

# 基于随机森林和人工神经网络构建种植体周炎的 诊断模型

杨浩然<sup>1,2</sup> 陈宇翔<sup>1,2</sup> 赵安娜<sup>1,2</sup> 程婷婷<sup>1,2</sup> 周建忠<sup>1,2</sup> 李自良<sup>1,2</sup>

1. 昆明医科大学附属口腔医院, 昆明 650000; 2. 云南省口腔医学重点实验室, 昆明 650000

**[摘要]** **目的** 本研究旨在揭示种植体周炎发生发展过程中参与调控的关键基因, 并通过随机森林 (RF) 和人工神经网络 (ANN) 构建种植体周炎的诊断模型。**方法** 本研究从 GEO 数据库中获取 GSE33774、GSE106090 和 GSE57631 数据集。对 GSE33774 和 GSE106090 数据集进行差异表达和功能富集分析, 通过蛋白质互作网络 (PPI) 和 RF 筛选出关键基因, 利用 ANN 建立种植体周炎的诊断模型, 并在 GSE33774 和 GSE57631 数据集中进行验证。同时, 构建转录因子-基因相互作用网络和转录因子-微小 RNA (miRNA) 调控网络。**结果** 本研究共筛选出 124 个参与调控种植体周炎的差异表达基因 (DEGs)。富集分析结果表明, DEGs 主要和免疫受体活性蛋白及细胞因子受体活性相关, 主要参与白细胞和中性粒细胞迁移的过程。PPI 和 RF 筛选出 6 个关键基因, 分别为 CD38、CYBB、FCGR2A、SELL、TLR4 和 CXCL8。受试者操作特征曲线 (ROC) 表明 ANN 模型具有较好的诊断性。本研究还发现 FOXC1、GATA2 和 NF- $\kappa$ B1 可能是种植体周炎中重要的转录因子, hsa-miR-204 可能是关键的 miRNA。**结论** RF 和 ANN 构建的种植体周炎的诊断模型可信度高, CD38、CYBB、FCGR2A、SELL、TLR4 和 CXCL8 是潜在的诊断标志物。FOXC1、GATA2 和 NF- $\kappa$ B1 可能是种植体周炎中重要的转录因子, hsa-miR-204 作为关键的 miRNA 在其中扮演着重要角色。

**[关键词]** 种植体周炎; 生物信息学; 随机森林; 人工神经网络; 诊断模型

**[中图分类号]** R78 **[文献标志码]** A **[doi]** 10.7518/hxkq.2024.2023275



本文链接 开放科学标识码

## Construction of a diagnostic model based on random forest and artificial neural network for peri-implantitis

Yang Haoran<sup>1,2</sup>, Chen Yuxiang<sup>1,2</sup>, Zhao Anna<sup>1,2</sup>, Cheng Tingting<sup>1,2</sup>, Zhou Jianzhong<sup>1,2</sup>, Li Ziliang<sup>1,2</sup>

1. Stomatological Hospital of Kunming Medical University, Kunming 650000, China; 2. Yunnan Provincial Key Laboratory of Stomatology, Kunming 650000, China

Supported by: Yunnan Provincial Health and Family Planning Commission Medical Reserve Talent Program (H20-17054)

Correspondence: Li Ziliang, E-mail: 1752114604@qq.com

**[Abstract]** **Objective** This study aimed to reveal critical genes regulating peri-implantitis during its development and construct a diagnostic model by using random forest (RF) and artificial neural network (ANN). **Methods** GSE-33774, GSE106090, and GSE57631 datasets were obtained from the GEO database. The GSE33774 and GSE106090 datasets were analyzed for differential expression and functional enrichment. The protein-protein interaction networks (PPI) and RF screened vital genes. A diagnostic model for peri-implantitis was established using ANN and validated on the GSE33774 and GSE57631 datasets. A transcription factor-gene interaction network and a transcription factor-micro-

RNA (miRNA) regulatory network were also established.

**Results** A total of 124 differentially expressed genes (DEGs) involved in the regulation of peri-implantitis were screened. Enrichment analysis showed that DEGs were mainly associated with immune receptor activity

**[收稿日期]** 2023-08-24; **[修回日期]** 2024-01-17

**[基金项目]** 云南省卫生和计划生育委员会医学后备人才项目 (H-2017054)

**[作者简介]** 杨浩然, 硕士, E-mail: 2360495671@qq.com

**[通信作者]** 李自良, 副教授, 博士, E-mail: 1752114604@qq.com

and cytokine receptor activity and were mainly involved in processes such as leukocyte and neutrophil migration. The PPI and RF screened six essential genes, namely, CD38, CYBB, FCGR2A, SELL, TLR4, and CXCL8. The receiver operating characteristic curve (ROC) indicated that the ANN model had an excellent diagnostic performance. FOXC1, GATA2, and NF- $\kappa$ B1 may be essential transcription factors in peri-implantitis, and hsa-miR-204 may be a key miRNA. **Conclusion** The diagnostic model of peri-implantitis constructed by RF and ANN has high confidence, and CD38, CYBB, FCGR2A, SELL, TLR4, and CXCL8 are potential diagnostic markers. FOXC1, GATA2, and NF- $\kappa$ B1 may be essential transcription factors in peri-implantitis, and hsa-miR-204 plays a vital role as a critical miRNA.

**[Key words]** peri-implantitis; bioinformatics; random forest; artificial neural network; diagnostic model

种植体周炎是指以菌斑为始动因素的种植体周围黏膜炎症以及种植体周围骨组织的进行性丧失<sup>[1-2]</sup>。种植体周炎的发病率从4.7%到43%不等<sup>[3-4]</sup>。手术是治疗种植体周炎的主要方法,其中种植体周围骨缺损的修复一直是重点和难点<sup>[5]</sup>。然而,现有治疗种植体周围骨缺损的方法的可靠性还需要更多的证据来证明。目前还没有特异性治疗种植体周炎的方法<sup>[6]</sup>。

种植体周炎是由菌斑引起的炎症性疾病。当细菌攻击宿主时,免疫细胞通过产生趋化因子诱导种植体周围的软组织和骨组织产生局部炎症反应,导致黏膜炎症和骨吸收<sup>[7-8]</sup>。炎症免疫反应是促进种植体周围骨吸收的主要机制,包括诱导破骨细胞的形成和活化<sup>[9]</sup>,然而,种植体周炎的发病机制尚未阐明<sup>[10]</sup>。因此,寻找种植体周炎发生发展的关键基因对于探索其潜在的发病机制和有效的治疗策略具有重要意义。

近年来随着微阵列高通量测序技术的快速发展,发现越来越多与种植体周炎相关的基因,为种植体周炎的诊断和治疗提供大量的依据。同时,机器学习技术也在全面迈向临床医学领域。机器学习旨在通过对数据信息的自主学习和分析,对未知预测样本进行高精度的分类或判别。目前,主要有决策树(decision trees, DT)、人工神经网络(artificial neural network, ANN)、支持向量机(support vector machine, SVM)、随机森林(random forest, RF)等多种算法<sup>[11-14]</sup>。在当前的疾病诊断中,使用机器学习分类算法进行疾病诊断是比较流行的<sup>[15]</sup>。近年来,许多学者<sup>[16-17]</sup>利用机器学习方法构建分类模型,取得了良好的效果。

本研究通过生物信息学方法从基因表达库(gene expression omnibus, GEO)中筛选出种植体周炎牙龈组织与健康牙龈组织之间的差异表达基因(differently expressed genes, DEGs),对DEGs进行差异表达和功能富集分析,构建一个蛋白质相互作用网络(protein-protein interaction networks, PP-

I)来筛选枢纽基因。用RF进一步筛选出靶基因,通过合成数据构建一个ANN模型,并在数据集集中进行验证。

## 1 材料和方法

### 1.1 数据下载与处理

本研究中的数据集来源于GEO数据库(<https://www.ncbi.nlm.nih.gov/geo/>),使用关键词“peri-implantitis”进行搜索,生物类型是智人。纳入标准:包括已建立的种植体周炎样本作为种植体周炎组和健康牙龈或骨组织样本作为对照组的数据集。最终筛选出GSE33774、GSE106090和GSE57631三个数据集。其中,GSE33774包括8个健康牙龈组织样本和7个种植体周炎组织样本,GSE106090包括6个健康牙龈组织样本和6个种植体周炎组织样本,GSE57631包括2个健康骨组织样本和6个种植体周炎组织样本。在本研究中,训练模型使用的是软组织样本GSE33774和GSE106090,而验证模型使用的是软组织样本GSE33774和骨组织样本GSE57631。对基因芯片数据集的原始数据进行背景校正和归一化处理。分析工作流程如图1所示。

### 1.2 差异表达分析

使用“limma”软件包分析GSE33774和GSE106090数据集中种植体周炎组与对照组之间基因表达的差异<sup>[18]</sup>。 $P < 0.05$ 且 $|\log_{2}FC$ (变化倍数) $\geq 1$ 为显著的DEGs。本研究使用“ggplot2”软件包绘制DEGs的火山图;将GSE33774和GSE106090合并为一个数据集,使用“pheatmap”软件包绘制DEGs的热图<sup>[19]</sup>。

### 1.3 基因本体论(gene ontology, GO)与通路富集分析

为进一步研究差异基因的功能和作用途径,使用R语言“clusterProfiler”软件包对124个DEGs进行基因组百科全书(Kyoto encyclopedia of ge-

nes and genomes, KEGG) 与 GO 分析。富集分析结果使用 Benjamini-Hochberg 校正法, 阈值设置为  $P < 0.05$ 。GO 分析分为分子功能 (molecular function, MF)、生物学过程 (biological process, BP)

及细胞成分 (cellular component, CC) 3 类<sup>[20]</sup>。用 KEGG 来识别基因富集的途径<sup>[21]</sup>。富集分析结果使用 “dplyr”、“ggplot2” 软件包和在线网站 (<https://hiplot.com.cn/>) 进行可视化。

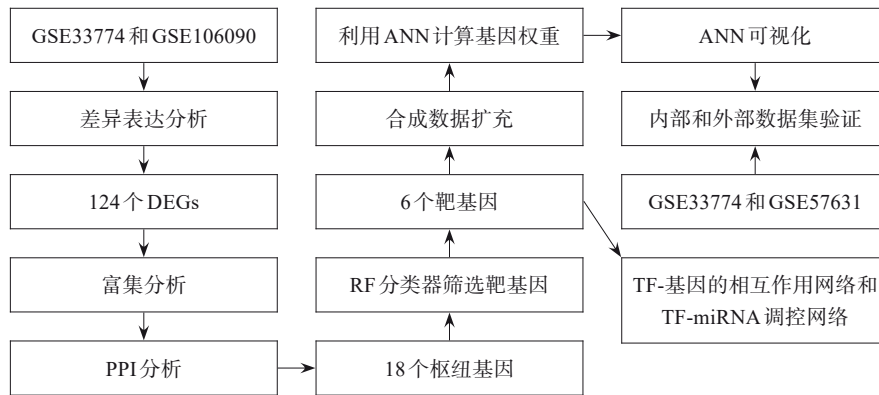


图 1 研究流程图

Fig 1 Flowchart of the study

### 1.4 构建 PPI 及筛选枢纽基因

STRING 数据库 (<https://cn.string-db.org/>) 是一个基于公共数据库和文献信息的 PPI 数据库<sup>[22]</sup>。本研究将 DEGs 输入到 STRING 数据库, 显著阈值设置为最低交互得分  $> 0.4$  (低置信度)。Cytoscape 软件用来构建 PPI 并分析其拓扑特征<sup>[23]</sup>, 其中 Cytoscape 内置的 4 种算法 (Degree、EPC、MCC 和 MNC) 用来筛选 PPI 中的枢纽基因<sup>[24]</sup>; 4 种算法排名前 20 的基因取交集, 得到最终的枢纽基因。

### 1.5 RF 筛选靶基因

使用 RF 筛选基因时, 首先需要确定该数据使用 RF 时最佳的 DT 数量, 以此保证模型的错误率能够达到最低且此时模型的错误率也相对稳定。接下来, 使用递减精度法 (基尼系数法) 从 RF 模

型中获得维度重要性值, 选择重要性值大于 1 且排名前 6 的基因作为种植体周炎的靶基因用于后续模型构建。本研究通过 “RandomForest” 软件包建立 RF 模型。

### 1.6 合成数据扩充

本研究合成数据基于 Montanez<sup>[25]</sup> 创建的 SDV 库: 一个用于生成合成数据的开源库。用 Python 的 SDV 库在基于数据集 GSE33774 和 GSE106090 的基础上生成合成数据, 并将合成数据用于模型的训练上。SDV 主要是使用递归条件参数聚合的机器学习算法来进行合成数据的实现。当用户提供数据和架构后, 将模型拟合到数据中, 从拟合的模型中随机生成新的综合数据 (图 2)。

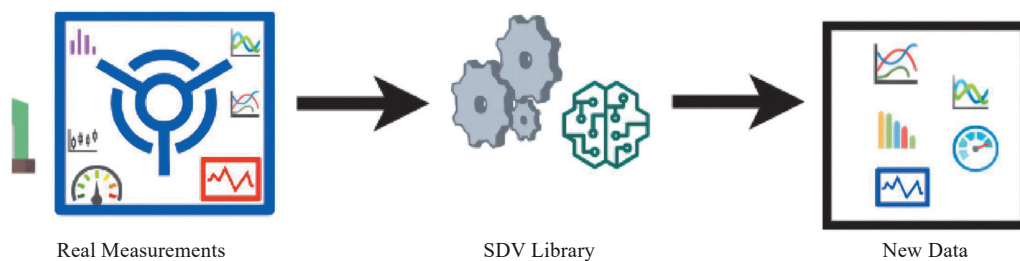


图 2 SDV 工作流程图

Fig 2 SDV workflow diagram

### 1.7 构建和验证 ANN 模型

本研究通过合成数据中 6 个基因的表达分数为特征来训练 ANN 模型, 对样本是否患有种植体周炎进行预测。在选择参数时, 使用多少层和多少个神经元并没有固定的规则。以灵敏度和特异度为坐标轴, 绘制受试者操作特征曲线 (receiver

operating characteristic curve, ROC); 通过曲线下面积 (area under the curve, AUC), 可以评估 ANN 模型的诊断性能。本研究使用 “neuralnet” 软件包建立 ANN 模型, 利用 R 包 “pROC” 生成 ROC 曲线。

1.8 内部和外部数据集验证

在GSE33774和GSE57631数据集中分别绘制ROC曲线,通过AUC值分别验证每个基因和ANN模型的诊断性能。

1.9 转录因子(transcription factors, TF)-基因相互作用网络和TF-微小RNA(microRNA, miRNA)核心调控网络

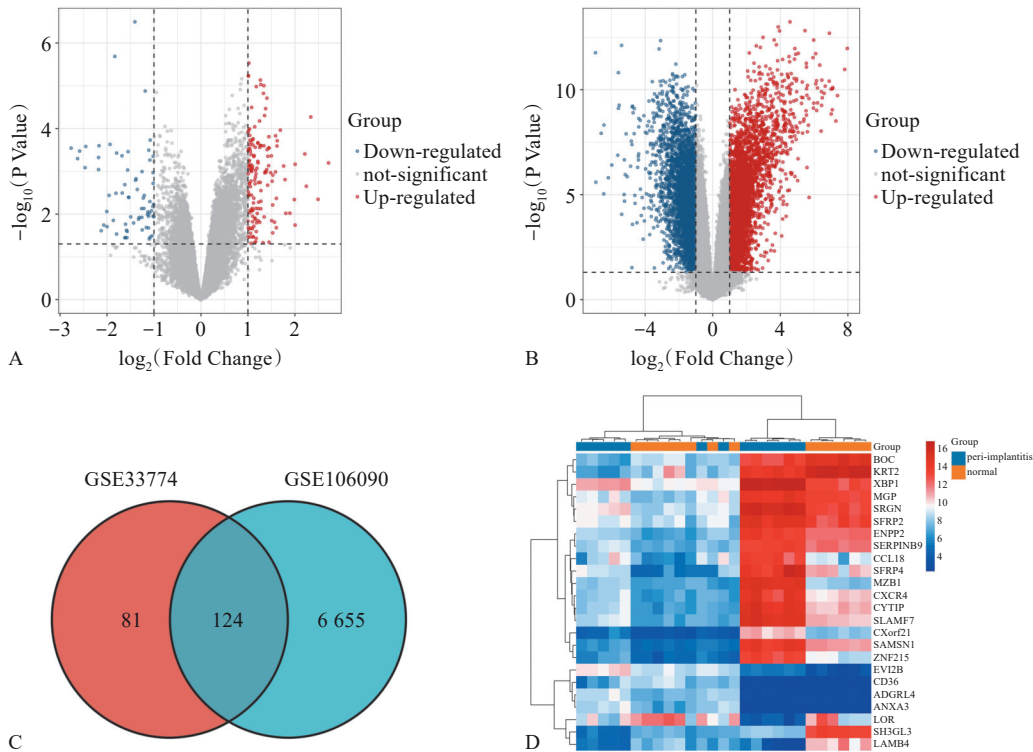
TF是一群能调控基因表达的蛋白质分子。miRNA属于非编码RNA,主要参与基因转录后水平的调控。根据筛选出的重要基因构建了TF-基因相互作用网络和TF-miRNA调控网络。TF的靶点来自JASPAR数据库(https://jaspar.genereg.net/); TF-miRNA相互作用信息从RegNetwork数据库(https://regnetworkweb.org/)中获取。NetworkAna-

lyst数据库(https://www.networkanalyst.ca/)生成了TF-基因相互作用网络和TF-miRNA调控网络<sup>[26]</sup>,并通过Cytoscape软件进行可视化。

2 结果

2.1 差异表达分析结果

对GSE33774和GSE106090数据集中种植体周炎牙龈和健康牙龈的基因数据行差异分析,得到有意义的上调基因89个,下调基因35个,共124个DEGs。利用GSE33774和GSE106090数据集的DEGs绘制了火山图(图3A、B)。将GSE33774和GSE106090合并成一个数据集(图3C),绘制出前24个DEGs的热图(图3D)。



A、B: 分别表示GSE33774和GSE106090中DEGs的火山图,其中红色代表在种植体周炎样本中显著高表达的基因,蓝色代表在健康牙龈组织中显著高表达的基因,灰色代表无明显变化的基因; C: GSE33774和GSE106090中共同DEGs的维恩图; D: 合成数据集中前24个DEGs的热图。

图3 差异表达分析结果

Fig 3 Differential expression analysis results

2.2 GO和通路富集分析结果

对124个DEGs进行了GO和KEGG富集分析。GO分析结果显示,DEGs在CC中主要富集在质膜外侧、分泌颗粒膜和膜囊泡上。MF主要包括免疫细胞受体活性、细胞因子受体活性和细胞因子结合位点等。在BP中,DEGs主要参与白细胞迁移、髓系白细胞迁移和中性粒细胞迁移等过程。

KEGG通路主要包括疟疾、利什曼病、阿米巴病等(图4)。

2.3 PPI分析结果

在STRING数据库中输入124个DEGs,通过Cytoscap软件构建PPI(最小置信度大于0.4)(图5A),其中基因按照DC值排列。构建的网络包含75个节点和348条边。利用MCODE插件分析PPI

中的重要模块，一个由15个节点和90条边组成的模块被认为是PPI中的一个重要的簇（图5B）。对簇中15个基因进行了GO本体和通路富集分析。结果显示，BP包括白细胞迁移、髓系白细胞迁移和中性粒细胞迁移；MF包括NAD+核苷酸酶生成和NAD(P)+核苷酸活性等；KEGG通路包括脂质与动脉粥样硬化、疟疾和军团菌病等（图5C）。通过CytoHubba的4种算法（Degree、EPC、MCC和MNC）筛选PPI中的枢纽基因，将4种算法排序前20名的基因取交集，最终得到18个枢纽基因（图5D）。这18个基因均为上调基因，包括PTPRC、TLR4、IL6、IL1B、CXCR4、MRC1、PECAM1、SELL、MMP9、FCGR2A、CXCL8、CYBB、CD-38、CXCL1、C3AR1、CXCR1、TREM1、IRF4。

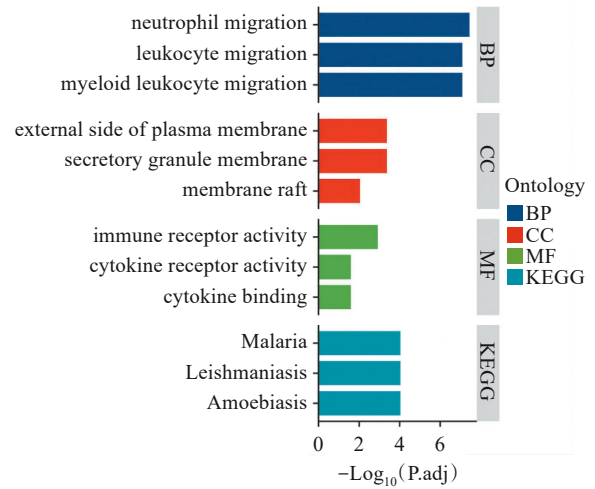


图4 DEGs的功能富集分析结果

Fig 4 Results of functional enrichment analysis of DEGs

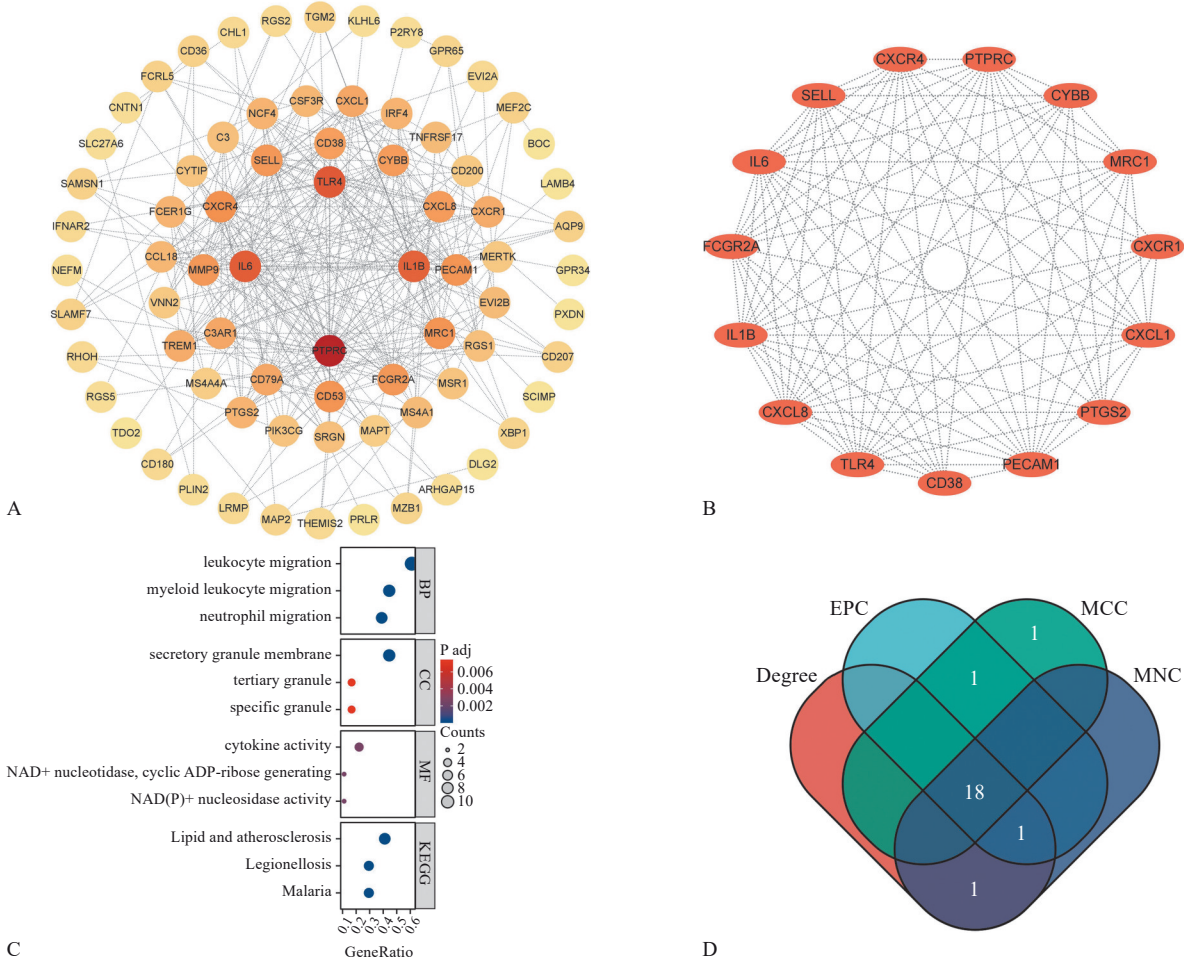


图5 PPI构建和分析结果

Fig 5 PPI construction and analysis results

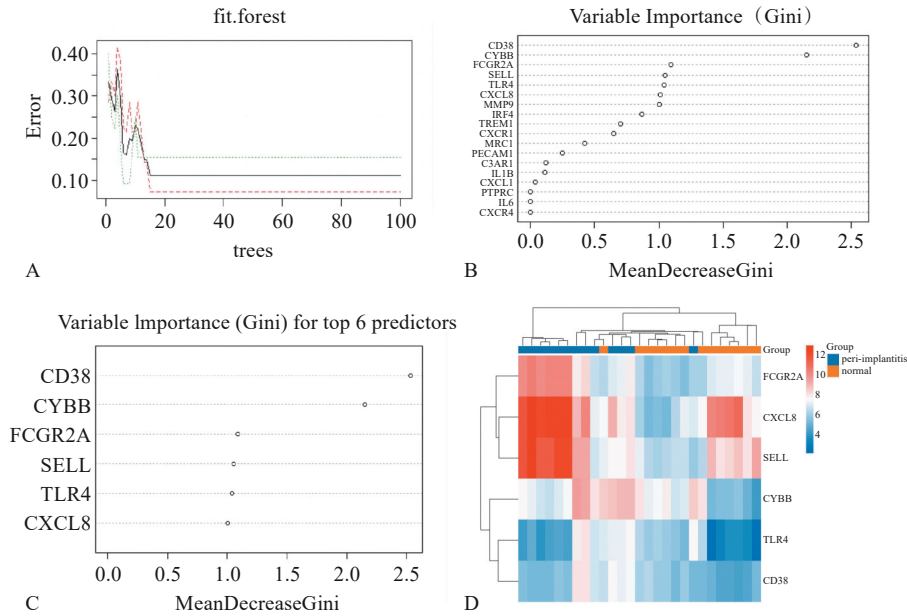
### 2.4 RF筛选靶基因的结果

为了获得靶基因，将PPI筛选出的18个枢纽基因输入到RF中，得出随机森林树的数量与模型

误差之间的相关图（图6A），选择由15棵DT组成的RF模型对数据进行训练。依据基因表达分数对预测结果影响的重要性程度进行排序（图6B），选

取重要程度最高的前6个基因, 它们分别是 CD38、CYBB、FCGR2A、SELL、TLR4 和 CX-

CL8 (图6C), 将6个重要基因进行无监督层次聚类并绘制热图 (图6D)。



A: DT数量对模型误差的影响图; B、C表示基于基尼系数法从RF模型中获得维度重要性值, X轴表示基尼系数, Y轴表示基因变量; D: 6个特征基因的无监督分层聚类结果热图。

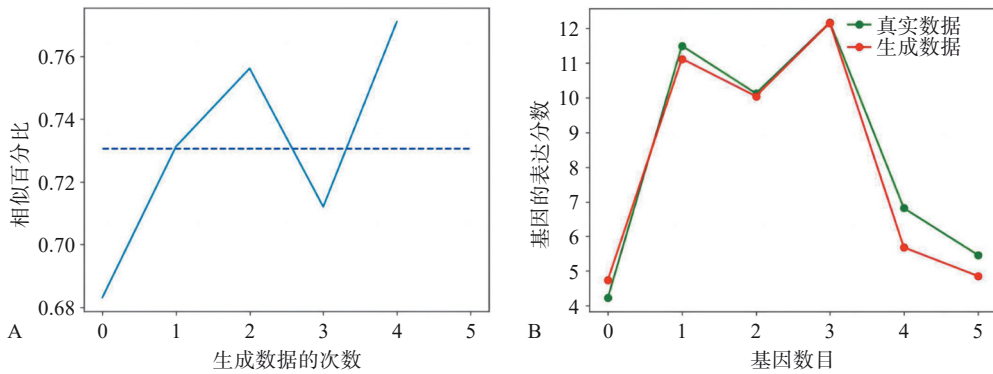
图 6 RF筛选基因结果

Fig 6 Results of RF screening of genes

2.5 合成数据扩充结果

本研究使用SDV合成数据, 以GSE33774和GSE106090结合得到的27个样本为主样本, 生成另外27个虚拟合成样本, 最后使用54个样本训练

ANN模型。SDV库生成的数据与真实数据的相似度平均可达73% (图7A)。在对生成的某组数据与真实数据进行比较时, 发现SDV可以生成几组与真实数据组高度相似的数据 (图7B)。



A: SDV生成数据与真实数据的相似百分比图; B: 折线图。

图 7 生成合成数据结果

Fig 7 Generate synthetic data results

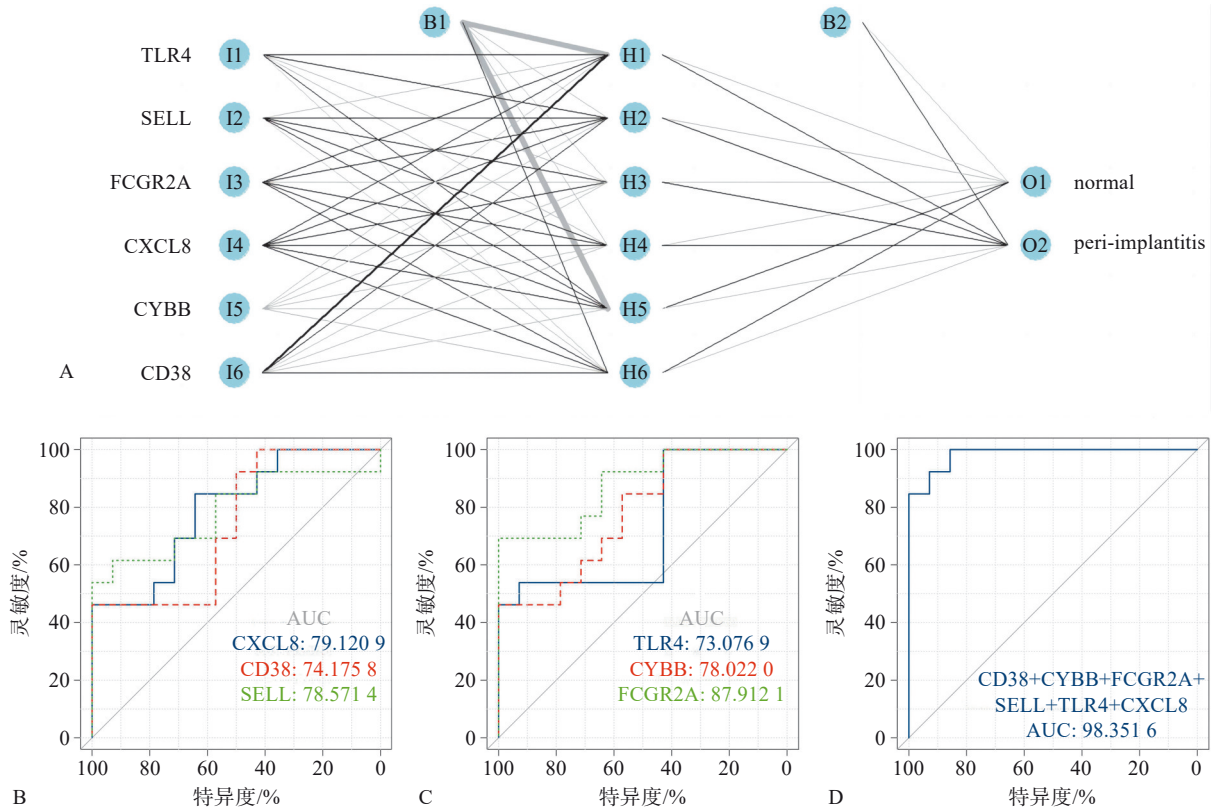
2.6 ANN模型可视化结果

本研究选择RF模型中筛选出的6个关键基因的表达分数构建ANN模型, 以诊断样本是否患有种植体周炎。由于样本数量不足以支持ANN训练和验证所需的样本量, 因此使用SDV扩展数据以获得54个样本。隐藏层使用6个神经元, 训练好的ANN模型结构见图8A。在GSE33774和

GSE106090合成的数据集中, CYBB、FCGR2A、SELL、TLR4、CXCL8和CD38的AUC值分别为78.022 0%、87.912 1%、78.571 4%、73.076 9%、79.120 9%和74.175 8% (图8B、C), 整个模型的AUC值为98.352% (图8D)。使用5倍交叉验证的方法对ANN模型进行验证, 结果表明在5次交叉验证中ANN模型的AUC值分别为84.9%、82.9%、

83.8%、79.7% 和 83.7%，平均 AUC 值为 82.8% (图9)。以上结果表明，本研究构建的ANN模型

具有较好的泛化能力和鲁棒性。



A: ANN模型可视化图; B: CXCL8、CD38和SELL基因的ROC曲线; C: CYBB、FCGR2A和TLR4的ROC曲线; D: 该诊断模型的ROC曲线。

图 8 ANN模型可视化及其诊断性能结果

Fig 8 Results of ANN model visualization and diagnostic performance

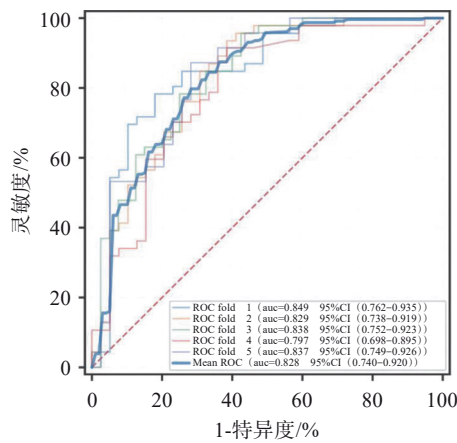


图 9 5折交叉验证结果

Fig 9 5-fold cross-validation results

2.7 内部和外部数据集验证结果

在GSE33774数据集中，CYBB、FCGR2A、TLR4、SELL、CXCL8和CD38 6个基因在ROC曲线中的AUC值均大于80% (图10A、B); 取6个基因的平均表达分值绘制ROC曲线，整体的AUC值为96.428 6% (图10C)。在外部数据集GSE-

57631中，6个基因在ROC曲线中的AUC值大于70% (图10D、E); 取6个基因的平均表达分值绘制ROC曲线，整体的AUC值为83.333 3% (图10F)。上述验证结果表明，本研究构建的种植体周炎诊断模型具有良好的诊断性能。

2.8 TF-基因相互作用网络和TF-miRNA调控网络

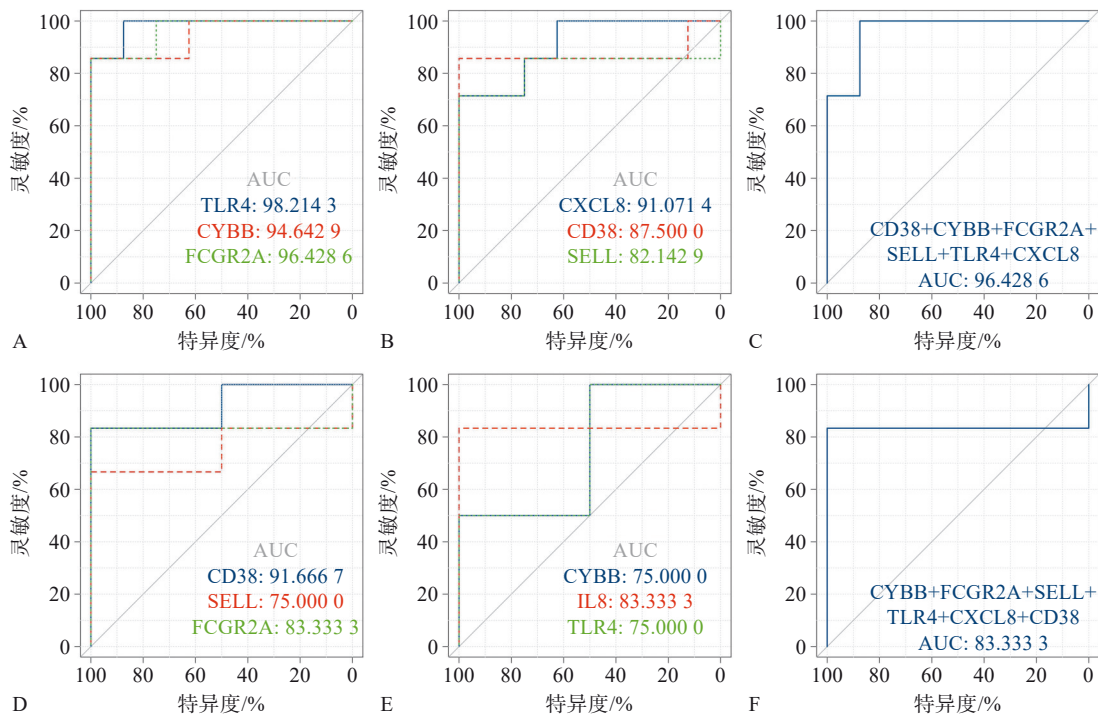
本研究构建的TF-基因相互作用网络包含27个节点和41条边 (图11A)。FOXC1和GATA2与其他基因的相互作用率较高。TF-miRNA调控网络包含97个节点和107条边 (图11B)。NF-κB1是Degree和Betweenness分数最高的TF，hsa-miR-204是Degree和Betweenness分数最高的miRNA。

3 讨论

本研究共获得了124个DEGs。GO结果显示，在BP过程中，DEGs主要富集在白细胞迁移、髓系白细胞迁移、中性粒细胞迁移和粒细胞迁移过程中。白细胞，包括中性粒细胞和髓系白细胞，

在感染和炎症反应中起着至关重要的作用。中性粒细胞能促进促炎因子的释放,进而导致种植体周炎中的软硬组织损伤<sup>[27]</sup>。种植体周炎中M1极化巨噬细胞的数量显著增加,主要参与由细菌引起的抗炎反应<sup>[28]</sup>。对于种植体周炎来说,白细胞的迁移是一个复杂的过程,任何影响白细胞迁移的因素,如病原体的毒性、宿主免疫状态以及种植体表面的物理和化学特性,都可能影响种植体周

炎的发展<sup>[29]</sup>。因此,理解白细胞在种植体周炎中的迁移及其作用机制,对于改善种植体周炎的诊断和治疗具有重要意义。在MF过程中,DEGs主要富集在免疫受体活性、细胞因子受体活性和细胞因子结合。免疫受体、细胞因子受体及细胞因子结合在种植体周炎中起关键作用,它们一起参与调节免疫反应,并影响种植体周炎的进程<sup>[30]</sup>。



A、B、C表示CYBB、FCGR2A、TLR4、SELL、CXCL8和CD38在数据集GSE33774中的ROC曲线; D、E、F表示CYBB、FCGR2A、TLR4、SELL、CXCL8和CD38在数据集GSE57631中的ROC曲线。

图 10 内部和外部数据集验证结果

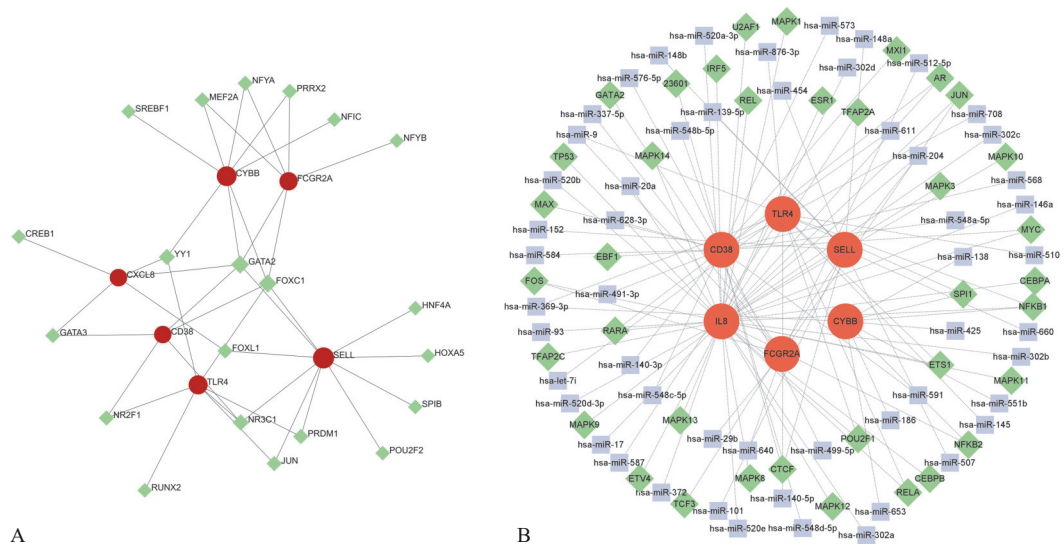
Fig 10 Internal and external dataset validation results

本研究中,通过PPI和RF筛选出了6个靶基因,分别是TLR4、CXCL8、CD38、FCGR2A、CYBB和SELL。TLR4是一种重要的模式识别受体,可通过识别和结合病原体的病原相关分子模式(pathogen-associated molecular pattern, PAMPs)来引发机体的免疫炎症反应。Deng等<sup>[31]</sup>研究发现,TLR4通过调节B细胞的浸润增加了核因子κB受体活化因子配体(receptor activator of nuclear factor kappa-B ligand, RANKL)/骨保护素(osteoprotegerin, OPG)的比例和肿瘤坏死因子-α(tumor necrosis factor-α, TNF-α)的表达,加重了牙龈卟啉单胞菌感染引起的种植体周炎的炎症反应。先前的研究表明,miR-146a通过调节TLR2/TLR4信号传导增强了对种植体周围骨吸收的抑制<sup>[32]</sup>,同时参与了TLR5信号传导的Wnt4a诱导的炎性细胞

因子的产生,并导致种植体周炎中细胞外基质的分解<sup>[33]</sup>。CXCL8在炎症反应过程中发挥着重要作用,主要通过CXCR1和CXCR2两种受体结合来介导炎症反应<sup>[34-35]</sup>。CXCL8和其受体的结合激活磷脂酰肌醇3-激酶(phosphatidylinositol-3-kinases, PI3K)/Akt、丝裂原活化蛋白激酶(mitogen-activated protein kinase, MAPK)、NF-κB等下游信号传导,引导中性粒细胞向炎症部位迁移,促进炎症反应的发生<sup>[36]</sup>。Zhang等<sup>[37]</sup>的研究表明,促炎细胞因子白细胞介素(interleukin, IL)-6、IL-1β和CXCL8在种植体周炎中上调。先前的研究<sup>[38]</sup>表明,患者龈沟液中IL-1β、CXCL8和TNF-α的表达水平可以作为种植体周围炎的诊断标志之一。CD38是一种细胞表面分子,可通过产生环ADP核糖(cyclic ADP-ribose, cADPR)来激活PI3K-Akt通

路，促进炎症因子的产生和分泌<sup>[39]</sup>。对种植体周炎中记忆B细胞的亚群进行了定量分析，发现相比牙周炎，种植体周炎具有更高活性的记忆B细胞（CD27+CD38hi 细胞比例较高）<sup>[40]</sup>。FCGR-2A 基因是人体免疫球蛋白G2（IgG2）抗体的唯一受体，它可以在多种先天免疫细胞中表达，如巨噬

细胞和淋巴细胞，并调节细胞识别和吞噬作用<sup>[41]</sup>。研究<sup>[42]</sup>证实，FcγR II a 的基因多态性可影响IgG1和IgG2的结合，降低免疫系统对抗微生物的能力，增加对炎症的易感性。研究<sup>[43]</sup>发现FcγR II a-H131-R、FcγR III a-V158F和FcγR III b-I232-T 3个位点的多态性与慢性牙周炎和种植体周炎密切相关。



A: TF-基因相互作用网络图，红色节点代表基因，其他节点代表TF；B: TF-miRNA 调控网络图，红色节点代表基因，紫色节点代表 miRNA，绿色节点代表 TF。

图 11 TF-基因相互作用网络和 TF-miRNA 调控网络结果

Fig 11 Results of visualization of TF-gene interaction networks and TF-miRNA regulatory networks

此外，本研究首次发现CYBB和SELL基因与种植体周炎的发生有关。CYBB基因编码的是一种叫做细胞色素 b (-245) β亚基的蛋白，它是烟酰胺腺嘌呤二核苷酸磷酸（nicotinamide adenine dinucleotide phosphate, NADPH）氧化酶复合体中的一部分。NADPH氧化酶是一种负责细胞内氧化反应的重要酶，它在免疫细胞中起着关键作用<sup>[44]</sup>。CYBB基因的缺陷会使得NADPH氧化酶功能受损，导致细胞内的活性氧物质无法生成，进一步影响巨噬细胞、中性粒细胞等免疫细胞的杀菌和损伤效应<sup>[45]</sup>。笔者推测，在种植体周炎中，CYBB基因的异常表达或突变可能会降低细胞色素 b (-245) β亚基的功能，减少免疫细胞对感染病原体的杀灭效应，进而导致炎症反应的持续和进展。SELL基因编码的是选择素L（L-selectin）蛋白。选择素L主要存在于白细胞表面，它参与了白细胞的滚动、黏附和迁移过程<sup>[46]</sup>。在炎症反应中，选择素L能够与细胞黏附分子结合，通过与血管内皮细胞相互作用，使白细胞停滞在炎症部位并渗透到组织内<sup>[47]</sup>。通过选择素L的表达和功能调节，可以控制白细胞的黏附和迁移，从而影响种植体周

炎的发生和发展。进一步研究上述基因与种植体周炎之间的潜在联系，有助于深入理解其发病机制，并指导制定个体化治疗和预防策略。

本研究发现FOXC1、GATA2和NF-κB1可能是种植体周炎发生发展中重要的TF，has-miR-204可能是关键的miRNA。研究<sup>[48]</sup>表明，FOXC1过表达可以显著减少炎症细胞（如中性粒细胞和巨噬细胞）的浸润，并减少IL-1β和TNF-α等炎症因子的产生。FOXC1还可以抑制NF-κB活化，进一步抑制炎症反应。同时，FOXC1过表达抑制了细胞凋亡相关蛋白Caspase-3的活化，并增加了细胞凋亡抑制蛋白Bcl-2的表达<sup>[49]</sup>。除了抑制炎症反应，FOXC1可以调控血管内膜细胞的分化和迁移，从而影响血管新生的过程<sup>[50]</sup>。FOXC1在种植体周炎中可能起到了抑制炎症反应、抗凋亡和调控血管新生的作用。GATA2是一种在多个细胞类型中表达的TF，它参与调控多种细胞的生物学过程，包括细胞增殖、分化和炎症反应等<sup>[51]</sup>。GATA2通过与NF-κB的相互作用、调控STAT3的表达、与PU.1形成转录复合物以及蛋白激酶C（protein kinase C, PKC）和MAPK信号通路的调控等机制和

通路,参与调节促炎因子(如IL-1 $\beta$ 、IL-6和TNF- $\alpha$ )和趋化因子(如CXCL2和CCL2)的转录,从而调控炎症反应的强度和规模<sup>[52-53]</sup>。NF- $\kappa$ B1是NF- $\kappa$ B家族的成员之一。在炎症过程中,炎症信号激活NF- $\kappa$ B1,使其从胞质中转移到细胞核,结合到靶基因的启动子区域上,从而启动基因的转录<sup>[54]</sup>。在牙周炎中,炎症反应引发的细胞损伤和氧化应激可能导致细胞凋亡,而NF- $\kappa$ B1的活化可促使细胞存活和阻止凋亡的发生<sup>[55]</sup>。因此,NF- $\kappa$ B1可能在种植体周炎中参与了细胞凋亡的调控。has-miR-204(miR-204)是一种miRNA,在种植体周炎中发挥重要作用。has-miR-204的上调可以促使巨噬细胞向M2型极化,有助于抑制炎症反应的发生和发展<sup>[56]</sup>。同时,has-miR-204可以通过抑制细胞周期相关蛋白的表达,如cyclin D1、CDK4等,来抑制细胞的增殖;has-miR-204可以通过靶向调控一些关键的TF,如MEF2C、FOXC1等,来调节细胞分化的过程;has-miR-204可以通过靶向调控自噬相关基因,如Beclin-1、LC3等,来调节细胞的自噬过程<sup>[57-60]</sup>。总之,has-miR-204在种植体周炎中可能起到了抗炎作用,并参与了细胞增殖、分化和自噬等过程。进一步的研究可以帮助更好地理解TF和miRNA在种植体周炎中的调控机制,并为种植体周炎的预防和治疗提供新的思路和策略。

本研究使用RF算法来筛选靶基因。与传统的分类算法相比,RF的分类精度更高<sup>[61]</sup>。首先,RF能够避免过拟合问题,从而提高分类精度。其次,RF能够灵活处理连续或离散变量,无需进行归一化处理,从而减少数据预处理的步骤。最后,RF能够容忍异常值和缺失值,避免个体差异对模型的过度影响。另一方面,ANN的结构通常包括输入层、隐藏层和输出层。每一层都有其特定的功能,分工明确<sup>[62]</sup>。首先,ANN通过激活函数来模拟非线性特征,具有其他线性建模方法所不具备的优势。其次,ANN通过简单的网络效应可以实现复杂的功能。最后,ANN通过调整权值和阈值来拟合高维复杂函数,完成对基因诊断疾病的高维近似<sup>[63]</sup>。在本研究中,将输入层设置为“基因”,输出层设置为“疾病”,模拟基因和疾病间的前向连接,实现通过基因来诊断疾病。

种植体周炎是一种常见的生物学并发症,其主要特征是炎症和进行性边缘骨吸收。然而,关于种植体周炎的发病机制目前存在较大争议。因此,构建一种基因诊断模型对种植体周炎进行准确诊断具有重要科学意义和应用价值。该模型分

析种植体周炎病检组织中特定基因的表达情况,寻找与种植体周炎相关的生物标志物或异常表达的基因,可作为早期诊断和评估工具。相比传统临床检查方法,基因诊断模型可提供较灵敏、准确、可靠的结果。通过对病检组织样本进行基因表达分析,也可深入了解种植体周炎的发病机制,为治疗和预防提供理论依据。最后,通过识别与种植体周炎相关的特定基因变异或异常表达,可以制定个体化治疗方案,提高治疗效果和预后。总之,建立种植体周炎的基因诊断模型可加深对发病机制的理解,为临床诊断、治疗和预防提供新方法和策略,最终改善患者的口腔健康。

在本研究中,图3D的热图显示部分种植体周炎样本与其他种植体周炎样本之间存在显著差异,但与健康牙龈样本相似。这可能是因为在基因表达水平上,炎症样本和健康样本之间存在共同的表达模式。然而,为了确定这些样本是否具有代表性,需要进行更全面的分析和评估,包括样本数量、采样时机、个体差异、全身系统性疾病等因素。在本研究中,验证模型使用的是骨组织样本GSE57631,而训练模型使用的是软组织样本GSE33774和GSE106090。这两个组织来源之间的基因表达模式和生物过程可能存在差异,因此需要进行额外的数据预处理或采用其他策略来解决组织差异的影响,如数据标准化、特征选择或域适应方法等。本研究对所有数据进行了标准化处理,以最大限度地减少组织来源差异对模型验证的影响,确保模型的鲁棒性和泛化能力。然而,本研究也存在一些局限性。由于种植周炎样本数据集较小,因此通过Python合成数据来建立模型。尽管取得了不错的效果,但今后的研究中,将增加更多的训练集以提升模型的性能。

综上所述,本研究利用RF和ANN构建了种植体周炎的诊断模型。CD38、CYBB、FCGR2A、SELL、TLR4和CXCL8是潜在的诊断标志物。同时,本研究发现FOXC1、GATA2和NF- $\kappa$ B1可能是种植体周炎中重要的TF,hsa-miR-204作为关键的miRNA在其中扮演着重要角色。这些发现可以为种植体周炎的发病机制提供新的见解。

利益冲突声明:作者声明本文无利益冲突。

#### [参考文献]

- [1] Salvi GE, Cosgarea R, Sculean A. Prevalence and mechanisms of peri-implant diseases[J]. J Dent Res, 2017,

- 96(1): 31-37.
- [2] Caton JG, Armitage G, Berglundh T, et al. A new classification scheme for periodontal and peri-implant diseases and conditions—Introduction and key changes from the 1999 classification[J]. *J Periodontol*, 2018, 89 (Suppl 1): S1-S8.
- [3] Sahrman P, Gilli F, Wiedemeier DB, et al. The microbiome of peri-implantitis: a systematic review and meta-analysis[J]. *Microorganisms*, 2020, 8(5): 661.
- [4] Derks J, Tomasi C. Peri-implant health and disease. A systematic review of current epidemiology[J]. *J Clin Periodontol*, 2015, 42 Suppl 16: S158-S171.
- [5] Schwarz F, Jepsen S, Obreja K, et al. Surgical therapy of peri-implantitis[J]. *Periodontol 2000*, 2022, 88(1): 145-181.
- [6] Listl S, Frühauf N, Dannewitz B, et al. Cost-effectiveness of non-surgical peri-implantitis treatments[J]. *J Clin Periodontol*, 2015, 42(5): 470-477.
- [7] Petković-Curcin A, Matić S, Vojvodić D, et al. Cytokines in pathogenesis of peri-implantitis[J]. *Vojnosanit Pregl*, 2011, 68(5): 435-440.
- [8] Corrêa MG, Pimentel SP, Ribeiro FV, et al. Host response and peri-implantitis[J]. *Braz Oral Res*, 2019, 33(suppl 1): e066.
- [9] Insua A, Monje A, Wang HL, et al. Basis of bone metabolism around dental implants during osseointegration and peri-implant bone loss[J]. *J Biomed Mater Res A*, 2017, 105(7): 2075-2089.
- [10] Fragkioudakis I, Tseleki G, Doufexi AE, et al. Current concepts on the pathogenesis of peri-implantitis: a narrative review[J]. *Eur J Dent*, 2021, 15(2): 379-387.
- [11] Che D, Liu Q, Rasheed K, et al. Decision tree and ensemble learning algorithms with their applications in bioinformatics[J]. *Adv Exp Med Biol*, 2011, 696: 191-199.
- [12] Jiřík M, Moulisová V, Hlaváč M, et al. Artificial neural networks and computer vision in medicine and surgery [J]. *Rozhl Chir*, 2022, 101(12): 564-570.
- [13] Ma CL, Yuan YB. A novel support vector machine with globality-locality preserving[J]. *Sci World J*, 2014, 2014: 872697.
- [14] Pfeifer B, Holzinger A, Schimek MG. Robust random forest-based all-relevant feature ranks for trustworthy AI [J]. *Stud Health Technol Inform*, 2022, 294: 137-138.
- [15] Lindon JC, Holmes E, Bollard ME, et al. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis[J]. *Biomarkers*, 2004, 9(1): 1-31.
- [16] Tian Y, Yang J, Lan M, et al. Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure[J]. *Aging (Albany NY)*, 2020, 12(24): 26221-26235.
- [17] Sun D, Peng H, Wu Z. Establishment and analysis of a combined diagnostic model of Alzheimer's disease with random forest and artificial neural network[J]. *Front Aging Neurosci*, 2022, 14: 921906.
- [18] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [19] Lai D, Ma W, Wang J, et al. Immune infiltration and diagnostic value of immune-related genes in periodontitis using bioinformatics analysis[J]. *J Periodontol Res*, 2023, 58(2): 369-380.
- [20] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine[J]. *Nucleic Acids Res*, 2021, 49(D1): D325-D334.
- [21] Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms[J]. *Nucleic Acids Res*, 2021, 49(D1): D545-D551.
- [22] Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. *Nucleic Acids Res*, 2021, 49(D1): D605-D612.
- [23] Otasek D, Morris JH, Bouças J, et al. Cytoscape automation: empowering workflow-based network analysis[J]. *Genome Biol*, 2019, 20(1): 185.
- [24] Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. *BMC Syst Biol*, 2014, 8(Suppl 4): S11.
- [25] Montanez A. SDV: an open source library for synthetic data generation[D]. Cambridge: Massachusetts Institute of Technology, 2018.
- [26] Zhou G, Soufan O, Ewald J, et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis[J]. *Nucleic Acids Res*, 2019, 47(W1): W234-W241.
- [27] Baseri M, Radmand F, Hamed R, et al. Immunological aspects of dental implant rejection[J]. *Biomed Res Int*, 2020, 2020: 7279509.

- [28] Fretwurst T, Garaicoa-Pazmino C, Nelson K, et al. Characterization of macrophages infiltrating peri-implantitis lesions[J]. *Clin Oral Implants Res*, 2020, 31(3): 274-281.
- [29] Li Y, Ling J, Jiang Q. Inflammasomes in alveolar bone loss[J]. *Front Immunol*, 2021, 12: 691013.
- [30] Kensara A, Hefni E, Williams MA, et al. Microbiological profile and human immune response associated with peri-implantitis: a systematic review[J]. *J Prosthodont*, 2021, 30(3): 210-234.
- [31] Deng S, Hu Y, Zhou J, et al. TLR4 mediates alveolar bone resorption in experimental peri-implantitis through regulation of CD45+ cell infiltration, RANKL/OPG ratio, and inflammatory cytokine production[J]. *J Periodontol*, 2020, 91(5): 671-682.
- [32] Pan K, Hu Y, Wang Y, et al. RANKL blockade alleviates peri-implant bone loss and is enhanced by anti-inflammatory microRNA-146a through TLR2/4 signaling[J]. *Int J Implant Dent*, 2020, 6(1): 15.
- [33] Zhang Q, Liu J, Ma L, et al. Wnt5a is involved in LOX-1 and TLR4 induced host inflammatory response in peri-implantitis[J]. *J Periodontal Res*, 2020, 55(2): 199-208.
- [34] Vacchini A, Mortier A, Proost P, et al. Differential effects of posttranslational modifications of CXCL8/interleukin-8 on CXCR1 and CXCR2 internalization and signaling properties[J]. *Int J Mol Sci*, 2018, 19(12): 3768.
- [35] Gabellini C, Trisciuglio D, Desideri M, et al. Functional activity of CXCL8 receptors, CXCR1 and CXCR2, on human malignant melanoma progression[J]. *Eur J Cancer*, 2009, 45(14): 2618-2627.
- [36] Tong H, Ke JQ, Jiang FZ, et al. Tumor-associated macrophage-derived CXCL8 could induce ER $\alpha$  suppression via HOXB13 in endometrial cancer[J]. *Cancer Lett*, 2016, 376(1): 127-136.
- [37] Zhang X, Wang Z, Hu L, et al. Identification of potential genetic biomarkers and target genes of peri-implantitis using bioinformatics tools[J]. *Biomed Res Int*, 2021, 2021: 1759214.
- [38] Aleksandrowicz P, Brzezińska-Błaszczyk E, Kozłowska E, et al. Analysis of IL-1 $\beta$ , CXCL8, and TNF- $\alpha$  levels in the crevicular fluid of patients with periodontitis or healthy implants[J]. *BMC Oral Health*, 2021, 21(1): 120.
- [39] Glaría E, Valledor AF. Roles of CD38 in the immune response to infection[J]. *Cells*, 2020, 9(1): 228.
- [40] Chen D, Wu X, Liu Q, et al. Memory B cell as an indicator of peri-implantitis status: a pilot study[J]. *Int J Oral Maxillofac Implants*, 2021, 36(1): 86-93.
- [41] Munde EO, Okeyo WA, Raballah E, et al. Association between Fc $\gamma$  receptor II A, III A and III B genetic polymorphisms and susceptibility to severe malaria anemia in children in western Kenya[J]. *BMC Infect Dis*, 2017, 17(1): 289.
- [42] Elmer BM, Swanson KA, Bangari DS, et al. Gene delivery of a modified antibody to A $\beta$  reduces progression of murine Alzheimer's disease[J]. *PLoS One*, 2019, 14(12): e0226245.
- [43] Saremi L, Esmacilzadeh E, Ghorashi T, et al. Association of Fc gamma-receptor genes polymorphisms with chronic periodontitis and Peri-Implantitis[J]. *J Cell Biochem*, 2019, 120(7): 12010-12017.
- [44] Taylor JP, Tse HM. The role of NADPH oxidases in infectious and inflammatory diseases[J]. *Redox Biol*, 2021, 48: 102159.
- [45] Denson LA, Jurickova I, Karns R, et al. Clinical and genomic correlates of neutrophil reactive oxygen species production in pediatric patients with Crohn's disease[J]. *Gastroenterology*, 2018, 154(8): 2097-2110.
- [46] Ivetic A. A head-to-tail view of L-selectin and its impact on neutrophil behaviour[J]. *Cell Tissue Res*, 2018, 371(3): 437-453.
- [47] Siddiqui K, George TP, Mujammami M, et al. The association of cell adhesion molecules and selectins (VCAM-1, ICAM-1, E-selectin, L-selectin, and P-selectin) with microvascular complications in patients with type 2 diabetes: a follow-up study[J]. *Front Endocrinol (Lausanne)*, 2023, 14: 1072288.
- [48] Xia S, Qu J, Jia H, et al. Overexpression of Forkhead box C1 attenuates oxidative stress, inflammation and apoptosis in chronic obstructive pulmonary disease[J]. *Life Sci*, 2019, 216: 75-84.
- [49] Wang J, Li W, Zheng X, et al. Research progress on the forkhead box C1[J]. *Oncotarget*, 2017, 9(15): 12471-12478.
- [50] Ji Z, Chen S, Cui J, et al. Oct4-dependent FoxC1 activation improves the survival and neovascularization of mesenchymal stem cells under myocardial ischemia[J]. *Stem Cell Res Ther*, 2021, 12(1): 483.
- [51] Aktar A, Heit B. Role of the pioneer transcription factor GATA2 in health and disease[J]. *J Mol Med (Berl)*, 2023, 101(10): 1191-1208.
- [52] Takai J, Shimada T, Nakamura T, et al. Gata2 heterozy-

- gous mutant mice exhibit reduced inflammatory responses and impaired bacterial clearance[J]. *iScience*, 2021, 24(8): 102836.
- [53] Xie M, Li Z, Li X, et al. Identifying crucial biomarkers in peripheral blood of schizophrenia and screening therapeutic agents by comprehensive bioinformatics analysis [J]. *J Psychiatr Res*, 2022, 152: 86-96.
- [54] Mitchell JP, Carmody RJ. NF- $\kappa$ B and the Transcriptional Control of Inflammation[J]. *Int Rev Cell Mol Biol*, 2018, 335: 41-84.
- [55] Cao N, Liu X, Hou Y, et al. 18- $\alpha$ -glycyrrhetic acid alleviates oxidative damage in periodontal tissue by modulating the interaction of Cx43 and JNK/NF- $\kappa$ B pathways [J]. *Front Pharmacol*, 2023, 14: 1221053.
- [56] Teng H, Chen S, Fan K, et al. Dexamethasone liposomes alleviate osteoarthritis in miR-204/-211-deficient mice by repolarizing synovial macrophages to M2 phenotypes [J]. *Mol Pharm*, 2023, 20(8): 3843-3853.
- [57] Zhang N, Zhang RF, Zhang AN, et al. MiR-204 promotes fracture healing via enhancing cell viability of osteoblasts[J]. *Eur Rev Med Pharmacol Sci*, 2018, 22(1 Suppl): 29-35.
- [58] Li N, Guo X, Liu L, et al. Molecular mechanism of miR-204 regulates proliferation, apoptosis and autophagy of cervical cancer cells by targeting ATF2[J]. *Artif Cells Nanomed Biotechnol*, 2019, 47(1): 2529-2535.
- [59] Grieco FA, Schiavo AA, Brozzi F, et al. The miRNAs miR-211-5p and miR-204-5p modulate ER stress in human beta cells[J]. *J Mol Endocrinol*, 2019, 63(2): 139-149.
- [60] Ye G, Wang P, Xie Z, et al. IRF2-mediated upregulation of lncRNA HHAS1 facilitates the osteogenic differentiation of bone marrow-derived mesenchymal stem cells by acting as a competing endogenous RNA[J]. *Clin Transl Med*, 2021, 11(6): e429.
- [61] Rigatti SJ. Random forest[J]. *J Insur Med*, 2017, 47(1): 31-39.
- [62] Carleo G, Troyer M. Solving the quantum many-body problem with artificial neural networks[J]. *Science*, 2017, 355(6325): 602-606.
- [63] Arji G, Safdari R, Rezaeizadeh H, et al. A systematic literature review and classification of knowledge discovery in traditional medicine[J]. *Comput Methods Programs Biomed*, 2019, 168: 39-57.

(本文编辑 杜冰)

## 《口腔美学区种植临床精要》出版发行

书籍名称：口腔美学区种植临床精要

主编：付钢、黄元丁

出版日期：2023年11月

出版社：人民卫生出版社

内容简介：本书由重庆医科大学口腔医学院付钢和黄元丁两位长期在临床一线工作的教授联合主编，将其团队20年的临床经验和研究成果通过认真总结编写成书。主要内容包括：1) 前牙美学区种植的特点；2) 前牙美学区的解剖及生理；3) 以患者为中心的前牙区美学种植原则；4) 前牙美学区的术前评估和方案设计；5) 前牙美学区的外科基本技巧；6) 维持理念下的即刻种植；7) 重构理念下的前牙美学区骨增量技术；8) 前牙美学区的软组织增量及管理；9) 前牙美学区的修复技术。

