基于深度学习的《古兰经》基础诵读规则发音错误检测

艾哈迈德·阿尔·哈雷雷*，克鲁德·阿尔·贾拉德

叙利亚德拉省阿拉伯国际大学信息与通信工程系。

*通讯作者　　thor。电子邮箱：201710632@aiu.edu.sy；贡献作者：k-aljallad@aiu.edu.sy；

摘要

在伊斯兰教中，诵读者必须遵循一套称为泰吉威德诵读规则的发音规则，以便按照天使吉卜利勒传授给先知穆罕默德的方式诵读《古兰经》。传统上，要学会正确运用这些规则，需要由持有资格认证且经验丰富的人来检测发音错误。由于世界各地的穆斯林人数不断增加，如今泰吉威德教师的数量已不足以满足每位穆斯林日常练习诵读的需求。因此，为了帮助诵读者以比传统学习方式更轻松、更短的时间正确诵读《古兰经》，许多研究致力于自动检测泰吉威德诵读规则中的发音错误。以往的研究普遍存在三个问题。第一，大多数研究只关注机器学习算法。第二，研究使用的是私有数据集，缺乏可供比较的基准。第三，尽管语音信号属于时间序列，但研究并未充分考虑输入数据的序列特性。为解决这些问题，我们提出了一种方案，将梅尔频率倒谱系数（MFCC）特征与利用时间序列的长短期记忆（LSTM）神经网络相结合，以检测泰吉威德诵读规则中的发音错误。此外，我们在公开数据集 QDAT 上进行了实验。该数据集包含超过 1500 条语音，涵盖三条泰吉威德诵读规则（独立拉伸[1], 正午紧密[2], 和隐藏[3]）的正确与错误诵读。据我们所知，尚无研究论文使用过 QDAT 数据集。我们将所提出的 LSTM 模型与当前最先进技术（SoTA）中采用的传统机器学习算法进行了性能比较。结合时间序列的 LSTM 模型明显优于传统机器学习方法。LSTM 在 QDAT 数据集上针对三条规则（独立拉伸、正午紧密和隐藏）取得的准确率分别为 96%、95% 和 96%。

关键词：深度学习、语音识别、梅尔频率倒谱系数（MFCC）、泰吉威德规则、计算机辅助语言学习、发音错误检测

1. 引言

由于语音在人类声音交流中具有重要作用，计算机辅助语言学习（CALL）系统在过去十年中获得了极大的重视。CALL 系统是利用信息技术来教授和学习一门新语言的计算机应用程序[1], ，这种语言通常不是学习者的母语，有时也被称为第二语言。此类系统旨在使学习过程更加轻松，因为传统学习方法往往面临诸多挑战，例如学习成本高昂，或缺乏足够数量的外语教师。CALL 系统包括计算机辅助发音训练（CAPT）系统，该系统能够诊断和识别发音错误，并提供有用的反馈 [2], 。识别发音错误使语言学习者能够准确评估自身发音的正确性。

[1]独立拉伸：指位于单词末尾的延长字母，以及其后一个单词开头的字母“哈姆扎”。总共延长四或五个音拍[24]。[2] 正午紧密：在重读的“正午”字母中，鼻音发音需要持续两个音拍[24]。[3] 隐藏：辅音字母“正午”或坦温以非重读状态发音，持续两个音拍，介于清晰发音与消隐之间[24]。

发音错误检测已在多种语言领域受到广泛关注 [3], [4]。然而，与其他语言相比，阿拉伯语受到的关注较少 [5], ，尽管阿拉伯语十分重要，因为《古兰经》是以阿拉伯语降示的，而且世界各地所有非阿拉伯穆斯林都学习阿拉伯语，以履行其宗教义务并正确诵读《古兰经》。

《古兰经》的正确诵读需要遵循一套称为泰吉威德诵读规则的发音规则。遵循这些规则，意味着像先知穆罕默德那样诵读经文，以保留其含义。传统的泰吉威德诵读学习需要由真人对学习者的诵读进行反馈。指导教师必须持有许可，才能教授《古兰经》。此外，持有泰吉威德许可的人不能立即担任指导教师，因为他们必须积累丰富经验，才能以较高的准确率识别错误。因此，古兰经诵读规则指导教师数量不足，无法帮助所有穆斯林每天练习诵读，并为每次诵读练习提供反馈。

尽管在学习过程中有真人教师具有众所周知的优势，但也存在一些缺点，因此自动化学习过程可以帮助克服这些缺点。这类系统将成为清真寺教师的助手，而非替代者。此类系统的目标是帮助尽可能多的学习者在更短时间内背诵《古兰经》，并每天纠正诵读错误。

阿拉伯语发音错误检测领域的大多数研究都是针对一般阿拉伯语开展的 [5]–[10], ，而只有部分研究人员专注于检测《古兰经》诵读中的发音错误。遗憾的是，这些研究人员全部使用了私有数据集，且没有任何人将其数据集公开，以供其他研究人员使用 [11]–[20]。我们使用了一个公开数据集，因为我们希望未来的研究人员能够将他们的研究结果与我们的结果进行比较，从而使自动化流程在实际应用中更加完善、高效。我们的实验针对一组全新的泰吉威德规则开展，据我们所知，这些规则尚未在任何研究论文中得到研究。通过使用长短期记忆神经网络，我们在检测这些规则相关的发音错误方面取得了高达 96% 的准确率。

2. 相关研究

尽管关于《古兰经》诵读识别的研究数量有限，但近年来这一领域已逐渐受到研究人员的关注。许多研究致力于识别诵读的基本要素、字母和单词 [13], [14], [16], [19], ，或检测泰吉威德规则中的发音错误 [11], [12],[15], [17], [18], [20]。这些论文采用了大量不同的技术，其中一些是传统方法，例如隐马尔可夫模型（HMM） [16], [19] 和高斯混合模型（GMM） [14], ；另一些则是机器学习模型，例如支持向量机 [12], [18], [20] 和多层感知机（MLP） [15], [17],，具体如下：

Tabbal 等人在 [19] 开发了一种分隔器，可自动从音频文件中分离经文，并开发了用于阿拉伯语和《古兰经》诵读的自动语音识别器（ASR）。他们在一个私有数据集上使用了隐马尔可夫模型分类器和梅尔频率倒谱系数（MFCC）特征；该数据集仅包含一小时的忠诚章诵读音频。该系统取得的最高准确率为女性 85%、男性 90%。

Al-Ayyoub 等人利用 [18] 机器学习构建了一个用于古兰经诵读规则自动识别的系统。该系统能够判断以下八项泰吉威德规则的诵读正确性：

双唇鼻音同化[4], 双唇鼻音隐读[5], 重读 Lam[6], 轻读 Lam[7], 鼻音 Noon 规则[8] （Noon）、鼻音 Noon 同化至 Meem、鼻音 Noon 同化至 Waw 和鼻音 Noon 同化至 Ya'。作者使用了一个包含 3,071 个音频文件的数据集，每个文件均包含八种规则之一的录音，且每段录音只涉及一种规则（既包括规则的正确使用，也包括错误使用）。此外，作者比较了多种特征提取技术，例如线性预测编码（LPC）、梅尔频率倒谱系数（MFCC）、多信号小波包分解（WPD）和卷积受限玻尔兹曼机（CRBM）。在分类方面，作者采用了多种分类器，例如K近邻（KNN）、支持向量机（SVM）、人工神经网络（ANN）和随机森林（RF），其中支持向量机（SVM）的准确率达到96%。

Alagrami 等人在 [20] 中提出了一种利用阈值评分和支持向量机（SVM）的系统，用于自动识别四种不同的泰吉威德规则（Edgham Meem、Ekhfaa Meem、takhfeef Lam 和轻读 Lam），其准确率达到 99%，其中采用滤波器组进行特征提取。所使用的数据集包含约 657 条来自阿拉伯语母语者和非母语者的记录，每条规则包含 160 条记录，且每条记录均对应于该规则的正确发音或错误发音。

Ahmad 等人在 [17] 开发了一种泰吉威德分类模型。该方案聚焦于一组称为"静音努恩规则"的诵读规则，尤其是"合并音"及其带鼻音和不带鼻音的规则。特征提取和分类过程采用了梅尔频率倒谱系数（MFCC）和多层感知机（MLP）；神经网络则使用带动量的梯度下降、弹性反向传播和 Levenberg-Marquardt 优化算法进行训练。Levernberg Marquardt 算法取得了最高的测试准确率（77.7%），其次是带动量的梯度下降（76.7%）和弹性反向传播（73.3%）。所用数据集包含两位著名诵读者诵读的 300 个音频文件，每个文件均为其中一条泰吉威德规则的诵读。

在 [16], Rahman 等人提出了一种自动泰吉威德检查系统，以帮助儿童学习正确诵读《古兰经》。该系统使用梅尔频率倒谱系数（MFCC）进行特征提取，并使用隐马尔可夫模型（HMM）进行分类和识别。借助 HMM 算法，系统可以将儿童的诵读与数据库中存储的教师正确诵读进行比较，从而识别并标出其中的任何差异或不一致。不过，该系统仅支持《古兰经》中的一章——《开端章》。

在 [15], Hassan 等人介绍了一种识别大震音[9] (القلقلة الكبرى)发音的系统，该系统使用多层感知机（MLP）作为分类器，并使用梅尔频率倒谱系数作为特征。所用数据集仅包含《古兰经》中五个不同单词的 50 个诵读采样点，每个单词各有十条发音正确和错误的记录，取得的结果介于 95% 至 100% 之间。

Putra 等人 [14] 开发了一种语音识别系统，作为《古兰经》经文诵读学习软件。该系统使用梅尔频率倒谱系数（MFCC）进行特征提取，并采用高斯混合模型作为分类器。为了测试纠正的可靠性和准确性，研究人员收集了一个数据集，其中十位说话者分别以错误和正确的方式诵读若干经文。对于字母发音（阿拉伯字母），纠正准确率达到 90%；对于诵读规则（"合并音""Ikhfa'"或"Idhar"），达到 70%；对于发音与诵读规则的组合，达到 60%。

[4] 双唇鼻音同化：将两个词通过"م"合并；如果一个词以"م"开头，后面接一个静音"م"，则使用"鼻音" [49]. [5] 双唇鼻音隐读：如果[ب]后面接一个静音"م"，则在隐读静音"م"的同时发出"鼻音"，然后继续发[ب] [49]. [6] 重读 Lam：如果真主或 Allahum 一词前有"Fatha"（ً）或"Dhamma"（ٌ），其中的"Lam"就要重读。[49]. [7]轻读 Lam：如果真主或 Allahum 一词前有"Kasrah"（ٍ），其中的"Lam"就要轻读。[49].[8]Edgam Noon：如果任何一个"合并音"字母出现在静音"ن"或"坦温"之后，诵读者应发出后一个字母，而不是"ن"或"坦温" [49]. [9] قَلْقَلَةْ：指字母发音结束时声音的振动。قَلْقَلَةْ的字母：قطب جد [49].

Muhammad 等人 [13] 提出了 E-Hafiz 系统，以方便学习诵读《古兰经》。研究使用梅尔频率倒谱系数（MFCC）、矢量量化（VQ）以及向量间距离计算来提取特征、减少特征向量的数量，并将结果与阈值进行比较。研究使用了 10 段由专家诵读的《古兰经》前 5 个苏拉的录音数据集，男性、儿童和女性的识别准确率分别达到 92%、90% 和 86%。

纳哈尔等人于 [12] 采取了不同的方法，提出了一种识别模型，用于识别《古兰经》诵读中的“诵读方式”，其准确率达到96%。这是因为根据取自 [21], 的第5041号圣训传述，《古兰经》有七种主要的诵读模式，即所谓的“古兰经诵读法”，这些诵读法被公认为最流行的《古兰经》诵读方式，此外还有三种补充性的诵读方式。该模型采用了梅尔频率倒谱系数（MFCC）特征和支持向量机（SVM）。作者构建了一个包含10个类别的数据集，每个类别代表一种《古兰经》诵读类型或“诵读方式”，总计包含258个波形文件。

表1研究人员之间的比较

Work	诵读规则	数据集	特征提取	模型	结果
[19]	《忠诚章》与最重要的诵读规则	约 1 小时音频的诵读的《古兰经》章节 Al-伊赫拉斯（Ikhlass）for 不同诵读者	Mel-频率倒谱系数（梅尔频率倒谱系数）	隐藏马尔可夫模型（隐马尔可夫模型）基于 Sphinx	85% - 90%
[15]	一条泰吉威德规则（قلقلة）	100 个采样点针对 قلقلة 发音	Mel-频率倒谱系数(梅尔频率倒谱系数)	多层感知机（多层感知机）	95% - 100%
[11]	检查发音部位（Makharij）(place of 阿拉伯语的语音片段）音素和音位变体。	Ten 组合方式字母的发音来自 10 位用户，从而导致 20,000 采样点。	Mel-频率倒谱系数（MFCC）	Mean 平方误差（MSE）	最高可达 100% 准确适用于单对单模式
[18]	八条泰吉威德诵读规则(双唇鼻音同化,双唇鼻音隐读,重读 Lam, 轻读 Lam, 鼻音 Noon 同化至 Noon, 鼻音 Noon 同化至 Meem, 鼻音 Noon 同化至 Waw 和 鼻音 Noon 同化至 Ya’))	3071 个音频文件，每个文件包含一段录音恰好属于八条规则中的一条	卷积的受限玻尔兹曼机器（CRBM），梅尔频率倒谱系数 WPD, HMM-SPL	支持向量机器（支持向量机）	96.4%
[20]	四条泰吉威德规则（إدغامميم，إخفاء 米姆（Meem）、轻读拉姆（Takhfeef Lam）、轻读 Lam	657 录音内容来自 4种不同的规则总数,	Mel-频率倒谱	支持向量机器（支持向量机）	99%

		约有 160 条记录 对于每条规则	系数 （梅尔频率倒谱系数）		
[17]	“合并音”规则及其与不带“鼻音”	300 个音频文件	Mel-频率倒谱系数（MFCC）	多层感知机 多层感知机（多层感知机）	77.7%
[16]	《开端章》。	一项正确诵读的 苏拉特·阿尔-开端章。	Mel-频率倒谱系数（梅尔频率倒谱系数）	隐藏马尔可夫模型（隐马尔可夫模型）	N/A
[13]	经文的词级位置在 the 数据集	由 10 个组成专家诵读的前五个苏拉中的《古兰经》	Mel-频率倒谱系数（梅尔频率倒谱系数）	阈值基于欧几里得氏的距离	86% - 92%
[14]	字母级及部分诵读规则	一种声音 录制一段 a 《古兰经》 诵读专家	Mel-频率倒谱系数（梅尔频率倒谱系数）	高斯混合模型（高斯混合模型）	60% - 90%
[12]	将诵读分类为十种“古兰经诵读法”之一	258个波形文件	Mel-频率倒谱系数（梅尔频率倒谱系数）	支持向量机（支持向量机）	99%

3. 方法论

本文提出了一种深度学习模型，利用长短期记忆神经网络检测一组新的泰吉威德规则中的发音错误。数字语音信号被视为一系列音频帧。每帧由一组采样点 [22], 组成，因此发音错误可能出现在若干个连续帧上。长短期记忆网络具有内部记忆，能够利用这一帧序列做出正确判断，因此适合用于检测发音错误；与其他研究相比，后者在做出判断时未能充分考虑帧序列 [23] [12], [15], [17], [18], [20]。

我们提出的解决方案框架如图1所示：首先对语音信号进行预加重并从中提取特征，然后将这些特征输入深度学习模型，最后判断泰吉威德规则的发音正确与否。以下各小节将详细说明每个步骤。

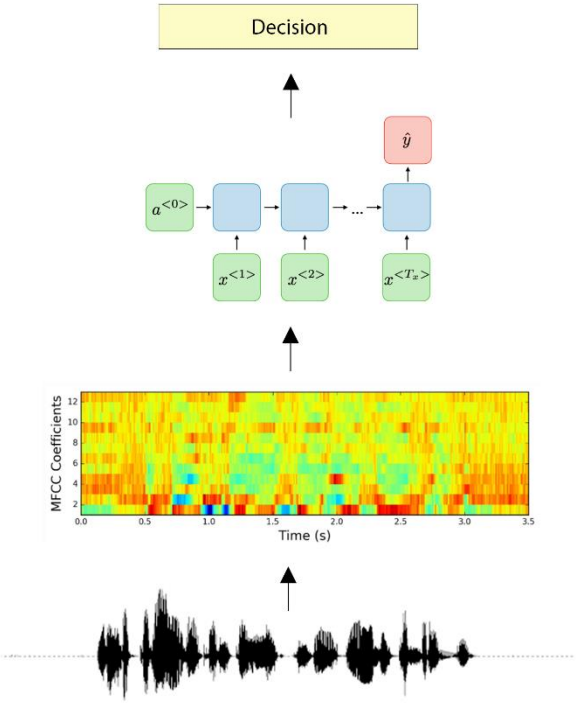

图1 我们提出的解决方案框架

3.1 数据集

尽管已有许多研究致力于识别《古兰经》中的泰吉威德规则，但这些研究均未公开其数据集供其他研究人员使用。本研究采用了一个公开数据集 QDAT [24]，其中包含超过 1500 段正确和错误诵读三种泰吉威德规则的语音：独立拉伸[10]、正午紧密[11]和隐藏[12]。据我们所知，此前尚无其他研究人员对此开展过研究。这些音频片段由不同性别和不同年龄组的人员录制，内容包括《筵席章》第 109 节经文的一部分。录音通过 WhatsApp 在线录制，采用 WAV格式，采样率为 11kHz，单声道，位分辨率为 16 位。

3.2 预加重

预加重是一种信号处理技术，通过提高高频频带的振幅并降低低频频带的振幅，来补偿人类语音产生过程中高频的衰减。这项技术有助于声学模型更好地检测音素，并降低这些模型对噪声的敏感度 [25]。公式 1 中的一阶有限脉冲响应（FIR）滤波器是最常用的预加重滤波器。

$$H(z) = 1 - \alpha z^{-1} \tag{1}$$

[10]独立拉伸：指最后一个词中的延长字母，而哈姆扎则位于其后紧接着的第一个词中。持续时间为四拍或五拍。[11] 正午紧密：必须通过两个节拍来体现重读正午音中的鼻音。 [12] 隐藏：指以介于清晰发音与省略之间的方式发出静音辅音字母或坦温，不加重音，同时保持两个节拍的鼻音。

其中，α 为预加重系数，在本研究中取值为 0.97。

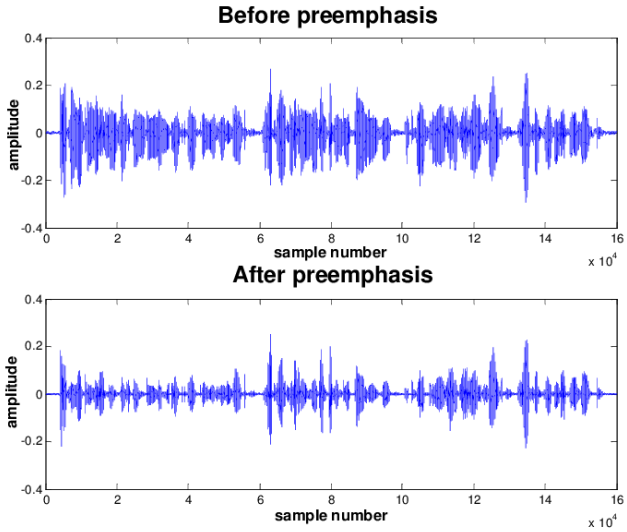

图 2 强调前后的信号 [26]

3.3 特征提取

计算梅尔频率倒谱系数（MFCC）是语音识别应用中最广泛、最主要的频谱特征提取方法 [27]，，该方法基于傅里叶变换和梅尔频率尺度的频域，以模拟人耳的尺度 [28]。Tewari [29] 指出了提取 MFCC 特征的流程，如图 3 所示。

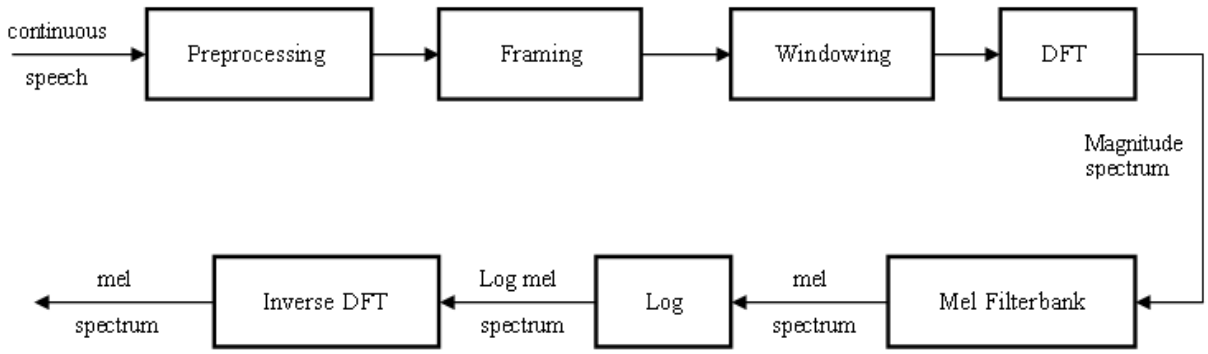

图 3 MFCC（梅尔频率倒谱系数）计算步骤框图 [29]

3.3.1 框架构建

分帧是将语音划分为多个帧，每帧持续时间为 20 至 40 毫秒，其中每帧由若干采样点组成，并可用一个向量表示。该过程通过将非平稳信号转换为准平稳信号，使其能够进行傅里叶变换，从时域转换到频域，从而获得用于语音分析的稳定声学特性 [30]。

3.3.2 加窗

加窗过程用于减小并消除信号每帧开头和结尾处非预期的衰减，这种衰减会在对信号进行离散傅里叶变换时在高频域产生噪声。最著名的函数之一是汉明窗 [25]，本文采用了该函数，其定义如下方程 2 所示：N 是窗长，n 是帧中的第 n^{th} 个采样点。

其中 0 ≤ n ≤ N-1：

$$W_n = \begin{cases} 0.52 - 0.46\cos(\frac{2\pi n}{N-1}), & 0 \le n \le N-1 \\ 0 \quad , & others \end{cases} \tag{2}$$

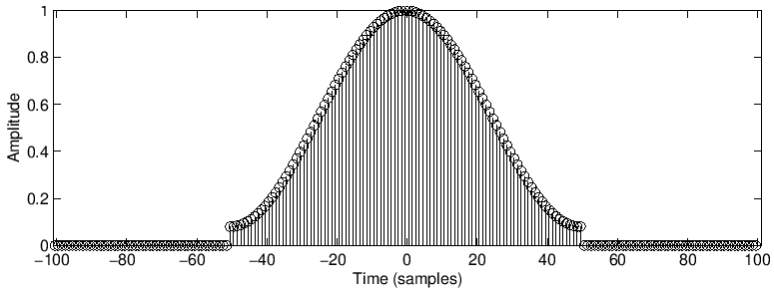

图 4 汉明窗 [31]

3.3.3 离散傅里叶变换（DFT）

对每一帧进行离散傅里叶变换，将信号从时域转换到频域，以计算频谱。通常使用快速傅里叶变换（FFT）算法来计算 DFT，而不是直接使用 DFT，[32] 因为它的复杂度低于 DFT。如公式 3 所示，DFT 的输入是加窗后的信号 $Y1[k]$，输出是复数 $Y2[n]$，表示该频率分量的幅值和相位。

$$S[n] = \sum_{k=0}^{N-1} \mathrm{Y}[k] e^{-2\pi jkn/N} \tag{3}$$

其中，N 是采样点数，$k = 0$、1、2、3、……、$N-1$，而$S[n]$ 是$Y[k]$的傅里叶变换。

3.3.4 梅尔滤波器组

与高频相比，人类对语音中的低频成分更敏感，也能从中获取更多有用信息。人类听觉系统对音调物理频率的感知更接近对数关系，而非线性关系，这正是 Mel 标度的特点。它展示了频率（Hz）与 Mel 标度之间的对应关系：1 kHz 以下的频率间隔近似线性，1 kHz 以上则呈对数间隔 [33]。这一过程通过多个三角滤波器组实现，这是最常用的滤波器形状，其带宽会随频率升高而增大。Hz到Mel 的映射如公式 4 [34] 所示：

$$Frequency\ (Mel\ Scale) = 2592 * log10\ (1 + f/700) \tag{4}$$

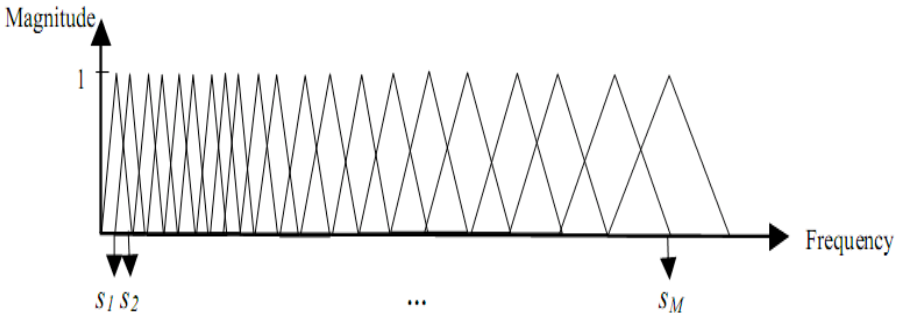

图 5 梅尔滤波器组构建 [35]

3.3.5 对数

正如人类听觉系统对低频下的频率差异更敏感一样，对于低能量信号，人类对能量差异也比对高能量信号时更敏感；人类对信号强度的响应呈对数关系 [36]。因此，对 Mel 滤波器组的输出应用对数函数，以模拟人类的听觉。

3.3.6 离散傅里叶变换的逆变换（IDFT）

由于滤波器组之间存在重叠，前述步骤输出的能量彼此相关，因此需要对其应用逆离散傅里叶变换，这是获得梅尔频率倒谱系数特征的最后一步。该步骤通过公式 5 [25], 得到特征向量，其中 $Y[n]$ 为倒谱系数，$x[k]$ 表示上一步中梅尔滤波器生成的每个语音片段的对数值，N 为梅尔频率倒谱系数的数量。

$$Y[n] = \sum_{k=0}^{N-1} x[k] \cos\left[\frac{n\left\langle k - \frac{1}{2}\right\rangle \pi}{N}\right], n = 1,2,3 \ldots, N \tag{5}$$

3.4 长短期记忆网络（LSTM）

循环神经网络（RNN）是一种人工神经网络，由 Rumelhart 等人提出，[37],能够处理序列数据或时间序列数据。RNN 通常用于序数或时序问题，其独特之处在于具备“记忆”功能，可以利用早期输入中的数据来影响当前的输入和输出。传统的深度神经网络则不同，它们将输入与输出视为彼此独立。RNN 存在梯度消失 [38] 和梯度爆炸 [39] 问题，因此 Hochreiter 和 Schmidhuber 提出了长短期记忆网络（LSTM）[23] , 该网络克服了这些问题，并能够学习长期依赖关系。LSTM 已被应用于语言翻译、自然语言处理、语音识别、图像字幕生成等复杂问题领域 [40]–[43]。

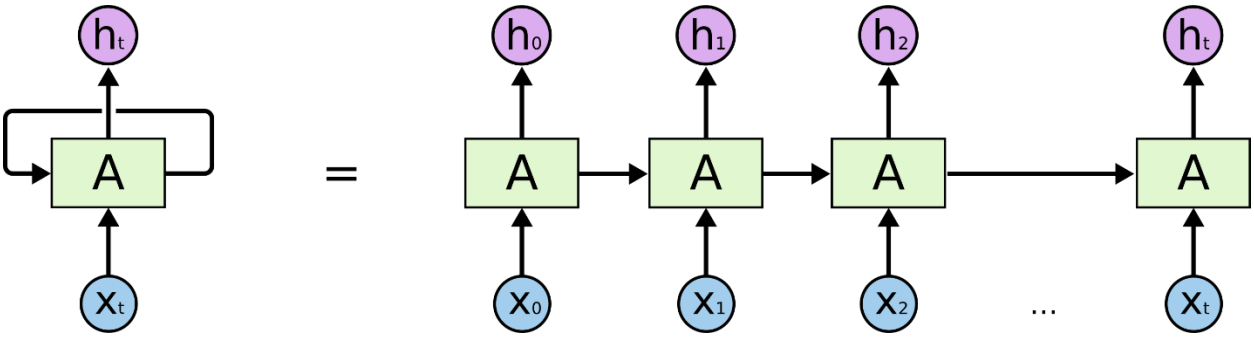

图 6 循环神经网络结构 [44]

LSTM 的核心思想是细胞状态，它在序列处理过程中携带相关信息，发挥网络记忆的作用。信息通过称为门的神经网络添加到细胞状态或从中删除；这些门由逐点乘法运算和 sigmoid 神经网络层组成 [23]。

图 7 展示了长短期记忆网络（LSTM）单元的架构，其中每个 LSTM 单元由 C_t 时刻的细胞状态 t 以及三个主要门组成：输入门 i_t、遗忘门 f_t 和输出门 o_t 。

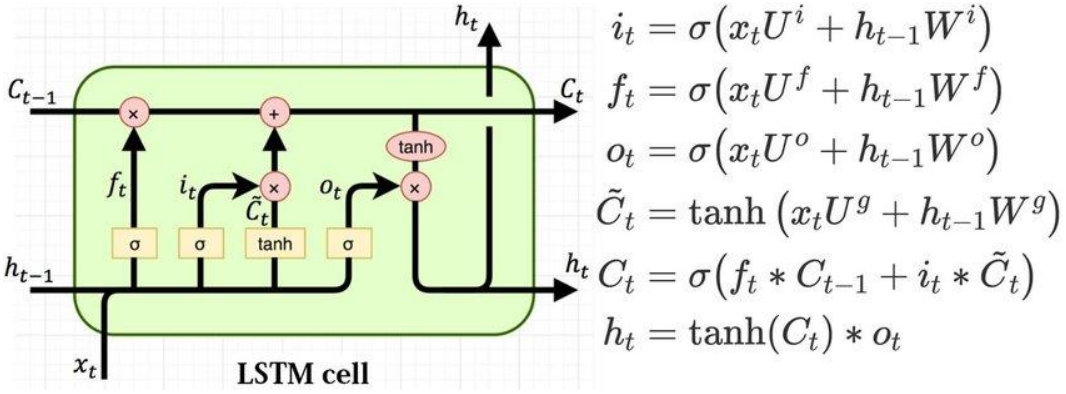

图 7 长短期记忆网络（LSTM）架构及方程 [45]

3.5 实验设置

为提取梅尔频率倒谱系数（MFCC）特征，我们对 32 毫秒的帧应用了汉明窗函数，并使用了 40 个梅尔滤波器组。这些数值是在参数调优过程中经过多次实验后确定的。数据集分别按 90% 和 10% 划分为训练集和测试集。训练集用于训练深度学习模型，该模型由三个长短期记忆网络（LSTM）层组成，每层包含 256 个单元；此外还包括三个全连接层，分别包含 16、16 和 8 个单元，并均使用 RELU 函数作为激活函数；两个 Dropout 层，丢弃率为 0.2；最后，为每条规则设置一个全连接层作为分类层，并使用 sigmoid 函数作为激活函数。该模型还采用了二元交叉熵损失函数和学习率为 0.0005 的 RMSprop 优化器。

4. 结果与讨论

我们将深度学习模型的结果与先前研究中提出的传统机器学习方法在 QDAT 数据集上的结果进行了比较，因为先前研究使用的是私有数据集。

表2 通过深度学习和机器学习模型所达到的准确率

	独立拉伸	Hide	正午紧密
随机森林	91%	83%	91%
KNN	91%	81%	94%
SVM	81%	71%	85%
逻辑回归	88%	79%	91%
长短期记忆网络（我们的模型）	**96%**	**95%**	**96%**

我们可以从以下事实来解释这些结果：在语音识别等复杂问题中，使用深度学习优于使用传统机器学习算法 [46]；此外，随着训练数据规模的增加，深度学习能够取得更好的结果 [47][48]。此外，长短期记忆网络的使用对准确率的影响最大，其表现明显优于其他研究中的方法，因为与其他研究人员所使用的传统机器学习算法相比，它能够更好地考虑声学帧的序列信息。

表3 结果详情

	独立拉伸	Hide	正午紧密
准确率	96%	95%	96%
召回率	95%	95%	98%
F1分数	95%	95%	97%

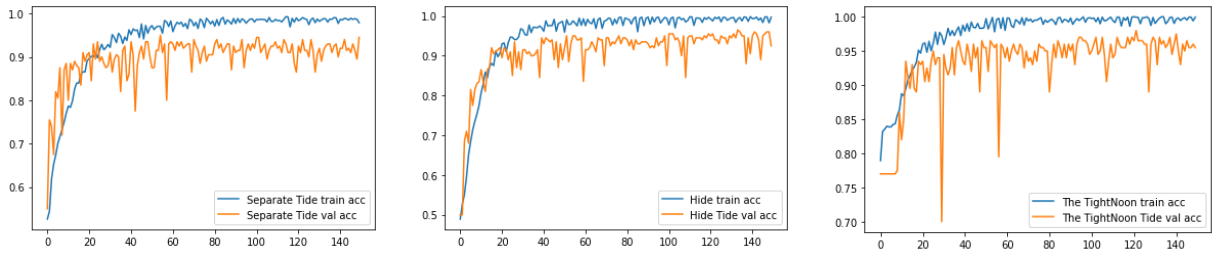

图 8 准确率曲线

5. 结论

正确诵读泰吉威德规则是阅读《古兰经》的必要条件。尽管由于传统诵读学习方法缺少足够的人工指导者，自动检测这些规则中的发音错误十分必要，但研究人员对此关注甚少。因此，许多论文致力于自动化泰吉威德学习。然而，所有先前研究都存在三个共同问题：第一，它们只关注机器学习；第二，它们使用私有数据集，缺乏可供比较的基准；第三，尽管语音信号是一系列帧，它们却未能充分考虑输入数据的时序特性。为解决这些问题，我们提出了一种方案，将梅尔频率倒谱系数（MFCC）特征与利用时间序列的长短期记忆网络（LSTM）神经网络相结合，以检测诵读规则中的发音错误。此外，我们在公开数据集 QDAT 上进行了实验，并将结果与先前研究中的传统机器学习算法进行了比较；我们的方案表现更优。所提出的方案取得了出色的结果，对所有规则的检测准确率均达到 96%。

6. 声明

作者贡献

AAH 进行了文献综述、开展了实验并撰写了稿件。KAJ 承担了指导工作，并参与了研究构思与分析。所有作者均阅读并批准了最终稿件。

资金投入

作者声明，他们未获得任何资助。

数据可用性

本研究使用的数据集可在以下位置获取：

https://www.kaggle.com/datasets/annealdahi/quran-recitation

伦理审批和参与知情同意

作者已获得伦理审批并取得参与知情同意。

发表同意

作者同意发表。

利益冲突

作者声明不存在任何竞争性利益。

致谢

我要感谢我的父母一直以来给予我的支持。

参考文献

[1] Michael. Levy，《计算机辅助语言学习：背景与概念化》，牛津大学出版社，第298页，1997年，访问日期：2022年9月29日。[在线]。获取地址：https://www.amazon.com/Computer-Assisted-Language-Learning-Context-Conceptualization/dp/019823631X[2] C. Tsurutani，《计算机辅助发音训练与评估（CAPTA）程序》，载于计算机辅助外语教学与学习，IGI Global，2013年，第276–288页。doi: 10.4018/978-1-4666-2821-2.ch016.[3] C. Agarwal 和 P. Chakraborty，《英语计算机辅助发音训练（CAPT）的工具与技术综述》，《教育与信息技术》*2019*年第*24*卷第*6*期，第24卷，第6期，第3731–3743页，2019年7月，doi: 10.1007/S10639-019-09955-7.[4] J. H. Woo 和 H. Choi，《基于人工智能的语言学习工具系统性综述》，2021年10月，访问日期：

：

2022年10月3日。[在线]。可访问：http://arxiv.org/abs/2111.04455

[5] F. Nazir、M. N. Majeed、M. A. Ghazanfar 和 M. Maqsood，《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的发音错误》，IEEE Access，第7卷，第52589–52608页，2019年，doi：10.1109/ACCESS.2019.2912648。[6] J. Farooq 和 M. Imran，《利用机器学习检测阿拉伯字母发音部位的发音错误》，2021年计算、电子与电气工程国际会议，ICE Cube 2021论文集，2021年，doi：10.1109/ICECUBE53880.2021.9628251。[7] M. Maqsood、H. Habib、S. Anwar、M. ali Ghazanfar 和 T. Nawaz，《易混淆阿拉伯语音素对的基于分类器的发音错误检测系统比较研究》，The Nucleus，第54卷，第2期，第114–120页，2017年6月。[8] M. Maqsood、H. Habib 和 T. Nawaz，《使用支持向量机检测阿拉伯语音素发音错误的完整系统》，International Journal of Computer Science and Network Security，第16卷，第3期，第30页，2016年9月。[9] M. Algabri、H. Mathkour、M. Alsulaiman 和 M. A. Bencherif，《面向非母语者阿拉伯语语音的发音错误检测与诊断及发音部位级反馈生成》，Mathematics 2022，第10卷，第2727页，第10卷，第15期，第2727页，2022年8月，doi：10.3390/MATH10152727。[10] S. Akhtar 等，《利用深度卷积神经网络特征提升非母语学习者阿拉伯语单词的发音错误检测》，Electronics 2020，第9卷，第963页，第9卷，第6期，第963页，2020年6月，doi：10.3390/ELECTRONICS9060963。[11] N. Arshad 等，《利用梅尔频率倒谱系数识别《古兰经》诵读中的发音部位》，International Journal of Intelligent Information Processing，第4卷，第45–53页，2013年10月，doi：10.4156/ijiip.vol4.issue2.5。[12] R. M. Al-Khatib、M. Alshanaq、M. Barhoush 和 K. Nahar，《基于支持向量机学习模型的高效《古兰经》诵读识别器》，Jordanian Journal of Computers and Information Technology，第6卷，第392–414页，2020年10月，doi：10.5455/jjcit.71-1593380662。[13] A. Muhammad、Z. Qayyum、W. Muhammad、S. Tanveer、A. Martinez-Enriquez 和 A. Syed，《E-Hafiz：帮助穆斯林诵读与背诵《古兰经》的智能系统》，Life Sci J，第9卷，第534–541页，2012年10月，访问日期：2022年10月1日。[在线]。可访问：https://www.researchgate.net/publication/291155414_E- Hafiz_智能_系统_以_帮助_穆斯林_进行_诵读_与_背诵_《_古兰经[14] B. Putra、B. Tris Atmaja 和 D. Prananto，《为《古兰经》经文诵读学习软件开发语音识别系统》，IJID（International Journal on Informatics for Development），第1卷，第14页，2012年10月，doi：10.14421/ijid.2012.01203。[15] H. A. Hassan、N. H. Nasrudin、M. N. M. Khalid、A. Zabidi 和 A. I. Yassin，《使用多层感知器进行大震音发音识别的模式分类》，ISCAIE 2012 - 2012 IEEE计算机应用与工业电子研讨会，第209–212页，2012年，doi：10.1109/ISCAIE.2012.6482098。[16] M. Ab. Rahman、I. A. A. Kassim、T. Ab. Rahman 和 S. Z. M. Muji，《为儿童学习《古兰经》开发自动泰吉威德检查系统》，Evolution in Electrical and Electronic Engineering，第2卷，第1期，第165–176页，2021年5月，doi：10.30880/eeee.2021.02.01.019。[17] F. Ahmad、S. Z. Yahya、Z. Saad 和 A. R. Ahmad，《利用人工神经网络进行泰吉威德分类》，2018年智能应用、通信与网络国际会议，SmartNets 2018，2018年7月，doi：10.1109/SMARTNETS.2018.8707394。

[18] M. Al-Ayyoub、N. A. Damer 和 I. Hmeidi，《利用深度学习自动判定基本《古兰经》诵读规则的正确应用》，《阿拉伯国际信息技术期刊》，第 15 卷，2018 年 4 月，访问日期：2022 年 10 月 1 日。[在线]。网址：https://www.researchgate.net/publication/326010238_Using_Deep_Learning_for_Automatically_Det ermining_Correct_Application_of_Basic_Quranic_Recitation_Rules[19] H. Tabbal、W. el Falou 和 B. Monla，《利用语音识别技术对音频文件中的《古兰经》经文划分系统进行分析与实现》，《2006年国际信息与通信技术会议：从理论到应用》论文集（ICTTA 2006），第 2 卷，第 2979–2984 页，2006 年，DOI：10.1109/ICTTA.2006.1684889。[20] A. M. Alagrami 和 M. M. Eljazzar，《SMARTAJWEED：阿拉伯语《古兰经》诵读规则的自动识别》，《计算机科学与信息技术（CS & IT）》，第 145–152 页，2020 年 12 月，DOI：10.5121/CSIT.2020.101812。[21] M. ibn I. Bukhārī 和 M. Muhsin. Khan，《布哈里圣训实录：圣训实录含义的阿拉伯语-英语译本》，Kazi Publications，第 6 卷，1997 年。[22] M. R. Schroeder，《语音信号》，第 105–108 页，1999 年，DOI：10.1007/978-3-662-03861-1_7。[23] S. Hochreiter 和 J. Schmidhuber，《长短期记忆》，《神经计算》，第 9 卷，第 8 期，第 1735–1780 页，· 1997 年 11 月，DOI：10.1162/NECO.1997.9.8.1735。[24] H. M. Osman、B. S. Mustafa 和 Y. Faisal，《QDAT：一个用于《古兰经》诵读的数据集》，《伊斯兰计算机科学与技术应用国际期刊》，第 1–9 页，2021 年，访问日期：2022 年 10 月 3 日。[在线]。网址：https://www.researchgate.net/publication/350785609_QDAT_A_data_set_for_Reciting_the_Quran[25] J. W. Picone，《语音识别中的信号建模技术》，《IEEE会刊》，第 81 卷，第 9 期，第 1215–1247 页，1993 年，DOI：10.1109/5.237532。[26] S. Guha、A. Das、P. Singh、A. Ah madian、N. Senu 和 R. Sarkar，《基于和声搜索与裸鼹鼠算法的混合特征选择方法：用于音频信号中的口语语言识别》，IEEE Access，第 8 卷，第 182868–182887 页，2020 年 1 月，DOI：10.1109/ACCESS.2020.3028121。[27] T. Ganchev、N. Fakotakis 和 K. George，《说话人验证任务中各种梅尔频率倒谱系数（MFCC）实现方法的比较评估》，SPECOM 论文集，第 1 卷，2005 年 10 月。[28] I. Kamarulafizam、S. H. Salleh、J. M. Najeb、A. K. Ariff 和 A. Chowdhury，《利用梅尔频率倒谱系数（MFCC）和时频分布进行心音分析》，IFMBE论文集，第 15 卷，第 402–405 页，2007 年，DOI：10.1007/978-3- 540-68017-8_102/COVER。[29] V. Tiwari，《梅尔频率倒谱系数（MFCC）及其在说话人识别中的应用》，《新兴技术国际期刊》，第 1 卷，第 1 期，第 19–22 页，2010 年，访问日期：2022 年 10 月 3 日。[在线]。网址：https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.689.4627&rep=rep1&type=pdf[30] J. Benesty、M. M. Sondhi 和 Y. A. Huang，《Springer语音处理手册》，Springer，2008 年，DOI：10.1007/978-3-540-49127-9。[31] J. O. Smith、S. University. C. for Computer Research in Music、Acoustics 和 S. University. D. of Music，《音频信号频谱处理》。W3K，2011 年。[在线]。网址：https://books.google.nl/books?id=NXRMcgAACAAJ[32] D. N. Rockmore，《快速傅里叶变换：全家人都能使用的算法》，《计算机科学与工程》，第 2 卷，第 1 期，第 60–64 页，2000 年，DOI：10.1109/5992.814659。

[33] S. S. Stevens、J. Volkmann 和 E. B. Newman，《用于测量心理声学音高的量表》，*J Acoust Soc Am*，第 8 卷，第 3 期，第 185 页，2005 年 6 月，doi：10.1121/1.1915893。[34] J. R. Deller、J. H. L. Hansen 和 J. G. Proakis，《语音信号的离散时间处理》，*IEEE Press*，第 908 页，2000 年。[35] B. Putra、B. Atmaja 和 D. Prananto，《为《古兰经》经文诵读学习软件开发语音识别系统》，*IJID*（*International Journal on Informatics for Development*），第 1 卷，第 2 期，第 14 页，2012 年 12 月，doi：10.14421/ijid.2012.01203。[36] D. Jurafsky 和 J. H. Martin，《语音与语言处理：自然语言处理、计算语言学和语音识别导论》，第 934 页，2000 年，访问日期：2022 年 10 月 3 日。[在线]。可访问：http://14.139.156.51:8080/jspui/bitstream/123456789/8730/1/8.Language%20and%20speech%20processing-%20Mariani.pdf[37] D. E. Rumelhart、G. E. Hinton 和 R. J. Williams，《通过反向传播误差学习表征》，*Nature 1986 323:6088*，第 323 卷，第 6088 期，第 533–536 页，1986 年，doi：10.1038/323533a0。[38] S. Hochreiter，《循环神经网络学习过程中的梯度消失问题及其解决方案》，*International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*，第 6 卷，第 107–116 页，1998 年 10 月，doi：10.1142/S0218488598000094。[39] Y. Bengio、P. Simard 和 P. Frasconi，《使用梯度下降学习长期依赖关系十分困难》，*IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*，第 5 卷，第 157–166 页，1994 年 10 月，doi：10.1109/72.279181。[40] S. Fernández、A. Graves 和 J. Schmidhuber，《循环神经网络在判别式关键词检测中的应用》，*Lecture Notes in Computer Science (including subseries Lecture Notesin Artificial Intelligence and Lecture Notes in Bioinformatics)*，第 4669 卷（LNCS），第 2 部分，第 220–229 页，2007 年，doi：10.1007/978-3-540-74695-9_23/COVER。[41] A. Graves 和 J. Schmidhuber，《使用双向长短期记忆网络及其他神经网络架构进行逐帧音素分类》，*Neural Networks*，第 18 卷，第 5–6 期，第 602–610 页，2005 年 7 月，doi：10.1016/j.neunet.2005.06.042。[42] A. Graves 和 J. Schmidhuber，《使用多维循环神经网络进行离线手写识别》，*Adv Neural Inf Process Syst*，第 21 卷，2008 年。[43] P. Malhotra、L. Vig、G. Shroff 和 P. Agarwal，《用于时间序列异常检测的长短期记忆网络》，*undefined*，2015 年，访问日期：2022 年 10 月 4 日。[在线]。可访问：http://www.i6doc.com/en/。[44] P. Radhakrishnan，《循环神经网络简介》，2017 年 8 月 20 日。https://towardsdatascience.com/introduction-to-recurrent-neural-network-27202c3945f3（访问日期：2022 年 10 月 4 日）。[45] S. Varsamopoulos、K. Bertels 和 C. Almudever，《设计基于神经网络的表面码解码器》。2018 年 10 月。访问日期：2022 年 10 月 4 日。[在线]。可访问：https://www.researchgate.net/publication/329362532_Designing_neural_network_based_decoders_for surface codes_ _[46] A. Baruwa、M. Abisiga、I. Gbadegesin 和 A. Fakunle，《借助神经架构搜索实现端到端语音识别》，*ArXiv*，2019 年 12 月，doi：10.48550/arxiv.1912.05946。

[47] C. Sun、A. Shrivastava、S. Singh 和 A. Gupta，《重新审视深度学习时代数据的惊人效用》。第 843–852 页，2017 年。访问日期：2022 年 10 月 7 日。[在线]。可访问：https://openaccess.thecvf.com/content_iccv_2017/html/Sun_Revisiting_Unreasonable_Effectiveness_ICCV_2017_paper.html[48] A. Mathew、P. Amudha 和 S. Sivakumari，《深度学习技术：概述》，智能系统与计算进展，第 1141 卷，第 599–608 页，2021 年，doi: 10.1007/978-981-15-3383-9_54/COVER.
[49] 《在线学习泰吉威德规则 | Firdaws Academy》。https://firdawsacademy.com/learning-tajweed- rules-online/（访问日期：2022 年 10 月 25 日）。