

水工常态混凝土多目标性能概率预测及可信度评价

徐潇航¹, 储大勇¹, 田 丹², 杨华美², 胡张莉¹, 刘加平¹

(1. 东南大学材料科学与工程学院, 南京 211189; 2. 中国长江三峡集团有限公司水工程材料与应用技术
湖北省重点实验室, 武汉 430010)

摘 要: 水工工程常态混凝土配合比设计需同时满足施工工作性、早龄期强度发展、标准龄期强度达标和体积稳定性等要求, 现有预测方法主要以绝对数值预测, 难以面向材料非均质性、施工扰动和试验误差引起的各类性能波动。本工作构建了包含 457 组样本的水工常态混凝土多目标性能数据库, 以水泥、粉煤灰、骨料等原材料信息及配合比等 45 维变量为输入, 以坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩为输出目标; 提出分层多任务异方差概率神经网络模型(HMT-HPNN), 该模型通过共享表征、层级任务表征、任务特异分支和异方差概率子网络, 实现多目标均值预测与样本相关不确定性估计。结果表明, HMT-HPNN 的点预测精度与经典神经网络模型处于同一较优水平, 各模型在 4 个预测目标上的平均 NRMSE 均小于 0.078, 具备可靠的基础回归能力; 与基于蒙特卡洛 Dropout 的多层感知机(MLP+MC Dropout)相比, HMT-HPNN 表现出更优的不确定性表征能力, 各目标平均 90% 预测区间覆盖率(PICP90)与 90% 名义置信水平更为接近, 总体覆盖率达 90.49%; 4 个目标的负对数似然(NLL)均更低, 降低幅度均超过 40%; 同时, HMT-HPNN 在测试集中综合预测误差随综合预测不确定性等级升高而递增, 模型输出的不确定性具有误差风险指示能力; 针对 46 组同源工程验证数据, HMT-HPNN 对 4 个目标的 R^2 分别为 0.70、0.94、0.90 和 0.87, 平均 NLL 优于测试集结果, 且不确定性等级与预测误差间具有一致性。本工作可为水工常态混凝土性能预判、配合比设计和性能风险感知提供数据驱动支撑。

关键词: 概率神经网络; 多目标预测; 水工常态混凝土; 不确定性; 可信度

中图分类号: TU528 文献标志码: A 文章编号: 0454-5648(2026)10-3464-19

网络出版时间: 2026-09-03



水利水电工程中, 水工混凝土需在施工与服役过程中满足工作性、强度、耐久性及抗裂等多方面要求^[1-3]。水工常态混凝土是水工工程中的重要结构材料, 其配合比设计不仅要保证新拌混凝土在运输、浇筑和振捣过程中的施工适应性, 还要兼顾早龄期强度发展、标准龄期强度达标以及硬化后的体积变形控制。上述性能分别对应施工成型、分层施工与早期承载、强度等级评定及长期服役稳定性等关键工程需求, 且在实际设计中通常受到目标值、允许波动范围和强度保证率等约束。《水工混凝土配合比设计规程》(DL/T 5330—2015)将常态混凝土界定为混凝土拌和物坍落度为 10~100 mm 的混凝土, 并要求配合比设计明确混凝土强度等级及保证率、工作性等基本要求^[4]; 同时, 水工混凝土配合比需通过试配、调整 and 性能检验确定满足设计与施工要求

的配合比^[1, 2]。同时, 《水工混凝土施工规范》(SL 677—2014)要求水工混凝土满足抗裂等设计要求, 并提出应从原材料选择、配合比设计、施工安排、温度控制、养护和表面保温等方面采取综合措施, 防止混凝土裂缝^[2]。因此, 面向水工工程的水工常态混凝土性能设计并非单一性能预测问题, 而是需要对坍落度、早龄期抗压强度、标准龄期抗压强度和干燥收缩等关键指标进行协同预测与综合控制; 其中, 干燥收缩用于表征硬化混凝土的收缩变形特征及潜在抗裂风险^[2, 4-7]。

近年来, 机器学习为水泥基材料复杂性能预测提供了新的方法基础。混凝土性能受原材料组成、配合比参数、龄期条件和工程来源等因素共同影响, 组分-性能关系通常表现出显著的非线性、多变量耦合和数据离散性。相关研究表明, 机器学习已广

收稿日期: 2026-05-15。 修订日期: 2026-07-04。

基金项目: 中国长江三峡集团有限公司科研项目(202203397); 湖北省科技创新人才计划项目(2024DJA036); 江苏省研究生科研与实践创新计划项目(KYCX24_0384)。

第一作者: 徐潇航(1998—), 男, 博士研究生。

通信作者: 杨华美(1986—), 女, 博士, 副教授。

Received date: 2026-05-15. Revised date: 2026-07-04.

First author: XU Xiaohang (1998—), male, Doctoral candidate.

E-mail: seumaturity@163.com

Correspondent author: YANG Huamei (1986—), female, Ph.D., Associate Professor.

E-mail: yang_huamei@ctg.com.cn

泛应用于混凝土科学中的性能预测、配合比设计和材料行为分析,为复杂水泥基材料建模提供了有效的数据驱动工具^[8-12]。在具体性能预测方面,已有研究围绕多类混凝土的抗压强度、坍落度、干燥收缩及水工混凝土力学性能建立了人工神经网络、支持向量机、随机森林、梯度提升树和深度学习等模型,验证了机器学习在单一或特定性能预测中的有效性^[11-17]。随着工程应用对多性能协同服役的要求不断提高,多目标回归、多输出神经网络及多性能预测模型逐渐被用于混凝土强度、耐久性和工作性等指标的同步预测,以利用不同性能目标之间潜在的相关性,并减少多个独立模型建模造成的信息割裂^[18-23]。然而,多目标同步预测若仍只给出确定性数值结果,则难以反映在相近配合比、原材料条件和试验条件下,由材料非均质性、拌和与成型扰动、测试误差及数据分布差异引起的性能波动范围。因此,预测不确定性可理解为模型在给出性能预测值的同时,对预测结果可能偏离真实值的程度或范围进行量化,通常可通过预测标准差、预测区间及其覆盖情况进行表征。现有研究通过蒙特卡洛 Dropout、异方差不确定性建模、深度集成和校准/保形预测等方法,为预测模型由确定性点估计向不确定性量化发展提供了技术基础^[24-30],使混凝土性能预测从关注预测值是否准确进一步拓展至预测结果是否可靠。

然而,现有研究在面向水工常态混凝土的多目标可信预测方面仍存在不足。首先,工程标准和实际配合比设计通常以目标值、允许波动范围及强度保证率等形式控制水工混凝土性能,而多数机器学习模型仍以单一数值点预测为主,难以表征材料非均质性、施工扰动和测试误差引起的性能离散性^[24-30]。其次,已有混凝土多性能预测研究多采用独立模型分别建模,或通过简单并列输出结构实现多目标预测,难以充分反映不同性能指标之间既相互关联又具有差异化主控机制的材料特征^[4-7, 18-23]。最后,水工混凝土数据通常具有样本规模有限、工程来源集中和分布差异显著等特点^[8, 12, 13],仅依赖常规随机划分评价模型性能,难以充分反映模型在同源工程条件下的泛化能力、预测区间覆盖可靠性及不确定性-误差一致性^[28-30]。

基于此,本工作面向水工常态混凝土,建立多目标性能数据库,输入特征包括水泥的物理化学性质、粉煤灰的物理化学性质、骨料的物理化学性质和配合比参数 4 类,共 45 维原始变量,输出目标包括坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干

燥收缩。针对水工常态混凝土多性能协同预测与结果可信度量化需求,构建分层多任务异方差概率神经网络(HMT-HPNN),通过共享表征、层级任务表征、任务特异分支和异方差概率子网络,同时实现四个目标的均值预测与样本相关不确定性估计。为系统评价模型性能,本工作将 HMT-HPNN 与单目标多层感知机(Single-target MLP)、多输出多层感知机(Multi-output MLP)和基于蒙特卡洛 Dropout 的多层感知机(MLP+MC Dropout)进行对比,从预测准确性、不确定性、区间覆盖能力及可信度等方面开展评价,并基于 46 组水工常态混凝土同源工程验证数据进一步检验 HMT-HPNN 模型的工程泛化能力。本工作旨在形成一种兼顾材料多目标性能预测和可信度评价的智能建模方法,为面向重力坝工程水工常态混凝土的性能预判、配合比设计和性能风险感知提供数据驱动支撑。

1 水工常态混凝土数据库建立

本工作构建了面向工程配合比设计与性能表征的水工常态混凝土数据库。数据库样本来源于水工工程中掺粉煤灰水工常态混凝土的配合比设计、原材料性能测试及相应力学、变形与工作性能试验结果。数据库共包含 457 组样本,每组样本由输入特征和输出目标两部分组成。其中,输入端主要描述原材料组成、材料物理化学性质和配合比参数;输出端则对应水工常态混凝土设计与施工中重点关注的关键性能指标。

1.1 特征体系与目标分布

水工常态混凝土性能受胶凝材料组成、配合比参数、骨料性质和外加剂作用等因素共同控制^[1, 4-7]。为表征材料组成与性能响应关系,本工作将数据库输入变量划分为水泥物理化学相关变量、粉煤灰物理化学相关变量、骨料物理化学相关变量和配合比变量 4 类,共 45 维原始输入特征。水泥相关变量包括强度等级、比表面积、密度、矿物组成、化学组成、水化热及 28 d 胶砂力学性能等,用于表征水泥熟料矿物、水化活性和早期放热特征等;粉煤灰相关变量包括细度、密度、需水量比、烧失量及化学组成等,用于描述粉煤灰颗粒特性、活性组分和对拌合物需水性的影响;配合比变量包括水、水泥、粉煤灰、细骨料、粗骨料、减水剂、引气剂掺量等,直接反映单位体积混凝土中各组分用量;骨料相关变量包括细度模数、表观密度、吸水率等,用于表征骨料级配、体积占比和吸水特性等工作性、强

度和体积变形的影响。上述特征体系基本覆盖了水工常态混凝土配合比设计中影响拌合物状态、硬化强度和变形性能的主要材料因素，为后续建模提供了输入基础^[1, 3, 4–7]。

本工作选取坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩作为输出目标。4 个性能目标能够较完整地反映水工常态混凝土配合比设计中从新拌状态到硬化性能的关键需求^[1–5, 15, 16]。图 1 给出了各目标性能在样本全集上的分布特征。坍落度主要分布在 40~80 mm 范围内，整体处于水工常态混凝土常见工作性区间，均值与中位数较为接近，说明数据库样本以适用于常态浇筑施工的低至中等

流动性混凝土为主；7 d 抗压强度主要集中在 10~35 MPa 范围内，并向高强度区间形成一定延伸，说明数据库覆盖了不同早期强度发展水平的配合比；28 d 抗压强度主要集中在 25~45 MPa 之间，整体呈连续单峰分布，表明数据库覆盖了水工常态混凝土中较典型的标准龄期强度水平，同时包含一定数量的高强度样本；28 d 干燥收缩主要集中在 280~380 $\mu\text{ε}$ 范围内，同时低收缩区间仍有一定样本分布，整体呈现出较强的离散性和非对称特征。综合来看，4 个目标在量纲、数值范围和分布形态上存在明显差异，同时目标分布中的离散性和局部偏态也说明水工常态混凝土各性能响应具有波动特征。

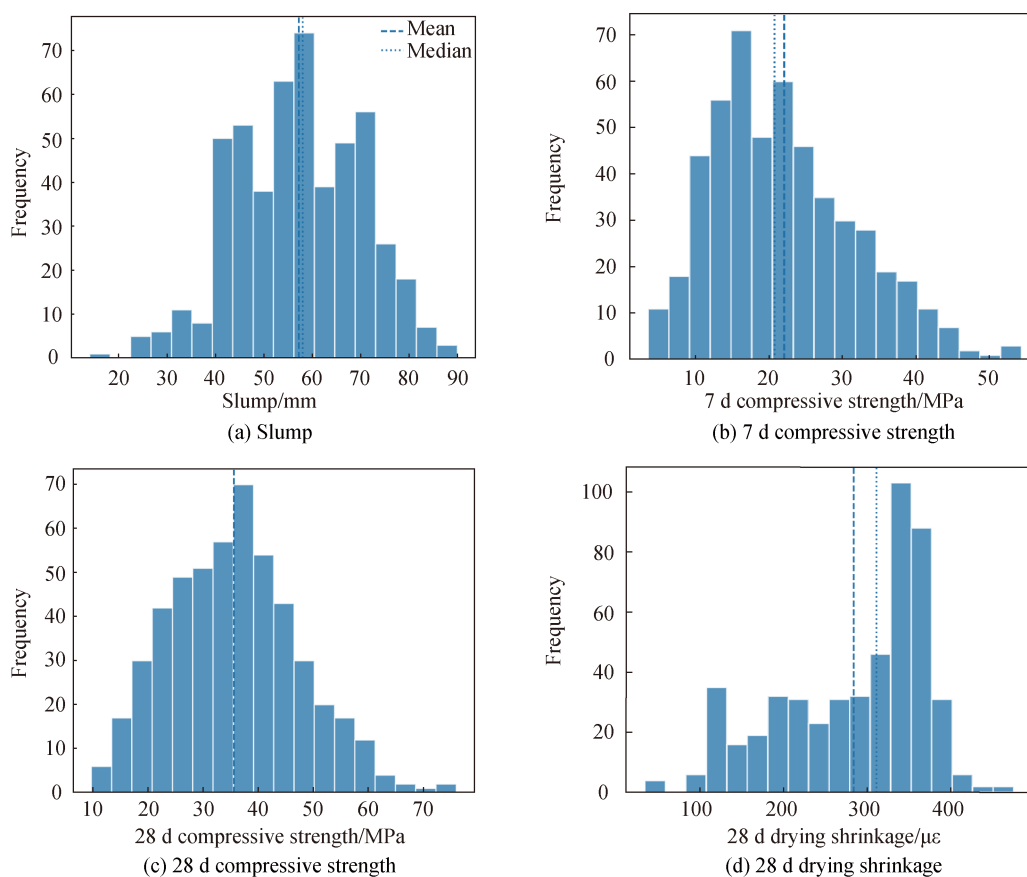


图 1 目标性能的分布特征

Fig. 1 Distribution characteristics of the target properties

1.2 数据相关性分析

为分析数据库中目标性能之间的关联特征，以及输入变量与输出目标之间的初步响应关系，本工作采用 Pearson 相关系数对数据库进行相关性分析，Pearson 相关系数主要反映变量间的线性统计关系，进行相关性分析主要是为了揭示数据库的基本结构特征，并为后续模型建立提供参考^[31, 32]。

图 2 给出了各目标性能间的相关性矩阵。7 d

抗压强度与 28 d 抗压强度之间的相关系数为 0.92，呈现显著正相关关系，7 d 强度主要反映早期水化产物生成和初始承载骨架形成程度，28 d 强度则进一步受持续水化、粉煤灰火山灰反应及孔结构细化过程的影响，二者高度相关表明强度发展具有明确的时序连续性，这为后续 HMT-HPNN 模型中采用共享表征提供依据^[4–7, 18–23]；坍落度与 7 d 抗压强度、28 d 抗压强度之间的相关系数分别为 0.14 和 0.13，

均处于弱相关水平, 坍落度与抗压强度虽在水胶比和胶凝材料组成层面存在一定联系, 但其性能影响机制并不完全一致, 因此未表现出明显线性相关性; 28 d 干燥收缩与坍落度、7 d 抗压强度和 28 d 抗压强度的相关系数分别为 0.27、0.19 和 0.21, 说明干燥收缩与工作性及强度目标之间也仅存在弱线性关系, 干燥收缩受孔结构、水分迁移、骨料约束、胶凝材料组成等因素共同影响, 其变化不能简单由强度或坍落度表征^[4-7, 15, 16]。

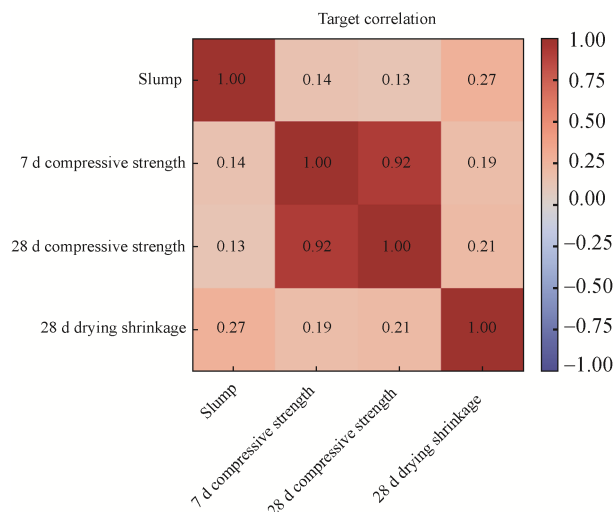


图 2 性能目标间相关性矩阵

Fig. 2 Correlation matrix among the target properties

图 3 展示了输入特征与各目标性能间相关性较高的变量。总体来看, 不同目标对应的主要相关变量存在明显差异, 说明水工常态混凝土工作性、强

度发展和体积稳定性受不同材料因素控制。对于坍落度, 骨料吸水率、单位用水量、粉煤灰需水量比和减水剂掺量等变量表现出较高相关性, 骨料吸水率会改变拌合物中的有效自由水含量, 并影响浆体对骨料表面的包裹状态^[4-7]; 单位用水量和减水剂掺量直接调节浆体流动性及颗粒分散程度; 粉煤灰需水量比则反映粉煤灰颗粒形貌、细度和表面性质对拌合物需水性的影响。对于 7 d 和 28 d 抗压强度, 水泥用量、7 d 水化热、水泥 SO_3 含量等变量与强度目标呈现较明显正相关, 水泥用量增加通常提高单位体积内水化产物生成潜力; 水化热反映胶凝体系早期反应活性, 对早期强度发展及后续结构形成具有指示作用^[6, 7, 11, 13]; 水泥 SO_3 含量与铝酸盐相水化调控和早期凝结硬化过程有关; 粉煤灰的密度、碱含量、需水量比和部分化学组成等变量与强度表现出负相关或弱负相关, 说明粉煤灰在水工常态混凝土中具有双重作用: 一方面, 可降低水化热, 改善后期结构致密化并提升大体积混凝土温控适应性; 另一方面, 较高掺量或较低活性粉煤灰可能在早期产生稀释效应, 从而降低早期强度水平。对于 28 d 干燥收缩, 单位用水量、水泥用量和骨料吸水率等变量与其存在较明显关联, 单位用水量增加会提高硬化浆体中可蒸发水含量及毛细孔收缩潜势; 水泥用量增加通常意味着浆体体积比例提高, 在骨料约束相对不足时可能增大体积变形敏感性; 骨料吸水率则可能通过影响有效水胶比、内部水分交换和界面过渡区性质, 对收缩变形产生作用^[4, 5, 15, 16]。

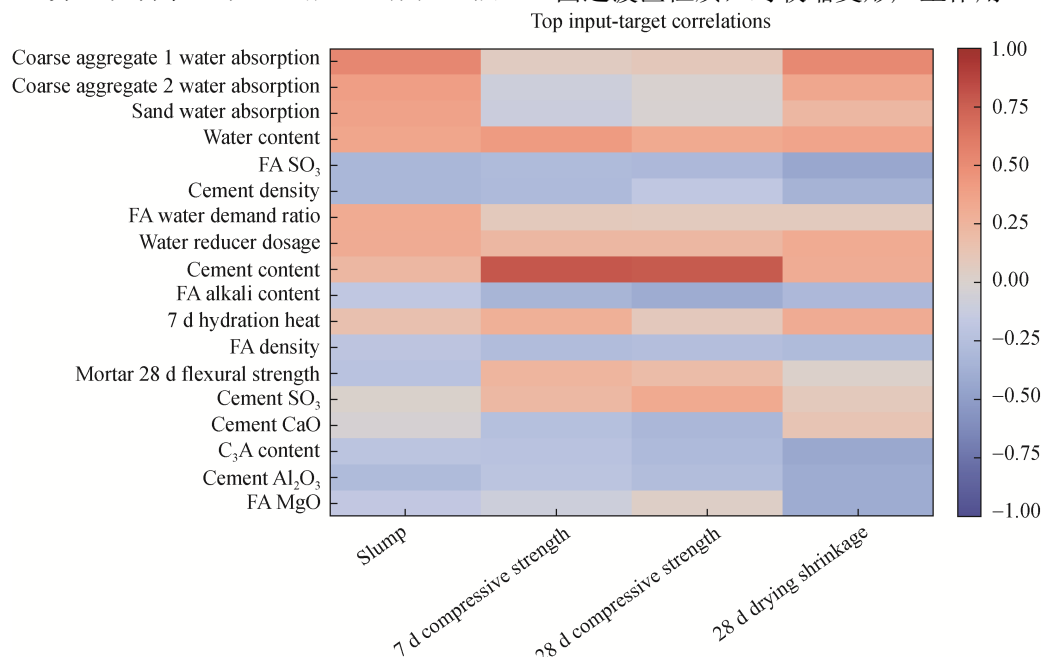


图 3 输入特征与输出目标间 Top 相关性矩阵

Fig. 3 Top correlation matrix between input features and output targets

1.3 数据预处理

本工作所采用的模型均为基于梯度优化的神经网络模型^[32, 33]。由于数据库中输入特征和输出目标具有不同量纲与数量级,若直接采用原始数据训练,较大尺度变量易在梯度更新和损失函数中占据主导,导致模型收敛不稳定,并影响多目标损失的均衡优化。因此,在构建各模型前需对输入特征和输出目标均进行标准化处理。对任一变量 z , 其标准化形式为:

$$z' = \frac{z - \mu_{\text{train}}}{s_{\text{train}}} \quad (1)$$

式中: z 和 z' 分别为标准化前后的变量值; μ_{train} 和 s_{train} 分别为该变量在训练集中的均值和标准差。测试集和工程验证集均采用训练集的均值和标准差进行变换,模型训练、交叉验证和超参数搜索均在标

准化空间中完成,模型预测结果在评价前反标准化至原始物理量纲^[32, 33]。

2 研究流程及模型建立

本工作具体流程如图 4 所示。首先,基于 457 组水工常态混凝土多目标性能数据库构建开发集,输入特征与输出目标同 1.1 节中所阐述。开发集按 8:2 比例划分为 365 组训练集和 92 组测试集,训练集用于模型训练、交叉验证和超参数优化,测试集用于最终模型准确性与不确定性评价;此外,46 组水工常态混凝土同源工程的验证数据用于检验 HMT-HPNN 在工程验证样本上的泛化能力。用于评价的测试集与同源工程验证集始终采用完整训练集的均值和标准差进行标准化变换,不参与任何模型训练、交叉验证与超参数搜索。

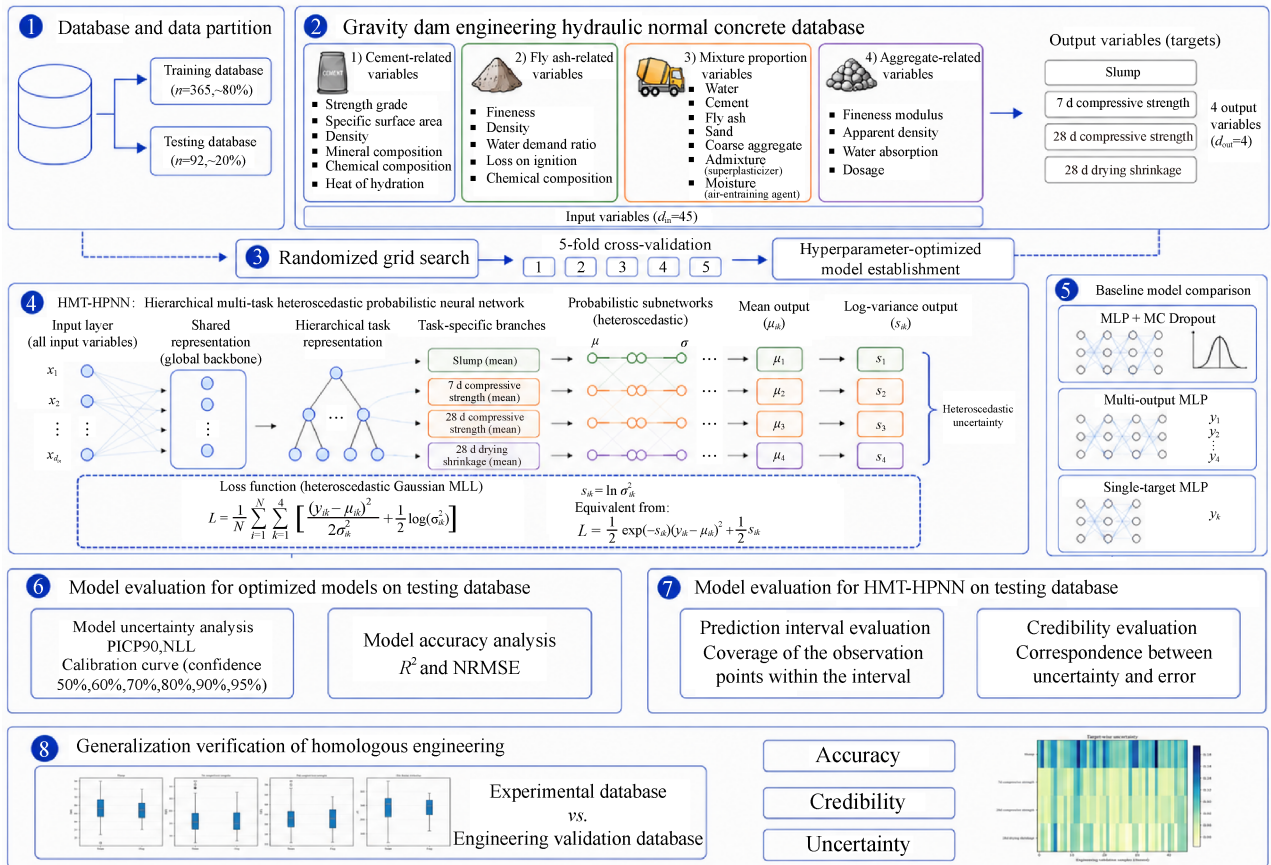


图 4 研究流程及 HMT-HPNN 网络结构

Fig. 4 Research workflow and network architecture of HMT-HPNN

在模型构建阶段,建立 HMT-HPNN,并选取 Single-target MLP、Multi-output MLP 和 MLP+MC Dropout 作为对比模型。通过与 Single-target MLP 及 Multi-output MLP 的对比,从独立目标映射、多输出联合映射的层面检验 HMT-HPNN 的预测准确

性;通过与 MLP+MC Dropout 的对比,从概率映射的层面检验 HMT-HPNN 的不确定性表征能力。为保证不同模型评价结果的可比性,所有模型均采用相同的数据划分方式和标准化处理流程。训练集内部进一步以 28 d 抗压强度分位数为依据建立固定 5

折分层交叉验证划分,用于随机网格搜索和模型选择^[34, 35];而为避免数据泄漏,每一折的标准化均值与标准差仅由该折对应的训练子集计算,并用于变换该折的验证子集,验证子集信息不参与标准化参数的估计。

完成超参数选择后,各模型均基于完整训练集重新训练最终模型,并在测试集上开展统一评价。评价内容包括 2 方面:1) 分析各模型的点预测精度;2) 针对 MLP+MC Dropout 和 HMT-HPNN,进一步采用 PICP90、NLL 和校准曲线评价预测不确定性质量。在确定 HMT-HPNN 预测精度与预测不确定性质量的基础上,针对 HMT-HPNN 构建 90%预测区间分析区间对于观察点的覆盖情况,并基于训练集 5 折内的预测结果确定低、中、高不确定性分级阈值,分析综合不确定性与综合误差间的一致性情况。最后,将训练完成的 HMT-HPNN 直接应用于同源工程验证集,从准确性、不确定性和可信度 3 个层面评价其工程泛化能力。

2.1 HMT-HPNN

2.1.1 概率神经网络 对于水工常态混凝土而言,即使配合比、原材料参数和试验条件相近,由于材料非均质性、拌和与成型扰动及测试误差等因素影响,各目标性能仍会存在一定离散性。因此,仅采用确定性点预测难以满足工程中对性能波动范围和结果可信度的评价需求^[24-30, 36]。

本工作采用异方差概率神经网络描述多目标性能预测问题。对于第 i 个样本和第 k 个目标,模型不再仅输出预测均值 $\mu_{i,k}$,而是同时输出预测方差 $\sigma_{i,k}^2$,并假定目标变量服从高斯分布:

$$y_{i,k} \sim \mathcal{N}(\mu_{i,k}, \sigma_{i,k}^2) \quad (2)$$

式中: $\mu_{i,k}$ 表示模型对目标性能的条件均值预测, $\sigma_{i,k}$ 表示该样本在对应目标上的预测标准差。为保证方差为正并提高训练稳定性,网络实际输出对数方差:

$$s_{i,k} = \ln \sigma_{i,k}^2 \quad (3)$$

模型训练采用异方差高斯负对数似然损失函数。忽略常数项后,其形式为:

$$\mathcal{L}_{\text{NLL}} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \left[\frac{1}{2} \exp(-s_{i,k}) (y_{i,k} - \mu_{i,k})^2 + \frac{1}{2} s_{i,k} \right] \quad (4)$$

式中: N 为样本数, $K=4$ 为输出目标数量。该损失函数同时约束预测误差和预测方差:当预测误差较大时,模型可通过增大方差来降低误差项惩罚,但

方差项本身也会受到 $\frac{1}{2} s_{i,k}$ 的约束,从而避免无意义地扩大预测区间。由此,模型能够在均值预测的基础上学习样本相关的不确定性^[26, 32, 36, 37]。

2.1.2 模型网络结构 HMT-HPNN 的网络结构详见图 4 中的模块 4。模型输入为 45 维水工常态混凝土特征变量,网络首先通过共享表征模块提取不同目标共同依赖的材料组成和配合比特征,该部分用于捕捉胶凝材料组成、骨料特性和外加剂作用等因素对多目标性能的共性影响,避免各目标完全独立建模造成信息割裂^[18-23]。在共享表征之后,模型设置层级任务表征模块,以进一步组织不同性能目标之间的关联关系,坍落度反映新拌混凝土工作性,7 d 抗压强度和 28 d 抗压强度共同表征强度发展过程,28 d 干燥收缩则对应硬化后体积稳定性,不同目标既受部分共同材料因素控制,又具有各自的主导机制。因此,层级任务表征模块在共享信息基础上保留任务差异,使模型能够同时利用目标间相关性和任务特异性。输出端设置 4 个任务特异分支,分别对应坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩,每个任务分支后连接异方差概率子网络,并输出该目标的预测均值 μ_k 和对数方差 s_k ,HMT-HPNN 对每个样本可同时获得 4 个目标的点预测结果及其对应的预测标准差。该结构不仅能进行多目标均值预测,还能针对不同样本和不同目标给出差异化不确定性估计。相比普通 Multi-output MLP,其关键区别在于:完全共享表征与完全任务特异分支间,引入了一个按目标物理属性分组的中间表征层级,并在输出端为每个目标配置异方差概率子网络。

通过上述结构,HMT-HPNN 将水工常态混凝土多目标性能预测转化为多任务异方差概率回归问题。模型输出的预测均值用于评价坍落度、抗压强度和干燥收缩的预测准确性;预测标准差则用于构建 90%预测区间、计算不确定性评价指标,并形成综合预测不确定性。

2.2 MLP+MC Dropout

MLP+MC Dropout 的优势在于实现简单,能够在不显著改变基础神经网络结构的条件下获得概率预测结果,为对比 HMT-HPNN 的概率预测能力,研究中选取这一经典概率预测神经网络,作为概率预测的基线^[24-26]。该模型在普通多层感知机中引入 Dropout 层,通过在训练阶段随机失活部分神经元降低模型对特定特征组合的依赖,从而改善模型泛

化能力。与常规 Dropout 仅在训练阶段启用不同, MC Dropout 在测试阶段仍保持 Dropout 激活, 并对同一样本进行多次随机前向传播, 由多次预测结果构成经验预测分布, 以近似表征模型预测不确定性^[24, 26]。

对于输入样本 x_i , 设模型第 t 次随机前向传播得到第 k 个目标的预测值为 $\hat{y}_{i,k}^{(t)}$, 共进行 T 次随机采样, 则预测均值可表示为:

$$\mu_{i,k} = \frac{1}{T} \sum_{t=1}^T \hat{y}_{i,k}^{(t)} \quad (5)$$

预测标准差可表示为:

$$\sigma_{i,k} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\hat{y}_{i,k}^{(t)} - \mu_{i,k})^2} \quad (6)$$

式中: $\mu_{i,k}$ 用于表征目标性能的点预测结果; $\sigma_{i,k}$ 用于表征由随机网络子结构引起的预测不确定性。研究中 MLP+MC Dropout 采用与多输出 MLP 相近的基础网络结构, 输入层接收 45 维特征变量, 输出层同时给出 4 个目标的预测值。训练阶段以标准化目标值的均方误差为损失函数, 测试阶段通过多次随机采样获得各目标的预测均值和标准差。

2.3 Single-target MLP

Single-target MLP 由输入层、若干隐藏层和单一输出神经元组成, 隐藏层通过非线性激活函数提取输入特征与目标性能之间的非线性关系。模型训练以均方误差作为损失函数:

$$\mathcal{L}_{\text{MSE},k} = \frac{1}{N} \sum_{i=1}^N (y_{i,k} - \hat{y}_{i,k})^2 \quad (7)$$

式中: $y_{i,k}$ 为第 i 个样本第 k 个目标的实测值, N 为训练样本数。由于每个目标均独立训练, Single-target MLP 不受其他目标损失项影响, 能够较充分地拟合单一性能目标的输入与输出关系^[32, 33]。Single-target MLP 结构清晰、目标函数单一, 在研究中作为确定性基线模型。该模型针对 4 个目标性能分别建立独立的多层感知机, 即每个模型仅学习 45 维输入特征与单一目标性能之间的映射关系。对于第 k 个目标, 其预测形式可表示为:

$$\hat{y}_{i,k} = f_k(x_i; \theta_k) \quad (8)$$

式中: x_i 为第 i 个样本的输入特征; $\hat{y}_{i,k}$ 为第 i 个样本第 k 个目标的预测值; $f_k(\cdot)$ 表示对应目标的 MLP 模型; θ_k 为该模型的可训练参数。

2.4 Multi-output MLP

Multi-output MLP 由输入层、共享隐藏层和输出层组成。共享隐藏层用于提取不同目标共同依赖的材料组成、配合比和骨料特征; 输出层同时给出

目标性能的预测值。模型训练采用目标均方误差的平均值作为损失函数:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K (y_{i,k} - \hat{y}_{i,k})^2 \quad (9)$$

式中: K 为输出目标数量。研究中采用 Multi-output MLP 作为确定性多目标基线模型^[18–23], 其与 Single-target MLP 分别建立 4 个独立模型不同, Multi-output MLP 采用一个统一网络同时预测 4 个目标性能。其预测形式可表示为:

$$\hat{\mathbf{y}}_i = f(\mathbf{x}_i; \theta) \quad (10)$$

式中: x_i 为第 i 个样本的 45 维输入特征, $\hat{\mathbf{y}}_i = [\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}, \hat{y}_{i,4}]$ 为 4 个目标的预测结果, $f(\cdot)$ 表示多输出 MLP 模型, θ 为网络可训练参数。

2.5 随机网格搜索

神经网络模型的预测性能受网络深度、隐藏层神经元数量、学习率、正则化参数、Dropout 比例、批量大小和训练轮数等超参数影响。为降低人为设定超参数对模型结果的影响, 本工作采用随机网格搜索对 Single-target MLP、Multi-output MLP、MLP+MC Dropout 和 HMT-HPNN 进行统一优化。随机网格搜索在预设超参数空间内随机抽取若干组候选组合, 相较于穷举网格搜索, 可在有限计算成本下覆盖更丰富的参数配置^[35]。

确定性模型 Single-target MLP 与 Multi-output MLP 以 5 折平均 NRMSE 指标作为随机网格搜索优化超参数的主要选择依据, 以保证模型在目标性能上的整体预测精度; 概率预测模型 HMT-HPNN 与 MLP+MC Dropout 在随机网格搜索优化超参数时的选择依据则综合考虑 NRMSE 和 NLL, 使模型在保持均值预测能力的同时具备较合理的不确定性表征能力。

2.6 评价指标

本工作从预测准确性、不确定性质量和可信度 3 个层面建立评价指标体系。准确性指标用于衡量模型预测均值与实测值之间的偏差; 不确定性指标用于评价概率模型预测区间和预测分布的合理性; 可信度指标用于分析模型输出的不确定性是否能够反映实际预测误差风险。

2.6.1 准确性 研究中采用决定系数(R^2)、平均绝对误差(MAE)和归一化均方根误差(NRMSE)评价模型点预测精度。对于第 k 个目标, R^2 和 MAE 分别定义为:

$$R_k^2 = 1 - \frac{\sum_{i=1}^N (y_{i,k} - \hat{y}_{i,k})^2}{\sum_{i=1}^N (y_{i,k} - \bar{y}_k)^2} \quad (11)$$

$$\text{MAE}_k = \frac{1}{N} \sum_{i=1}^N |y_{i,k} - \hat{y}_{i,k}| \quad (12)$$

式中: $y_{i,k}$ 和 $\hat{y}_{i,k}$ 分别为第 i 个样本第 k 个目标的实测值和预测值; $\bar{y}_k$ 为该目标实测值均值; N 为样本数量。 R^2 越接近 1, 说明模型对目标变量的解释能力越强; MAE 越小, 说明预测值与实测值之间的平均绝对偏差越低。

由于坍落度、抗压强度和干燥收缩的量纲及数值范围不同, 本工作进一步采用 NRMSE 衡量不同目标之间可比较的相对误差:

$$\text{NRMSE}_k = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_{i,k} - \hat{y}_{i,k})^2}}{y_{k,\max}^{\text{train}} - y_{k,\min}^{\text{train}}} \quad (13)$$

式中: $y_{k,\max}^{\text{train}}$ 和 $y_{k,\min}^{\text{train}}$ 分别为第 k 个目标在训练集中的最大值和最小值。NRMSE 消除了不同目标数值尺度的影响, 其数值越小表示模型预测误差越低。对于多目标模型, 本工作同时计算 4 个目标 NRMSE 的平均值, 用于评价模型整体准确性^[31, 32]。

2.6.2 确定性 对于 MLP+MC Dropout 和 HMT-HPNN 两类概率预测模型, 本工作采用 90% 预测区间覆盖率(PICP90)、负对数似然(NLL)和校准曲线评价不确定性质量。假定第 k 个目标的预测均值为 $\mu_{i,k}$, 预测标准差为 $\sigma_{i,k}$, 则 90% 预测区间可表示为:

$$[\mu_{i,k} - 1.645\sigma_{i,k}, \mu_{i,k} + 1.645\sigma_{i,k}] \quad (14)$$

PICP90 定义为真实值落入 90% 预测区间的样本比例:

$$\text{PICP90}_k = \frac{1}{N} \sum_{i=1}^N I(y_{i,k} \in [\mu_{i,k} - 1.645\sigma_{i,k}, \mu_{i,k} + 1.645\sigma_{i,k}]) \quad (15)$$

式中: $I(\cdot)$ 为指示函数。PICP90 越接近 0.90, 说明模型给出的 90% 预测区间与名义置信水平越一致; 明显低于 0.90 表示区间存在欠覆盖, 明显高于 0.90 则可能说明区间偏保守。90% 预测区间采用均值 $\pm 1.645\sigma$ 的形式, 其依据为高斯分布假设。对于 HMT-HPNN, 其基于异方差高斯负对数似然进行训练, 显式假设条件分布服从高斯分布, 预测均值 μ 与标准差 σ 对应分布参数, 区间构造与模型假设一致; 对于 MLP+MC Dropout, 其通过多次随机前向传播得到的预测分布, 在连续回归问题中可用高斯分布进行有效近似。

NLL 用于综合评价预测均值误差和预测方差合理性, 其表达式为:

$$\text{NLL}_k = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{2} \ln(2\pi\sigma_{i,k}^2) + \frac{(y_{i,k} - \mu_{i,k})^2}{2\sigma_{i,k}^2} \right] \quad (16)$$

NLL 能同时惩罚较大的预测误差和不合理的方差估计, 数值越低表示模型预测分布与实测数据越匹配^[24, 27-30, 36, 37]。

为进一步评价不同置信水平下预测区间的校准能力, 本工作绘制校准曲线。设置信水平为 $c = 50\%, 60\%, 70\%, 80\%, 90\%, 95\%$, 分别计算对应预测区间的经验覆盖率, 并与理想覆盖线进行比较。当校准曲线越接近理想对角线时, 说明模型在不同置信水平下的不确定性估计越可靠^[28, 36]。

2.6.3 可信度 为将多目标预测不确定性转化为样本级可信度评价指标, 本工作定义综合预测不确定性 (Composite uncertainty) 和综合预测误差 (Composite error)。对于第 i 个样本, 综合预测不确定性 U_i 由 4 个目标的预测标准差归一化合成:

$$U_i = \sqrt{\frac{1}{K} \sum_{k=1}^K \left(\frac{\sigma_{i,k}}{y_{k,\max}^{\text{train}} - y_{k,\min}^{\text{train}}} \right)^2}, K = 4 \quad (17)$$

该指标仅由模型输出的预测标准差计算得到, 不使用真实误差, 反映了模型对当前样本多目标预测结果的整体不确定程度。

综合预测误差 E_i 用于表征样本级真实预测偏差, 其定义为:

$$E_i = \sqrt{\frac{1}{K} \sum_{k=1}^K \left(\frac{|y_{i,k} - \mu_{i,k}|}{y_{k,\max}^{\text{train}} - y_{k,\min}^{\text{train}}} \right)^2}, K = 4 \quad (18)$$

该指标将 4 个目标的绝对误差经训练集目标范围归一化后合成为单一样本级误差标量, 数值越大表示该样本在多目标预测中的整体偏差越大。

本工作基于 HMT-HPNN 在训练集 5 折内预测中得到的 U_i 分布, 为考虑样本数量的均衡性, 取其 1/3 和 2/3 分位数作为低、中、高不确定性分组阈值, 并将该固定阈值应用于测试集和同源工程验证集。随后, 通过分析不同不确定性组的 E_i 变化, 以及 U_i 与 E_i 之间的一致性, 评价模型输出的不确定性是否具有误差风险指示能力^[24, 27-30, 36]。

3 结果与讨论

3.1 各模型性能预测的准确性分析

图 5 给出了优化后 Single-target MLP、Multi-output MLP、MLP+MC Dropout 和 HMT-HPNN 在训练集和测试集上的准确性评价结果。总体来看, 4

类神经网络模型均能较好地建立输入特征与各性能目标之间的非线性映射关系, 说明所构建的特征体

系能够有效表征水工常态混凝土工作性、强度发展和体积稳定性的主要影响因素。

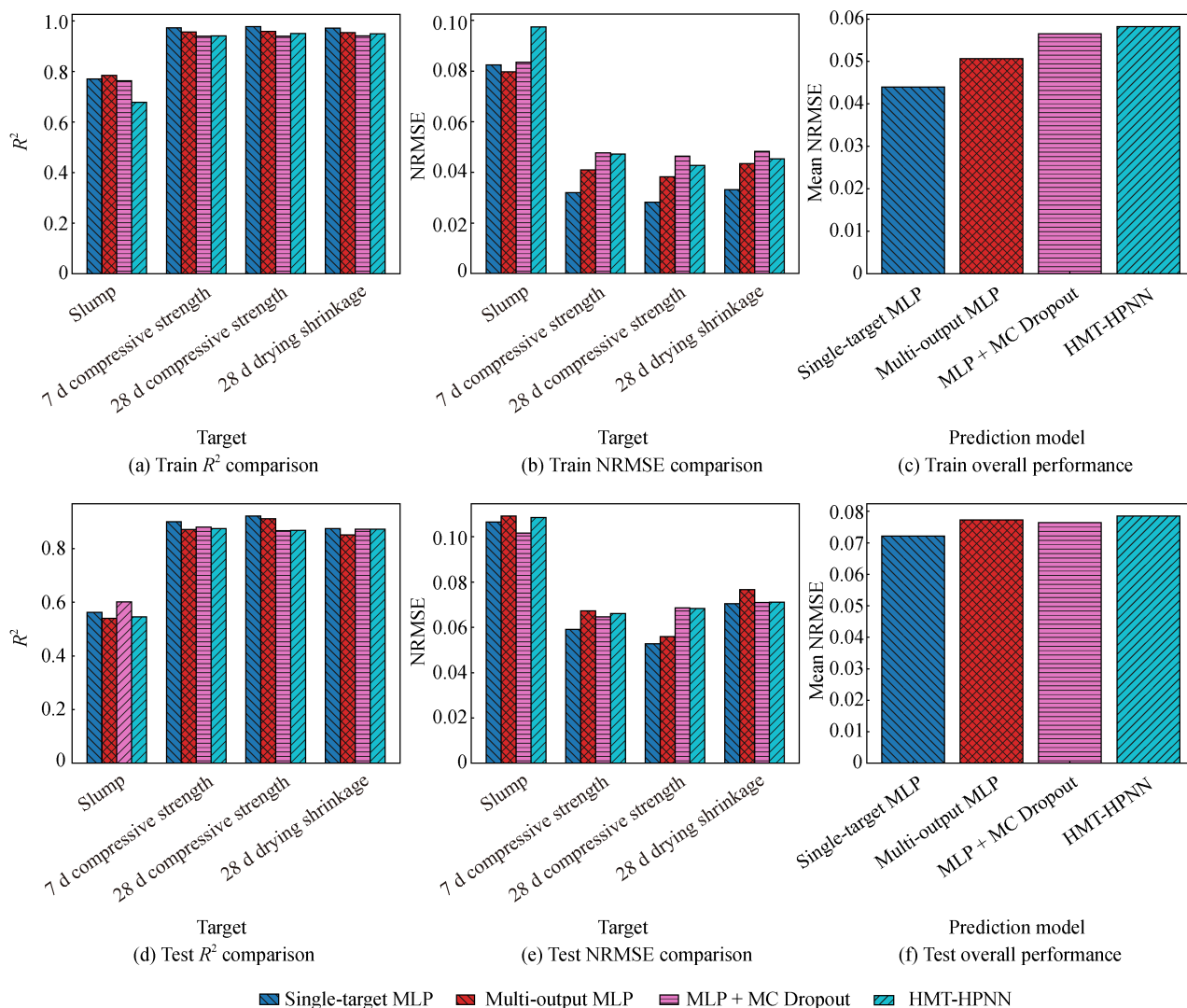


图5 优化后的各模型在训练集和测试集上的准确性对比
Fig. 5 Accuracy comparison of the optimized models on the training and testing sets

从整体误差水平看, Single-target MLP 在测试集上的平均 NRMSE 最低, 约为 0.072, 表现出较高的点预测精度。这与其独立建模方式有关, 即每个目标单独训练模型, 损失函数仅服务于单一性能指标, 因而能够较充分地拟合各目标自身的数据分布。Multi-output MLP、MLP + MC Dropout 和 HMT-HPNN 在测试集上的平均 NRMSE 分别约为 0.077、0.076 和 0.078, 整体差异较小, 均处于同一误差水平。这表明 HMT-HPNN 虽然并非点预测精度最优模型, 但其在引入异方差概率输出后仍保持与常规确定性神经网络相近的基础回归能力, 满足后续开展不确定性分析和可信度评价的精度前提。

从不同目标看, 7 d 抗压强度、28 d 抗压强度和

28 d 干燥收缩的预测结果整体优于坍落度。HMT-HPNN 在测试集上对 7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩的 R^2 分别约为 0.875、0.868 和 0.872, 对应 NRMSE 分别约为 0.066、0.068 和 0.071, 说明该模型能够较稳定地捕捉胶凝材料组成、配合比参数和骨料性质对强度发展及体积变形的影响。

相比之下, 坍落度是 4 个目标中预测难度较高的指标。4 类模型在测试集上的坍落度 R^2 约为 0.54~0.60, NRMSE 约为 0.102~0.109, 明显低于强度和干燥收缩目标。坍落度属于新拌混凝土工作性指标, 除受水胶比、单位用水量、砂率、骨料级配和外加剂掺量影响外, 还对拌和均匀性和施工扰动

等较为敏感,上述因素中部分变量难以在工程数据库中完全量化,因此水工常态混凝土工作性目标通常表现出更高的数据离散性和预测不确定性。

训练集与测试集对比结果进一步反映了各模型的泛化特征。*Single-target MLP* 在训练集上的拟合误差最小,但测试集误差有所增加,说明独立模型在提高训练拟合能力的同时,对训练样本分布具有更强依赖性。*HMT-HPNN* 训练集平均 *NRMSE* 约为 0.058,测试集平均 *NRMSE* 约为 0.078,训练集与测试集误差对应关系较为稳定,未表现出明显异常过拟合。该结果说明,*HMT-HPNN* 在共享表征、层级任务表征和异方差概率输出的共同约束下,能够在保持多目标均值预测能力的同时维持较好的泛化稳定性。

综合来看,4类模型均可实现水工常态混凝土各目标性能的有效预测。*HMT-HPNN* 的点预测精度虽不是所有模型中最优,但与 *Single-target MLP* 和 *Multi-output MLP* 差异较小,已具备可靠的基础预

测能力。

3.2 概率预测模型的不确定性分析

在准确性分析基础上,进一步比较 *MLP+MC Dropout* 和 *HMT-HPNN* 两类概率预测模型的不确定性表征能力。*MLP+MC Dropout* 通过测试阶段随机失活形成多个子网络,并利用预测结果离散性近似模型不确定性,主要反映参数扰动效应;*HMT-HPNN* 通过异方差学习建立输入特征与预测方差的映射,实现样本条件相关的不确定性建模。通过 *PICP90*、*NLL* 及校准曲线差异,对比两类概率预测模型的整体表现。

图6给出了两类模型在测试集上的校准曲线,表1列出了对应的 *PICP90* 和 *NLL* 指标。总体来看,*HMT-HPNN* 在目标性能上的预测区间覆盖率和概率分布拟合质量整体优于 *MLP+MC Dropout*,说明 *HMT-HPNN* 的异方差概率输出能够更有效地刻画水工常态混凝土多目标性能预测中的样本相关不确定性。

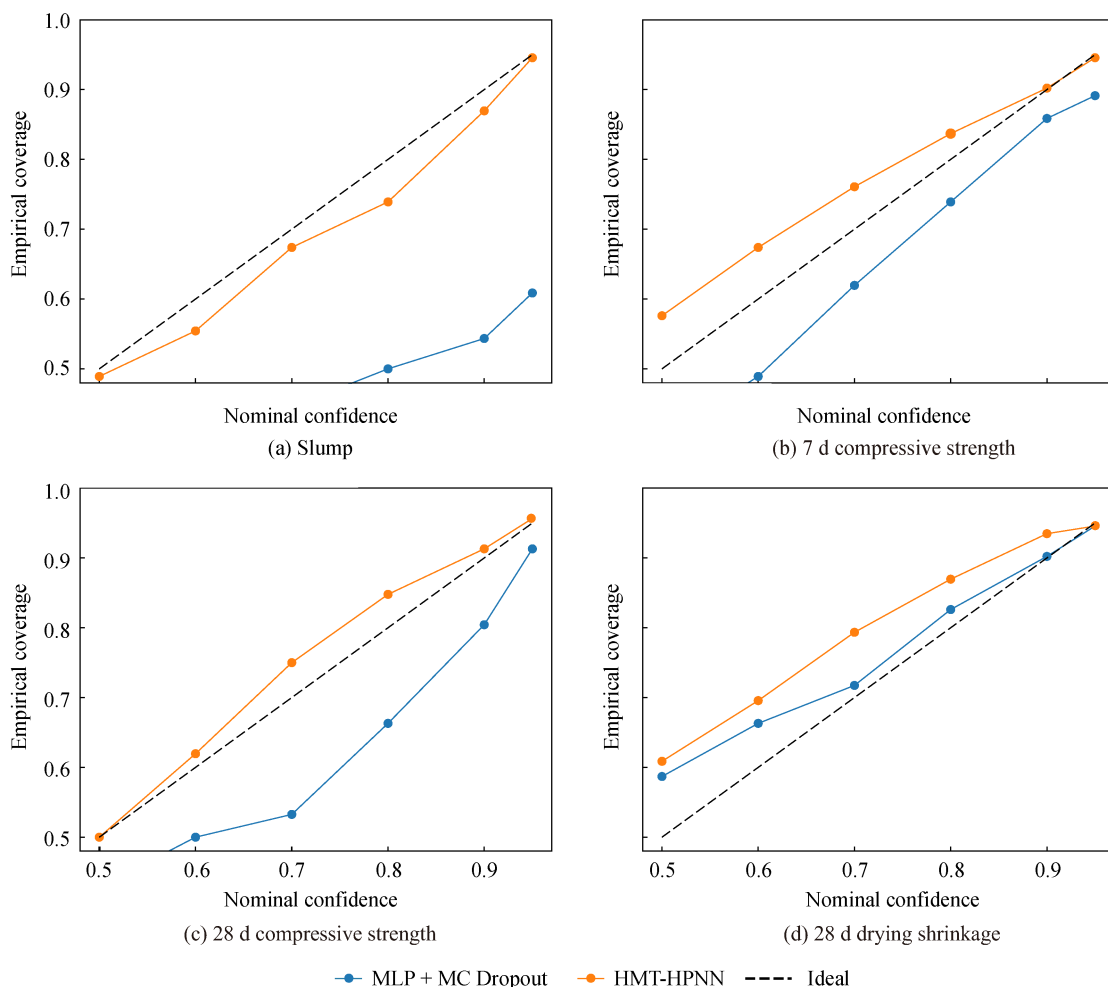


图6 各概率预测模型的不确定性校准曲线

Fig. 6 Uncertainty calibration curves of the probabilistic prediction models

表 1 各概率预测模型的不确定性指标量化对比

Table 1 Quantitative comparison of uncertainty metrics for the probabilistic prediction models

Probability prediction model	Target	PICP90	NLL
MLP + MC Dropout	Slump	54.35%	1.625 3
MLP + MC Dropout	7 d compressive strength	85.87%	0.326 2
MLP + MC Dropout	28 d compressive strength	80.43%	0.537 7
MLP + MC Dropout	28 d drying shrinkage	90.22%	0.251 4
HMT-HPNN	Slump	86.96%	0.959 4
HMT-HPNN	7 d compressive strength	90.22%	0.198 3
HMT-HPNN	28 d compressive strength	91.30%	0.270 2
HMT-HPNN	28 d drying shrinkage	93.48%	0.142 5

MLP + MC Dropout—multilayer perceptron with Monte Carlo Dropout; HMT-HPNN—hierarchical multi-task heteroscedastic probabilistic neural network; PICP90—coverage probability of the 90% prediction interval; NLL—negative log-likelihood.

从 PICP90 结果看, MLP+MC Dropout 在坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩上的覆盖率分别为 54.35%、85.87%、80.43% 和 90.22%, 除干燥收缩外均低于 90% 名义置信水平, 表明其预测区间存在不同程度的欠覆盖。HMT-HPNN 对应覆盖率分别为 86.96%、90.22%、91.30% 和 93.48%, 整体更接近 90% 名义置信水平。其中, 7 d 和 28 d 抗压强度的覆盖率与目标置信水平匹配较好, 说明模型能够较合理地表征强度类目标的预测波动范围; 坍落度覆盖率虽略低于 90%, 但较 MLP+MC Dropout 明显提高, 反映出异方差建模对工作性目标不确定性识别的改善。

从 NLL 结果看, HMT-HPNN 在 4 个目标上的 NLL 分别为 0.959 4、0.198 3、0.270 2 和 0.142 5, 均低于 MLP+MC Dropout 的 1.625 3、0.326 2、0.537 7 和 0.251 4。NLL 同时受预测均值误差和预测方差合理性影响, 数值降低说明 HMT-HPNN 并非单纯通过扩大区间提高覆盖率, 而是在预测均值与预测方差之间形成了更合理的平衡。对于 28 d 干燥收缩, 两类模型的校准曲线均较接近理想线, HMT-HPNN 在部分置信水平下表现出略偏保守的区间估计, 但其不确定性量化结果仍保持较优结果, 说明该目标的概率预测质量总体稳定。

校准曲线进一步表明, HMT-HPNN 在多数置信水平下的经验覆盖率更接近理想对角线, 尤其在 7 d 抗压强度和 28 d 抗压强度上表现出较好的校准一致性。相比之下, MLP + MC Dropout 的校准曲线在多个目标上位于理想线下方, 说明其由随机 Dropout 采样得到的预测标准差不足以充分覆盖实测波动。该差异与两类模型的不确定性建模机制有关: MLP+MC Dropout 的不确定性主要来自测试阶段随

机网络子结构采样, 更偏向于表征模型参数扰动; HMT-HPNN 则直接学习输入样本与目标方差之间的映射关系, 能够针对不同配合比样本和不同性能目标输出差异化标准差, 因此更适合描述水工常态混凝土中由材料组成、性能响应等差异引起的异方差特征。

综合 PICP90、NLL 和校准曲线结果可见, HMT-HPNN 在保持可接受点预测精度的基础上, 进一步提高了预测区间覆盖能力和概率校准质量。该结果说明, HMT-HPNN 的优势主要体现在不确定性量化层面, 其输出的预测标准差具有较好的统计合理性。

3.3 HMT-HPNN 模型的预测区间及可信度评价

在前述准确性和不确定性分析基础上, 进一步对 HMT-HPNN 的预测区间及可信度评价能力进行分析。图 7 给出了 HMT-HPNN 在测试集上对各性能目标的 90% 预测区间, 表 2 汇总了各目标的区间覆盖情况。整体来看, 4 个目标共 368 个观测值中有 333 个落入 90% 预测区间内, 总体覆盖率为 90.49%, 与 90% 名义置信水平基本一致, 说明 HMT-HPNN 输出的异方差不确定性能够较好地转化为具有统计意义的区间预测结果。

从单目标结果看, 坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩在预测区间上观察点的覆盖率分别为 86.96%、90.22%、91.30% 和 93.48%。其中, 7 d 和 28 d 抗压强度的覆盖率与名义水平较为接近, 表明模型对强度类目标的区间估计具有较好的校准性。坍落度覆盖率略低于 90%, 同样反映出坍落度的部分波动难以由数据库变量完全表征。28 d 干燥收缩覆盖率略高于 90%, 说明模型对体积稳定性目标的区间估计相对稳健。

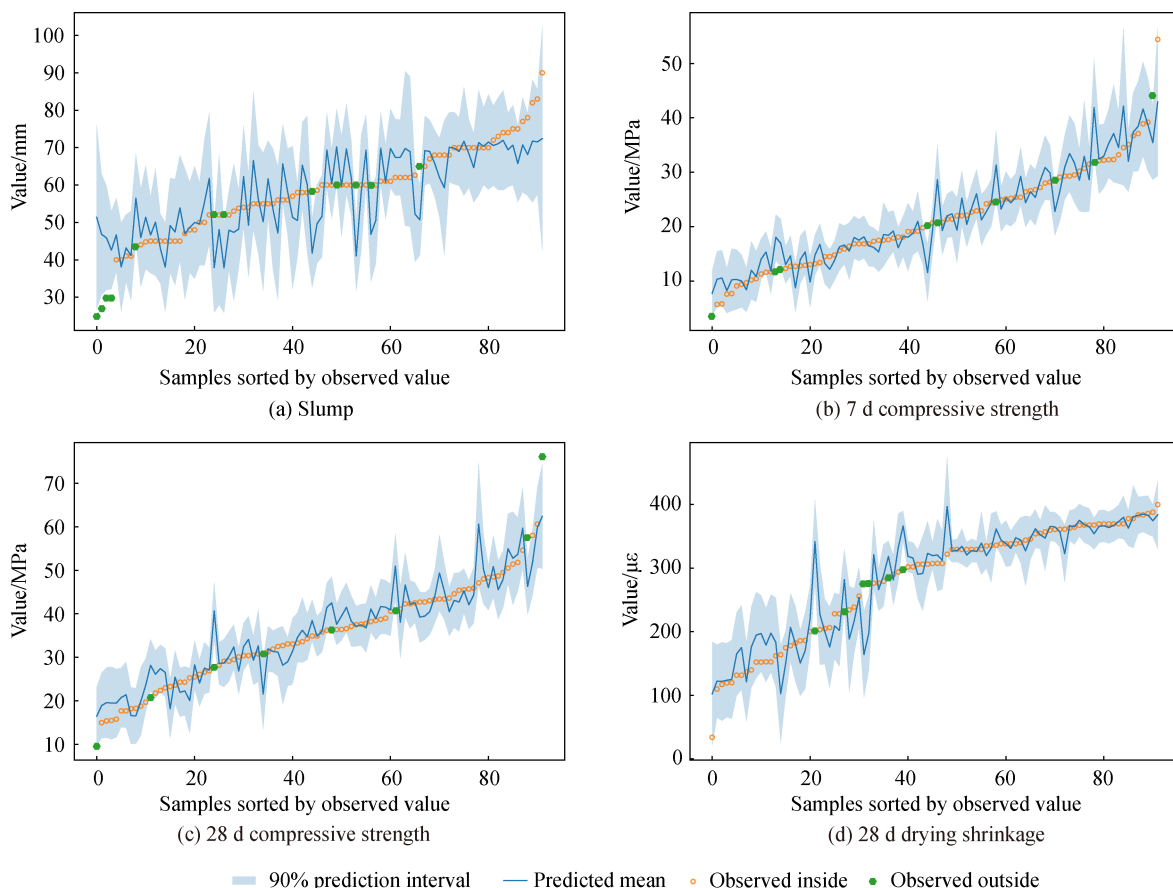


图 7 HMT-HPNN 模型对各目标性能的 90%预测区间及观测值覆盖情况

Fig. 7 90% prediction intervals and observation coverage of HMT-HPNN model for the target properties

表 2 HMT-HPNN 模型 90%预测区间对测试集上观察点的覆盖情况

Table 2 Coverage of the 90% prediction interval of the HMT-HPNN model for the observed points in the test set

Target	Inside the interval	Total number of observed points	Coverage rate
Slump	80	92	86.96%
7 d compressive strength	83	92	90.22%
28 d compressive strength	84	92	91.30%
28 d drying shrinkage	86	92	93.48%
Overall	333	368	90.49%

从预测区间形态看, HMT-HPNN 给出的区间宽度并非固定不变, 而是随样本特征和目标类型发生变化, 体现出样本相关的异方差特征。对于坍落度, 预测区间整体相对较宽, 说明模型能够识别工作性预测中较高的不确定性水平; 对于 7 d 和 28 d 抗压强度, 区间宽度整体较为稳定, 并在局部高响应或偏离样本处适度扩大; 对于 28 d 干燥收缩,

区间在低值段及局部波动样本附近有所加宽, 表明模型对体积变形指标的离散性具有一定表征能力。上述结果说明, HMT-HPNN 不仅能够输出预测均值, 还可为不同样本提供差异化的风险边界, 使模型结果更接近工程中预测值与允许波动范围共同表达的实际需求。

为进一步评价 HMT-HPNN 预测不确定性是否具有误差风险指示作用, 研究基于训练集 5 折内预测结果计算综合预测不确定性, 并以其 1/3 和 2/3 分位数作为低、中、高不确定性分组阈值。其不确定性分布确定的低-中不确定性阈值为 0.064 8、中-高不确定性阈值为 0.077 1; 按此固定阈值, 测试集 92 组样本中低、中、高不确定性组的样本数分别为 36、20、36 组。图 8 给出了不同不确定性分组下的综合预测误差, 图 9 展示了各目标标准化误差在不同分组中的分布。结果表明, 低、中、高不确定性组的平均综合误差(Composite error)呈单调递增趋势, 约由 0.05 增至 0.07 和 0.09, 说明模型给出的综合预测不确定性与实际综合误差之间具有较好的一致性。

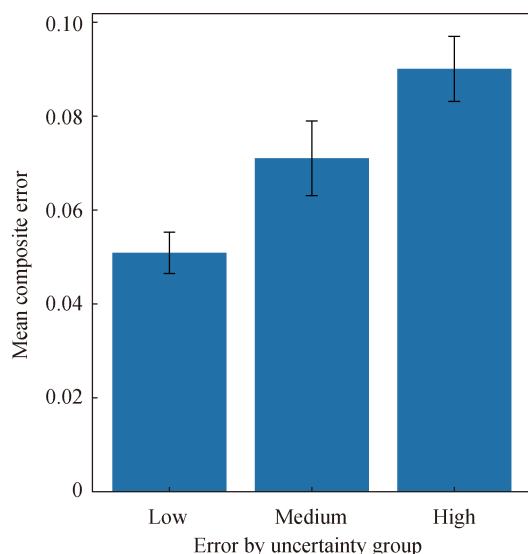


图8 不确定性分组下的综合预测误差比较

Fig. 8 Comparison of composite error across uncertainty groups

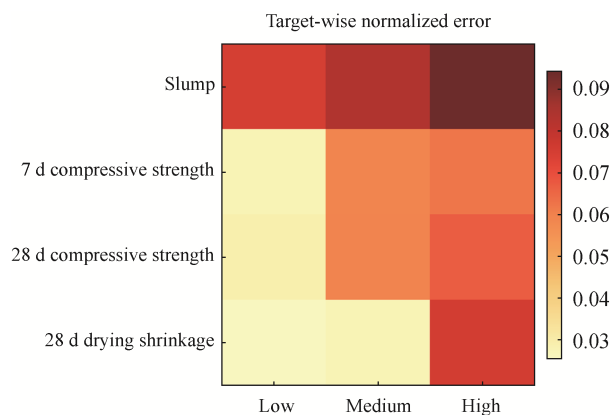


图9 不确定性分组下各目标的标准化误差分布

Fig. 9 Distribution of target-wise normalized errors across uncertainty groups

从目标层面看, 坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩在高不确定性组中的标准化误差整体高于低不确定性组, 这表明不确定性等级不仅同综合误差层面有效对应, 也能反映不同性能目标的误差风险差异。坍落度在各分组中的误差水平相对较高, 进一步说明工作性目标具有更强的随机波动特征; 强度类目标随不确定性等级升高表现出较为清晰的误差递增关系, 说明 HMT-HPNN 对强度预测风险的分层识别较为稳定; 干燥收缩在高不确定性组中误差增大, 也反映出模型能够对体积稳定性预测中的高风险样本给出相应的不确定性响应。

综合预测区间与可信度评价结果可见, HMT-HPNN 的概率输出具有较好的工程解释性。1) 90%

预测区间总体覆盖率与名义水平一致, 表明模型能够为 4 个目标提供可靠的区间预测; 2) 基于综合不确定性(Composite uncertainty)建立的低、中、高不确定性分组能够有效区分样本实际误差水平, 说明模型不确定性输出具有误差风险指示能力。

3.4 同源工程的泛化性验证

为检验 HMT-HPNN 在工程场景中的适用性, 本工作将超参数优化后的 HMT-HPNN 模型直接应用于 46 组水工常态混凝土同源工程验证集。该验证集为同源水工工程中试验得到的水工常态混凝土数据, 不参与模型训练、交叉验证和超参数搜索, 仅用于评价 HMT-HPNN 的工程泛化能力。按训练集确定的固定阈值, 工程验证样本中低、中、高不确定性组的样本数分别为 16、7、23 组。

图 10 比较了训练集与同源工程验证集在 4 个目标性能上的分布。总体来看, 验证集的坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩均位于训练集主体分布范围内, 中位数和四分位区间与训练集保持较好对应, 未出现明显目标分布偏移。其中, 验证集坍落度分布略低且更集中, 强度目标与训练集主体区间重叠程度较高, 28 d 干燥收缩也处于训练集覆盖范围内。上述结果说明, 工程验证集与模型构建所用训练集具有较好的数据同源性和可比性, 可用于评价模型在相近材料体系和工程条件下的泛化表现。

图 11 给出了 HMT-HPNN 在同源工程验证集上的预测结果。模型对坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩的 R^2 分别为 0.70、0.94、0.90 和 0.87, 对应 MAE 分别为 5.78 mm、1.88 MPa、2.62 MPa 和 16.61 $\mu\epsilon$ 。两类强度目标的预测点主要分布在 $y=x$ 参考线附近, 表明模型对胶凝材料组成、配合比参数及骨料性质与强度响应之间的关系具有较好的迁移能力; 28 d 干燥收缩同样保持较高的预测相关性, 说明模型对体积稳定性目标的学习结果能够在同源工程样本中延续; 坍落度的泛化精度相对较低, 这与测试集结果一致。总体而言, HMT-HPNN 在工程验证集上保持了较好的多目标均值预测能力。

表 3 比较了 HMT-HPNN 模型在实验数据集与同源工程验证集上的不确定性评价结果。在工程验证集上, 坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩的 PICP90 分别为 91.30%、97.83%、95.65%和 93.48%, 均达到且高于 90%名义置信水平, 说明模型在独立工程样本上的 90%预测区间具

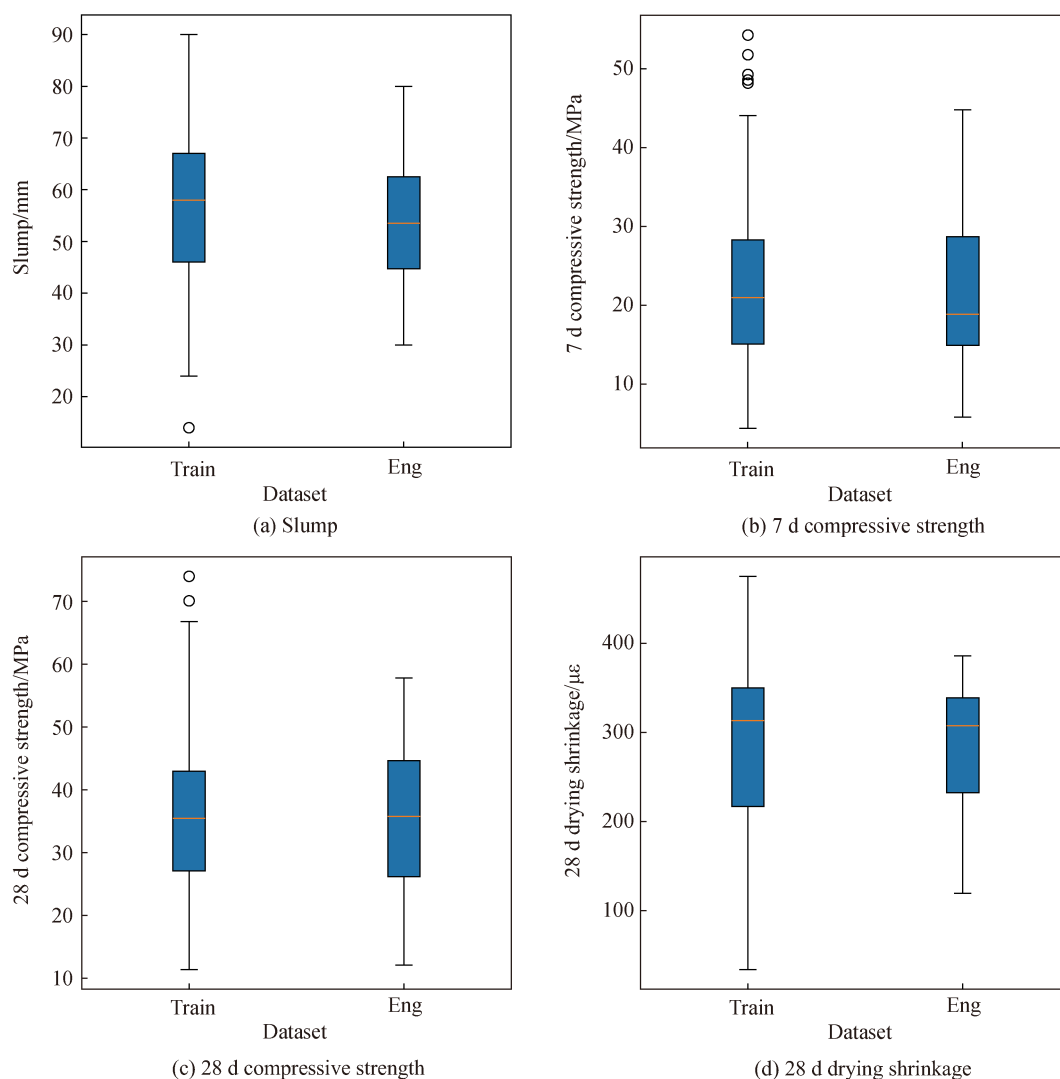


图 10 训练集与同源工程验证集各目标性能的分布对比

Fig. 10 Distribution comparison of the target properties between the training set and the homologous engineering validation set

有较稳定的覆盖能力。对应 NLL 分别为 0.800 8、0.052 6、0.168 7 和 0.028 4, 均低于实验测试集结果, 表明模型在工程验证集上的预测分布与实测值具有较好匹配性。各目标的 PICP90 虽高于 90%, 说明区间估计具有一定保守性, 但结合较低的 NLL 可知, 该结果并非单纯由过度放宽区间造成, 而是模型在工程样本上保持了较合理的均值与方差平衡。

图 12 展示了工程验证集中各样本在不同目标上的预测不确定性分布。不同样本和不同目标之间的不确定性水平存在明显差异, 说明 HMT-HPNN 在工程验证集上仍能输出样本相关的异方差信息, 坍落度的不确定性整体相对较高, 符合其受现场拌和和施工因素影响较强的特点; 7 d 和 28 d 抗压强度的不确定性相对较低, 反映出同源工程条件下强度响应规律较为稳定; 干燥收缩的不确定性在不同

样本间存在一定波动, 说明体积变形目标仍受材料组成、浆体体积分数和骨料约束等多种因素共同影响。

本工作中将基于训练集 5 折内结果所确定的低、中、高不确定性阈值应用于工程验证集, 以验证 HMT-HPNN 的可信预测能力。图 13 结果表明, 低、中、高不确定性组的平均 composite error 分别约为 0.040、0.057 和 0.075, 综合误差随不确定性等级升高而递增。该结果说明, HMT-HPNN 在训练集上学习得到的可信度评判能力能够在工程验证样本中保持有效, 模型输出的 Composite uncertainty 对实际预测误差仍具有风险指示作用。

综合来看, 工程验证结果表明, HMT-HPNN 不仅在实验数据集内部具有较好的准确性和不确定性表征能力, 也能够同源工程数据上保持较稳定的泛化表现。

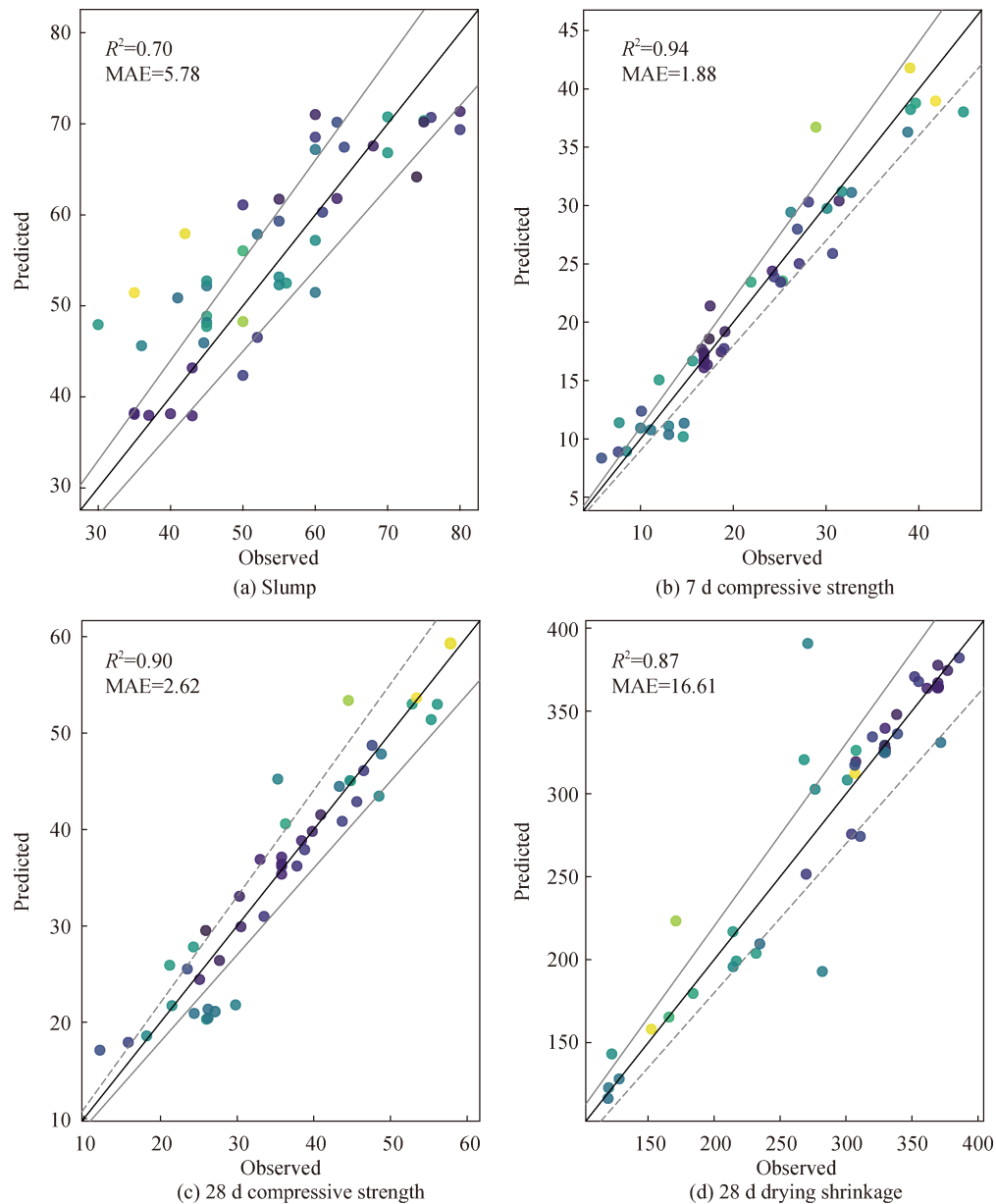


图 11 HMT-HPNN 模型在同源工程验证集上的预测精度泛化结果

Fig. 11 Generalization results of the predictive accuracy of HMT-HPNN model on the homologous engineering validation set

表 3 HMT-HPNN 模型在实验数据集与同源工程验证集上的不确定性指标对比

Table 3 Comparison of uncertainty metrics of HMT-HPNN model on the experimental dataset and the homologous engineering validation dataset

Data source	Probability prediction model	Target	PICP90	NLL
Experimental dataset	HMT-HPNN	Slump	86.96%	0.959 4
Experimental dataset	HMT-HPNN	7 d compressive strength	90.22%	0.198 3
Experimental dataset	HMT-HPNN	28 d compressive strength	91.30%	0.270 2
Experimental dataset	HMT-HPNN	28 d drying shrinkage	93.48%	0.142 5
Engineering validation dataset	HMT-HPNN	Slump	91.30%	0.800 8
Engineering validation dataset	HMT-HPNN	7 d compressive strength	97.83%	0.052 6
Engineering validation dataset	HMT-HPNN	28 d compressive strength	95.65%	0.168 7
Engineering validation dataset	HMT-HPNN	28 d drying shrinkage	93.48%	0.028 4

MLP + MC Dropout—multilayer perceptron with Monte Carlo Dropout; HMT-HPNN—hierarchical multi-task heteroscedastic probabilistic neural network; PICP90—coverage probability of the 90% prediction interval; NLL—negative log-likelihood.

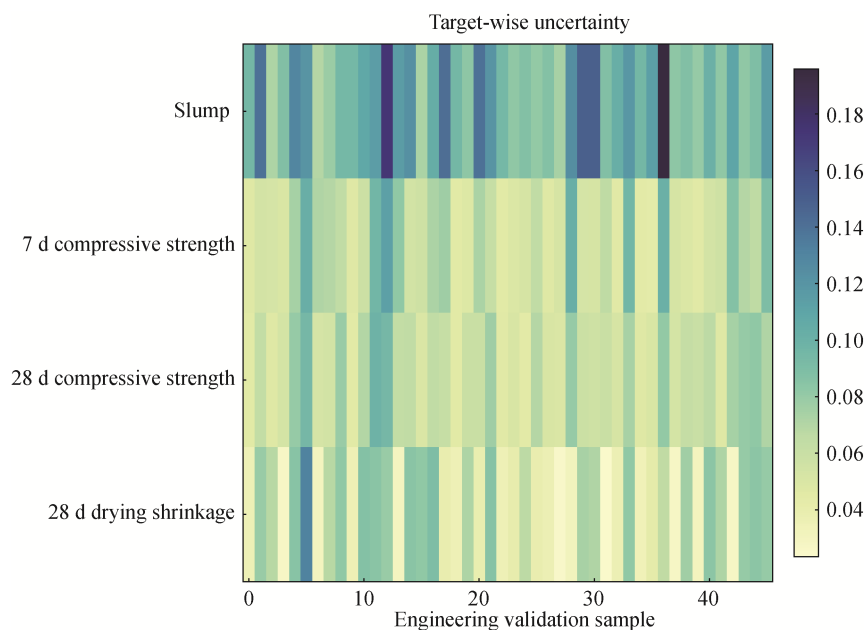


图 12 工程验证集中各目标综合不确定性分布

Fig. 12 Distribution of target-wise composite uncertainty in the engineering validation set

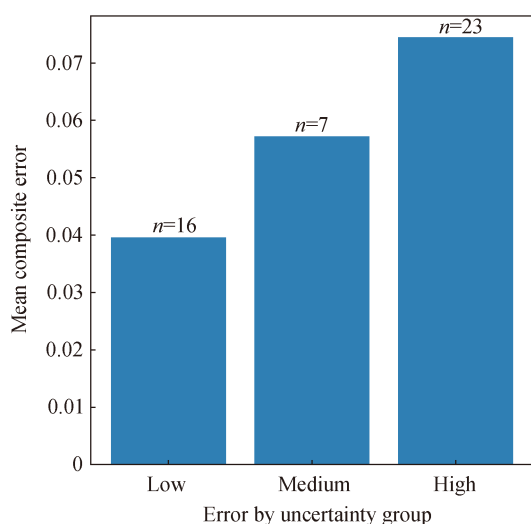


图 13 工程验证集中不确定性分组的综合误差比较

Fig. 13 Comparison of composite errors across uncertainty groups in the engineering validation set

4 结论

本工作面向水工常态混凝土, 构建了以 45 维原始特征为输入, 以坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩为输出的多目标性能数据库, 并提出 HMT-HPNN, 从准确性、不确定性、可信度及工程泛化性等方面进行了系统评价, 以应对水工常态混凝土多性能协同概率预测与结果可信度量化需求。主要结论如下:

1) HMT-HPNN 在水工常态混凝土多性能数据上表现出与经典神经网络相近的回归能力, 虽然其

点预测精度并非 4 类模型中最优, 但其测试集平均 NRMSE 与 Single-target MLP、Multi-output MLP 和 MLP+MC Dropout 处于同一较优水平。HMT-HPNN 对 7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩均取得较高预测精度, 说明该模型能够有效学习水工常态混凝土配合比、胶凝材料组成和骨料特性与多目标性能之间的非线性关系。

2) 与经典概率预测网络 MLP+MC Dropout 相比, HMT-HPNN 具有更优的不确定性表征能力。其在坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩上的 PICP90 分别为 86.96%、90.22%、91.30% 和 93.48%, 整体更接近 90% 名义置信水平; 同时, 4 个目标的 NLL 均低于 MLP+MC Dropout, 说明 HMT-HPNN 能够在预测均值和预测方差之间形成更合理的概率分布估计。

3) HMT-HPNN 的 90% 预测区间总体覆盖率为 90.49%, 与名义置信水平基本一致, 其异方差输出能够有效转化为可靠的区间预测结果。HMT-HPNN 在测试集中的 Composite error 随 Composite uncertainty 等级升高而递增, 说明模型输出的不确定性具有误差风险指示能力, 可用于样本级可信度评价。

4) 在 46 组同源工程验证数据上, HMT-HPNN 仍保持较好的泛化性能。模型对坍落度、7 d 抗压强度、28 d 抗压强度和 28 d 干燥收缩的 R^2 分别为 0.70、0.94、0.90 和 0.87, PICP90 均达到且高于 90%。同时, 综合预测误差在工程验证集上仍表现出随不确

定性等级递增的趋势, HMT-HPNN 在同源工程中具有较好的准确性、不确定性和可信度泛化能力。

(第二十八届中国科协年会学术论文)

参考文献:

- [1] 国家能源局. DL/T 5330—2015 水工混凝土配合比设计规程[S]. 北京: 中国电力出版社, 2015.
- [2] 中华人民共和国水利部. SL 677—2014 水工混凝土施工规范[S]. 北京: 中国水利水电出版社, 2015.
- [3] 中华人民共和国水利部. SL/T 352—2020 水工混凝土试验规程[S]. 北京: 中国水利水电出版社, 2021.
- [4] MEHTA P K, MONTEIRO P J M. Concrete: Microstructure, Properties, and Materials[M]. 4th ed. New York: McGraw-Hill, 2014.
- [5] KOSMATKA S H, WILSON M L. Design and Control of Concrete Mixtures[M]. 15th ed. Skokie: Portland Cement Association, 2011.
- [6] NAYAK D K, ABHILASH P P, SINGH R, et al. Fly ash for sustainable construction: A review of fly ash concrete and its beneficial use case studies[J]. Clean Mater, 2022, 6: 100143.
- [7] TERMKHAJORNKIT P, NAWA T, NAKAI M, et al. Effect of fly ash on autogenous shrinkage[J]. Cem Concr Res, 2005, 35(3): 473–482.
- [8] LI Z Z, YOON J, ZHANG R, et al. Machine learning in concrete science: Applications, challenges, and best practices[J]. npj Comput Mater, 2022, 8: 127.
- [9] 刘晓, 王思迈, 卢磊, 等. 机器学习预测混凝土材料耐久性的研究进展[J]. 硅酸盐学报, 2023, 51(8): 2062–2073.
- LIU Xiao, WANG Simai, LU Lei, et al. J Chin Ceram Soc, 2023, 51(8): 2062–2073.
- [10] 龙武剑, 罗盛禹, 程博远, 等. 机器学习算法用于自密实混凝土性能设计的研究进展[J]. 材料导报, 2024, 38(11): 104–113.
- LONG Wujian, LUO Shengyu, CHENG Boyuan, et al. Mater Rep, 2024, 38(11): 104–113.
- [11] YOUNG B A, HALL A, PILON L, et al. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions: New insights from statistical analysis and machine learning methods[J]. Cem Concr Res, 2019, 115: 379–388.
- [12] DONG Y H, TANG J H, XU X H, et al. A new method to evaluate features importance in machine-learning based prediction of concrete compressive strength[J]. J Build Eng, 2025, 102: 111874.
- [13] ZHANG M X, LI M C, SHEN Y, et al. Multiple mechanical properties prediction of hydraulic concrete in the form of combined damming by experimental data mining[J]. Constr Build Mater, 2019, 207: 661–671.
- [14] ÖZTAŞ A, PALA M, ÖZBAY E, et al. Predicting the compressive strength and slump of high strength concrete using neural network[J]. Constr Build Mater, 2006, 20(9): 769–775.
- [15] BAL L, BUYLE-BODIN F. Artificial neural network for predicting drying shrinkage of concrete[J]. Constr Build Mater, 2013, 38: 248–254.
- [16] HILLOULIN B, UMUNNAKWE R. Machine learning-aided prediction of shrinkage in modern concrete: Focus on mix proportions and SCMs[J]. J Build Eng, 2024, 98: 111410.
- [17] 徐潇航, 胡张莉, 刘加平, 等. 基于机器学习回归模型的三峡大坝混凝土强度预测[J]. 材料导报, 2023, 37(2): 41–49.
- XU Xiaohang, HU Zhangli, LIU Jiaping, et al. Mater Rep, 2023, 37(2): 41–49.
- [18] TAFSESE W Z, ESPINOSA-LEAL L. Multitarget regression models for predicting compressive strength and chloride resistance of concrete[J]. J Build Eng, 2023, 72: 106523.
- [19] 王琰帅, 万承鹏, 董必钦, 等. 基于混合反向传播神经网络的双输出预测模型构建[J]. 硅酸盐学报, 2026, 54(3): 894–908.
- WANG Yanshuai, WAN Chengpeng, DONG Biqin, et al. J Chin Ceram Soc, 2026, 54(3): 894–908.
- [20] BORCHANI H, VARANDO G, BIELZA C, et al. A survey on multi-output regression[J]. WIREs Data Min & Knowl, 2015, 5(5): 216–233.
- [21] CARUANA R. Multitask learning[J]. Mach Learn, 1997, 28(1): 41–75.
- [22] RUDER S. An overview of multi-task learning in deep neural networks[EB/OL]. (2017–06–15)[2026–07–29]. <https://arxiv.org/abs/1706.05098>.
- [23] ZHANG Y, YANG Q. A survey on multi-task learning[J]. IEEE Trans Knowl Data Eng, 2022, 34(12): 5586–5609.
- [24] GAL Y, GHAHRAMANI Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning[C]//Proceedings of the 33rd International Conference on Machine Learning. 2016: 1050–1059.
- [25] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. J Mach Learn Res, 2014, 15(1): 1929–1958.
- [26] KENDALL A, GAL Y. What uncertainties do we need in Bayesian deep learning for computer vision?[C]//ADV NEURAL INF PROCESS SYST. 2017, 30: 5574–5584.
- [27] LAKSHMINARAYANAN B, PRITZEL A, BLUNDELL C. Simple and scalable predictive uncertainty estimation using deep ensembles[C]//ADV NEURAL INF PROCESS SYST. 2017, 30: 6402–6413.
- [28] GUO C, PLEISS G, SUN Y, WEINBERGER K Q. On calibration of modern neural networks[C]//PROC 34TH INT CONF MACH LEARN. 2017, 70: 1321–1330.
- [29] TAMULY P, NAVA V. Machine learning based conformal predictors for uncertainty-aware compressive strength estimation of concrete[J]. Constr Build Mater, 2025, 487: 141844.
- [30] HAN B, WU Y Q, LIU L L. Prediction and uncertainty quantification of compressive strength of high-strength concrete using optimized machine learning algorithms[J]. Struct Concr, 2022, 23(6): 3772–3785.
- [31] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The elements of statistical learning: Data mining, inference, and prediction[M]. 2nd ed. New York: Springer, 2009.
- [32] BISHOP C M. Probability Distributions[M]//Pattern Recognition and Machine Learning. New York: Springer, 2006: 68–126.
- [33] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Cambridge, Mass.: MIT Press, 2016.
- [34] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//PROC 14TH INT JOINT CONF ARTIF INTELL. 1995: 1137–1145.
- [35] BERGSTRA J, BENGIO Y. Random search for hyper-parameter optimization[J]. J Mach Learn Res, 2012, 13: 281–305.
- [36] MURPHY K P. Probabilistic Machine Learning: An Introduction[M]. Cambridge: MIT Press, 2022.
- [37] GNEITING T, RAFTERY A E. Strictly proper scoring rules, prediction, and estimation[J]. J Am Stat Assoc, 2007, 102(477): 359–378.

作者贡献声明:

徐潇航: 论文构思、设计研究方法/实验方案、综合数据并推导结论、撰写初稿

储大勇: 调查研究、设计模型算法、分析数据并推导结论、绘制图表

田 丹：调查研究、数据管理、验证

胡张莉：监督指导、审阅与修订

杨华美：调查研究、数据管理、验证

刘加平：提供资源、获取资助、项目管理

Probabilistic Multi-target Performance Prediction and Credibility Evaluation of Hydraulic Normal Concrete

XU Xiaohang¹, CHU Dayong¹, TIAN Dan², YANG Huamei², HU Zhangli¹, LIU Jiaping¹

(1. School of Materials Science and Engineering, Southeast University, Nanjing 211189, China; 2. Hubei Key Laboratory of Water Engineering Materials and Application Technology China Three Gorges Corporation, Wuhan 430010, China)

Extended Abstract

Introduction Hydraulic normal concrete is required to satisfy multiple performance demands during mixture proportioning, construction, and long-term service. Its fresh-state workability affects transportation, placement, and vibration; early-age compressive strength governs construction sequence and early load-bearing capacity; standard-age compressive strength is directly related to strength grade assessment; and drying shrinkage reflects dimensional stability of hardened concrete. In engineering practice, these properties are not controlled only by deterministic target values, but are usually associated with allowable fluctuations, quality assurance requirements, and strength guarantee rates. Owing to the multiphase and heterogeneous nature of concrete, the measured properties may still fluctuate even under similar mixture proportions, raw material conditions, and curing conditions. Therefore, deterministic point prediction is insufficient for describing the reliability of performance predictions.

Machine learning has provided an effective data-driven approach for modeling the nonlinear relationship between mixture composition and concrete properties. Previous studies have demonstrated its applicability in predicting compressive strength, slump, shrinkage, durability-related indicators, and other engineering properties of cement-based materials. Multi-output and multitask learning methods further enable simultaneous prediction of multiple performance indicators by exploiting shared information among related objectives. However, most existing models for concrete performance prediction still focus mainly on deterministic outputs, while uncertainty quantification, prediction interval construction, and credibility evaluation remain insufficiently explored, especially for hydraulic normal concrete. To address this issue, this study developed a hierarchical multi-task heteroscedastic probabilistic neural network (HMT-HPNN) for probabilistic multitarget performance prediction and credibility evaluation of hydraulic normal concrete.

Methods A multitarget performance database of hydraulic normal concrete was established. The database contained 457 samples, with 45 raw input variables related to cement, fly ash, aggregates, and mixture proportions. Four output targets were selected, including slump, 7 d compressive strength, 28 d compressive strength, and 28 d drying shrinkage. The development database was divided into a training set and a test set at an 8:2 ratio, corresponding to 365 and 92 samples, respectively. In addition, 46 homologous engineering samples were used as an engineering validation set to evaluate the generalization ability of the proposed model in an engineering scenario.

Four neural network models were constructed and compared, including Single-target MLP, Multi-output MLP, MLP + MC Dropout, and HMT-HPNN. The Single-target MLP was used as an independent deterministic modeling baseline, while the Multi-output MLP was adopted as a conventional deterministic multi-output baseline. The MLP + MC Dropout was selected as a classical probabilistic neural-network baseline, in which multiple stochastic forward passes were used to estimate predictive uncertainty. The proposed HMT-HPNN consisted of a shared representation module, a hierarchical task representation module, task-specific branches, and heteroscedastic probabilistic subnetworks. For each target, the model simultaneously output the predictive mean and log-variance, and was trained using a heteroscedastic Gaussian negative log-likelihood (NLL) loss.

To ensure a fair comparison, all models were optimized using the same data partition and preprocessing strategy. The training set was further divided using fixed five-fold stratified cross-validation based on the quantiles of 28 d compressive strength. Randomized grid search was adopted for hyperparameter optimization. Deterministic models were mainly selected according to the five-fold average NRMSE, while probabilistic models were selected by jointly considering prediction accuracy and negative log-likelihood. The final models were retrained on the complete training set and evaluated on the test set. The evaluation metrics included R^2 , MAE, NRMSE, PICP90, NLL, calibration curves, 90% prediction intervals, composite uncertainty, composite error, and credibility classification.

Results and discussion The target distribution and correlation analysis showed that the four properties had different statistical characteristics and material-control mechanisms. The 7 d and 28 d compressive strengths exhibited a strong positive correlation, reflecting the continuity of strength development. In contrast, slump and 28 d drying shrinkage showed weak linear correlations with strength targets, indicating that workability and dimensional stability should be treated as independent prediction targets rather than being inferred from compressive strength alone. For prediction accuracy, the Single-target MLP achieved the lowest average test NRMSE among the four models. However, the overall accuracy differences among Single-target MLP, Multi-output MLP, MLP +

MC Dropout, and HMT-HPNN were relatively small. HMT-HPNN retained reliable regression capability after introducing probabilistic outputs. On the test set, its R^2 values for 7 d compressive strength, 28 d compressive strength, and 28 d drying shrinkage were approximately 0.875, 0.868, and 0.872, respectively. Slump showed relatively lower prediction accuracy, which was associated with its higher sensitivity to aggregate moisture content, water demand, admixture dispersion, and construction-related disturbances.

For uncertainty quantification, HMT-HPNN outperformed MLP+MC Dropout. The PICP90 values of HMT-HPNN for slump, 7 d compressive strength, 28 d compressive strength, and 28 d drying shrinkage were 86.96%, 90.22%, 91.30%, and 93.48%, respectively, which were closer to the nominal confidence level of 90%. In addition, HMT-HPNN achieved lower NLL values for all four targets, indicating a more appropriate balance between prediction error and predicted variance. The calibration curves also showed that HMT-HPNN generally provided better agreement between nominal confidence levels and empirical coverage rates. The 90% prediction intervals generated by HMT-HPNN achieved an overall coverage rate of 90.49% on the test set, demonstrating that the heteroscedastic uncertainty output could be effectively transformed into reliable prediction intervals. To evaluate the risk-indication ability of model uncertainty, composite uncertainty was calculated from the normalized predictive standard deviations of the four targets, while composite error was calculated from the normalized absolute prediction errors. The testing results showed that composite error increased with the composite uncertainty level, confirming that the uncertainty output of HMT-HPNN was informative in identifying sample-level prediction risk.

For the homologous engineering validation set, HMT-HPNN maintained satisfactory generalization performance. The R^2 values for slump, 7 d compressive strength, 28 d compressive strength, and 28 d drying shrinkage were 0.70, 0.94, 0.90, and 0.87, respectively. The PICP90 values for all four targets reached or exceeded 90%, and an effective correspondence was observed between the level of uncertainty and the prediction error. These results indicate that HMT-HPNN maintained its accuracy, uncertainty quantification capability, and credibility evaluation ability under homologous engineering conditions.

Conclusions A probabilistic multitarget prediction framework was developed for hydraulic normal concrete. The proposed HMT-HPNN simultaneously predicted slump, 7 d compressive strength, 28 d compressive strength, and 28 d drying shrinkage by integrating shared representation, hierarchical task representation, task-specific branches, and heteroscedastic probabilistic subnetworks. Although the point prediction accuracy of HMT-HPNN was not the highest among all compared models, its testing accuracy was comparable to that of conventional neural network models, confirming its reliable regression capability. Compared with MLP + MC Dropout, HMT-HPNN provided superior uncertainty quantification, with PICP90 values closer to the nominal 90% confidence level and lower NLL values for all four targets. The overall coverage rate of the 90% prediction intervals reached 90.49%, indicating that the model uncertainty could be effectively converted into reliable prediction intervals. The composite uncertainty was consistent with the composite prediction error, which confirmed the ability of HMT-HPNN to identify sample-level prediction risks. Independent homologous engineering validation further demonstrated the generalization capability of HMT-HPNN. The model maintained satisfactory accuracy, reliable prediction interval coverage, and effective credibility classification on 46 engineering validation samples. Therefore, the proposed framework provides a data-driven method for multitarget performance prediction, prediction interval construction, and credibility evaluation of hydraulic normal concrete, and can support mixture proportion design and performance risk assessment.

Keywords probabilistic neural network; multitarget prediction; hydraulic normal concrete; uncertainty; credibility