

DOI: 10.16410/j.issn1000-8365.2024.3179

基于数据合成与机器学习的 6DM 气缸体 复杂铸件缺陷预测

王传胜¹, 冯相灿¹, 潘徐政², 高峰¹, 刘冰¹, 李岩¹, 韩宇¹,

钟东彦¹, 付煜¹, 计效园², 周建新²

(1. 一汽铸造有限公司, 吉林 长春 130062; 2. 华中科技大学材料成形与模具技术国家重点实验室, 湖北 武汉 430074)

摘要: 在汽车核心零部件制造等关键领域, 复杂铸件出现缺陷的后果尤为严重, 因此对复杂铸件进行缺陷预测并提高其生产质量刻不容缓。本文针对实际铸造过程中采集到的 6DM 气缸体复杂铸件生产数据中气孔、砂眼等缺陷类别的数据量严重不平衡问题, 对基于数据合成与机器学习的 6DM 气缸体复杂铸件缺陷预测进行研究, 梳理了人工神经网络与复杂铸件缺陷预测的研究现状, 结合企业现场生产情况, 开展了需求分析, 获取 6DM 气缸体复杂铸件生产数据。并基于 SMOTE(synthetic minority oversampling technique)算法, 创建了合成数据集, 采用合成数据集作为训练模型的数据集, 预测准确率达到 99.37%。结果表明, 构建的复杂铸件缺陷预测模型能够准确预测复杂铸件缺陷。

关键词: 6DM 气缸体; 缺陷预测; 不平衡数据; 数据合成; SMOTE 算法

中图分类号: TG245; TP181

文献标识码: A

文章编号: 1000-8365(2024)02-0173-06

Defect Prediction of 6DM Cylinder Block Complex Castings Based on Data Synthesis and Machine Learning

WANG Chuansheng¹, FENG Xiangcan¹, PAN Xuzheng², GAO Feng¹, LIU Bing¹, LI Yan¹, HAN Yu¹,
ZHONG Dongyan¹, FU Yu¹, JI Xiaoyuan², ZHOU Jianxin²

(1. FAW Foundry Co., Ltd, Changchun 130062, China; 2. State Key Laboratory of Materials Processing and Die & Mould Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The problems caused by defects in complex castings are particularly serious in automotive core part manufacturing and other key areas, which makes it urgent to predict the defects of complex castings and improve their production quality. In this paper, aiming at the problem of serious imbalance in the production data of complex 6DM cylinder block castings, such as those of pores and sand holes collected during the actual casting process, the defect prediction of complex 6DM cylinder block castings based on data synthesis and machine learning was studied, and the research status of artificial neural networks and complex casting defect prediction was combed. Combined with the on-site production situation of enterprises, demand analysis was carried out, and the production data of 6DM cylinder block complex castings were obtained. The synthetic dataset created based on the synthetic minority oversampling technique (SMOTE) algorithm was adopted as the dataset of the training model, which achieved a prediction accuracy of 99.37%. The results show that the constructed defect prediction model can accurately predict the defects in complex castings.

Key words: 6DM cylinder block; defect prediction; unbalanced data; data synthesis; SMOTE algorithm

各类复杂铸件广泛应用于航空、航天、轨道交通、工程机械等国民经济多个领域, 其中, 以机匣、发动机缸体、转向桥等为代表的典型复杂铸件, 生产中受各种因素影响, 铸件中产生气孔、砂眼、冷隔、

缩孔等缺陷, 从而导致品质铸件产量低。因此, 对铸件进行缺陷预测, 找出关键原因、制定解决方法, 有利于降低铸件产生缺陷的概率, 提高生产效率和经济效益。目前, 铸造企业通常采用 X 射线探伤^[1]、超

收稿日期: 2023-07-03

基金项目: 国家重点研发计划(2020YFB1710100); 国家自然科学基金(51905188, 52090042, 51775205)

作者简介: 王传胜, 1970 年生, 正高级工程师。主要从事铸造企业数字化规划、系统开发工作。Email: wcs_fc@faw.com.cn

引用格式: 王传胜, 冯相灿, 潘徐政, 等. 基于数据合成与机器学习的 6DM 气缸体复杂铸件缺陷预测[J]. 铸造技术, 2024, 45(2): 173-178.

WANG C S, FENG X C, PAN X Z, et al. Defect prediction of 6DM cylinder block complex castings based on data synthesis and machine learning[J]. Foundry Technology, 2024, 45(2): 173-178.

声波探伤^[2]、机器视觉^[3]等技术对生产出的复杂铸件进行缺陷检测,但是对于铸造前的复杂铸件缺陷预测并没有较为成熟的技术支撑,所以实现复杂铸件的缺陷预测具有重要意义。

人工神经网络(artificial neural networks, ANN)是一种非线性自适应信息处理算法,基于大量生产数据(如工艺参数等),通过构建神经网络进行机器学习,能够快速、精确地完成复杂铸件缺陷预测,效率较高。张井柱等^[4]为了提高铝合金铸件的质量和生产效率,提出了一种基于均值标准蜂窝算法的参数化智能优化方法,建立了工艺参数优化模型。Lu等^[5]为了设计一种能够快速准确地预测无定形成形能力的方法,构建了人工神经网络和卷积神经网络两种模型,预测准确率分别达到 77.623% 和 71.693%,比 13 个准则的预测准确率高出 19% 以上。张建明等^[6]为了快速、精确预测铸件缩孔和缩松缺陷,引入机器学习,采用特定预处理方法处理铸造过程数据,多模态数据输入方法提高数据维数,全连通卷积神经网络快速计算铸件缩孔和缩松缺陷。大连理工大学陈相吉^[7]采用 YOLOv3 作为缺陷检测的基础算法,提出了一种快速准确的增压器铸造缺陷检测算法。燕山大学栗子豪^[8]针对低压铸造生产过程中的时延性问题,提出了多变量时延长短时记忆网络模温预测模型和多变量门控循环单元网络模温预测模型,并用烟花算法分别对预测模型参数进行优化,实现了低压铸造生产过程模温的精准预测和模型参数的自动寻优。Wang 等^[9]为了利用量子计算的并行计算来解决神经网络中计算的难点,基于脉冲耦合神经网络(PCNN),提出了量子脉冲耦合神经网络(QPCNN)。Kumar^[10]为了识别非线性动力系统,提出了一种新型的高阶上下文分层递归 pi-sigma 神经网络。Solís-Pérez 等^[11]提出了一种基于双曲切线和哈利勒适宜指数函数的新型传递函数。Shi 等^[12]首先综述了近年来求解旅行推销员问题的神经网络方法,包括 Hopfield 神经网络、图神经网络和具有强化学习的神经网络。喻会等^[13]针对 3 种合金

的磨损量问题,构建了以浇注温度、补料顺序和合金成分为输入,耐磨性为输出的神经网络优化模型。Shen 等^[14]提出了可作为神经网络中 tanh 替代方案的新型激活函数 tanhLU。

综上所述,人工神经网络在铸造领域有很好的应用前景。本文选用人工神经网络方法对铸件进行缺陷预测,找出导致铸件出现缺陷的关键原因,制定科学的解决方法,进而指导生产过程,有利于降低铸件产生缺陷概率,提高生产效率和经济效益。

1 铸件缺陷与企业需求分析

1.1 6DM 气缸体铸件介绍

6DM 气缸体是某铸造公司的重点产品,铸件净重 274 kg,主要壁厚 5 mm,是某柴油机厂奥威 CA6DM 系列发动机缸体毛坯铸件,是为解放重卡提供动力的 11-13L 大马力柴油机,直列 6 缸水冷湿式缸套发动机,年纲领 10 万套左右,满足国六标准,奥威 CA6DM 发动机搭载在一汽解放、青岛解放等的主流、高端重卡车型上,6DM 气缸体铸件三维图如图 1 所示。

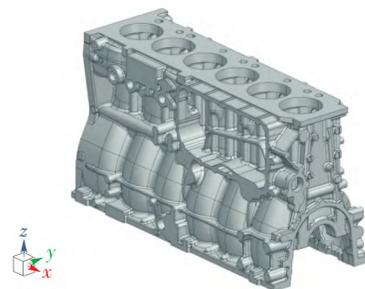


图 1 6DM 气缸体铸件三维图
Fig.1 3D diagram of 6DM cylinder block casting

1.2 企业生产现场铸件缺陷情况

气缸体是引导活塞在缸内进行直线往复运动的圆筒形金属机件,属于关键器件,铸造过程中出现的缺陷可能会引发严重后果。因此,对于 6DM 气缸体铸件进行缺陷预测刻不容缓。企业生产现场铸件缺陷情况如图 2 所示。

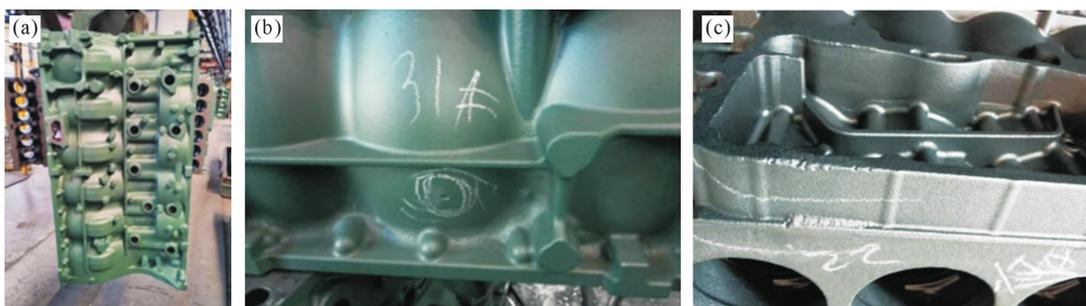


图 2 生产现场铸件缺陷情况:(a)合格铸件,(b)铸件气孔缺陷,(c)铸件砂眼缺陷
Fig.2 Defects in castings at the production site: (a) qualified castings, (b) porosity defect in castings, (c) blister defects in castings

1.3 企业需求分析

调研获取的6DM气缸体复杂铸件生产数据,包括无缺陷、气孔缺陷、砂眼缺陷3类数据。根据企业需求,最终建立的神经网络缺陷预测模型应满足以下目标:①按照缺陷种类进行预测,而非笼统进行有无缺陷预测;②需要足够的预测准确率,预计准确率95%以上。

2 数据采集、数据分析与数据合成

2.1 数据采集

经过人工及机器采集,得到6DM气缸体复杂铸件的生产和质量数据,采集到的6DM气缸体部分原始数据如表1所示。其中包括100条气孔缺陷数据,272条砂眼缺陷数据,1504条无缺陷数据。样本属性数量共22个,作为复杂铸件缺陷预测模型的输入;样本标签数量共3个,作为复杂铸件缺陷预测模型的输出。样本标签用独热编码的形式表示,“100”代表此样本对应铸件无缺陷,“010”代表此样本对应铸件有气孔缺陷,“001”代表此样本对应铸件有砂眼缺陷。经过简单的数据预处理,将6DM气缸体复杂铸件的生产数据展示如表2~3所示。表2中数据代表各成分的质量分数。

2.2 数据分析

大多数分类学习方法都有一个共同的基本假设,即不同类别的训练样本数量相等。如果不同类型的训练样本数量有细微的差异,通常影响不大。但是如果差异过大,会给学习过程带来麻烦。例如,如果反例有900个,而正例只有100个,则该学习方法只需要返回一个总是预测新样本为反例的学习器,准确率就可以达到90%,但是以此训练出的学习器是没有价值的,因为它不能预测任何正例。

针对经过简单预处理得到的数据集,有100条气孔缺陷数据,272条砂眼缺陷数据,1504条无缺陷数据,仅有约1000条有缺陷数据;若把缺陷分类,则每种缺陷拥有的数据会更少。这是典型的不平衡数据,利用不平衡数据训练出的预测模型无法作出准确预测。

2.3 数据合成

针对本文研究的数据集,如果使用欠采样随机丢弃样本,可能会丢失一些重要信息。因此,选择过采样方法增加各类缺陷数据,使其数据的数量与无缺陷数据接近。对数据集中的少数类样本进行过采样,也就是将少数类样本相加,使分类问题中每个类的样本数量接近。但是,过采样不能简单对少数类样

表1 6DM气缸体部分原始数据
Tab.1 Raw data of the 6DM cylinder block

缺陷类型	缺陷部位	浇注日期	w(C) /%	w(Si) /%	w(Mn) /%	w(S) /%	w(P) /%	w(Cr) /%	w(Cu) /%	w(Mo) /%	w(Sn) /%	w(V) /%	w(Zn) /%	w(Al) /%	w(Pb) /%	浇注温度 /°C	砂芯终强度 /MPa	砂芯总重 /kg	树脂加入量 /%
砂芯疏松	5-C1	2021/4/8	3.15	1.94	0.64	0.08	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0	1428	1.27	270	1.30
57/砂芯砂眼	6-C1	2021/3/19	3.15	1.94	0.64	0.08	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0	1426	1.33	270	1.30
		2021/3/19	3.15	1.89	0.60	0.11	0.01	0.18	0.66	0.01	0.07	0.06	0.05	0	0.01	1436	1.22	270	1.30
		2021/3/19	3.15	1.87	0.72	0.09	0.01	0.18	0.66	0.01	0.08	0.06	0.07	0	0.01	1429	1.34	270	1.30
57/砂芯砂眼	6-C1	2021/3/19	3.12	1.91	0.69	0.09	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0.01	1426	1.25	270	1.30
58/砂芯砂眼	6-C1	2021/3/19	3.12	1.91	0.69	0.09	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0.01	1426	1.25	270	1.30
59/砂芯砂眼	6-C1	2021/3/19	3.12	1.91	0.69	0.09	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0.01	1426	1.25	270	1.30
60/砂芯砂眼	6-C1	2021/3/19	3.12	1.91	0.69	0.09	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0.01	1426	1.25	270	1.30
61/砂芯砂眼	6-C1	2021/3/19	3.12	1.91	0.69	0.09	0.01	0.18	0.65	0.01	0.07	0.06	0.06	0	0.01	1426	1.25	270	1.30

注:C元素含量由碳硅热分析得到

表2 6DM气缸体复杂铸件的成分
Tab.2 Chemical composition of the complex castings of the 6DM cylinder block

元素	C	Si	Mn	S	P	Cr	Cu	Mo	Sn	V	Zn	Al	Pb	(mass fraction/%)	
含量	3.15	1.89	0.60	0.11	0.01	0.18	0.66	0.01	0.07	0.06	0.05	0.004	0.01		

表3 6DM气缸体复杂铸件的生产数据与质量数据
Tab.3 Production and quality data of complex castings of the 6DM cylinder block

生产数据										质量数据(独热编码)		
浇注温度 /°C	砂芯强度 /MPa	烘干后水分 /%	型砂紧实率 /%	型砂剪切强度 /MPa	型砂水分 /%	造型上型压强 /MPa	造型下型压强 /MPa	砂型硬度 (HC)	无缺陷	气孔	砂眼	
1436	1.22	0.13	36	2.8	2.7	9.4	9.4	61	100	010	001	

本进行重复采样(即复制样本),否则会导致过拟合。过采样的代表性算法是 SMOTE 算法,其流程如图 3 所示。SMOTE 算法核心思想如图 4 所示。

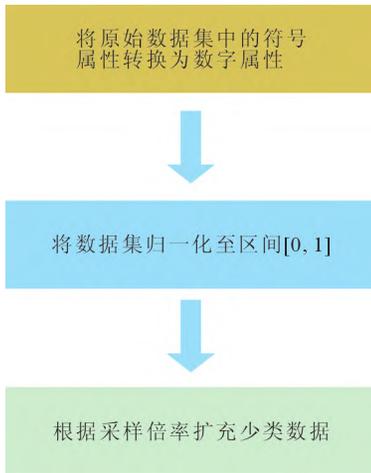


图 3 SMOTE 算法流程
Fig.3 SMOTE flow chart

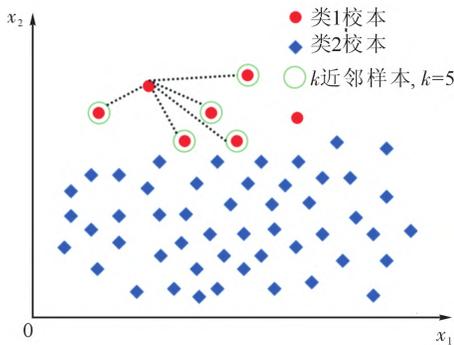


图 4 SMOTE 算法核心思想
Fig.4 Diagram of the SMOTE core idea

SMOTE 算法的生成过程为:①对于少数类中的每个样本 x_i ,以欧氏距离为标准,计算其到少数类样本集中所有样本的距离,得到其 k 个最近邻;②根据样本不平衡比设定一个抽样比 n 。对于每个少数类样本 x_i ,从其 k 个邻居中随机选取几个样本,假设所选邻居为 x_{zi} ;③对于每个随机选出的近邻 x_{zi} ,分别与原样本按照式(1)构建新的样本:

$$x_n = x_i + \text{rand}(0, 1) \times (x_{zi} - x_i) \quad (1)$$

经过简单预处理的数据集,包含无缺陷数据 1 504 条,气孔缺陷数据 100 条,砂眼缺陷数据 272 条,属于不平衡数据。经过 SMOTE 算法进行数据合成,得到总样本数据量 4 636 条,每种类别,即无缺陷、气孔缺陷、砂眼缺陷的数据量各 1 500 条左右。至此,解决了数据不平衡的问题。数据合成前后数据量对比如表 4 所示。

3 企业需求分析与模型效果验证

3.1 模型构建

神经网络缺陷预测模型使用 python 语言进行

表 4 数据合成前后数据量对比

Tab.4 Comparison of data volume before and after data synthesis

	无缺陷数据	气孔缺陷	砂眼缺陷
数据合成前样本数量	1 504	100	272
数据合成后样本数量	1 504	1 501	1 631

构建,以下是具体的模型构建步骤:

(1)导入需要的第三方 python 库,包括开放源代码软件库 Tensorflow、科学计算库 Numpy、数据可视化库 Matplotlib、读取数据文件库 Xlrd。

(2)加载数据,利用读取数据文件库 Xlrd 读取 Excel 表中的数据。

(3)将数据导入 Numpy 数组,创建全 0 的 Numpy 数组 Input 和 Output,分别将输入数据和输出数据导入数组 Input 和 Output。

(4)将数据归一化处理,因为数据拥有多个特征属性,各个属性之间数量级差异很大,如果不作归一化处理,直接用原数据集作为数据集,会对训练出的模型的性能造成影响。

(5)将归一化处理后的数据导入数组 Data。

(6)随机打乱 Data 中的数据,以 2:1 的比例划分训练集和测试集。

(7)设置学习率、迭代次数和显示间隔,设置模型参数初始值。超参数学习率以 0.01~0.001 为宜,本文选取学习率 0.01,迭代次数分别为 1 000、1 100、1 200、1 300 4 组,显示间隔统一为每训练 100 次显示输出训练集和测试集的准确率和交叉熵损失。

(8)训练模型:定义 4 个空列表,分别用来记录训练集和测试集的准确率和交叉熵损失;定义梯度带,计算得到训练集和测试集的输出和交叉熵损失,并将结果追加到空列表中;迭代学习模型参数,更新模型参数。

(9)将训练结果可视化。

3.2 模型效果验证

利用经过 SMOTE 算法合成的 6DM 气缸体数据集,对复杂铸件缺陷预测模型进行训练,并更改模型参数,得到 4 组在不同训练次数下的模型预测准确率与损失值的输出,损失值和准确率随迭代次数的详细数值如表 5 所示。

表 5 损失值和准确率随迭代次数的详细数值

Tab.5 Detailed values of the loss value and accuracy with respect to the number of iterations

迭代次数	训练集准确率	训练集损失值	测试集准确率	测试集损失值
1 000	99.28%	3.42%	99.35%	3.71%
1 100	99.45%	2.97%	98.90%	4.28%
1 200	99.32%	3.18%	99.61%	2.97%
1 300	99.42%	3.12%	99.16%	3.07%

训练次数为 1 000 时,训练集预测准确率为 99.28%,测试集预测准确率为 99.35%;训练次数为 1 100 时,训练集预测准确率为 99.45%,测试集预测准确率为 98.90%;训练次数为 1 200 时,训练集预测准确率为 99.32%,测试集预测准确率为 99.61%;训练次数为 1 300 时,训练集预测准确率为 99.42%,测试集预测准确率为 99.16%。对模型预测结果进行分析,训练集与测试集的预测准确率均大于 98%,达到企业预期预测准确率 95%;损失值均小于

5%,模型未出现过拟合问题。结果可视化展现如图 5 所示。

在迭代过程中,训练集和测试集的交叉熵损失随迭代次数的变化曲线均严格递减,说明基于数据合成的缺陷预测模型预测出的结果与数据的实际分布情况吻合,性能优良。准确率随迭代次数的变化曲线在迭代过程中,训练集和测试集的准确率均严格递增,且曲线光滑,未出现异常点,说明建立的缺陷预测模型能够进行准确预测,性能优良。

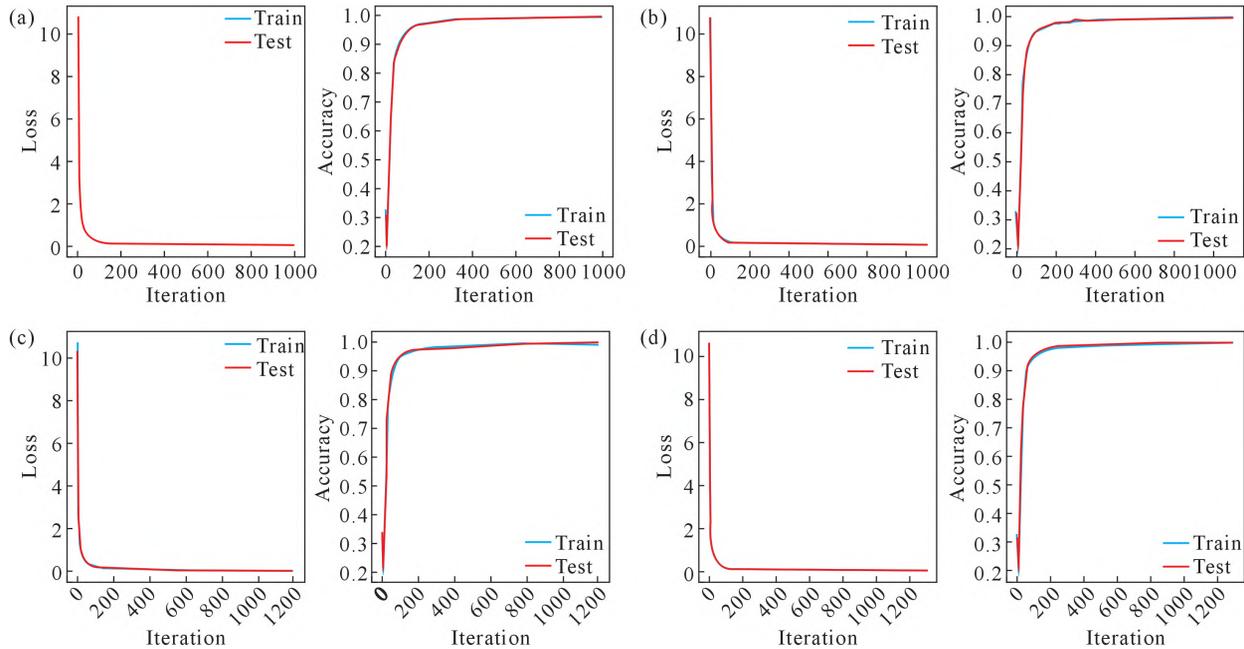


图 5 模型准确率与损失值可视化结果:(a) 训练次数 1 000,(b) 训练次数 1 100,(c) 训练次数 1 200,(d) 训练次数 1 300
Fig.5 Visualization results of the model accuracy and loss value: (a) 1 000 times, (b) 1 100 times, (c) 1 200 times, (d) 1 300 times

4 结语

基于 SMOTE 算法,创建了合成数据集,解决了数据不平衡问题,为复杂铸件缺陷预测模型提供训练数据集。针对企业对复杂铸件缺陷预测的要求,构建了复杂铸件缺陷预测模型,平均预测准确率达到 99.37%,说明构建的复杂铸件缺陷预测模型能够准确进行复杂铸件缺陷预测。

参考文献:

- [1] 武博,计效园,孙晓龙,等. 基于 X 射线图像的铸件缺陷智能检测与识别方法研究进展[J]. 铸造, 2022, 71(6): 709-718.
WU B, JI X Y, SUN X L, et al. Research progress on intelligent detection and recognition methods for casting defects based on X-Ray images[J]. Foundry, 2022, 71(6): 709-718.
- [2] 张元华,王玲玲. 铸件超声波探伤的缺陷分析研究——评《大型铸锻件及结构件超声波探伤》[J]. 铸造, 2021, 70(1): 133.
ZHANG Y H, WANG L L. Study on defect analysis of ultrasonic flaw detection of castings-comment on ultrasonic flaw detection of large cast and forged parts and structural parts[J]. Foundry, 2019, 70(1): 133.

- [3] 郭志明,李杰,李爽,等. 机器视觉技术在铸件缺陷检测中的应用[J]. 内燃机与动力装置, 2022, 39(3): 91-95.
GUO Z M, LI J, LI S, et al. Application of machine vision technology in casting defect detection[J]. Internal Combustion Engine & Powerplants, 2022, 39(3): 91-95.
- [4] 张井柱,翁月宏. 铝合金铸造工艺的 Metropolis 准则蜂群算法优化[J]. 机械设计与制造, 2021(8): 240-245.
ZHANG J Z, WENG Y H. Aluminum alloy casting technology optimization based on bee colony algorithm fused metropolis criterion[J]. Machinery Design & Manufacture, 2021, 366(8): 240-245.
- [5] LU F, LIANG Y C, WANG X Y, et al. Prediction of amorphous forming ability based on artificial neural network and convolutional neural network[J]. Computational Materials Science, 2022, 210: 111464.
- [6] 张建明,廖敦明,孙飞. 基于全连接卷积神经网络的铸造缩松缩孔缺陷快速预测[J]. 特种铸造及有色合金, 2020, 40(8): 841-845.
ZHANG J M, LIAO D M, SUN F. Fast prediction method of casting shrinkage defects based on fully convolutional networks[J]. Special Casting & Nonferrous Alloys, 2020, 40(8): 841-845.
- [7] 陈相吉. 基于深度学习的增压器铸造缺陷检测算法研究[D]. 大连:大连理工大学, 2020.
CHEN X J. Research on the defect detection for turbocharger cast-

- ing based on deep learning[D]. Dalian: Dalian University of Technology, 2020.
- [8] 栗子豪. 基于循环神经网络的模温预测模型研究[D]. 秦皇岛:燕山大学, 2021.
- LI Z H. Research on model temperature prediction based on recurrent neural network[D]. Qinhuangdao: Yanshan University, 2021.
- [9] WANG Z B, XU M Z, ZHANG Y N. Quantum pulse coupled neural network[J]. *Neural Networks*, 2022, 152: 105-117.
- [10] KUMAR R. A Lyapunov-stability-based context-layered recurrent pi-sigma neural network for the identification of nonlinear systems[J]. *Applied Soft Computing*, 2022, 122: 108836.
- [11] SOLÍS-PÉREZ J E, HERNÁNDEZ J A, PARRALES A, et al. Artificial neural networks with conformable transfer function for improving the performance in thermal and environmental processes[J]. *Neural Networks*, 2022, 152: 44-56.
- [12] SHI Y, ZHANG Y Y. The neural network methods for solving traveling salesman problem[J]. *Procedia Computer Science*, 2022, 199: 681-686.
- [13] 喻会, 肖凌. 基于神经网络的建筑装饰用铝青铜铸造性能优化研究[J]. *热加工工艺*, 2021, 50(7): 81-84.
- YU H, XIAO L. Study on casting performance optimization of aluminum bronze for building decoration based on neural network[J]. *Hot Working Technology*, 2021, 50(7): 81-84.
- [14] SHEN S L, ZHANG N, ZHOU A N, et al. Enhancement of neural networks with an alternative activation function tanhLU[J]. *Expert Systems with Applications*, 2022, 199: 117181.