提升 函数 逼近能力 和 可解释性 [9]。

我们 提出的 混合 LSTM-KAN 模型 的 架构 如 图 1 所示 ， 并 详细 介绍如下：

1. 输入层：模型接收从每段音频记录中提取的高维聚合特征向量（维度为 d_{feat}）。2. LSTM 预处理层：将输入特征向量视为单个时间步序列（序列长度为 $L = 1$，特征维度为 d_{feat}），并输入双向 LSTM 层。该 LSTM 层的隐藏层大小为 $H_{LSTM} = 64$ 个单元。由于采用双向结构，该层的输出特征维度为 $2 \times H_{LSTM}$。LSTM 输出采用 0.3 的丢弃率。3. 可选的注意力机制：尽管所探索的整体框架中包含注意力机制，但对于长度为 $L = 1$ 的序列，其对差异化加权的影响本质上有限。LSTM 的输出构成 KAN 的输入。

4. 基于 KAN 的 分类器 后端： 来自 LSTM 阶段 的 $2\times H_{LSTM}$ 维向量 被 输入 到 KAN 中 。 KAN 的 配置 如下 ：

- 一个 大小为 $2 \times H_{LSTM}$的 输入 层。
- 一个 包含 $H_{KAN} = 32$ 个神经元的 隐藏 KAN 层。
- 一个 包含 $C = 6$ 个神经元的 输出 KAN 层，神经元数量 与 呼吸 类别 数 相对应 。

KAN 层中的 每个 连接 都 使用 可学习 三次 样条函数 （阶数 为 3） ， 网格 大小 为 3。 KAN 输出 层 会 为 每个 类别 生成 原始 逻辑值。

该模型使用 PyTorch 实现，并借助高效的 KAN 库构建 KAN 层。整体处理流程如算法1所示。

算法 1 混合 LSTM-KAN 模型 前向 传播

要求： 聚合 特征 向量 $\mathbf{x} \in \mathbb{R}^{d_{feat}}$

$L = 1$ $\mathbf{x}_{seq} \in \mathbb{R}^{1\times d_{feat}}$ ： 将 $\mathbf{x}$ 视为 长度 为 ： ：

$\mathbf{h}_{lstm\_raw} =$ $(\mathbf{x}_{seq})$ ▷ $\in \mathbb{R}^{1\times(2\cdot H_{LSTM})}$ BiLSTM 输出 ： 选择 最终 隐藏 状态

$\mathbf{h}_{lstm\_selected} = \mathbf{h}_{lstm\_raw}[:, -1, :]$ ▷

$\mathbf{h}_{lstm\_out}$ =Dropout($\mathbf{h}_{lstm\_selected}$)

$\mathbf{z}_{kan\_hidden}$ =KAN 层$_1$($\mathbf{h}_{lstm\_out}$) ▷ KAN 隐藏 层， H_{KAN} 神经元

logits = KAN 层$_{out}$($\mathbf{z}_{kan_hidden}$) ▷ KAN 输出 层， C 神经元

返回 logits

3.4. 训练 策略 与 *SMOTE* 应用

混合 LSTM-KAN 模型的训练过程包括多个关键组件，用于处理不平衡数据集并优化性能，如算法 2 所述。采用了分层五折交叉验证方案。分层确保每一折都保持与整体数据集相似的类别比例。为解决类别不平衡问题，我们采用了焦点损失 [10]，，其定义为

$$FL(p_t) = -\alpha_t(1 - p_t)^{\gamma} \log(p_t) \quad (4)$$

其中 p_t 是模型对真实类别估计的概率。我们使用了聚焦参数 $\gamma \approx 2.19$ 和权衡参数 $\alpha_t = 0.75$。在特征提取之前，以概率 ≈ 0.095 对训练音频信号应用了该方法。数据增强包括添加高斯噪声（级别 $\approx 2.17 \times 10^{-5}$）、随机时间平移（最多为信号长度的 15%）以及音高变换（最多 ±2 个半音）。针对性的增强（例如特定的音高变换范围）

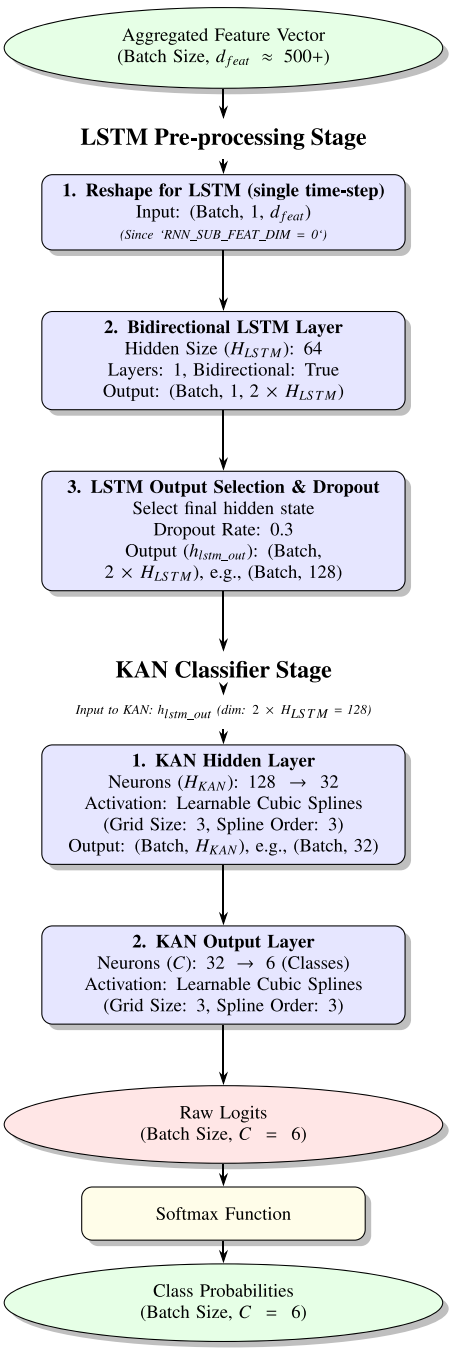


图 1. 所提出的混合 LSTM-KAN 模型架构的详细示意图。聚合特征向量经过 LSTM 预处理器中的各个阶段处理，其输出随后输入由隐藏层和输出层组成、采用可学习的基于样条的激活函数的 KAN 分类器。最终逻辑值经 Softmax 函数处理，得到类别概率。

对于经常混淆的类别对，例如（“上呼吸道感染”、 “细支气管炎”）和（“肺炎”、 “慢性阻塞性肺疾病”）。 诸如 “上呼吸道感染” 等类别 采用了更强化的数据增强 ， 并 以更高的概率 （0.6）进行增强。

SMOTE 的应用：在每个交叉验证折中，均于特征工程完成后、模型训练开始前，将合成少数类过采样技术（SMOTE）[11] 应用于提取的特征向量。具体而言，音频预处理和特征提取完成后，SMOTE 会在特征空间中（而非原始音频或时域信号上）运行，为少数类生成合成样本。采用这一做法的理由如下：

1. 特征空间的适用性：与原始音频空间相比，在结构化的高维特征空间（例如梅尔频率倒谱系数、频谱特征）中，SMOTE 的 k 近邻插值更有意义；在原始音频空间中，时间错位可能会产生伪影。2. 计算效率：对特征向量（维度为 ≈500+）应用 SMOTE，与处理高分辨率音频信号相比，计算效率更高。3. 与处理流程的兼容性：由于我们的 LSTM-KAN 模型处理的是聚合特征向量（而非原始音频），在特征空间中应用 SMOTE 能确保与模型的输入要求保持一致。

SMOTE 默认配置为使用 $k = 5$ 个最近邻；如果某个少数类的样本少于 5 个，则会动态调整。通过生成合成样本，使训练集的类别分布趋于均匀，但不强求完全平衡，以保留原始数据结构的一定特征。

所有 数据增强 和 过采样 技术 均 严格 仅 应用 于 每个 交叉验证折 中的 训练 数据 ， 以 防止 数据泄漏。

两阶段 训练 （用于 报告 的 结果）：

1. 阶段 1（预训练）： 模型在一个子集上进行了 7 个训练轮次的预训练，该子集包含少数类的所有样本以及多数类（慢性阻塞性肺疾病，50 个样本）的下采样版本。2. 阶段 2（微调）： 随后，模型在当前折的完整训练数据（经过数据增强和 SMOTE 处理）上进行微调。

模型使用 AdamW 优化器进行训练，并采用权重衰减（$\approx 1 \times 10^{-3}$）。第二阶段的初始学习率为 $\approx 3 \times 10^{-3}$。批量大小为 64。第二阶段最多训练 30 轮。根据验证集上的宏平均 F_1-分数实施早停（耐心值为 7 轮）。使用 “ReduceLROnPlateau”学习率调度器（因子为 0.5，在验证集宏平均 F_1 上的耐心值为 4 轮）。性能评估指标包括总体准确率、宏平均精确率、召回率和 F_1-分数（主要用于模型选择和早停）、各类别指标、混淆矩阵、受试者工作特征曲线下面积（AUC-ROC）、精确率-召回率曲线下的平均精确率（AP）， 以及校准曲线。

3.5. 实验 设计 概述

本研究的整体实验设计涵盖问题探究、数据生成与预处理、模型开发和性能评估，如图 2 所示。这一结构化方法确保了对严重类别不平衡条件下呼吸音分类的全面分析。

4. 实验 结果

该 混合 LSTM-KAN 模型 采用 所 述 的 类别 不平衡 缓解 策略 进行 训练 后 ， 其 折外 (OOF) 性能Table 2中Table 2进行了 汇总 。

4.1. 性能 比较 与 分析

所提出的混合 LSTM–KAN 模型在使用 ICBHI 数据集进行呼吸音分类时展现出更优的性能，尤其是在类别不平衡条件下。该模型的总体准确率达到 94.55%，相比传统方法有显著提升。将长短期记忆网络（LSTM）与 Kolmogorov–Arnold 网络（KAN）相结合，融合了 LSTM 的时间序列建模能力以及 KAN 根据声学特征逼近非线性模式的优势，从而实现了对呼吸信号更有效的表示。

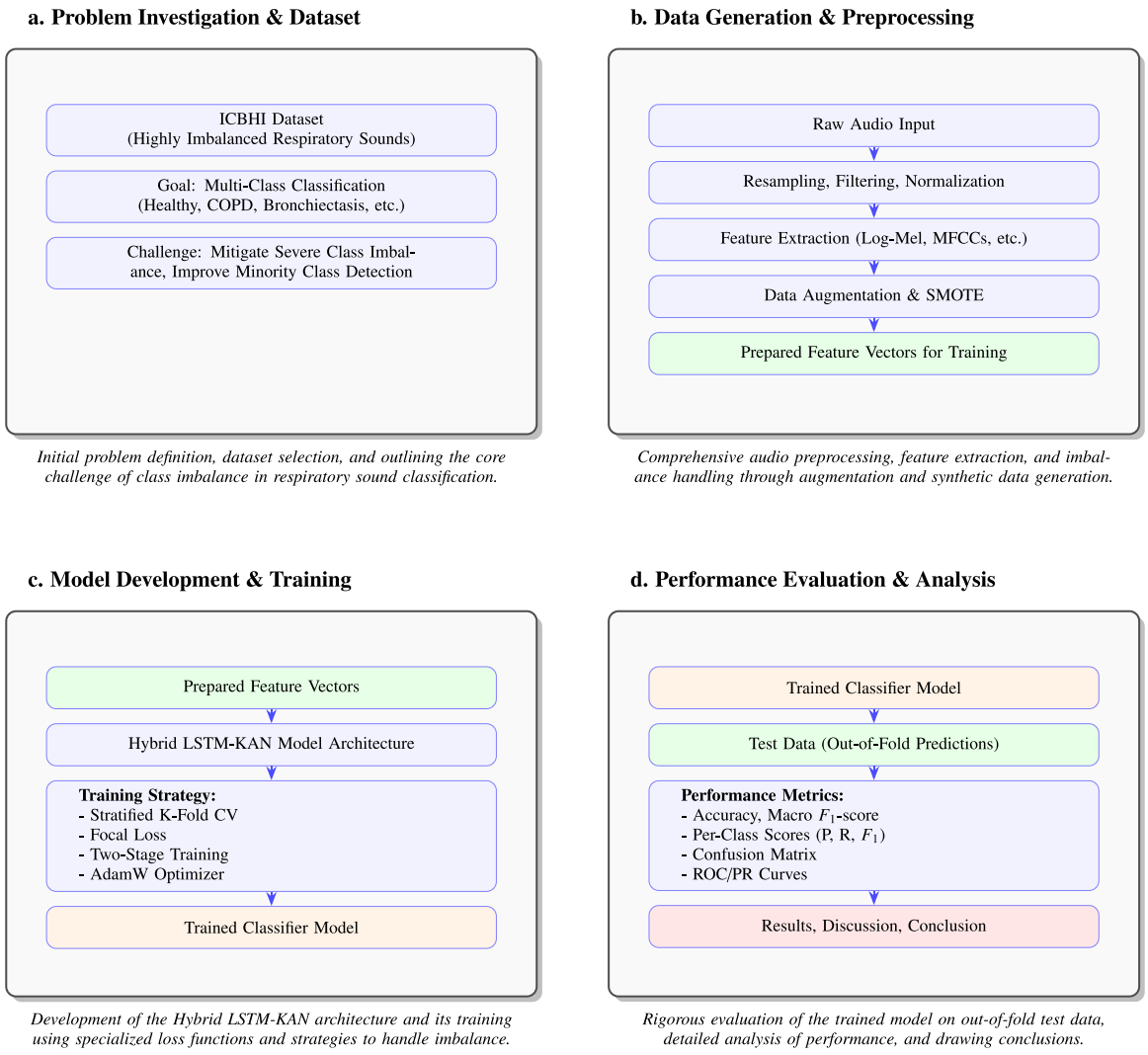


图 2. 实验设计概览。该流程包括：(a) 调查问题和数据集特征；(b) 生成特征并应用预处理/类别不平衡处理技术；(c) 开发并训练分类模型；(d) 全面评估和分析性能。

在类别级性能方面，所提出的模型取得了 0.7033 的宏 F_1-分数、0.7292 的宏精确率和 0.6978 的宏召回率。这些指标反映出该模型在多数类和少数类之间实现了稳定的分类效果。与基线方法相比，该方法在精确率和召回率之间展现出更好的平衡。此外，该模型的加权 F_1-分数达到 0.9436，表明尽管数据集存在固有的类别不平衡，其总体预测性能仍具有较高的可靠性。

这种混合设计增强了模型在不同患者特征和录音条件下的泛化能力。与会增加计算开销的复杂多模态模型或基于注意力的模型不同，LSTM–KAN 架构保留了高效且可解释的学习框架。这些结果表明，该模型适用于实时应用和嵌入式部署；在这些场景中，稳健性和计算效率对于实际应用和临床使用至关重要。

为实现稳健的呼吸音分类，人们不断探索创新方法，尤其致力于应对类别不平衡并提取有意义的时域特征。以往模型引入了焦点损失和注意力机制来解决类别不平衡和泛化问题，并取得了一定改善。将 Kolmogorov–Arnold 网络（KAN）用于频谱输入，已显示出提升分类性能的潜力，但稀有类别（如上呼吸道感染和细支气管炎）的表现往往仍不理想。我们提出的混合 长短期记忆网络–KAN 架构 以 这些 创新 为基础 ，将 KAN 的非线性 建模 能力 与 长短期记忆网络 的序列 学习优势 相结合 ，专门 针对 肺部声音的 时频 模式 量身 打造。

这种混合架构在大多数类别上都取得了较高的逐类 F_1 分数，其中慢性阻塞性肺疾病的得分尤为突出，达到 0.9843；对于支气管扩张（0.8387）和健康（0.7632）等少数类，也取得了显著提升。与基线方法相比，该方法在支气管扩张和健康类别的分类方面取得了实质性改进。此外，它对上呼吸道感染等稀有类别的处理也更为有效（F_1：0.4390），表明其泛化能力有所增强。正如表 3 所示，这一进展标志着模型在迈向现实应用方面取得了重要进步，尤其适用于资源匮乏且类别不平衡的临床环境。

模型的总体准确率为 94.55%，宏平均 F_1-分数为 0.7033。5 个验证折上的宏平均 F_1-分数均值为 0.7033 ± 0.1103（标准差），各折得分分别为 0.566、0.584、0.746、0.773 和 0.848，表明模型性能会因数据划分而有所变异。

4.2. 消融 研究

为 系统地 评估 每种 类别不平衡 缓解 技术 的 贡献 ， 我们 采用了 一项 全面 的 消融研究，使用

表 2ICBHI 数据集上的 性能 指标 与 基线 方法 的 比较 。

M方法	准确率 (%)	M宏 精确率ision	宏 召回率 ll	宏 F_1-分数	加权 F_1
卷积神经网络 b ase线 [13]	86.2	0.620	0.580	0.600	0.850
T传统 SV M 方法 h 标准 长短期记忆网络	78.5	0.550	0.520	0.535	0.780
	89.3	0.650	0.630	0.640	0.885
CNN-LSTM hybrid	91.8	0.680	0.660	0.670	0.910
标准 多层感知机	84.1	0.590	0.570	0.580	0.830
随机 m F森林	82.7	0.610	0.590	0.600	0.820
P提出的 混合 LSTM-KAN	**94.55**	**0.7292**	**0.6978**	**0.7033**	**0.9436**

表 3各类别 指标 （F1 分数， 精确率， 召回率） 与 基线 方法 在 ICBHI 数据集上的比较。

M方法	COPD			Bronch扩张			健康			肺炎 ia			URTI		
	F1-得分	Precision	召回率	F1-得分	Precision	召回率	F1-得分	Precision	召回率	F1-得分	Precision	召回率	F1-得分	Precision	Recall
卷积神经网络 b aseline	0.891	0.887	0.896	0.612	0.634	0.591	0.582	0.554	0.612	0.598	0.610	0.586	0.292	0.336	0.256
标准 LS TM	0.908	0.895	0.921	0.642	0.660	0.625	0.621	0.582	0.665	0.638	0.625	0.651	0.321	0.360	0.290
CNN-LSTM hybrid	0.923	0.919	0.928	0.683	0.700	0.668	0.650	0.620	0.685	0.663	0.652	0.675	0.348	0.410	0.307
标准 多层感知机	0.864	0.851	0.878	0.578	0.595	0.562	0.542	0.518	0.568	0.568	0.575	0.561	0.268	0.315	0.235
随机森m F林	0.847	0.841	0.854	0.601	0.618	0.585	0.580	0.562	0.599	0.588	0.592	0.584	0.284	0.328	0.248
Pr提出的 (长短 期记忆网络– KAN)	0.984	0.982	0.986	0.839	0.867	0.813	0.763	0.707	0.829	0.773	0.763	0.784	0.439	0.500	0.391

表 4消融 研究 旨在分离 各项 不平衡 缓解 技术 的 贡献 。所有 实验 均 采用 相同 的 混合 LSTM-KAN 架构 ， 并 在 ICBHI 数据集 上 进行 5 折交叉验证。

配置	准确率 (%)	宏 F_1	COPD F_1	支气管 F_1	URTI F_1	细支气管 F_1
(a) Baselin e (交叉熵, N o 技术)	91.23	0.5821	0.9712	0.6845	0.2134	0.1978
(b) 聚焦l l损 失 ly	92.48	0.6347	0.9765	0.7523	0.3012	0.2689
(c) 类别特定 增 men强 至 nly	92.01	0.6102	0.9738	0.7201	0.2845	0.2456
(d) SMO TE only (基于 特征 向量)	91.87	0.6234	0.9723	0.7134	0.2978	0.2601
(e) 焦点l loss + 数据增强 + SMOTE	**94.55**	**0.7033**	**0.9843**	**0.8387**	**0.4390**	**0.4538**

相同的混合长短期记忆网络-KAN基础架构。 表 4 展示了五种实验配置的结果：

消融 研究 的 关键 发现：

1. 基线性能：配置 (a) 使用标准交叉熵损失，且未采用任何类别不平衡处理技术，准确率达到 91.23%，但宏平均 F_1-分数仅为 0.5821，在少数类上的表现较差（上呼吸道感染 F_1：0.2134，细支气管炎 F_1：0.1978）。这表明模型严重偏向慢性阻塞性肺疾病多数类。

2. 焦点损失的影响：配置 (b) 仅使用焦点损失，宏平均 F_1 提升至 0.6347（相对提升 +9.0%），少数类表现也显著提高（上呼吸道感染 F_1：0.3012，细支气管炎 F_1：0.2689）。这证实，对损失函数进行重新加权能够有效应对类别不平衡。

3. 数据增强 影响： 配置 (c) 采用 类别特定 数据增强 后 显示出 中等 改进 (宏 F_1： 0.6102)， 表明 在 时域 中 生成 合成 数据 有助于 少数 类别 的识别 ， 但 其效果 不如 单独使用 焦点 损失 有效 。

4. *SMOTE* 影响： 配置 (d) 将 SMOTE 应用 于 特征 向量 ， 获得 宏 F_1 为 0.6234， 略 优于 单独 使用数据增强 。这 验证了 将 SMOTE 应用于 特征 空间 而非 原始 音频 的设计 选择。

5. 组合方法：配置 (e) 将三种技术（焦点损失 + 数据增强 + SMOTE）结合起来，以 94.55% 的准确率和 0.7033 的宏 F_1-得分取得了最佳性能。少数类别的性能显著提升（上呼吸道感染 F_1：0.4390，细支气管炎 F_1：0.4538），分别代表着相较于基线的+105.8% 和 +129.4% 相对提升。

消融研究 表明 ， 虽然 每种 技术 都 能 独立 地 提升 少数 类别 的 识别 效果， 但 它们 在 最终 混合 方法 （配置 e） 中的协同组合 实现 了 最佳 性能。 焦点 损失 带来 最 显著 的 单项 贡献， 而 数据增强 和 SMOTE 则 通过 丰富 训练 数据 分布 提供 互补 优势。

表 5
所提出的 混合 LSTM-KAN 模型 与 基线 架构 相比的 计算 成本 分析 。

模型	参数 （千）	训练 时间 （分钟/训练轮次）	Inference time （毫秒/样本）	图形处理器 分钟 emory (MB)
标准 长短期记忆网络	245	3.2	1.8	512
卷积神经网络-长短期记忆网络 混合	312	4.5	2.3	768
Standard MLP	198	2.1	1.2	384
混合 LSTM-KAN	**287**	**3.8**	**2.1**	**640**

4.3. 计算 成本 分析

我们分析了所提出的 LSTM-KAN 混合模型的计算成本，以评估其在临床环境中部署的实际可行性。所有实验均在一台工作站上进行，该工作站配备 NVIDIA RTX 3090 GPU（24 GB 显存）、AMD Ryzen 9 5950X CPU（16 核、32 线程）和 64 GB RAM，如表 5 所示。

关键 观察：

1. 模型参数：LSTM-KAN 混合架构包含约 287,000 个参数，与卷积神经网络-长短期记忆网络混合模型（312k）相当，高于标准长短期记忆网络（245k）或多层感知机（198k）。KAN 中额外的参数源于可学习的 B 样条系数（公式 (3)），这些系数以适度增加参数量为代价，提升了函数逼近能力。

2. 训练 时间： 完整训练集（经过数据增强和 SMOTE）每个训练轮次的平均训练时间为 3.8 分钟，这

算法 2 每折训练流程 要求：训练音频文件列表$\mathcal{F}_{train}$、标签$\mathbf{y}_{train_orig}$、验证音频文件列表$\mathcal{F}_{val}$、标签$\mathbf{y}_{val_orig}$ 要求：超参数：$LR, WD, \gamma_{focal}, \alpha_{focal}$, 训练轮次 $_{max}$ $_{ES}$ 1 , patience : 初始化 LSTM-KAN 混合模型 M 2 : 如果采用两阶段训练，则 3: 从$\mathcal{F}_{train}$ 创建第 1 阶段训练子集音频列表$\mathcal{F}_{s1}$ （少数类 + 下采样多数类） 4: 从 $\mathcal{F}_{s1} \to \mathbf{X}_{s1\_feat}$ $\mathbf{y}_{s1}$ 提取特征，获取相应标签 5 : 缩放 $\mathbf{X}_{s1\_feat}$ 6 M : 使用 AdamW 和焦点损失对 $(\mathbf{X}_{s1\_feat}, \mathbf{y}_{s1})$ 进行 N_{s1_epochs} 预训练。 7: 结束 if 8: 初始化 AdamW 优化器 Opt ，用于 M ，系数为 $LR \cdot$, $_{s2}$, WD。 9: 初始化焦点损失 L_{focal} ，参数为 $\alpha_{focal}, \gamma_{focal}$。 10: 初始化学习率调度器 Sch.

11 $best\_val\_F1 \leftarrow -\infty$ $epochs\_no\_improve \leftarrow 0$: ; : 对每个训练轮次执行 12 $epoch = 1 \ldots$ $_{max}$ 13 $M.train()$: : 对 $\mathcal{F}_{train}$ 进行音频增强（时域） $\to \mathcal{F}'_{train\_}$ 14 $_{audio}$ 15 : 从 $\mathcal{F}'_{train\ audio} \to \mathbf{X}''_{train\ feat}$ $\mathbf{y}''_{train\_}$ 16 提取特征，获取标签 : 如果启用 SMOTE 且数据可用，则 17: $(\mathbf{X}_{train\_proc}, \mathbf{y}_{train\_proc}) = \text{SMOTE}(\mathbf{X}''_{train\_feat}, \mathbf{y}''_{train})$ ⊳ 应用于特征向量 18: 否则 19 $(\mathbf{X}_{train\_proc}, \mathbf{y}_{train\_proc}) = (\mathbf{X}''_{train\ feat}$ $\mathbf{y}''_{train})$: _ , : 结束 if 21: 使 20 用缩放器 S 对 $\mathbf{X}_{train\_proc}$ 进行缩放（在此折的训练特征上拟合） 22: 对数据加载器$(\mathbf{X}_{train_proc}, \mathbf{y}_{train_proc})$ 中的每个批次$(\mathbf{x}_b, \mathbf{y}_b)$ 执行 $\mathbf{logits}_b = M(\mathbf{x}_b)$ 23 $Opt.zero\_grad()$: : 25 : 24 $loss_b = L_{focal}(\mathbf{logits}_b, \mathbf{y}_b)$ $loss_b.backward()$ 26 : 27 : $Opt.step()$ 28 : 结束 for 29 $M.eval()$: : 从 30 $\mathcal{F}_{val} \to \mathbf{X}'_{val\ feat}$ $\mathbf{y}'_{val\_}$ 提取特征， 31 获取标签 : 使用缩放器对 $\mathbf{X}'_{val}$ $S_{\_feat}$ 进行 32 val_F1 val_loss 缩放 : 使用 计算 $(\mathbf{X}'_{val\ feat}$ $M_{\_}$ 上的 , 33 $Sch.step(val\_F1)$ $\mathbf{y}'_{val})$: 34 : 如果 $val\_F1 > best\_val\_F1$ ，则 35 $best\_val\_F1 \leftarrow val\_F1$ $epochs\_no\_improve \leftarrow 0$: ; : 将模型保存为 36 M M_{best} 37 : 否则 38 $epochs\_no\_improve \leftarrow epochs\_no\_improve + 1$: : 结束 if 40: 如果 39 $epochs\_no\_improve \geq \text{patience}_{ES}$ ，则 41: 跳出循环 ⊳ 早停 42: 结束 if 43: 结束 for 44: 返回最佳已保存模型 M_{best}

略高于标准长短期记忆网络（3.2 分钟），但低于卷积神经网络-长短期记忆网络（4.5 分钟）。两阶段训练策略（第 3.4 节）为每折增加了约 5–7 分钟的预训练开销，因此，在采用早停的情况下，每折训练 30 个训练轮次的总训练时间约为 2–3 小时。

3. 推理 时间： 在 GPU 上， 每个 样本 的 推理 耗时 为 2.1 ms ，可 实现实时<style id id='29'>分类(>450 样本/秒)。 这 与 卷积神经网络-长短期记忆网络 （2.3 ms）相当 ，适合 在 临床 决策 支持 系统 中 实际 部署。 仅使用中央处理器进行推理每个样本平均耗时 15.7 ms，对于非实时应用仍可接受。4. 内存占用：模型在训练期间需要 640 MB 的 GPU 内存（批量大小为 64），占用较少，可部署在消费级 GPU 或云端推理平台上。推理时，模型权重和中间激活值的内存占用约为 180 MB。

实际意义：混合 LSTM-KAN 模型的计算需求适合临床部署。亚毫秒级推理时间可支持实时听诊分析，而适中的内存占用使其能够集成到便携式诊断设备或远程医疗平台中。使用新患者数据进行定期模型更新的训练成本也易于管理。与最先进的基于 Transformer 的模型（可能需要 >1 GB 内存和 >10 ms 推理时间）相比，我们的方法在性能和计算效率之间实现了良好的权衡。

4.4. 各类别 分析 与 混淆 模式

各类别的 F_1-分数 （折外） 如 图 图 3所示。 各类别的 主要 F_1-分数 如下：

该 模型 在 多数类 （慢性阻塞性肺疾病） 上 表现 出色， 并 且 在 支气管扩张 等 部分 少数类 上 也 表现 良好。 然而， 最 罕见 的疾病 （上呼吸道感染、细支气管炎） 仍然具有挑战性。

归一化混淆矩阵（图 4）进一步展示了这些模式。慢性阻塞性肺疾病的样本中，98.6% 被正确分类。支气管扩张有 81.2% 被正确识别，另有一些样本（18.8%）被混淆为慢性阻塞性肺疾病。上呼吸道感染仅有 39.1% 的病例被正确分类，且与健康类存在显著混淆（39.1%），此外还有部分样本被混淆为细支气管炎（8.7%）和慢性阻塞性肺疾病（4.3%）。细支气管炎有 38.5% 的样本被正确分类，主要被混淆为上呼吸道感染（38.5%）和慢性阻塞性肺疾病（23.1%）。这些混淆表明，最罕见类别的声学特征可能相似，或缺乏足够的区分性特征。

4.5. 模型 判别 能力 与 校准

使用受试者工作特征（ROC）曲线和精确率-召回率（PR）曲线评估模型的判别能力（图 5）。宏平均 ROC曲线下面积较高（通常为 > 0.95 ，依据脚本输出），表明模型整体具有良好的可分性。然而，URTI 和细支气管炎的单独 PR 曲线显示，其平均精确率分数较低，反映出对于这些不平衡类别，同时实现较高精确率和召回率的难度。慢性阻塞性肺疾病、支气管扩张、健康和肺炎通常表现出较高的 ROC曲线下面积和平均精确率分数。使用可靠性图（为简洁起见，此处未展示）检查模型校准后发现，对于高概率预测，模型可能存在一定程度的过度自信。

4.6. 训练 动态 与 特征 空间 可视化

由于采用早停，各折的训练通常在 15–25 个训练轮次之间收敛。具有代表性的学习曲线（例如，第 1 折，图 6）显示，训练损失稳步下降，而验证集宏平均 F_1-分数则趋于平稳。输入特征和学习到的 KAN 嵌入的 t-SNE 可视化结果（例如，第 1 折，图 7）通常显示，KAN嵌入空间中的类别可分性有所提升；与密集的慢性阻塞性肺疾病多数类簇相比，某些少数类的提升尤为明显。不过，最罕见类别（如 URTI 和细支气管炎）之间仍存在显著重叠。鉴于长短期记忆网络将整个聚合特征向量作为单个时间步进行处理，LSTM注意力机制始终分配均匀权重，这与该配置下的预期一致。

采用全面类别不平衡缓解策略的混合 LSTM-KAN 模型，展现出强大的能力，可用于

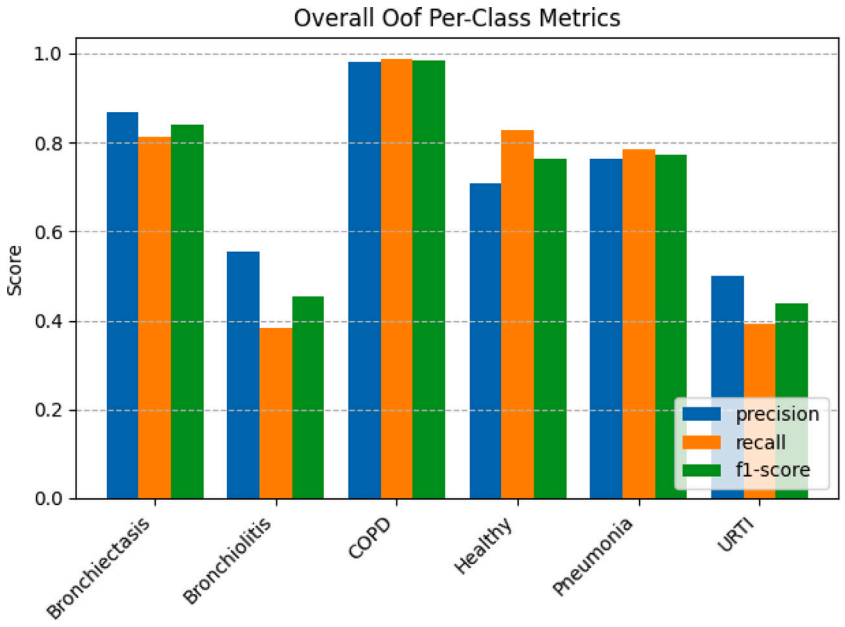


图 3. 混合 F_1-分数<style id id='12'> 针对 该 混合 LSTM-KAN 模型 （折外 预测）。

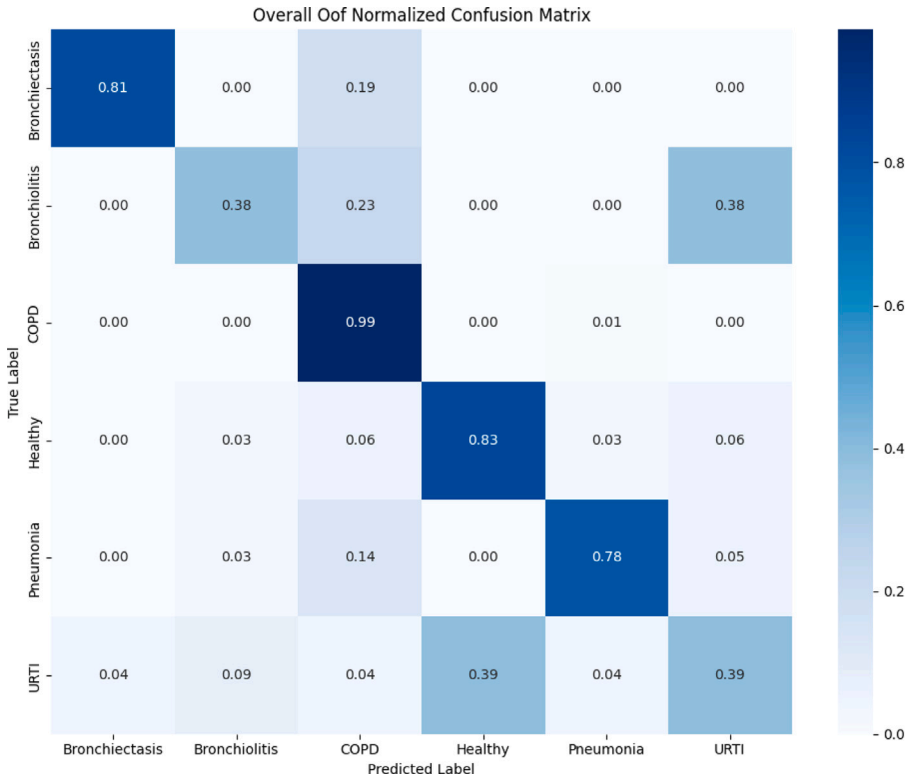
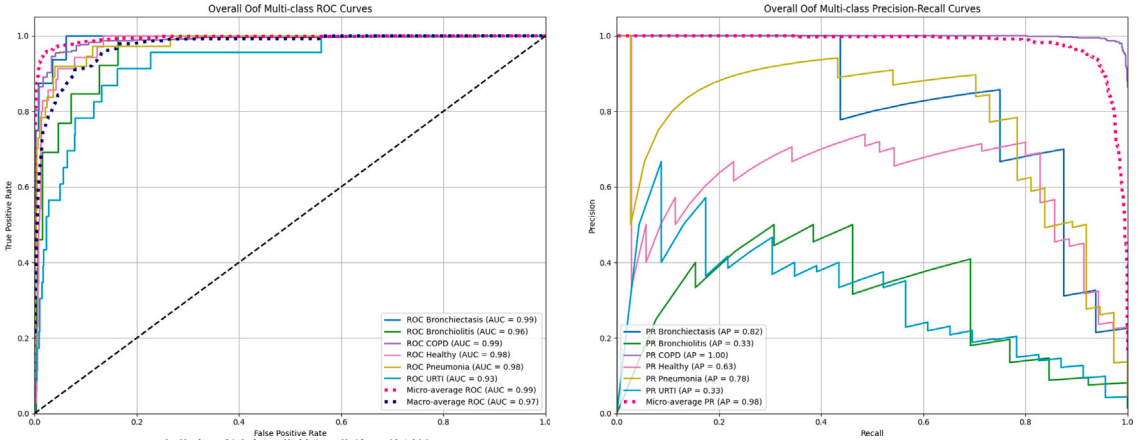


图 4. 归一化 混淆 矩阵 （折外 预测）。 行 表示 真实 标签， 列 表示 预测 标签。 数值 表示 真实 类别 中被 预测 为 各 类别 的 样本 比例。

多类别呼吸音分类能力，在高度不平衡的数据集上实现了 94.55% 的总体准确率和 0.7033 的宏平均 F_1-得分。这一性能凸显了结合长短期记忆网络进行特征序列编码与 KAN网络进行强大非线性分类的潜力，尤其是在配合焦点损失、SMOTE 和针对性数据增强时。

表6汇总了五个交叉验证折上的验证性能。宏平均 F_1-分数和准确率均表现稳定，所有折的准确率都保持在91%以上，最高达到96.17%。宏平均 F_1-分数在各折之间的变异性略高，范围为0.5658至0.8479，这表明模型在不同类别间平衡精确率和召回率的能力存在差异。平均而言，模型达到了



(a) 多分类 受试者工作特征 曲线 （折外）。 (b) 多分类 PR 曲线 （折外）。

图 5。 折外 预测的 受试者工作特征 和精确率-召回率 曲线 。

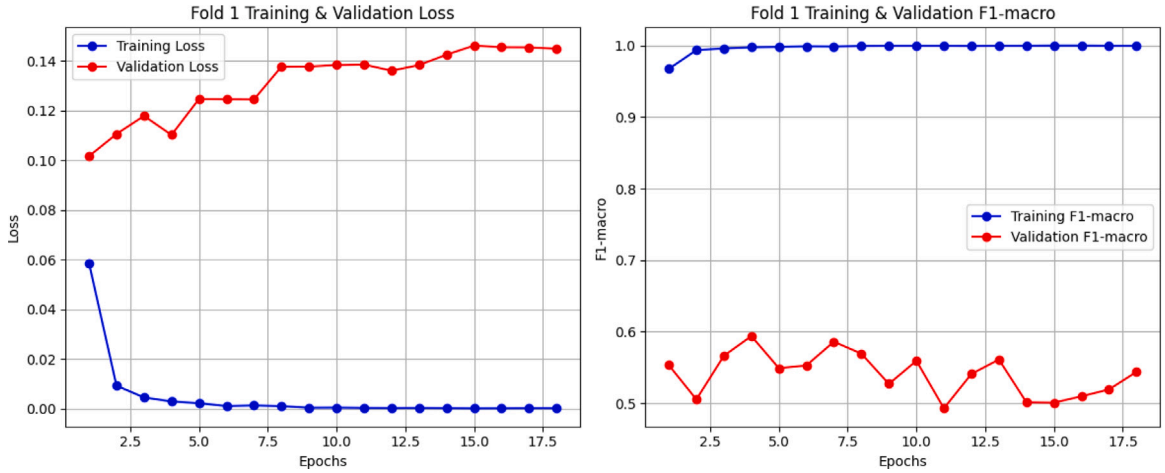
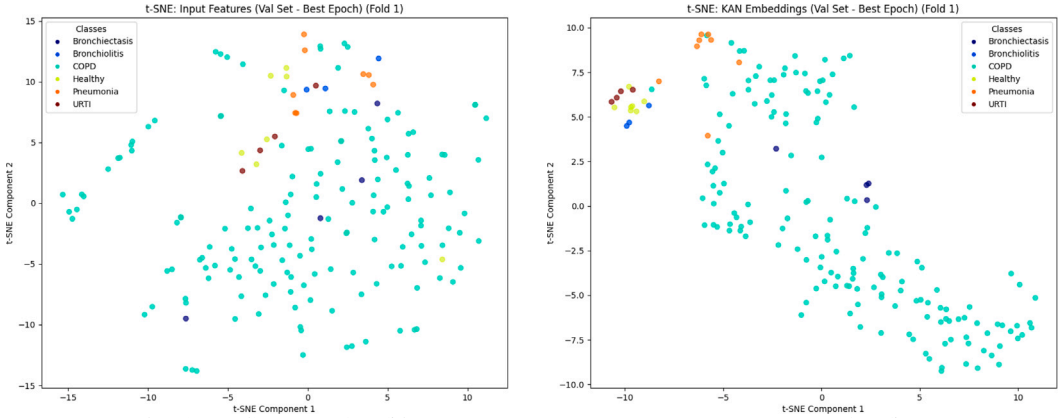


图 6。 第 1 折的训练 和验证 曲线 （代表性），展示 损失 和宏平均 F_1-分数 随 训练轮次的变化 .



(a) 输入 特征 （第 1 折，验证 集）。 (b) KAN 嵌入 (折 1， 验证 集)。

图 7. 折 1（验证 集 - 最佳 训练轮次）特征 空间的 t-SNE 可视化 。

表 6 各 交叉验证 折 （验证 集）上的 性能 指标 。

Fold	宏 F_1-分数	准确率
1	0.5658	92.93%
2	0.5838	91.85%
3	0.7462	95.63%
4	0.7730	96.17%
5	0.8479	96.17%
Mean	**0.7033**	**94.55%**
Std Dev	**0.1103**	**0.0181**

宏平均 F_1-分数为 0.7033，标准差为 0.1103；准确率为 94.55%，标准差为 1.81%。这些结果证实，该模型不仅准确，而且在各折上的表现也相当稳健，尽管从 F_1-分数可以看出，类别级平衡在不同折之间存在一定差异。

表 6 展示 了 在 验证 集 上 获得 的 准确率 和 宏平均 F_1-分数 ，针对 5 折交叉验证中的 每一 折 分别给出结果。

该模型在识别多数类（慢性阻塞性肺疾病， F_1：0.98）方面表现出色，在若干少数类上也取得了令人称赞的表现，包括支气管扩张（F_1：0.84）、健康（F_1：0.76）和肺炎（F_1：0.77）。然而，最罕见的类别——上呼吸道感染（F_1：0.44）和细支气管炎（F_1：0.45）——仍然难以识别。

5. 讨论

5.1. KAN 的 呼吸 特征 可解释性

Kolmogorov–Arnold 网络（KAN）架构的一个主要优点是，可以通过可视化学习到的单变量样条函数 $\phi_{i,j}(x)$ （式（3））来实现其可解释性。与采用固定非线性激活函数（例如 ReLU、Sigmoid）的传统多层感知机（MLP）不同，KAN 可学习的 B 样条能够提供显式且可视化的变换，将输入特征映射到输出神经元。

特征到类别 映射 洞见： 通过 检查 KAN 输出 层 学习到 的 样条函数 ，我们 可以 解读 特定 呼吸 特征 （编码 在 LSTM 输出 中） 如何 影响 疾病 分类 。例如：

- 与谱质心和带宽相关的特征（可指示哮鸣音或喘鸣）可能呈现陡峭的非线性样条曲线，并连接至哮喘或支气管扩张等类别。
- 时域特征（如过零率，与呼吸相位转换相关）对于健康类别预测可能呈现平滑、单调的样条曲线。
- 梅尔频率倒谱系数衍生特征能够捕捉音色和纹理，可能呈现复杂的多峰样条曲线，以区分慢性阻塞性肺疾病和肺炎。

临床相关性：这种可解释性对临床应用至关重要，因为它能让临床医生理解模型作出特定预测的原因。例如，如果KAN样条曲线显示，较高的谱质心值会显著提高支气管扩张的概率，这与已知的临床知识相符：支气管扩张气道会产生高频哮鸣音。这种透明度有助于建立对人工智能诊断系统的信任，并促进数据科学家与医学专家之间的合作。

与黑箱模型的比较：标准深度学习分类器（例如全连接多层感知机、卷积神经网络）缺乏这种程度的可解释性。Grad-CAM 或 SHAP 等技术可以提供事后解释，但无法揭示模型学到的明确函数关系。KAN 基于样条的架构具有内在可解释性，无需额外的解释方法。

未来 工作 中 可解释性： 在 未来 的研究中， 我们 计划 对 每 个 类别 开展 样条函数 详细 分析， 并 关联 学习到的 变换 与 已知 的 呼吸 疾病 声学 生物标志物。 与 肺科医生 合作 以 验证 这些 解读 将 进一步 增强 所 提出模型 的 临床 实用性 。

5.2. 临床 意义 与 部署 可能性

所提出的 LSTM-KAN 混合模型展现出显著潜力，可作为智能听诊工具部署于真实临床环境中。基于人工智能的医疗系统近期进展表明，要弥合研究原型与临床实践之间的差距，必须应对若干关键挑战：模型可解释性、计算效率、与现有工作流程的整合，以及医学专家的验证 [20]。

整合 到 临床 决策 支持 系统 （CDSS）：

1. 床旁诊断：模型推理速度快（每个样本 2.1 毫秒，见第 4.3 节），可在患者问诊期间实时进行呼吸音分类。临床医生可使用连接到运行该模型的平板电脑或智能手机的便携式数字听诊器，即时获取诊断建议。

2. 远程医疗应用：在远程医疗场景中，患者可以在家中使用低成本数字听诊器录制肺音。录制的音频可以传输至基于云的服务器，由模型执行分类，并将结果发送回医疗服务提供者。这对于肺科医生资源有限的医疗服务不足地区尤其有价值。

3. 筛查项目：该模型可部署于大规模筛查项目中，用于在高危人群（例如吸烟者、产业工人）中早期检测呼吸系统疾病，如慢性阻塞性肺疾病或结核病。对多数类具有较高的敏感度（慢性阻塞性肺疾病 F_1：0.98），可确保对常见疾病进行可靠检测。

挑战 ： 临床 部署：

1. 数据异质性：临床环境在录音设备、声学条件（例如环境噪声）和患者人口统计特征方面存在显著差异。该模型基于 ICBHI 数据集进行训练，该数据集包含多样化的录音条件，但仍需在多中心数据集上进行进一步验证，以确保其泛化性。

2. 监管 审批： 基于 AI 的 诊断系统 的部署 需要 获得监管 许可（例如 美国的美国食品药品监督管理局 批准 和 欧洲的 CE 标志 ）。这 需要 开展严格的 临床 试验 ，以证明其 安全性 、 有效性 ，并且对传统诊断 方法 具有 非劣效性。

3. 医学专家验证：尽管模型的准确率较高（94.55%），仍需由肺科医生进行临床验证。我们提出一种分两个阶段的验证方法：（a）回顾性验证：临床专家审查模型基于历史患者数据作出的预测，并评估其与真实诊断的一致性。（b）前瞻性临床试验：在医院环境中部署该模型，并将其诊断建议与临床医生的实时诊断进行比较。我们将评估评估者间一致性（科恩 κ 系数）和诊断一致性等指标。

4. 稀有类别性能：模型对稀有类别的表现一般（上呼吸道感染 F_1：0.44，细支气管炎 F_1：0.45），限制了其检测这些疾病的效用。在临床部署中，系统应对这些类别的低置信度预测发出标记，并建议由专科医生进行随访检查。

部署 架构： 一种 实用的 部署 架构 可以 包括：

•边缘计算：对于延迟敏感型应用（例如急诊室），模型可以运行在与数字听诊器部署在同一地点的边缘设备上（例如配备 GPU 加速器的 NVIDIA Jetson、Raspberry Pi）。•基于云的推理：对于远程医疗，模型托管在云服务器（例如 AWS、Azure）上，并通过 REST API 接入电子健康记录（EHR）系统。•混合方法：采用混合架构，在设备端进行初步分类（使用轻量级模型变体），对于不确定的病例，则将其提交至基于云的服务器，使用完整的混合 LSTM-KAN 模型重新分析。

由医学专家 进行 验证 ： 为 确保 临床 可靠性， 我们 提出 以下 验证 方案 ：

1. 数据收集：与医院合作，收集确诊患者的前瞻性肺音录音数据集（通过影像学检查、肺量测定或活检确诊）。 2. 盲法评估：向由 3–5 名肺科医生组成的评审组提供模型预测结果和音频录音，但不透露模型输出。临床医生独立诊断每个病例。 3. 一致性分析：使用科恩 κ 系数、灵敏度、特异度和阳性／阴性预测值等指标，计算模型与临床医生之间的评定者间一致性。 4. 失败案例分析：针对模型预测结果与临床医生共识不一致的病例，开展详细的声学分析，以识别混杂因素（如噪声、呼吸伪影），并据此优化模型。

长期愿景：借鉴医学影像领域成功部署人工智能的经验[20]，，我们展望未来，呼吸音分类将融入常规临床实践，类似于自动心电图（ECG）解读辅助心脏病学家。该系统不会取代临床医生，而是增强其诊断能力，提供第二诊疗意见并减少诊断错误。

5.3. 优势、 局限性 与 未来 工作

优势：

1. 有效协同运用多种类别不平衡缓解技术（焦点损失、SMOTE、数据增强），消融研究（第 4.2 节）对此进行了验证。2. 创新地应用混合 LSTM-KAN 架构，利用 KAN 基于样条的可解释性进行呼吸音分类。3. 全面的特征工程，提取频谱、时间和倒谱特征。4. 采用稳健的交叉验证，并进行详细的逐类别性能分析。

局限性：

1. 由于训练样本不足，极稀有类别（上呼吸道感染、细支气管炎）的分类仍然困难。2. 对整段录音进行特征聚合，可能导致细粒度时间信息丢失。3. 研究结果基于单一公开数据集（ICBHI 2017）；仍需在独立数据集上进行外部验证。4. 缺乏医学专家在真实场景中的临床验证。

未来 工作：

1. 高级 合成 数据 生成： 探索 生成式 对抗式 网络 （生成对抗网络） 或 扩散 模型 ， 以 生成 高质量 合成 样本 ， 用于 稀有 类别。

2. 端到端时间建模：开发可直接处理帧级频谱图的模型，以保留细粒度的时间动态，而不是聚合特征。3. 深化KAN可解释性：与肺科医生合作，详细分析学得的样条函数，并将其与已知的声学生物标志物关联起来。4. 外部验证：在独立的呼吸音数据集（例如医院特定数据集、使用不同录音设备采集的数据集）上测试模型，以评估其泛化性。5. 临床试验：开展涉及医学专家验证的前瞻性临床研究，以评估模型在真实场景下的诊断准确性、临床实用性和安全性。6. 多模态整合：将呼吸音分析与其他诊断模态（例如肺量测定、胸部X光片、患者病史）相结合，以全面评估疾病。

6. 结论

本文提出了一种用于多类别呼吸音分类的混合LSTM-KAN深度学习模型，专门应对严重的类别不平衡问题。通过结合这一新颖架构、全面的特征工程、焦点损失、SMOTE 和有针对性的数据增强，该模型在不平衡的 ICBHI 数据集上取得了较高的总体准确率（94.55%）和显著的宏平均 F_1 分数（0.7033）。消融研究表明，不平衡缓解技术的协同组合显著优于单独方法；完整的混合方法相较于基线，使宏平均 F_1 分数相对提高了 +20.8%。计算成本分析证实，该模型适用于实时部署：每个样本的推理时间为 2.1 毫秒，且内存需求适中。

尽管模型在极稀有类别（上呼吸道感染、细支气管炎）上的表现仍有提升空间，但本研究展示了一种可行且稳健的方法，可用于改善呼吸系统疾病的自动诊断。研究结果有助于开发智能听诊工具，并凸显了先进神经网络架构结合有针对性的数据处理策略，在应对具有挑战性的真实世界医学分类任务方面的潜力。未来工作将聚焦于与医学专家开展临床验证、使用外部数据集进行评估，以及在真实世界的远程医疗和即时诊断系统中部署，为人工智能辅助呼吸系统疾病诊断的实际应用铺平道路。

CRediT 作者 贡献 声明

Nithinkumar K.V.: 撰写 – 综述 & 编辑， 撰写 – 初稿， 可视化，验证， 正式 分析， 数据 整理， 概念化。 **Anand R.:** 撰写 – 综述 & 编辑， 撰写 – 初稿， 可视化， 正式 分析， 数据 整理， 概念化。

利益冲突声明

作者 声明 他们 不存在 已知的 利益 冲突 或 个人 关系， 这些 关系 可能 会 影响 本文 所 报告 的 研究 成果。

致谢

作者 感谢 Amrita Vishwa Vidyapeetham （印度 科imbatore），感谢其 为本研究 提供 必要的 资源 和 支持。

数据 可用性

该 呼吸 音 数据集 使用 于 本 研究 可 在此获取： ： https://bhichallenge.med.auth.gr/ICBHI_2017_挑战赛。

参考文献

[1] D. Adeloye, S. Chua, C. Lee, C. Basquill, A. Papana, E. Theodoratou, H. Nair, D. Gasevic, D. Sridhar, H. Campbell, K.Y. Chan, A. Sheikh, I. Rudan, COPD患病率的全球和区域估计：系统综述与荟萃分析, J. Glob. Health 5 (2015) 020415. [2] 世界卫生组织，《十大死因》，2020，https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.（访问于2024年）。[3] A. Bohadana, G. Izbicki, S.S. Kraman, 肺部听诊基础，N. Engl. J. Med. 370 (2014) 744–751. [4] A.M. Osman, S. Al-Shehri, M.H. Temsah, A. Al-Eyadhy, A. Alrabiaah, F. Al-Sohime, A.Al. Mutair, R. Halwani, COVID-19患者的远程听诊：一项宝贵却被低估的工具, Telemed. E-Health 27 (2021) 98–103. [5] S. Kumar, A.Y. Shaikh, D.R. Vora, S.B. Aher, 基于机器学习的声音诊断流程，Int. J. Online Biomed. Eng. 20 (2024) 148–165, http: //dx.doi.org/10.3991/ijoe.v20i10.49585. [6] X. Wang, G. Qi, 基于机器学习的声音诊断和疾病预测，IEEE Access 12 (2024) 22788–22803, http://dx.doi.org/10.1109/ACCESS.2024.3361943. [7] H. He, E.A. Garcia, 从不平衡数据中学习，IEEE Trans. Knowl. Data Eng. 21 (2009) 1263–1284. [8] B.M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R.P. Paiva, I. Chouvarda, P. Carvalho, N. Maglaveras, 用于评估呼吸音分类算法的开放获取数据库，Physiol. Meas. 40 (2019) 035001. [9] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T.Y. Hou, M. Tegmark, KAN：Kolmogorov–Arnold网络，2024，arXiv预印本 arXiv: 2404.19756. [10] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, 用于密集目标检测的焦点损失，载于：IEEE国际计算机视觉会议论文集，IEEE，2017，第2980–2988页。

[11] N.V. Chawla、K.W. Bowyer、L.O. Hall、W.P. Kegelmeyer，SMOTE：合成少数类过采样技术，《人工智能研究杂志》16（2002）321–357。[12] R. Anand、S. Samiappan、M. Prabukumar，利用灰狼优化对数字有限脉冲响应滤波器进行微调以实现峰值性能，《科学报告》14（1）（2024）12675。[13] C. Potes、S. Parvaneh、A. Rahman、B. Conroy，基于特征的分类器与基于深度学习的分类器集成，用于异常心音检测，载于：2016年计算心脏病学会议，CinC，IEEE，2017年，第621–624页。[14] J. Salamon、J.P. Bello，用于环境声音分类的深度卷积神经网络与数据增强，《IEEE信号处理快报》24（2017）279–283。[15] R. Anand、S.V. Lakshmi、D. Pandey、B.K. Pandey，利用心电图生物医学指标进行心律失常检测的增强型ResNet-50深度学习模型，《进化系统》15（1）（2024）83–97。[16] D.S. Park、W. Chan、Y. Zhang、C.C. Chiu、B. Zoph、E.D. Cubuk、Q.V. Le，SpecAugment：一种用于自动语音识别的简单数据增强方法，2019年，arXiv预印本arXiv:1904.08779。[17] G. Haixiang、L. Yijing、J. Shang、G. Mingyun、H. Yuanyue、G. Bing，从类别不平衡数据中学习：方法与应用综述，《专家系统与应用》73（2017）220–239。[18] A. Sivasubramanian、D. Sasidharan、V. Sowmya、V. Ravi，利用轻量级CNN注意力深度学习架构进行高效特征提取，用于超声胎儿切面分类，《物理与工程科学医学》（2025）1–15。[19] N.V. Jacob、V. Sowmya、E.A. Gopalakrishnan、R. Ramachandran、A.V. Pillai，基于扰动的可解释人工智能模型用于伤口愈合分级与跨语料库分类的印度数据库，《生物医学计算方法与程序》（2025）108981。[20] K. Arumugam、R. Swathika、S.S. Thasni Tabriz，基于机器学习的心血管疾病诊断，使用心音图信号，载于：2024年智能计算与系统工程第二届国际会议，SCSE，IEEE，2024年，第1–6页，http://dx.doi.org/10.1109/SCSE61872.2024.10550572。