

# Extreme Fuzzy Broad Learning System: Algorithm, Frequency Principle, and Applications in Classification and Regression

## 极限模糊广义学习系统：算法，频率原理及其应用分类和回归

Junwei Duan , Shiyi Yao, Jiantao Tan, Yang Liu ,  
Long Chen , Senior Member , IEEE, Zhen Zhang ,  
and C. L. Philip Chen , Fellow, IEEE

2024 年 11 月 7 日

### 摘要

作为神经网络的有效替代方法，广义学习系统（BLS）因其高效的性能、出色的表现以及较短的训练过程，在分类和回归任务中受到越来越多的关注。然而，随着节点数量达到饱和点并继续增加，BLS 的性能将不会继续提高，甚至可能下降。此外，之前关于神经网络的研究通常忽视了神经网络良好泛化能力的原因。为了解决这些问题，本文首次提出了极限模糊 BLS（E-FBLS），这是一种新型的级联模糊 BLS，其中多个模糊 BLS 块被组合或级联在一起。此外，原始数据输入到每个 FBLS 块，而不是前一个块。为了进一步验证 E-FBLS 的有效性，我们使用了残差学习方法。从频域的角度，我们还发现了 E-FBLS 中频率原理的存在，这为神经网络的泛化能力提供了良好的可解释性。在经典的分类和回归数据集上的实验结果表明，提出的 E-FBLS 在处理分类和回归任务时的精度优于传统的 BLS。当块数增加到一定程度时，精度得到了提高。此外，我们验证了 E-FBLS 的频率原理，表明 E-FBLS 可以快速获得低频成分，同时随着 FBLS 块数量的增加，高频成分会逐渐调整。

**关键词：**广义学习系统（BLS）；分类；神经网络；频率原理；模糊极限学习机（ELM）；回归

# 目录

<b>1</b>	<b>引言</b>	<b>1</b>
<b>2</b>	<b>相关工作</b>	<b>2</b>
2.1	经典 BLS . . . . .	2
2.2	堆叠 BLS . . . . .	3
2.3	模糊 BLS . . . . .	4
<b>3</b>	<b>极限模糊广义学习系统</b>	<b>6</b>
3.1	E-FBLS 算法 . . . . .	6
3.2	E-FBLS 中的频率原则 . . . . .	8
<b>4</b>	<b>实验</b>	<b>10</b>
4.1	分类数据集实验 . . . . .	10
4.2	回归数据集上的实验 . . . . .	11
4.3	EBLS 中频率原则的验证 . . . . .	12
<b>5</b>	<b>讨论</b>	<b>13</b>
<b>6</b>	<b>结论</b>	<b>14</b>
<b>7</b>	<b>参考文献</b>	<b>15</b>

# 1 引言

深度学习作为最重要的机器学习技术之一，广泛应用于计算机视觉、模式识别和其他多个领域。近年来，提出了许多深度网络，例如卷积神经网络 (CNN) [19]、残差网络 (ResNet) [13]、递归神经网络 (RNN) [33] 和注意力机制变换器 [10]。尽管深度神经网络取得了重大突破，但其迭代模型构建过程通常非常耗时，因为复杂的深层结构需要调整大量超参数。为了解决这些问题，提出了许多方法，例如优化 [?] 和网络结构 [?]。宽度学习 (BLS) 是一种高效的广义神经网络，受到了越来越多的关注。它是基于随机向量函数链接神经网络 (RVFLNN) [4] 提出的。在各个领域，已对经典 BLS 进行了一些重要的扩展，以拓宽 BLS 的应用范围。例如，Jin 等 [17] 提出了一种图正则化的 BLS，采用流形学习，使得输出权重受到约束，从而学习更具判别性的信息。在解决非线性问题和分析模糊数据集方面，Feng 等提出了一种模糊 BLS [12]，将 BLS 的特征节点替换为一组 TS 模糊子系统。还有一些基于 BLS 的模型与其他方法结合提出。基于 RNN 和长短期记忆 (LSTM) [9] 分别扩展了递归 BLS 和门控 BLS。此外，为了处理无标签数据，开发了半监督 BLS [34] 和无监督 BLS [22]，以根据经典 BLS 从有标签或无标签数据中提取特征。为应对噪声和异常值，Chu 等人 [8] 提出了一种加权 BLS (WBLS)，以提高泛化能力和鲁棒性。通过结合自空间学习 (SPL)，Liu 等 [20] 提出了自适应 BLS (SSBLS)，以学习适当的优先权重，重新调整每个样本的重要性。凭借快速的效率和强大的性能，BLS 已广泛应用于医疗脑电信号分析 [35]、牵引系统故障诊断 [7]、高光谱图像分类 [18]、发作间期癫痫放电检测 [16] 等领域。

到目前为止，BLS 的级联变体已经得到了广泛的发展。通常，Zhang 等人 [11] 提出了一种修改版 BLS，具有级联增强节点和密集连接。在每个模块中，第一个增强节点的输入由所有功能节点和前一个模块的最后一个增强节点组成。此外，Xu 等人 [28] 提出了 R-BLS，通过循环连接每组中的增强节点，以捕捉时间序列的动态特征。此外，Ye 等人 [32] 设计了一种自适应深级级联 BLS (DCBLS)，通过将特征修改节点和映射节点分别转换为  $n$  和  $m$  的深级级联结构。还有，SBLS [21] 通过堆叠多个 BLS 块提出了一个更深的结构。后续 BLS 块的输入是原始数据或前一个基本 BLS 块的输出。

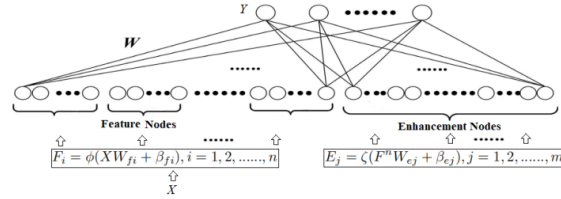


图 1: 经典 BLS 结构

BLS 是一种单层结构的神经网络，包括特征节点和增强节点，如图 1 所示。为了确保模型的高效性，BLS 采用伪逆方法来逼近期望输出。同时，在 [5] 中已数学证明了其通用逼近能力，这使得其在分类和回归任务中具有良好的表现。此外，增量学习方法也在 [4] 中用于更新 BLS 的参数，以实现优异的性能。在处理分类和回归任务时，BLS 模型非常强大，但当节点数量达到饱和点并继续增加时，BLS 的性能不会持续提高，甚至可能下降。为了解决 BLS 中特征节点和增强节点扩展的

局限性，并进一步提高性能，我们提出了一种新型的级联广义神经网络，可以同时进一步扩展广泛维度和宽度维度。同时，我们还专注于从频域角度对神经网络进行解释。本文的主要贡献如下：

(1) 极限模糊 BLS (E-FBLS) 作为一种新型的级联结构神经网络, 首次通过水平组合或合并多个模糊 BLS 块, 并结合级联残差学习方法提出。在这种方式下, 特征节点和增强节点的数量可以进一步扩展, 以便在广阔空间中更好地提取特征。

(2) 提出的 E-FBLS 方法有效地继承了 FBLS 的优势, 能够通过集成多个模糊 BLS 块从原始数据中提取潜在的模糊特征。此外, 专家的知识可以通过模糊规则有效地融合到 E-FBLS 的 FBLS 块中, 从而更好地适应数据的不确定性。

(3) 本文首次揭示了 E-FBLS 中的频率原理, 这有助于从频域的角度加强对 E-FBLS 的解释。E-FBLS 中存在的频率原理不仅为神经网络的泛化能力提供了良好的可解释性, 而且可能为扩展广义神经网络的应用提供了重要的启示。

(4) 提出的 E-FBLS 方法成功地解决了分类和回归问题, 并由于其极其广泛的表示能力, 性能显著提升。此外, 实验结果也验证了 E-FBLS 在分类和回归问题上的表现优于其他先进的神经网络方法。

本文的其余部分组织如下: 第二部分首先介绍了经典的 BLS, 并简要回顾了 SBLS 和 FBLS。第三部分主要介绍了提出的 E-FBLS 及其频率原理。在第四部分, 展示了基于不同数据集的实验结果, 以评估 E-FBLS 的性能并验证其频率原理, 同时进行讨论和分析。最后, 第五部分给出了结论。

## 2 相关工作

### 2.1 经典 BLS

经典 BLS 采用了 RVFLNN 的结构 [24]。BLS 的结构如图 1 所示。与 RVFLNN 不同, BLS 的输入层和输出层没有直接连接。在 BLS 中, 特征节点是通过映射函数  $\phi$  获得的, 然后与增强节点共同构成隐藏层。

首先, 输入数据记为  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , 标签矩阵记为  $Y = \{y_1, y_2, y_3, \dots, y_n\}$ 。  $n$  组特征映射节点可以通过以下公式获得:

$$F_i = \phi(XW_{fi} + \beta_{fi}), \quad i = 1, 2, \dots, n \quad (1)$$

其中,  $W_{fi}$  和  $\beta_{fi}$  是从给定分布中随机采样的。而特征映射层可以表示为  $F^n = [F_1, F_2, \dots, F_n]$ 。其次,  $m$  组增强节点的设计如下:

$$E_j = \zeta(F^n W_{ej} + \beta_{ej}), \quad j = 1, 2, \dots, m \quad (2)$$

其中  $W_{ej}$  和  $\beta_{ej}$  是随机生成的。因此, 我们将增强层表示为  $E^m = \{E_1, E_2, \dots, E_m\}$ 。在 [4] 中, BLS 的隐藏层表示为  $[F^n | E_m]$ , 模型由下式建立:

$$Y = HW \quad (3)$$

其中  $W$  为连接隐藏层和输出的权值, 可由下式计算

$$W = H^+ Y = \lim_{\lambda \rightarrow 0} (\lambda I + H^T H)^{-1} H^T Y \quad (4)$$

需要说明的是, 隐藏层的权值是随机生成的。在某种程度上, 可能造成不确定的结果。因此, 为了降低不确定性, 我们通常设置固定的随机种子 [4]。此外, 当增加一组增强节点时, BLS 模型

可以通过增量学习进行更新。与深度网络相比，BLS 具有更快的训练速度，并且在分类和回归任务的几个经典数据集上取得了更好的性能 [11]。

## 2.2 堆叠 BLS

在这里，我们简要讨论堆叠 BLS (SBLS) [21]，它由多个基础 BLS 模块组成。在 SBLS 中，后续 BLS 模块的输入是通过前一个 BLS 模块的输入和输出之间的差值生成的残差值。最终，整个 SBLS 模型的输出是每个模块输出的总和。基于经典的残差学习理论，SBLS 的深度结构能够有效提高原始 BLS 模型的准确性。图 2 展示了具有三个模块的 SBLS 的特殊结构。下面给出了 SBLS 的具体算法。

首先，描述 SBLS 中单个模块的算法。假设一个包含  $n$  个模块的 SBLS 模型，其中对于第  $i$  个模块，我们用  $u_i$  表示输入，用  $v_i$  表示输出。单个模块的输入和输出之间的映射函数由  $f(\cdot)$  表示。因此，根据公式 (3)，第  $i$  个模块的数学表达式由以下公式表示：

$$v_i = f(u_i) = (F^n | E^m) W_i, \quad i = 1, 2, \dots, n \quad (5)$$

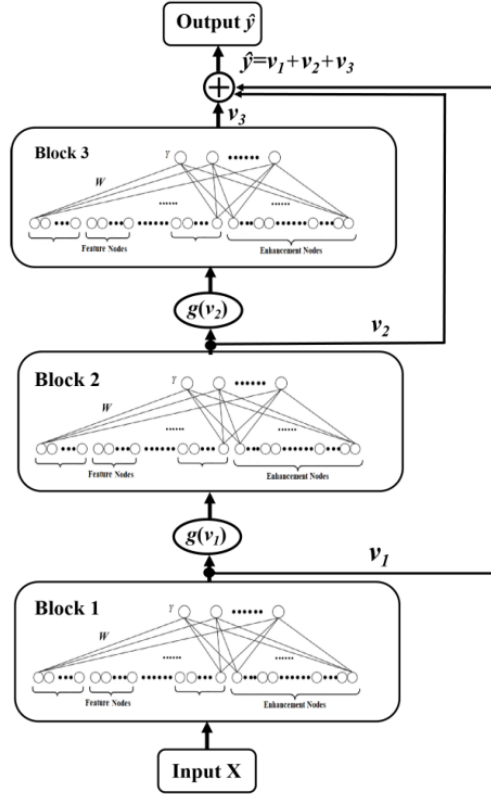


图 2: Enter Caption

其中， $F^n$  和  $E^m$  可通过公式 (1) 和 (2) 计算得到， $W_i$  是第  $i$  个模块的权重矩阵，连接了隐藏层和输出层。关键的是， $W_i$  可以通过优化以下经典的岭回归问题来获得：

$$\arg \min_{\lambda} : \|v_i - y_i\|_2^2 + \lambda \|W_i\|_2^2 \quad (6)$$

其中， $y_i$  是第  $i$  个模块的期望输出。公式 (6) 可以通过伪逆法，如公式 (4) 所示来求解。

接下来，明确 SBLS 中各模块之间的数学关系。输入  $u_i$  可以通过以下公式计算得到：

$$u_i = \begin{cases} X & \text{if } i = 1 \\ g(v_{i-1}) & \text{if } i = 2, 3, \dots, n \end{cases} \quad (7)$$

其中， $X$  是输入数据， $g(\cdot)$  是恒等函数，这意味着  $u_i = v_{i-1}$ 。实际上，你可以选择其他激活函数作为  $g(\cdot)$  来获得更好的结果。每个模块的期望输出  $y_i$  的计算公式如下：

$$u_i = y_i = \begin{cases} Y & \text{if } i = 1 \\ Y - \sum_{k=1}^{i-1} v_k & \text{if } i = 2, 3, \dots, n \end{cases} \quad (8)$$

其中， $Y$  是标签矩阵。因此，整个模型的输出表示为

$$\hat{y} = \sum_{i=1}^n v_i \quad (9)$$

到目前为止，SBLS 模型中包含  $n$  个模块的算法已经介绍完毕。此外，当准确率开始下降时，我们应该停止继续堆叠模块。

SBLS 与 ResNet 相似，构造简单，易于优化，通过堆砌一定数量的深层结构块可以提高其精度 [21]。此外，每个块都可以采用基本的 BLS 结构或其变体，与 ResNet 相比，可以更快、更有效地训练模型。

### 2.3 模糊 BLS

在本部分中，对模糊 BLS (FBLS) 做一个基础介绍 [12]。模糊 BLS 大体上保留了 BLS 的结构，但将特征节点替换为 TS 模糊子系统。最终输出通过整合来自模糊子系统和增强节点的输入生成 (见图 3)。值得注意的是，模糊 BLS 去掉了稀疏自编码器，该自编码器用于在特征层微调权重，降低了模型的复杂度。设模糊子系统的数量为  $n$ ， $m$  代表增强节点的数量。输入数据表示为

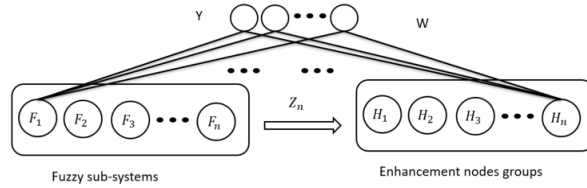


图 3: Enter Caption

$X = (x_1, x_2, \dots, x_N)^T \in R^{N \times M}$ ，其中  $x_s = (x_{s1}, x_{s2}, \dots, x_{sM})$ ,  $s = 1, 2, \dots, N$ ，输入数据通过使用一阶 Takagi-Sugeno 模糊模型映射到第  $i$  个模糊系统。假设  $x_{st}$  是 Takagi-Sugeno 模糊模型中的一阶多项式， $s = 1, 2, \dots, N$ 。它可以通过以下方式计算：

$$z_{sk}^i = \sum_{t=1}^M \alpha_{kt}^i x_{st} \quad (10)$$

其中， $\alpha_{ikt}$  是系数， $k = 1, 2, \dots, K_i$  表示第  $i$  个模糊模型中的编号。我们通过选择高斯函数来定义隶属度函数，具体形式如下：

$$\mu_{kt}^i(x_{qt}) = e^{-\left(\frac{x_{qt}^i - c_{kt}^i}{\sigma_{kt}^i}\right)^2} \quad (11)$$

其中,  $c_{ikt}$  和  $\sigma_{ikt}$  分别是高斯隶属度函数的中心和宽度。为了获得  $K_i$  个聚类中心, 采用了  $k$ -means 算法。由于  $k$ -means 算法在初始时的不确定性, 导致每个模糊子系统从输入中选择不同的中心, 这有助于提取更多特征, 因为所有模糊子系统作为一个整体工作。因此, 每条规则的加权激活水平可以按如下方式计算:

$$\omega_{sk}^i = \frac{\prod_{t=1}^M \mu_{kt}^i(x_{st})}{\sum_{k=1}^{K_i} \prod_{t=1}^M \mu_{kt}^i(x_{st})} \quad (12)$$

其中,  $\prod_{t=1}^M \mu_{kt}^i(x_{st})$  代表第  $i$  个模糊子系统中第  $k$  条模糊规则的 Fire Strength(FS)。在将它

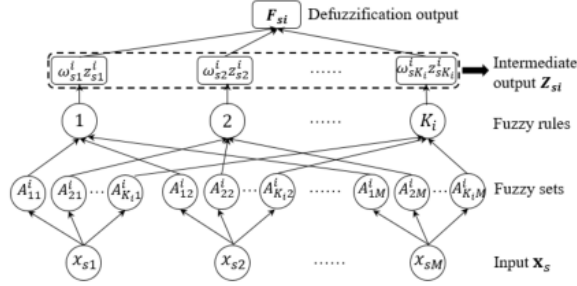


图 4: Enter Caption

们合并成一个值作为模糊子系统的去模糊化输出前, 我们在第  $i$  个模糊子系统中建立一个由每个模糊规则的输出组成的向量。然后将所有模糊子系统的中间向量提供给增强节点层, 进行额外的非线性修正。第  $i$  个模糊子系统的结构如图 4 所示。第  $i$  个模糊子系统中第  $s$  个训练样本的中间输出定义为:

$$\mathbf{Z}_{si} = (\omega_{s1}^i z_{s1}^i, \omega_{s2}^i z_{s2}^i, \dots, \omega_{sK_i}^i z_{sK_i}^i) \quad (13)$$

而第  $i$  个模糊子系统对所有训练样本  $X$  的输出矩阵表示为:

$$\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \dots, Z_{N_i})^T \in R^{N \times K_i}, \quad i = 1, 2, \dots, n. \quad (14)$$

$n$  个模糊子系统的中间输出矩阵为:

$$\mathbf{Z}_n = (Z_1, Z_2, \dots, Z_n) \in R^{N \times (K_1 + K_2 + \dots + K_n)} \quad (15)$$

然后, 增强节点对  $\mathbf{Z}_n$  进行非线性变换, 以保持输入的特性, 表示为:

$$H_j = \xi_j \left( \mathbf{Z}_n^T \mathbf{W}_{h_j} + \beta_{h_j} \right) \quad (16)$$

其中,  $H_j$  代表从  $\mathbf{Z}_i$  转换得到的增强节点。将模糊系统的输出  $\mathbf{Z}_n$  与适当的增强节点, 这些增强节点是从  $[0, 1]$  区间随机生成的 ( $j = 1, 2, \dots, m$ ), 并使用权重和偏置项  $\mathbf{W}_h^j$  和  $\beta_h^j$ 。而  $\xi_j$  是 Sigmoid 激活函数。然后, 所有增强节点的组可以表示为:

$$\mathbf{H}_m = (H_1, H_2, \dots, H_m) \quad (17)$$

第  $i$  个模糊子系统中, 第  $s$  个训练样本的输出向量为:

$$\mathbf{F}_{si} = \sum_{t=1}^M \alpha_{kt}^i x_{st} (\omega_{s1}^i, \dots, \omega_{sK_i}^i) \begin{pmatrix} \delta_{11}^i & \dots & \delta_{1C}^i \\ \vdots & & \vdots \\ \delta_{K_i1}^i & \dots & \delta_{K_iC}^i \end{pmatrix}. \quad (18)$$

接下来，第  $i$  个模糊子系统的所有训练样本的输出向量可以表示为

$$F_i = (F_{1i}, F_{2i}, \dots, F_{Ni})^T \equiv D\Omega^i\delta^i \in R^{N \times C} \quad (19)$$

其中  $D = \text{diag}\{\sum_{t=1}^M \alpha_{kt}^i x_{1t}, \dots, \sum_{t=1}^M \alpha_{kt}^i x_{Nt}\}$ ，并且

$$\Omega^i = \begin{pmatrix} \omega_{11}^i & \dots & \omega_{1K_i}^i \\ \vdots & \ddots & \vdots \\ \omega_{N1}^i & \dots & \omega_{NK_i}^i \end{pmatrix}, \quad \delta^i = \begin{pmatrix} \delta_{11}^i & \dots & \delta_{1C}^i \\ \vdots & \ddots & \vdots \\ \delta_{K_i1}^i & \dots & \delta_{K_iC}^i \end{pmatrix}.$$

模糊子系统  $r$  的组合输出的计算表示为

$$F^n = \sum_{i=1}^n F_i = \sum_{i=1}^n D\Omega^i\delta^i = D(\Omega^1, \dots, \Omega^n) \begin{pmatrix} \delta^1 \\ \vdots \\ \delta^n \end{pmatrix} \equiv D\Omega\Delta \in R^{N \times C} \quad (20)$$

其中表示 Fire Strength (FS)  $\omega_{sk}^{ij}$  的矩阵表示为  $\Omega = (\Omega^1, \dots, \Omega^n) \in R^{N \times (K_1 + K_2 + \dots + K_n)}$ ，将稍后确定的参数表示为  $\Delta = ((\delta^1)^T, \dots, (\delta^n)^T)^T \in R^{(K_1 + K_2 + \dots + K_n) \times C}$ 。

最后，FBLS 的输出记为  $Y$ ，等于

$$\hat{Y} = F^n + H^m W_e = D\Omega\Delta + H^m W_e = (D\Omega, H^m) \begin{pmatrix} \Delta \\ W_e \end{pmatrix} \equiv (D\Omega, H^m) W. \quad (21)$$

假设  $P = (D\Omega, H^m)$ ，则

$$W = (P)^+ Y. \quad (22)$$

其中，向量  $p$  可以通过公式 (4) 计算得出。至此，模糊 BLS 的过程和理论已经全部介绍完成。

FBLS 在分类和回归任务中表现良好，其精度可以通过使用 Takagi-Sugeno (TS) 模糊规则处理原始数据来提高，从而充分提取原始数据的特征 [12]。此外，FBLS 通过伪逆法计算连接特征层和增强层输出到最终输出层的权重，继承了 BLS 快速计算的特点。

### 3 极限模糊广义学习系统

#### 3.1 E-FBLS 算法

为了进一步提升性能，同时继续增加 BLS 中的神经节点，本文提出了极限模糊广义学习系统 (E-FBLS)。在这一部分中，E-FBLS 通过水平分组或合并多个 FBLS 块，并结合级联残差学习的方式构建。E-FBLS 的结构如图 6 所示。原始数据被输入到每个 FBLS 块，而不是前一个 FBLS 块的输出。不同于神经网络中在深度维度上堆叠层次，我们在更宽的维度上级联模糊 BLS 块。与 SBLS 的结构相比，E-FBLS 从宽度方向上将原始数据传递到每个基本的 FBLS 块，并同时继承了 SBLS 的带残差连接的深度扩展。此外，E-FBLS 用 FBLS 块替换了 BLS 块，以便从原始数据中提取更多特征，更好地适应模糊数据集。E-FBLS 的最终期望输出可以通过级联残差学习方法来逼近。

级联残差学习的概念可以用向量来说明，如图 5 所示。E-FBLS 模型的期望输出向量  $Y$  表示为一条直线。 $y_1$  和  $y_2$  分别是第一个和第二个块的输出。虚线表示  $Y$  和  $y_1$  的残差，这也是第二个 FBLS 块的期望输出。与  $y_1$  相比，显然最终输出  $y_1 + y_2$  更接近  $Y$ 。因此，我们可以得出结论，E-FBLS 的多块模型输出相比单个 FBLS 块更接近期望输出。因此，从数学理论的角度来看，我们提出的算法也是有效的。



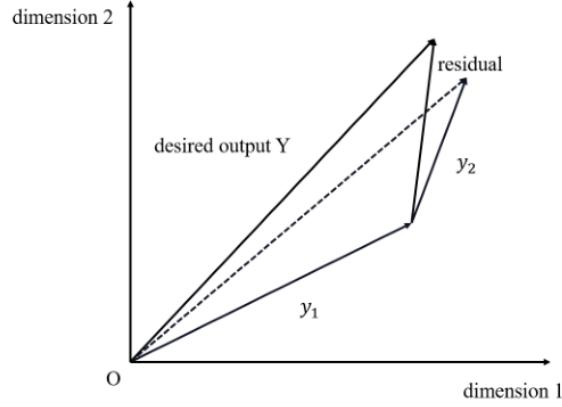


图 5: Enter Caption

E-FBLS 具体描述如下。首先描述单个 FBLS 块的结构。输入数据和相应的标签分别表示为  $X = \{x_1, x_2, \dots, x_n\}$  和  $Y = \{y_1, y_2, \dots, y_n\}$ 。我们假设一个具有  $r$  个块的 E-FBLS 模型，其中第  $i$  个块的输入和输出分别表示为  $p_i$  和  $q_i$ 。这里  $i = 1, 2, \dots, r$ 。此外，隐藏层表示为  $D_i = [F^n | E^m]$ ，其中  $F^n = [F_1, \dots, F_n]$  和  $E^m = [E_1, \dots, E_m]$ 。  $F_j$  和  $E_k$  可以通过公式 [35] 和 [8] 计算得到。第  $i$  个 FBLS 块的输出计算如下：

$$q_i = D_i W_i \quad (23)$$

其中  $W_i$  是连接隐藏层和输出层的权重，可以通过伪逆计算如下：

$$W_i = D^+ y_i = \lim_{\lambda \rightarrow 0} (\lambda I + D^T D)^{-1} D^T Y \quad (24)$$

其中  $y_i$  是第  $i$  个块的期望输出，稍后将详细解释。

其次，如何级联多个 FBLS 块也进行了详细说明。我们假设第  $i$  个块的输入和输出之间的映射函数为  $q_i = G_i(p_i)$ 。图 6 显示了具有  $r$  个块的 E-FBLS 模型的结构，由此可以看出，E-FBLS 可以视为一种块级增量学习方法，因为每个 FBLS 块的输入  $p_i = X$  是恒定的。第  $i$  个块的期望输出可以通过 (12) 计算得到

$$y_i = \begin{cases} Y & \text{if } i = 1 \\ Y - \sum_{k=1}^{i-1} q_k & \text{if } i = 2, 3, \dots, r \end{cases} \quad (25)$$

其中  $\sum_{k=1}^{i-1} q_k = \sum_{k=1}^{i-1} G_k(p_k)$ 。  $y_i$  是第  $(i-1)$  个块的残差，由  $(y_{i-1} - q_{i-1})$  计算得到。因此，具有  $r$  个块的整个模型的最终输出表示为

$$\hat{Y} = \sum_{i=1}^r q_i \quad (26)$$

并且

$$q_i = G(p_i) = D_i W_i. \quad (27)$$

为了确定 FBLS 块的数量，我们设置一个条件，如式 (28) 所示。当该条件满足时，FBLS 块的数量将不再增加，从而得到 E-FBLS 中的 FBLS 块数量：

$$\|y_i\|_2^2 \leq \delta \quad (28)$$

其中,  $\delta$  是允许的误差。

在某种程度上, E-FBLS 也可以看作是 FBLS 在块级别上的广义版本。随着 E-FBLS 中 FBLS 块数量的增加, E-FBLS 的性能将逐渐提高。E-FBLS 的整个过程总结如算法 1 所示。

---

**Algorithm 1** E-FBLS

---

**Input:** dataset input  $X$ , the number of blocks of E-FBLS  $r$ , the permitted error  $\delta$

**Output:** the output  $\hat{Y}$  of E-FBLS model

---

```

1: for  $i = 1$  to  $r$  do
2:   Randomly generate  $W_{hi}, \beta_{hi}$ 
3:   Calculate  $F_n$  and  $E_m$  by (18)(16) with  $p_i = X$ 
4:   Calculate  $y_i$  by (25)
5:   Calculate  $W_i$  by (24)
6:   Calculate  $q_i$  by (27)
7:   if  $\|y_i\|_2^2 \leq \delta$  then
8:     break
9:   end if
10: end for
    Calculate  $\hat{Y} = \sum_{i=1}^r q_i$ 

```

---

### 3.2 E-FBLS 中的频率原则

E-FBLS 中的频率原则主要是分析神经网络在对输入数据和模型输出应用傅里叶分析时的表现。在 DNN 和 BLS 中存在频率原则, E-FBLS 中也同样存在这一原则, 即 E-FBLS 可以优先快速捕捉低频成分, 并在增加 FBLS 块数量的扩展过程中逐渐拟合高频成分。为了证明 E-FBLS 可

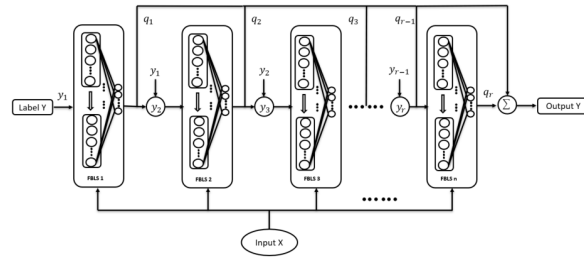


图 6: Enter Caption

以有效地学习提取低频成分并更好地拟合高频成分的频率原则, 我们需要首先理解 Xu 等人 [30] 提出的频率原则的概念。根据 [30], 频率是指响应频率。

E-FBLS 中的频率原则详细介绍如下。对于一个训练数据集  $(X, Y)$ , 我们在计算主成分之前, 对  $X$  应用如下变换 (见式 (29)):

$$x'_i = x_i - \frac{1}{n} \sum_{j=1}^n x_j, \quad i = 1, 2, \dots, n. \quad (29)$$

然后我们将  $X' = \{x'_1, x'_2, \dots, x'_n\}$  表示为用于计算主成分的新数据。 $x_{p1}$  是沿着  $p1$  方向投影后的  $X'$ , 其计算公式为 (17):

Symbol	Variable
$\bar{X}$	Input training data
$Y$	Input training targets
$F$	Output matrix of fuzzy subsystems
$E$	Output matrix of enhancement layer
$W$	Parameter matrix
$a$	Random weight matrix in fuzzy subsystems
$\mu$	Membership function
$w$	Weighted fire strength
$c$	Center of Gaussian membership function
$\sigma$	Width of Gaussian membership function
$q_i$	Output of the $i$ -th block matrix
$\ddot{\delta}$	Permitted error
$N_r$	The number of fuzzy rules per group
$N_t$	The number of fuzzy systems
$N_c$	The number of enhancement nodes
$\frac{1}{2} = \frac{1}{2} = \frac{1}{2}$	Trade-off parameter to $\ \mathbf{W}\ ^2$

表 1: Variables and Symbols with Descriptions

$$x_{p1} = X'p1 \quad (30)$$

其中,  $p1$  是通过计算协方差矩阵  $X'X'^T$  对应最大特征值的特征向量得到的第一个主成分向量 [29]。

为了得到第一个主成分, 我们对数据集  $Y$  的目标进行傅里叶变换。为简化说明, 假设只考虑  $Y$  的第一维, 并将其表示为  $Y^{(1)}$  [6]。然后我们可以计算:

$$F(Y^{(1)})(\gamma) = \sum_{i=0}^{n-1} Y_i^1 e^{-2\pi j \gamma x_{p1}} \quad (31)$$

其中,  $\gamma$  是频率索引,  $\gamma \in Z$ 。而  $F$  是非均匀快速傅里叶变换 (NUFFT)。我们可以得到频域中目标函数的表达式, 表示为  $\mathcal{F} = \{F(Y^{(1)})(1), F(Y^{(1)})(2), \dots, F(Y^{(1)})(\gamma)\}$ , 其中  $\gamma \in [1, 50]$ 。类似地,  $\hat{F}$  表示频域中拟合函数的表达式, 其等于  $\{\hat{F}(\hat{Y}^{(1)})(1), \hat{F}(\hat{Y}^{(1)})(2), \dots, \hat{F}(\hat{Y}^{(1)})(\gamma)\}$ , 其中  $\hat{Y}^{(1)}$  是 (26) 中的第一维。

为了检验  $\mathcal{F}$  和  $\hat{F}$  之间的差异, 我们通过以下公式计算每个频率峰值索引的相对误差 [31]:

$$\Delta\mathcal{F}(\gamma_0) = \frac{|\mathcal{F}(\gamma_0) - \hat{F}(\gamma_0)|}{|\mathcal{F}(\gamma_0)|} \quad (32)$$

其中,  $\gamma_0$  是从  $\gamma$  中选择的频率峰值索引集合。

通过上述过程, 我们可以获得在频域中, FBLS 块输出与对应目标函数之间的相对误差。为了验证 E-FBLS 中频率原理的存在性, 我们在添加一个 FBLS 块后计算每个频率峰值索引的误差。随着块数增加, 相对误差的变化反映了 E-FBLS 对不同频率的响应特性。

## 4 实验

在本节中，进行实验以验证 E-FBLS 的性能。与几种基于经典分类和回归数据集的机器学习算法进行比较。此外，还验证了 E-FBLS 中频率原理的存在。本文中的所有实验均在一台配备 1.6 GHz Intel i5-8265U CPU 的笔记本电脑上进行。

### 4.1 分类数据集实验

在本节中，针对三个经典分类数据集 [3]（如乳腺癌威斯康星（WBC）、乳腺癌威斯康星诊断（WDBC）和葡萄酒数据集（WINE））进行了多分类实验。

为了测试 E-FBLS 模型的性能，我们选择了传统的机器学习方法和相关的 BLS 级联变体进行比较。为了展示 E-FBLS 的有效性，我们选择了支持向量机（SVM）进行比较，[27]，它是最传统且具有代表性的机器学习方法之一。此外，极限学习机（ELM）[14] 具有与 BLS 相似的扁平结构，并且在分类和回归任务中表现良好，具有代表性，因此被选中进行比较。FBLS 和堆叠 BLS 都是 BLS 的变体。此外，堆叠 BLS 是一种基于 BLS 的神经网络，具有典型的深度级联结构。因此，我们选择 FBLS 和堆叠 BLS 进行比较。此外，BLS、EBLS 和 FBLS 也被选为消融实验。由于 BLS、FBLS 和 SBLS 方法已在文献 [4, 12, 21] 中验证了它们相对于深度学习方法的优越性，因此我们在实验中省略了与深度学习方法的比较。

Datasets	No. of samples		Attributes	Classes
	Training	Testing		
WBC	350	349	9	2
WDBC	393	276	30	2
WINE	112	67	13	3

表 2: Dataset details

在实验中，ELM 的所有激活函数设置为 sigmoid。支持向量机（SVM）选择径向基函数（RBF）[25] 作为核函数。WBC、WDBC 和 WINE 数据集简要介绍如下，数据集的详细信息列在表 I 中。

1) WBC 数据集：WBC 原始数据集 [23]，该数据集包含乳腺癌样本，由威斯康星医学院的 William H.W. Olberg 博士提供。所有数据来自临床病例，每个病例有 10 个属性。前九个属性是测试指标，每个属性值为 1 到 10 之间的整数，其中 1 表示最正常的测试指标，10 表示最异常的。第十个是分类属性，表示肿瘤是否是恶性的。WBC 是机器学习中的经典分类数据集，因此通过 WBC 数据集展示 E-FBLS 的性能是可行的。

2) WDBC 数据集：WDBC 数据集 [23]。WDBC 数据集中的样本分为两组：1) 具有良性症状的患者，2) 恶性诊断的患者。

3) WINE 数据集：WINE 数据集包含来自意大利同一地区三种不同葡萄酒品种的化学分析结果。与 WBC 和 WDBC 这样的二元数据集不同，这是一个多类数据集。

在 E-FBLS 中，我们将每组模糊规则的数量表示为  $N_r$ ，模糊系统组的数量表示为  $N_t$ ，增强节点的数量表示为  $N_e$ 。为了在该数据集上获得 E-FBLS 模型的最佳结果，我们调整每个块的  $N_r$ 、 $N_t$  和  $N_e$ ，以更好地表示数据特征。

实验设置如下：ELM 的隐藏节点数设置为 40，激活函数选择 tansig。对于 SVM，选择 RBF 作为其核函数，并将正则化参数  $C$  固定为 1。

BLS 的三个参数  $N_f$ 、 $N_g$ 、 $N_e$  设置为 (10, 10, 500)。同样地，SBLS 的参数  $N_f$ 、 $N_w$  和  $N_e$  设置为 (10, 10, 3000)。基于训练好的 FBLS 作为第一个模块，我们为 FBLS 堆叠了两个模块，其参

Method	Testing Accuracy	Training Time (s)
ELM	93.41%	0.934
SVM	91.80%	0.036
BLS	85.39%	0.010
SBLS	85.87%	0.059
EBLS	86.53%	0.132
FBLS	97.99%	0.088
E-FBLS (Ours)	98.57%	0.583

表 3: Performance Comparison of Different Methods

Method	Testing Accuracy	Training Time (s)
ELM	92.87%	0.153
SVM	90.50%	0.037
BLS	93.48%	0.003
SBLS	93.53%	0.082
EBLS	93.84%	0.117
FBLS	94.20%	0.052
E-FBLS (Ours)	94.93%	0.183

表 4: Classification Accuracy on WDBC Dataset

数  $N_r$ 、 $N_t$  和  $N_e$  的搜索范围分别为  $[1, 5]$ 、 $[1, 20]$  和  $[1, 50]$ 。

WBC、WDBC 和 WINE 数据集上各种方法的分类结果见表 3–5。从这三个表格中可以看出，尽管 E-FBLS 的训练速度较慢，但在测试准确率方面，E-FBLS 的性能优于其他方法。与 SBLS 相比，E-FBLS 训练时间翻倍的主要原因可能是 E-FBLS 需要更多的隐藏层节点来学习特征。此外，与 WDBC 数据集上的分类结果相比，所有方法在 WBC 数据集上的训练时间更长，主要是因为该数据集更具挑战性。为了进一步展示 E-FBLS 的优势，表 4 中 E-FBLS 模型的最终结果表明，通过增加 E-FBLS 中 FBLS 模块的数量，分类性能可以进一步提高。

## 4.2 回归数据集上的实验

在这一部分中，我们对 Abalone、House-value 和 Forest-fires 三个回归数据集进行了实验，具体介绍如下：

1) Abalone：根据七种不同特征，通过物理测量预测鲍鱼的年龄。所有样本分为两个集合，2302 个样本用于训练，1880 个样本用于测试。

2) House-Value：该数据集收集了房屋年龄、与最近地铁站的距离、步行范围内的便利店数量及地理位置等信息，以预测单位面积的房价。选择 201 个样本作为训练数据，213 个样本作为测试数据。

3) Forest-Fires：该数据集包含了葡萄牙蒙特西地区的 517 次自然公园火灾事件，记录了每次事件的日期、月份、坐标、燃烧面积，以及降雨量、温度、湿度、风速等气象数据，以预测火灾面积。该数据集包括 517 个样本，其中 272 个样本用于训练，270 个样本用于测试。

为了展示回归结果，我们仅设置三个 BLS 模块在 E-FBLS 中训练，并绘制了测试数据的残差 [2]，并使用决定系数  $R^2$  作为指标。 $R^2$  的值介于 0 和 1 之间，可以更好地评估模型。简而言之，

Method	Testing Accuracy	Training Time (s)
ELM	93.63%	0.101
SVM	94.27%	0.015
BLS	94.03%	0.025
SBLS	94.86%	0.068
EBLS	95.52%	0.114
FBLS	97.01%	0.066
E-FBLS (Ours)	98.51%	0.180

表 5: Classification Accuracy on WINE Dataset

表 6: Classification Results of E-FBLS on WBC, WDBC and WINE

Dataset	No. of FBLS blocks	Testing Accuracy	$N_r$	$N_t$	$N_e$	Training Time(s)
WBC	1	97.99%	2	3	5	0.086
	2	98.28%	2	8	20	0.256
	3	98.57%	2	10	30	0.583
WDBC	1	94.20%	2	3	10	0.052
	2	94.57%	2	5	15	0.119
	3	94.93%	2	6	20	0.182
WINE	1	97.01%	2	6	15	0.048
	2	97.88%	2	15	15	0.112
	3	98.51%	2	6	20	0.163

模型的预测效果越好，其  $-R^2$  值越接近 1。三组数据集上的残差图和实验结果见图 7 和表 7-9。从图 7 可以看出，几乎所有的蓝点都在红线之间，表明预测值接近真实值。观察实验结果，我们可以得出结论：与其他六种机器学习算法相比，E-FBLS 的表现更为优异。

### 4.3 EBLS 中频率原则的验证

在本部分中，通过对第一个主成分的傅里叶分析，证明了 E-FBLS 中频率原则的存在。我们计算了训练好的 E-FBLS 模型与上述六个数据集的目标函数之间的相对误差 [26]。为了便于说明，在每个频率峰值索引处的结果被组合为一个列向量，每次添加一个 BLS 模块时计算，并将相对差异矩阵显示为伪彩色图 [6]。

为了计算每个频率索引的频率幅度，我们在  $[1, 50]$  的范围内搜索合适的  $\gamma$  值。由于我们为 E-FBLS 设置了三个模块以适应数据集，我们需要计算三次  $F$  以获得相对误差的变化趋势。六个数据集的相对误差结果如图 8 所示，图中显示了随着模块数量的增加，每个索引的误差总体上呈下降趋势。对于特定的频率峰值索引，误差越小，拟合函数对该频率成分的逼近程度越高。换句话说，

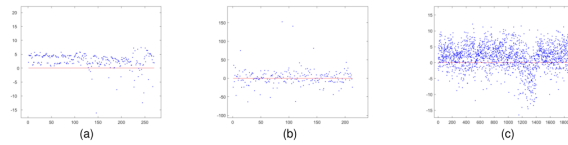


图 7: Enter Caption

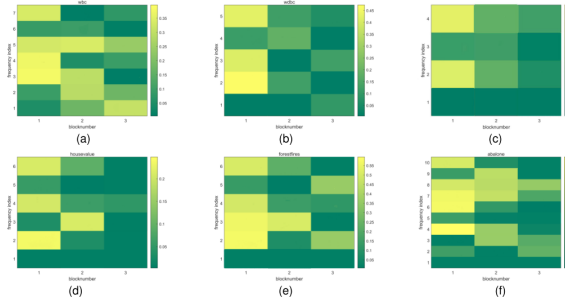


图 8: Enter Caption

模型在最低误差的成分上被训练至期望状态，而不是在其他成分上。

详细讨论中需指出，E-FBLS 对不同频率成分具有不同的响应能力。图中显示，低峰值索引的色块（底部的一到两行）比其他色块绿得更深，这表明在低索引处误差较小，尤其是在图 8(c)–(f) 中的  $y$ 。因此，可以得出结论：E-FBLS 在训练开始时能够更快更好地拟合低频成分。此外，高频值索引的色块在开始时较黄，随着模块的增加逐渐变暗。这反映了尽管 E-FBLS 在开始时难以检测到高频成分，但可以通过添加模块逐渐进行匹配。

综上所述，E-FBLS 能够快速捕捉低频成分，同时随着 FBLS 模块的增加逐渐拟合高频成分。因此，我们可以得出结论：频率原则在 E-FBLS 模型中确实存在。E-FBLS 中频率原则的存在证明也为神经网络在频率域的可解释性提供了支持。从频率成分的实际意义上解读频率原则也具有意义。低频成分代表数据的框架，较为稳定；而高频成分是数据变化较大的部分，对输出影响较大。因此，E-FBLS 模型更容易学习稳定的部分，并逐渐拟合变化较大的部分。E-FBLS 中的频率原则可以有效用于求解泊松方程和其他任务。

## 5 讨论

与不同的 BLS 级联变体相比，我们的研究具有几个重要的区别，这些区别为其新颖性做出了贡献。

**方法差异：**我们选择了一种不同的级联模式，以更广泛的方式从原始数据中提取更多信息，同时在 E-FBLS 中保持深度方向。原始数据被输入到每个 FBLS 模块中，而不是前一个 FBLS 模块的输出。

**理论支持：**我们进一步验证了 E-FBLS 中频率原则的存在，该原则通常被忽视，并为神经网络的泛化提供了良好的可解释性。

**性能提升：**通过进行回归和分类实验，E-FBLS 在六种最先进的模型中表现更优，并且与其他深度学习方法相比，保持了 BLS 的高速性。

尽管如此，我们的研究仍存在一些局限性。提出的 E-FBLS 方法在数据爆炸发生时难以处理高维数据。E-FBLS 方法中采用的 K-means 算法无法实现良好的聚类效果。如我们所知，如果使用 K-means 算法对高维数据进行聚类，所有点之间的距离会相似，“最近点”的概念将消失，因此可能没有意义 [1]。此外，对于大规模数据，训练所提出的 E-FBLS 模型的耗时将非常长。这是因为当大规模数据被直接输入到多个 FBLS 模块中时，需要更多的模糊规则和子系统来捕捉输入的特征。

在未来，我们将把一种先进的聚类算法整合到提出的 E-FBLS 方法中，从而能够处理高维数据。此外，更好的大规模数据特征提取方法也是另一个研究方向。

## 6 结论

本文提出了一种基于 FBLS 和级联残差学习的新型级联广义神经网络 E-FBLS。E-FBLS 高效，因为当添加新的 FBLS 模块时，较低的 FBLS 模块架构保持不变，并保留了 FBLS 的优点。由于其相比 SBLS 更广的结构，E-FBLS 能够实现更好的特征表示。此外，在分类和回归数据集上的实验结果验证了 E-FBLS 相比于其他六种经典机器学习算法（如 SVM 和 ELM）[15] 的优越性。此外，E-FBLS 中频率原则的存在得到了证实。E-FBLS 能够快速学习低频成分，而随着模块数量的增加，高频成分逐渐适应。E-FBLS 中存在的频率原则不仅为神经网络的泛化提供了良好的可解释性，而且对神经网络应用的扩展具有重要的启发意义。



## 7 参考文献

### 参考文献

- [1] C. C. Aggarwal, A. Hinneburg, and D. A. Keim. On the surprising behavior of distance metrics in high dimensional space. In *Proc. 8th Int. Conf. Database Theory*, pages 420–434, London, U.K. Berlin, Germany, 2001. Springer.
- [2] T. Chau and A. K. C. Wong. Pattern discovery by residual analysis and recursive partitioning. *IEEE Trans. Knowl. Data Eng.*, 11(6):833–852, Nov./Dec. 1999.
- [3] B. B. Chaudhuri and U. Bhattacharya. Efficient training and improved performance of multi-layer perceptron in pattern classification. *Neurocomputing*, 34(1–4):11–27, Sep. 2000.
- [4] C. L. P. Chen and Z. Liu. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Trans. Neural Netw. Learn. Syst.*, 29(1):10–24, Jan. 2018.
- [5] C. L. P. Chen, Z. Liu, and S. Feng. Universal approximation capability of broad learning system and its structural variations. *IEEE Trans. Neural Netw. Learn. Syst.*, 30(4):1191–1204, Apr. 2019.
- [6] G.-Y. Chen, M. Gan, C. L. P. Chen, H.-T. Zhu, and L. Chen. Frequency principle in broad learning system. *IEEE Trans. Neural Netw. Learn. Syst.*, 33(11):6983–6989, Nov. 2022.
- [7] H. Chen and B. Jiang. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.*, 21(2):450–465, Feb. 2020.
- [8] F. Chu, T. Liang, C. L. P. Chen, X. Wang, and X. Ma. Weighted broad learning system and its application in nonlinear industrial process modeling. *IEEE Trans. Neural Netw. Learn. Syst.*, 31(8):3017–3031, Aug. 2020.
- [9] J. Du, C.-M. Vong, and C. L. P. Chen. Novel efficient rnn and lstm-like architectures: Recurrent and gated broad learning systems and their applications for text classification. *IEEE Trans. Cybern.*, 51(3):1586–1597, Mar. 2021.
- [10] A. Vaswani et al. Attention is all you need. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 30, pages 1–11, 2017.
- [11] L. Zhang et al. Analysis and variants of broad learning system. *IEEE Trans. Syst., Man, Cybern., Syst.*, 52(1):334–344, Jan. 2022.
- [12] S. Feng and C. L. P. Chen. Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification. *IEEE Trans. Cybern.*, 50(2):414–424, Feb. 2020.
- [13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 770–778, Jun. 2016.
- [14] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, 42(2):513–529, Apr. 2012.

- [15] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3):489–501, Dec. 2006.
- [16] Z. Huang and J. Duan. Gfbls: Graph-regularized fuzzy broad learning system for detection of interictal epileptic discharges. *Eng. Appl. Artif. Intell.*, 125, Oct. 2023. Art. no. 106763.
- [17] J. Jin, Z. Liu, and C. L. P. Chen. Discriminative graph regularized broad learning system for image recognition. *Sci. China Inf. Sci.*, 61(11):1–14, Nov. 2018.
- [18] Y. Kong, X. Wang, Y. Cheng, and C. Chen. Hyperspectral imagery classification based on semi-supervised broad learning system. *Remote Sens.*, 10(5):685, Apr. 2018.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 25, pages 1–9, 2012.
- [20] L. Liu, L. Cai, T. Xie, and Y. Wang. Self-paced broad learning system. *IEEE Trans. Cybern.*, 53(6):4029–4042, Jun. 2022.
- [21] Z. Liu, C. L. P. Chen, S. Feng, Q. Feng, and T. Zhang. Stacked broad learning system: From incremental flattened structure to deep model. *IEEE Trans. Syst., Man, Cybern., Syst.*, 51(1):209–222, Jan. 2021.
- [22] Z. Liu, J. Zhou, and C. L. P. Chen. Broad learning system: Feature extraction based on k-means clustering algorithm. In *Proc. 4th Int. Conf. Inf., Cybern. Comput. Social Syst. (ICCSS)*, pages 683–687, Jul. 2017.
- [23] C. B. D. Newman. Uci repository of machine learning databases, 1998. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [24] Y.-H. Pao, G.-H. Park, and D. J. Sobajic. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing*, 6(2):163–180, Apr. 1994.
- [25] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural Comput.*, 3(2):246–257, Jun. 1991.
- [26] S. Ruder. An overview of gradient descent optimization algorithms, 2016. arXiv:1609.04747.
- [27] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Process. Lett.*, 9(3):293–300, Jun. 1999.
- [28] M. Xu, M. Han, C. L. P. Chen, and T. Qiu. Recurrent broad learning systems for time series prediction. *IEEE Trans. Cybern.*, 50(4):1405–1417, Apr. 2020.
- [29] Z.-Q. J. Xu. Frequency principle in deep learning with general loss functions and its potential application. 2018. arXiv:1811.10146.
- [30] Z.-Q. J. Xu, Y. Zhang, T. Luo, Y. Xiao, and Z. Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. 2019. arXiv:1901.06523.
- [31] Z.-Q. J. Xu, Y. Zhang, and Y. Xiao. Training behavior of deep neural network in frequency domain. In *Proc. 26th Int. Conf. Neural Inf. Process.*, pages 264–274, Sydney, NSW, Australia, 2019. Springer.

- [32] H. Ye, H. Li, and C. L. P. Chen. Adaptive deep cascade broad learning system and its application in image denoising. *IEEE Trans. Cybern.*, 51(9):4450–4463, Sep. 2021.
- [33] W. Zaremba, I. Sutskever, and O. Vinyals. Recurrent neural network regularization, 2014. arXiv:1409.2329.
- [34] H. Zhao, J. Zheng, W. Deng, and Y. Song. Semi-supervised broad learning system based on manifold regularization and broad network. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 67(3):983–994, Mar. 2020.
- [35] Y. Zheng, X. Qin, Z. Xi, and B. Chen. Mixed-norm based broad learning system for eeg classification. In *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pages 7092–7095, Jul. 2019.