将图神经网络应用于数值数据的路线图：来自胶凝材料的见解 Mahmuda Sharmin[1], Taihao Han[1*], Jie Huang[1], Narayanan Neithalath[2], Gaurav Sant[3] Aditya Kumar[1]

1. 密苏里科技大学，罗拉，密苏里州，美国 65409 2. 亚利桑那州立大学可持续工程与建筑环境学院，坦佩，亚利桑那州，美国 85287

3. 加利福尼亚大学洛杉矶分校土木与环境工程系，洛杉矶，加利福尼亚州，美国 90095

***通讯作者**

Taihao Han

助理研究教授，密苏里科技大学材料科学与工程系，麦克纳特楼248A室，
地址：密苏里州罗拉市北毕晓普街1400号，邮编65409；电子邮箱：thy3b@mst.edu；电话：573-341-6051；传真：573-341-6934

摘要

机器学习（ML）已日益应用于混凝土研究，以优化性能和配合比设计。然而，将机器学习应用于胶凝材料面临的一项主要挑战，是现有数据库的规模有限且多样性不足。一种有前景的解决方案是开发多模态数据库，整合数值数据和图形数据。水泥研究中的传统机器学习框架通常仅限于单一数据模态。图神经网络（GNN）代表了新一代神经网络架构，能够从以图形式构建的数据中学习——通过不规则或依赖拓扑结构的连接来捕捉关系，而非依赖固定的空间坐标。图神经网络虽专为图形数据设计，但也可用于从数值数据集中提取相关性，并有望将物理定律直接嵌入其架构，从而实现可解释且受物理信息指导的预测。本研究是最早将图神经网络应用于混凝土设计的研究之一，尤其着重于建立一条清晰且可复现的路径，利用K近邻（K-NN）方法将表格数据转换为图表示。我们系统地优化模型超参数并进行特征选择，以提升预测性能。图神经网络的表现与基准随机森林相当；多项研究已证明，基准随机森林能够对胶凝材料作出可靠预测。总体而言，本研究为从传统机器学习转向先进人工智能架构提供了基础路线图。所提出的框架为未来多模态、物理信息驱动的图神经网络模型奠定了坚实基础，使其能够捕捉复杂的材料行为，并加速胶凝材料的设计与优化。

关键词：图神经网络；数值数据；混凝土；抗压强度；K近邻

1.0. 引言

水泥是一种多组分的复杂材料，主要由阿利特、贝利特、铝酸三钙和铁铝酸四钙组成。即使这些相的比例只有细微变化，也会对最终性能产生深远影响。此外，原材料的质量——尤其是来自不同地区的矿物——会导致不同生产批次和制造地点之间存在差异，进而造成性能不一致。尽管水泥从古代至今一直得到应用，但我们对其基本性质的科学认识仍不完整。例如，尽管已有大量研究，水泥水化的控制机制仍存在争议 [1,2]。此外，水泥组成与性能之间的关系高度非线性，这进一步增加了理解和研究水泥化学的复杂性。尽管面临这些挑战，研究人员仍尝试开发分析模型，以建立水泥组成与其性能之间的关联 [3–6]。近年来，辅助胶凝材料（SCMs）已广泛用于部分替代混凝土中的硅酸盐水泥，以降低成本和环境影响。一些辅助胶凝材料（例如石英和石灰石）大多呈惰性，主要作为水化产物的额外成核位点。它们对胶凝材料性能的影响相对简单。然而，许多其他材料（例如粉煤灰和矿渣）会与水泥发生积极反应。它们的化学组成和结晶度因来源而有显著差异，导致反应活性和反应机制（例如火山灰反应和硫酸盐缓凝）存在较大差异。因此，开发能够准确捕捉水泥、辅助胶凝材料及其他参数（例如养护条件）之间复杂相互作用的分析模型，仍是水泥科学面临的一项重大挑战。

为应对这一挑战，研究人员广泛采用机器学习（ML）作为预测和优化胶凝材料性能的强大工具。机器学习提供了一种数据驱动的方法，可替代传统分析模型，无需显式的机理方程，即可发现所有影响参数与性能之间复杂的潜在相关性。大量研究已证明，机器学习能够以配合比设计为输入，准确预测力学性能 [7–9], 流变行为 [10–12], 水化动力学 [13–15], 以及耐久性 [16–18] 。在

此基础上，研究人员近期还探索了将微观结构 [19–21] 和光谱 [22,23] 成像数据用作机器学习输入，以预测混凝土性能。机器学习从基于图像的数据中学习的能力，为将微观结构与宏观行为联系起来开辟了新途径，并提供了难以通过传统统计或分析方法获得的见解。尽管取得了这些进展，机器学习应用于胶凝材料时面临的主要挑战之一仍是数据库规模有限。获取一个实验数据点可能需要大量实验室工作，通常涉及材料制备、养护、测试和数据处理等多个阶段，因此汇集包含数千条记录的数据集既耗时又昂贵。由此，许多现有机器学习模型都是基于规模较小的同质数据集训练的，这限制了它们的泛化能力和鲁棒性。克服这一局限的一项可行策略是开发多模态数据库，整合数值数据和基于图像的数据。通过系统地收集并整合已发表文献中的数据，可以构建此类数据库，从而广泛涵盖材料、组成和测试条件。整合异构数据源不仅能扩大数据库规模并增加数据变异性，还能更全面地表示胶凝材料体系。开发多模态数据库的另一项优势在于，数值数据和图形数据能够从不同角度呈现同一性能。数值数据可捕捉配合比设计与加工条件之间的定量关系，而图形数据则反映潜在的微观结构和形貌。整合这些不同但相关的信息形式，有助于提升机器学习模型的泛化能力和可靠性。然而，目前用于水泥研究的机器学习框架通常仅针对单一数据模态进行设计——要么处理数值数据，要么处理基于图像的输入——因此难以充分利用不同数据类型所蕴含的互补信息。因此，亟需开发能够同时从数值数据和图形数据中学习的先进多模态机器学习架构。

图神经网络（GNN）是一类新型神经网络，专门用于从以图结构表示的数据中提取并学习特征 [24–26]。与在欧几里得空间中运行的传统卷积神经网络不同，后者的数据点（或神经元）排列在规则网格上，且间距

均匀，例如图像中的像素；GNN 则在非欧几里得空间中运行，其中数据元素之间的关系由连接节点的边定义，而非由空间邻近性决定。这使 GNN 能够发现潜在相关性：元素之间的相互作用通过复杂、不规则或依赖拓扑结构的连接发生，而不是由固定的空间坐标决定。从本质上说，GNN 将卷积的概念从欧几里得域推广到不规则图域，从而能够从分子网络、社交网络或材料微观结构等数据结构中提取特征。图中的每个节点代表一个实体（例如，原子、矿物相或微观结构的体素），边则定义它们在物理、化学或空间上的关系。通过迭代执行消息传递操作，每个节点聚合来自其邻居节点和边的信息，以更新自身的表示，使网络能够学习数据中的潜在相关性和层级依赖关系。这种架构使 GNN 尤其适用于通过学习分子结构、微观结构和分子动力学中的物理与化学信息来预测材料性能 [27–30]。GNN 模型已用于优化混凝土结构 [31,32] 并预测水泥制造的能耗 [33]。然而，GNN 模型尚未用于关联胶凝材料的组成与性能。它在将混凝土的微观结构转换为图表示方面具有很大潜力，其中节点表示单个相、孔隙或特征，边则编码它们的几何或物理化学相互作用。通过从这类结构化表示中学习，GNN 能够有效捕捉局部连通模式、长程依赖关系和复杂的拓扑关系，而这些往往是传统机器学习模型所忽略的。

近期研究表明，图神经网络并不局限于图形数据，也能有效适配表格数据库并从中学习。多位研究人员提出了创新策略，将表格数据转换为图结构，使图神经网络能够在关系语境中处理和解读数值特征。例如，Li等人 [34] 提出了多种图构建方法（即将数据作为节点、将特征作为节点，以及将数据和特征都作为节点），以将表格数据集转换为图表示。Hong等人 [35] 开发了一种基于图神经网络的架构，旨在实现信息交换，其中每个节点都会聚合多个数值特征的信息。信息量最大的节点

嵌入随后被传递至后续神经网络层，用于预测材料的熔化温度。Guo 等 [36] 构建了一种多重图，用于建模样本之间的多方面关系，并设计了一种多重图神经网络，使其能够通过跨层交互学习增强的高维特征表示。然而，如何将图神经网络应用于数值数据仍缺乏明确的路径，尤其是在材料科学和水泥化学领域。

如前所述，要提高胶凝材料数据的质量和多样性，需要构建整合数值数据与图形数据的多模态数据库。在此背景下，图神经网络是学习复杂相关性并在此类异构数据域中生成可靠预测的理想方案。图神经网络的另一项优势在于，它们在定义节点、边和连接规则方面具有内在的灵活性，因此能够将物理定律和领域知识明确融入模型。例如，可以在不同层级定义节点，并依据具有物理意义的相互作用有选择地建立边。此外，还可以针对特定边设计定制的激活函数，以反映已知相关性，确保模型预测符合物理规律。将物理原理直接嵌入网络架构的能力，使图神经网络成为开发下一代可解释物理信息模型、推动胶凝材料研究发展的变革性工具。

本研究是较早应用图神经网络来学习组成-性能关联并设计混凝土的研究之一，重点探索在表格数据库中应用图神经网络的策略。尽管本研究目前仅聚焦于表格数据，但它标志着朝着开发多模态（即数值和图形）框架迈出了重要一步。为将表格输入转换为图结构，本研究采用K近邻模型将表格数据记录映射到图空间，从而定义图神经网络的节点和边。研究系统地确定了K近邻的最优邻居数量。此外，本研究开展特征重要性分析，以优化输入选择，并评估特征修改对图神经网络性能的影响。研究将图神经网络模型的预测性能与基准随机森林模型进行比较。总体而言，本研究建立了一个

基础框架，推动从传统机器学习向先进人工智能模型转型，从而实现准确预测，并为胶凝材料构建具有物理意义的图神经网络架构。

2.0. 图神经网络（GNN）

图神经网络模型是一种机器学习模型，旨在从图结构数据中学习相关性。图神经网络模型可以直接处理实际应用中有用的图类型，例如循环图、无环图、有向图和无向图。与多层神经网络、卷积神经网络或循环神经网络等传统神经网络不同，这些网络处理的是固定大小的向量或网格输入，[37], 图神经网络旨在通过一种称为消息传递的机制，从非欧几里得图结构数据中学习[38–40]。这一机制是图神经网络的核心过程，使信息能够在相连的节点之间流动，从而让网络通过多层聚合学习局部依赖关系（来自直接邻居）和全局图拓扑结构。

在图神经网络中的图里，实体表示为节点（V），而它们之间的关系或相似性则表示为边（E）。这种图表示使模型能够利用样本之间的结构相似性来提高学习准确率。在消息传递开始之前，每个节点$i \in V$ 都会初始化为一个 d 维特征向量，用以表示其输入属性，

$$x_i = h_i^0 \in \mathbb{R}^d \qquad \text{公式 1}$$

其中，x_i 表示节点i 的输入特征向量，$\mathrm{h}_i^{(0)}$ 表示其在图神经网络输入层的初始状态。在消息传递过程中，每个节点i通过聚合其邻居节点$j \in N(i)$的信息，更新其在第k层的隐藏信息或特征 $\mathrm{h}_i^{(k)}$。这是一个迭代过程，每个节点在聚合邻居信息的同时，通过结合自身信息来更新自身。消息传递分为两个步骤：聚合和组合，表示为

$$a_i^{(k)} = \mathrm{AGGREGATE}^{(K)}\{ h_j^{(k-1)} : \mathrm{j} \in N(i)\} \qquad \textbf{Eq. 2}$$

$$h_i^{(k)} = \sigma\left(W^{(k)} \cdot \mathrm{COMBINE}^{(K)}\{ h_i^{(k-1)}, a_i^{(k)}\}\right) \qquad \textbf{Eq. 3}$$

其中，$N(i)$是节点i 的邻居节点集合，$\mathrm{h}_i^{(k-1)}$ 是该节点在上一次迭代$k-1$ 中的特征。对于每个节点，*AGGREGATE*函数都会收集其所有邻居节点 j∈ $N(i)$的特征 $\mathrm{h}_j^{(k-1)}$ ，并生成一条中间聚合消息 $a_i^{(k)}$。随后，*COMBINE*函数将这条聚合消息与该节点自身先前的表示 $\mathrm{h}_i^{(k-1)}$ 合并，生成更新后的节点特征表示 $\mathrm{h}_i^{(k)}$。然后，将每个节点的更新结果乘以可学习权重矩阵 $W^{(k)}$，将聚合信息转换为更新后的节点表示，并应用非线性激活函数 σ ，为模型引入非线性。本研究采用了修正线性单元（*ReLU*）激活函数，该函数已广泛应用于 GCN 和 GraphSAGE 等现代图神经网络架构中[41,42]。*ReLU*能够有效缓解深度神经网络中常见的梯度消失问题，同时保持计算简洁并产生稀疏激活。这些特性可加快训练过程中的收敛，并提升泛化性能。

给定图结构和节点信息后，图神经网络可根据预测任务的类型，在三个层级上运行：节点级、边级和图级[40,43]。

- 在节点级任务中，模型根据单个节点及其邻居的特征，估计该节点的属性，例如进行分类或回归。

- 在边级任务中，模型预测两个相连节点之间的关系或交互，包括边是否存在、边的类型或强度。

- 在图级任务中，模型使用均值、求和或注意力池化，将所有节点的嵌入整合为单一的图表示，进而预测分子活性或材料类别等全局属性。

为实现消息传递，发展出了不同的图神经网络架构，并采用各种聚合和更新方案。这些架构共享同一概念基础，即消息传递框架，但在如何聚合、归一化或加权邻居信息方面各不相同。总体而言，它们使图神经网络能够从复杂的非欧几里得数据结构中学习层次化的关系表示。

- 图卷积网络（GCN）：GCN 使用归一化邻接矩阵进行特征传播。每个节点通过结合自身特征与邻居节点的特征来更新其表示，然后应用线性变换和非线性激活 [41]。GCN 最适用于同配图，即相连节点具有相似属性的图。

- GraphSAGE：GraphSAGE 通过采样固定数量的邻居、聚合其特征，并将其与节点自身的特征拼接后进行线性变换，从而学习节点嵌入。这使其能够对未见节点或图进行归纳学习 [42]。得益于其采样策略，GraphSAGE 非常适用于大规模或动态的图。

- 图注意力网络（GAT）：GAT 利用自注意力在聚合过程中为邻居节点分配不同的重要性权重，从而实现自适应特征传播 [44]。GAT 对异构图十分有效，因为邻居的重要性各不相同。

- 消息传递神经网络（MPNN）：MPNN 定义了一种通用的消息传递框架，由消息阶段和更新阶段组成。边特征可以无缝集成其中，使其非常适用于分子和物理系统 [39]。MPNN 非常适用于边数量众多的图，例如分子图，因为键细节至关重要。

本研究采用 PyTorch Geometric 库实现结合均值聚合的 GraphSAGE 架构，以执行有监督节点级回归。在拼接部分，采用均值聚合。所构建图中的每个节点代表一种混凝土配合比设计，输出节点表示抗压强度。为定义样本之间的关系（GNN 中的边），在归一化特征空间中应用 K 近邻模型 [45,46]，在组成相似的节点之间创建边。该策略将表格数据转换为图结构，使图神经网络能够在相似节点之间传播信息并学习关系依赖。通过添加互反边，将所得图构建为无向图，从而确保相连节点之间进行双向信息交换。

使用 K近邻可能具有挑战性，因为其需要调节超参数。关键超参数包括邻居数量、距离函数和加权函数 [46]。本文在较宽的 K 值范围（2 至 30）内开展了敏感性研究，以考察图连通性对模型性能的影响。较低的 K近邻值可能导致连接稀疏和欠拟合，而较高的值则可能因纳入过多邻居信息而导致过平滑。K近邻构建使模型能够结合每条数据记录最相似邻居的情况进行解释，而不是孤立地处理数据记录。采用欧几里得函数 [47,48] 计算数据记录之间的距离。

选择 GraphSAGE 层是因为它能够进行归纳学习，这意味着无需在整个图上重新训练，它也能泛化到未见数据或新节点 [42,44]。与依赖固定邻接矩阵的 GCN 等谱方法不同 [41,43], , GraphSAGE 会学习聚合函数，从节点邻域中采样并组合信息。这种方法尤其适用于混凝土混合物等数据集，因为其中每个节点（样本）相互独立，但彼此之间存在组成关系。

本文的网络架构由三个 GraphSAGE 卷积层组成，每层包含 128 个隐藏神经元，后接批归一化、ReLU激活和随机失活（0.25），以防止过拟合。模型参数使用 Adam优化器进行优化，损失函数为均方误差（MSE），这与其他基于回归的图神经网络研究一致 [41, 42]。为确保可复现性，采用固定的随机种子（42），使模型初始化、数据划分和训练结果在多次运行中保持一致。通过多层消息传递和非线性变换，网络学会将节点特征和邻域信息映射为预测的抗压强度。

为展示所实现的图神经网络模型的工作流程，**图1**呈现了用于模型构建、训练和评估的完整流程。该过程首先对输入数据进行归一化，并使用 K近邻构建图，随后训练 GraphSAGE 模型，通过消息传递聚合邻居节点的信息。在每次前向传播过程中，节点特征会被传递并组合，以形成更新后的节点表示，同时计算训练损失并

通过反向传播和权重更新来最小化训练损失，过程与传统神经网络类似。持续监控模型在验证集上的性能，一旦检测到性能提升便保存模型。如果连续 30 个训练轮次均未出现性能提升，则提前终止训练，以防止过拟合。最后，重新加载性能最佳的模型，在未见过的数据上进行测试和评估，然后将预测结果转换回原始物理单位（兆帕）。

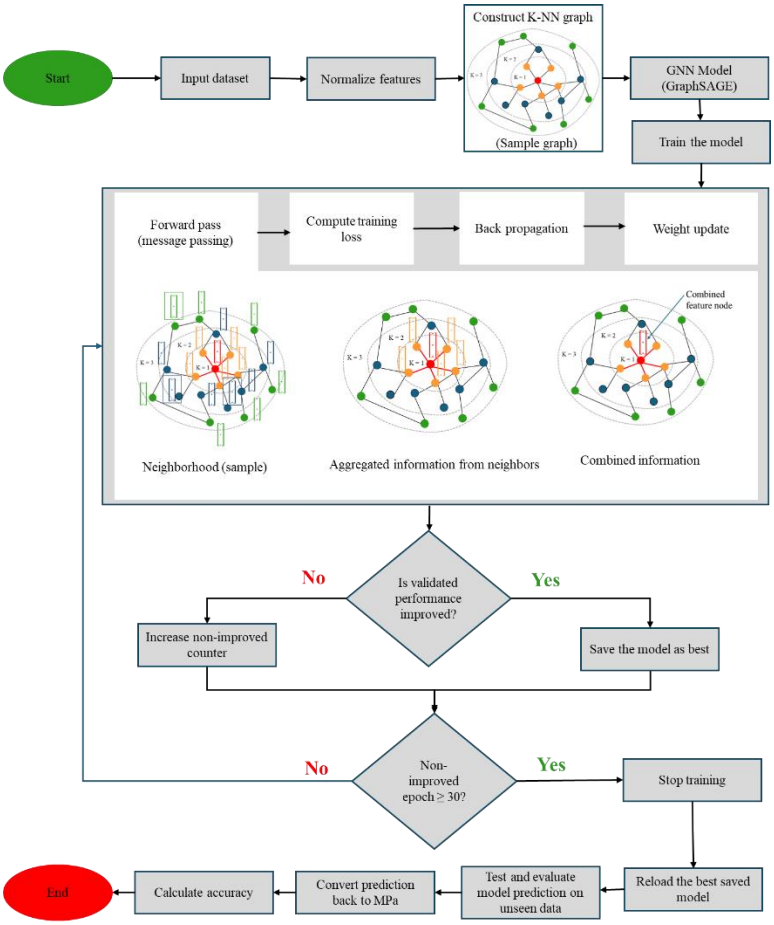


图1：利用配合比设计预测混凝土抗压强度的图神经网络模型工作流程，涵盖数据预处理、模型实现和性能评估。

3.0. 数据库收集、数据处理和随机森林（RF）模型

本研究所用的数据库是来自 UCI机器学习存储库 [49] 的混凝土抗压强度数据集。这个成熟的数据库已被广泛用于评估各种机器学习模型的性能。由于该数据库相对简单且结构清晰，读者可以专注于理解图神经网络模型的实现及其性能。此外，

使用这一基准数据库，有助于直接比较图神经网络与传统机器学习模型（即随机森林）。采用此类标准基准对于评估所提出的图神经网络框架的优势与局限至关重要。

混凝土数据库包含 1030 条数据记录，每条记录包含八个输入变量和一个输出变量，数据库中没有缺失值。输入变量包括水泥含量（kg/m^3）、粉煤灰含量（kg/m^3）、矿渣含量（kg/m^3）、高效减水剂（kg/m^3）、细骨料含量（kg/m^3）、粗骨料含量（kg/m^3）、含水量（kg/m^3）以及龄期（天）。输出变量为时变抗压强度（MPa）。有关数据范围和分布的统计参数可见我们先前的研究 [7]。在构建图并将数据输入图神经网络模型之前，我们进行了数据预处理，以保持数值一致性并使特征具有可比性。对数据库进行了归一化[50–53]，确保输入变量的取值范围为 0 至 1。这使每个特征都能对图生成和消息传递发挥同等作用，从而减少数据库带来的偏差。

为训练和验证图神经网络模型，将抗压强度数据库随机划分为三个数据集：（1）70% 的数据记录作为训练数据集；（2）15% 的数据记录作为验证数据集；（3）其余 15% 的数据记录作为测试数据集。为评估图神经网络模型的性能，采用了四种统计指标：决定系数（R^2）、平均绝对误差（*MAE*）、均方根误差（*RMSE*）和平均绝对百分比误差（*MAPE*）。这些指标用于量化预测抗压强度与实测抗压强度之间的差异。

本研究采用随机森林模型，以便与图神经网络模型进行直接性能比较。下文简要介绍随机森林，更多详情请参阅我们此前的研究[14,54,55]。随机森林是一种基于分类与回归树的集成学习方法，通过两个组成部分提升预测性能：自助聚合（装袋法）和随机化 [56,57]。具体而言，随机森林使用随机生成的训练数据子集构建多棵决策树，每个子集均通过对父数据集中的数据记录进行有放回抽样生成。在每棵树的各个节点，随机森林会随机选择一部分输入变量进行节点分裂，从而引入额外的随机性。随机森林允许树生长至最大深度，从而

提高各棵树之间的多样性。在预测过程中，随机森林会聚合所有树的输出（回归任务中通常取平均值），以生成最终输出。

4.0. 结果与讨论

本研究实现了两类图神经网络任务：节点级回归和图级回归。研究重点是节点级回归，因为它更符合数据库的性质，并能在训练过程中有效保留混合设计信息。作为对比，我们还使用相同的数据集进行了图级回归。特征选择对于确定机器学习的预测性能至关重要。在胶凝材料领域，同一含义的特征可以通过多种方式表示，例如，水和水泥含量可以作为独立输入，也可以合并表示为水胶比。为降低计算需求并提高模型效率，通常希望尽可能减少输入变量的数量。因此，本研究首先评估不同的特征组合，以考察它们对预测准确度的影响。

采用随机森林模型，利用基尼重要性分数评估每个输入变量对预测性能的影响。关于如何使用随机森林模型计算变量重要性的详细信息，参见我们此前的研究 [58]。**图2**展示了各变量对抗压强度预测准确度的影响。正如预期，水泥含量、含水量、高效减水剂掺量和养护龄期对抗压强度的影响最为显著。水泥通过水化反应成为强度的主要来源；含水量决定孔隙率，而孔隙率与强度直接相关；适量的高效减水剂则可提高堆积密度并减少夹带空气。此外，由于水化产物持续形成，抗压强度通常会随龄期增长。在这些主导变量之后，高炉矿渣、细骨料、粗骨料和粉煤灰含量对抗压强度的影响相对较小。

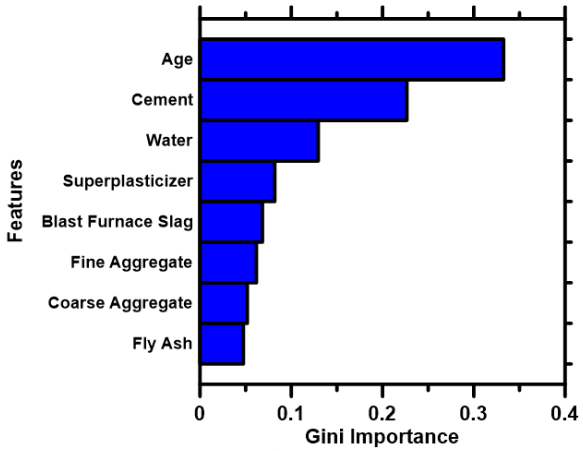


图2：随机森林模型确定的影响混凝土抗压强度的输入变量的重要性。变量按重要性降序排列。

基于特征重要性分析和水泥化学领域知识，我们将图神经网络模型的输入变量重新划分为五个特征组。这些特征组既包括移除影响较小的变量，也包括将相关输入变量合并为单一代表性特征。**表1**总结了这五个特征组及其对应的变量。该方法的主要目标是评估图神经网络模型在减少输入变量数量的同时，能否保持可靠的预测结果。原始数据集被保留为特征组A。B组包括**图2**所示的4个最重要输入变量。为减轻特征中的多重共线性，根据水泥化学将若干输入变量合并为单一输入变量，从而形成特征组C。特征组D通过将相似材料合并为一组进行设计。在特征组D的基础上，特征组E进一步对相似材料进行合并。

表1：图神经网络模型中使用的五个特征组的定义，旨在减少输入变量的数量，并将多个相关特征整合为统一的表示。

特征组	数量特征	输入特征
A	8	1. 水泥；2. 矿渣；3. 粉煤灰；4. 水；5. 高效减水剂；6. 粗骨料；7. 细骨料；8. 龄期
B	4	1. 水泥；2. 水；3. 高效减水剂；4. 龄期
C	7	1. 胶凝材料 = 水泥 + 矿渣 + 粉煤灰；2. 骨料 = 粗 + 细骨料；3. 流动性 = 水 + 高效减水剂；4. 水胶比 = 水/胶凝材料；5. 骨料/胶凝材料 = 骨料/胶凝材料；6. 高效减水剂/胶凝材料 = ；7. 龄期
D	7	1. 水泥；2. 矿渣；3. 粉煤灰；4. 水；5. 高效减水剂；6. 骨料 = 粗骨料 + 细骨料；7. 龄期

E 6 1. 水泥；2. 补充胶凝材料 = 矿渣 + 粉煤灰；3. 水；4. 高效减水剂；5. 骨料 = 粗骨料 + 细骨料；6. 龄期

图神经网络模型应用于全部五个特征组。为获得最佳性能，我们对超参数进行了仔细调优。在本研究中，最关键的超参数是 K近邻中的最近邻数量，它决定了每个节点所连接的边数。**图3** 展示了各特征组中邻居数量与模型准确率之间的关系，而**表2**汇总了各组在最佳邻居数量下对应的统计性能指标。

图3 表明，预测性能对邻域大小的选择高度敏感。当邻域较小时，生成的图较为稀疏——许多节点只有很少甚至没有连接，这限制了消息传递，并增加了预测的方差。相反，当邻域大小变大时（K > 8），图会变得过于稠密，大多数节点几乎与其他所有节点相连。这种过度连接会导致过平滑，这是 GraphSAGE 等消息传递图神经网络中的常见问题；由于从邻居节点聚合了过多信息，节点表示会变得过于相似 [42,43,59]。在大多数特征组中，最佳邻域大小范围为 3 至 5，即 3–5。由于 A 和 E 组的准确度相对较高，因此有必要进一步研究更稠密的邻域配置（如**图3c**和**3d**所示）。然而，当邻域大小超过这一范围时，预测准确度会明显下降。最佳邻域大小似乎取决于两个关键因素：数据库大小和输入变量的数量。对于数据集较大或具有更多输入特征的数据集，可能需要更大的邻域大小，才能有效捕捉具有代表性的关系。例如，仅包含四个输入变量的 B组在邻域大小为 2 时取得了最佳性能。未来的研究应进一步验证这种关系，以便为胶凝材料图神经网络模型中邻域大小的选择建立更具普适性的指导原则。

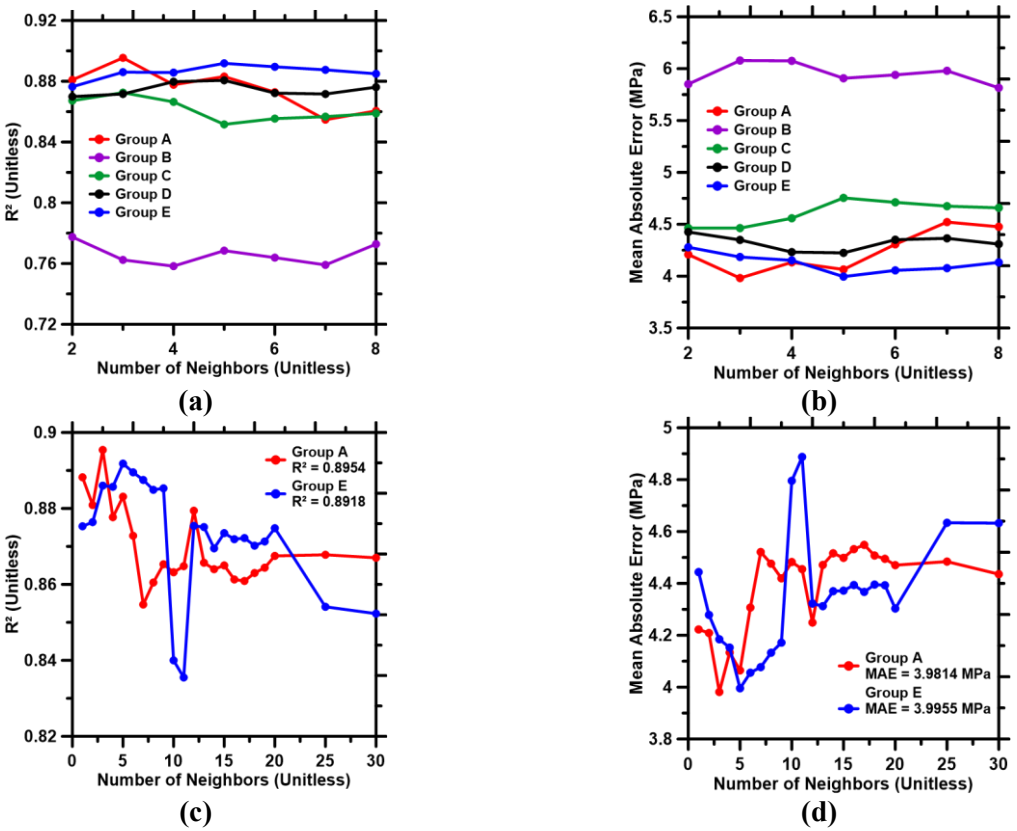


图3：不同特征组的图神经网络模型预测性能随邻居数量的变化。(a) 和 (b) 展示了较小邻域大小下所有特征组的模型性能，而 (c) 和 (d) 展示了较大邻域大小下特征组A和E的性能。

表2：不同特征组和任务类型下图神经网络模型的预测性能及其最佳邻居数量。

图神经网络类型和特征组	数量邻居	R^2（无单位）	平均绝对误差（兆帕）	平均绝对百分比误差（%）	均方根误差（兆帕）
节点级别 - A	3	0.8954	3.98	14.48%	5.07
节点级别 - B	2	0.7775	5.85	20.40%	7.39
节点级别 - C	3	0.8725	4.46	16.54%	5.59
节点级别 - D	5	0.8806	4.22	15.29%	5.41
节点级别 - E	5	0.8918	3.99	15.08%	5.15
图级别 - A	3	0.8138	5.29	19.48%	6.98

在**表2**中可以清楚看出，邻域大小为 3 的特征组A取得了最佳预测性能，R^2 为 0.8954，平均绝对误差为 3.98 兆帕。这种优异性能是意料之中的，因为 A 组包含所有直接影响抗压强度的关键混合料设计参数。相比之下，仅包含四个关键特征的 B组表现出最低的预测准确度，

无论邻域大小如何，这都证实，缺少关键的配合比设计信息会限制模型学习隐藏输入-输出相关性的能力。尽管使用的输入变量较少，C、D 和 E 组仍表现出良好的预测性能。这些组的成功源于将多个原始特征组合成复合输入，使模型能够在降低计算成本的同时保留关键信息。表现优异的 A 和 E 组将用于后续研究。如表所示，采用特征组A的图级任务表现不及节点级模型。这种性能下降归因于图级任务中的全局池化操作，该操作会对整个图中的信息进行平均，从而丢失样本特有的细节。后续章节将进一步分析图级任务。

为确保模型性能达到最佳，训练轮数也是一个需要仔细调优的关键超参数。在神经网络中，训练轮次指训练期间整个训练数据集通过模型的次数。每个训练轮次都会根据预测值与实际值之间的均方误差更新内部权重函数，使模型逐步逼近损失函数的全局最小值。训练轮数不足可能导致欠拟合，而训练轮数过多则可能造成过拟合并降低泛化能力。

图 4展示了模型经过多个训练轮次后，训练数据集、验证数据集和测试数据集的均方根误差变化情况。蓝色、橙色和绿色曲线分别对应训练数据集、验证数据集和测试数据集。由于采用了早停条件，各图在不同的训练轮次处结束——如果连续 30 个训练轮次的验证损失未能至少改善 $1\times10^{-}4$，就会停止训练。对于节点级和图级回归，图神经网络模型在特征组A中采用的邻域大小为 3，在特征组E中采用的邻域大小为 5，因为这些配置分别能为相应的数据集提供最准确的预测。

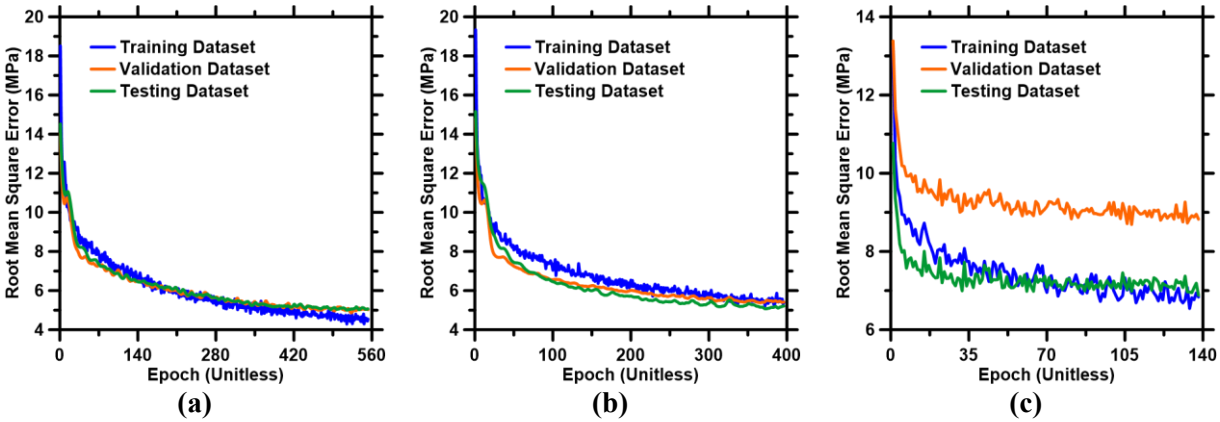


图 4：图神经网络模型的预测准确度随训练轮数的变化：（a）使用特征组A、邻域大小为 3 的节点级任务；（b）使用特征组E、邻域大小为 5 的节点级任务；（c）使用特征组A、邻域大小为 3 的图级任务。

在**图 4a**中，均方根误差在前 100 个训练轮次中迅速下降，表明模型快速捕捉到了配合比设计与抗压强度之间的非线性关系。随着训练的进行，曲线逐渐趋于平缓，表明模型收敛且性能稳定。均方根误差持续下降至约 280 个训练轮次，此后，三条曲线均趋于稳定——在第 523 个训练轮次，训练数据集约为 4.69 兆帕，验证数据集约为 5.08 兆帕，测试数据集约为 5.07 兆帕。三个数据集的结果高度一致，表明模型实现了良好的泛化，且没有过拟合。验证损失在第 523 个训练轮次达到最低值，根据早停准则，训练在第 553 个训练轮次停止。**图 4b**也呈现出类似趋势：均方根误差曲线更早趋于平缓，约在 200 个训练轮次时稳定下来；在第 366 个训练轮次，训练、验证和测试数据集的误差分别稳定在约 5.41 兆帕、5.38 兆帕和 5.15 兆帕，训练则在第 396 个训练轮次终止。该情形下收敛速度更快，反映出与特征组A相比，特征组E的输入变量更少，使模型能够更高效地学习相关性，同时降低计算需求。

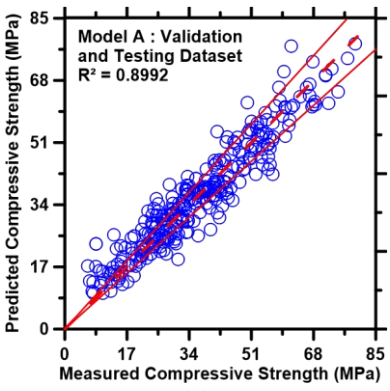
在**图 4c**中，图级图神经网络比节点级模型早得多地达到收敛，在约第 108 个训练轮次时便趋于稳定，但均方根误差值明显更高——训练数据集为 6.93 兆帕，验证数据集为 8.69 兆帕，测试数据集为 6.98 兆帕。这种快速稳定与整体误差偏高的情况表明，图表示与底层问题定义之间存在结构不匹配。在本研究中，每个数据点代表一种独立的混凝土混合料

设计及其抗压强度。然而，当所有样本被合并到单个全局图中，并通过全局池化操作进行处理时，模型实际上会在每次前向传播中，将多个各不相同的混合料压缩为单个图级表示。这种池化步骤会造成表示瓶颈：多样的节点级特征被聚合为固定大小的嵌入，必然导致信息损失，并减少有效独立训练样本的数量。这种行为符合图神经网络中广为人知的“过度挤压”和“池化瓶颈”现象：跨越关系疏松或相互独立的节点进行过度消息聚合，会限制模型表达能力并降低预测准确率 [60–62]。因此，图级模型表现出过拟合迹象，反映在验证数据集上的预测误差较高。这凸显了全局池化的局限性：当数据集中的每条记录代表独立观测而非相互依赖的实体时，不宜采用全局池化。

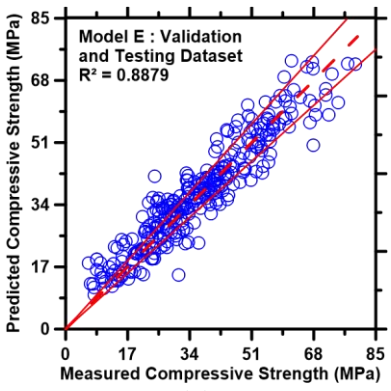
尽管图神经网络模型的误差指标表明其预测性能可靠，但其能否达到成熟机器学习模型的水平仍不确定。为此，我们还将同一抗压强度数据库应用于随机森林模型，该模型广泛用于预测混凝土性质。随机森林模型采用与图神经网络模型相同的数据集和划分策略进行训练。通过十折交叉验证优化随机森林模型的超参数，并将其余 30% 的数据保留为测试数据。为进行直接的一对一比较，图神经网络模型的预测结果涵盖验证数据集和测试数据集，这两者对应于随机森林测试数据集所用的同一 30% 数据。由于验证数据集未用于训练，因此对于图神经网络模型而言，它仍是盲数据，尽管该数据集被用于超参数调优。此方法确保能够公平比较两个模型的预测准确度和泛化能力。

图5比较了图神经网络模型和随机森林模型所得抗压强度预测结果的预测性能。节点级图神经网络模型表现良好，其中，特征组A的$R^2 = 0.8992$，特征组E的$R^2 = 0.8879$。在节点级模型中，数据点与对角虚线高度一致，表明图神经网络有效捕捉了配合比设计与强度之间的非线性关系。相比之下，图级模型的数据点更加分散（$R^2 = 0.8039$），反映出

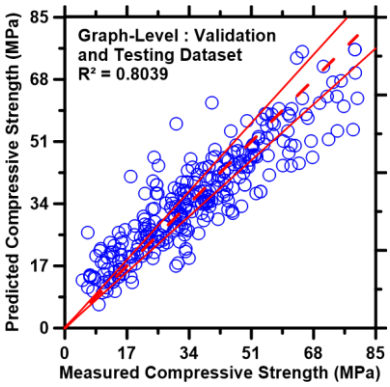
较弱的预测性能以及全局池化过程中样本特定信息的丢失。与RF基准相比，随机森林模型实现了最高的总体预测准确率（R^2 = 0.9016）。尽管存在这一细微差异，图神经网络仍表现出相当的预测性能。不同于随机森林等传统的基于特征的模型，图神经网络通过消息传递学习高阶依赖关系和特征间交互，从而能够更好地推广到盲数据。此外，图拓扑结构天然编码了材料相似性，提升了物理可解释性 [42]。尽管具有这些优势，图神经网络仍存在两个主要缺点。首先，由于消息传递过程中需要反复进行稀疏-稠密矩阵运算并执行不规则数据访问，其训练需要显著更多的计算资源和内存 [63]。其次，图神经网络对图构建参数较为敏感，例如邻居数量可能会显著影响模型稳定性和性能 [64]。随机森林模型以极低的计算成本提供了快速、可靠且准确度出色的基线，而图神经网络则在大型或关系型材料数据集方面提供了更高的灵活性、可解释性和可扩展性；尽管需要更多计算资源，这些特性仍是推进数据驱动的材料设计的关键属性。这些特点使图神经网络成为传统机器学习模型在胶凝材料领域中颇具潜力的替代方案。



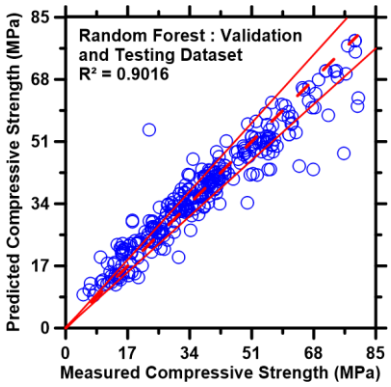
(a)



(b)



(c)



(d)

图5：图神经网络模型对混凝土抗压强度的预测结果（（a）节点级：数据模型A（K近邻=3），（b）节点级：数据模型E（K近邻=5），（c）图级图神经网络（K近邻=3）；（d）随机森林。红色虚线表示理想预测值，红色实线表示± 10%误差界限。

5.0. 结论

本研究将图神经网络应用于胶凝材料表格数据库，建立了一条清晰且可复现的路径，将传统配合比设计输入转化为图结构表示。通过采用K近邻策略进行图构建，并结合依据领域知识确定的特征选择方案，本研究表明，图神经网络能够有效学习决定混凝土抗压强度的非线性多维关系。

本研究将每条数据记录视为一个节点，以构建节点级图神经网络模型。根据数据点之间的相似性，使用 K近邻建立边，从而有效地将表格数据集转换为图结构表示。这种方法为将图神经网络模型应用于数值数据提供了清晰且可复现的途径。此外，本研究还系统评估了特征选择，以平衡计算效率和预测准确率。结果表明，纳入完整混合设计信息的模型取得了最高的预测准确度。然而，将若干相关输入变量整合为一个代表性特征，可显著缩短训练时间，同时保持相近的准确率。相比之下，图级图神经网络因全局池化过程中的信息损失而表现出较低的预测准确度，这说明对于每条记录代表一种独立配合比设计的数据集，节点级架构更为适用。

总体而言，图神经网络模型的表现与传统机器学习模型相当，展现出其在推动胶凝材料数据驱动建模方面的巨大潜力。

图神经网络在推动材料信息学发展方面具有独特优势。其灵活的图拓扑结构能够显式融入物理定律、分层材料描述符和多尺度关系，而这些能力是传统机器学习模型所无法实现的。这些特点使图神经网络成为构建新一代多模态、物理信息驱动人工智能模型的有前景基础。尽管目前仍存在局限，尤其是对图构建的敏感性以及较高的计算需求，但已展现出的准确性和可解释性凸显了图神经网络在胶凝材料领域的变革潜力。

总体而言，本研究为开发先进的人工智能架构奠定了基础路线图，旨在通过单一学习框架整合数值、微观结构和物理化学信息。本文提出的方法和获得的见解，是迈向可扩展、可解释且由物理信息驱动的图神经网络模型的关键一步，有望改进混凝土设计、加速材料发现，并深化对胶凝材料体系的科学认识。

致谢

作者感谢 Kummer Institute（密苏里科技大学）和美国国家科学基金会（NSF-DMR: 2034856 和 2228782）提供的经费支持。

数据可用性

本研究使用的数据库下载自加州大学欧文分校机器学习资料库。

作者贡献

Mahmuda Sharmin： 研究；形式分析；撰写——初稿；方法学**Taihao Han：** 概念化；研究；撰写——初稿**Gaurav Sant：** 监督；撰写——审阅与编辑；经费获取

Narayanan Neithalath：指导；撰写——审阅与编辑；以及经费筹措

Aditya Kumar：概念构思；指导；撰写——审阅与编辑；以及经费筹措

参考文献

[1] J.W. Bullard, H.M. Jennings, R.A. Livingston, A. Nonat, G.W. Scherer, J.S. Schweitzer, K.L. Scrivener, J.J. Thomas, 水泥水化机制，《水泥与混凝土研究》41 (2011) 1208–1223. https://doi.org/10.1016/j.cemconres.2010.09.011. [2] E. John, B. Lothenbach, 水泥水化机制的演变——综述，Journal of Materials Science 58 (2023年) 9805–9833. https://doi.org/10.1007/s10853-023-08651-9. [3] T. Oey, A. Kumar, G. Falzone, J. Huang, S. Kennison, M. Bauchy, N. Neithalath, J.W. Bullard, G. Sant, 水活度对硅酸三钙水化速率的影响，《美国陶瓷学会会刊》99 (2016年) 2481–2492. https://doi.org/10.1111/jace.14181. [4] B. Lothenbach, T. Matschei, G. Möschner, F.P. Glasser, 温度对波特兰水泥水化和孔隙率影响的热力学建模，《水泥与混凝土研究》38 (2008) 1–18. https://doi.org/10.1016/j.cemconres.2007.08.017. [5] A. Çolak, 一种估算波特兰水泥混凝土抗压强度的新模型，《水泥与混凝土研究》36 (2006) 1409–1413. https://doi.org/10.1016/j.cemconres.2006.03.002. [6] C.-J. Haecker, E.J. Garboczi, J.W. Bullard, R.B. Bohn, Z. Sun, S.P. Shah, T. Voigt, 波特兰水泥浆体线弹性性质建模，《水泥与混凝土研究》35 (2005) 1948–1960. https://doi.org/10.1016/j.cemconres.2005.05.001. [7] R. Cook, J. Lapeyre, H. Ma, A. Kumar, 混凝土抗压强度预测：混合机器学习模型与独立模型性能的关键比较，《土木工程材料杂志》31 (2019) 04019255. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902.[8] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake，利用 Shapley 加性解释（SHAP）揭示机器学习在混凝土抗压强度预测中的黑箱特性：一种新方法，《建筑材料案例研究》16 (2022年) e01059. https://doi.org/10.1016/j.cscm.2022.e01059. [9] G.A. Lyngdoh, M. Zaki, N.M.A. Krishnan, S. Das，借助缺失数据插补和可解释机器学习预测混凝土强度，Cement and Concrete Composites 128 (2022年) 104414. https://doi.org/10.1016/j.cemconcomp.2022.104414. [10] S. Nazar, J. Yang, X.-E. Wang, K. Khan, M.N. Amin, M.F. Javed, F. Althoey, M. Ali，利用机器学习和 SHapely Additive exPlanations（SHAP）估算碱激发砂浆的强度、流变参数及原材料成分的影响，Construction and Building Materials 377 (2023年) 131014. https://doi.org/10.1016/j.conbuildmat.2023.131014. [11] O. Canbek, Q. Xu, Y. Mei, N.R. Washburn, K.E. Kurtis，预测石灰石煅烧黏土水泥（石灰石煅烧黏土水泥）的流变性：通过机器学习将组成和水化动力学与屈服应力联系起来，《水泥与混凝土研究》160 (2022年) 106925. https://doi.org/10.1016/j.cemconres.2022.106925. [12] Z. Tariq, M. Murtaza, M. Mahmoud，利用机器学习技术开发含纳米黏土添加剂的G级水泥新型流变模型，《ACS 欧米伽》5 (2020年) 17646–17657. https://doi.org/10.1021/acsomega.0c02122.

[13] J. Lapeyre, T. Han, B. Wiles, H. Ma, J. Huang, G. Sant, A. Kumar，机器学习能够快速预测多组分水泥基体系的水化动力学，《Scientific Reports》11 (2021) 3922。https://doi.org/10.1038/s41598-021-83582-6。[14] T. Han, S.A. Ponduru, R. Cook, J. Huang, G. Sant, A. Kumar，利用深度学习设计和发现可持续水泥基胶凝材料：从小型数据库中学习并开发闭式解析模型的策略，Frontiers in Materials 8 (2022) 796476。https://doi.org/10.3389/fmats.2021.796476。[15] Y. Yang, Z. Cheng，集成机器学习与 Shapley 加性解释在评估 C-S-H晶种加速水泥水化能力方面的应用，Journal of Materials Science 59 (2024) 2852–2871。https://doi.org/10.1007/s10853-024-09380-3。[16] W.Z. Taffese, E. Sistonen，机器学习在钢筋混凝土结构耐久性和使用寿命评估中的应用：最新进展与未来方向，Automation in Construction 77 (2017) 1– 14。https://doi.org/10.1016/j.autcon.2017.01.016。[17] H. Chen, Y. Cao, Y. Liu, Y. Qin, L. Xia，提升严寒地区混凝土的耐久性：基于机器学习的配合比优化，Construction and Building Materials 371 (2023) 130644。https://doi.org/10.1016/j.conbuildmat.2023.130644。[18] M. Zhang, R. Kang，预测混凝土材料耐久性的机器学习方法：综述，Advances in Cement Research 37 (2025) 502–517。https://doi.org/10.1680/jadcr.24.00133。[19] M. Liang, K. Feng, S. He, Y. Gan, Y. Zhang, E. Schlangen, B. Šavija，利用机器学习模型生成水泥浆体微观结构，Developments in the Built Environment 21 (2025) 100624。https://doi.org/10.1016/j.dibe.2025.100624。[20] Y. Yu, G. Geng，用于水泥浆体和掺矿物掺合料体系背散射电子图像相分割的深度学习方法，Cement and Concrete Composites 155 (2025) 105810。https://doi.org/10.1016/j.cemconcomp.2024.105810。[21] J. Guo, M. Li, L. Wang, B. Yang, L. Zhang, Z. Chen, S. Han, L. Garcia-Hernandez, A. Abraham，利用三维微观结构图像和深度置信网络估算水泥抗压强度，Engineering Applications of Artificial Intelligence 88 (2020) 103378。https://doi.org/10.1016/j.engappai.2019.103378。[22] J. Velez-Alvarez, W. Cordero, K. Arboleda, N. Rios，通过光学光谱学和基于机器学习的图像识别确定水泥质量，Results in Optics 13 (2023) 100525。https://doi.org/10.1016/j.rio.2023.100525。[23] S. Li, Z. Fang, H. Xu, Y. Sakai，基于高光谱成像的结合水演变与机器学习辅助水泥浆体早期水化程度估算，Construction and Building Materials 470 (2025) 140597。https://doi.org/10.1016/j.conbuildmat.2025.140597。[24] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, R. Barzilay，图神经网络，Nature Reviews Methods Primers 4 (2024) 17。https://doi.org/10.1038/s43586-024-00294-7。[25] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu，图神经网络综合综述，《IEEE神经网络汇刊》32 (2021) 4–24。https://doi.org/10.1109/TNNLS.2020.2978386。[26] F. Liang, C. Qian, W. Yu, D. Griffith, N. Golmie，图神经网络及其应用综述，Wireless Communications and Mobile Computing 2022 (2022) 9261537。https://doi.org/10.1155/2022/9261537。[27] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, P. Friederich，面向材料科学与化学的图神经网络，Commun Mater 3 (2022) 1–18。https://doi.org/10.1038/s43246-022-00315-6。[28] X. Shi, L. Zhou, Y. Huang, Y. Wu, Z. Hong，图神经网络在原子尺度材料科学中的应用综述，Materials Genome Engineering Advances 2 (2024) e50。https://doi.org/10.1002/mgea.50。[29] V. Fung, J. Zhang, E. Juarez, B.G. Sumpter，材料化学领域图神经网络的基准测试，Npj Comput Mater 7 (2021) 84。https://doi.org/10.1038/s41524-021-00554-0。

[30] M. Maurizi, C. Gao, F. Berto，利用图神经网络预测材料和结构中的应力、应变及变形场，《科学报告》12 (2022) 21834。https://doi.org/10.1038/s41598-022- 26424-3。[31] Y. Fei, W. Liao, X. Lu, H. Guan，利用知识增强型图神经网络估算钢筋混凝土建筑的建筑材料用量，《计算机辅助土木与基础设施工程》39 (2024) 518–538。https://doi.org/10.1111/mice.13094。[32] F. Liu, I.L. Al-Qadi，基于图神经网络的三维沥青混凝土路面响应建模，《国际路面工程杂志》26 (2025) 2464201。https://doi.org/10.1080/10298436.2025.2464201。[33] G. Shi, S. Pan, R. Zou, A. Yu，STGNets：用于水泥工业制造流程能耗预测的时空图神经网络，《粉体技术》434 (2024) 119280。https://doi.org/10.1016/j.powtec.2023.119280。[34] C.-T. Li, Y.-C. Tsai, C.-Y. Chen, J.C. Liao，表格数据学习的图神经网络：分类体系与发展方向综述，《ACM计算综述》。58 (2025) 16:1-16:51。https://doi.org/10.1145/3744918。[35] Q.-J. Hong, S.V. Ushakov, A. van de Walle, A. Navrotsky，利用图神经网络模型预测熔化温度：从古代矿物到新材料，《美国国家科学院院刊》119 (2022) e2209630119。https://doi.org/10.1073/pnas.2209630119。[36] X. Guo, Y. Quan, H. Zhao, Q. Yao, Y. Li, W. Tu，TabGNN：用于表格数据预测的多重图神经网络，(2021)。https://doi.org/10.48550/arXiv.2108.09127。[37] Y. LeCun, Y. Bengio, G. Hinton，深度学习，《自然》521 (2015) 436–444。https://doi.org/10.1038/nature14539。[38] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini，图神经网络模型，《IEEE神经网络汇刊》20 (2009) 61–80。https://doi.org/10.1109/TNN.2008.2005605。[39] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl，面向量子化学的神经消息传递，(2017)。https://doi.org/10.48550/arXiv.1704.01212。[40] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu，图神经网络综合综述，IEEE Transactions on Neural Networks and Learning Systems 32 (2021) 4–24。https://doi.org/10.1109/TNNLS.2020.2978386。[41] T.N. Kipf, M. Welling，利用图卷积网络进行半监督分类，(2017)。https://doi.org/10.48550/arXiv.1609.02907。[42] W. Hamilton, Z. Ying, J. Leskovec，大型图上的归纳式表示学习，(2018)。https://doi.org/ 10 48550 1706 02216 。/arXiv. . J. [43] Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun，图神经网络：方法与应用综述，《AI Open》1 (2020) 57–81。https://doi.org/10.1016/j.aiopen.2021.01.001。[44] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio，图注意力网络，(2018)。https://doi.org/10.48550/arXiv.1710.10903。[45] S. Kang，基于图神经网络的K近邻学习，《数学》9 (2021) 830。https://doi.org/10.3390/math9080830。[46] L. Jiang, Z. Cai, D. Wang, S. Jiang，改进K近邻分类方法综述，载于：第四届模糊系统与知识发现国际会议（FSKD 2007），2007：第679–683页。https://doi.org/10.1109/FSKD.2007.552。[47] P.-E. Danielsson，欧几里得距离映射，《计算机图形学与图像处理》14 (1980) 227–248。https://doi.org/10.1016/0146-664X(80)90054-4。[48] A. Singh, B. Pandey，一种基于欧几里得距离的KNN计算方法，用于评估肝损伤程度，载于：2016年发明计算技术国际会议（ICICT），2016：第1–4页。https://doi.org/10.1109/INVENTIVE.2016.7823222。[49] I.-C. Yeh，利用人工神经网络对高性能混凝土强度进行建模，《水泥与混凝土研究》28 (1998) 1797–1808。

[50] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, L. Shao, 《深度神经网络训练中的归一化技术：方法论、分析与应用》, 《IEEE模式分析与机器智能汇刊》45（2023年）10173–10196。https://doi.org/10.1109/TPAMI.2023.3250241。[51] K. He, X. Zhang, S. Ren, J. Sun, 《用于图像识别的深度残差学习》, 载于：2016 IEEE计算机视觉与模式识别会议（CVPR）, 2016年：第770–778页。https://doi.org/10.1109/CVPR.2016.90。[52] Y.A. LeCun, L. Bottou, G.B. Orr, K.-R. Müller, 《高效反向传播》, 载于：G. Montavon, G.B. Orr, K.- R. Müller（编）, 《神经网络：行业诀窍（第二版）》, 施普林格，柏林、海德堡，2012年：第9–48页。https://doi.org/10.1007/978-3-642-35289-8_3。[53] A. Krizhevsky, 《从微型图像中学习多层特征》, 技术报告0，多伦多大学，多伦多，安大略省（2009年）。[54] T. Han, A. Siddique, K. Khayat, J. Huang, A. Kumar, 《一种用于预测和优化再生骨料混凝土弹性模量的集成机器学习方法》, Construction and Building Materials 244（2020年）118271。https://doi.org/10.1016/j.conbuildmat.2020.118271。[55] T. Han, R. Bhat, S.A. Ponduru, A. Sarkar, J. Huang, G. Sant, H. Ma, N. Neithalath, A. Kumar, 《利用深度学习预测含粉煤灰胶凝材料的水化与性能》, 《水泥与混凝土研究》165（2023年）107093。https://doi.org/10.1016/j.cemconres.2023.107093。[56] L. Breiman, 《装袋预测器》, Machine Learning 24（1996年）123–140。https://doi.org/10.1007/BF00058655。[57] L. Breiman, 《随机森林》, Machine Learning 45（2001年）5–32。https://doi.org/10.1023/A:1010933404324。[58] T. Han, B.K. Aylas-Paredes, J. Huang, A. Goel, N. Neithalath, A. Kumar, 《石灰石煅烧黏土水泥力学性能的预测：一种针对水泥化学定制的随机森林方法》, 《矿物》13（2023年）1261。https://doi.org/10.3390/min13101261。[59] T.K. Rusch, M.M. Bronstein, S. Mishra, 《图神经网络过平滑现象综述》（2023年）。https://doi.org/10.48550/arXiv.2303.10993。[60] D. Grattarola, D. Zambon, F.M. Bianchi, C. Alippi, 《理解图神经网络中的池化》, IEEE Transactions on Neural Networks and Learning Systems 35（2024年）2708–2718。https://doi.org/10.1109/TNNLS.2022.3190922。[61] U. Alon, E. Yahav, 《论图神经网络的瓶颈及其实际影响》（2021年）。https://doi.org/10.48550/arXiv.2006.05205。[62] D. Mesquita, A.H. Souza, S. Kaski, 《重新思考图神经网络中的池化》（2020年）。https://doi.org/10.48550/arXiv.2010.11418。[63] H. Yuan, Y. Liu, Y. Zhang, X. Ai, Q. Wang, C. Chen, Y. Gu, G. Yu, 《图神经网络训练系统的全面评估：数据管理视角》, Proceeding of the VLDB Endowment 17（2024年）1241–1254。https://doi.org/10.14778/3648160.3648167。[64] H.T. Kose, J. Nunez-Yanez, R. Piechocki, J. Pope, 《面向可重构系统的计算高效图神经网络综述》, 《信息》15（2024年）377。https://doi.org/10.3390/info15070377。