基于集成的框架：阿拉伯语音素发音错误检测

Sükrü Selim Calık[a]， Ayhan Kucukmanisa[b] 和ZeynepHilalKilimci[c]，*

[a]*Maviay Consultancy Company*，科贾埃利大学科技园，*41275*，科贾埃利，土耳其[b]电子与通信工程系，科贾埃利大学，*41001*，土耳其[c]信息系统工程系，科贾埃利大学，*41001*，土耳其

arXiv:2301.01378v1 [cs.SD] 3 Jan 2023

文章信息

关键词：计算机辅助语言学习 阿拉伯语发音检测 集成学习 机器学习 投票分类器

摘要

计算机辅助语言学习（CALL）系统负责判定发音错误，并确保向用户提供反馈。在本研究中，我们提出了一种集成模型，可有效识别阿拉伯语音素的发音错误，并辅助阿拉伯语学习。据我们所知，这是首次全面尝试运用集成学习技术和传统机器学习模型来判定阿拉伯语音素的发音错误。为考察特征提取技术的影响，我们将梅尔频率倒谱系数（MFCC）和梅尔频谱图分别与各学习算法结合。为展示所提模型的效果，我们请 11 位不同的人士分别发出阿拉伯语音素中的 29 个字母，其中 8 个为哈菲兹。通过添加噪声、时间偏移、时间拉伸和音高偏移等方法扩充数据集。大量实验结果表明，采用梅尔频谱图特征提取技术的投票分类器作为集成算法，取得了显著的分类效果，准确率达到 95.9%。

1. 引言

如今，计算机辅助语言学习（CALL）系统十分普及，因为它们有助于人们提升语言能力和发音水平。CALL 主要关注非母语学习者语音中的发音错误等问题。CALL 可用于执行多种任务，例如检测发音错误、语音识别和发音评分。此外，已有许多研究将语音处理应用于各种语言，以辅助语言学习（Cucchiarini 等，1998年；Neumeyer 等，2000年；Minematsu，2004年）。计算机科学，尤其是人工智能的发展，使得人们能够借助 CALL 开展大量研究（Cucchiarini 等，1998年；Neumeyer 等，2000年；Minematsu，2004年；Adnan 和 Zamari，2012年；Hu 等，2013年；Ehsani 和 Knodt，1998年；Arias 等，2010年）。

对于某种语言，使用不同语言的说话者在说话时容易出现发音错误，因为口腔肌肉无法准确发出该语言特有的细微语音差别。因此，研究人员通常致力于研究英语、荷兰语和法语等多种语言中的发音错误，而关于阿拉伯语的研究文献则较为有限。尽管如此，近年来有关阿拉伯语的研究逐渐增多（Abufanas，2013年）、Khan 等人（2013年）、（Muaad 等人，2021年）、（Shaalan 1，2005）、（Nazir 等人，2019年）、（Ziafat 等人，2021年）。阿拉伯语是使用最广泛的语言之一（Lane，2019年），拥有超过 2.9 亿母语使用者和 1.32 亿非母语使用者，也是联合国（UN）的六种官方语言之一。阿拉伯语主要有两种变体：古典阿拉伯语（CA）和现代标准阿拉伯语（MSA）。古典阿拉伯语是《古兰经》使用的语言，而现代标准阿拉伯语则是《古兰经》语言的改良版本，用于日常交流。《古兰经》语言的发音规则有明确的规定，以确保词语含义准确无误。本研究结合机器学习模型和集成学习模型，致力于检测《古兰经》语言中音素的发音错误。

集成学习是一种通过使用多个分类器，相较于单个分类器取得更好效果的技术。换言之，它将一组单独的分类器整合起来，构建基于集成的分类器。生成和整合是集成学习策略的主要阶段。

*通讯作者：Zeynep Hilal Kilimci sselimcalik1@gmail.com（S.S. Calık）；ayhan.kucukmanisa@kocaeli.edu.tr（A. Kucukmanisa）；zeynep.kilimci@kocaeli.edu.tr（Z. H. Kilimci）开放研究者与贡献者身份识别码：

在集成的生成阶段，会基于训练集构建由单个学习器组成的多样化数据集。在整合阶段，通过合并各个独立分类器的结果得出最终决策。集成策略的主要方法是生成多个学习器，并结合各个独立分类器的决策，使基学习器的整合能够提升单个分类器的效果（Rokach，2010年），（Polikar，2006a），（Gopikaand Azhagusundari，2014年），（Renetal.，2016年），Kilimciet al. (2016)。

本研究着重于检测阿拉伯语音素的发音错误，并观察传统机器学习算法和集成学习技术与特征提取方法结合时所产生的影响。为此，我们采用支持向量机、K近邻、决策树、朴素贝叶斯和随机森林分类器，并结合不同的特征提取技术进行评估，具体包括梅尔频率倒谱系数和梅尔频谱图。随后，通过多数投票、提升法、装袋法、与逻辑回归堆叠、与随机森林堆叠以及与极端随机树堆叠等方法，整合各个分类器的决策。实验结果表明，采用梅尔频谱图进行特征提取时，将多数投票作为集成技术可取得显著效果，准确率达到95.3%。

本文的主要贡献如下：•为加速未来的研究，我们与哈菲兹合作创建了自己的阿拉伯字母数据集。•与高性能深度学习模型相比，在计算负载较低的情况下实现更高的准确率。•复杂度低，适用于容量有限的嵌入式平台。

本文其余部分安排如下：第2 节介绍有关阿拉伯语音素发音错误的研究。第3 节介绍本研究采用的单个机器学习分类器和集成学习方法。第4 节给出实验结果。第5节对本文进行讨论并总结。

2. 文献综述

本　　节简要介绍有关阿拉伯语音素发音错误的研究　　d.

Muhammad 等人(2010年) 提出了一种非实时电子哈菲兹系统，旨在避免《古兰经》出现发音错误。为此，系统执行了数据准备、特征提取和建模阶段。在数据准备步骤中实施静音去除和预加重方法后，特征提取阶段依次采用分帧、加窗、离散傅里叶变换、梅尔滤波器组、取对数和离散傅里叶逆变换技术。随后，系统计算相似度以检测发音错误。作者在五名参与者身上将所提出的模型与一个同类网络应用程序进行比较，并指出该模型有助于提升他们的背诵表现。在另一项研究中，(Elsayed 和 Fathy，2019年) 旨在开发一种自动化模型，根据哈菲兹的诵读情况评估其背诵水平。该系统采用梅尔频率倒谱系数作为特征提取技术，并使用矢量量化进行降维。作者指出，实验结果表明，使用所提出的系统评估《古兰经》背诵水平可获得显著的准确率。在另一项研究中（Putra 等人，2012），研究人员开发了一种用于学习《古兰经》的语音识别模型，将高斯混合模型与梅尔频率倒谱系数特征提取技术相结合。研究采用了多种语音信号处理技术，包括最低采样频率、帧分块、加窗、离散傅里叶变换、动态40阶倒谱系数和频谱能量，以及梅尔频率倒谱系数。作者报告的实验结果显示，该模型在发音方面的准确率为70%，在泰吉威德诵读规则方面的准确率为90%，而在发音与泰吉威德诵读规则综合评估方面的准确率为60%。

(Arafa 等人，2018年）提出了一种能够发现发音错误的语音识别系统。该数据集由89名学生组成，其中46名为女性。28个阿拉伯语音素各发音10次。收集890条话语后，提取梅尔频率倒谱系数，用于通过五种不同的机器学习模型进行建模。这些模型分别是K近邻、支持向量机、朴素贝叶斯、多层感知机和随机森林。实验结果表明，随机森林技术的准确率达到85.02%，优于其他机器学习模型。在另一项近期研究中，Nazir 等人，2019年）提出了一种基于发音错误检测的模型，该模型利用深度卷积神经网络的特性检测阿拉伯语音素。首先，从CNN的第4层至第7层提取特征。随后，在训练阶段采用K近邻、支持向量机和神经网络分类器。他们比较了使用手工设计特征与使用深度特征时的实验结果。最后，作者强调，在采用深度特征时，所提出的学习方法显著提升了发音错误判定过程的准确率，达到92.2%。在一项研究中，Akhtar 等人，2020年）提出了一种利用深度卷积神经网络特征来改进

非母语学生阿拉伯语单词发音错误检测的模型。为利用深度特征，从Alex Net的第6层、第7层和第8层提取特征。随后，在训练阶段采用n近邻、支持向量机和随机森林。为展示深度特征的影响，作者还评估了使用梅尔频率倒谱系数提取的特征。采用相同的三种模型，并比较两种方法的结果。他们强调，所提出的模型准确率达到93.2%，能够判定阿拉伯语单词的不正确发音。Maqsood 等人，2016年）集中研究了利用声学语音特征（APF）和支持向量分类器检测阿拉伯语音素的发音错误。为此，作者重点采用声学语音特征，而不是基于置信度度量的分数。该数据集由100名说话者的500条话语构成。这些特征包括均方根能量（RMSE）、梅尔频率倒谱系数、低能量、频谱特征、过零率、基频和统计特征。随后，采用支持向量机确定阿拉伯语音素的发音错误。论文最后指出，支持向量机的准确率达到97.5%。

在另一项研究中，（Farooq 和 Imran，2021年）重点研究阿拉伯字母发音部位的发音错误判定。为此，研究将相对谱变换-感知线性预测特征提取技术与隐马尔可夫模型（HMM）相结合。研究人员收集了 29 个阿拉伯字母的 1,486 个样本。研究采用多种技术与 HMM 配合，例如使用高斯混合模型计算观测概率，或采用前向算法。作者总结指出，所提出的模型能够识别发音错误，识别率达到 85%。在研究（Nazir 等人，2021年）中，研究人员考察了一种基于支持向量机的系统，旨在检测亚洲说话者对 28 个阿拉伯语音素的发音错误。该系统的准确率达到 88%。（Asif 等人，2021年）借助深度神经网络，着重研究阿拉伯语元音的正确发音。由于样本数量有限，研究人员构建了新的数据集，并采用多种技术进行扩充。该数据集包含 85 名说话者的录音，其中 43 名为女性，只有 4 名是非母语说话者。收集到 6,229 条录音后，研究人员进行了预处理，包括降噪、数据分段、重采样和静音截断。为检测阿拉伯语元音的发音错误，研究人员构建了优化的卷积神经网络模型。该模型在古典阿拉伯语音素发音分类任务中的准确率达到 95.77%。（Maqsood 等人，2017年）利用声学语音特征（即梅尔频率倒谱系数），对用于检测阿拉伯语音素发音错误的不同分类器进行了比较分析。研究采用了以下机器学习算法：随机森林、朴素贝叶斯、AdaBoost 和 k近邻。作者报告称，随机森林模型表现最佳，准确率达到 95.4%。

（Shareef 和 Al-Irhayim，2022）专注于详细比较特征提取模型，以检测阿拉伯语障碍语音。特征提取技术基于多种小波变换。为检测阿拉伯语障碍语音，研究构建了长短期记忆网络和卷积神经网络-长短期记忆网络模型。梅尔频率倒谱系数与长短期记忆网络相结合，取得了最佳分类准确率，达到 93%；其次是卷积神经网络-长短期记忆网络，准确率为 91%。（Algabri 等人，2022）为非阿拉伯语母语者设计了发音错误检测系统，旨在向他们提供语音障碍反馈。该系统利用基于深度学习的模型，在词语和句子层面为阿拉伯语学习者提供反馈。实验结果表明，最佳模型的音素检测错误率接近 4%，发音错误检测的 F1 分数约为 70%。（Yang 等人，2022）提出了一种面向英语的改进型发音错误检测系统，该系统基于在线伪标签和 wav2vec 模型。具体而言，该方法使用未标注的第二语言语音执行伪标签流程，并采用预训练的自监督学习模型（wav2vec）来改进微调方法。实验结果显示，与传统的离线伪标签技术相比，该方法取得了显著更好的分数。该模型使音素错误率降低了近 5%，并使第二语言语音发音错误检测的 F1 分数提升了约 2.5%。（Peng 等人，2022）提出了一种文本感知模型，通过提高音频特征相对于无关文本特征得分的权重来检测发音错误。为此，该方法在 TIMIT 和 L2-Arctic 英语数据集中采用了对比学习流程。研究者报告称，与基线相比，所提出的模型提升了约 4%。

我们的研究与上述文献中的工作主要有以下不同：我们专注于基于集成的策略，与高性能深度学习模型相比，该策略能够以较低的计算负载实现更高的准确率。此外，我们自行编制了一个阿拉伯字母数据集，以促进未来的研究。

3. 提出的框架

本文提出了一种基于集成机器学习的阿拉伯语音素发音错误检测方法。图 1 展示了所提出系统的流程图。该系统主要包含三个阶段。第一步是进行预处理，以提高声音质量和后续步骤的性能。接下来，对从麦克风获取的原始音频信号进行特征提取。最后，考察基于机器学习的方法在该问题上的性能，然后通过对这些方法采用集成学习方法来提升性能。本文使用的阿拉伯语音素是图 2 中展示的阿拉伯字母。该图呈现了字母的独立形式及其发音。

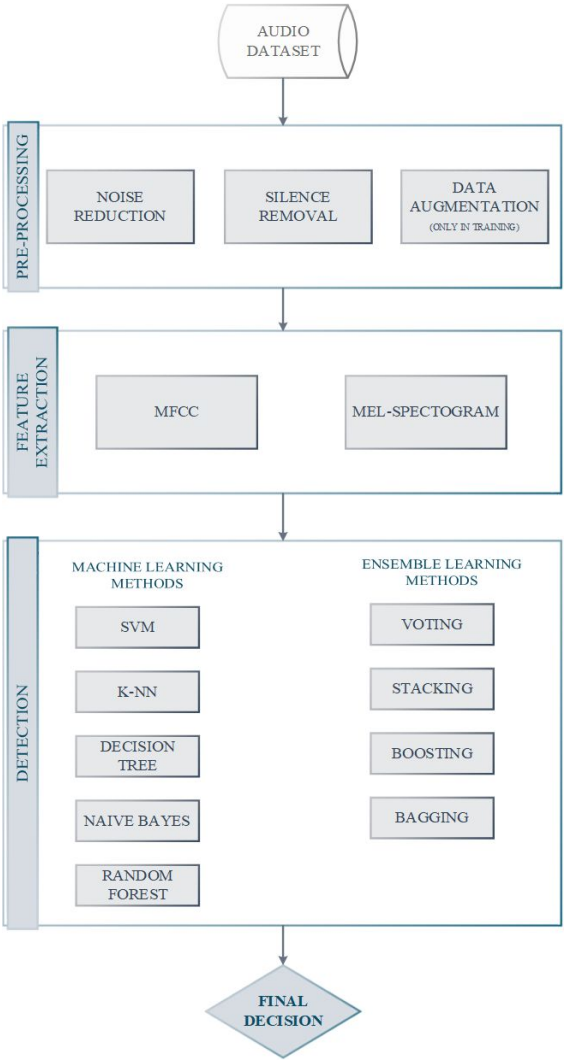

图1： 所提出模型的流程图。

3.1. 预处理

预处理步骤旨在从语音与噪声的混合声音中分离出干净语音，并仅截取含有音频信号的部分。降噪依赖一种称为“频谱门控”的方法，它是噪声门的一种形式，主要用于改善语音识别。随后进行静音去除。语音开始的标志是声音电平升至某个 dB 值以上。在本研究中，将音频信号首次升至 30 dB 以上、随后降至 30 dB 以下的片段截取出来，作为算法的输入。

Isolated Form	ح	ج	ث	ت	ب	ا
Pronunciation	haa	jiim	thaa	taa	baa	alif
Isolated Form	س	ز	ر	ذ	د	خ
Pronunciation	siin	zaay	raa	thaal	daal	kha
Isolated Form	ع	ظ	ط	ض	ص	ش
Pronunciation	ayn	thaa	taa	daad	saad	shiin
Isolated Form	م	ل	ك	ق	ف	غ
Pronunciation	miim	laam	kaaf	qaaf	faa	ghayn
Isolated Form		لا	ي	و	ه	ن
Pronunciation		lamelif	yaa	waaw	ha	nuun

图2：阿拉伯字母的独立形式及其发音。

图 3展示了降噪和静音去除操作前后的一个输入信号样本。作为最后一项预处理操作，通过数据增强增加数据量。所使用的数据增强技术包括添加噪声、时间偏移、时间拉伸和音高偏移。此过程不用于测试集，仅用于训练集。

3.2. 特征提取

人类语音具有各种鲜明特征。然而，由于噪声较多且数据量较大，通常不倾向于直接处理原始音频信号。研究发现，从音频信号中提取鲜明特征，并将其作为基础机器学习模型的输入，比直接将原始音频信号作为输入能获得更好的性能。梅尔频率倒谱系数（Dave，2013年）和梅尔频谱图（Stevens 等人，1937）是从音频信号中提取特征的两种常用技术。

梅尔频谱图由128个梅尔特征和窗口长度为2048的快速傅里叶变换（FFT）构成。获得的特征维度因音频时长而异。因此，需要将特征向量的大小缩放至单一维度。获得梅尔特征后，通过取平均值将其缩放至 1×128 （缩放前的维度例如为 128×48, 128×40 ）。最终，问题转化为对大小为 1×128的特征向量进行分类。

MFCC特征是经过一系列处理后得到的，具体包括：对音频信号加窗、应用离散傅里叶变换（DFT）、计算振幅对数、在梅尔刻度上扭曲频率，以及应用逆离散余弦变换（DCT）。经过这些处理，可得到39个特征，其中最后20个特征的区分度更高。本研究使用这最后20个特征作为输入特征。与梅尔频谱图类似，音频文件大小会导致特征维度发生变化。因此，对所有特征取平均值后，问题便转化为对大小为 1×20 的数据进行分类。

图4展示了特征向量的缩放过程，图 5 展示了两个字母（“Alif”和“Baa”）的平均梅尔频谱图和MFCC特征。在图 5中，不同颜色代表不同哈菲兹的语音。如

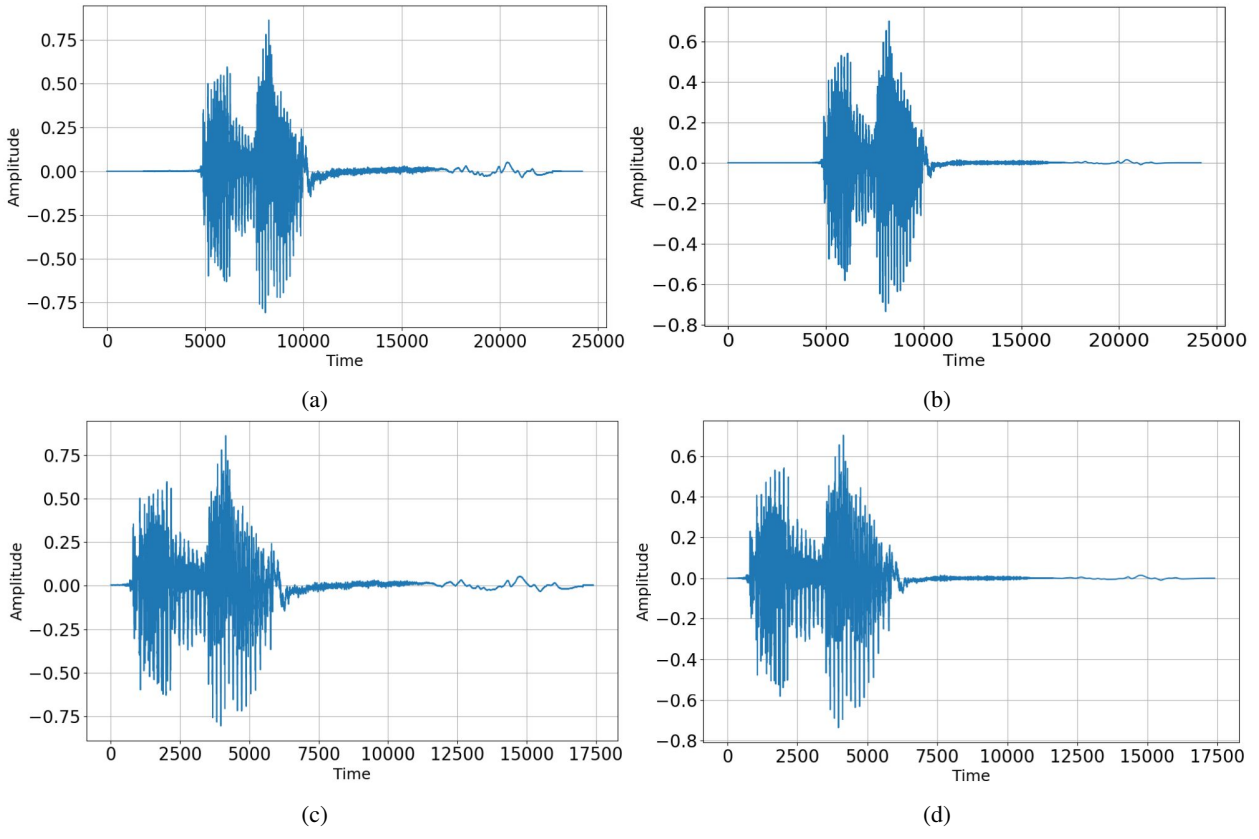

图3： 降噪和静音去除预处理操作：(a) 原始音频信号；(b) 降噪后；(c) 降噪并去除静音后；(d) 降噪和静音去除后。

从该图可以看出，即使不同哈菲兹说的是同一个字母，所形成的特征向量也会呈现不同的特征。此外，同一个人在发不同字母时，其特征差异并不明显。由于这些困难，分类方法的性能变得更加重要。

3.3. 检测

3.3.1. 机器学习方法

机器学习可定义为人工智能算法，这些算法能够从数据中推断和预测，以模仿人类的学习方式。各种机器学习算法可用于解决分类、回归和聚类任务。本研究重点介绍适用于分类问题的常用方法，包括支持向量机（SVM）、k近邻（k-NN）、决策树（DT）、朴素贝叶斯和随机森林。

支持向量机（Cristianini 和 Shawe-Taylor，2000年）是一种监督学习方法。根据算法运行时数据的类型，也可以使用核函数。这样便可执行线性和非线性分类操作。其目标是用一个超平面将所有数据分开。然而，如果数据无法完全分开，就不能用单个平面对其进行分类。因此，需要使用不同的核函数。超平面周围会确定一个间隔。间隔的大小会直接影响分类性能。间隔可通过“C”超参数进行控制。C 越大，间隔越窄。此外，如果模型过拟合，则需要减小 C。本研究采用线性核，并将 0.02 作为 C 参数。

k近邻 (Mucherino 等人，2009)基本上是根据训练集中的最近“k”个邻居，确定类别未知的数据所属的类别。通过测量测试数据与训练数据之间的距离，确定最近的“k”个最近邻。随后，根据这些标签确定测试数据的类别值。在本研究中，将 3、4 和 5 作为“k”值进行评估。

决策树**的基本用途**(Alp，2019年是将数据集划分为更小的子组，使其在视觉上更易于

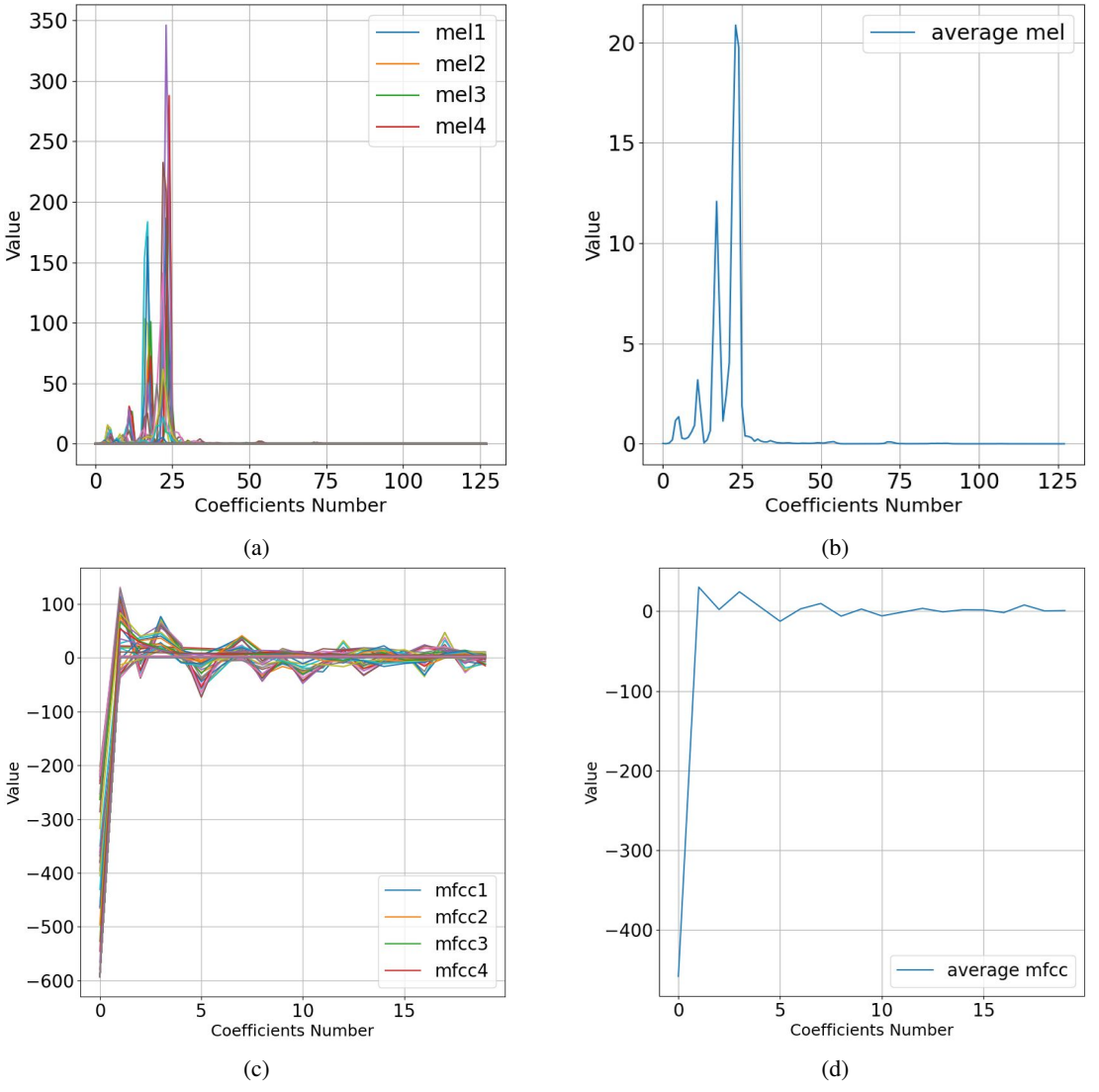

图 4： 特征向量的缩放（a）原始梅尔频谱图图像（b）平均梅尔频谱图图像（c）原始梅尔频率倒谱系数图像（d）平均梅尔频率倒谱系数图像。

在特定规则（决策规则）的框架下理解。由于算法的输出是一个视觉上类似于树的流程图，因此称为决策树。决策树有 4 种基本结构：根节点、节点、分支和叶节点（终端节点）。根节点是分类过程的起点。如果观测数据具有同质结构，它们自然会属于同一类别，分类过程也会在根节点处分支之前结束。对于异质观测数据，根节点会根据最能将观测数据划分为不同类别的特征分成两个或多个分支，并创建新的节点。树中最后一个不再分支的节点是终端节点，表示观测数据被分配到的类别。

朴素贝叶斯(Webb等人，2010年）分类基于贝叶斯定理。它用于估计一组特定特征属于某个特定类别的概率。其目标是通过概率计算，选择概率最高的决策。每个属性都被视为独立于该类别中的其他属性。根据不同属性划分不同类别。与更复杂的方法相比，朴素贝叶斯分类器的运行速度极快。

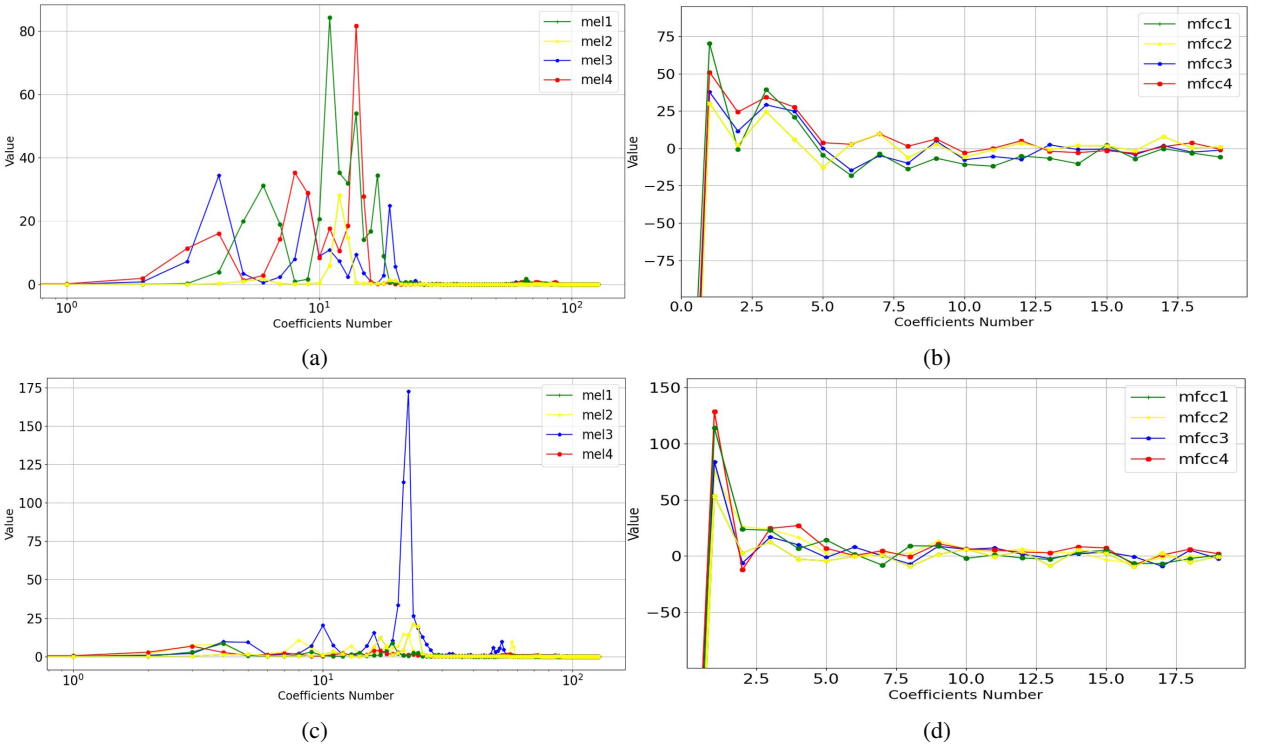

图5： “Alif”和“Baa”字母的平均特征向量：(a) “Alif”的平均梅尔频谱特征；(b) “Alif”的平均梅尔频率倒谱系数特征；(c) “Baa”的平均梅尔频谱特征；(d) “Baa”的平均梅尔频率倒谱系数特征。

3.3.2. 集成学习技术

集成学习是一种利用多种学习算法来提高分类框架整体准确率的方法。与单个学习模型相比，这种策略通常能取得更好的结果。文献中最常用的集成学习技术包括装袋法、堆叠、提升法和投票法。

装袋法 是应用最广泛的集成算法之一（L.，1996年）。它是自助聚合的简称。该方法通过重采样实现多样化：从整个训练数据集中随机、有放回地抽取多个训练数据子集。每个子集用于训练集成学习器中的一个不同基学习器。最终决策通过结合各个学习器的决策并进行多数投票得出。

提升法 （Freund 和 Schapire，1996年）是另一种集成学习模型，被认为是装袋法的替代方案。该方法的核心思路是生成一组独立分类器，这些分类器使用经过加权处理的数据子集。该方法通过迭代，在训练数据集的不同分布上运行基学习器来实现。第一步中，所有实例的权重相同；之后，每次迭代都会根据前一个基学习器的训练误差更新分类错误实例的权重。每个学习器都使用从更新后的训练数据集中选取的实例子集。随后，前序学习器预测错误的实例会比预测正确的实例更频繁地被选取。最终结果通过对基学习器预测的类别进行加权多数投票得出。文献中采用的变体包括 AdaBoost、AdaBoost.M1、AdaBoost.M2、AdaBoost.R、Arcing 和 Real Adaboost（L.，2010年）、（Polikar，2006b）、（Gopika D.Azhagusundari，2014年）、（Ren Y.，2016年）。本文采用 AdaBoost.M1。

堆叠是将多种类型的模型拟合到同一数据集上，然后利用元层模型学习如何以最佳方式整合预测结果的过程（Džeroski，2004年）。该过程主要分为两个步骤。第一步是利用训练集生成一组独立分类器。因此，元数据集由独立分类器的决策组成。在构建元数据集之后，使用该数据对元层学习器进行建模

，以获得泛化估计。在我们的工作中，元数据集由原始训练样本和独立分类器的决策共同构成。此外，在本研究中，堆叠法与不同的学习算法结合使用，即逻辑回归、随机森林和极端随机树。

在**多数投票**中，每个模型都会为每个测试样本做出一个预测（投票），最终结果的预测值由获得票数最多的模型给出。

随机森林(Breiman，2001年）是一组决策树分类器。它是装袋法的一种具体实现，其中每个独立分类器都是一棵决策树。装袋法用于为每棵基决策树选择训练子集。随机森林采用的划分任务不同于广为人知的决策树策略；在该策略中，每个节点都根据所有其他属性中最优的属性进行划分。在随机森林方法中，首先随机选取一组属性子集，然后在该属性子集中确定最佳划分。除装袋法外，这种方法还为模型引入了额外的随机性。由于样本空间和属性空间中都引入了随机性，随机森林能够应对过拟合问题。

图6 展示了集成架构的概览。

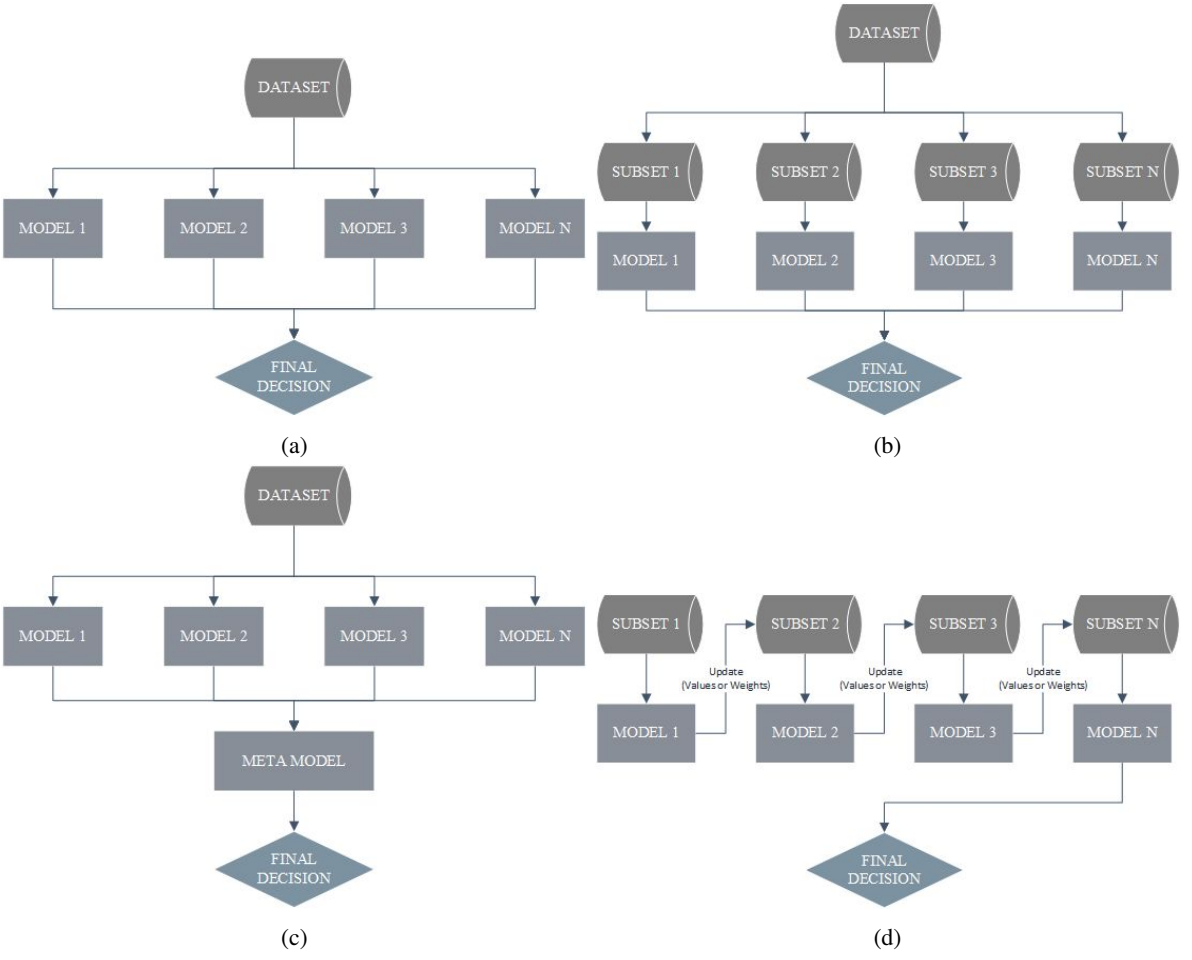

图 6： 集成架构（a）投票法（b）装袋法（c）堆叠（d）提升法。

4. 实验结果

本研究通过采集 11 位说话者的音频样本构建了一个原始数据集，其中 8 位是哈菲兹。数据集由阿拉伯字母组成。研究人员要求说话者逐个朗读每个字母，并对所有字母进行录音。原始数据集包含 290 个样本。通过添加噪声、时间偏移、时间拉伸、

表 1基于 MFCC 特征的机器学习方法评估结果

方法	准确率	精确率	召回率	F1分数
SVM	**0.758**	**0.829**	**0.785**	**0.806**
k近邻 (k=3)	0.466	0.527	0.466	0.494
k近邻 (k=4)	0.308	0.310	0.308	0.309
k近邻 (k=5)	0.307	0.34	0.307	0.322
决策树	0.714	0.735	0.710	0.722
朴素贝叶斯	0.392	0.445	0.392	0.417
随机森林	0.756	0.78	0.754	0.766

表 2使用梅尔频谱图特征的机器学习方法评估结果

方法	准确率	精确率	召回率	F1分数
SVM	0.952	**0.970**	0.952	0.960
k近邻（k=3）	**0.953**	0.968	**0.953**	**0.960**
k近邻（k=4）	0.698	0.734	0.698	0.716
k近邻（k=5）	0.670	0.700	0.67	0.684
决策树	0.827	0.852	0.823	0.837
朴素贝叶斯	0.538	0.68	0.538	0.600
随机森林	0.938	0.947	0.938	0.942

音高变换增强方法，将数据集的样本数量增加至 1450 个。数据集分为 80% 的训练集和 20% 的测试集。采用五折交叉验证程序，以确保所提模型的得分不受训练集和测试集子集划分方式的影响。所提方法的性能通过公式 (1)、(2)、(3) 和 (4) 进行衡量。在这些公式中，如果将每个类别命名为 C_x，TP（真正例）表示属于 C_x 的音频被正确分类为 C_x。FP（假正例）是所有被分类为 C_x 的非 C_x 样本。TN（真负例）表示所有未被分类为 C_x 的非 C_x 样本。FN（假负例）表示所有未被分类为 C_x 的 C_x 样本。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

基于机器学习的方法的评估结果见表1和表2。表1展示了将MFCC特征用作机器学习方法输入时的效果。同样，表2展示了将梅尔频谱图特征用作输入时的效果。在这些表中，粗体数值表示最佳结果。根据表1，支持向量机优于其他方法。在表2中，可以认为k值设为3的k近邻表现良好。综合评估表1和表2可知，最高结果由k近邻方法取得，其输入特征为梅尔频谱图特征。

将集成学习方法应用于表1和表2所列的方法。集成学习方法的评估结果见表3和表4。在这些表中，粗体数值表示最佳结果。查看表3可知，与其他方法相比，采用随机森林的堆叠方法表现最佳。表4清楚地表明，投票法取得了最佳结果。

表 3采用 MFCC 特征的集成学习方法评估结果

方法	准确率	精确率	召回率	F1分数
投票法	**0.772**	0.794	0.764	0.778
堆叠（逻辑回归）	0.700	0.720	0.704	0.711
堆叠（随机森林）	0.761	**0.804**	**0.768**	**0.785**
堆叠（极端随机树）	0.731	0.767	0.747	0.756
提升法	0.744	0.793	0.740	0.765
装袋法	0.750	0.781	0.762	0.771

表 4采用梅尔频谱图特征的集成学习方法评估结果

方法	准确率	精确率	召回率	F1分数
投票法	**0.959**	**0.969**	**0.957**	**0.962**
堆叠（逻辑回归）	0.866	0.893	0.862	0.877
堆叠（随机森林）	0.951	0.960	0.947	0.953
堆叠（极端随机树）	0.938	0.944	0.939	0.941
提升法	0.871	0.932	0.855	0.891
装袋法	0.943	0.953	0.935	0.943

表 5基于深度学习的方法的性能评估

方法	准确率	精确率	召回率	F1分数	时间（秒）
(Abdoli 等，2018年)	0.910	0.920	0.910	0.915	0.470
(Choi 等，2018年)	0.900	0.900	0.900	0.900	0.612
所提方法	**0.959**	**0.969**	**0.957**	**0.962**	**0.091**

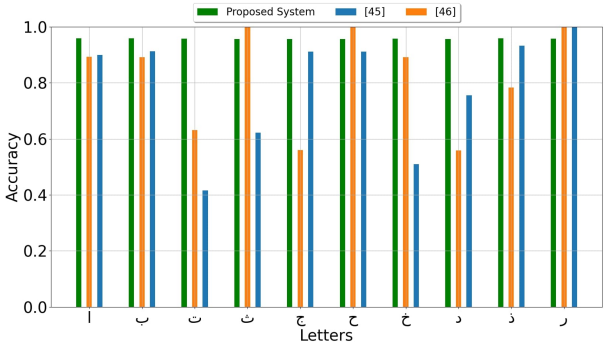

图 7： 基于深度学习的方法的基于字母的性能评估。

表5从处理速度和检测性能两方面，对文献中近期提出的方法与本文所提方法在声音分类方面进行了比较。比较使用的是本研究范围内构建的数据集。表5中用于比较的方法均为基于深度学习的方法。这些方法按照论文中指定的原始网络结构和参数实现。在此表中，所提方法指以梅尔频谱图特征为输入的投票集成方法。如表5所示，所提方法的性能优越，且处理速度在其他方法中最快。此外，还分析了表5所列方法基于字母的性能。图7 展示了（Abdoli 等人，2019年）、Choi 等人（2018年）的方法以及所提方法在10个字母上的表现。可以看出，尽管其他方法在某些字母上的性能大幅下降，所提方法的表现仍然稳定。表5和图7 表明，凭借所提方法，在阿拉伯语音素发音错误检测方面可以超越流行的深度学习方法。

所提出的集成方法。此外，值得注意的是，由于复杂度较高，基于深度学习的方法通常需要更长的训练时间和处理（推理）时间。阿拉伯语发音错误检测是一项具有挑战性的问题。值得注意的是，所提出的框架能够以较低的计算负载，在该问题上实现更优的性能。

5. 讨论与结论

本研究提出了一种集成学习方法，用于检测阿拉伯语音素的发音错误。在所提方法中，首先采用梅尔频率倒谱系数和梅尔频谱图方法进行特征提取。随后，传统方法和基于集成学习的方法以这些特征为输入，并在为本研究创建的原始数据集上评估其性能。评估中性能最高的方法还与基于深度学习的方法进行了比较。实验结果表明，将投票分类器用作集成算法，并采用梅尔频谱特征提取技术，可取得显著成果，准确率达到 95.9%。未来的研究将着重于扩充数据集并采用迁移学习技术。

参考文献

Abdoli, S., Cardinal, P., Lameiras Koerich, A., 2019年。使用一维卷积神经网络进行端到端环境声音分类。Expert Systems with Applications 136, 252–263。Abufanas, O., 2013年。面向第二语言学习者的阿拉伯语计算机辅助语言学习系统。International Journal of Educational and Pedagogical Sciences 7, 3189–3193。Adnan, A.H.M., Zamari, Z.M., 2012年。计算机辅助自主语言学习：印度尼西亚、马来西亚和新西兰从业者的观点。Procedia-Social and Behavioral Sciences 67, 49–60。Akhtar, S., Hussain, F., Raja, F.R., Ehatisham-ul haq, M., Baloch, N.K., Ishmanov, F., Zikria, Y.B., 2020年。利用深度卷积神经网络特征改进非母语学习者的阿拉伯语单词发音错误检测。Electronics 9, 963。Algabri, M., Mathkour, H., Alsulaiman, M., Bencherif, M.A., 2022。针对非母语者阿拉伯语语音的发音错误检测与诊断及发音动作层级反馈生成。Mathematics 10, 2727。Alp, S., O.E., 2019年。机器学习中的分类方法及R语言应用。Nobel Akademik Yayıncılık，土耳其伊斯坦布尔。Arafa, M.N., Elbarougy, R., Ewees, A., Behery, G., 2018年。用于支持阿拉伯语音素发音的语音识别数据集。International Journal of Image, Graphics and Signal Processing 10, 31。Arias, J.P., Yoma, N.B., Vivanco, H., 2010年。面向计算机辅助语言学习的自动语调评估。Speech communication 52, 254–267。Asif, A., Mukhtar, H., Alqadheeb, F., Ahmad, H.F., Alhumam, A., 2021年。利用深度学习对古典阿拉伯语音素进行发音分类的方法。Applied Sciences 12, 238。Breiman, L., 2001年。随机森林。Machine Learning 45, 5–32。Choi, K., Fazekas, G., Sandler, M., Cho, K., 2018年。用于音乐标签分类的深度神经网络音频信号预处理方法比较，第1870–1874页。Cristianini, N., Shawe-Taylor, J., 2000年。支持向量机导论。剑桥大学出版社，英国剑桥。Cucchiarini, C., Wet, F.d., Strik, H., Boves, L., 1998年。通过自动语音识别技术评估荷兰语发音。Dave, N., 2013年。语音识别中的特征提取方法：线性预测倒谱、感知线性预测和梅尔频率倒谱系数。International Journal for Advance Research in Engineering and Technology 1, 1–4。Džeroski, S., Z.B., 2004年。通过堆叠组合分类器是否优于选择最佳分类器？Machine Learning 54, 255–273。Ehsani, F., Knodt, E., 1998年。计算机辅助语言学习中的语音技术：一种新型CALL范式的优势与局限。Language Learning & Technology 2, 54–73。Elsayed, E.K., Fathy, D., 2019年。基于OWL本体的系统对古兰经诵读进行评估。Int. Arab J. Inf. Technol. 16, 970–977。Farooq, J., Imran, M., 2021年。利用机器学习检测阿拉伯字母发音部位上的发音错误，载于：2021年计算、电子与电气工程国际会议（ICE Cube），IEEE，第1–6页。Freund, Y., Schapire, R., 1996年。新型提升算法的实验，载于：国际机器学习会议；意大利，第148–156页。Gopika, D., Azhagusundari, B., 2014年。分类任务中集成方法的分析。Gopika D. Azhagusundari, B., 2014年。分类任务中集成方法的分析。International Journal of Advanced Research in Computer and Communication Engineering 3, 7423–7427。Hu, W., Qian, Y., Soong, F.K., 2013年。基于新型DNN的高质量发音评估，面向计算机辅助语言学习（CALL），载于：Interspeech，第1886–1890页。Khan, A.F.A., Mourad, O., Mannan, A.M.K.B., Dahan, H.B.A.M., Abushariah, M.A., 2013年。面向计算机辅助语言学习的阿拉伯语发音自动评分，载于：2013年第一届通信、信号处理及其应用国际会议（ICCSPA），IEEE，第1–6页。Kilimci, Z.H., Akyokus, S., Omurca, S.I., 2016年。针对土耳其语和英语文本的同质集成分类器有效性，载于：2016年智能系统与应用创新国际研讨会（INISTA），IEEE，第1–7页。L., B., 1996年。装袋预测器。Machine learning 24, 123–140。L., R., 2010年。基于集成的分类器。Artificial Intelligence Review 33, 1–39。Lane, J., 2019年。世界上使用人数最多的十大语言。Babbel Magazine 6。

Maqsood, M., Habib, H., Anwar, S., Ghazanfar, M.A., Nawaz, T., 2017年。针对易混淆项的基于分类器的发音错误检测系统比较研究。The Nucleus 54, 114–120。Maqsood, M., Habib, H.A., Nawaz, T., Haider, K.Z., 2016年。基于支持向量机的完整阿拉伯语音素发音错误检测系统。Inter- national Journal of Computer Science and Network Security (IJCSNS) 16, 30。Minematsu, N., 2004年。基于学习者发音结构与目标语言词汇结构兼容性的发音评估，载于：第八届国际口语处理会议。Muaad, A.Y., Jayappa, H., Al-antari, M.A., Lee, S., 2021年。Arcar：一种新型深度学习计算机辅助识别方法，用于字符级阿拉伯语文本表示与识别。Algorithms 14, 216。Mucherino, A., Papajorgji, P.J., PM., P., 2009年。K近邻分类。Data Mining in Agriculture 34, 83–106。Muhammad, W.M., Muhammad, R., Muhammad, A., Martinez-Enriquez, A., 2010年。面向《古兰经》诵读者的语音内容匹配系统，载于：2010年第九届墨西哥国际人工智能会议，IEEE，第148–153页。Nazir, F., Majeed, M.N., Ghazanfar, M.A., Maqsood, M., 2019年。利用深度卷积神经网络特征和基于迁移学习的模型检测阿拉伯语音素发音错误。IEEE Access 7, 52589–52608。Nazir, F., Majeed, M.N., Ghazanfar, M.A., Maqsood, M., 2021年。基于亚洲说话者错误频率的阿拉伯语发音错误检测系统。Mehran University Research Journal Of Engineering & Technology 40, 279–297。Neumeyer, L., Franco, H., Digalakis, V., Weintraub, M., 2000年。发音质量自动评分。Speech communication 30, 83–93。Peng, L., Gao, Y., Lin, B., Ke, D., Xie, Y., Zhang, J., 2022。文本感知端到端发音错误检测与诊断。arXiv预印本 arXiv:2206.07289。Polikar, R., 2006a。用于决策的基于集成的系统。IEEE Circuits and systems magazine 6, 21–45。Polikar, R., 2006b。用于决策的基于集成的系统。IEEE Circuits and Systems Magazine 6, 21–45。Putra, B., Atmaja, B., Prananto, D., 2012年。开发用于古兰经经文诵读学习软件的语音识别系统。IJID（International Journal on Informatics for Development）1, 14–21。Ren, Y., Zhang, L., Suganthan, P.N., 2016年。集成分类与回归的最新进展、应用及未来方向。IEEE Computational intelligence magazine 11, 41–53。Ren Y., Zhang L., S.P.N., 2016年。集成分类与回归的最新进展、应用及未来方向。IEEE Computational Intelligence Magazine 11, 41–53。Rokach, L., 2010年。基于集成的分类器。Artificial intelligence review 33, 1–39。Shaalan 1, K.F., 2005年。面向阿拉伯语学习者的智能计算机辅助语言学习系统。Computer Assisted Language Learning 18, 81–109。Shareef, S.R., Al-Irhayim, Y.F.M., 2022年。阿拉伯语语音障碍特征提取技术比较。Al-Rafidain Engineer- ing Journal (AREJ) 27, 190–197。Stevens, S.S., Volkmann, J., B., N.E., 1937年。用于测量基频心理量值的量表。Journal of the Acoustical Society of America 8, 185–190。Webb, G.I., Keogh, E., R., M., 2010年。朴素贝叶斯。Encyclopedia of Machine Learning 15, 713–714。Yang, M., Hirschi, K., Looney, S.D., Kang, O., Hansen, J.H., 2022。采用基于wav2vec2的动量伪标签改进发音错误检测，以评估口音程度和可懂度。arXiv预印本 arXiv:2203.15937。Ziafat, N., Ahmad, H.F., Fatima, I., Zia, M., Alhumam, A., Rajpoot, K., 2021年。利用深度学习检测阿拉伯字母的正确发音。Applied Sciences 11, 2508。