

基于卷积神经网络的液化预测模型及可解释性分析

龙 潇^{1,2}, 孙 锐^{1,2}, 郑 桐^{1,2}

(1. 中国地震局工程力学研究所 地震工程与工程振动重点实验室, 黑龙江 哈尔滨 150080;

2. 地震灾害防治应急管理部重点实验室, 黑龙江 哈尔滨 150080)

摘 要: 常规液化判别方法通常是半经验方法, 存在人为因素干扰, 成功率及均衡性不佳。现有的机器学习方法缺乏足够的样本支撑, 存在一定的局限性。通过整合液化数据集, 选取修正标准贯击数、细粒含量、土层深度、地下水位深度、总上覆应力、有效上覆应力、门槛加速度、循环剪应力比、剪切波速、震级与地表峰值加速度 11 个液化特征建立卷积神经网络 (convolutional neural network, 简称 CNN) 模型。引入边界合成少数过采样技术消除不平衡数据集的影响。将 CNN 模型与随机森林模型、逻辑回归模型、支持向量机模型、极致梯度提升模型和规范方法进行对比, 并结合沙普利加性解释 (SHapley Additive exPlanations, 简称 SHAP) 分析输入特征对预测结果的影响趋势。结果表明, CNN 模型准确率达 92.58%, 各项指标均优于其他 4 种机器学习模型和规范方法。对 SHAP 结果分析可知, 修正标准贯击数小于 15 的土层液化概率较高, 循环剪应力比 CSR 小于 0.25 的土层更不易液化。各因素的影响规律均符合现有认知, 预测模型合理可靠。

关 键 词: 机器学习; 液化预测; 卷积神经网络; 边界合成少数过采样技术; 沙普利加性解释 (SHAP)

中图分类号: TU43

文献标识码: A

文章编号: 1000-7598 (2024) 09-2741-13

Convolutional neural network-based liquefaction prediction model and interpretability analysis

LONG Xiao^{1,2}, SUN Rui^{1,2}, ZHENG Tong^{1,2}

(1. Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin, Heilongjiang 150080, China; 2. Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin, Heilongjiang 150080, China)

Abstract: Conventional methods for liquefaction discrimination are often semi-empirical, prone to human factors, with low success rates and balance. Moreover, current machine learning approaches lack ample sample support and have specific limitations. Through the integration of liquefaction datasets, 11 features were chosen: corrected standard penetration test blow count, fine content, soil layer depth, groundwater table depth, total overburden stress, effective overburden stress, threshold acceleration, cyclic shear stress ratio, shear wave velocity, earthquake magnitude, and peak ground acceleration. A convolutional neural network (CNN) model was established. The Borderline SMOTE technique was introduced to address the issue of imbalanced datasets. The CNN model was compared against random forest, logistic regression, support vector machine, extreme gradient boosting models, and methods specified in Chinese codes. Furthermore, the SHapley Additive exPlanations (SHAP) algorithm was utilized to examine the influence trends of input features on the prediction outcomes. The results demonstrated that the CNN model attained an accuracy of 92.58%, outperforming all metrics of the four machine learning models and the method specified in Chinese codes. Examination of the SHAP results unveiled that soil layers with corrected standard penetration blow numbers below 15 exhibited a higher liquefaction probability, whereas layers with cyclic stress ratios under 0.25 were less prone to liquefaction. The influence patterns of each factor align with current understanding, indicating the prediction model's credibility and reliability.

Keywords: machine learning; liquefaction prediction; convolutional neural network; borderline synthesis minority oversampling technique (borderline SMOTE); SHapley Additive exPlanations (SHAP)

收稿日期: 2023-10-25

录用日期: 2024-01-10

基金项目: 中国地震局工程力学研究所基本科研业务费专项资助项目 (No.2020C04); 黑龙江省自然科学基金联合引导项目 (No.LH2020E019)。

This work was supported by the Scientific Research Fund of Institute of Engineering Mechanics, China Earthquake Administration (2020C04) and the Heilongjiang Provincial Natural Science Foundation Joint Guidance Project of China (LH2020E019).

第一作者简介: 龙潇, 男, 1997 年生, 硕士研究生, 主要从事岩土地震工程方面的研究。E-mail: iem_lx@163.com

通讯作者: 郑桐, 男, 1984 年生, 博士, 助理研究员, 主要从事边坡稳定性及抗震加固方面的研究工作。E-mail: zhengt0928@163.com

1 引言

土壤液化是指地震动作用下土的孔隙水压力积累导致土体剪切强度下降、土体完全失去强度与刚度的现象,是造成震害的主要因素之一^[1]。因此,准确判别土壤液化对地震灾害防治工作具有重要意义。

在传统液化判别方法中,以基于标准贯入试验^[2]、静力触探试验^[3]、剪切波速试验^[4]等现场试验的半经验法最为常用。这类方法直接利用经验公式对液化的某些重要影响因素进行运算,并与标准值进行比较,具有较强的实用性。但土壤液化的影响因素众多,与判别结果之间的非线性关系较强,并且传统液化判别方法自身适应性较差,判别结果不均衡。因此,建立一种可以综合考虑多参数、受人为因素影响小的判别方法十分关键。

随着人工智能的不断发展,机器学习方法凭其优异的学习能力广泛应用于各个领域。在岩土工程领域中,Goh^[5]和蔡煜东等^[6]较早地将 BP 神经网络应用于砂土液化预测。刘红军等^[7]采用共轭梯度法等措施对 BP 网络进行改进。陈国兴^[8]、Hanna^[9]等将径向基函数与人工神经网络相结合,以增强模型的非线性逼近能力。但人工神经网络通常为全连接结构,对于复杂地质、土壤和地震样本的学习能力较弱。胡记磊^[10]、Schnidt^[11]等基于贝叶斯理论建立液化概率预测模型,可以对数据不完备样本进行预测。刘年平^[12]、颜可珍^[13]等分别基于 Fisher 判别理论和距离判别分析法构建液化预测模型,并给出相应的判别函数,可以直观了解到不同特征对于判别结果的影响。但此类模型预测结果的可靠性对数据的典型性依赖较强,且文中验证集数据较少,模型性能有待进一步检验。除此之外,还有许多常见的二分类机器学习模型,如支持向量机模型^[14]、随机森林模型^[15]、极致梯度提升模型^[16]。但由于土层参数的多样性、地震作用的不确定性、液化影响因素的复杂性及缺乏充分的数据支撑,使得这些模型对高维数据中重要特征的提取能力以及对大规模液化样本的识别能力等均存在一定程度的不足。同时随着地质勘探与震后调查技术的愈发完善,记录在册的液化数据也将愈发充足。因此,建立一个高效处理大规模数据、避免过拟合、符合现实逻辑的模型至关重要。

2006 年,Hinton 等^[17]提出通过增加隐藏层和“逐层训练”方法克服深度模型难以训练的问题,从而极大地推动了深度学习的研究。卷积神经网络(convolutional neural network,简称 CNN)模型便

是经典的深度学习模型。相较于一般模型,其特有的卷积层、池化层等操作可以有效捕捉数据重要特征,并且减少冗余信息,使模型具有更强的表达能力,同时,模型复杂度降低,受数据间相关性影响更小,更易于训练^[18]。王自法等^[19]以单台站的初至竖向地震波为主输入,建立了基于深层卷积神经网络的震级快速估算模型。骆勇鹏等^[20]建立了基于一维卷积神经网络的结构损伤识方法,并用数值模型验证了该方法的准确性及应用可行性。Yuan 等^[21]同样利用一维卷积神经网络建立了快速震害评估模型,但鲜有学者将其与液化预测问题相结合。

对于一般的分类问题,当数据集中存在样本类别数量不均衡的情况时,会对模型学习产生一定影响,常采用上采样方法对少数类进行补充^[22]。在液化判别的机器学习模型中过去往往采用 k -近邻上采样算法^[16]。但该算法无法识别异常值,易误导模型的特征学习。

此外,现有的机器学习液化预测模型普遍侧重于精确率,为黑盒模型,缺乏对各因素影响规律的分析,难以与工程实际相联系。这也是机器学习技术服务工程的最大阻碍。

目前已有对黑盒模型做可解释性分析的算法,如沙普利加性解释(SHapley Additive explanations,简称 SHAP)^[23]方法、不可知模型局部解释(local interpretable model-agnostic explanation,简称 LIME)^[24]方法。其中 LIME 算法只具备局部可解释性,且对参数设置十分敏感,不同的样本可能导致完全不同的解释。SHAP 可从对单个样本出发做局部解释,也可基于所有数据样本给出全局解释。在地震工程领域中 SHAP 已有较多应用^[25-27]。但在液化判别方面还较为缺乏。

本文将 Hanna^[9]、Cetin^[28]等公布的数据进行整合,以形成包含多场地、多震次且数量可观的数据集;利用卷积神经网络建立性能可靠的液化预测模型;使用合成上采样算法对训练集中少数类数据进行补充,消除不平衡数据集对模型的不利影响^[29];结合可解释性算法分析模型中数据特征与预测结果之间的关系。分析结果将有助于了解黑盒模型中各液化特征对模型预测结果的影响,为未来的模型发展、液化数据收集以及抗液化措施的选择提供有价值的参考信息。

2 数据库建立

2.1 数据来源

1999 年土耳其 Kocaeli 和中国台湾集集地区分

别发生 7.4 级和 7.6 级地震,引起了显著的土壤液化,导致建筑物遭受严重的结构性破坏^[30-31]。2007 年, Hanna 等^[9]对以上两次地震的现场液化数据进行汇总、整理(以下简称, Hanna2007)。数据中包含场地的细粒含量、可液化层埋深、标准贯入击数、地表峰值加速度等 12 项基本场地信息。

2018 年, Cetin 等^[28]通过对以往文章、工程或现场勘察报告进行汇总,收集了 1944—1995 年美国、日本、阿根廷、菲律宾和中国等的 26 个不同国家和地区的液化数据(以下简称 CEA2018),并在 Cetin 等^[32]公布的数据基础上,对循环剪应力折减系数等进行了修正。

2.2 数据整合

Hanna2007 数据集中只有集集与 Kocaeli 地震的数据,相对集中,代表性不强。CEA2018 数据集虽包含不同时间多个地区的数据,但数量较少。这样的数据训练的模型或精度不足,或将只能学习到某些地区的特性,从而难以应用于其他区域。因此,为使液化判别模型不局限于区域,只着眼于数据本身,本文将前述数据集进行整合。

CEA2018 数据中,剪切波速为修正剪切波速,而 Hanna2007 数据为实测剪切波速。为统一尺度,本文依据 Andrus 等^[33]提出的计算方法,将其转化为实测剪切波速。表达式为

$$V_{S1} = V_S (100/\sigma'_V)^{0.25} \quad (1)$$

式中: V_S 为实测剪切波速; V_{S1} 为修正剪切波速; σ'_V

为有效上覆压力。

Hanna2007 数据集中包含门槛加速度,其表达式为

$$\frac{a_t}{g} = \frac{V_S^2 (G/G_{\max})_t \gamma_t}{0.65 g z \gamma_d} \quad (2)$$

式中: a_t 为门槛加速度; g 为重力加速度; γ_t 为门槛应变,设为 0.01%^[34]; $(G/G_{\max})_t$ 为阈值处的模量折减系数,假定为 0.8^[34]; G 为土体有效剪切模量; $G_{\max}$ 为小应变下土体剪切模量; γ_d 为应力折减系数; z 为土样深度。依据以上公式计算 CEA2018 数据中缺少的门槛加速度值。

本文将上述数据进行对比整合,形成了由 28 个不同地区、11 个影响因素组成的液化数据集,其符号与含义如表 1 所示。数据集中液化数据与非液化数据比值为 315:411,数据概览见表 2。

表 1 数据集信息
Table 1 Dataset information

特征名称	单位	特征含义
z	m	土样深度
$(N_1)_{60}$	—	修正的标准贯入击数
FC	%	有效粒径小于 0.075 mm 的细粒含量
d_w	m	地下水位深度
σ_v	kPa	总上覆应力
σ'_v	kPa	有效上覆应力
a_t	g	门槛加速度
CSR	—	循环剪应力比
V_s	m/s	实测剪切波速
M_v	—	震级
$a_{\max}$	g	地表峰值加速度

表 2 数据概览表
Table 2 Data summary

z/m	$(N_1)_{60}$	FC/%	d_w/m	σ_v/kPa	σ'_v/kPa	a_t/g	CSR	$V_s/(m \cdot s^{-1})$	M_v	$a_{\max}/g$	是否液化
9.00	48.00	5.00	0.77	167.30	85.00	0.10	0.44	250.00	7.40	0.40	否
1.00	3.00	74.00	0.78	18.00	15.80	0.14	0.29	105.00	7.40	0.40	是
1.80	5.00	86.00	0.78	32.80	22.60	0.08	0.36	105.00	7.40	0.40	是
3.40	2.00	85.00	0.78	62.40	36.20	0.03	0.42	85.00	7.40	0.40	否
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6.50	43.83	7.00	9.84	125.28	90.97	0.11	0.21	241.97	7.00	0.24	否
4.05	6.81	5.00	2.00	78.16	44.40	0.10	0.25	176.19	7.90	0.23	是
5.05	25.84	5.00	6.41	97.63	67.29	0.10	0.23	195.49	7.00	0.25	否

3 建立 CNN 模型

3.1 卷积神经网络基本原理

卷积神经网络是一种带有卷积结构的前馈神经网络,其基本结构由输入层、卷积层、池化层、全连接层与输出层组成^[35]。本文模型使用的语言环

境是 Python 3.9,编译器为 Jupyter notebook,框架为 PyTorch,使用 GPU 处理,结构如图 1 所示。依据惯例将数据集分成三等份,两份用于模型训练,一份用于模型测试。当下大多常用的场地液化判别方法为“非零即一”的确定性评估,不利于工程师根据工程的重要性做出相应的决策。本文采用

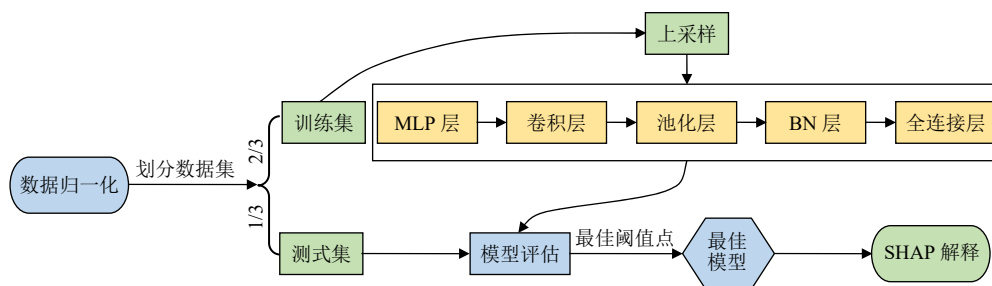


图1 本文模型结构图

Fig.1 Structural diagram of the model in this paper

Sigmoid 函数作为输出层激活函数，它可以将变量映射至 $(0, 1)$ ，以概率的形式输出结果。表达式如下：

$$\text{Sigmoid}(x) = 1 / (1 + e^{-x}) \quad (3)$$

式中： x 为自变量。为使卷积神经网络达到更好的训练效果，本文将多层感知机（multi-Layer perceptron，简称 MLP）层与 CNN 模型结合。

3.2 批归一化

深度神经网络训练的过程中高层单元的输入分布会因各层权重的更新而变化剧烈，使得高层要不断地重新适应底层参数更新，模型训练难度大大增加，本文使用批量归一化（batch normalization，简称 BN）方法，对每层输入分布进行把控。实现方法如下：

（1）将小批次数据白化处理：

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (4)$$

式中： $\hat{x}_i$ 为白化处理后的输入； x_i 为输入数据； μ_β 为平均值； σ_B 为方差； ε 为常数。

（2）进行尺度缩放和偏移操作：

$$y_i = \alpha \hat{x}_i + \beta \quad (5)$$

式中： y_i 为缩放和偏移后的数据； α 、 β 为可学习参数。通过以上处理，使得标准化后的数据通过可学习参数的尺度缩放和偏移，仍可保留前端网络学习到的特征分布^[36]，降低模型调参难度。

3.3 类不平衡

地震现场获取的数据基本属于类不平衡数据，易导致模型预测偏向多数类，产生学习偏差^[26]。

合成少数类过采样技术（synthetic minority over-sampling technique，简称 SMOTE）^[22]是一种经典的上采样算法，通过合成新的少数类样本数据以平衡少数类与多数类样本的比例。边界合成少数类过采样技术（borderline SMOTE）则是 SMOTE

的进阶版^[37]，其把数据点分成 3 种类型：安全（Safe）、危险（Danger）、噪声（Noise）。若基点附近的 k 个样本点中一半以上为同类别数据，则认为该基点为 Safe，无需在该基点附近合成新样本；若 k 个样本中没有同类别样本，则认为该基点为 Noise，不适合在附近合成新样本；若 k 个样本中有同类别数据但不足一半，则认为该基点为 Danger，应在该基点周围补充数据，补充方式采用 SMOTE 算法。

本文依据上述方法对数据集中的少数类进行分类，对判定为 Danger 类型的数据点进行上采样，以此补充本文中的少数类数据，从而平衡模型对不同类别的关注度。

3.4 特征选择

为探究各个因素对模型预测结果的影响规律，将数据库中 11 个因素作为数据特征全部引入模型。为避免模型梯度爆炸以及更好的特征学习，采用最佳归一化将数据映射到 $[0, 1]$ 区间。

土层埋深是表征场地条件的参数之一，通常情况下，土层越深，其抵抗液化的能力越强；地下水位深度对土体是否饱和具有显著意义，地下水位越高，土体越容易处于饱和状态，发生液化的可能性就越大；标准贯入击数是国内外场地液化判别最常用指标^[38]，常用于表征土体的抗液化能力；对于细粒含量，Seed 等^[38]提出，标贯击数一定时，土体抗液化强度随细粒含量的增大而提高。也有学者的试验结果显示，随着细粒含量的增大，抗液化强度反而降低^[39]。虽然结果尚有争论，不过可以确定的是细粒含量会影响土体的抗剪强度；付海清等^[40]通过试验证明，其余条件相同时，上覆压力越大，饱和砂土的抗液化强度就越高。因此，上覆压力是影响场地液化的敏感因素；剪切波速是表征土的动力特性的指标。已有的研究表明，当场地实测剪切波速大于 300 m/s 时，认为土体处于密实状态，不易液化^[41]；循环剪应力比反映地震等外荷载的作用；地表

水平峰值加速度与地震震级是反映地震动强度的最直观的参数。相同的场地条件下,地震动强度越大,发生液化的概率也随之增大。假设门槛地表峰值加速度为砂土层中引起门槛剪应变的地表峰值加速度。Dobry 等^[42]提出,当地震诱发的剪应变小于门槛剪应变时,土层结构完好,不发生液化现象。因此,门槛加速度也可作为判别场地液化与否的因素。以上 11 个影响因素均可对土壤是否液化起到一定的作用,因此被纳入本文模型。

图 2 给出了各特征的分布箱型图。箱型图可直观反映数据的离散程度,所代表统计含义的示意如图 2(l)所示, A 代表液化数据或非液化数据。从箱型图可以看出,大多数特征的中位数均偏于箱子的一侧,各特征的数据分布呈偏态分布,细粒含量的箱型最长,表示其分布的离散性最强。总体来看,多数非液化数据的分布范围大于液化数据分布范围,地下水深度、门槛加速度和震级分布较为集中且液化数据与非液化数据的差异性较小。

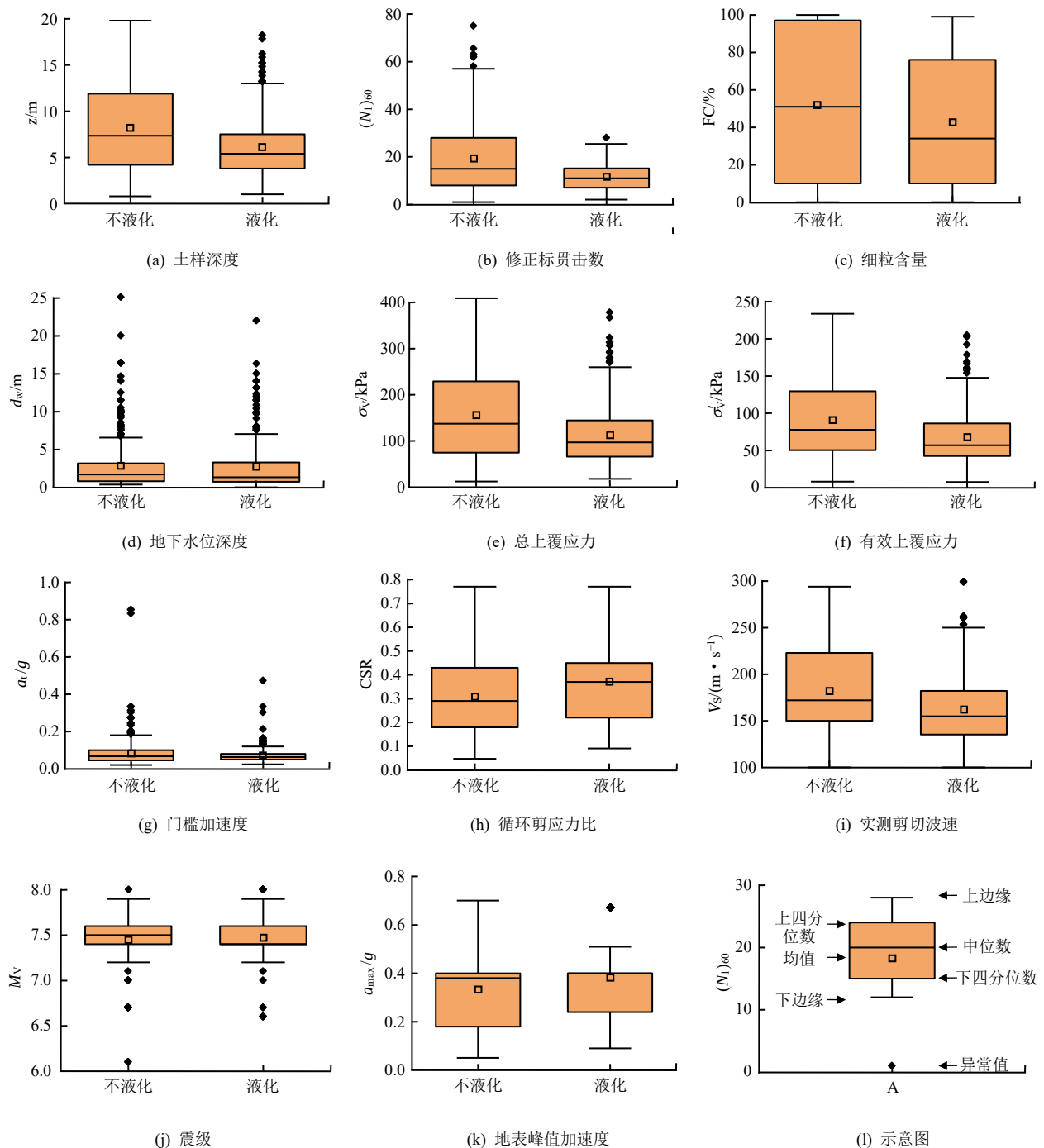


图 2 本文数据集中液化参数箱型图

Fig.2 Box plots of liquefaction parameters in this dataset

4 模型评估

4.1 评估指标

首先将液化假定为 Positive（正），非液化假定为 Negative（负）^[43]。本文利用混淆矩阵（confusion matrix），计算精确率 Pre、召回率 Rec、准确率 Acc 作为评估指标，公式如下：

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (6)$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (7)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\% \quad (8)$$

式中：TP 为液化点被正确识别的数量；FP 为误报为非液化点的数量；FN 为漏报的液化点数量；TN 为非液化点被正确识别的数量。对召回率和精确率求加权平均值 F1 可得

$$\text{F1} = 2\text{Pre} \cdot \text{Rec} / (\text{Pre} + \text{Rec}) \quad (9)$$

F1 取值为 0~1，越接近 1，说明模型能够在正确预测样本的同时，最大化找出所有的正例。但 F1 值过高也存在着模型过拟合的可能。因此引入受试者工作特征（receiver operating characteristic，简称 ROC）曲线。真阳性率 TPR（true positive rate）为

曲线横坐标，代表模型对少数类样本的识别能力。假阳性率 FPR（false positive rate）为纵坐标，代表模型将多数类样本判错的比例。TPR 与 FPR 计算公式如下：

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (10)$$

$$\text{FPR} = \text{FP} / (\text{TN} + \text{FP}) \quad (11)$$

但仅凭 ROC 曲线难以直观分辨模型优劣，因此采用特征曲线下面积 AUC（area under curve）表示曲线下的面积。AUC 的取值区间为 0~1，值越高，模型预测性能越好。对于二分类问题，ROC 曲线亦可给出约登指数 YI（Youden index），如图 3(a) 所示。YI 可以认为是二分类问题的最佳概率阈值。当模型的 AUC 值接近 1 时，认为通过设定最佳阈值，可以使模型达到最佳性能^[16]。

4.2 最佳阈值点

模型训练时初始阈值为 0.5。输出大于 0.5 时判定为液化，反之为不液化。为比较不同机器学习算法对液化判别的适用性，采用了极致梯度提升（XGBoost）、随机森林（random forest，简称 RF）、逻辑回归（logistic regression，简称 LR）和支持向量机（support vector machine，简称 SVM）4 种机器学习方法进行比较。各模型 ROC 曲线见图 3。

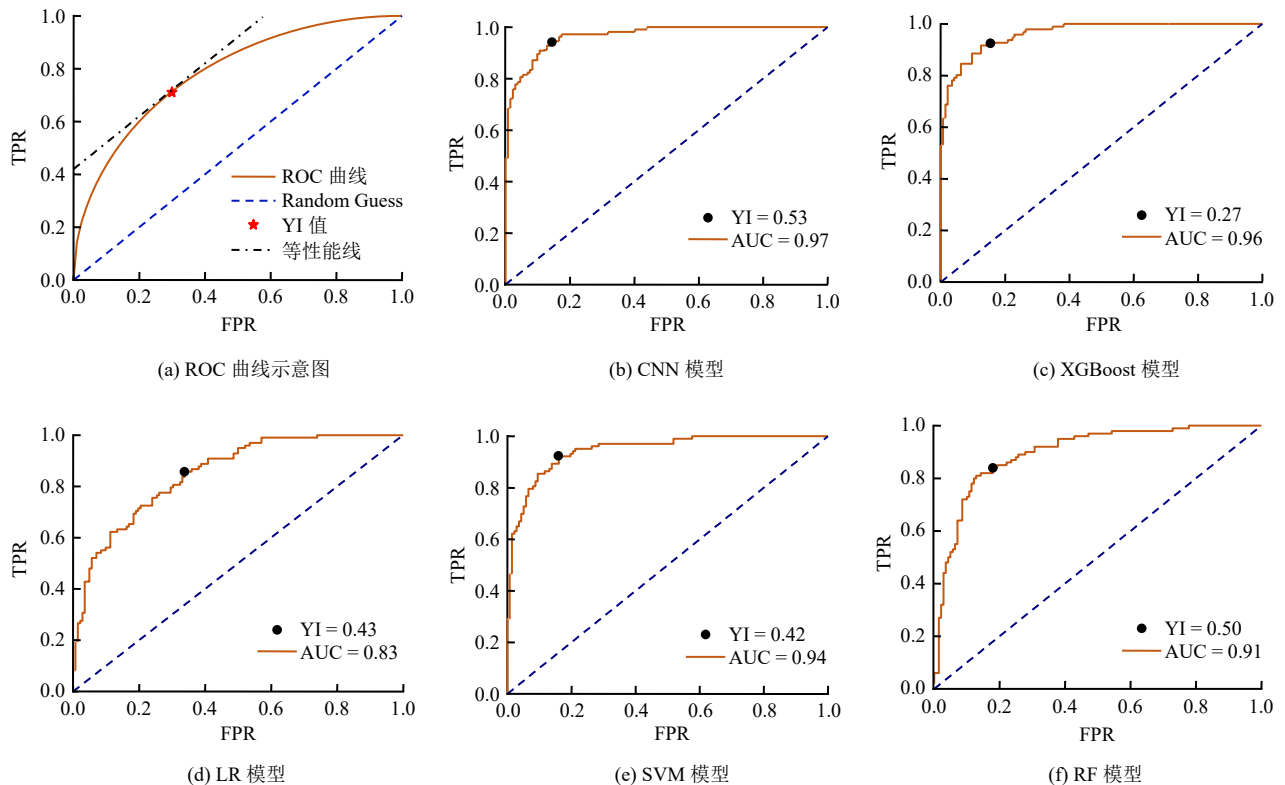


图3 ROC 曲线及其最佳阈值点

Fig.3 Receiver operating characteristic of models and its optimal threshold point

除 LR 模型外,其余 4 种模型的 AUC 值均在 0.9 以上, CNN 模型以及 XGBoost 模型的 AUC 值达到 0.95 以上。其中 CNN 模型与 RF 模型的最佳概率阈值 YI 值分别为 0.53 与 0.50,属于相对无偏模型。而 XGBoost 的 YI 值为 0.27,过于偏驳,液化评估结果明显趋于保守。

4.3 性能评估

为统一标准,均结合上采样算法进行训练且考虑最佳阈值对模型的影响,将 YI 值分别代入对应模型,测试集结果如表 3 所示。将 CNN 模型与 4 种机器学习模型进行对比, CNN 模型的各项指标均具明显优势。在准确率达到 92.58% 的情况下, F1 值为 0.93, AUC 值为 0.97,十分接近 1。表明在多场地,多震次的复杂数据中, CNN 模型拥有更加高效的识别能力,精度更高,且误判率更低。

表 3 模型判别结果

Table 3 Classification evaluation results of models

模型	准确率/%	精确率/%	召回率/%	F1 值	AUC
CNN	92.58	90.83	93.16	0.93	0.97
XGBoost	88.75	82.73	91.92	0.87	0.96
LR	72.92	62.69	84.85	0.72	0.83
SVM	86.25	80.02	88.89	0.84	0.94
RF	82.08	75.93	82.83	0.79	0.91

利用测试集数据对 CNN 模型与我国《建筑抗震设计规范》(GB50011—2010) [44] (以下简称建规法) 方法判别成功率进行对比,结果如表 4 所示。建规法对于液化的成功率为 83.65%,而非液化的成功率只有 42.64%,结果过于保守。而 CNN 模型预测结果更为均衡,液化和非液化成功率均在 89% 以上,准确率更高。

表 4 CNN 模型与我国建筑抗震设计规范判别结果对比

Table 4 Comparison of CNN model with Chinese code for building seismic design discrimination results

类别	准确率/%		样本数
	CNN	建规法	
液化	94.23	83.65	104
非液化	89.70	42.64	136

5 神经网络模型的可解释性分析

5.1 SHAP 值

深度学习模型通常在分类与拟合问题上具有优越的性能,但因其内部运行机制难以观察和解释,无法服务于工程问题。SHAP 是基于合作博弈论的 Shapley 值构建的可加性解释算法,为特征归因法的

一种。Shapley 所做的是量化每个个体对结果的贡献。SHAP 则是量化每个特征对模型输出的贡献值。

本文使用的是 SHAP 中的期望梯度算法 (expected gradients) [45],适用于任何具有梯度的模型。它是积分梯度算法 (integrated gradients) [46] 的拓展,旨在解释模型当前的预测与模型在给定基线输入时做出的预测之间的差异。同时通过积分的形式模拟特征值未知的情况,确定基线值。

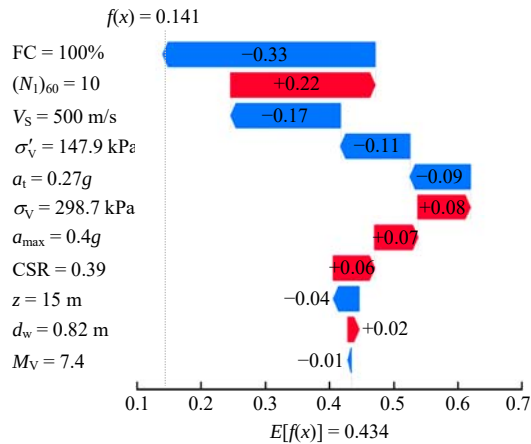
5.2 模型的局剖解释

单个样本的 SHAP 值瀑布图可以给出各因素对预测结果的具体贡献及各因素的影响程度排序,即样本预测的局部解释。图 4 为两个典型样本的局部解释。图中 $E[f(x)]$ 为模型基线值, $f(x)$ 为平衡点,均由模型实际输出转化得到 [46]; 液化影响因素以 SHAP 绝对值从大到小排列。平衡点是从基线值出发,在各影响因素的 SHAP 值的推动下得到的,平衡点大于基线值为液化样本,小于基线值为非液化样本。SHAP 值即为该特征值在模型液化预测中做出的贡献,其绝对值的大小由箭头的长度表示。蓝色与红色分别代表该特征值对模型预测的液化概率起负向作用和正向作用。

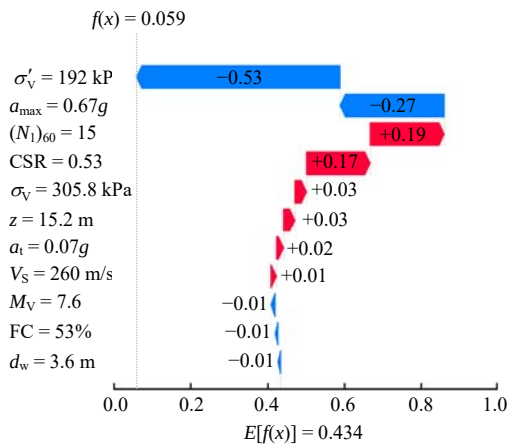
图 4(a)为模型成功预测的非液化样本;图 4(b)为误判样本,即实际为液化而模型预测为非液化。图 4(a)中对预测结果影响最大的 3 个因素为细粒含量、标贯击数和剪切波速。该样本细粒含量为 100%,不易液化,因而其 SHAP 值为负,对液化起主要负向作用;修正标贯击数为 10,相对较低,SHAP 值为正,对液化预测结果起正向作用;剪切波速为 500 m/s,远高于 300 m/s,基本可以认为土体是密实状态 [41],对土体液化的作用是负向的,SHAP 值为负;同时,有效上覆压力较大,不利于液化,门槛加速度为 0.27g,即产生较大应变所需的最小加速度较大,同样不利于液化,因此这两个特征的 SHAP 值为负。地表峰值加速度为 0.4g, CSR 为 0.39,均较大,利于液化,其 SHAP 值均为正。以上因素综合作用下,该样本平衡点为 0.141 小于基线值,判别为不液化,判别结果正确。

图 4(b)表示的误判样本,有效上覆压力与地表峰值加速度的 SHAP 绝对值最大。有效上覆压力的值为 192 kPa,为高上覆压力水平,土层不易液化,其 SHAP 值为负,对预测结果起负向作用。地表峰值加速度为 0.67g,同时循环剪应力比为 0.53,表明土层受到的地震动强度较大,应对液化预测结果起正向作用,而地表峰值加速度的 SHAP 值为负。因此,模型错估了地表峰值加速度对预测的影响,从

而导致误判。



(a) 正确预测样本



(b) 误判样本

图4 典型样本的 SHAP 局部解释

Fig.4 Local explanations for typical samples using SHAP

5.3 模型的全局解释

除了局部解释外, SHAP 算法还可以通过计算所有样本的 SHAP 值的平均绝对值, 给出模型的特征密度散点图, 即模型的全局解释。如图 5 所示, 具体的特征值用点表示, 点的颜色由蓝到红代表特征值由小到大。横坐标为特征值对应 SHAP 值, 即影响的权重。垂直方向上, 重叠区域显示了点的 SHAP 值沿纵轴的分布密度, 并且将特征从高到低依重要性排序。

从图 5 中可以看出, 标贯击数、细粒含量、循环剪应力比、地表峰值加速度对液化概率的预测影响程度较大, 基本与现有认知一致。震级的影响程度最小, 考虑为震级种类较为单调, 导致震级对模型预测影响较小。

细粒含量越高, 意味着土中的粉粒或黏粒含量越高, 在同等动荷载作用下, 土体的孔隙水压力增

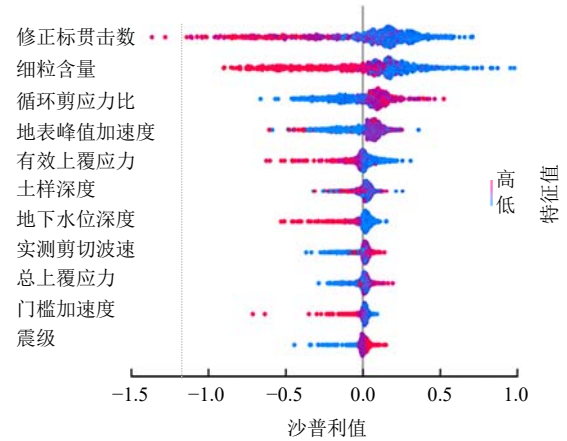


图5 特征散点图

Fig.5 SHAP summary plot

长越慢^[47], 相应的 SHAP 值为负, 液化发生的可能性就越低; 地下水位值较低时, 浅层土层的含水率较高, 地震作用下液化发生的概率也较高, SHAP 值为正; 门槛加速度值越大, 意味着土层产生临界剪应变所需外力越大, 土层不易变形, 液化敏感性也就随之降低^[42], 相应地, 其 SHAP 值越小; 从土层埋深的散点图中可以看到, 在纵轴的两侧, 蓝色点与红色点交错分布。而剪切波速对液化的影响也需结合土质进行考虑。因此, 二者的发展趋势需结合特征依赖图在之后进行解释。

5.4 单因素 SHAP 散点图分析

通过特征依赖性散点图分析各样本特征值对预测的影响规律, 如图 6 所示, 横坐标为样本实际值, 纵坐标为各样本对应的 SHAP 值。因篇幅有限, 本文仅展示特征重要性排序前 6 位特征。图 6(a)为修正标贯击数与 SHAP 值关系。在修正标贯击数小于等于 15 时, 样本的 SHAP 值基本处于正值, 对液化预测起正向作用; 当修正标贯击数值大于等于 25 时, SHAP 值均为负数, 且 SHAP 的绝对值有随着修正标贯击数的增大而增大的趋势, 表明修正标贯击数大于 25 时, 土层不易液化; 图 6(b)整体呈现出 SHAP 值随细粒含量增大而减小的关系, 在细粒含量在 80%~100%区间内快速下降, 即对预测结果的负向作用快速增强。在 40%~80%区间 SHAP 值趋势平缓, SHAP 值接近 0, 对液化发生概率影响较小。细粒含量小于 40%对应的 SHAP 值均为正值, 更易液化。

图 6(c)中 CSR 小于等于 0.25 对应的 SHAP 值基本为负, 不易液化; 而 CSR 大于 0.25 时, SHAP 值为正, 较易产生液化。整体上 SHAP 值随着 CSR 的增大而单调递增, 表明 CSR 越大, 模型预测的液

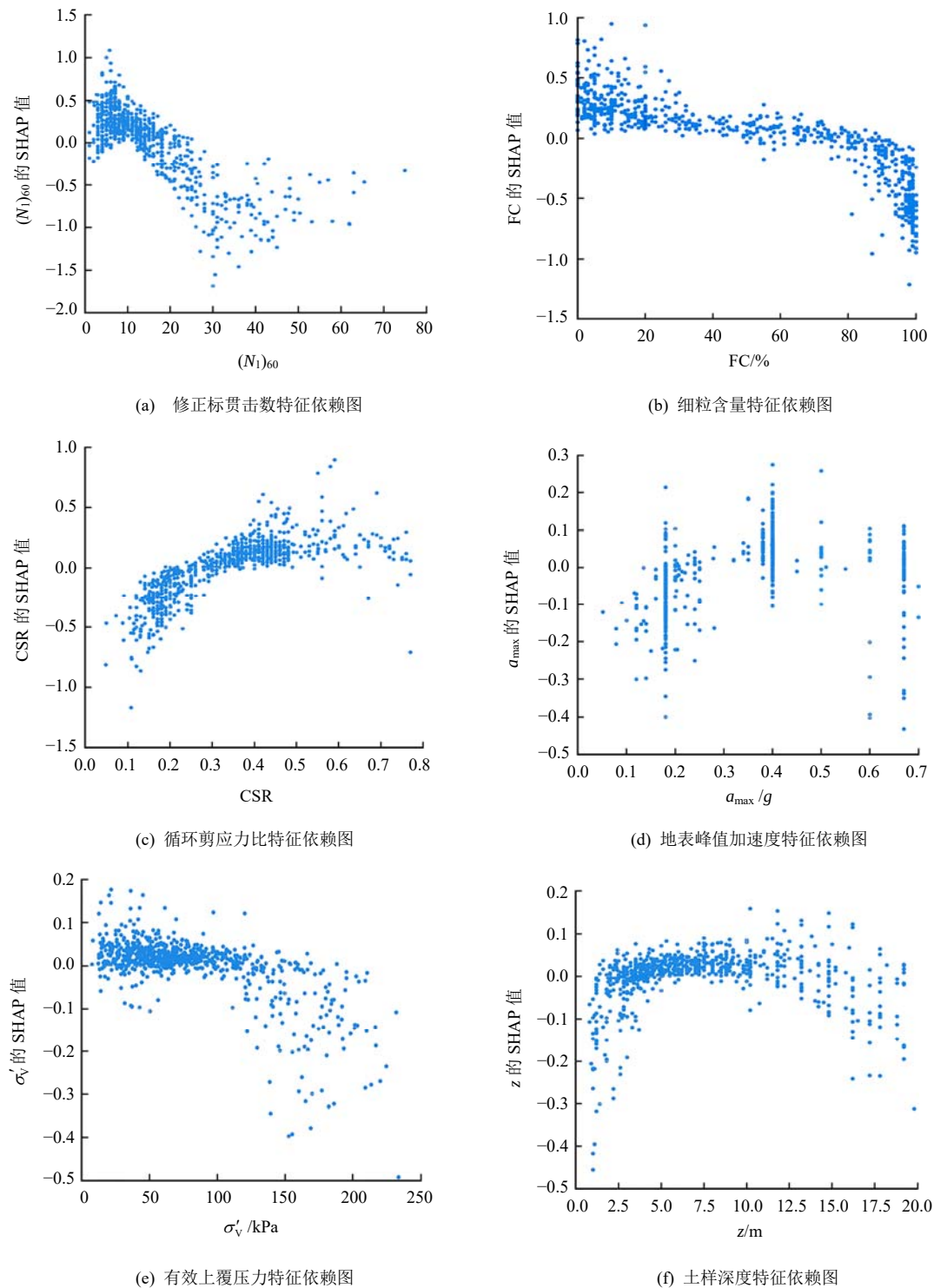


图6 SHAP 依赖性散点图

Fig.6 SHAP dependence scatter plots

化发生概率越大。图 6(d)中, 地表峰值加速度的 SHAP 值在 $0.10g \sim 0.40g$ 区间内逐步升高, 在 $0.25g$ 附近升为正值, 对液化预测起正向作用。在 $0.40g \sim 0.70g$ 区间 SHAP 值则较为分散, 但大多数为正值, 少数为负值。

图 6(e)中, SHAP 值的绝对值基本小于 0.2, 当有效上覆压力小于等于 120 kPa 时, 整体呈随有效

上覆压力的增大, 逐渐向 0 靠拢。当有效上覆压力大于 120 kPa 时, SHAP 值降至负值, 液化发生的可能性降低。图 6(f)中, 埋深的 SHAP 值总体趋势类似抛物线形式, 在 5 m 内 SHAP 值从 -0.5 快速增长至 0.1, 10 m 之后大部分数据的 SHAP 值呈现出降低趋势。因此在 10 m 以内深度对液化预测的影响是随深度的增加由负向作用变为微弱的正向作

用,而后逐渐又变为负向作用。

5.5 交互性 SHAP 散点依赖图分析

液化各影响因素之间存在一定的相互作用关系,因此,可以通过对单因素依赖图中的点进行

色来研究其相互作用对液化预测结果的影响,如图 7 所示。图中,横轴的特征值由坐标表示,右侧纵轴的特征值由点的颜色表示、由蓝到红表示特征值逐渐变大,左侧纵轴表示 SHAP 值。

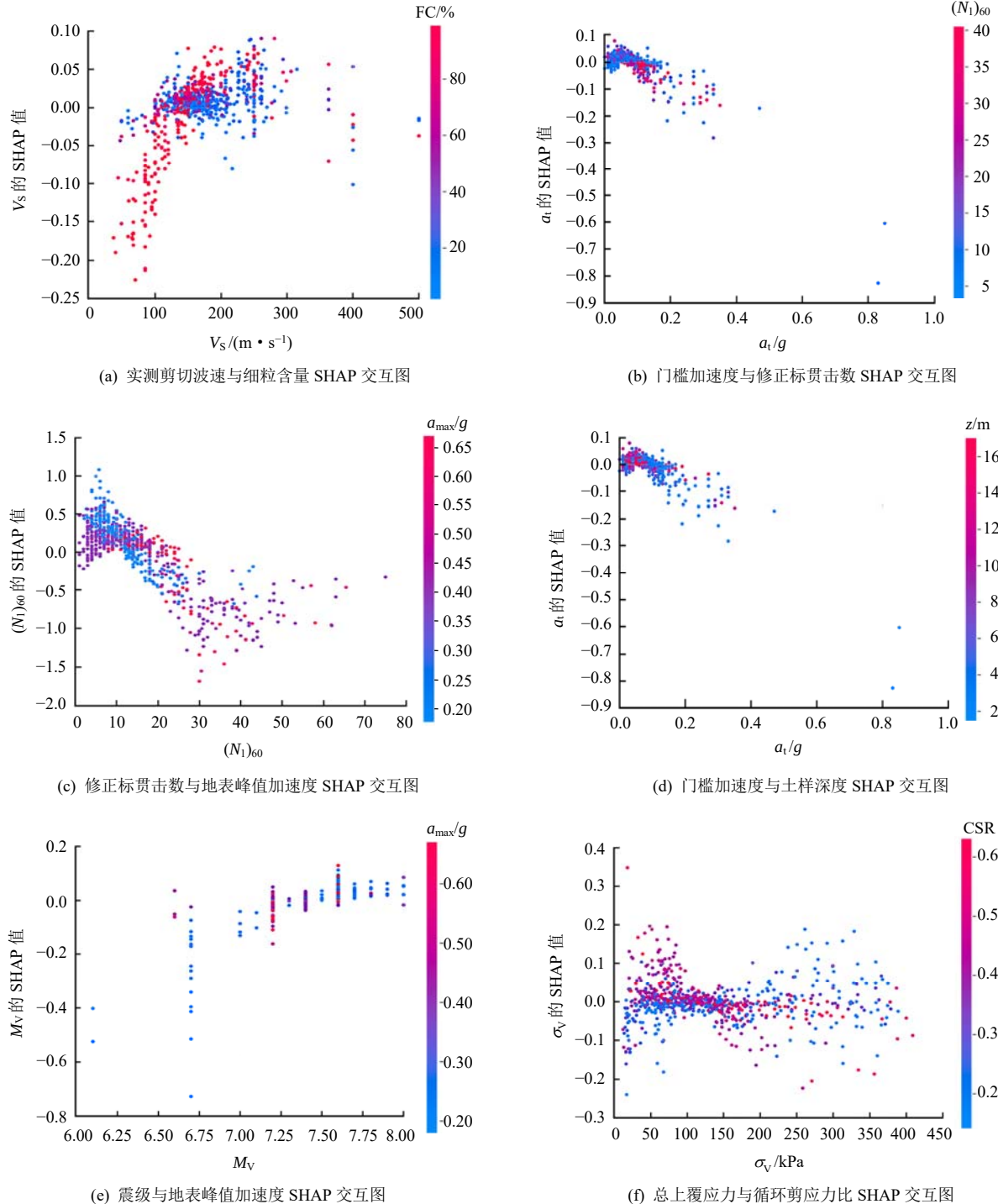


图 7 交互性 SHAP 散点依赖图
Fig.7 SHAP dependence scatter plots with interaction

从图 7(a)可以看出,剪切波速小于等于 150 m/s 时,细粒含量较高,SHAP 值为负值,对土体液化存在负向作用。剪切波速大于 150 m/s 时,细粒含量较小,液化发生概率较高,SHAP 值为正值。因

此,剪切波速小的样本对应的细粒含量高,是导致剪切波速小的样本 SHAP 值为负、不易液化的主要原因。在图 7(b)中,修正标贯击数值小的点集中在门槛加速度较小的区域,相应的 SHAP 值为正,即

较松散的土层其门槛加速度也小,更易液化。在门槛加速度等于 $0.1g$ 时, SHAP 值由正值变为负值,修正标贯击数也有相应增大的趋势。从图 7(c)中可以观察到, SHAP 值变化范围最大,对土体液化影响显著。图中主要有 3 种颜色的样本点、分别代表 3 种峰值加速度值,由蓝到红依次为较小值、中间值、较大值,反映了地震动强度的大小。对于 3 种类型的样本, SHAP 值均呈现随修正标贯击数增大而下降的趋势。当修正标贯击数约为 25 时, SHAP 值由正值转变为负值,对模型预测的液化发生概率的影响也由正向作用转变为负向作用。

图 7(d)反映的是门槛加速度与埋深的相互关系:埋深较浅时, SHAP 值随门槛加速度的增大而减小,土体液化的可能性降低;埋深较大时,样本的门槛加速度较小,集中于 $(0 \sim 0.2)g$, SHAP 值为正值,液化发生的可能性增大。

图 7(e)中,震级在 7.0 级及以下时,地表峰值加速度也较小, SHAP 值均为负值,液化发生概率较低;在震级为 7.5 级以上时, SHAP 值为正,对液化发生起主导作用。图 7(f)显示,在总上覆压力为 $0 \sim 130$ kPa、且循环剪应力比较大时, SHAP 值大于 0,模型预测的液化发生概率增加;循环剪应力比较小时, SHAP 值小于 0,模型预测的液化发生概率降低;当总上覆应力大于 130 kPa 时,与之相反。

6 结 论

本研究通过整合 Hanna2007 与 CEA2018 数据集,构建了一个包含多场地、多震次且数量较为可观的数据集。通过引入 MLP 层和卷积结构组成 CNN 模型,在模型架构中加入 BN 层简化模型调优的难度,采用 borderline SMOTE 上采样法补充训练集中的少数类数据来解决类不平衡问题,结合 SHAP 算法进行了可解释性分析。主要结论如下:

(1) CNN 模型预测准确率达 92.58%,与 XGBoost、LR、SVM 与 RF 这 4 种机器学习模型和建规法相比,预测性能十分优异。本文模型的最佳液化概率临界点为 0.53,为相对无偏模型,对液化数据与非液化数据均有良好的学习能力。同时,特征曲线下面积 AUC 值为 0.97、接近于 1,表明模型对液化样本和非液化样本的预测能力比较均衡。

(2) 利用 SHAP 算法对模型进行了局部和全局解释性分析,结果表明修正标贯击数、地表峰值加速度、循环剪应力比以及细粒含量是液化预测评估的重要影响因素,各因素对模型预测的影响趋势符

合现有认知,证明了本文模型的合理性和可靠性。

(3) 通过对单因素依赖图分析发现,修正标贯击数小于 15 的土体更易发生液化,修正标贯击数大于 25 的土体发生液化的可能性较低;就地震动强度而言,当循环剪应力比小于 0.25 时,土体发生液化的概率较低,循环剪应力比与液化发生概率的关系较为清晰。

参 考 文 献

- [1] 汪闻韶. 土的液化机理[J]. 水利学报, 1981(5): 22-34.
WANG Wen-sao. Mechanism of soil liquefaction[J]. *Journal of Hydraulic Engineering*, 1981(5): 22-34.
- [2] 王维铭, 陈龙伟, 郭婷婷, 等. 基于中国砂土液化数据库的标准贯入试验液化判别方法研究[J]. 岩土力学, 2023, 44(1): 279-288.
WANG Wei-ming, CHEN Long-wei, GUO Ting-ting, et al. Analysis of standard penetration test-based liquefaction evaluation methods using Chinese liquefaction database[J]. *Rock and Soil Mechanics*, 2023, 44(1): 279-288.
- [3] 张思宇, 李兆焱, 袁晓铭. 基于静力触探试验的液化判别新方法[J]. 岩土力学, 2022, 43(6): 1596-1606.
ZHANG Si-yu, LI Zhao-yan, YUAN Xiao-ming. A new method for evaluating liquefaction based on cone penetration test[J]. *Rock and Soil Mechanics*, 2022, 43(6): 1596-1606.
- [4] 杨洋, 孙锐. 基于剪切波速的液化可能性等级评估表法[J]. 岩土力学, 2023, 44(增刊 1): 634-644.
YANG Yang, SUN Rui. Liquefaction probability criteria table based on shear wave velocity[J]. *Rock and Soil Mechanics*, 2023, 44(Suppl.1): 634-644.
- [5] GOH A. Seismic liquefaction potential assessed by neural networks[J]. *Journal of Geotechnical and Geoenvironmental Engineering Division* 1995, 120(9): 1467-80.
- [6] 蔡煜东, 宫家文, 姚林声. 砂土液化预测的人工神经网络模型[J]. 岩土工程学报, 1993, 15(6): 53-58.
CAI Yu-dong, GONG Jia-wen, YAO Lin-sheng. Artificial neural network models for soil liquefaction prediction in sandy soils[J]. *Chinese Journal of Geotechnical Engineering*, 1993, 15(6): 53-58.
- [7] 刘红军, 薛新华. 砂土地震液化预测的人工神经网络模型[J]. 岩土力学, 2004, 25(12): 1942-1950.
LIU Hong-jun, XUE Xin-hu. Artificial neural network model for prediction of seismic liquefaction of sand soil[J]. *Rock and Soil Mechanics*, 2004, 25(12): 1942-1950.
- [8] 陈国兴, 李方明. 基于径向基函数神经网络模型的砂土液化概率判别方法[J]. 岩土工程学报, 2006, 28(3): 301-305.
CHEN Guo-xing, LI Fang-ming. Probabilistic estimation of sand liquefaction based on neural network model of

- radial basis function[J]. **Chinese Journal of Geotechnical Engineering**, 2006, 28(3): 301-305.
- [9] HANNA A M, URAL D, SAYGILI G. Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data[J]. **Soil Dynamics and Earthquake Engineering**, 2007, 27(6): 521-540.
- [10] 胡记磊, 唐小微, 裘江南. 基于贝叶斯网络的地震液化概率预测分析[J]. **岩土力学**, 2016, 37(6): 1745-1752.
HU Ji-lei, TANG Xiao-wei, QIU Jiang-nan. Prediction of probability of seismic-induced liquefaction based on Bayesian network[J]. **Rock and Soil Mechanics**, 2016, 37(6): 1745-1752.
- [11] SCHMIDT J, MOSS R. Bayesian hierarchical and measurement uncertainty model building for liquefaction triggering assessment[J]. **Computers and Geotechnics**, 2021, 132: 103963.
- [12] 刘年平, 王宏图, 袁志刚, 等. 砂土液化预测的 Fisher 判别模型及应用[J]. **岩土力学**, 2012, 33(2): 554-557.
LIU Nian-ping, WANG Hong-tu, YUAN Zhi-gang, et al. Fisher discriminant analysis model of sand liquefaction and its application[J]. **Rock and Soil Mechanics**, 2012, 33(2): 554-557.
- [13] 颜可珍, 刘能源, 夏唐代. 基于判别分析法的地震砂土液化预测研究[J]. **岩土力学**, 2009, 30(7): 2049-2052.
YAN Ke-zhen, LIU Neng-yuan, XIA Tang-dai. Discriminant analysis model for prediction of sand soil liquefaction during earthquake[J]. **Rock and Soil Mechanics**, 2009, 30(7): 2049-2052.
- [14] 刘勇健. 基于聚类-二叉树支持向量机的砂土液化预测模型[J]. **岩土力学**, 2008, 29(10): 2764-2768.
LIU Yong-jian. Support vector machine model for predicting sand liquefaction based on clustering binary tree algorithm[J]. **Rock and Soil Mechanics**, 2008, 29(10): 2764-2768.
- [15] 彭刘亚, 解惠婷, 冯伟栋. 基于随机森林算法的砂土液化预测方法[J]. **物探与化探**, 2020, 44(6): 1429-1434.
PENG Liu-ya, XIE Hui-ting, FENG Wei-dong. The method of predict sand liquefaction based on random forest algorithm[J]. **Geophysical and Geochemical Exploration**, 2020, 44(6): 1429-1434.
- [16] ZHAO Z, DUAN W, CAI G, et al. CPT-based fully probabilistic seismic liquefaction potential assessment to reduce uncertainty: integrating XGBoost algorithm with Bayesian theorem[J]. **Computers and Geotechnics**, 2022, 149: 104868.
- [17] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. **Science**, 2006, 313(5786): 504-507.
- [18] HUANG J T, LI J, GONG Y. An analysis of convolutional neural networks for speech recognition[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2015: 4989-4993.
- [19] 王自法, 廖吉安, 王延伟, 等. 基于深层卷积神经网络的震级快速估算方法[J]. **地球物理学报**, 2023, 66(1): 272-288.
WANG Zi-fa, LIAO Ji-an, WANG Yan-wei, et al. A fast magnitude estimation method based on deep convolutional neural networks[J]. **Chinese Journal of Geophysics**, 2023, 66(1): 272-288.
- [20] 骆勇鹏, 王林堃, 廖飞宇, 等. 基于一维卷积神经网络的结构损伤识别[J]. **地震工程与工程振动**, 2021, 41(4): 145-156.
LUO Yong-peng, WANG Lin-kun, LIAO Fei-yu, et al. Vibration-based structural damage identification by 1-dimensional convolutional neural network[J]. **Earthquake Engineering and Engineering Dynamics**, 2021, 41(4): 145-156.
- [21] YUAN X, TANKSLEY D, LI L, et al. Faster post-earthquake damage assessment based on 1D convolutional neural networks[J]. **Applied Sciences**, 2021, 11(21): 9844.
- [22] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. **Journal of Artificial Intelligence Research**, 2002, 16: 321-357.
- [23] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]//Conference and Workshop on Neural Information Processing Systems. Long Beach: NIPS Press, 2017: 4765-4774.
- [24] RIBEIRO M T, SINGH S, GUESTRIN C. "Why should I trust you?": explaining the predictions of any classifier[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA: Association for Computing Machinery, 2016: 1135-1144.
- [25] 齐婉婉, 孙锐, 郑桐, 等. 融合 XGBoost 和 SHAP 的地表峰值加速度预测分析模型[J]. **岩土工程学报**, 2023, 45(9): 1934-1943.
QI Wan-wan, SUN Rui, ZHENG Tong, et al. Prediction and analysis model of ground peak acceleration based on XGBoost and SHAP[J]. **Chinese Journal of Geotechnical Engineering**, 2023, 45(9): 1934-1943.
- [26] HU J, DING Y, ZHANG H, et al. A real - time seismic intensity prediction framework based on interpretable ensemble learning[J]. **Seismological Research Letters**, 2023, 94(3): 1579-1602.
- [27] WANG X, WANG Z, WANG J, et al. Machine learning based ground motion site amplification prediction[J]. **Frontiers in Earth Science**, 2023, 11: 1053085.

- [28] CETIN K O, SEED R B, KAYEN R E, et al. SPT-based probabilistic and deterministic assessment of seismic soil liquefaction triggering hazard[J]. **Soil Dynamics and Earthquake Engineering**, 2018, 115: 698-709.
- [29] YAZDI J S, KALANTARY F, YAZDI H S. Investigation on the effect of data imbalance on prediction of liquefaction[J]. **International Journal of Geomechanics**, 2013, 13(4): 463-466.
- [30] 张敏政. 土耳其伊兹米特地震震害考察[J]. 工程抗震与加固, 2000(4): 34-38.
ZHANG Min-zheng. Damage assessment of the Izmit Earthquake in Turkey[J]. **Earthquake Resistant Engineering and Retrofitting**, 2000(4): 34-38.
- [31] 叶耀先. 台湾 1999 年 9·21 集集大地震简况[J]. 建筑知识, 1999(6): 16.
YE Yao-xian. Overview of the Taiwan 9·21 Jiji earthquake in 1999[J]. **Architectural Knowledge**, 1999(6): 16.
- [32] CETIN K O, SEED R B, DER KIUREGHIAN A, et al. Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential[J]. **Journal of Geotechnical and Geoenvironmental Engineering**, 2004, 130(12): 1314-1340.
- [33] ANDRUS R D, STOKOE II K H. Liquefaction resistance of soils from shear-wave velocity[J]. **Journal of Geotechnical and Geoenvironmental Engineering**, 2000, 126(11): 1015-1025.
- [34] HARDIN BOBBY O, DRNEVICH VINCENT P. Shear modulus and damping in soils: measurement and parameter effects[J]. **Journal of the Soil Mechanics and Foundations Division**, 1972, 98(6): 603-624.
- [35] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
ZHOU Fei-yan, JIN Lin-peng, DONG Jun. Review of convolutional neural network[J]. **Chinese Journal of Computer**, 2017, 40(6): 1229-1251.
- [36] SANTURKAR S, TSIPRAS D, ILYAS A, et al. How does batch normalization help optimization?[C]//32nd Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018.
- [37] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[M]//Advances in Intelligent Computing. Berlin: Springer, 2005: 878-887.
- [38] SEED B H, TOKIMATSU K, HARDER L F, et al. Influence of SPT procedures in soil liquefaction resistance evaluations[J]. **Journal of Geotechnical Engineering**, 1985, 111(12): 1425-1445.
- [39] 李方圆, 董林, 夏坤, 等. 细粒含量对砂土液化势影响探讨[J]. 地震工程与工程振动, 2022, 42(2): 244-251.
- LI Fang-yuan, DONG Lin, XIA Kun, et al. Discussion on the effects of fines content on liquefaction potential of sandy soils[J]. **Earthquake Engineering and Engineering Dynamics**, 2022, 42(2): 244-251.
- [40] 付海清, 伊纪祥, 冯陆伟, 等. 考虑上覆压力影响的现场液化试验研究[J]. 岩土工程学报, 2024, 46(3): 558-566.
FU Hai-qing, YI Ji-xiang, FENG Lu-wei, et al. Study on the in-situ liquefaction test considering the effect of overburden pressure[J]. **Chinese Journal of Geotechnical Engineering**, 2024, 46(3): 558-566.
- [41] 邱志刚, 薄景山, 罗奇峰. 土壤剪切波速与标贯击数关系的统计分析[J]. 自然灾害学报, 2012, 21(2): 102-107.
QIU Zhi-gang, BO Jing-shan, LUO Qi-feng. Statistical analysis of relationship between shear wave velocity and standard penetration test blow count[J]. **Journal of Nature Disaster**, 2012, 21(2): 102-107.
- [42] DOBRY R, LADD R S, YOKEL F Y, et al. Prediction of pore water pressure buildup and liquefaction of sands during earthquakes by the cyclic strain method: NBS BSS 138[R]. Gaithersburg: National Bureau of Standards, 1982.
- [43] YAZDI J, MOSS R. Nonparametric liquefaction triggering and postliquefaction deformations[J]. **Journal of Geotechnical and Geoenvironmental Engineering**, 2012, 21(2): 102-107.
- [44] 中国建筑科学研究院. GB 50011—2010 建筑抗震设计规范[S]. 北京: 中国建筑工业出版社, 2010.
China Academy of Building Research. GB 50011—2010 Code for seismic design of buildings[S]. Beijing: China Architecture & Building Press, 2010.
- [45] ERION G, JANIZEK J D, STURMFELS P, et al. Improving performance of deep learning models with axiomatic attribution priors and expected gradients[J]. **Nature Machine Intelligence**, 2021, 3(12): 620-631.
- [46] SUNDARARAJAN M, TALY A, YAN Q. Axiomatic attribution for deep networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney: [s. n.], 2017: 3319-3328.
- [47] 赵学亮, 陈馨睿, 丁鹏程, 等. 动荷载下不同细粒含量粉土孔压发展模式研究[J]. 武汉理工大学学报(交通科学与工程版), 2022, 46(2): 264-269.
ZHAO Xue-liang, CHEN Xin-rui, DING Peng-cheng. Study on pore pressure development model of different fine-grained content silt under dynamic loading[J]. **Journal of Wuhan University of Technology (Transportation Science & Engineering)**, 2022, 46(2): 264-269.