

## 基于深度学习的时间序列分类研究综述

任利强 贾舒宜\* 王海鹏 王子玲

(海军航空大学 烟台 264001)

**摘要:** 时间序列分类(TSC)是数据挖掘领域中最重要且最具有挑战性的任务之一。深度学习技术在自然语言处理和计算机视觉领域已取得革命性进展,同时在时间序列分析等其他领域也显示出巨大的潜力。该文对基于深度学习的时间序列分类的最新研究成果进行了详细综述。首先,定义了关键术语和相关概念。其次,从多层感知机、卷积神经网络、循环神经网络和注意力机制四个网络架构角度分类总结了当前最新的时间序列分类模型,及各自优点和局限性。然后,概述了时间序列分类在人体活动识别和脑电图情绪识别两个关键领域的最新进展和挑战。最后,讨论了将深度学习应用于时间序列数据时未解决的问题和未来研究方向。该文为研究者了解最新基于深度学习的时间序列分类研究动态、新技术和发展趋势提供了参考。

**关键词:** 深度学习; 时间序列; 神经网络; 分类; 综述

中图分类号: TN911.7; TN99; TP183

文献标识码: A

文章编号: 1009-5896(2022)YU-0001-23

DOI: 10.11999/JEIT231222

## A Review of Research on Time Series Classification Based on Deep Learning

REN Liqiang JIA Shuyi WANG Haipeng WANG Ziling

(Naval Aviation University, Yantai 264001, China)

**Abstract:** Time Series Classification(TSC) is one of the most important and challenging tasks in the field of data mining. Deep learning techniques have achieved revolutionary progress in natural language processing and computer vision, and have also demonstrated great potential in areas such as time series analysis. A detailed review of the latest research advances in deep learning-based TSC is provided in this paper. Firstly, key terms and related concepts are defined. Secondly, the latest time series classification models are classified from four perspectives of network architectures: multilayer perceptron, convolutional neural networks, recurrent neural networks, and attention mechanisms, along with their respective advantages and limitations. Additionally, the latest developments and challenges in time series classification in the fields of human activity recognition and electroencephalogram-based emotion recognition are outlined. Finally, the unresolved issues and future research directions when applying deep learning to time series data are discussed. This paper provides researchers with a reference for understanding the latest research dynamics, new technologies, and development trends in the deep learning-based time series classification field.

**Key words:** Deep learning; Time series; Neural networks; Classification; Review

### 1 引言

时间序列分析被认为是21世纪数据挖掘领域的十大挑战性研究问题之一<sup>[1]</sup>。时间序列分类(Time Series Classification, TSC)是时间序列分析中的关键任务之一<sup>[2]</sup>。TSC旨在构建一个机器学习模型,

用于预测连续有序的实值观测序列的类别标签。TSC具有广泛的应用,如人体活动识别(Human Activity Recognition, HAR)<sup>[3,4]</sup>、基于电子健康记录的诊断<sup>[5]</sup>以及系统监测<sup>[6]</sup>等。此外,作为时间序列基准数据集的加利福尼亚大学河滨分校时间序列数据集<sup>[7]</sup>(University of California Riverside, UCR)和东英吉利大学时间序列数据集<sup>[8]</sup>(University of East Anglia, UEA)包含类型的多样性,也充分说明了TSC的广泛应用领域。

近年,深度学习在计算机视觉和自然语言处理领域取得了巨大的成功。相对于传统的机器学习方法,深度学习能够自动地从原始数据的低级属性中

收稿日期: 2023-11-03; 改回日期: 2023-12-21; 网络出版: 2023-12-27

\*通信作者: 贾舒宜 ganlingxian@sina.com

基金项目: 国家自然科学基金(62076249), 山东省自然科学基金(ZR2020MF154)

Foundation Items: The National natural Science Foundation of China (62076249), The natural Science Foundation of Shandong Province (ZR2020MF154)

学习并抽取更高层次的有意义性特征。因此,深度学习在TSC领域也逐渐受到关注,更有学者认为深度学习对于视觉和语音等领域的影响会在TSC研究领域中得到复制。文献[9]讨论了亚历克斯网络(Alex Krizhevsky Network, AlexNet)对计算机视觉的影响,并指出鉴于数据的相似性,深度学习在TSC领域有巨大的改进潜力。文献[10]是最早发表的基于深度学习的TSC综述之一,具有很大的影响力。作者将TSC分为生成模型和判别模型两大类,通过实验验证了这两大类中当前最先进性能的模型,结果显示判别模型中残差网络(Residual Network, ResNet)<sup>[11]</sup>是最准确的TSC深度学习模型。随后,同一研究团队提出了时间序列分类集成深度卷积神经网络(InceptionTime)<sup>[9]</sup>,该方法的性能优于之前综述中表现最佳的ResNet。自从InceptionTime之后,已有大量深度学习论文提出了TSC算法,其中大多数发表在最近3年。特别是,近年来备受关注且在时间序列数据中显示出出色的短程和长程依赖关系建模能力的注意力模型,这些模型非常适用于时间序列建模<sup>[12,13]</sup>。许多注意力的变体模型已被提出,以解决时间序列建模中的特定挑战,并成功应用于时间序列分类<sup>[14,15]</sup>。鉴于注意力机制和时间序列分类的新网络结构的出现,本综述旨在提供关于最新方法的介绍,以对时间序列分类领域的深度学习最新研究进行系统和全面的总结。特别是,本文将对尚未在文献[10]中进行系统综述的基于深度学习的TSC方法进行介绍,这将有益于时间序列分析社区的发展。

本文首先定义并简要介绍了时间序列分类任务,随后概述了经典的非深度学习时间序列分类方法。接下来,从网络架构和应用领域两个角度提出一个用于TSC的新分类法,并总结了基于多层感知器、基于卷积神经网络、基于循环神经网络和基于注意力的TSC模型架构,以及用于提高性能的改进方法。然后,总结了TSC在人体活动识别领域的研究现状。最后,探讨了深度学习时间序列分类任务中尚未解决的问题并展望了未来的研究方向。

## 2 背景和定义

本节首先定义了必要的关键术语和概念,然后介绍了在时间序列分类研究领域广泛应用的两个基准数据集,进一步对经典非深度学习时间序列分类算法进行概述,最后介绍了用于TSC任务的深度神经网络模型的新分类方法。

### (1) 相关定义。

将集合中的时间序列数量定义为 $m$ ,观测通道数或维度定义为 $d$ ,序列长度定义为 $n$ 。

**定义1(时间序列(Time Series, TS))** 时间序列 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ 是 $n$ 个数据点的有序序列, $\mathbf{a}_i$ 表示 $\mathbf{A}$ 的第 $i$ 个值。

在上面定义中,如果时间序列中的每个点 $\mathbf{a}_i \in \mathbf{A}$ 都表示单个值( $\mathbf{a}_i \in \mathbb{R}$ ),则该序列是单变量时间序列(Univariate Time Series, UTS)。如果每个点 $\mathbf{a}_i \in \mathbf{A}$ 表示在同一时间点观测到的多个变量(例如,温度、湿度、气压等),那么每个点本身就是长度为 $d$ 的向量 $\mathbf{a}_i \in \mathbb{R}^d$ ,则将该序列称为多变量时间序列(Multivariate Time Series, MTS)。

**定义2(多变量时间序列)** 多变量时间序列 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \in \mathbb{R}^{(d \times n)}$ 是 $n$ 个向量的列表,其中每个 $\mathbf{a}_i$ 都是具有 $d$ 个通道(有时称为维度)的向量。标量 $a_{k,i} \in \mathbb{R}$ 表示第 $k$ 个通道的第 $i$ 个观测值。

通常,也可以将MTS视为一组 $d$ 个时间序列,且假设 $\mathbf{a}_i$ 中的所有观测值都在同一时间点或空间点观察到。

**定义3(数据集)** 数据集 $D = (\mathbf{X}, \mathbf{Y}) = (\mathbf{A}^{(i)}, \mathbf{y}^{(i)})_{i \in [1, 2, \dots, m]}$ 是包含 $m$ 个时间序列和一组预定义的离散类标签 $C$ 的集合。其中, $\mathbf{A}^{(i)}$ 可以是单变量或多变量时间序列,其标签为 $\mathbf{y}^{(i)} \in C$ 。

**定义4(时间序列分类)** TSC是一种监督学习任务,其目的是在数据集 $D = (\mathbf{X}, \mathbf{Y}) = (\mathbf{A}^{(i)}, \mathbf{y}^{(i)})_{i \in [1, 2, \dots, m]}$ 中学习目标变量与一组时间序列之间的关系,其中 $\mathbf{y}^{(i)}$ 表示每个 $\mathbf{A}^{(i)}$ 的目标变量。TSC预测时间序列的分类值,将时间序列数据归类为有限的类别。通常,对于TSC来说,目标变量 $\mathbf{y}^{(i)}$ 是一个独热编码的向量。

在深度学习的背景下,TSC模型是一个神经网络,其执行以下函数将输入的时间序列映射到目标变量

$$f_L(\theta_L, \mathbf{X}) = f_{L-1}(\theta_{L-1}, f_{L-2}(\theta_{L-2}, \dots, f_1(\theta_1, \mathbf{X}))) \quad (1)$$

其中, $f_l$ 表示第 $l$ 层的非线性激活函数,其中 $l \in [1, 2, \dots, L]$ 。TSC的目标是训练神经网络模型,将时间序列数据集 $D$ 映射到具有 $C$ 个类别标签的集合 $Y$ 。在训练完成之后,神经网络输出一个包含 $C$ 个值的向量,估计了时间序列 $X$ 属于每个类别的概率。通常在神经网络的最后一层使用Softmax激活函数来实现这一目标。Softmax函数为所有相关类别估计概率,使其在所有类别上总和为1。交叉熵损失通常用于训练具有Softmax输出或分类类型的神经网络。

(2) 时间序列数据集。UCR<sup>[7]</sup>和UEA<sup>[8]</sup>时间序列存档分别是单变量和多变量时间序列分类基准数据集。UCR数据集首次在2002年提出,自此一直

是时间序列挖掘领域重要的开源数据集资源。该数据集在2015年经过一次重大更新之前，一共包含了46个类别的数据集。在2015年的更新中，数据集的数量其达到了85个，并且数据进行了标准化和去噪处理。最新的更新于2018年完成，数据集数量扩展至128个。此次更新中，序列数据被标注，每个数据集样本都带有样本类别标签。UEA数据集于2018年发布，包含来自不同应用领域的30个多变量数据集(其中等长度时间序列数据集26个)，例如心电图、运动分类、光谱分类等。这些数据集在维度数量(2~1 345)、时间序列数量(12~30 000)、时间序列类别数量(2~39)和时间序列长度(8~17 984)等方面各不相同。表1总结了UCR和UEA数据集详细信息，其中包括时间序列维度、数据集数量、类别数量、训练集大小、序列长度以及来源领域。UCR<sup>[7]</sup>和UEA<sup>[8]</sup>时间序列存档分别是单变量和多变量时间序列分类基准数据集。UCR数据集首次在2002年提出，自此一直是时间序列挖掘领域重要的开源数据集资源。该数据集在2015年经过一次重大更新之前，一共包含了46个类别的数据集。在2015年的更新中，数据集的数量其达到了85个，并且数据进行了标准化和去噪处理。最新的更新于2018年完成，数据集数量扩展至128个。此次更新中，序列数据被标注，每个数据集样本都带有样本类别标签。UEA数据集于2018年发布，包含来自不同应用领域的30个多变量数据集(其中等长度时间序列数据集26个)，例如心电图、运动分类、光谱分类等。这些数据集在维度数量(2~1 345)、时间序列数量(12~30 000)、时间序列类别数量(2~39)和时间序列长度(8~17 984)等方面各不相同。表1总结了UCR和UEA数据集详细信息，其中包括时间序列维度、数据集数量、类别数量、训练集大小、序列长度以及来源领域。

### (3) 基于深度学习的TSC分类。

近年来，深度学习领域取得了显著进展，这促使研究人员开始探索将深度学习用于解决TSC中的复杂问题。基于深度学习的TSC方法可以分为两种主要类型：生成式和判别式。生成式方法通常被视为基于模型的方法<sup>[2]</sup>，其目标是在训练分类器之前找到合适的时间序列表示。Serra等人<sup>[16]</sup>提出了一种通用的神经网络编码器，用于将可变长度的时间

序列转换为固定长度的表示。此外，Banerjee等人<sup>[17]</sup>使用了深度置信网络结合迁移学习方法以无监督的方式对时间序列的潜在特征进行建模。Aswolinskiy等人<sup>[18]</sup>还使用回声状态网络来重建原始时间序列，以在训练分类器之前学习适当的表示。由于需要额外的非监督训练步骤，实现生成式方法通常更复杂。此外，生成方法通常比判别方法效率低，这是由于判别方法是直接将原始时间序列映射到类别概率分布。

本综述主要关注端到端判别式方法，因为其直接将原始时间序列输入深度学习模型，而无需进行任何繁琐的预处理。为了对现有基于深度学习的时序分类模型进行系统的总结，本文提出一种基于网络架构和应用领域的分类方法，如图1所示，并在本文的第3和第4节中进行了讨论。

## 3 深度学习模型

### (1) 多层感知机。

全连接网络(Fully Connected network, FC)是最简单的神经网络架构是，也称为多层感知器(Multilayer Perceptron, MLP)。如图2所示， $l-1$ 层的所有神经元都与下一 $l$ 层的所有神经元连接，其中 $l \in [1, L]$ 。权重用来建模神经网络中的连接关系。将非线性应用于输入 $\mathbf{A}^{l-1}$ 的一般公式为

$$\mathbf{A}^l = f(\mathbf{W}^l \times \mathbf{A}^{l-1} + b) \quad (2)$$

其中， $\mathbf{A}^l$ 表示第 $l$ 层神经元的激活， $\mathbf{A}^1$ 等于输入序列 $X$ 。此外， $\mathbf{W}$ 和 $b$ 分别表示神经元的权重和偏置， $f$ 表示非线性激活函数。在MLP模型中，层数和神经元数量为超参数。

MLP在处理时间序列数据时存在的一个主要局限是它们不适合捕捉这种类型数据中的时间依赖关系。MLP按照固定和预定的顺序处理输入数据，而不考虑输入值之间的时间关系。如图2所示，每个时间步骤都被单独加权，并且时间序列元素被独立处理。为了解决这个问题，一些研究将MLP和其他特征提取器相结合，如动态时间规整(Dynamic Time Warping, DTW)<sup>[19]</sup>。动态时间规整神经网络(Dynamic Time Warping Neural Network, DTW-NN)利用DTW的弹性匹配技术来动态对齐网络层的输入与权重。通过这种方式，DTW-NN能够解决时间序列识别前馈神经网络架构中的时间扭曲和

表1 UCR和UEA时间序列数据集详细信息

数据集	维度	数量	类别数量	训练集大小	序列长度	类型
UCR	1	128	2~60	16~8926	24~2 709	图像轮廓、传感器读数、动作分类、心电图、电子设备和模拟数据等
UEA	2-1345	30	2~39	12~30 000	8~17 984	心电图、运动分类、光谱分类等

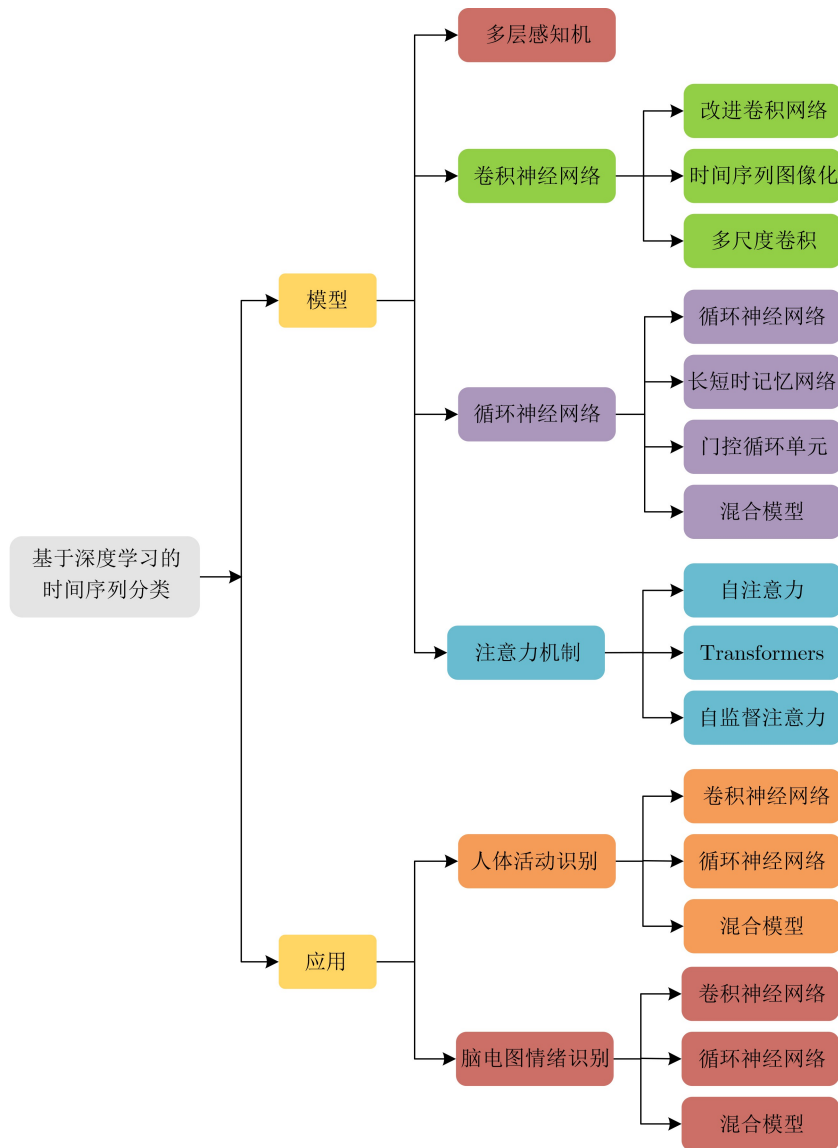


图1 基于网络结构和应用领域的深度学习时序分类体系结构

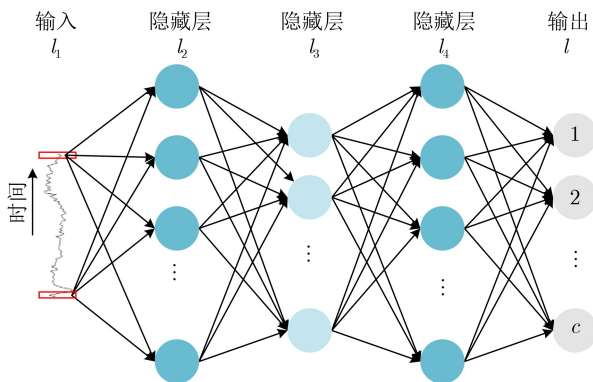


图2 用于单变量时序分类的多层感知

可变模式长度问题<sup>[19]</sup>。类似地, Tabassum等人<sup>[20]</sup>提出了一种名为简单快速词分割集成(Simple And Fast segmented word Embedding-based, SAFE)的分类器, 用于将原始时间序列转化为符号表示,

并基于该符号表示生成单词序列。然后, 根据生成的单词序列, 训练了两层MLP分类器。

尽管上述模型尝试解决MLP模型无法捕捉时间依赖关系的问题, 但它们在捕捉时间不变特征方面仍存在局限性<sup>[11]</sup>。此外, MLP模型无法以多尺度方式处理输入数据。时间序列数据通常在不同尺度上呈现出不同的模式和结构, 如长期趋势和短期波动。MLP模型无法捕捉这些模式, 因为其只能处理具有固定长度表示的输入数据。许多其他深度学习模型更适合处理时间序列数据, 例如循环神经网络(Recurrent Neural Networks, RNNs)和卷积神经网络(Convolutional Neural Network, CNNs), 这些模型专门设计用于捕捉时间序列数据中的时间依赖性和局部模式。

(2) 基于卷积神经网络模型。

CNN最初由Fukushima<sup>[21]</sup>于1982年提出, 其灵

感来自于动物视觉皮层的结构和功能，并已广泛应用于视觉模式识别。然而，在图形处理单元(Graphics Processing Unit, GPU)出现之前，由于对计算资源的要求，CNN无法处理大规模图像。随着GPU技术的发展，Krizhevsky等人<sup>[22]</sup>实现了一种高效的基于GPU的深度卷积神经网络架构AlexNet，并在2012年赢得了大规模视觉识别挑战赛(ImageNet large scale visual recognition challenge, ImageNet)竞赛，使得卷积神经网络重新受到关注。

研究者提出了许多CNN架构的变体，但它们的主要组成部分非常相似。以LeCun卷积神经网络(LeCun convolutional neural Network, LeNet-5)<sup>[23]</sup>为例，它由3种类型的层组成：卷积层、池化层和全连接层。卷积层用于学习输入数据的特征表示。图3显示了t-LeNet<sup>[24]</sup>网络的架构，它是LeNet的时间序列特定版本。该图显示了卷积层由多个卷积核组成，用于计算不同的特征图。具体来说，每个特征图中的每个神经元都与前一层中称为感受野的一组相邻神经元相连。首先，将输入与学习的卷积核进行卷积，然后对卷积结果应用逐元素的非线性激活函数，即可以创建特征图。值得注意的是，输入的所有空间位置共享每个特征图的卷积核，并使用多个卷积核来获取整个特征图。在位置 $(i, j)$ 处， $l$ 层的第 $k$ 个特征图的特征值由式(3)计算

$$Z_{i,j,k}^l = (\mathbf{W}_k^l)^T \mathbf{A}_{i,j}^{l-1} + b_k^l \quad (3)$$

其中， $\mathbf{W}_k^l$ 和 $b_k^l$ 分别是第 $l$ 层第 $k$ 个卷积核的权重向量和偏置项， $\mathbf{A}_{i,j}^{l-1}$ 是位于 $(i, j)$ 位置的第 $l$ 层的输入块。生成特征图 $Z_{i,j,k}^l$ 的卷积核 $\mathbf{W}_k^l$ 是共享的。权重共享机制具有减少模型复杂性和使网络更容易训练的几个优点。设 $f(\cdot)$ 表示非线性激活函数。卷积特征 $Z_{i,j,k}^l$ 的激活值计算为

$$\mathbf{A}_{i,j,k}^l = f(Z_{i,j,k}^l) \quad (4)$$

最常用的激活函数是Sigmoid函数、双曲正切函数tanh和ReLU函数。如图3所示，池化层通常放

置在两个卷积层之间，以降低特征图的分辨率并实现平移不变性。在经过多个卷积阶段后，通常会存在一个或多个全连接层，用于进行高级推理。在CNN的最后一层，通常是输出层，常使用Softmax分类器进行分类任务<sup>[25]</sup>。

(a) 改进卷积神经网络。

自2012年AlexNet<sup>[22]</sup>在计算机视觉领域图像分类任务中取得重大突破以来，CNN已经历了多次改进，例如使用更深的网络，应用更小更高效的卷积核，添加池化层以降低特征图的维数，以及使用批归一化提高训练的稳定性<sup>[25]</sup>。由于CNN在这些不同领域的成功，研究人员也开始将其应用于时间序列分类<sup>[10,13,26]</sup>。本节介绍了第一类，我们将其称为改进的时间序列分类CNNs。该部分讨论的论文大都是没有特定预处理或数学特性的转换，例如将序列转换为图像或使用多尺度卷积，因此它们不属于其他任何一个类别。

多通道深度卷积神经网络(Multi-Channel Deep Convolutional Neural Network, MC-DCNN)<sup>[27]</sup>是首个将CNN应用于时间序列分类的模型。该模型针对多变量数据特点对传统的深度CNN进行了改进：每个通道(输入维度)都独立应用卷积。输入序列的每个维度都经过两个具有ReLU激活函数的卷积层，并在每个卷积层后紧跟一个长度为2的最大池化层。所有维度的输出被串联在一起传递给一个全连接层，最后通过Softmax分类器进行分类。类似于MC-DCNN，人体活动识别MC-CNN<sup>[28]</sup>提出了一个3层CNN模型。与MC-DCNN不同，该模型将1D卷积同时应用于所有输入通道，以捕捉早期阶段的时间和空间关系。Zhao等人<sup>[26]</sup>在UCR时间序列数据挖掘数据集的最早版本上使用了两阶段版本的MC-CNN架构。作者还进行了消融研究，以评估在不同数量的卷积核和池化类型下CNN模型的性能。

全卷积网络(Fully Convolutional Networks, FCN)<sup>[29]</sup>和ResNet<sup>[30]</sup>是两种常用于图像和视频识别任务的深度神经网络，它们已经被改进用于端到端

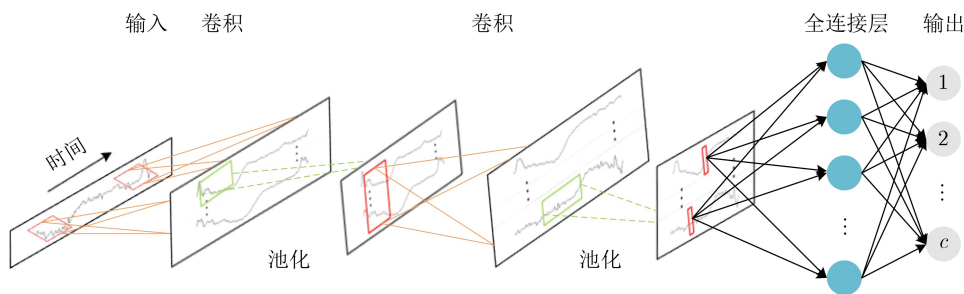


图3 t-LeNet 时间序列特定版本网络架构

的时间序列分类<sup>[11]</sup>。FCN是CNN的一种变体,设计用于处理任意大小的输入,而不像传统CNN那样受固定大小输入的限于。这是通过使用全局平均池化(Global Average Pooling, GAP)<sup>[29]</sup>替换传统CNN中的全连接层以实现。Wang等人<sup>[11]</sup>将FCN应用于单变量时间序列分类,包含3个卷积块,每个块包含1个卷积层,后跟批归一化和ReLU激活。每个块分别使用128, 256, 128个卷积核,卷积核长度分别为8, 5, 3。最后一个卷积块的输出通过GAP层进行平均,并传递给最终的Softmax分类器。GAP层减小了输入空间维度,且保留了通道信息的特性,这使其可以与类激活图(Class Activation Map, CAM)<sup>[31]</sup>结合使用,以突出输入中对预测类别最重要的区域。

类似于FCN, Wang等人<sup>[11]</sup>也将ResNet用于单变量时间序列分类。ResNet是最深的网络结构之一,包含3个残差块,后跟1个GAP层和1个Softmax分类器。在每个块之间, ResNet使用残差连接以减小深度学习模型中的梯度消失效应。每个残差块的结构类似于FCN结构,包含3个卷积层,后跟批归一化和ReLU激活函数。每个卷积层使用64个卷积核,卷积核长度分别为8,5,3。在85个单变量时间序列分类数据集上, ResNet被证明是最准确的深度学习时间序列分类网络结构之一<sup>[2]</sup>。此外,文献<sup>[32]</sup>还提出了将ResNet和FCN结合的方法,以充分利用两个网络的优势。

除了调整网络架构外,一些研究还致力于修改卷积核以更好地适应时间序列分类任务。扩张卷积神经网络(Dilated Convolutions Neural Networks, DCNNs)<sup>[33]</sup>是一种卷积神经网络的类型,它使用扩张卷积来增加网络的感受野,而不增加参数数量。扩张卷积在卷积核的元素之间创建间隔,并执行卷积操作,从而覆盖输入的更大区域。这使得网络能够捕捉数据中的长距离依赖关系,非常适合时间序列分类任务<sup>[34]</sup>。最近,分离型卷积神经网络(Disjoint-CNN)<sup>[35]</sup>显示将1维卷积核分解为不相交的时间和空间组件可以在几乎不增加计算成本的情况下提高准确性。该网络首先应用不相交的时间卷积,然后再进行空间卷积其中时间卷积扩展了输入通道的数量,而空间卷积则将扩展后的隐藏状态重新投影回原始大小,以捕捉时空交互关系。

#### (b) 时间序列图像化处理。

在时间序列分类中,一种常见的方法是将时间序列数据转化为固定长度的表示,如向量或矩阵,然后输入到深度学习模型中。然而,对于长度变化或具有复杂时间依赖性的时间序列数据,这可能是

具有挑战性的。解决这个问题的方法之一是将时间序列数据表示为类似图像的形式,其中每个时间步被视为图像中的一个单独通道。这使得模型能够学习数据内部的空间关系,而不仅仅是时序关系。这里的“空间”指的是时间序列在单个时间步中不同变量或特征之间的关系。

Wang等人<sup>[36]</sup>将单变量时间序列数据编码为不同类型的图像,然后通过CNN进行处理。这种基于图像的框架开创了时间序列深度学习新方法的新分支,其中将图像转换视为特征工程技术之一。作者提出了两种将时间序列转化为图像的方法,分别为格拉姆角场(Gramian Angular Field, GAF)和马尔可夫转移场(Markov Transition Field, MTF)。GAF将时间序列数据表示为极坐标,并使用各种运算将这些角度转换为对称矩阵,而MTF则使用数据点从一个时间步到另一个时间步的转换概率来编码矩阵元素。单变量时间序列首先被转换为这两种类型的图像,并组合成一个双通道图像。最后,使用改进的CNN模型对时间序列图像进行分类。在其他研究中,提出了各种转换方法,包括递归图(Recurrence Plots, RP)<sup>[37]</sup>、格拉姆差角场(Gramian Angular Difference Field, GADF)<sup>[38]</sup>和格拉姆加和场(Gramian Angular Summation Field, GASF)<sup>[39]</sup>,用于将时间序列转换为输入图像,以揭示原始时间序列中隐含的特征和模式。

Hatami等人<sup>[37]</sup>提出了一种基于RP的表示方法,将时间序列转化为2维图像,并使用包含两阶段卷积和两个全连接层的深度CNN模型对生成的RP图像进行分类。随着计算机视觉领域出现性能更好的网络架构,他们使用预训练的Inception v3<sup>[40]</sup>将GADF图像映射到一个2 048维的向量空间,并使用具有3个隐藏层的MLP,以及Softmax激活函数进行时间序列的分类<sup>[38]</sup>。在相同的框架下,Chen等人<sup>[41]</sup>采用了相对位置矩阵和VGGNet<sup>[42]</sup>来通过转换2维图像对时间序列数据进行分类。实验结果表明,通过使用两个时间戳之间的相对位置,将1元时间序列数据转换为2维图像,可以获得良好的性能。类似地, Yang等人<sup>[39]</sup>使用了3种图像编码方法: GASF, GADF和MTF,将多变量时间序列数据编码成2维图像。研究表明,简单结构的卷积神经网络在分类任务上的性能与VGGNet相当。

总的来说,将时间序列数据表示为2维图像可能会导致原始时间序列数据中时间关系和模式信息损失,从而使模型难以准确地对数据进行分类。Chen等人<sup>[41]</sup>还指出,使用特定的转换方法(如GASF, GADF和MTF)并没有明显改善预测结果。

## (c) 多尺度卷积。

本部分讨论的论文是对输入时间序列应用了多尺度卷积核，或者在不同输入序列尺度上应用常规卷积。多尺度卷积神经网络(Multi-scale Convolutional Neural Networks, MCNN)<sup>[43]</sup>和t-LeNet<sup>[24]</sup>被认为是最早对输入时间序列进行预处理以在多尺度序列而不是原始序列上应用卷积的模型。

MCNN和t-LeNet的设计受到计算机视觉模型的启发，由用于图像识别任务的模型改进而来。这些模型可能不适合于时间序列分类任务，或者可能表现不如专门为时间序列分类任务设计的模型。潜在原因是这些模型中使用了渐进式池化层，这些层通常用于计算机视觉模型，以减小输入数据的尺寸并使其更容易处理。然而，当应用于时间序列数据时，这些池化层可能不是有效的，并且可能限制模型的性能。

MCNN结构简单，包括2个卷积层和1个池化层，然后是一个全连接层和Softmax层。然而，该方法涉及大量的数据预处理。具体来说，在训练之前，需要利用滑动窗口提取时间序列子序列，然后，子序列将经历恒等映射、下采样和平滑3次转换，这导致单变量输入时间序列转变为多变量序列。最后，转换后的输出被输入CNN模型以训练分类器。t-LeNet使用两种数据增强技术：窗口切片(Window Slicing, WS)和窗口扭曲(Window Warping, WW)，以防止过拟合。WS用于确保提取出相同长度的子序列，而WW使用了一种扭曲技术以挤压或扩展时间序列。因此，给定长度为 $L$ 的输入时间序列，首先通过WW进行膨胀( $\times 2$ )，然后通过WW进行压缩( $\times 1/2$ )，从而产生长度为 $L$ ,  $2L$ ,  $1/2L$ 的3个时间序列，这些序列经过WS提取等长度的子序列进行训练。最后，由于MCNN和t-LeNet都为每类提取的子序列预测一个分类器，所以采用多数投票法来获得完整时间序列的类别预测。

Inception<sup>[44]</sup>最初被提出用于端到端图像分类，现在已经发展为Inception-v4<sup>[45]</sup>，其中Inception与残差连接相结合，以进一步提高性能。受Inception架构的启发，Liu等人<sup>[46]</sup>设计了一个多变量卷积神经网络(MultiVariate Convolutional Neural Network, MVCNN)，其使用3种尺度的卷积核，包括 $2 \times 2$ ,  $3 \times 3$ 和 $5 \times 5$ ，提取传感器之间的相互作用特征。Inception-ResNet<sup>[47]</sup>架构包括卷积层，然后是Inception模块和残差块。Inception模块用于学习数据的多个尺度和多个层面，使网络能够捕获更复杂的模式。残差块用于学习网络输入和输出之间的残差或差异，从而提高性能。

InceptionTime<sup>[9]</sup>在UCR基准测试中达到最先进的性能。它是一个由5个相同结构的深度学习分类器组成的集成模型，每个分类器由两个级联的Inception模块组成。通过在每个模型中随机化初始权重值，实现了模型之间的多样性。每个Inception模块首先通过具有长度和步幅为1的瓶颈层来降低多变量时间序列的维度，同时保持相同的长度。然后，对瓶颈层的输出应用不同长度的1维卷积以提取不同尺寸的模式。同时，还对原始时间序列应用最大池化，以增加模型对小扰动的鲁棒性。卷积和最大池化层的输出被堆叠在一起，形成一个新的多变量时间序列，然后传递到下一层。在每个Inception块之间使用残差连接以减少梯度消失效应。第2个Inception块的输出经过GAP层后传入Softmax分类器。

由于InceptionTime在时间序列分类中的优异表现，提出了各种扩展，例如脑电信号Inception模型(ElectroEncephalography inception, EEGinception)<sup>[48]</sup>和全卷积Inception模型(Inception Fully Convolutional Network, InceptionFCN)<sup>[49]</sup>。与InceptionTime类似，EEGinception使用多个Inception层和残差连接作为其骨干网络。此外，为了应对训练数据的限制，作者提出了基于噪声添加的脑电图信号数据增强方法，可以提高平均准确度。InceptionFCN则结合了Inception模块和全卷积网络这两种广泛使用的深度学习技术。与InceptionTime不同，张雅雯等人<sup>[50]</sup>提出了一种名为多尺度残差全卷积神经网络(Multi-scale Residual Full Convolutional Neural Network, MRes-FCN)的结构，该结构使用了3个全卷积块和残差块作为骨干网络，其特点在于全卷积块和残差块使用串行多尺度卷积核提取具有辨别性的多尺度特征。表2总结了基于CNN的时间序列分类模型。

## (3) 循环神经网络模型。

## (a) 循环神经网络。

RNNs是一类专门设计用于处理时间序列和其他序列数据的神经网络类型。该网络架构可以处理可变长度的输入并产生可变长度的输出。这一特点是通过在不同层之间建立共享参数的有向连接实现。用于时间序列分类的RNN模型根据输出可以被分类为序列到序列和序列到单一输出两种类型。图4展示了序列到序列架构的RNN模型，每个输入子序列都有一个对应的输出。另一方面，在序列到单一输出的架构中，只使用 $\mathbf{y}^T$ 做出决策，忽略其他输出。

在每个时间步 $t$ 上，RNNs通过以下方式更新隐藏向量 $\mathbf{h}$

表2 基于CNN的时间序列分类模型总结

模型	提出年份	基准架构	模型特点
自适应模型			
MC-DCNN <sup>[26]</sup>	2014	2-Stage Conv	每个通道独立卷积
MC-CNN <sup>[27]</sup>	2015	3-Stage Conv	所有通道1D卷积
Zhao et al. <sup>[28]</sup>	2015	2-Stage Conv	所有通道1D卷积
FCN <sup>[11]</sup>	2017	FCN	使用GAP替代FC层
ResNet <sup>[11]</sup>	2017	ResNet 9	使用3个残差块
Res-CNN <sup>[32]</sup>	2019	ResNet+FCN	使用1个残差块+FCN
DCNNs <sup>[34]</sup>	2019	4-Stage Conv	使用扩张卷积
Disjoint-CNN <sup>[35]</sup>	2021	4-Stage Conv	分离型时空卷积
时间序列转换为图像			
Wang&Oates <sup>[36]</sup>	2015	Tiled CNN	格拉姆角场和马尔可夫转移场图像编码
Hatami et al. <sup>[37]</sup>	2017	2-Stage Conv	递归图图像编码
Karimi et al. <sup>[38]</sup>	2018	Inception V3	格拉姆差角场图像编码
RPMCNN <sup>[41]</sup>	2019	VGGNet, 2-Stage Conv	相对位置矩阵图像编码
Yang et al. <sup>[39]</sup>	2019	VGGNet	格拉姆差角场、格拉姆加和场和马尔可夫转移场图像编码
多尺度卷积操作			
MCNN <sup>[43]</sup>	2016	2-Stage Conv	恒等映射、下采样和平滑预处理
t-LeNet <sup>[24]</sup>	2016	2-Stage Conv	挤压和扩展预处理
MVCNN <sup>[46]</sup>	2019	4-stage Conv	基于Inception V1卷积
Inception-ResNet <sup>[47]</sup>	2021	ResNet	基于Inception V1卷积
InceptionTime <sup>[9]</sup>	2019	Inception V4	
EEG-inception <sup>[48]</sup>	2021	InceptionTime	多分类器集成模型
Inception-FCN <sup>[49]</sup>	2021	InceptionTime + FCN	
MRes-FCN <sup>[50]</sup>	2022	FCN + ResNet	使用多个串行多尺度卷积核

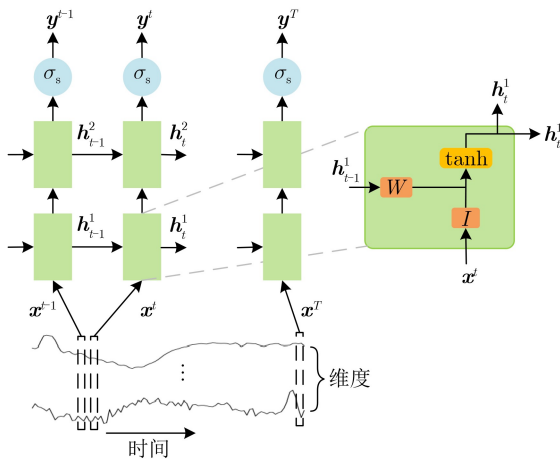


图4 两层循环神经网络的架构

$$h_t = \tanh(\mathbf{W}h_{t-1} + \mathbf{I}x^t) \quad (5)$$

其中,  $\mathbf{X} = \{x^1, \dots, x^{t-1}, x^t, \dots, x^T\}$  包含所有观测,  $\tanh$  表示双曲正切函数,  $\mathbf{W}$  和  $\mathbf{I}$  分别表示递归权重和投影矩阵, 隐藏状态  $h$  用于进行预测, 如式(6)所示

$$y^t = \sigma_s(\mathbf{W}h_{t-1}) \quad (6)$$

其中,  $\sigma_s$  是Softmax函数, 用于提供可能类别的归

一化概率分布。如图4所示, 隐藏状态  $h$  可用于堆叠RNNs, 以构建更深的网络

$$h_t^l = \sigma(\mathbf{W}h_{t-1}^l + \mathbf{I}h_t^{l-1}) \quad (7)$$

其中,  $\sigma$  是sigmoid函数。作为将每个时间步输入RNN的替代方法, 数据可以被划分为具有  $\omega$  个观测值的时间窗口, 每个时间窗口会被标记为在该窗口内的大多数观测对应的标签。

Dennis等人<sup>[51]</sup>提出了一种用于时间序列分类的双层RNN。在第1层中, 将输入序列分割为多个独立的RNNs, 以提高模型的并行性, 在第2层中利用第1层的输出来捕获长期依赖性。此外, Hermans等人<sup>[52]</sup>展示了更深层次的循环神经网络可以在复杂的时间任务上执行分层处理, 并比浅层版本更有效地捕获时间序列结构。RNNs通常使用一种称为通过时间反向传播(BackPropagation Through Time, BPTT)的迭代训练方法进行训练。当在时间上展开时, RNNs类似于具有共享参数的非常深的网络。

在RNN中, 通过在不同RNN单元之间共享权重, 梯度会在每个时间步上累加以训练模型。因此, 由于链式法则, 梯度经历连续的矩阵乘法, 要么呈指



数级的缩小导致梯度消失的问题，要么出现指数级的增大导致梯度爆炸的问题<sup>[53]</sup>。为了解决该问题，研究者们提出了可用于深层架构的网络，如长短时记忆网络(Long Short Term Memory, LSTM)<sup>[54]</sup>和门控循环单元(Gated Recurrent Unit, GRU)<sup>[55]</sup>。

(b) 长短时记忆网络。

LSTM<sup>[54]</sup>通过引入门控记忆单元来解决标准循环神经网络中常见的梯度消失/梯度爆炸问题。LSTM利用隐藏向量 $h$ 和记忆向量 $c$ 来控制每个时间步的状态更新和输出。具体而言，时间步 $t$ 上的计算如式(8)所示<sup>[56]</sup>

$$\left. \begin{aligned} \tilde{c}_t &= \tanh(\mathbf{W}^c \mathbf{h}_{t-1} + \mathbf{I}^c \mathbf{x}^t) \\ \mathbf{i}_t &= \sigma(\mathbf{W}^i \mathbf{h}_{t-1} + \mathbf{I}^i \mathbf{x}^t) \\ \mathbf{f}_t &= \sigma(\mathbf{W}^f \mathbf{h}_{t-1} + \mathbf{I}^f \mathbf{x}^t) \\ \mathbf{o}_t &= \sigma(\mathbf{W}^o \mathbf{h}_{t-1} + \mathbf{I}^o \mathbf{x}^t) \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{c}_t \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{aligned} \right\} \quad (8)$$

其中， $\mathbf{c}_t$ 是细胞状态门， $\mathbf{i}_t$ 、 $\mathbf{f}_t$ 和 $\mathbf{o}_t$ 分别是输入门、遗忘门和输出门的激活向量。 $\sigma$ 是sigmoid函数， $\odot$ 表示逐元素乘积。 $\mathbf{W}^c$ 、 $\mathbf{W}^i$ 、 $\mathbf{W}^f$ 和 $\mathbf{W}^o$ 分别代表循环权重矩阵，而 $\mathbf{I}^c$ 、 $\mathbf{I}^i$ 、 $\mathbf{I}^f$ 和 $\mathbf{I}^o$ 分别代表投影矩阵。

由于其设计特性，LSTM非常适用于涉及序列数据的问题，例如语言翻译<sup>[57]</sup>、视频表示学习<sup>[58]</sup>和图像字幕生成<sup>[59]</sup>。时间序列分类问题也不例外，通常采用类似语言翻译的模型序列到序列注意力网络(Sequence-to-Sequence with Attention, S2SwA)<sup>[60]</sup>将两个LSTM，一个是编码器，另一个是解码器，以序列到序列的方式结合起来，用于时间序列分类。在该模型中，编码器LSTM接受任意长度的输入时间序列，并从原始数据中提取信息，基于这些信息，解码器LSTM构建固定长度的序列，可视为自动提取的用于分类的特征。

(c) 门控循环单元。

GRU是另一种广泛使用的循环神经网络变体<sup>[55]</sup>。尽管GRU的提出时间晚于LSTM，但它具有更简单的架构。与LSTM不同，GRU只包括两个门，即重置门和更新门，这使得它在计算上更高效，同时需要更少的数据来实现泛化，其定义为

$$\left. \begin{aligned} \mathbf{z}_t &= \sigma(\mathbf{W}^z \mathbf{h}_{t-1} + \mathbf{I}^z \mathbf{x}^t) \\ \mathbf{r}_t &= \sigma(\mathbf{W}^r \mathbf{h}_{t-1} + \mathbf{I}^r \mathbf{x}^t) \\ \tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}[\mathbf{h}_{t-1} \odot \mathbf{r}_t, \mathbf{x}^t]) \\ \mathbf{h}_t &= (\mathbf{1} - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \end{aligned} \right\} \quad (9)$$

类似于S2SwA<sup>[60]</sup>，基于GRU的序列自编码器

被设计用来处理时间序列分类问题<sup>[61]</sup>。使用GRU作为编码器和解码器，处理不同长度的输入，从而产生固定大小的输出。通过在大规模无标签数据上对参数进行预训练，还提高了模型的准确性。

(d) 混合模型。

在时间序列分类中，通常将CNN和RNN结合起来，因为它们具有互补的优势，可以提高模型的性能。如前所述，CNN非常适合从数据中学习空间关系，例如时间序列中不同时间步的通道之间的模式和相关性。这使得它们能够从时间序列数据中学习有用的特征，有助于提高分类性能。另一方面，RNN非常适合从数据中学习时间依赖关系，这使其能够捕捉时间序列数据的动态特性并进行更准确的分类。将CNN和RNN的优势结合起来，可以从时间序列数据中同时学习空间和时间特征，提高时间序列分类模型的性能。此外，这两种模型可以一起训练，使它们能够相互学习并提高模型的整体性能。

已经提出了各种扩展模型，如多变量长短期记忆全卷积网络(Multivariate Long Short-Term Memory Fully Convolutional Network, MLSTM-FCN)<sup>[62]</sup>、多尺度卷积注意力机制长短期记忆模型(Multi-scale Convolution and Attention mechanism based Long Short-Term Memory, MCA-LSTM)<sup>[63]</sup>、时间序列注意力原型网络(Time series attentional prototype network, TapNet)<sup>[64]</sup>和多变量时间序列的半监督时空表示学习模型(Semi-supervised spatio-temporal representation learning on Multivariate Time Series, SMATE)<sup>[65]</sup>用于处理时间序列数据。MLSTM-FCN将单变量LSTM-FCN模型扩展到多变量情况。与LSTM-FCN类似，多变量版本由LSTM块和完全卷积块组成，用于从输入序列中提取特征。还在FCN块中添加了一个挤压和激励块，可以在前一层的输出特征图上执行一种自注意力操作<sup>[62]</sup>。MCA-LSTM<sup>[63]</sup>不仅使用注意力模块融合通道信息以获取特征的重要性并分配注意力权重，而且使用了多尺度卷积模块提取序列的多尺度特征。另外两种用于多变量时间序列分类的方法时间序列注意力原型网络(TapNet)和半监督时空网络(SMATE)<sup>[64,65]</sup>结合并试图利用传统基于距离的方法和深度学习方法的各自优势。

MLSTM-FCN, MCA-LSTM, TapNet和SMATE均采用双网络架构设计。输入分别输入到CNN和RNN模型中，它们的输出在全连接层之前进行整合，用于最终的任务。然而，在特征提取过程中，一个分支无法充分利用另一个分支的隐藏状态，因为

最终的分类结果是通过连接两个分支的输出生成的。这激发了不同类型的架构，试图在CNN和RNN模型之间进行逐层集成，例如GCRNN<sup>[66]</sup>和CNN-LSTM<sup>[67]</sup>。

尽管循环神经网络通常用于时间序列预测，但在时间序列分类方面应用RNN的研究较少，主要有以下4个原因：

①RNN在长时间序列上训练时通常会遇到梯度消失和梯度爆炸的问题。这是因为RNN在长时间序列上的训练中，梯度可能会变得非常小或非常大，导致模型难以训练。

②由于训练和并行化困难，RNN的计算成本较高，因此研究人员很少使用它们。

③循环架构主要设计用于从先前的数据中学习，以便做出关于未来的预测。这使其不太适合直接用于时间序列分类问题。

④RNN模型可能无法有效地捕捉和利用长序列中的长程依赖关系。

(4) 基于注意力机制模型。

如前所述，CNN在包括计算机视觉、自然语言处理和语音识别在内的各种应用中已经成为最成功的深度学习框架之一。虽然CNN模型可以捕获局部时空相关性，但无法有效地捕获和利用长距离依赖关系。此外，CNN只考虑数据点的局部顺序，而不考虑整体顺序。因此，许多最近的研究将RNN如LSTMs与CNN结合使用以捕捉这些信息<sup>[62,64,68]</sup>。然而，RNN模型在计算上的代价较高，并且在捕捉长距离依赖性方面的能力有限<sup>[13,69]</sup>。

相对而言，注意力模型可以捕捉长距离依赖关系，其更广泛的感受野提供了更多上下文信息，从而提高了模型的学习能力。注意力机制旨在通过关注重要特征并抑制不必要的特征来提高网络的表示能力。注意力模型在自然语言处理领域已经获得成功<sup>[69,70]</sup>，同时也有许多研究尝试将其应用于计算机视觉<sup>[71-73]</sup>和时间序列分析<sup>[11,13,14,74]</sup>等领域。

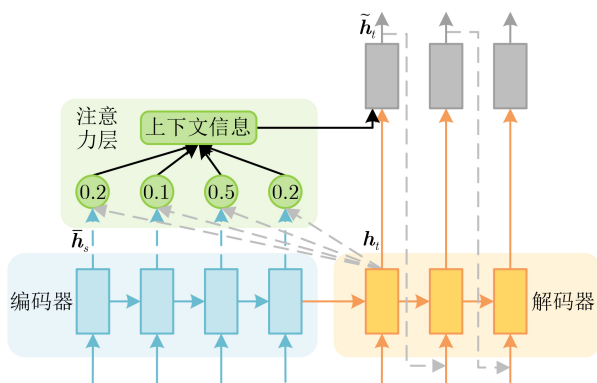


图5 自注意力机制

(a) 注意力机制。

注意力机制最初是由Bahdanau等人<sup>[75]</sup>提出，用于改进神经网络机器翻译中的编码器-解码器模型<sup>[76]</sup>性能。在神经网络机器翻译中，编码器将源语句编码成潜在空间中的向量，而解码器将潜在向量解码成目标语句。如图5所示，注意力机制允许解码器通过上下文向量 $c_t$ 关注源中的每个单词。对于该模型，注意力向量 $\alpha_t$ 是根据当前目标隐藏状态 $h_t$ 与每个源隐藏状态 $\bar{h}_s$ 进行比较所得，其计算公式为<sup>[77]</sup>

$$\alpha_t(s) = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_{s'} \exp(\text{score}(h_t, \bar{h}_{s'}))} \quad (10)$$

其中，score评分函数用于将目标隐藏状态 $h_t$ 与每个源隐藏状态 $\bar{h}_s$ 进行比较，并将结果标准化以得到注意力权重。score评分函数定义为

$$\text{score}(h_t, \bar{h}_s) = h_t^T W \bar{h}_s \quad (11)$$

如式(11)所示，评分函数被参数化为一个前馈神经网络，其与模型的所有其他组件一起进行联合训练。模型直接计算软注意力，并通过损失函数的梯度进行反向传播<sup>[75]</sup>。

给定注意力权重，上下文向量 $c_t$ 通过对所有源隐藏状态的加权平均来计算，计算公式为

$$c_t = \sum_s \alpha_{ts} \bar{h}_s \quad (12)$$

因此，计算路径为 $h_t \rightarrow \alpha_t \rightarrow c_t \rightarrow \tilde{h}_t$ ，然后使用Softmax函数进行预测<sup>[77]</sup>。

注意力机制已被证明在各种自然语言处理任务中非常有效，因为它能够捕捉文本中的长期依赖关系<sup>[69]</sup>。近期的研究还证明了它在时间序列分类任务中的有效性<sup>[78-80]</sup>。正如本文提到的，注意力模块已被嵌入到编码器-解码器模型中以提高模型性能。然而，目前只有编码器和注意力模块用于时间序列分类。早期的时间序列分类模型采用了与自然语言处理模型相似的骨干结构，并使用基于循环的模型，如RNN<sup>[81]</sup>，GRU<sup>[78]</sup>和LSTM<sup>[82,83]</sup>对输入序列进行编码。例如，多视图注意力网络(Multi-view Attention Network, MuVAN)在每个输入维度上独立应用了双向GRU作为编码器，并将所有表示传入一个自注意力块中<sup>[78]</sup>。

鉴于CNN模型的出色性能，许多研究尝试在应用注意力之前使用CNN对时间序列进行编码<sup>[79,84,85]</sup>。交叉注意力稳定全卷积神经网络(Cross Attention Stabilized Fully Convolutional Neural network, CA-SFCN)<sup>[14]</sup>和局部感知可解释卷积注意力网络(Locality Aware Explainable Convolutional AT-

tention network, LAXCAT)<sup>[79]</sup>运用了注意机制长期依赖关系来处理多变量时间序列分类任务。CA-SFCN结合了FCN和两种注意机制—时间注意力和变量注意力，它们相互作用以捕捉长距离依赖关系和变量交互作用。LAXCAT也使用了时间和变量注意力来识别有信息性的变量以及用于分类的有信息模式的时间间隔。

已经提出了几种注意力模型以提高网络性能<sup>[86,87]</sup>，包括Squeeze-and-Excitation(SE)<sup>[88]</sup>，其专注于通道注意力，并经常用于对时间序列数据进行分类<sup>[78]</sup>。SE模块允许整个网络使用全局信息，以选择性地关注有信息的特征图并抑制不重要的特征图<sup>[88]</sup>。每个特征图的权重在网络的每一层都是自动学习的，而SE块可以增强整个网络中的特征区分能力。多尺度注意力卷积神经网络(Multi-scale Attention Convolutional Neural Network, MACNN)<sup>[80]</sup>通过使用不同大小的卷积核以捕捉沿时间轴的不同尺度的信息，并生成不同尺度的特征图。然后，使用SE块通过自动学习每个特征图的重要性来增强有用的特征图并抑制不重要的特征图。

#### (b) Transformers。

近年来，Transformer在自然语言处理以及计算机视觉任务中不断取得突破，成为深度学习领域的基础模型。其具有编码器-解码器结构，接受源语言的词序列作为输入，并生成目标语言的翻译文本<sup>[69]</sup>。编码器和解码器都由多个相同的模块组成。每个编码器模块由多头自注意力模块和前馈网络(position-wise Feed Forward Network, FFN)组成，而每个解码器模块在多头自注意力模块和FFN之间插入了交叉注意力模块。与循环神经网络不同，Transformer使用位置编码来建模序列信息。

Transformer架构基于点积操作来寻找各输入片段之间的关联或相关性。图6为Transformer的多头注意力模块，包括八个注意力头，输入序列为两个时间步。如图6所示，Transformer中的注意力操作从输入 $x_i$ 构建3个不同的线性加权向量，分别称为查询 $q_i$ 、键 $k_i$ 和值 $v_i$

$$q_i = W_q x_i, k_i = W_k x_i, v_i = W_v x_i \quad (13)$$

其中， $W_q$ 、 $W_k$ 和 $W_v$ 是可学习的权重矩阵。输出向量 $z_i$ 计算公式为

$$z_i = \sum_j \text{softmax} \left( \frac{q_i^T k_j}{\sqrt{d_q}} \right) v_j \quad (14)$$

需要注意的是，值向量 $v_i$ 的加权取决于位置 $i$ 上查询向量 $q_i$ 与位置 $j$ 上键向量 $k_j$ 之间的相关性映射。随着输入嵌入规模的增大，点积的数值也会增加。由于Softmax函数对较大的数值非常敏感，注意力权重通过查询向量和键向量维度 $d_q$ 的平方根进行缩放。输入数据可能包含多个层次的相关信息，学习过程可能会受益于以多种不同的方式处理输入数据。为此，引入了多个注意力头，它们并行地对同一输入进行操作，并使用不同的权重矩阵 $W_q$ 、 $W_k$ 和 $W_v$ 来提取输入数据之间的不同层次的相关性。

用于分类的Transformer通常采用简单的编码器结构，包括注意力层和前馈层。SAnD<sup>[89]</sup>架构首次采用类似于标准Transformer<sup>[69]</sup>的多头注意力机制对临床时间序列进行分类。该模型同时使用位置编码和稠密插值嵌入技术，将时间顺序融入表示学习中。在另一项振动信号分类的研究中<sup>[90]</sup>，使用频率系数和短时傅里叶变换(STFT)频谱等时频特征作为输入嵌入到transformers中。还有一项基于多头注意力的研究<sup>[91]</sup>应用于原始光学卫星时间序列分

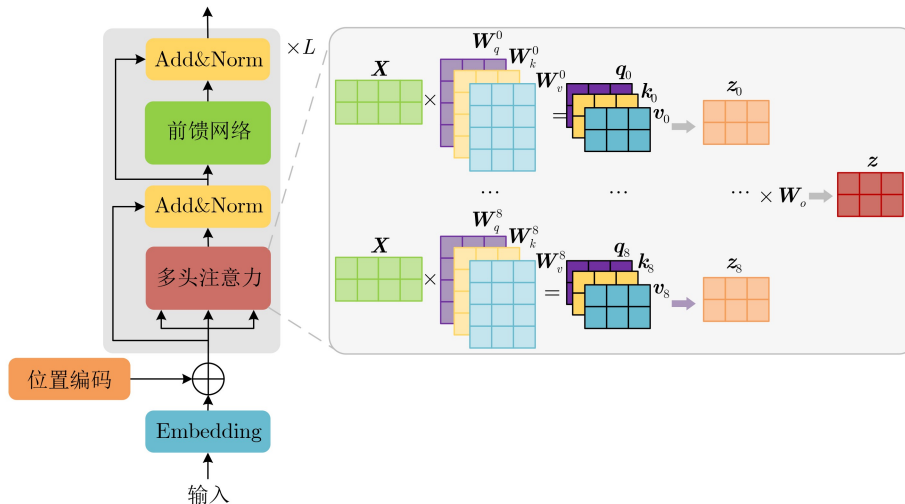


图6 多头注意力模块

类,采用高斯过程插值<sup>[92]</sup>嵌入方法,相较于卷积和循环神经网络获得更好的表现。门控Transformer网络(Gated Transformer Networks, GTN)<sup>[93]</sup>使用了两个塔式多头注意力来捕获输入序列中的判别信息,并通过一个名为gating的可学习矩阵将两个塔的输出合并。为了增强Transformer在时间序列分类中的局部感知能力,灵活的多头线性注意力(FMLA)结合了可变形卷积块、在线知识蒸馏以及随机掩码以减少噪声<sup>[94]</sup>。时频域多尺度交叉注意力融合网络(TFFormer)将时间序列时频域谱序列输入Transformer,使用时频域交叉注意力模块加强时间序列时域和频域之间的信息交互,使模型挖掘更深层次的特征,从而提高了模型分类精度。时频域多尺度交叉注意力融合网络<sup>[95]</sup>(TFFormer)将时间序列的时频域谱序列输入到Transformer中,通过时频域交叉注意力模块来增强时间序列的时域和频域之间的信息交互,从而使模型能够挖掘更深层次的特征并提高分类精度。

### (c) 自监督注意力模型。

自监督学习是一种在训练过程中使用自动标注而非人工标注的数据集进行模型学习的方法。当手动标注数据困难或成本高,或者有大量可用数据用于训练时,这种方法非常有用。在时间序列分类的背景下,可以通过自动生成时间序列数据的标签来应用自监督学习。例如,可以训练模型来预测序列中的下一个时间步,或者预测某个时间步的时间序列值。然后可以使用这些标签来训练模型,以从时间序列数据中学习有用的特征,从而提高其分类性能。

随着基于Transformer的自监督学习模型(如BERT<sup>[70]</sup>)的发展,许多研究尝试采用相同的结构来处理时间序列分类问题<sup>[74]</sup>。BENDER<sup>[74]</sup>设计了一种用于时间序列的编码器替代了wav2vec,以利用相同的结构处理时间序列数据。BENDER的研究表明,如果我们有大量的EEG数据,预训练模型可以有效地对具有不同硬件记录的EEG序列进行建模。类似地,基于Transformer的语音到时间序列(Voice-to-Series, V2Sa)模型使用大规模预训练的语音处理模型来解决时间序列分类等下游问题<sup>[96]</sup>。另一项研究引入了一个基于Transformer的框架(TransFormer-based framework, TST),将标准Transformer应用于多变量时间序列领域<sup>[15]</sup>。TST只使用Transformer的编码器部分,并以无监督的方式对其进行预训练,使用按比例遮蔽的数据。最后,预训练模型在分类和回归等下游任务中进行微调。最近,TEST<sup>[97]</sup>通过设计适用于大语言模型(Large Language Models, LLMs)的时间序列嵌入

表3 基于注意力的时间序列分类模型总结

模型	提出年份	Embedding	注意力
自适应模型			
MuVAN <sup>[78]</sup>	2018	Bi-GRU	注意力
ChannelAtt <sup>[81]</sup>	2018	RNN	注意力
GeoMAN <sup>[82]</sup>	2018	LSTM	注意力
Multi-Stage-Att <sup>[83]</sup>	2020	LSTM	注意力
CT_CAM <sup>[84]</sup>	2020	FCN + Bi-GRU	注意力
CA-SFCN <sup>[14]</sup>	2020	FCN	注意力
RTFN <sup>[85]</sup>	2021	CNN + LSTM	注意力
LAXCAT <sup>[79]</sup>	2021	CNN	注意力
MACNN <sup>[80]</sup>	2021	Multi-scale CNN	注意力
Transformers			
SAnD <sup>[89]</sup>	2018	线性Embedding	多头注意力
T2 <sup>[91]</sup>	2021	高斯过程	多头注意力
GTN <sup>[93]</sup>	2021	线性Embedding	多头注意力
TRANS_tf <sup>[90]</sup>	2021	时频特征	多头注意力
FMLA <sup>[94]</sup>	2022	可变形卷积	多头注意力
TFFormer <sup>[95]</sup>	2023	线性Embedding	多头注意力
自监督注意力			
BENDER <sup>[71]</sup>	2021	Wav2Vec 2.0+ 自监督	多头注意力
TST <sup>[15]</sup>	2021	线性Embedding+自监督	多头注意力
TARNet	2022	线性Embedding+自监督	多头注意力
TEST <sup>[97]</sup>	2023	线性Embedding+自监督	多头注意力

方法,来激活LLM对时间序列数据的处理能力。尽管实验结果表明该方法无法显著超越当前专为TSC任务定制的最佳模型,但通过监督微调,将LLM与时间序列的特性进行对其,引导LLM适应时间序列,然后以下游预测任务为导向进一步对模型进行微调,使得模型能够更好地适配各类不同域的数据及不同的下游任务的思路,具有极大的研究潜力<sup>[98]</sup>。

在时间序列分类中,当存在大量可用数据且手动标注困难时,自监督学习是一种有效的方法。然而,由于自动生成的标签可能无法准确反映数据中的真实潜在关系,所学习的特征和预测质量可能不如监督学习产生的好。TARNet<sup>[99]</sup>使用了通过分类任务学习得到的监督注意力得分分布,以屏蔽时间戳,而非随机遮盖输入序列的时间戳。表3总结了基于注意力的时间序列分类模型。

## 4 应用

时间序列分类在多个领域拥有广泛的应用,如人类活动识别、脑电图情绪识别和系统监测等。本综述重点关注时间序列分类在人类活动识别和医疗诊断中脑电图情绪识别的应用,并对这些领域的最新发展和挑战进行了简要概述。

(1)人体活动识别最新进展和挑战。

HAR是通过传感器或其他仪器收集的数据进行分析来识别或监测人类活动<sup>[100]</sup>。近年来,可穿戴技术和物联网的迅猛发展不仅导致活动数据的大量收集<sup>[101]</sup>,也使得应用这些数据来改善人类生活的质量和安全的应用程序部署变得更加容易<sup>[100]</sup>。因此,HAR是一个重要的研究领域,其应用包括医疗保健、健身监测、智能家居<sup>[102]</sup>以及辅助生活<sup>[103]</sup>等。用于收集HAR数据的设备可以分为视觉设备或基于传感器的设备。基于传感器的设备又可以进一步分为对象传感器(例如嵌入到物体中的RFID)、环境传感器(运动传感器、固定位置的WiFi或蓝牙设备)以及可穿戴传感器<sup>[3]</sup>。然而,大多数HAR研究使用可穿戴传感器或视觉设备的数据<sup>[102]</sup>。此外,通过视觉设备数据中进行人体活动识别需要使用计算机视觉技术,超出了本综述的范围。因此,本节将回顾基于可穿戴传感器的HAR方法。

在可穿戴设备中,主要使用的传感器包括加速度计、陀螺仪和磁传感器<sup>[101]</sup>,每个传感器在时间上收集3维空间数据。惯性测量单元(IMUs)是将这3个传感器结合在一个单元中的可穿戴设备<sup>[102]</sup>。通常,研究可穿戴设备的数据会使用身体不同部位上的多个IMUs来收集<sup>[104]</sup>。为了创建适用于HAR建模的数据集,传感器数据被分成(通常是等大小的)时间窗口<sup>[105]</sup>。然后,任务是学习一个将每个时间窗口的多元传感器数据映射到一组活动的函数。因此,数据形成了多维时间序列,可以应用时间序列分类方法。

鉴于本综述的研究范围,本节仅提供了对使用深度学习进行HAR研究的简要概述。用于HAR的深度学习的方法包括CNN和RNN,以及混合的CNN-RNN模型。尽管一些模型包括注意力模块,但本文没有找到任何使用完全注意力或Transformer模型的研究。

(a) 卷积神经网络。

用于HAR的卷积核中,最常见的一种是 $k \times 1$ 核。该卷积核将 $k$ 个时间步一起卷积,依次沿着输入特征中的每个时间序列移动<sup>[106]</sup>,因此虽然权重在输入特征之间共享,但特征之间没有混合。最后一个卷积层的输出被展平,并通过全连接层进行处理,然后进行最终的分类。Ronao等人<sup>[107]</sup>对HAR的CNN模型进行了全面评估,评估了使用的层数、卷积核数量和卷积核大小的变化对模型的影响。输入数据是从智能手机的加速度计和陀螺仪传感器中收集的。Ignatov<sup>[108]</sup>使用了一个单层的CNN,并在传递到全连接层之前用统计特征增强了提取的特

征。该架构对于短时间序列(1 s)非常有效,因此对于实时活动建模非常有用。Li等人<sup>[103]</sup>使用了一种具有更大卷积核尺寸和更浅层数的一维卷积神经网络(One-Dimensional Convolutional Neural Network, 1D-CNN)进行特征提取。然后,他们通过拼接捕获到的基本特征,并使用两个全连接层进行特征融合。最后,他们通过softmax对活动进行分类。

上述方法的一个缺点是它要求在所有输入特征之间共享权重。这在使用从多个设备收集的数据时可能不是最优的。在这种情况下,可以为每个设备使用独立的CNN,这样可以对特征进行独立的加权。同样,由于每个传感器通常是三轴的,可以为每个轴使用单独的CNN<sup>[109]</sup>。然后,每个CNN提取的特征被连接起来,并通过全连接层<sup>[109]</sup>或注意力头<sup>[110]</sup>进行处理。

区别于上述两种最常见方法的思路,DCNN<sup>[111]</sup>通过使用离散傅立叶变换来预处理传感器数据,将惯性测量单元(IMU)数据转换为频率信号,然后使用2维卷积来提取组合的时间和频率特征。Lee等人<sup>[112]</sup>对三轴加速度计数据进行了预处理,转换为幅度向量,然后并行使用不同核大小的CNN来提取不同尺度的特征。Xu等人<sup>[113]</sup>在2维卷积神经网络(Two-Dimensional Convolutional Neural Network, 2D-CNN)和ResNet模型中使用了可变形卷积,发现这些模型表现比非可变形卷积模型更好。

(b) 循环神经网络。

在HAR领域,已经提出了多种LSTM模型。Mura<sup>[114]</sup>设计并比较了3种多层LSTM模型,分别是单向LSTM、双向LSTM和级联LSTM(第1层为双向,其后为单向)。在每种情况下,所有时间步长的输出被用作分类层的输入。Zeng等人<sup>[115]</sup>在LSTM中添加了两个注意力层,即LSTM之前的传感器注意力层和LSTM之后的时间注意力层。他们还包括了一个称为“连续注意力”的正则化项,以平滑注意力权重之间的过渡。Guan<sup>[116]</sup>通过在每个训练时期保存模型,然后根据验证集的结果选择最佳数量的模型,创建了一个集成LSTM模型,以减少模型的方差。

(c) 混合模型。

许多最近的研究集中在混合模型上,结合了CNN和RNN。Deep-ConvLSTM<sup>[101]</sup>由4个时间卷积层和两个LSTM层组成,作者发现它的性能比等效的CNN(将LSTM层替换为全连接层)更好。由于LSTM层的参数比全连接层少,DeepConvLSTM模型也更小。Singh等人<sup>[117]</sup>使用CNN对空间数据进行编码(即每个时间戳的传感器读数),然后使用单个LSTM

层对时间数据进行编码,最后使用自注意力层对时间步加权。他们发现这个模型的表现优于使用CNN层中的时间卷积的等价模型。Challa等人<sup>[118]</sup>提出使用3个具有不同核大小的并行1D-CNN,然后是2个双向LSTM层和1个全连接层。Nafea等人<sup>[119]</sup>也使用了具有不同核大小和双向LSTMs的1D-CNNs。然而,他们为CNN和LSTMs使用了独立的分支,将每个分支中提取的特征合并到最终的全连接层中。Mekruksavanich<sup>[120]</sup>比较了一个4层的CNN-LSTM模型与一个较小的CNN-LSTM模型和LSTM模型,发现额外的卷积层可以提高性能。Chen等人<sup>[121]</sup>是另一个多层模型。它使用并行的1D-CNN,每个都具有不同的卷积核和池化大小,以提取与不同类型活动相关的不同类型的特征。紧随其后的是一个组合的1D-CNN,然后是两个LSTM层。Mekruksavanich<sup>[122]</sup>组合了4种不同的模型:一个CNN,一个LSTM,一个CNN-LSTM和一个ConvLSTM模型。该模型不仅能够识别用户正在执行的活动,还能够识别执行活动的参与者。

一些混合模型使用GRUs而不是LSTMs。ICGNet<sup>[122]</sup>的CNN部分灵感也来自著名的Inception模块。它由4个并行的卷积分支组成,每个分支通过使用不同尺寸的卷积核在多个尺度上捕获数据中的信息。卷积模块的输出经过展平,并传入两个堆叠的GRU层。最后,通过softmax激活函数生成所有活动类别的概率分布,得到分类结果。AttNens使用快速傅立叶变换生成频率特征,然后分别对每个时间步进行卷积。注意力层用于对提取的频率特征进行加权。然后,这些特征通过带有时间注意力的GRU层进行处理,以提取时间特征。CNN-BiGRU<sup>[124]</sup>使用CNN层从传感器数据中提取空间特征,然后1个或多个GRU层提取时间特征。模型的最后部分是一个全连接模块,包括1个或多个隐藏层和1个Softmax输出层。

## (2) 脑电图情绪识别最新进展和挑战。

情绪对于人类的决策、规划、推理和其他心理状态都起着至关重要的作用。人类的情绪可以通过多种来源来识别,如面部表情、语言、行为或生理信号<sup>[125]</sup>。使用生理信号可以增强情绪识别的客观性和可靠性。与外周生理信号相比,脑电图(EEG)是一种非侵入性的生理信号,可直接测量情绪状态下的脑电活动<sup>[126]</sup>。EEG信号具有较高的时间分辨率、更快的数据采集和传输速度,以及更低的成本。此外,EEG是一种自发和非主观的生理信号,可以客观地反映人类的情绪状态<sup>[127]</sup>,因此在情绪识别研究中被广泛应用。由于脑电图信号的非平稳

性、非线性特性以及受肌肉活动和电极接触不良引起的伪影影响,基于脑电图的情绪识别具有极大的挑战性。EEG信号为典型的多维时间序列,可以将时间序列分类方法用于脑电图情绪识别。

鉴于本综述的研究范围,本节其余部分仅提供了对使用深度学习进行脑电图情绪识别研究的简要概述。同样,用于脑电图情绪识别的深度学习方法可以分为CNN、RNN以及混合的CNN-RNN模型。尽管部分研究使用Transformer模型,但相关研究目前较少。

### (a) 卷积神经网络。

在情绪识别研究中,EEG信号通常是1维信号,可以使用1D-CNN架构的卷积层来处理。卷积层在情绪识别中起着关键作用,可以识别EEG信号中的模式。此外,池化层和FC层用于降维和信号分类。针对不同类别样本数量不平衡的问题,Chen等人<sup>[128]</sup>提出了一种基于1D-CNN和Borderline-SMOTE数据增强方法的情感识别模型。作者使用短时傅里叶变换提取频域特征,并引入Borderline-SMOTE数据增强方法来平衡EEG信号数据集。最后,他们使用1D-CNN模型对情感的两个维度(愉悦度和唤醒度)进行了三分类任务。Gao等人<sup>[129]</sup>提出了一种具有优先级概念的梯度优先粒子群优化(GPSO)方法,用于选择深度学习模型结构。基于GPSO的1D-CNN模型能够实现显著提高脑电图情绪识别的准确率,并提供了一个对于具有不确定网络深度的情感识别任务的CNN模型的有效自动优化框架。Maheshwari<sup>[130]</sup>等人提出了一种基于节律选择的1D-CNN模型用于使用多通道EEG信号进行自动情感识别。该模型包括8个1维卷积层、3个1维平均池化层、4个批次归一化层、3个空间dropout层、2个dropout层、1个1维全局平均池化层、3个全连接层和1个用于分类的softmax层。与2D-CNN架构相比,1D-CNN架构具有更少的超参数,因此更容易在硬件资源有限的情况下实现。由于2D-CNN架构是CNN的默认架构,因此它们广泛用于医学数据的分割和分类,2D-CNN架构也已在一些情绪识别研究中成功实现。通常,首先使用高级预处理技术将1维EEG信号转换为2D图像。在后续步骤中,2D-CNN架构根据EEG信号提取特征并对2D图像进行分类。Wang等人<sup>[131]</sup>提出了一种基于脑电图的新型深度神经网络用于情绪分类,所提出的2D-CNN使用两个不同大小的卷积核沿着时间方向和空间方向提取情感相关特征。Khare等人<sup>[132]</sup>提出通过使用不同的CNNs进行特征的自动提取和分类方法。首先,该方法使用时频表示将滤波后的脑电

图信号转换为图像，然后这些图像被输入到可配置的CNN和预训练的AlexNet, ResNet50和VGG16进行训练。实验结果显示，可配置的CNN需要的学习参数非常少，且具有更高的精度。2D-CNN在脑电图信号情感识别方面具有多种优势，它们擅长识别空间模式，能够有效捕捉医疗数据中的复杂关系，非常适合情感识别。然而，1D-CNN和2D-CNN的一个主要缺点是在面对复杂且高度变化的数据时可能表现不佳。

#### (b) 循环神经网络。

RNN是一种常用于时间序列分析的深度学习模型，主要的RNN架构包括RNN, LSTM和GRU。已经成功提出多种基于RNN的EEG信号情感识别模型。为了提高通过EEG信号进行情感识别模型的性能，Algarni等人<sup>[133]</sup>提出了一种基于深度学习的EEG信号情感识别方法。该方法包括数据选择、特征提取、特征选择和分类四个阶段。作者提取了统计特征、小波特征和Hurst指数特征，并使用BGWO进行特征选择来优化模型的性能。在分类阶段，采用了堆叠的双向长短时记忆(Bidirectional Long Short-Term Memory networks, BiLSTM)模型来识别人类的情感。由于使用了一系列改进的方法，所提出的方法相对于传统的LSTM具有更高的准确性。Sharma等人<sup>[134]</sup>采用LSTM从经过特征提取和降维处理后的数据中提取情绪变化特征，用于在线情绪识别。所提模型由4层组成，包括序列输入层、双向LSTM层、全连接层、softmax分类层。R2G-STNN<sup>[135]</sup>结合了空间和时间神经网络模型，通过分层学习过程来提取空间-时间EEG特征。采用BiLSTM网络来分析脑区内部和脑区间的空间关系，并引入区域关注层来调整不同脑区的贡献。脑电信号是动态时间序列，时间信息通常携带重要的情绪信息，循环神经网络对于识别不同的情绪非常有帮助。然而，RNN在EEG信号分类中的一个关键局限是在训练过程中会出现梯度消失问题。由于EEG信号通常包含复杂的时间依赖关系，RNN的这些限制可能导致模型在EEG信号分类任务中整体性能下降。

#### (c) 混合模型。

许多最新的研究聚焦于混合型模型，这些模型融合了CNN和RNN各自优势。Xiao等人<sup>[136]</sup>将原始脑电信号转换为4维空间-频谱-时间表示。首先使用CNN处理频谱和空间信息，并在空间和频谱维度上应用注意力机制以捕捉重要的大脑区域和频率带。然后利用BiLSTM模型集成时间注意力机制，以揭示4维表示中不同时间片的重要性。Kang等人<sup>[137]</sup>

使用CNN在时间和频率轴上从EEG频谱图在时频TF域中提取与情绪相关的特征，并将两个轴上整合的信息拼接起来，然后送入LSTM进行分类。Iyer等人<sup>[138]</sup>使用了基于CNN和LSTM的混合模型，提取的EEG特征分别输入并行的CNN, LSTM和混合模型，然后对这3个模型的预测结果进行集成得到最终预测结果。Kim等人<sup>[139]</sup>利用具有独立分支的基于注意力机制LSTM网络和CNN分别提取EEG信号的特征，然后合并后输入最终的全连接层中进行分类。区别于上述两种采用串行或并行连接方式的CNN和LSTM混合模型，Rajpoot等人<sup>[140]</sup>提出了一种新型的深度学习框架，可以独立于受试者进行情感识别。该框架由两个部分组成。首先，他们提出了一种带有通道注意自动编码器的无监督LSTM，用于获得受试者不变的潜在向量空间。其次，他们引入了一种具有注意机制的CNN，用于在获得的低维潜在空间中执行独立于受试者的情感识别任务。EEGFuseNet<sup>[141]</sup>通过集成CNN, RNN和生成对抗网络(Generative Adversarial Networks, GAN)来整合不同来源EEG的特征信息。具体来说，CNN从原始EEG信号中提取特征，RNN通过探索时间邻接处的潜在特征关系来增强特征表示。为了提高训练性能，GAN被纳入通过无监督方式动态更新来改进CNN-RNN网络的训练过程。

Transformer模型是深度学习体系结构中的一种较新架构，近年来引起了医学研究人员的关注。该模型基于编码器-解码器结构，注意力层起着关键作用。Transformer主要用于医学数据的分割或分类，未来的研究可以探索使用Transformer模型来分析从EEG信号中识别情绪的方法。一种潜在的方法是在预处理阶段将EEG信号转换为2D数据，然后应用Transformer模型，例如视觉Transformer (Vision Transformer, ViT)<sup>[142]</sup>进行情绪识别。

## 5 基于深度学习的时间序列分类研究趋势

近年来，深度学习一直是TSC研究中最活跃的领域，发表的相关论文数量不断增加。然而，尚未出现TSC领域的“Alexnet”，即一种能够主导其他所有方法的模型。深度学习算法在时间序列分类领域仍有很大的改进的空间，待解决的问题和未来研究趋势主要包括：

(1) 如何更好地处理不等长度时间序列：在实际应用中，时间序列通常具有不等长度。然而，目前很少有关于如何最好地处理不等长度时间序列的研究。此外，现有模型通常假设所有时间序列具有相同的采样频率，即序列的每个数据点是以相同的

时间间隔测量得到的。针对上述问题，如何在数据预处理层面或网络架构设计层面采取有针对性的措施，以满足时间序列分类精度的需求，是值得考虑的一个方向。

(2) 如何设计最佳的网络架构：深度学习目前仍是TSC研究中最活跃的领域。有学者认为深度学习对于视觉和语音等领域的影响可能会在TSC研究中得以复制，进一步优化和改进现有的深度学习模型或者探索新的模型结构，使其更适应时间序列数据的特点，从而提升时间序列分类模型的性能将是未来有希望的研究方向。

(3) 如何提升模型的可解释性：深度学习模型存在“黑盒效应”，使其决策过程缺乏透明性。然而，在医学和金融等关键的领域，这种不透明性可能导致不可接受的错误决策。因此，研究可解释性的深度学习模型以及使用可视化技术突出显示输入时间序列的哪些部分对于特定类别识别的贡献是值得研究的一个方向。

(4) 如何有效地处理类别不平衡：目前，大多数时间序列分类模型的训练都使用类别平衡的数据集。然而，在许多实际应用场景中，存在时间序列数据类别不平衡的情况，例如医学诊断、工业设备健康监测，无线通信或雷达信号识别以及金融欺诈检测等领域。通常，某些类别的样本数量明显少于其他类别，这可能导致模型在训练和评估时出现偏差。因此，如何有效处理类别不平衡问题在时间序列分类中非常重要，也是值得考虑的一个方向。

(5) 如何利用大语言模型处理时间序列：预训练大型语言模型具有强大的表示学习能力。最近，一些研究开始关注如何利用这些大型语言模型来处理时间序列，包括探索从零开始设计并预训练适用于时间序列的基础大模型，然后根据下游任务对模型进行微调，或者设计相应机制对时间序列进行适配，使其能够适用于现有的语言模型。因此，基于现有的语言模型处理时间序列的各类下游任务是未来的研究热点之一。

(6) 构建大型通用标签数据集：区别于计算机视觉领域的ImageNet数据集，其中包含1000个类别。时间序列分类领域目前面临的一个重要挑战，即缺乏一个大型通用标签数据集。为了有效评估时间序列分类深度学习模型，迫切需要构建一个时间序列大型通用标签数据集。

## 6 结束语

本文回顾了时间序列分类相关概念，并介绍了广泛应用于时间序列分类研究的基准数据集。然

后，从多层感知机、卷积神经网络、循环神经网络和注意力机制四个网络架构的角度对当前最新的时间序列分类模型进行分类回顾和总结，并指出它们各自的优缺点。此外，讨论了时间序列分类在人体活动识别和脑电图情绪识别两个关键领域的最新研究进展和挑战。最后，从6个角度总结了基于深度学习的时间序列分类模型未来的发展趋势。总体而言，在时间序列分类领域，使用深度网络架构及其改进方法取得了显著进展，并将在解决各种实际问题中继续发挥重要作用。本文希望这篇综述能够推动深度学习技术在时间序列分类研究中的进一步发展。

## 参考文献

- [1] YANG Qiang and WU Xindong. 10 Challenging problems in data mining research[J]. *International Journal of Information Technology & Decision Making*, 2006, 5(4): 597–604. doi: [10.1142/S0219622006002258](https://doi.org/10.1142/S0219622006002258).
- [2] BAGNALL A, LINES J, BOSTROM A, *et al.* The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances[J]. *Data Mining and Knowledge Discovery*, 2017, 31(3): 606–660. doi: [10.1007/s10618-016-0483-9](https://doi.org/10.1007/s10618-016-0483-9).
- [3] ZHANG Shibo, LI Yaxuan, ZHANG Shen, *et al.* Deep learning in human activity recognition with wearable sensors: A review on advances[J]. *Sensors*, 2022, 22(4): 1476. doi: [10.3390/s22041476](https://doi.org/10.3390/s22041476).
- [4] CHEN Kaixuan, ZHANG Dalin, YAO Lina, *et al.* Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities[J]. *ACM Computing Surveys*, 2022, 54(4): 77. doi: [10.1145/3447744](https://doi.org/10.1145/3447744).
- [5] KHADEMI Z, EBRAHIMI F, and KORDY H M. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals[J]. *Computers in Biology and Medicine*, 2022, 143: 105288. doi: [10.1016/j.combiomed.2022.105288](https://doi.org/10.1016/j.combiomed.2022.105288).
- [6] ABANDA A, MORI U, and LOZANO J A. A review on distance based time series classification[J]. *Data Mining and Knowledge Discovery*, 2019, 33(2): 378–412. doi: [10.1007/s10618-018-0596-4](https://doi.org/10.1007/s10618-018-0596-4).
- [7] DAU H A, BAGNALL A, KAMGAR K, *et al.* The UCR time series archive[J]. *IEEE/CAA Journal of Automatica Sinica*, 2019, 6(6): 1293–1305. doi: [10.1109/JAS.2019.1911747](https://doi.org/10.1109/JAS.2019.1911747).
- [8] BAGNALL A, DAU H A, LINES J, *et al.* The UEA multivariate time series classification archive, 2018[EB/OL]. <https://arxiv.org/abs/1811.00075>, 2018.
- [9] ISMAIL FAWAZ H, LUCAS B, FORESTIER G, *et al.*



- Inceptiontime: Finding alexnet for time series classification[J]. *Data Mining and Knowledge Discovery*, 2020, 34(6): 1936–1962. doi: [10.1007/s10618-020-00710-y](https://doi.org/10.1007/s10618-020-00710-y).
- [10] ISMAIL FAWAZ H, FORESTIER G, WEBER J, *et al.* Deep learning for time series classification: A review[J]. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917–963. doi: [10.1007/s10618-019-00619-1](https://doi.org/10.1007/s10618-019-00619-1).
- [11] WANG Zhiguang, YAN Weizhong, and OATES T. Time series classification from scratch with deep neural networks: A strong baseline[C]. 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, USA, 2017: 1578–1585. doi: [10.1109/IJCNN.2017.7966039](https://doi.org/10.1109/IJCNN.2017.7966039).
- [12] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting[C]. The 35th AAAI Conference on Artificial Intelligence, 2021: 11106–11115. doi: [10.1609/aaai.v35i12.17325](https://doi.org/10.1609/aaai.v35i12.17325).
- [13] WEN Qingsong, ZHOU Tian, ZHANG Chaoli, *et al.* Transformers in time series: A survey[J]. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, Macao, China, 2022: 6778–6795. doi: [10.24963/ijcai.2023/759](https://doi.org/10.24963/ijcai.2023/759).
- [14] HAO Yifan and CAO Huiping. A new attention mechanism to classify multivariate time series[C]. The Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 2020: 1999–2005. doi: [10.24963/ijcai.2020/277](https://doi.org/10.24963/ijcai.2020/277).
- [15] ZERVEAS G, JAYARAMAN S, PATEL D, *et al.* A transformer-based framework for multivariate time series representation learning[C]. The 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021: 2114–2124. doi: [10.1145/3447548.3467401](https://doi.org/10.1145/3447548.3467401).
- [16] SERRA J, PASCUAL S, and KARATZOGLOU A. Towards a universal neural network encoder for time series[C]. Artificial Intelligence Research and Development - Current Challenges, New Trends and Applications, CCIA 2018, 21st International Conference of the Catalan Association for Artificial Intelligence, Alt Empordà, Spain, 2018: 120–129. doi: [10.3233/978-1-61499-918-8-120](https://doi.org/10.3233/978-1-61499-918-8-120).
- [17] BANERJEE D, ISLAM K, XUE Keyi, *et al.* A deep transfer learning approach for improved post-traumatic stress disorder diagnosis[J]. *Knowledge and Information Systems*, 2019, 60(3): 1693–1724. doi: [10.1007/s10115-019-01337-2](https://doi.org/10.1007/s10115-019-01337-2).
- [18] ASWOLINSKIY W, REINHART R F, and STEIL J. Time series classification in reservoir-and model-space[J]. *Neural Processing Letters*, 2018, 48(2): 789–809. doi: [10.1007/s11063-017-9765-5](https://doi.org/10.1007/s11063-017-9765-5).
- [19] IWANA B K, FRINKEN V, and UCHIDA S. DTW-NN: A novel neural network for time series recognition using dynamic alignment between inputs and weights[J]. *Knowledge-Based Systems*, 2020, 188: 104971. doi: [10.1016/j.knosys.2019.104971](https://doi.org/10.1016/j.knosys.2019.104971).
- [20] TABASSUM N, MENON S, and JASTRZĘBSKA A. Time-series classification with SAFE: Simple and fast segmented word embedding-based neural time series classifier[J]. *Information Processing & Management*, 2022, 59(5): 103044. doi: [10.1016/j.ipm.2022.103044](https://doi.org/10.1016/j.ipm.2022.103044).
- [21] FUKUSHIMA K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. *Biological Cybernetics*, 1980, 36(4): 193–202. doi: [10.1007/BF00344251](https://doi.org/10.1007/BF00344251).
- [22] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 1097–1105.
- [23] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [24] LE GUENNEC A, MALINOWSKI S, and TAVENARD R. Data augmentation for time series classification using convolutional neural networks[C]. ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, Riva Del Garda, Italy, 2016.
- [25] LI Zewen, LIU Fan, YANG Wenjie, *et al.* A survey of convolutional neural networks: Analysis, applications, and prospects[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(12): 6999–7019. doi: [10.1109/TNNLS.2021.3084827](https://doi.org/10.1109/TNNLS.2021.3084827).
- [26] ZHENG Yi, LIU Qi, CHEN Enhong, *et al.* Time series classification using multi-channels deep convolutional neural networks[C]. 15th International Conference on Web-Age Information Management, Macau, China, 2014: 298–310. doi: [10.1007/978-3-319-08010-9\\_33](https://doi.org/10.1007/978-3-319-08010-9_33).
- [27] YANG Jianbo, NGUYEN M N, SAN P P, *et al.* Deep convolutional neural networks on multichannel time series for human activity recognition[C]. The Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 3995–4001.
- [28] ZHAO Bendong, LU Huanzhang, CHEN Shangfeng, *et al.* Convolutional neural networks for time series classification[J]. *Journal of Systems Engineering and Electronics*, 2017, 28(1): 162–169. doi: [10.21629/JSEE.2017.01.18](https://doi.org/10.21629/JSEE.2017.01.18).
- [29] LONG J, SHELHAMER E, and DARRELL T. Fully convolutional networks for semantic segmentation[C]. 2015

- IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431–3440. doi: [10.1109/CVPR.2015.7298965](https://doi.org/10.1109/CVPR.2015.7298965).
- [30] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [31] ZHOU Bolei, KHOSLA A, LAPEDRIZA A, *et al.* Learning deep features for discriminative localization[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2921–2929. doi: [10.1109/CVPR.2016.319](https://doi.org/10.1109/CVPR.2016.319).
- [32] ZOU Xiaowu, WANG Zidong, LI Qi, *et al.* Integration of residual network and convolutional neural network along with various activation functions and global pooling for time series classification[J]. *Neurocomputing*, 2019, 367: 39–45. doi: [10.1016/j.neucom.2019.08.023](https://doi.org/10.1016/j.neucom.2019.08.023).
- [33] LI Yuhong, ZHANG Xiaofan, and CHEN Deming. CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1091–1100. doi: [10.1109/CVPR.2018.00120](https://doi.org/10.1109/CVPR.2018.00120).
- [34] YAZDANBAKHSHEH O and DICK S. Multivariate time series classification using dilated convolutional neural network[EB/OL]. <https://arxiv.org/abs/1905.01697>, 2019.
- [35] FOUMANI S N M, TAN C W, and SALEHI M. Disjoint-CNN for multivariate time series classification[C]. 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 2021: 760–769. doi: [10.1109/ICDMW53433.2021.00099](https://doi.org/10.1109/ICDMW53433.2021.00099).
- [36] WANG Zhiguang and OATES T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks[C]. AAI Workshop Papers 2015, Menlo Park, USA, 2015: 40–46.
- [37] HATAMI N, GAVET Y, and DEBAYLE J. Classification of time-series images using deep convolutional neural networks[C]. SPIE 10696, Tenth International Conference on Machine Vision, Vienna, Austria, 2018: 106960Y. doi: [10.1117/12.2309486](https://doi.org/10.1117/12.2309486).
- [38] KARIMI-BIDHENDI S, MUNSHI F, and MUNSHI A. Scalable classification of univariate and multivariate time series[C]. 2018 IEEE International Conference on Big Data (Big Data), Seattle, USA, 2018: 1598–1605. doi: [10.1109/BigData.2018.8621889](https://doi.org/10.1109/BigData.2018.8621889).
- [39] YANG C L, CHEN Zhixuan, and YANG Chenyi. Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images[J]. *Sensors*, 2019, 20(1): 168. doi: [10.3390/s20010168](https://doi.org/10.3390/s20010168).
- [40] SZEGEDY C, VANHOUCHE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2818–2826. doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- [41] CHEN Wei and SHI Ke. A deep learning framework for time series classification using relative position matrix and convolutional neural network[J]. *Neurocomputing*, 2019, 359: 384–394. doi: [10.1016/j.neucom.2019.06.032](https://doi.org/10.1016/j.neucom.2019.06.032).
- [42] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. 3rd International Conference on Learning Representations, San Diego, USA, 2015. doi: [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- [43] CUI Zhicheng, CHEN Wenlin, and CHEN Yixin. Multi-scale convolutional neural networks for time series classification[EB/OL]. <https://arxiv.org/abs/1603.06995>, 2016.
- [44] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [45] SZEGEDY C, IOFFE S, VANHOUCHE V, *et al.* Inception-v4, inception-ResNet and the impact of residual connections on learning[C]. The 31th AAAI Conference on Artificial Intelligence, San Francisco, USA, 2017: 4278–4284. doi: [10.1609/aaai.v31i1.11231](https://doi.org/10.1609/aaai.v31i1.11231).
- [46] LIU C L, HSAIO W H, and TU Y C. Time series classification with multivariate convolutional neural network[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(6): 4788–4797. doi: [10.1109/TIE.2018.2864702](https://doi.org/10.1109/TIE.2018.2864702).
- [47] RONALD M, POULOSE A, and HAN D S. iSPLInception: An inception-ResNet deep learning architecture for human activity recognition[J]. *IEEE Access*, 2021, 9: 68985–69001. doi: [10.1109/ACCESS.2021.3078184](https://doi.org/10.1109/ACCESS.2021.3078184).
- [48] SUN Jingyu, TAKEUCHI S, and YAMASAKI I. Prototypical inception network with cross branch attention for time series classification[C]. 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021: 1–7. doi: [10.1109/IJCNN52387.2021.9533440](https://doi.org/10.1109/IJCNN52387.2021.9533440).
- [49] USMANKHUJAEV S, IBROKHIMOV B, BAYDADAIEV S, *et al.* Time series classification with InceptionFCN[J]. *Sensors*, 2021, 22(1): 157. doi: [10.3390/s22010157](https://doi.org/10.3390/s22010157).
- [50] 张雅雯, 王志海, 刘海洋, 等. 基于多尺度残差FCN的时间序列分类算法[J]. *软件学报*, 2022, 33(2): 555–570. doi: [10.13328/j.cnki.jos.006142](https://doi.org/10.13328/j.cnki.jos.006142).
- ZHANG Yawen, WANG Zhihai, LIU Haiyang, *et al.* Time series classification algorithm based on multiscale residual

- full convolutional neural network[J]. *Journal of Software*, 2022, 33(2): 555–570. doi: [10.13328/j.cnki.jos.006142](https://doi.org/10.13328/j.cnki.jos.006142).
- [51] DENNIS D, ACAR D A E, MANDIKAL V, *et al.* Shallow RNN: Accurate time-series classification on resource constrained devices[C]. The 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, 2019: 12896–12906.
- [52] HERMANS M and SCHRAUWEN B. Training and analyzing deep recurrent neural networks[C]. The 26th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013: 190–198.
- [53] PASCANU R, MIKOLOV T, and BENGIO Y. On the difficulty of training recurrent neural networks[C]. The 30th International Conference on International Conference on Machine Learning, Atlanta, USA, 2013: 1310–1318.
- [54] HOCHREITER S and SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735–1780. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [55] CHUNG J, GULCEHRE C, CHO K H, *et al.* Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. <https://arxiv.org/abs/1412.3555>, 2014.
- [56] KAWAKAMI K. Supervised sequence labelling with recurrent neural networks[D]. [Ph. D. dissertation], Carnegie Mellon University, 2008.
- [57] SUTSKEVER I, VINYALS O, and LE Q V. Sequence to sequence learning with neural networks[C]. The 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 3104–3112.
- [58] DONAHUE J, ANNE HENDRICKS L, GUADARRAMA S, *et al.* Long-term recurrent convolutional networks for visual recognition and description[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 2625–2634. doi: [10.1109/CVPR.2015.7298878](https://doi.org/10.1109/CVPR.2015.7298878).
- [59] KARPATHY A and Fei-Fei L. Deep visual-semantic alignments for generating image descriptions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3128–3137. doi: [10.1109/CVPR.2015.7298932](https://doi.org/10.1109/CVPR.2015.7298932).
- [60] TANG Yujin, XU Jianfeng, MATSUMOTO K, *et al.* Sequence-to-sequence model with attention for time series classification[C]. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 2016: 503–510. doi: [10.1109/ICDMW.2016.0078](https://doi.org/10.1109/ICDMW.2016.0078).
- [61] MALHOTRA P, TV V, VIG L, *et al.* TimeNet: Pre-trained deep recurrent neural network for time series classification[C]. 25th European Symposium on Artificial Neural Networks, Bruges, Belgium, 2017.
- [62] KARIM F, MAJUMDAR S, DARABI H, *et al.* Multivariate LSTM-FCNs for time series classification[J]. *Neural Networks*, 2019, 116: 237–245. doi: [10.1016/j.neunet.2019.04.014](https://doi.org/10.1016/j.neunet.2019.04.014).
- [63] 玄英律, 万源, 陈嘉慧. 基于多尺度卷积和注意力机制的LSTM时间序列分类[J]. *计算机应用*, 2022, 42(8): 2343–2352. doi: [10.11772/j.issn.1001-9081.2021061062](https://doi.org/10.11772/j.issn.1001-9081.2021061062).
- XUAN Yinglu, WAN Yuan, and CHEN Jiahui. Time series classification by LSTM based on multi-scale convolution and attention mechanism[J]. *Journal of Computer Applications*, 2022, 42(8): 2343–2352. doi: [10.11772/j.issn.1001-9081.2021061062](https://doi.org/10.11772/j.issn.1001-9081.2021061062).
- [64] ZHANG Xuchao, GAO Yifeng, LIN J, *et al.* TapNet: Multivariate time series classification with attentional prototypical network[C]. The 34th AAAI Conference on Artificial Intelligence, New York, USA, 2020: 6845–6852. doi: [10.1609/aaai.v34i04.6165](https://doi.org/10.1609/aaai.v34i04.6165).
- [65] ZUO Jingwei, ZEITOUNI K, and TAHER Y. SMATE: Semi-supervised spatio-temporal representation learning on multivariate time series[C]. 2021 IEEE International Conference on Data Mining (ICDM), Auckland, New Zealand, 2021: 1565–1570. doi: [10.1109/ICDM51629.2021.00206](https://doi.org/10.1109/ICDM51629.2021.00206).
- [66] LIN Sangdi and RUNGER G C. GCRNN: Group-constrained convolutional recurrent neural network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 4709–4718. doi: [10.1109/TNNLS.2017.2772336](https://doi.org/10.1109/TNNLS.2017.2772336).
- [67] MUTEGEKI R and HAN D S. A CNN-LSTM approach to human activity recognition[C]. 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Fukuoka, Japan, 2020: 362–366. doi: [10.1109/ICAIIIC48513.2020.9065078](https://doi.org/10.1109/ICAIIIC48513.2020.9065078).
- [68] KARIM F, MAJUMDAR S, DARABI H, *et al.* LSTM fully convolutional networks for time series classification[J]. *IEEE Access*, 2018, 6: 1662–1669. doi: [10.1109/ACCESS.2017.2779939](https://doi.org/10.1109/ACCESS.2017.2779939).
- [69] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 6000–6010.
- [70] DEVLIN J, CHANG Mingwei, LEE K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding[C]. The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Minneapolis, USA, 2018: 4171–4186. doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
- [71] LIU Ze, LIN Yutong, CAO Yue, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows[C].

- 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 9992–10002. doi: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986).
- [72] CARON M, TOUVRON H, MISRA I, *et al.* Emerging properties in self-supervised vision transformers[C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 9630–9640. doi: [10.1109/ICCV48922.2021.00951](https://doi.org/10.1109/ICCV48922.2021.00951).
- [73] KHAN S, NASEER M, HAYAT M, *et al.* Transformers in vision: A survey[J]. *ACM Computing Surveys*, 2022, 54(10s): 200. doi: [10.1145/3505244](https://doi.org/10.1145/3505244).
- [74] KOSTAS D, AROCA-OUELLETTE S, and RUDZICZ F. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data[J]. *Frontiers in Human Neuroscience*, 2021, 15: 653659. doi: [10.3389/fnhum.2021.653659](https://doi.org/10.3389/fnhum.2021.653659).
- [75] BAHDANAU D, CHO K, and BENGIO Y. Neural machine translation by jointly learning to align and translate[C]. 3rd International Conference on Learning Representations, San Diego, USA, 2015. doi: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473).
- [76] CHO K, VAN MERRIËNBOER B, GULCEHRE C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]. The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014: 1724–1734. doi: [10.3115/v1/D14-1179](https://doi.org/10.3115/v1/D14-1179).
- [77] LUONG M T, PHAM H, and MANNING C D. Effective approaches to attention-based neural machine translation[C]. The 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015: 1412–1421. doi: [10.18653/v1/D15-1166](https://doi.org/10.18653/v1/D15-1166).
- [78] YUAN Ye, XUN Guangxu, MA Fenglong, *et al.* MuVAN: A multi-view attention network for multivariate temporal data[C]. 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018: 717–726. doi: [10.1109/ICDM.2018.00087](https://doi.org/10.1109/ICDM.2018.00087).
- [79] HSIEH T Y, WANG Suhang, SUN Yiwei, *et al.* Explainable multivariate time series classification: A deep neural network which learns to attend to important variables as well as time intervals[C]. The 14th ACM International Conference on Web Search and Data Mining, 2021: 607–615.
- [80] CHEN Wei and SHI Ke. Multi-scale attention convolutional neural network for time series classification[J]. *Neural Networks*, 2021, 136: 126–140. doi: [10.1016/j.neunet.2021.01.001](https://doi.org/10.1016/j.neunet.2021.01.001).
- [81] YUAN Ye, XUN Guangxu, MA Fenglong, *et al.* A novel channel-aware attention framework for multi-channel EEG seizure detection via multi-view deep learning[C]. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, USA, 2018: 206–209. doi: [10.1109/BHI.2018.8333405](https://doi.org/10.1109/BHI.2018.8333405).
- [82] LIANG Yuxuan, KE Songyu, ZHANG Junbo, *et al.* GeoMAN: Multi-level attention networks for geo-sensory time series prediction[C]. The Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 3428–3434. doi: [10.24963/ijcai.2018/476](https://doi.org/10.24963/ijcai.2018/476).
- [83] HU Jun and ZHENG Wendong. Multistage attention network for multivariate time series prediction[J]. *Neurocomputing*, 2020, 383: 122–137. doi: [10.1016/j.neucom.2019.11.060](https://doi.org/10.1016/j.neucom.2019.11.060).
- [84] CHENG Xu, HAN Peihua, LI Guoyuan, *et al.* A novel channel and temporal-wise attention in convolutional networks for multivariate time series classification[J]. *IEEE Access*, 2020, 8: 212247–212257. doi: [10.1109/ACCESS.2020.3040515](https://doi.org/10.1109/ACCESS.2020.3040515).
- [85] XIAO Zhiwen, XU Xin, XING Huanlai, *et al.* RTFN: A robust temporal feature network for time series classification[J]. *Information Sciences*, 2021, 571: 65–86. doi: [10.1016/j.ins.2021.04.053](https://doi.org/10.1016/j.ins.2021.04.053).
- [86] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al.* Spatial transformer networks[C]. The 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 2017–2025.
- [87] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module[C]. The 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 3–19. doi: [10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [88] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141. doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
- [89] SONG Huan, RAJAN D, THIAGARAJAN J, *et al.* Attend and diagnose: Clinical time series analysis using attention models[C]. The AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018: 4091–4098. doi: [10.1609/aaai.v32i1.11635](https://doi.org/10.1609/aaai.v32i1.11635).
- [90] JIN Cancan and CHEN Xi. An end-to-end framework combining time-frequency expert knowledge and modified transformer networks for vibration signal classification[J]. *Expert Systems with Applications*, 2021, 171: 114570. doi: [10.1016/j.eswa.2021.114570](https://doi.org/10.1016/j.eswa.2021.114570).
- [91] ALLAM JR T and MCEWEN J D. Paying attention to astronomical transients: Photometric classification with the time-series transformer[EB/OL]. <https://arxiv.org/abs/2105.06178v1>, 2021.

- [92] RASMUSSEN C E. Gaussian processes in machine learning[M]. ML Summer Schools 2003 on Advanced Lectures on Machine Learning, Tübingen, Germany, 2003: 63–71. doi: [10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4).
- [93] LIU Minghao, REN Shengqi, MA Siyuan, *et al.* Gated transformer networks for multivariate time series classification[EB/OL]. <https://arxiv.org/abs/2103.14438>, 2021.
- [94] ZHAO Bowen, XING Huanlai, WANG Xinhan, *et al.* Rethinking attention mechanism in time series classification[J]. *Information Sciences*, 2023, 627: 97–114. doi: [10.1016/j.ins.2023.01.093](https://doi.org/10.1016/j.ins.2023.01.093).
- [95] 王美, 苏雪松, 刘佳, 等. 时频域多尺度交叉注意力融合的时间序列分类方法[J/OL]. 计算机应用: 1–9. <http://www.joca.cn/CN/10.11772/j.issn.1001-9081.2023060731>, 2023.  
WANG Mei, SU Xuesong, LIU Jia, *et al.* Time series classification method based on multi-scale cross-attention fusion in time-frequency domain[J/OL]. *Journal of Computer Applications*: 1–9. <http://www.joca.cn/CN/10.11772/j.issn.1001-9081.2023060731>, 2023.
- [96] YANG C H H, TSAI Y Y, and CHEN P Y. Voice2Series: Reprogramming acoustic models for time series classification[C]. The 38th International Conference on Machine Learning, 2021: 11808–11819.
- [97] SUN Chenxi, LI Yaliang, LI Hongyan, *et al.* TEST: Text prototype aligned embedding to activate LLM's ability for time series[EB/OL]. <https://arxiv.org/abs/2308.08241>, 2023.
- [98] CHANG C, PENG W C, and CHEN T F. LLM4TS: Two-stage fine-tuning for time-series forecasting with pre-trained LLMs[EB/OL]. <https://arxiv.org/abs/2308.08469>, 2023.
- [99] CHOWDHURY R R, ZHANG Xiyuan, SHANG Jingbo, *et al.* TARNet: Task-aware reconstruction for time-series transformer[C]. The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, USA, 2022: 212–220. doi: [10.1145/3534678.3539329](https://doi.org/10.1145/3534678.3539329).
- [100] GUPTA N, GUPTA S K, PATHAK R K, *et al.* Human activity recognition in artificial intelligence framework: A narrative review[J]. *Artificial Intelligence Review*, 2022, 55(6): 4755–4808. doi: [10.1007/s10462-021-10116-x](https://doi.org/10.1007/s10462-021-10116-x).
- [101] ORDÓÑEZ F J and ROGGEN D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition[J]. *Sensors*, 2016, 16(1): 115. doi: [10.3390/s16010115](https://doi.org/10.3390/s16010115).
- [102] ARSHAD M H, BILAL M, and GANI A. Human activity recognition: Review, taxonomy and open challenges[J]. *Sensors*, 2022, 22(17): 6463. doi: [10.3390/s22176463](https://doi.org/10.3390/s22176463).
- [103] LI Yang, YANG Guanci, SU Zhidong, *et al.* Human activity recognition based on multienvironment sensor data[J]. *Information Fusion*, 2023, 91: 47–63. doi: [10.1016/j.inffus.2022.10.015](https://doi.org/10.1016/j.inffus.2022.10.015).
- [104] CHENG Xin, ZHANG Lei, TANG Yin, *et al.* Real-time human activity recognition using conditionally parametrized convolutions on mobile and wearable devices[J]. *IEEE Sensors Journal*, 2022, 22(6): 5889–5901. doi: [10.1109/JSEN.2022.3149337](https://doi.org/10.1109/JSEN.2022.3149337).
- [105] LARA O D and LABRADOR M A. A survey on human activity recognition using wearable sensors[J]. *IEEE Communications Surveys & Tutorials*, 2013, 15(3): 1192–1209. doi: [10.1109/SURV.2012.110112.00192](https://doi.org/10.1109/SURV.2012.110112.00192).
- [106] WANG Xing, ZHANG Lei, HUANG Wenbo, *et al.* Deep convolutional networks with tunable speed–accuracy tradeoff for human activity recognition using wearables[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1–12. doi: [10.1109/TIM.2021.3132088](https://doi.org/10.1109/TIM.2021.3132088).
- [107] RONAO C A and CHO S B. Human activity recognition with smartphone sensors using deep learning neural networks[J]. *Expert Systems with Applications*, 2016, 59: 235–244. doi: [10.1016/j.eswa.2016.04.032](https://doi.org/10.1016/j.eswa.2016.04.032).
- [108] IGNATOV A. Real-time human activity recognition from accelerometer data using convolutional neural networks[J]. *Applied Soft Computing*, 2018, 62: 915–922. doi: [10.1016/j.asoc.2017.09.027](https://doi.org/10.1016/j.asoc.2017.09.027).
- [109] ZENG Ming, NGUYEN L T, YU Bo, *et al.* Convolutional neural networks for human activity recognition using mobile sensors[C]. 6th International Conference on Mobile Computing, Applications and Services, Austin, USA, 2014: 197–205. doi: [10.4108/icst.mobicase.2014.257786](https://doi.org/10.4108/icst.mobicase.2014.257786).
- [110] ZHANG Haoxi, XIAO Zhiwen, WANG Juan, *et al.* A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention[J]. *IEEE Internet of Things Journal*, 2020, 7(2): 1072–1080. doi: [10.1109/JIOT.2019.2949715](https://doi.org/10.1109/JIOT.2019.2949715).
- [111] JIANG Wenchao and YIN Zhaozheng. Human activity recognition using wearable sensors by deep convolutional neural networks[C]. The 23rd ACM international conference on Multimedia, Brisbane, Australia, 2015: 1307–1310. doi: [10.1145/2733373.2806333](https://doi.org/10.1145/2733373.2806333).
- [112] LEE S M, YOON S M, and CHO H. Human activity recognition from accelerometer data using convolutional neural network[C]. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, Korea, 2017: 131–134. doi: [10.1109/BIGCOMP.2017.7881728](https://doi.org/10.1109/BIGCOMP.2017.7881728).
- [113] XU Shige, ZHANG Lei, HUANG Wenbo, *et al.* Deformable convolutional networks for multimodal human activity recognition using wearable sensors[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1–14. doi:

- [10.1109/TIM.2022.3158427](https://doi.org/10.1109/TIM.2022.3158427).
- [114] MURAD A and PYUN J Y. Deep recurrent neural networks for human activity recognition[J]. *Sensors*, 2017, 17(11): 2556. doi: [10.3390/s17112556](https://doi.org/10.3390/s17112556).
- [115] ZENG Ming, GAO Haoxiang, YU Tong, *et al.* Understanding and improving recurrent networks for human activity recognition by continuous attention[C]. The 2018 ACM International Symposium on Wearable Computers, Singapore, 2018: 56–63. doi: [10.1145/3267242.3267286](https://doi.org/10.1145/3267242.3267286).
- [116] GUAN Yu and PLÖTZ T. Ensembles of deep LSTM learners for activity recognition using wearables[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, 1(2): 11. doi: [10.1145/3090076](https://doi.org/10.1145/3090076).
- [117] SINGH S P, SHARMA M K, LAY-EKUAKILLE A, *et al.* Deep ConvLSTM with self-attention for human activity decoding using wearable sensors[J]. *IEEE Sensors Journal*, 2021, 21(6): 8575–8582. doi: [10.1109/JSEN.2020.3045135](https://doi.org/10.1109/JSEN.2020.3045135).
- [118] CHALLA S K, KUMAR A, and SEMWAL V B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data[J]. *The Visual Computer*, 2022, 38(12): 4095–4109. doi: [10.1007/s00371-021-02283-3](https://doi.org/10.1007/s00371-021-02283-3).
- [119] NAFEA O, ABDUL W, MUHAMMAD G, *et al.* Sensor-based human activity recognition with spatio-temporal deep learning[J]. *Sensors*, 2021, 21(6): 2141. doi: [10.3390/s21062141](https://doi.org/10.3390/s21062141).
- [120] MEKRUKSAVANICH S and JITPATTANAKUL A. LSTM networks using smartphone data for sensor-based human activity recognition in smart homes[J]. *Sensors*, 2021, 21(5): 1636. doi: [10.3390/s21051636](https://doi.org/10.3390/s21051636).
- [121] CHEN Ling, LIU Xiaozhe, PENG Liangying, *et al.* Deep learning based multimodal complex human activity recognition using wearable devices[J]. *Applied Intelligence*, 2021, 51(6): 4029–4042. doi: [10.1007/s10489-020-02005-7](https://doi.org/10.1007/s10489-020-02005-7).
- [122] MEKRUKSAVANICH S and JITPATTANAKUL A. Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models[J]. *Electronics*, 2021, 10(3): 308. doi: [10.3390/electronics10030308](https://doi.org/10.3390/electronics10030308).
- [123] SALEEM G, BAJWA U I and RAZA R H. Toward human activity recognition: A survey[J]. *Neural Computing and Applications*, 2023, 35(5): 4145–4182. doi: [10.1007/s00521-022-07937-4](https://doi.org/10.1007/s00521-022-07937-4).
- [124] MEKRUKSAVANICH S and JITPATTANAKUL A. Deep convolutional neural network with RNNs for complex activity recognition using wrist-worn wearable sensor data[J]. *Electronics*, 2021, 10(14): 1685. doi: [10.3390/electronics10141685](https://doi.org/10.3390/electronics10141685).
- [125] CHEN Xun, LI Chang, LIU Aiping, *et al.* Toward open-world electroencephalogram decoding via deep learning: A comprehensive survey[J]. *IEEE Signal Processing Magazine*, 2022, 39(2): 117–134. doi: [10.1109/MSP.2021.3134629](https://doi.org/10.1109/MSP.2021.3134629).
- [126] GU Xiaoqing, CAI Weiwei, GAO Ming, *et al.* Multi-source domain transfer discriminative dictionary learning modeling for electroencephalogram-based emotion recognition[J]. *IEEE Transactions on Computational Social Systems*, 2022, 9(6): 1604–1612. doi: [10.1109/TCSS.2022.3153660](https://doi.org/10.1109/TCSS.2022.3153660).
- [127] MAITHRI M, RAGHAVENDRA U, GUDIGAR A, *et al.* Automated emotion recognition: Current trends and future perspectives[J]. *Computer Methods and Programs in Biomedicine*, 2022, 215: 106646. doi: [10.1016/j.cmpb.2022.106646](https://doi.org/10.1016/j.cmpb.2022.106646).
- [128] CHEN Yu, CHANG Rui, and GUO Jifeng. Effects of data augmentation method borderline-SMOTE on emotion recognition of EEG signals based on convolutional neural network[J]. *IEEE Access*, 2021, 9: 47491–47502. doi: [10.1109/ACCESS.2021.3068316](https://doi.org/10.1109/ACCESS.2021.3068316).
- [129] GAO Zhongke, LI Yanli, YANG Yuxuan, *et al.* A GPSO-optimized convolutional neural networks for EEG-based emotion recognition[J]. *Neurocomputing*, 2020, 380: 225–235. doi: [10.1016/j.neucom.2019.10.096](https://doi.org/10.1016/j.neucom.2019.10.096).
- [130] MAHESHWARI D, GHOSH S K, TRIPATHY R K, *et al.* Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel EEG signals[J]. *Computers in Biology and Medicine*, 2021, 134: 104428. doi: [10.1016/j.combiomed.2021.104428](https://doi.org/10.1016/j.combiomed.2021.104428).
- [131] WANG Yuqi, ZHANG Lijun, XIA Pan, *et al.* EEG-based emotion recognition using a 2D CNN with different kernels[J]. *Bioengineering*, 2022, 9(6): 231. doi: [10.3390/bioengineering9060231](https://doi.org/10.3390/bioengineering9060231).
- [132] KHARE S K and BAJAJ V. Time–frequency representation and convolutional neural network-based emotion recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(7): 2901–2909. doi: [10.1109/TNNLS.2020.3008938](https://doi.org/10.1109/TNNLS.2020.3008938).
- [133] ALGARNI M, SAEED F, AL-HADHRAMI T, *et al.* Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM)[J]. *Sensors*, 2022, 22(8): 2976. doi: [10.3390/s22082976](https://doi.org/10.3390/s22082976).
- [134] SHARMA R, PACHORI R B, and SIRCAR P. Automated emotion recognition based on higher order statistics and deep learning algorithm[J]. *Biomedical Signal Processing*

- and Control*, 2020, 58: 101867. doi: [10.1016/j.bspc.2020.101867](https://doi.org/10.1016/j.bspc.2020.101867).
- [135] LI Yang, ZHENG Wenming, WANG Lei, *et al.* From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition[J]. *IEEE Transactions on Affective Computing*, 2022, 13(2): 568–578. doi: [10.1109/TAFFC.2019.2922912](https://doi.org/10.1109/TAFFC.2019.2922912).
- [136] XIAO Guowen, SHI Meng, YE Mengwen, *et al.* 4D attention-based neural network for EEG emotion recognition[J]. *Cognitive Neurodynamics*, 2022, 16(4): 805–818. doi: [10.1007/s11571-021-09751-5](https://doi.org/10.1007/s11571-021-09751-5).
- [137] KANG J S, KAVURI S, and LEE M. ICA-evolution based data augmentation with ensemble deep neural networks using time and frequency kernels for emotion recognition from EEG-data[J]. *IEEE Transactions on Affective Computing*, 2022, 13(2): 616–627. doi: [10.1109/TAFFC.2019.2942587](https://doi.org/10.1109/TAFFC.2019.2942587).
- [138] IYER A, DAS S S, TEOTIA R, *et al.* CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings[J]. *Multimedia Tools and Applications*, 2023, 82(4): 4883–4896. doi: [10.1007/s11042-022-12310-7](https://doi.org/10.1007/s11042-022-12310-7).
- [139] KIM Y and CHOI A. EEG-based emotion classification using long short-term memory network with attention mechanism[J]. *Sensors*, 2020, 20(23): 6727. doi: [10.3390/s20236727](https://doi.org/10.3390/s20236727).
- [140] ARJUN, RAJPOOT A S, and PANICKER M R. Subject independent emotion recognition using EEG signals employing attention driven neural networks[J]. *Biomedical Signal Processing and Control*, 2022, 75: 103547. doi: [10.1016/j.bspc.2022.103547](https://doi.org/10.1016/j.bspc.2022.103547).
- [141] LIANG Zhen, ZHOU Rushuang, ZHANG Li, *et al.* EEGFuseNet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional EEG with an application to emotion recognition[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021, 29: 1913–1925. doi: [10.1109/TNSRE.2021.3111689](https://doi.org/10.1109/TNSRE.2021.3111689).
- [142] PARVAIZ A, KHALID M A, ZAFAR R, *et al.* Vision transformers in medical computer vision—a contemplative retrospection[J]. *Engineering Applications of Artificial Intelligence*, 2023, 122: 106126. doi: [10.1016/j.engappai.2023.106126](https://doi.org/10.1016/j.engappai.2023.106126).
- 任利强：男，博士生，研究方向为信息融合、飞行数据智能处理。  
贾舒宜：女，副教授，研究方向为信息融合。  
王海鹏：男，教授，研究方向为信息融合。  
王子玲：女，副教授，研究方向为信息融合。

责任编辑：余蓉