以数据为中心的工程(2023), 4:
e9doi:10.1017/dce.2023.5

研究论文

利用机器学习进行混凝土的概率选择与设计

Jessica C. Forsdyke[1],Bahdan Zviazhynski[2],Janet M. Lees[1] 和Gareth J. Conduit[2]

[1]剑桥大学工程系，英国剑桥[2]剑桥大学卡文迪许实验室，英国剑桥通讯作者：
Jessica C. Forsdyke；电子邮箱：jf580@cam.ac.uk

收稿日期：2022年8月15日；修订日期：2023年3月1日；接受日期：2023年3月24日

关键词：碳化；混凝土；机器学习；基于性能的规范 n

摘要

由于组成材料存在自然变异性，且配合比的可能组合众多，开发环境影响较低且性能稳健的混凝土配合比颇具挑战。利用机器学习进行可靠的性能预测，有助于实现混凝土的基于性能的规范，减少材料低效，并提高混凝土施工的可持续性。在本研究中，我们开发了一种机器学习算法，可利用中间目标变量及其相关噪声来预测最终目标变量。我们将该方法应用于指定一种抗碳化性能高的混凝土配合比，以及另一种环境影响低的混凝土配合比。两种混合料还满足强度、密度和成本方面的目标。我们通过实验将所指定的混合料与其预测结果进行了验证。我们提出的通用方法能够利用机器学习中的噪声，在结构工程及其他领域具有广泛的应用前景。

影响声明

本文证明，即使使用稀疏且含噪数据集，也可以利用机器学习预测混凝土的性能。这对混凝土配合比的基于性能的规范具有重要应用价值——既能指定具有适当耐久性和高强度的混凝土，又能最大限度地减少隐含碳或成本。在需要耗时且成本高昂的试验的情况下，这一点尤其有益。本文开发并展示的机器学习方法可应用于更广泛的加速材料设计领域，从而能够针对每种特定应用快速设计定制材料。此外，许多其他垂直领域也存在信息嵌入噪声中的实例，包括自动驾驶车辆、增材制造和信息工程；在这些领域，机器学习为加速开发、理解和影响拓展提供了机会。

1. 引言

混凝土是世界上使用最广泛的建筑材料。只有水的消耗量超过了混凝土（Sedgwick，1991年）。混凝土非常适合用于建筑，因为它易于

J.C.F. 和 B.Z. 对本研究的贡献相同。本文获颁 OpenData 徽章，以表彰其透明实践。详情请参阅数据可用性声明。

© 作者，2023。由剑桥大学出版社出版。本文为开放获取文章，依据知识共享署名许可协议（http://creativecommons.org/licenses/by/4.0）发布，允许不受限制地重复使用、分发和复制，但须正确引用原文。

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

可调整——通过调整配合比设计中的比例和材料，可以改变包括密度、强度、耐久性和外观在内的材料性能。

传统混凝土配合比由三种主要成分组成：水泥、水和骨料。水泥在混凝土中充当胶结料，在经济和环境方面的成本都很高（Wassermann 等人，2009年）。水泥生产约占全球 CO_2 排放量的 52018年%，这既源于生产过程排放，也源于能源需求（Allwood 和 Cullen，2012年）。因此，混凝土的大部分由骨料组成，骨料可分为“细”（砂）骨料或“粗”（砾石和岩石）骨料。混凝土拌合并置于模具中后，水泥中的化合物与水发生水化反应（Neville，2011年），形成由孔隙、水化水泥浆体和骨料构成的复杂微观结构。这赋予混凝土“硬化状态”下的性能，包括抵抗所施加荷载（强度、刚度）和侵蚀性物质（耐久性）的能力。

在二氧化碳$_2$ 浓度不断上升的时代，二氧化碳$_2$ 与建成环境之间的相互作用日益受到关注（Talukdar 和 Banthia，2016年；Jiang 等人，2018年）。具体而言，二氧化碳（CO_2）会与混凝土中的水化产物发生化学反应（Papadakis 等人，1989年），这一过程称为混凝土碳化。CO_2 会从周围环境扩散进入混凝土。由此形成的碳化材料与未碳化混凝土相比，其微观结构发生了变化（Groves 等人，1991年；Greve-dierfeld 等人，2020年），且孔隙水的碱度较低（Papadakis 等人，1989年）。因此，这种碳化可能导致结构损伤，因为碱度的降低会加剧内部钢筋的腐蚀，而这些钢筋对于承受混凝土结构中的拉力荷载至关重要（Page，2007年）。最终，碳化是钢筋混凝土结构的主要劣化机制之一，因此成为一个值得重点关注的问题。

可以采用基于性能的方法或规定性方法，为特定应用选择最合适的混凝土。实施基于性能的混凝土规范流程时，会对混凝土的输出性能（例如强度或碳化系数）设定目标。这与规定性混凝土规范方法不同，后者会限制水泥含量或水灰比等输入参数（欧洲规范 BS EN 206 所采用的原则：2013+A1（英国标准协会，2016年）。目前，混凝土配合比设计方法（Abrams，1918年；ACI 第211委员会，1991年；Teychenné 等人，1997年；Wilson 和 Kosmatka，2011年）利用配合比与混凝土性能之间的经验关系，尤其是混凝土的强度，来确定混合料的配比。在采用基于性能的方法进行任何大规模混凝土应用时，都会使用上述方法之一设计多种混合料，然后浇筑试配混凝土，以验证特定混合料的最终性能是否符合要求。因此，基于性能的规范有助于选出效率更高、更优的混合料（Wally 等，2022年）。不过，试错过程既耗时又成本高昂。

机器学习有望捕捉输入（配合比）与输出（材料性能）之间复杂的多维关系，并减少混凝土设计中对试错方法的依赖。机器学习利用历史数据训练模型，之后可用于预测感兴趣量。这种方法已用于预测多种不同材料的性能（Bhadeshia et al., 1995; Sourmail et al., 2002; Agrawal et al., 2014; Wardet al., 2016年; Wardet al., 2017; Kim et al., 2018年），包括混凝土（Taffeseet al., 2015; Ben Chaabene et al., 2020年; Prayogoet al.,2020年; Liu et al., 2021; Tran et al., 2022年），证明了其能力与适用性。本文将运用机器学习进行混凝土设计。

机器学习方法还可以利用目标变量预测结果中的不确定性，着重关注最稳健的预测结果。例如，利用不确定性来设计最有可能满足目标标准的材料，已得到广泛展示和实验验证（Conduit 等，2017，2018、2019）。此外，不确定性数值本身也有助于预测关注的量（Goujon，2009；Zerva 等，2017；Zhang，2020；Zviazhynski 和 Conduit，2022）。在混凝土中，不同形状和尺寸的骨料随机分布，其外观可视为类似白噪声，从而导致碳化深度和抗压强度等性能出现变异性和不确定性。可以从某些测量性能的不确定性中提取信息，并用于

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

预测其他性能，从而获得更深入的洞见，并最终改进配合比设计。在本研究中，我们开发了一种能够从不确定性中提取信息的方法。我们利用该方法筛选出具有特定预期性能、满足可持续性和耐久性目标的混凝土配合比，为今后将机器学习应用于噪声中隐藏着信息的系统提供范例。

本文首先探讨实验室研究中制备的混凝土配合比的测量性能之间的关系，详见第 2 节。随后，我们介绍一种机器学习方法，该方法能够从噪声中提取信息，以帮助预测混凝土性能，详见第 3 节。接着，我们介绍该方法在混凝土规范中的应用，详见第 4 节，并在第 5 节中根据实验结果验证我们的方案。最后，我们在第 6 节中讨论该方法未来在混凝土相关领域及其他领域的应用。

2. 混凝土性能之间的关系

为建立稳固的机器学习模型基础，我们首先利用从 21 种实验室混凝土中实测收集的数据，探讨各种混凝土性能之间的关系，其中包括 Forsdyke 和 Lees（2021b）报告的两种混合料。该数据集（见补充材料）包含每批次组成材料（水泥、砾石、砂和水）的配比信息，以及所得混凝土的相关性能（碳化系数、环境影响、强度、密度和成本）。本文所述的所有混凝土均采用最大粒径为 10 mm 的碎石灰岩作为砾石，所用棱角砂的最大粒径为 4 mm（其中 80% 的砂颗粒小于 1 mm）。

这些混合料最初采用 BRE 方法进行配比设计（Teychenné 等人，1997年），该方法使用估算的每立方米质量，此处也提供了该数值。作为衡量混凝土耐久性的指标，21 种混凝土中有 16 种暴露于 4% CO_2 环境中。对每种混合料至少三个样本的碳化深度进行测量，并随时间记录，从而绘制碳化曲线并计算碳化系数，K。碳化系数越低，抗碳化性能越高。另一个变量“预处理时间”指的是，对于测量了碳化性能的试件，从水中养护结束到暴露于高浓度 CO_2之间的时间。还提供了以每千克混凝土所含 CO_2 千克数表示的环境影响估算值（kgeCO_2/kg），以及混凝土的立方体抗压强度、在空气中测得的饱水密度和每千克估算成本（£/kg）。

通过皮尔逊相关系数测得的混凝土数据集各项性能之间的相关性如图1所示。前九行表示输入性能，例如混凝土浇筑前假定的配合比和每立方米估算质量。最后五行表示最终混凝土的测量性能。特别值得关注的相关性已通过–标记框（a图1中突出显示。总体而言，图1 表明，混凝土数据集中许多变量之间存在相关性，这些相关性可用于拟合数据的机器学习模型并生成预测结果。接下来，我们将讨论各项性能之间值得关注的相关性。

2.1. 碳化系数

在 (a) 处可以观察到碳化系数与水泥含量之间存在很强的负相关关系，见图1。这是因为水泥比例越高，水化后混凝土基体中可碳化材料的体积就越大（Papadakis 等人，1991年），从而延长碳化达到特定深度所需的时间。碳化系数与碎石含量 (b) 之间，以及碳化系数与总骨料/水泥比 (c) 之间，均呈现很强的正相关关系。考虑到在固定体积的水化水泥浆体中，悬浮的骨料越多，浆体就会渗入混凝土越深，这种关系也合乎逻辑。

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

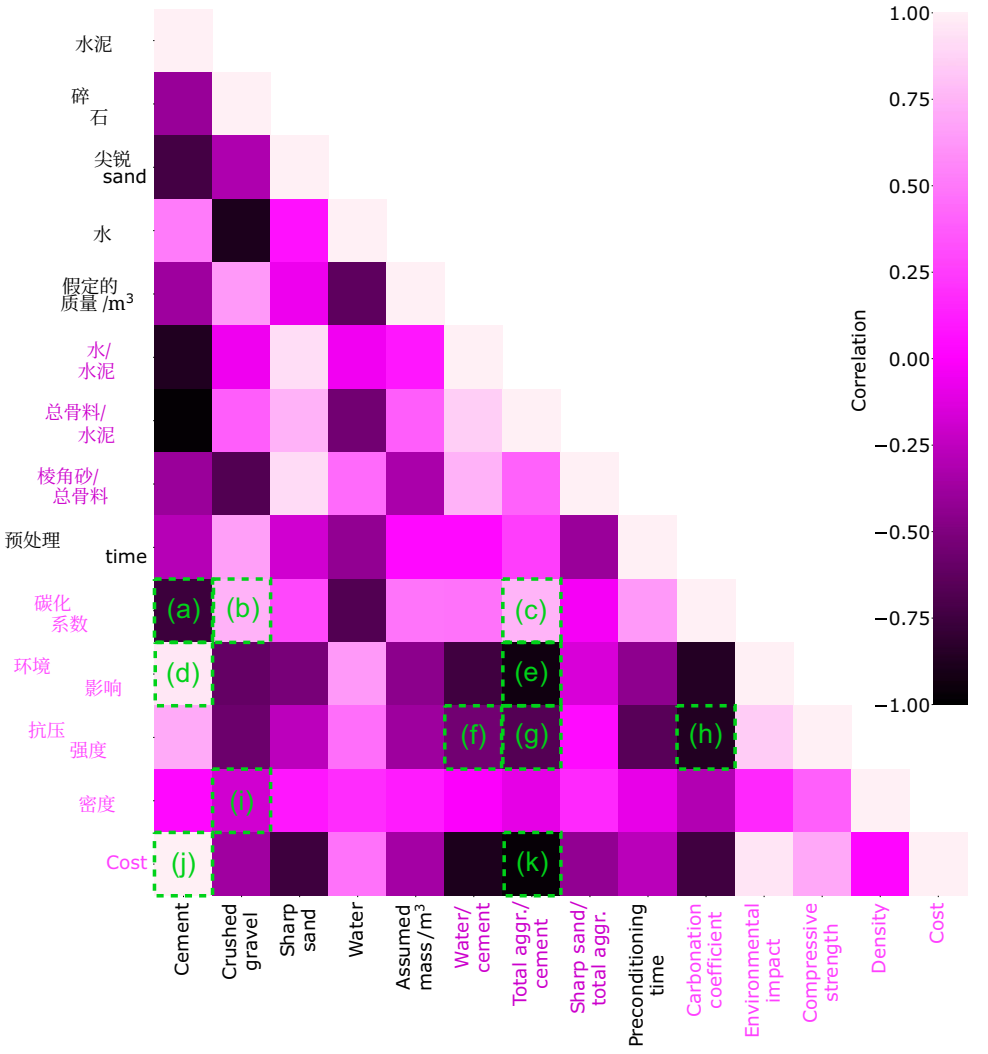


图1. 训练数据集中混凝土性能的相关性图。浅色表示强正相关，深色表示强负相关，中间色表示弱相关，具体如所示的色标。绿色方框标出了显著相关性。深粉色文字表示由混凝土配合比推导出的中间量，浅粉色文字表示目标变量。

2.2. 环境影响

众所周知，水泥是混凝土中造成大部分环境影响的组分。相比之下，作为惰性填充材料的骨料，其相对隐含碳较低（Neville，2011年）。图1 中也证实了这一点：环境影响与水泥含量 (d) 之间呈现很强的正相关关系，而环境影响与总骨料/水泥比 (e) 之间则呈现很强的负相关关系。鉴于这些相关性很强，我们预计机器学习对环境影响的预测会比对其他目标变量更准确，因为其他目标变量与组成变量的相关性较弱。

本研究假设每种组成成分的隐含排放量恒定，并据此计算环境影响。因此，环境影响是混凝土配合比的线性函数。然而，通常每种成分的隐含排放量取决于制造过程，

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

运输至施工现场以及固体废物产生量（Babor 等，2009；Sousa 和 Bogas，2021）。制造过程、运输和固体废物产生对环境影响的作用无法通过解析方法进行评估，因此，本研究展示了如何利用机器学习预测环境影响。

2.3. 抗压强度

图1 还显示，抗压强度与水灰比（f）之间，以及抗压强度与总骨料/水泥比（g）之间均呈负相关。普遍认为，混凝土强度取决于孔隙结构（Powers，1958年；Pantazopoulou 和 Mills，1995年；Chen 等，2013年）。当水灰比超过水泥完全水化所需的最低值时（Powers，1958年；Aïtcin，2016年），水灰比增大会降低混凝土强度，这是因为未与水化产物发生化学结合的水分子会形成微观孔隙（Pann 等，2004年）。总骨料/水泥比增大也会降低强度，这是由于水泥基体与骨料之间的黏结减弱（Poon 和 Lam，2008年），并且可能改变压实度（Marar 和 Eren，2011年），从而导致强度降低（Mindess 等，2003年）。此外，强度与碳化系数之间也呈现显著负相关（h），这支持了使用强度预测碳化行为的模型的有效性（Silva 等，2014年；Forsdyke 和 Lees，2021a）。

2.4. 密度

密度与输入参数之间似乎不存在明显的强相关性，因此我们预计，使用机器学习预测密度会比预测其他目标变量更困难。观察到的最强关系是密度与粗骨料（碎石）的绝对体积呈负相关（i）。混凝土在硬化状态下的密度受粒径分布影响，因为粒径分布决定了湿态下能否实现有效压实（Sims et al., 2019年）。本研究中各混合料的总骨料含量差异很小。因此，在该数据集中，粗骨料含量较高对应着细骨料含量较低（如棱角砂与碎石之间显著的负相关性所示）。这可能导致颗粒堆积不够理想，进而使粗骨料体积较高的混合料呈现出此处观察到的较低密度。其他未纳入此处考量的信息，例如骨料的粒径分布和光滑度，也可能影响密度，从而导致这些数据存在噪声。

2.5. 成本

观察到成本与水泥含量 (j) 之间存在很强的正相关，而成本与总骨料/水泥比 (k) 之间存在很强的负相关。这是意料之中的，因为水泥是混凝土中最昂贵的组成成分，而骨料作为填充材料，成本要低得多。与环境影响类似，我们预计机器学习对成本的预测将比对其他目标变量的预测更准确。

本研究假设每种组成成分的价格固定，并据此计算成本。因此，成本与环境影响类似，是混凝土配合比的线性函数。然而，通常每种组成成分的价格会受到法规、通货膨胀和供应情况的影响（Velumani and Nampoothiri，2018年；Ma et al.，2022年）。这些因素对成本的影响无法通过解析方法评估，因此本研究可作为基础，展示如何利用机器学习预测成本。

3. 机器学习方法论

机器学习算法利用已有数据进行训练，以做出预测。在本例中，预测的对象是混凝土配合比的性能。常用的机器学习算法包括*k*-均值聚类（Hastie 等人，2001），神经网络（Heskes，1997年）和高斯过程（Tancret，2013）。本文采用随机森林算法（Pedregosa 等人，2011年），因为它

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

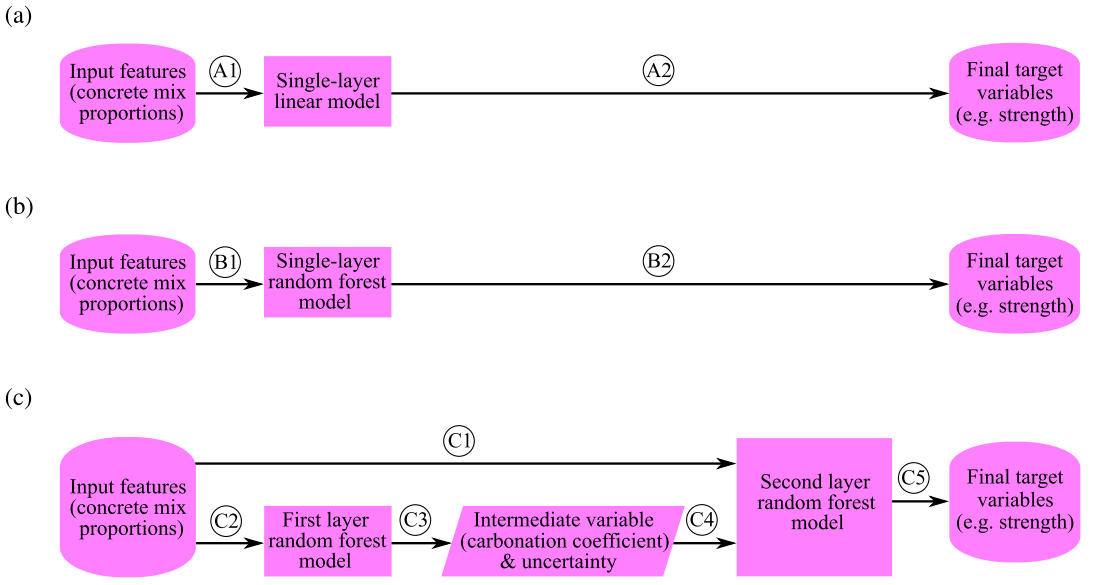


图2。（a）单层线性模型流程图。（b）单层随机森林模型流程图。（c）双层随机森林模型流程图。

在计算上成本低廉，并且能够稳健地发现非线性关系。我们比较了单层线性模型、单层随机森林模型，以及能够从混凝土数据中的噪声提取信息的双层随机森林模型。图 2 展示了这三种模型类型的信息流。

3.1. 单层线性模型

预测性能最基本的方法是对训练数据进行超平面拟合。超平面利用输入值（例如混凝土配合比）直接预测最终目标变量（例如强度）。图1展示了输入与输出性能之间线性相关的强度，为采用这种方法提供了依据。这种方法称为单层线性模型，由图2a中的流程 A1 和 A2 示意。

这种方法简单且稳健，但无法捕捉混凝土性能之间的非线性关系。如果最终目标变量与输入之间呈非线性依赖关系，线性模型的预测结果就不会准确。因此，我们还考虑一种能够捕捉数据中非线性的模型，即单层随机森林模型。

3.2. 单层随机森林模型

随机森林模型由一组相互独立且同分布的回归树组成（Loh，2011年）。在训练阶段，每棵回归树都会学习如何将输入变量映射到目标变量。随机森林算法通过自助法计算预测结果及其不确定性（Hastie 等人，2001）。对包含N 条记录的数据集进行自助抽样时，会通过有放回地随机抽取N 条记录生成新的子集。随后，每个子集都会用于训练随机森林中的一棵回归树。将各棵树的预测结果取平均，得到随机森林的预测结果；这些预测结果的标准差即为预测的不确定性。

随机森林中回归树的深度，以及因此而来的预测结果准确性，由随机森林的超参数决定。为了对盲数据实现最佳预测结果，我们对这些超参数进行调优，并使用留一法评估预测结果的准确性

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线发布

交叉验证（Hastie 等人，2001）在训练数据上进行。对于每个属性的预测结果，我们使用相同的超参数，以减轻过拟合。随机森林算法能够学习并利用各种目标变量及其不确定性之间的相关性，从而更准确地预测最终目标变量。

单层随机森林方法遵循图 2b中的流程 B1 和 B2，仅使用输入特征（混凝土配合比）来预测最终目标变量（例如强度）。

3.3. 双层随机森林模型

与传统的单层方法不同，双层随机森林模型中的信息流如图 2c所示。在双层方法中，我们首先以混凝土配合比为输入（流程 C1），训练第一个随机森林模型，以预测所有中间变量和输出变量（例如碳化系数）及其不确定性（流程 C2），采用 Zviazhynski 和 Conduit（2022年）所述的方法。随后，我们训练第二层随机森林模型，以混凝土配合比（流程 C3）以及其他近期预测的变量（包括不确定性值）（流程 C4）为输入来预测输出，例如强度（流程 C5）。

如果中间变量比目标变量更容易测量，那么将中间变量及其不确定性作为第二层随机森林模型的输入尤其有用。例如，与碳化系数相比，强度的测量成本更低，而且在混凝土中测量得更为普遍。因此，可以将强度作为双层随机森林模型中的中间变量来预测碳化系数，而无需进行耗时的碳化实验。另一方面，为了评估既有结构的状况，从结构中取出较大的芯样来测量抗压强度可能会造成严重破坏。作为替代，可以从较小、破坏性较低的样本中测量碳化深度等中间量，从而在减少结构损伤的同时估算强度。利用中间变量的期望值帮助预测最终变量的策略，先前已成功应用于材料与药物设计，以利用各种属性之间的关系（Verpoort 等人，2019年；Irwin 等人，2020a；Mahmoud 等人，2021），而我们现在则利用中间变量的不确定性。

机器学习需要捕捉的中间变量不确定性，源于实验不确定度导致训练数据不可避免的离散性。量化中间变量不确定性有两种方法。第一种方法是计算随机森林中回归树预测结果的标准差，该数值用于双层随机森林模型。第二种方法是计算实验测量值的标准差。在这两种情况下，给定变量的不确定性与该变量的单位相同。

为展示双层随机森林模型中不确定性的利用方式，我们考虑 图3 所示的示例。在这里，对于给定的水泥含量值 X，第一层随机森林模型预测相应的碳化系数值 Y 及其不确定性 σ_Y，如 图3 最左侧的图所示。然后，第二层随机森林模型利用碳化中的不确定性

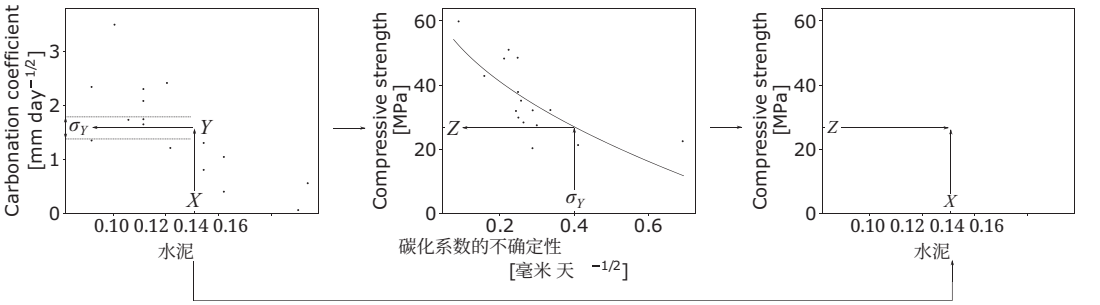


图3. 双层随机森林模型中不确定性的利用。从水泥含量 X 出发，模型利用碳化系数的不确定性来预测抗压强度 Z。

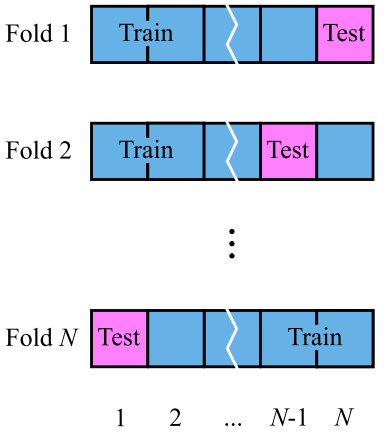
https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

系数，以预测相应的抗压强度值，*Z*，如图3中间的图所示。抗压强度是碳化系数不确定性的递减函数（斯皮尔曼相关系数为 −0：79），这使得两个模型能够协同工作，预测给定水泥含量下的抗压强度，如图3最右侧的图所示。

3.4. 模型的训练与测试

首先，使用留一交叉验证对三个备选模型的超参数进行调优。在该方法中，对于现有混凝土配合比数据中的每条记录，均使用在其余现有数据上训练的模型来预测混凝土性能（例如强度）（图4a）。然后，将这些预测结果与真实值进行比较，以获得留一交叉验证R^2, 决定系数（完美预测时为1，任意不准确的预测时为 $-\infty$），并以最大化该系数为目标。每个模型完成超参数调优后，各项性能对应的留一交叉验证 R^2 值见图4b中的表格。

单层线性模型对环境影响、抗压强度和成本的预测效果良好，因为这些指标近似为线性函数。然而，线性模型对碳化系数和密度的预测并不准确。单层随机森林模型改善了这些指标的预测，但代价是环境影响、强度和成本的预测准确性有所下降。正如第2节的数据分析所示，三个模型对环境影响和成本的交叉验证 R^2 值均最高，而密度的值通常最低。因此，能够利用目标变量之间相关性及其不确定性的双层模型，优于单层线性模型和单层随机森林模型。这表明，数据驱动机器学习对于性能预测是必要的，也证明了即使数据集仅包含21个样本，该方法仍然可行。单层线性模型、单层随机森林模型和双层随机森林模型的超参数见补充材料。



(a)

R^2	Single-layer linear model	Single-layer random forest model	Two-layer random forest model	Imputation model*
Carbonation coefficient	-3.00	0.44	0.71	**0.75**
Environmental impact	**> 0.99**	0.90	**> 0.99**	**> 0.99**
Compressive strength	0.80	0.67	0.81	**0.87**
Density	-1.45	0.27	0.50	**0.59**
Cost	**> 0.99**	0.95	**> 0.99**	**> 0.99**

*uses intermediate variables as inputs

(b)

图4。（a）留一交叉验证示意图。蓝色方块表示现有数据中的记录，洋红色方块表示每一折中的测试记录。（b）用于性能预测的留一交叉验证 R^2 值表。粗体数字表示给定性能对应的最佳 R^2 值。

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

为评估中间变量的影响，我们还测试了一种单层随机森林模型。该模型使用实验数据中的中间变量取值来预测目标变量。这种模型称为插补模型，已成功应用于材料和药物发现领域（Verpoort 等人，2019年；Whitehead 等人，2019年；Irwin 等人，2020a、2020b；Mahmoud 等人，2021）。插补模型利用中间变量与目标变量之间的相关性，因此其交叉验证 R^2 值高于单层线性模型和单层随机森林模型。插补模型无需预测中间变量，因此不会传播这些变量中的误差，从而使其交叉验证 R^2 值高于双层随机森林模型。然而，插补模型需要将实测混凝土性能作为输入，因此不适用于设计新型混凝土配合比并预测其性能。因此，在本研究后续部分中，我们采用双层随机森林模型。

双层随机森林模型可获取水泥类型、混凝土配合比、混凝土配合比的比值、假定质量/m^3, 和预处理时间。各特征对相应目标变量的预测相对重要性见补充材料。使用经过调优的超参数，在所有现有混合料数据上训练双层随机森林模型后，即可用该模型预测未见配合比的目标变量及其不确定性。随后将计算未见配合比满足既定目标标准的概率，并对最有可能满足给定目标的混合料进行实验验证。

4. 混凝土规范

随着机器学习模型就绪，我们已具备充分条件来探索完全未见过的混凝土配合比。下面，我们为混凝土规定两组具有挑战性的目标性质。为模拟施工现场的实际情况（当地材料的供应情况将决定可能的配合比设计组成），我们使用机器学习，在一系列假设配合比中，针对一组水灰比值网格寻找最佳配比。我们首先规定总含水量为 205 kg/m^3, ，并假设这些混合料的工作性要求保持不变。随后按照 BRE 配合比设计方法（Teychenné 等人，1997年）对这些假设配合比进行配料；该方法使用骨料的相对密度（本例中为 2.7）以及通过 600 微米筛的细骨料百分比（本例中为 70%），计算每种水灰比下其余组分的配比以及混凝土的总密度。这确保所有潜在配合比的比例都符合实际，且这些混合料切实可行。随后利用机器学习探索这一混合料空间，并预测这些混合料的广延性质；若不使用机器学习，这些性质原本需要通过实验获得。

对于第一个目标配合比 Low-K，我们重点降低碳化系数，从而提供高抗碳化性能。该配合比非常适用于暴露条件严酷的结构应用场景，此时需要较高的抗碳化性能，以保护钢筋免受碳化诱导腐蚀。

对于第二个目标配合比 Low-E，我们重点降低环境影响，同时保持相对较低的碳化系数；由于这两种性质呈负相关，这一目标颇具挑战性。该配合比非常适用于暴露条件中等的大规模结构应用场景，因为此类场景希望降低环境影响。

在这两种情景下，我们还要求满足最小强度、最大密度和最高财务成本方面的约束。两组目标标准汇总见表1。现在，我们计算每种假设配比成功满足目标标准的概率，以便重点关注实际应用中最稳健的混合料。首先，对于给定假设配比的每个目标变量（碳化系数、环境影响、抗压强度、密度和成本），我们构建机器学习预测结果的概率分布曲线，以反映预测中的不确定性。然后，计算该曲线在目标区域内的面积。最后，我们

https://doi.org/10.1017/dce.2023.5 剑桥大学出版社在线出版

表1. 从可用配合比设计中进行选择时采用的目标标准。

目标标准	Low-K	Low-E
碳化系数（mm $day^{-1/2}$）	< 1:2	< 2:4
环境影响（kg CO_2e/kg）	< 0:150	< 0:108
抗压强度 (MPa)	> 30	> 20
密度 (kg/m^3)	< 2,350	< 2,350
成本 (£/kg)	< 0:028	< 0:028

注：所选混合料是最有可能满足所有标准的混合料。

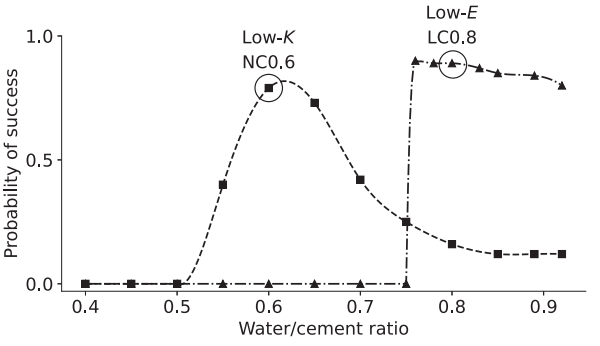


图 5. 满足 Low-K（虚线，方形标记）和 Low-E（点划线，三角形标记）标准的假设配合比系列的成功概率，按水灰比绘制。所选配合比已用圆圈标出。

将每个目标变量对应的面积相乘，得到给定配比成功满足目标标准的概率。我们寻求成功概率最大的配比。

图5展示了水灰比在0.4至0.95范围内的一系列假设配合比满足目标标准的成功概率。之所以特别改变水灰比，是因为图1显示它是配合比参数中的一个重要变量。曲线采用三次样条拟合绘制。

在候选混合料中，确定了水灰比为0.6的混合料（NC0.6）和水灰比为0.8的混合料（LC0.8）。总体而言，提高水灰比可改善环境影响，而降低水灰比则可提升抗碳化性。NC0.6在两者之间取得了折中，满足了*K* 中概述的低表1 标准，并具有较高的成功概率。对于低*E* 标准，环境影响的精确模型（见图4b中的表格）不确定性较小，促使图5 中水灰比超过0.75时的成功概率骤然上升。选择LC0.8，使其与边界保持足够距离，以避免拌和过程中的实验误差，同时维持较高的成功概率。选定的两种配合比设计见表2。

5. 实验验证

5.1. 制造

为验证所提出的两种混合料的性能，我们浇筑了若干个尺寸为 100 mm × 100 mm × 100 mm 的试块。这些混合料采用最大粒径为 10 mm 的碎石灰岩砾石，以及最大粒径为 4 mm、其中 80% 颗粒小于 1 mm 的棱角砂。所用组成材料与训练集中混凝土所用的相同。混凝土在盘式搅拌机中总共搅拌 5 min，然后转移至涂油的塑料立方体模具中。填模过程中，将模具置于振动台上振动，总计 12 s，以确保混凝土充分压实，随后用抹刀刮平

https://doi.org/10.1017/dce.2023.5 剑桥大学出版社在线出版

表2。各自最有可能满足相应目标标准的两种配比，因此建议进行实验验证。

目标标准	Low-K	Low-E
选定的配合比	NC0.6	LC0.8
成功概率	0.79	0.89
水泥（CEM I 52.5N）（%）	14.2	10.5
碎石（%）	48.9	48.2
棱角砂（%）	28.4	32.6
水（%）	8.5	8.5
水灰比	0.6	0.8
总骨料/水泥比	5.5	7.7

外露面。混凝土立方体试块在初凝 24 hr 后脱模；在此期间，试块覆以聚乙烯薄膜，以防止水分流失，随后在 20°C 的水中养护至龄期达到 28天。浇筑期间，出于实验目的，在 LC0.8 混合料中加入了微量染料（按质量计为 0.2%）。我们测量了碳化系数、立方体抗压强度和空气中水饱和密度。根据已发表的成本估算电子表格（Fibo Intercon，2019年）和 ICE 数据库（Jones and Hammond，2019年），分别采用各组成材料的系数计算了每种混合料的成本和隐含碳。假设质量测量的不确定度为 0.7%（质量为 15 kg 时，称重秤的精度为 0.1 kg）。其他实验值的不确定度则通过重复测量进行量化。

5.2. 单项性能验证

在本节中，我们测量并验证两种混凝土混合料的五项属性。两种混凝土配合比的五项属性（碳化系数、环境影响、强度、密度、成本）的机器学习预测和实验值呈现在图6中。对于两种混凝土配合比，所有预测结果均在标准误差范围内与相应的实验值一致。下面我们将讨论每项单独属性的测量和验证。

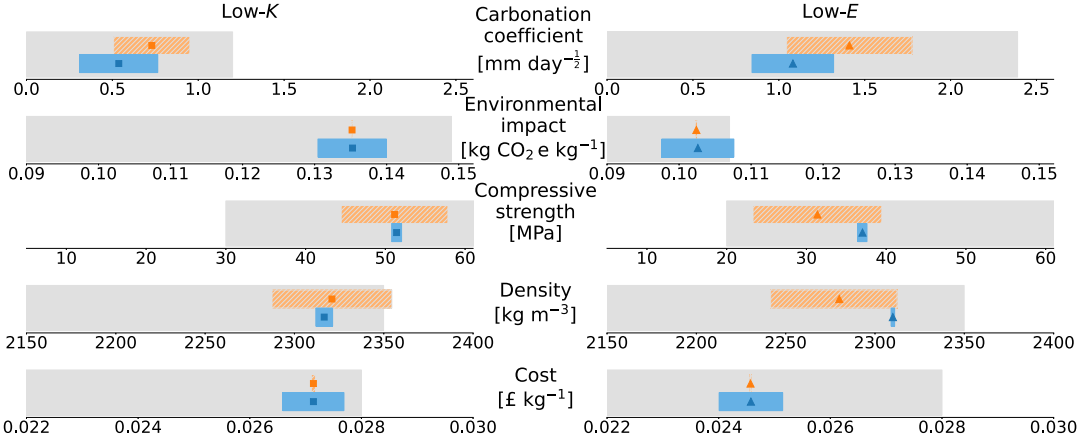


图6. 两种混凝土配合比各项性能的机器学习预测结果（橙色斜线填充）与实验结果（蓝色）汇总。柱形表示预测值和实验值的标准误差范围。灰色区域表示性能目标。

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

5.2.1. 碳化系数

当混凝土暴露于 CO_2时，水泥浆体基质的孔隙中会发生碳化反应。根据菲克'第一定律，该过程被建模为一维扩散（Kropp 和 Hilsdorf，1995），其中 CO_2, 的渗透深度（即碳化深度）x 的平方与暴露时间t成正比，比例系数为碳化系数K，而该系数本身是混凝土'的性质和环境 CO 浓度$_2$的函数。考虑所有边界条件后，该关系定义如下（Moreno，2013）：

$$x(t) = \sqrt{x(0)^2 + K^2 t}, \tag{1}$$

其中，t 为恒定 CO_2 浓度下的暴露时间，x (0) 为 t $=0$ 时的初始碳化深度（通常为 0 mm）。

由于大气浓度下的碳化过程极其缓慢，需要数年才能达到显著的碳化深度，因此需在较高 CO_2浓度下进行加速碳化试验，以评估不同混凝土配合比的相对性能。典型的试验浓度为 4%，由此得到的 K 值在本文中称为 4%加速碳化系数，$K_{4\%}$。

新鲜劈裂的混凝土表面上，未碳化材料呈现图7 中的粉红色，这是酚酞指示剂溶液显色所致。碳化材料仍呈灰色。与水泥浆基体和孔隙空间相比，骨料通常对 CO_2 的渗透性较低。因此，骨料会阻碍碳化在水泥浆中的推进，形成不均匀且曲折的碳化前沿（Huang 等，2012年；Shen 和 Pan，2017），如图7所示。为考虑这种曲折性，按照 BS 1881-210:2013（BSI，2013）所述方法，在碳化前沿上多个等距位置测量碳化深度。暴露时间为t时的碳化深度 x t( ) 取这些测量值的平均值，如图7所示。

将碳化深度数据转换为4%加速碳化系数，$K_{4\%}$ 方法是对每种配比的所有可用数据点进行曲线拟合：将$x\,t(\,)$与ffiffi$t\sqrt{}$拟合至方程（1），如图8所示。曲线拟合采用 BS EN 12390-12:2020（BSI，2020年）中的方法。碳化系数误差源于碳化前沿上$x\,t(\,)$的变异性。为估算该误差，我们对$x\,t(\,)+\sigma_x t(\,)$与ffiffi$t\sqrt{}$进行曲线拟合，其中$\sigma_x t(\,)$为给定t时碳化深度的标准差，据此求得碳化系数的上界。然后，用上界减去$K_{4\%}$ 值，以估算碳化系数误差。由于假设碳化系数服从正态分布，也可使用其下界得到相同的估算值。

可以看出，见图6 机器学习预测的碳化系数与实验测得值在标准误差范围内一致。我们注意到，机器学习预测结果是

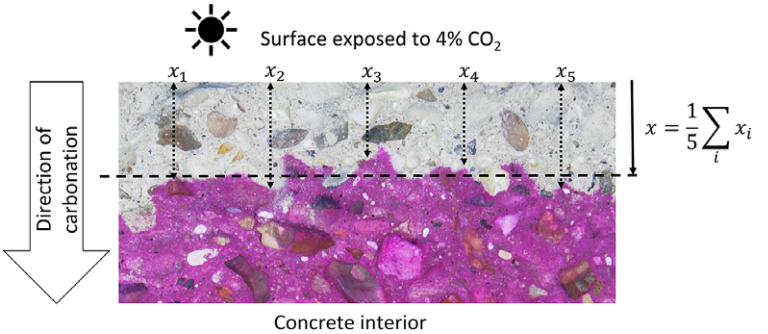


图7. 骨料/水泥比为6.9的混凝土试样，上表面暴露于4% CO_2 中49天，其他表面与试样其余部分接触。使用乙醇中的1%酚酞指示剂溶液显现碳化前沿（未碳化区域呈品红色）。

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

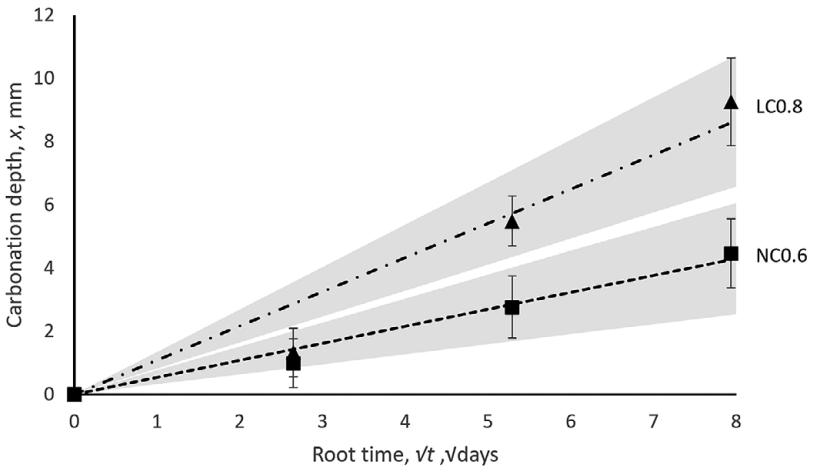


图8. 碳化试验结果，显示碳化系数的估计值（黑色线性拟合）和标准误差范围（灰色区域）。

对于两种混合料而言，该值均高于相应的实验值，这可能是由于训练数据中高碳化系数值占比较高所致。

5.2.2. 环境影响

新提出的混凝土配合比的环境影响通过公式（2）进行估算：

$$\mathrm{kgCO_2e/kg_{concrete}} = \sum_i e_i f_i, \tag{2}$$

其中，e_i 表示第 i^{th} 种材料每千克所含的隐含 CO_2，f_i 表示第 i^{th} 种材料在混凝土配合比中的质量分数。

各组成材料中隐含 CO_2 的系数 c_i取自 ICE 数据库（Jones and Hammond，2019年），并列于补充材料中。假设质量测量误差为 0.7%，根据实验系列所使用秤的精度，每称量 15 kg 的材料存在 ± 0.1 kg 的误差，据此估算该数值的实验误差。

图6显示，机器学习预测的环境影响与实验值高度一致。这是因为环境影响是混凝土组成比例的近似线性函数，因此可以直接进行预测。

5.2.3. 强度

对于每种新提出的混凝土配合比，在经过28天湿养护期后，按照BS EN 12390-3:2019（BSI，2019a）测量三个100 mm ×100mm × 100 mm立方体试件的立方体抗压强度，并假设立方体强度服从正态分布，计算立方体平均强度和标准差（Dayaratnam and Ranganathan，1976年）。这是现场混凝土最常测量的性能，因为它决定了采用该配合比建造的结构的承载能力。

从图6可以看出，机器学习预测的强度与实验测得值之间的差异在标准误差范围内。对于两种混合料，机器学习预测值均低于相应的实验值，这是因为训练集中水泥含量相近的混合料以较低强度值为主。

https://doi.org/10.1017/dce.2023.5 剑桥大学出版社在线出版

5.2.4. 密度

按照 BS EN 12390- 7:2019 (BSI, 2019年b) 所述方法测量混凝土在空气中的水饱和密度，试件为边长 100 mm 的立方体 $\times 100\,\text{mm} \times 100\,\text{mm}$。每种混合料测量三个立方体，并将这些测量值的标准差记录为实验误差。实验误差较低表明新拌混凝土拌和均匀，这也应能降低其他性能的变异性。

可以看出，在图6 中，机器学习对密度的预测结果与实验测量值在标准误差范围内一致。这是一项特别值得肯定的成果，因为双层模型在R^2 值在图4b 中低于其他性质。对于 NC0.6 Low-*K* 混合料，机器学习预测值高于实验值；而对于 LC0.8 Low-*E* 混合料，机器学习预测值低于实验值，也就是说，机器学习预测结果的方差大于实验值。这可能是因为密度随骨料粒径变化，即使是相似的混合料，其骨料粒径也可能不同，从而使密度难以预测。

5.2.5. 成本

混凝土的成本使用公式（3）计算。该值的实验误差采用与环境影响实验误差相同的假设进行估算。

$$£/\text{kg}_{\text{concrete}} = \sum_i c_i f_i, \tag{3}$$

其中，c_i 是第 i^{th} 种材料的每千克价格，f_i 是第 i^{th} 种材料在混凝土配合比中的质量分数。

在商业应用中，混凝土配合比的定价不仅在很大程度上取决于配合比，也取决于项目或应用的规模。因此，我们采用 Fibo Intercon 提供的2019年中型批次代表性数值，以便在优化时反映不同混凝土之间合理的相对数值，但这些数值不应视为绝对值。所用数值见补充材料。成本的实验误差估计为 0.7%，这是根据实验系列所用秤的精度，假设质量测量存在误差得出的；每称量 15 kg，误差为 $\pm$ 0.1 kg。

从图6 可以看出，机器学习预测的成本与实验值高度吻合。这是因为成本与环境影响类似，是混凝土组成比例的近似线性函数，因此相对容易预测。

6. 讨论

6.1. 对混凝土设计与规范的影响

本研究提出并经实验验证的混凝土，展示了机器学习方法预测混凝土性能的潜力。我们选取碳化系数、环境影响、强度、密度和成本作为混凝土规范中的约束示例，但这一方法也可扩展至其他性能，例如：新拌状态性能（如坍落度、湿密度）、结构响应（如刚度、抗折抗拉强度）或耐久性表现（如渗透性、孔隙率）。

基于性能的规范是混凝土耐久性设计中的一个日益受到关注的研究领域，其中包括抗碳化性能设计（Younsi 等，2011年；von Greve-Dierfeld 和 Gehlen，2016a、2016b；Wally 等，2022年）。基于性能的方法有望实现更加经济且可持续的结构（Teplý 和 Vořechovská，2009年）。然而，为满足性能要求，必须通过试验来证明混凝土性能，例如本研究所采用的抗压强度试验和加速碳化试验。利用以往试件性能方面的知识，机器学习能够在选择混合料时减少对此类试验中反复试错的需求，从而为建设项目节省宝贵时间。这还可以实现对多个不同参数进行选择，包括环境影响，这意味着在满足其他要求的同时仍可优先考虑可持续性

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

。这可能开启即时混凝土设计的时代，在施工现场制定定制化混合料，以提供最佳性能。

6.2. 机器学习的未来用途

本研究采用的机器学习方法可处理任意数量的输入特征和目标变量。该通用算法可应用于混凝土图像（Lemaire et al., 2005），其中碳化前沿的噪声或骨料的分布可能包含有关混凝土性能的信息，并有助于改进对这些性能的预测。该方法还可用于预测混凝土性能的时间动态。这有助于确定在整个生命周期内始终满足既定目标的混凝土配合比。

从噪声中提取信息的方法在混凝土设计之外的领域也具有潜在应用。其中一个领域是自动驾驶车辆。在这种情况下，测量目标物体距离时的不确定性包含有关其形状或类型的信息。例如，距离测量的不确定性较高，说明该物体可能是栅栏；不确定性较低则可能表明该物体是墙体。该方法还可应用于增材制造（Rasiya et al., 2021），利用金属粉末微观结构中的噪声来设计最佳熔化工艺。另一个潜在应用领域是信息工程，可利用传感器传输数据中的噪声进行目标跟踪（Płaczek and Bernaś,2014）。除工程领域外，该方法还可用于癌症研究，因为癌症已知会导致遗传混沌（Calin et al., 2003）。从这种混沌中提取的信息有望用于早期癌症检测。

6.3. 结论

本研究介绍了如何使用双层随机森林回归模型，筛选出最有可能成功达到各项目标性能的混凝土配合比设计。该模型能够从噪声中提取信息，因此特别适用于混凝土基体中骨料的随机分布；这种分布会导致混凝土性能（如强度和碳化系数）产生变异性。即使使用稀疏数据训练，该方法也表现有效：对于难以预测的密度，其留一交叉验证 R^2 值高于 0.50；对于环境影响和成本，数值则高于 0.99。这表明其总体性能优于单层线性模型和单层随机森林模型。对两种盲配合比的预测结果经过实验验证，误差在标准误差范围内。总体而言，这些结果表明，机器学习在未来有望利用噪声，实现针对多项性能的混凝土基于性能的设计，并应用于其他材料和领域。

致谢。作者感谢剑桥大学土木工程大楼实验室技术人员协助开展本文所述的实验工作。本文采用开放获取模式，网址为https://www.openaccess. cam.ac.uk。为实现开放获取，作者已对由此产生的任何作者接受稿版本采用知识共享署名（CC BY）许可协议。

作者贡献。概念构思：J.C.F.、B.Z.、J.M.L.、G.J.C.；数据整理：J.C.F.；正式分析：J.C.F.、B.Z.；经费获取：J.M.L.、G.J.C.；调查研究：J.C.F.；研究方法：J.C.F.、B.Z.；项目管理：J.M.L.、G.J.C.；软件：B.Z.；监督指导：J.M.L.、G.J.C.；验证：J.C.F.、B.Z.；可视化：J.C.F.、B.Z.；撰写—初稿：J.C.F.、B.Z.；撰写—审阅与编辑：J.M.L.、G.J.C.

利益冲突。G.J.C.声明，其作为机器学习公司 Intellegens Ltd. 的董事，存在潜在的财务利益冲突。其他作者声明不存在利益冲突。

数据可用性声明。本研究的相关数据见补充材料，也可在以下网址找到 https://doi.org/10.17863/CAM.92245。

经费声明。作者感谢工程与自然科学研究委员会（资助编号：EP/N017668/1 和 EP/N509620/1）、哈丁杰出研究生学者计划杠杆资助项目以及英国皇家学会（资助编号：URF\ R\ 201,002）提供的经费支持。资助方未参与研究设计、数据收集与分析、发表决定或稿件撰写。

补充材料。本文的补充材料可在以下网址获取：http://doi.org/10.1017/dce.2023.5

参考文献

Abrams DA（1918年）《混凝土混合料设计》。技术报告。芝加哥刘易斯学院结构材料研究实验室。网址：http://www2.cement.org/pdf_files/ls001.pdf。 [访问日期：15/11/22]。ACI 第211委员会（1991年）《普通混凝土、重混凝土和大体积混凝土配合比选用标准规范》（ACI 211.1-91）。技术报告，2002年重新批准。美国混凝土协会，密歇根州法明顿希尔斯。Agrawal A、Deshpande PD、Cecen A、Basavarsu GP、Choudhary AN 和 Kalidindi SR（2014）利用数据科学技术，根据钢材成分和加工参数预测其疲劳强度。《材料与制造创新集成》3，1–19。https://doi.org/10.1186/2193-9772-3-8 Aïtcin PC（2016）2——水泥水化的现象学。载于 Aitcin PC 和 Flatt RJ（编），《混凝土外加剂科学与技术》。剑桥：Woodhead 出版社（Elsevier），第15–25页。https://doi.org/10.1016/B978-0-08-100693-1.00002-3 Allwood JM 和 Cullen JM（2012）《可持续材料：睁开双眼》，第1版。剑桥：UIT Cambridge Ltd. 网址：https://www.uselessgroup.org/publications/book/chapters。 [访问日期：21/10/19]。Babor D、Plian D 和 Judele L（2009）混凝土的环境影响。《雅西理工学院学报：建筑学部分》LV（LIX）4，27–36。Ben Chaabene W、Flah M 和 Nehdi ML（2020）机器学习对混凝土力学性能的预测：批判性综述。《建筑与建材》260(7)，1–18。https://doi.org/10.1016/j.conbuildmat.2020.119889 Bhadeshia H、MacKay D 和 Svensson L（1995）C–Mn钢电弧焊缝的冲击韧性：贝叶斯神经网络分析。《材料科学与技术》11，1046–1051。https://doi.org/10.1179/mst.1995.11.10.1046 英国标准协会（2013）BS 1881-210:2013《硬化混凝土试验》第210部分：混凝土潜在抗碳化性的测定：加速碳化法。BSI Standards Limited。英国标准协会（2016）BS EN 206:2013+A1；《混凝土——规范、性能、生产和合格性》。英国标准协会（2019a）BS EN 12390-3:2019，《硬化混凝土试验》第3部分：试件的抗压强度。BSI Standards Limited。英国标准协会（2019b）BS EN 12390-7:2019，《硬化混凝土试验》第7部分：硬化混凝土的密度。BSI Standards Limited。英国标准协会（2020）BS EN 12390-12:2020，《硬化混凝土试验》第12部分：混凝土抗碳化性的测定——加速碳化法。BSI Standards Limited。Calin GA、Vasilescu C、Negrini M 和 Barbanti-Brodano G（2003）人类癌症中的遗传混沌与反混沌。《医学假说》60(2)，258–262。https://doi.org/10.1016/s0306-9877(02)00383-3 Chen X、Wu S 和 Zhou J（2013）孔隙率对水泥砂浆抗压强度和抗拉强度的影响。《建筑与建材》40，869–874。https://doi.org/10.1016/j.conbuildmat.2012.11.072 Conduit B、Illston T、Baker S、Vadegadde Duggappa D、Harding S、Stone H 和 Conduit G（2019）利用概率神经网络识别用于直接激光沉积的合金。《材料与设计》168，107644。https://doi.org/10.1016/j.matdes.2019.107644 Conduit B、Jones N、Stone H 和 Conduit G（2017）利用神经网络设计镍基高温合金。《材料与设计》131，358–365。https://doi.org/10.1016/j.matdes.2017.06.007 Conduit B、Jones N、Stone H 和 Conduit G（2018）利用神经网络对钼基合金进行概率设计。《Scripta Materialia》146，82–86。https://doi.org/10.1016/j.scriptamat.2017.11.008 Dayaratnam P 和 Ranganathan R（1976年）混凝土强度的统计分析。《建筑与环境》11(2)，145–152。https://doi.org/10.1016/0360-1323(76)90029-9 Fibo Intercon（2019）混凝土成本计算器。网址：https://fibointercon.com/wp-content/uploads/2019/06/Concrete- Calculator.xlsx（访问日期：2022年5月10日）。Forsdyke JC 和 Lees JM（2021a）对适用于自然条件试验的现有碳化系数模型进行分析。载于 Júlio E、Valença J 和 Louro AS （编），《混凝土结构：生态效率与性能的新趋势：2021 年 fib 研讨会论文集》，2021 年 6 月 12 日至16日在葡萄牙里斯本在线举行（第361–370页）。国际混凝土联合会（fib）。https://doi.org/10.17863/CAM.75504 Forsdyke JC 和 Lees JM（2021b）使用图像处理工具测量暴露于不同养护和预处理条件下混凝土的碳化深度。载于 Del Zoppo M 和 Colombo IG（编），《第二届 fib 意大利青年成员组混凝土与混凝土结构研讨会论文集》，2021年11月18日至19日在意大利罗马举行（第265–272页）。国际混凝土联合会（fib）。https://doi.org/10.17863/CAM.78987 Goujon B（2009）信息提取中的不确定性检测。载于 Angelova G 等（编），《2009年自然语言处理最新进展国际会议论文集》，2009年9月14日至16日在保加利亚博罗韦茨举行（第118–122页）。Incoma Ltd. https://aclanthology.org/R09-1 Greve-Dierfeld SV、Lothenbach B、Vollpracht A、Wu B、Huet B、Andrade C、Thiel C、Gruyaert E、Vanoutrive H、Sae IF、Provis JL、Kosmas K-CT、Maciej S 和 Natalia Z（2020）理解含辅助胶凝材料混凝土的碳化：国际材料与结构研究实验联合会 TC 281-CCC 批判性综述。《材料与结构》53，136。https://doi.org/10.1617/ s11527-020-01558-w Groves GW、Brough A、Richardson IG 和 Dobsont CM（1991年）硬化C3S水泥浆体因碳化而发生的结构渐进变化。《美国陶瓷学会学报》74(11)，2891–2896。https://doi.org/10.1111/j.1151- 2916.1991.tb06859.x

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

Hastie T、Tibshirani R 和 Friedman J（2001）《统计学习的要素》，第2版。纽约州纽约：施普林格。https://doi.org/10.1007/978-0-387-84858-7 Heskes T（1997 9, 176 182）对数意见池中的权重因子选择。《神经信息处理系统进展》–。Huang Q、Jiang Z、Zhang W、Gu X 和 Dou X（2012）粗骨料分布对混凝土碳化影响的数值分析。《建筑与建材》37，27–35。https://doi.org/10.1016/j.conbuildmat.2012.06.074 Irwin B、Levell J、Whitehead T、Segall M 和 Conduit G（2020a）深度学习在填补异质药物发现数据中的实际应用。《化学信息与建模杂志》60，2848。https://doi.org/10.1021/acs.jcim.0c00443 Irwin B、Mahmoud S、Whitehead T、Conduit G 和 Segall M（2020b）填补与预测：机器学习在药物发现中的应用。《未来药物发现》2(2)，FDD38。https://doi.org/10.4155/fdd-2020-0008 Jiang C、Gu X、Huang Q 和 Zhang W（2018）气候条件变化和交通荷载增加情况下混凝土桥梁的碳化深度预测。《水泥与混凝土复合材料》93，140–154。https://doi.org/10.1016/j.cemconcomp.2018.07.007 Jones CI 和 Hammond GP（2019）ICE数据库V3.0。可访问：http://www.circularecology.com/embodied-energy-and- carbonfootprint-database.html#.XcFb0pr7SUk。 [访问日期：02/06/22]。Kim K、Ward L、He J、Krishna A、Agrawal A 和 Wolverton C（2018）机器学习加速的高通量材料筛选：新型四元霍斯勒化合物的发现。《物理评论：材料》2，123801。https://doi.org/10.1103/PhysRevMaterials.2.123801 Kropp J 和 Hilsdorf H（编）（1995）国际材料与结构研究实验联合会报告12：《混凝土耐久性的性能标准：由国际材料与结构研究实验联合会技术委员会TC 116-PCD编写的现状报告——混凝土性能作为其耐久性的标准》。伦敦：E & FN Spon。Lemaire G、Escadeillas G 和 Rignot E（2005）使用图像分析过程评估混凝土表面。《建筑与建材》19(8)，604–611。https://doi.org/10.1016/j.conbuildmat.2005.01.025 Liu K、Alam MS、Zhu J、Zheng J 和 Chi L（2021）采用与群体智能算法相结合的人工神经网络预测再生骨料混凝土的碳化深度。《建筑与建材》301，124382。https://doi.org/10.1016/ j.conbuildmat.2021.124382 Loh W（2011 1, 14 23 10 1002 8）分类与回归树。《Wiley跨学科评论：数据挖掘与知识发现》–。https://doi.org/ . /widm。Ma M、Tam V、Le K 和 Osei-Kyei R（2022）影响再生混凝土价格的因素：批判性综述。《建筑工程杂志》46，103743。https://doi.org/10.1016/j.jobe.2021.103743 Mahmoud S、Irwin B、Chekmarev D、Vyas S、Kattas J、Whitehead T、Mansley T、Bikker J、Conduit G 和 Segall M（2021）使用深度学习填补感官性质数据。《计算机辅助分子设计杂志》35，1125–1140。https:// doi.org/10.1007/s10822-021-00424-3 Marar K 和 Eren Ö（2011）无外加剂条件下水泥含量和水灰比对新拌混凝土性能的影响。《国际物理科学杂志》6(24)，5752–5765。Mindess S、Young JF 和 Darwin D（2003）《混凝土》，第2版。霍博肯：Prentice Hall。Moreno EI（2013）高吸水性石灰岩骨料制成的混凝土的碳化系数。《材料科学与工程进展》，2013，734031。https://doi.org/10.1155/2013/734031 Neville AM（2011）《混凝土性能》，第5版。伦敦：Pearson Education。Page CL（2007）混凝土中钢筋的腐蚀与防护。载于Page MM和Page CL（编）《混凝土与水泥复合材料的耐久性》。剑桥：伍德海德出版有限公司，第136–186页。https://doi.org/10.1533/9781845693398.136 Pann K、Yen T、Tang C-W 和 Lin T（2004）基于水灰比和毛细孔隙率的新强度模型。《美国混凝土学会材料期刊》101(100)，249–250。Pantazopoulou SJ 和 Mills RH（1995）素混凝土力学响应的微观结构方面。《美国混凝土学会材料期刊》92(6)，605–616。https://doi.org/10.14359/9780 Papadakis VG、Vayenas CG 和 Fardis MN（1989）混凝土碳化问题的反应工程方法。《AICHE杂志》35(10)，1639–1650。https://doi.org/10.1002/aic.690351008 Papadakis VG、Vayenas CG 和 Fardis MN（1991）影响混凝土耐久性的物理和化学特性。《美国混凝土学会材料期刊》88(2)，186–196。https://doi.org/10.14359/1993 Pedregosa F、Varoquaux G、Gramfort A、Michel V、Thirion B、Grisel O、Blondel M、Prettenhofer P、Weiss R、Dubourg V、Vanderplas J、Passos A、Cournapeau D、Brucher M、Perrot M 和 Duchesnay É（2011）Scikit-learn：Python中的机器学习。《机器学习研究杂志》12，2825–2830。Płaczek B 和 Bernaś M（2014）无线传感器网络中基于不确定性的信息提取及其在控制应用中的应用。《自组织网络》14，106–117。https://doi.org/10.1016/j.adhoc.2013.11.009 Poon CS 和 La m CS（2008）骨料水泥比及骨料类型对预制混凝土砌块性能的影响。《水泥与混凝土复合材料》30(4)，283–289。https://doi.org/10.1016/j.cemconcomp.2007.10.005 Powers TC（1958）硬化硅酸盐水泥浆体的结构和物理性质。《美国陶瓷学会杂志》41(1)，7。Prayogo D、Santoso D、Wijaya D、Gunawan T 和 Widjaja J（2020）使用集成机器学习方法预测混凝土性能。《物理学杂志：会议系列》1625，012024。https://doi.org/10.1088/1742-6596/1625/1/012024

Rasiya G、Shukla A 和 Saran K（2021）增材制造——综述 [国际设计、材料与制造进展会议]。《今日材料：会议录》47，6896–6901。https://doi.org/10.1016/j.matpr.2021.05.181 Sedgwick J（1991）强大但敏感。《大西洋月刊》267，70–82。可访问：https://www.theatlantic.com/magazine/archive/1991/04/strong-butsensitive/668990/。 [访问日期：09/09/2020] Shen Q 和 Pan G（2017）不同骨料含量水泥基材料的碳化。《材料科学论坛》902，74–78。https://doi.org/10.4028/www.scientific.net/MSF.902.74 Silva A、Neves R 和 De Brito J（2014）钢筋混凝土碳化的统计建模。《水泥与混凝土复合材料》50，73–81。https://doi.org/10.1016/j.cemconcomp.2013.12.001 Sims I、Lay J 和 Ferrari JI（2019）混凝土骨料。载于《Lea水泥与混凝土化学》，第5版。阿姆斯特丹：Elsevier，第699–778页。https://doi.org/10.1016/B978-0-08-100773-0.00015-0 Sourmail T、Bhadeshia H 和 MacKay D（2002）奥氏体不锈钢蠕变强度的人工神经网络模型。《材料科学与技术》18，655–663。https://doi.org/10.1179/026708302225002065 Sousa V 和 Bogas J（2021）熟料和再生水泥生产中的能耗与碳排放比较。《清洁生产杂志》306，127277。https://doi.org/10.1016/j.jclepro.2021.127277 Taffese W、Sistonen E 和 Puttonen J（2015）CaPrM：使用机器学习方法的钢筋混凝土碳化预测模型。《建筑与建材》100，70–82。https://doi.org/10.1088/1742-6596/1625/1/012024 Talukdar S 和 Banthia N（2016）全球气候变化背景下混凝土基础设施的碳化：模型改进与代表性浓度路径情景评估。《土木工程材料杂志》28(4)，04015178。https://doi.org/10.1061/(asce)mt.1943-5533.0001438 Tancret F（2013）计算热力学、高斯过程与遗传算法：设计新合金的组合工具。《材料科学与工程中的建模与仿真》21，045013。https://doi.org/10.1088/0965-0393/21/4/045013 Teplý B 和 Vořechovská D（2009）《混凝土结构基于性能的设计：参数化方法》（1），1–16。Teychenné DC、Franklin R、Erntroy H 和 Marsh BK（1997）《普通混凝土配合比设计》，第2版。Construction Research Communications Ltd.，建筑研究院（BRE）。Tran VQ、Mai H-VT、To QT 和 Nguyen MH（2022）采用机器学习方法研究含粉煤灰混凝土的碳化深度。《结构混凝土》。https://doi.org/10.1002/SUCO.202200269 联合国气候变化框架公约（2018）世界水泥协会敦促采取气候行动。UNFCCC。可访问：https://unfccc.int/news/world-cement-association-urges-climate-action。 [访问日期：18/02/22]Velumani P 和 Nampoothiri N（2018）印度水泥价格预测：回归分析、趋势分析和人工神经网络的比较。《国际土木工程与技术杂志》9，1907–1915。Verpoort P、MacDonald P 和 Conduit G（2019）利用人工神经网络进行材料数据验证与填补。《计算材料科学》147，176–185。https://doi.org/10.1016/j.commatsci.2018.02.002 von Greve-Dierfeld S 和 Gehlen C（2016a）基于性能的耐久性设计，碳化第1部分——欧洲现行设计规则的基准比较。《结构混凝土》17(3)，309–328。https://doi.org/10.1002/SUCO.201600066 von Greve-Dierfeld S 和 Gehlen C（2016b）基于性能的耐久性设计，碳化第2部分——混凝土分类。《结构混凝土》17(4)，523–532。https://doi.org/10.1002/suco.201600067 Wally GB、Magalhães FC 和 Pinto da Silva Filho LC（2022）从规范性设计到基于性能的设计：耐久混凝土规定国际趋势概述。《建筑工程杂志》52，104359。https://doi.org/10.1016/j.jobe.2022.104359 Ward L、Agrawal A、Choudhary A 和 Wolverton C（2016）用于预测无机材料性能的通用机器学习框架。《NPJ计算材料》2，1–7。https://doi.org/10.1038/npjcompumats.2016.28 Ward L、Liu R、Krishna A、Hedge V、Agrawal A、Choudhary A 和 Wolverton C（2017）通过Voronoi镶嵌将晶体结构属性纳入形成能机器学习模型。《物理评论B》96，024104。https://doi.org/10.1103/ PhysRevB.96.024104 Wassermann R、Katz A 和 Bentur A（2009）最低水泥含量要求：必要条件还是神话？《材料与结构/Matériaux et Constructions》42(7)，973–982。https://doi.org/10.1617/s11527-008-9436-0 Whitehead T、Irwin B、Hunt P、Segall M 和 Conduit G（2019）使用深度学习填补测定生物活性数据。《化学信息与建模杂志》59，1197。https://doi.org/10.1021/acs.jcim.8b00768 Wilson ML 和 Kosmatka SH（2011）《混凝土混合料的设计与控制》，第15版。伊利诺伊州斯科基：波特兰水泥协会。Younsi A、Turcry P、Rozire E、Aït-Mokhtar A 和 Loukili A（2011）高粉煤灰含量混凝土的基于性能的设计与碳化。《水泥与混凝土复合材料》33(10)，993–1000。https://doi.org/10.1016/j.cemconcomp.2011.07.005 Zerva C、Batista-Navarro R、Day P 和 Ananiadou S（2017）利用不确定性连接并排序生物医学文献中的证据以进行模型整理。《生物信息学》33(23)，3784–3792。https://doi.org/10.1093/bioinformatics/btx466 Zhang L（2020）导数定价的一般框架。《数学金融杂志》10，255–266。https://doi.org/ 10.4236/jmf.2020.102016 Zviazhynski B 和 Conduit G（2022）揭示未见之处：挖掘噪声中隐藏的信息。《应用智能》。https://doi.org/ 10.1007/s10489-022-04102-1

引用本文：Forsdyke JC、Zviazhynski B、Lees JM 和 Conduit GJ （2023）。利用机器学习进行混凝土的概率选择与设计。以数据为中心的工程，4，e9。doi：10.1017/dce.2023.5

https://doi.org/10.1017/dce.2023.5 由剑桥大学出版社在线出版

A.

Table 1. *Training data. Values given are averages*

Mix	Cement type [CEM -]	Cement [%]	Crushed gravel [%]	Sharp sand [%]	Water [%]	Estimated mass/m^3 [kg/m^3]	Pre-conditioning time [days]	Carbonation coefficient, $K_{4\%}$ [mm/day$^{1/2}$]	Environmental impact [kgCO$_2$e/kg]	Compressive strength, f_{cube} [MPa]	Water sat. density [kg/m^3]	Cost [£/kg]
C25 I	IIA 32.5 R	11.1	52.8	28.1	7.8	2414	141	1.742	0.095	27.46	2337	0.025
C25 II	IIA 32.5 R	11.1	52.8	28.1	7.8	2414	138	2.301	0.095	21.30	2317	0.025
C25 III	IIA 32.5 R	11.1	52.8	28.1	7.8	2414	16	1.648	0.095	28.35	2300	0.025
C25 IV	IIA 32.5 R	11.1	52.8	28.1	7.8	2414	16	2.081	0.095	29.87	2297	0.025
C20 I	IIA 32.5 R	10.0	56.0	27.4	6.5	2448	141	3.493	0.087	22.49	2263	0.024
C18 I	IIA 32.5 R	12.0	52.9	27.3	7.8	2409	138	2.414	0.102	20.31	2260	0.026
C30 I	IIA 32.5 R	14.2	52.5	25.5	7.8	2409	16	1.045	0.119	35.14	2247	0.027
C40 I	IIA 32.5 R	17.4	51.4	23.4	7.8	2411	16	0.558	0.145	42.85	2220	0.029
C45 I	I 52.5 N	14.6	48.5	28.2	8.7	2344	-	-	0.139	49.90	2287	0.027
C55 I	I 52.5 N	13.7	51.5	27.2	7.6	2374	-	-	0.131	61.62	2333	0.027
C50 I	I 52.5 N	16.6	44.6	28.9	9.9	2312	-	-	0.157	53.48	2300	0.029
C20 II	I 52.5 N	10.3	47.7	33.2	8.7	2343	-	-	0.100	25.30	2237	0.024
C35 I	I 52.5 N	8.9	50.4	33.1	7.6	2374	-	-	0.088	38.76	2290	0.024
C30 II	I 52.5 N	9.2	50.9	32.4	7.3	2452	8	2.341	0.090	32.12	2240	0.024
C45 II	I 52.5 N	13.4	49.5	28.9	8.1	2421	8	1.304	0.128	48.26	2357	0.027
C30 III	I 52.5 N	9.2	50.9	32.4	7.3	2452	14	1.349	0.090	32.21	2343	0.024
C45 III	I 52.5 N	13.4	49.5	28.9	8.1	2421	14	0.806	0.128	48.53	2360	0.027
C25 V	I 52.5 N	10.6	48.3	32.1	9.1	2411	17	1.732	0.103	31.93	2310	0.025
C30 IV	I 52.5 N	12.2	48.8	29.9	9.1	2408	17	1.212	0.117	37.85	2307	0.026
C45 IV	I 52.5 N	14.2	48.9	27.9	9.0	2409	17	0.403	0.135	51.01	2330	0.027
C55 II	I 52.5 N	17.0	48.4	25.7	9.0	2409	17	0.062	0.161	59.83	2327	0.029

B.

表*2.* 单层线性模型、单层随机森林模型和双层随机森林模型的超参数。*Scikit-learn* 中各超参数的名称以斜体显示。双层随机森林模型每一层的超参数均相同。

	Single-layer linear model	Single-layer random forest model	Two-layer random forest model	Imputation model
树的数量 *n*_估计器	–	200	512	512
最大特征数 *max*_特征	8	3	6	13
最小样本数 拆分节点所需的 *min*_样本_拆分	–	2	2	2

C.

表 *3.* 用于预测各目标变量（行）的输入变量（列）的相对特征重要性。粉色越深，特征重要性越高。

	Cement type	Cement	Crushed gravel	Sharp sand	Water	Water/ cement	Total aggr./ cement	Sharp sand/ total aggr.	Precondi- tioning time
碳化系数	0.00	0.28	0.23	0.00	0.08	0.00	0.35	0.00	0.06
环境影响	0.04	0.45	0.27	0.24	0.00	0.00	0.00	0.00	0.00
抗压强度	0.06	0.22	0.19	0.00	0.00	0.22	0.22	0.00	0.09
密度	0.11	0.00	0.28	0.00	0.22	0.00	0.00	0.39	0.00
Cost	0.00	0.50	0.00	0.35	0.15	0.00	0.00	0.00	0.00

D.

表*4.* 假定的组成材料隐含排放量和价格。来源：*Fibo Intercon，2019年；Jones and Hammond，2019年*

材料，i	千克CO_2e/千克，e	£/kg，c
CEM IIA水泥	0.799	0.089
CEM I水泥	0.912	0.089
砾石/砂	0.007	0.018
水	0.0008	0.007