

## 基于BP神经网络预测儿童甲状腺疾病的模型研究

田娟<sup>1</sup>, 朱姝婧<sup>2</sup>, 陆强<sup>1</sup>, 李坤<sup>1</sup>, 张西学<sup>1</sup>

1. 山东第一医科大学医学信息工程学院, 山东 泰安 271016; 2. 泰安市疾病预防控制中心检验科, 山东 泰安 271016

**【摘要】目的:**构建儿童甲状腺疾病的预测模型。**方法:**根据某市疾病预防控制中心2013年~2016年采集的1 400名8~11岁儿童的体检数据及临床初步诊断结果作为研究数据,随机抽取其中的1 000名儿童作为训练样本,剩余的400名儿童作为测试样本,利用MATLAB R2018b软件编程实现三层BP神经网络模型。**结果:**当选择log&log组合作为隐含层和输出层的传递函数,隐含层节点数目选择8时,模型的分类正确率达到91.43%。**结论:**BP神经网络应用于儿童甲状腺疾病的预测,可以为疾病的防治工作提供理论依据。

**【关键词】**儿童甲状腺疾病;BP神经网络;疾病预测;分类正确率

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2020)10-1340-05

## Prediction model of pediatric thyroid disease based on back-propagation neural network

TIAN Juan<sup>1</sup>, ZHU Shujing<sup>2</sup>, LU Qiang<sup>1</sup>, LI Kun<sup>1</sup>, ZHANG Xixue<sup>1</sup>

1. School of Medical Information Engineering, Shandong First Medical University, Tai'an 271016, China; 2. Department of Laboratory, Tai'an Center for Diseases Prevention and Control, Tai'an 271016, China

**Abstract: Objective** To construct a prediction model of thyroid disease in pediatric patients. **Methods** The physical examination data and preliminary clinical diagnoses of 1 400 children aged 8-11 years were collected from 2013 to 2016 in a Center for Disease Prevention and Control. One thousand out of 1 400 children were randomly selected as training samples, and the remaining 400 were used as test samples. A three-layer back-propagation neural network model was constructed by MATLAB R2018b software. **Results** The classification accuracy of the model reached 91.43% when the combination of log&log was selected as the transfer function of the hidden layer and the output layer, with 8 nodes in the hidden layer. **Conclusion** Back-propagation neural network can be applied to the prediction of thyroid diseases in children, and provide a theoretical basis for the prevention and treatment of diseases.

**Keywords:** pediatric thyroid disease; back-propagation neural network; disease prediction; classification accuracy

### 前言

甲状腺疾病在临床是一种由多种因素导致的一类累及甲状腺的异质性疾病,其病因的不同,促使患者的临床表现和预后差异也相对较大,严重者最终发展为甲状腺癌,对患者的生命健康造成严重的威胁<sup>[1]</sup>。因此,早期发现甲状腺疾病、准确定性并予以有效治疗十分必要<sup>[2]</sup>。近些年,甲状腺疾病甚至甲状腺肿瘤的发病率逐年攀升,患病人数之多出乎了很多医生和健康专家的预测<sup>[3]</sup>。越来越多的人开始关注甲状腺健康的问题,甲状腺这个小小的器官变成了一个社会问题,连带着碘盐和海鲜也成了大家

关注的饮食热点。

医疗数字化技术、人工智能和大数据技术的发展,医疗模式逐渐从以治疗为主转变为以预防为主<sup>[4]</sup>。人们开始尝试从不同的角度结合多种数学模型进行疾病的预测,目的是发现疾病的潜在风险和趋势,对于疾病的预防、干预和管理具有重要作用。

人工神经网络(Artificial Neural Network, ANN)是一门近几年迅速发展起来的集神经科学、信息科学及计算机科学于一体的交叉学科,因应用简单、参数选择方便、分类结果准确性较高等优点被广泛应用。其中误差反向传播网络(Back-Propagation Network, 简称BP神经网络),是一种按误差逆传播算法训练的多层前馈网络,具有自适应性强、容错性高、组织能力强等优点<sup>[5]</sup>,在疾病预测方面具有广泛的应用空间,可以将其应用于儿童甲状腺疾病的预测和治疗过程。

**【收稿日期】**2020-04-11

**【基金项目】**泰安市科技发展规划(201730338)

**【作者简介】**田娟,讲师,硕士,研究方向:智能控制和模式识别, E-mail: jtian@sdfmu.edu.cn

基于此,本研究建立BP神经网络模型进行儿童甲状腺疾病预测。该研究不仅可以提高甲状腺疾病预测的准确率,为甲状腺疾病的诊断和治疗节省宝贵的时间,而且将其进行推广,可以扩大所能预测疾病的范围,进而对改善现代社会人们的健康状况和促进医学领域的智能化发展有参考价值。

1 基于BP神经网络预测甲状腺疾病

1.1 BP神经网络基础理论

BP神经网络由输入层、隐含层、输出层构成,通常使用两种非线性激活函数,一种称为Sigmoid函数或者Logistics函数,一种称为双曲正切函数。本研究选取Sigmoid函数作为激活函数,记作:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

设输入层有n个神经元,隐含层有p个神经元,输出层有q个神经元,定义如下:

输入变量为:

$$x_i = (x_1, x_2, \cdots, x_n), \quad i = 1, 2, \cdots, n \tag{2}$$

隐含层输出变量为:

$$h_j = (h_1, h_2, \cdots, h_p) = f(\sum_{i=1}^n w_{ij}x_i + a_j), j = 1, 2, \cdots, p \tag{3}$$

输出层实际输出变量为:

$$y_k = (y_1, y_2, \cdots, y_q) = \sum_{j=1}^p h_jw_{jk} + b_k, k = 1, 2, \cdots, q \tag{4}$$

输出层期望输出变量为:

$$d_k = (d_1, d_2, \cdots, d_q), \quad k = 1, 2, \cdots, q \tag{5}$$

误差函数为:

$$E = \frac{1}{2} \sum_{k=1}^q (d_k - y_k)^2 \tag{6}$$

设有m个训练样本,则误差函数为:

$$E = \frac{1}{2m} \sum_{l=1}^m \sum_{k=1}^q (d_k(l) - y_k(l))^2 \tag{7}$$

权值的更新公式为:

$$\begin{cases} w_{ij} = w_{ij} + \eta h_j(1 - h_j)x_i \sum_{k=1}^q w_{jk}(d_k - y_k) \\ w_{jk} = w_{jk} + \eta h_j(d_k - y_k) \end{cases} \tag{8}$$

偏置的更新公式为:

$$\begin{cases} a_j = a_j + \eta h_j(1 - h_j) \sum_{k=1}^q w_{jk}(d_k - y_k) \\ b_k = b_k + \eta (d_k - y_k) \end{cases} \tag{9}$$

其中,输入层与隐含层的权值为 $w_{ij}$ ,偏置为 $a_j$ ;隐含层与输出层的权值为 $w_{jk}$ ,偏置为 $b_k$ ,学习速率为 $\eta$ 。

1.2 甲状腺疾病预测模型研究

1.2.1 实验数据来源 甲状腺疾病种类繁多,且在大众的传统印象中,易患甲状腺疾病的都是成年人。但是近年来,儿童甲状腺疾病发病率越来越高。因此本研究以某市疾病预防控制中心2013年~2016年采集的1 400名8~11岁儿童的体检数据及临床初步诊断结果作为研究数据,随机抽取其中的1 000名儿童作为训练样本,剩余的400名儿童作为测试样本。其中部分体检数据如表1所示。

研究数据中的初步诊断结果是专科医生根据人群碘营养状况评价标准和《地方性甲状腺肿诊断标

表1 2013~2016年部分儿童体检数据表  
Tab.1 Physical examination data of some children in 2013-2016

性别	年龄/岁	体质量/kg	甲状腺大小/mm						甲状腺容积/mL	尿碘检测/ $\mu\text{g}\cdot\text{L}^{-1}$	盐碘检测/ $\text{mg}\cdot\text{kg}^{-1}$
			左长	左宽	左厚	右长	右宽	右厚			
男	8	41.0	32.70	13.10	8.32	32.20	12.90	9.09	3.515 784 605	234.47	26.80
男	8	31.5	33.40	14.20	10.10	34.20	11.10	7.13	3.591 024 009	361.95	26.10
男	9	58.0	34.20	12.20	10.20	35.10	11.30	11.40	4.204 388 970	377.37	26.80
男	9	42.0	35.00	12.30	7.87	34.20	10.30	10.30	3.360 813 927	329.41	27.30
男	9	72.0	37.10	14.20	10.70	37.10	14.50	10.90	5.508 801 290	263.64	25.60
男	10	42.0	26.60	11.30	9.71	33.00	11.60	10.80	3.378 325 592	241.98	29.80
女	10	30.0	33.70	8.20	8.32	34.20	10.50	6.78	2.267 512 617	17.07	28.10
女	10	25.0	26.00	9.70	9.41	32.20	11.10	12.00	3.191 213 918	281.27	29.30
女	9	49.5	33.50	13.90	8.94	35.90	14.00	9.86	4.367 784 013	239.29	23.90
女	9	42.0	33.00	11.40	12.20	32.00	10.00	9.40	3.639 269 560	422.18	17.77
女	9	46.0	33.30	9.86	80.60	34.50	8.93	8.01	13.858 333 900	282.08	27.70
女	8	31.5	34.00	9.25	8.47	33.60	12.20	9.24	3.090 256 908	440.22	23.80

准》<sup>[6]</sup>等规定做出的初步诊断,共分为6种不同的情况。因为样本集具有较强的不平衡性,为神经网络的训练和预测带来了更多的不确定性,所以本研究为方便后续研究为每种情况设定了一个标识,标注了每种情况包含的样本数量,如表2所示。

表2 初步诊断结果标识  
Tab.2 Identifications of preliminary diagnoses

标识	说明	样本数量
1	高碘性甲状腺肿	7
2	内源性甲状腺肿	6
3	缺碘性甲状腺肿	2
4	碘缺乏	243
5	正常	452
6	碘过量	690

1.2.2 数据归一化处理 本研究选用最大最小值法进行数据的归一化处理,如式(10)所示,旨在把变量值归一到区间[0, 1]内,如表3所示。

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(10)

其中, $x$ 是预测数据, $x_{\min}$ 和 $x_{\max}$ 分别是预测数据中的最小值与最大值。

表3 甲状腺疾病影响因素及量化赋值说明  
Tab.3 Influencing factors and quantitative assignments of thyroid diseases

编号	变量说明	量化赋值
X1	性别	0=男,1=女
X2	年龄	归一化处理
X3	甲状腺容积	归一化处理
X4	尿碘中位数	归一化处理
X5	盐碘检测	归一化处理
Y	甲状腺是否正常	Y=1 碘缺乏, Y=2 正常, Y=3 碘过量, Y=4 缺碘性甲状腺肿, Y=5 内源性 甲状腺肿, Y=6 高碘性甲状腺肿

1.2.3 疾病预测模型设计 本研究设计三层BP神经网络,其中输入层神经元个数为5,代表甲状腺检测的5个体检指标,输出层神经元个数为1,代表Y的不同取值。隐含层神经元个数根据式(11)实验获得:

$$n_1 = \sqrt{n + m} + a$$

(11)

其中, $n$ 为输入层的神经元个数, $m$ 为输出层的神经元个数, $a$ 为[1, 10]之间的常数。根据式(11)可得,隐

含层神经元个数在3~12内取值。  
1.2.4 疾病预测模型训练 训练过程如图1所示,具体如下:①确定好训练数据,对数据进行归一化处理,对权值和阈值进行初始化设置;②输入训练样本,分别计算隐含层和输出层的实际输出;③将实际输出与期望输出进行比较,计算误差;④根据式(8)、式(9)调整权值,直到权值达到稳定,使误差最小;⑤固定当前权值,训练结束。

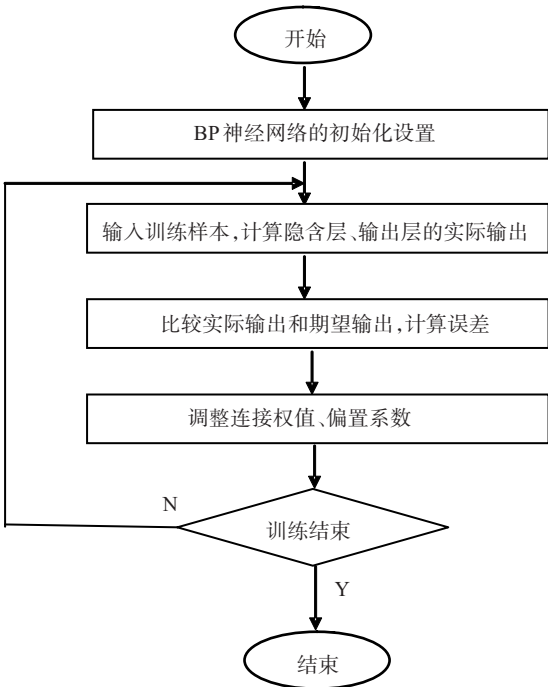


图1 神经网络的训练过程图  
Fig.1 Neural network training process

2 结果

本研究利用MATLAB R2018b软件编程实现训练样本的BP神经网络模型,用测试样本加以验证。为了检验模型的预测效果,本研究采用分类正确率进行检验,分类正确率越高,说明算法的性能越好,疾病的预测能力越好<sup>[7]</sup>。模型中设置的最大迭代次数为2 500次,当迭代次数达到2 500时,BP神经网络训练终止。

2.1 传递函数的确定

网络的训练过程中,保持其他参数值不变,选择不同的传递函数组合进行实验,观察不同函数组合对应的BP神经网络分类正确率,结果如图2所示。本研究选择函数log&log组合作为隐含层和输出层的传递函数。

2.2 隐含层神经元个数的确定

学习速率设为0.001,训练函数选择trainlm,传递函数选择log&log组合,训练目标设定为1e-17,在

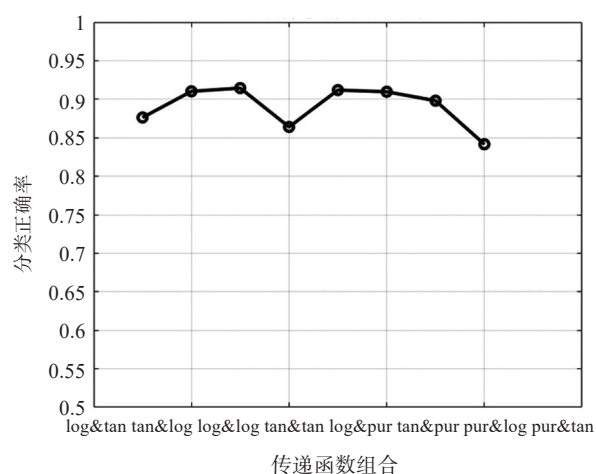


图2 不同传递函数组合下的分类正确率  
Fig.2 Classification accuracy under the combinations of different transfer functions

3~12 之间依次改变网络的隐含层神经元个数,每次重复 100 次实验,对实验结果取平均值和标准差,结果如图 3 所示。根据图 3,综合考虑各因素,本文最终选择隐含层神经元个数为 8。

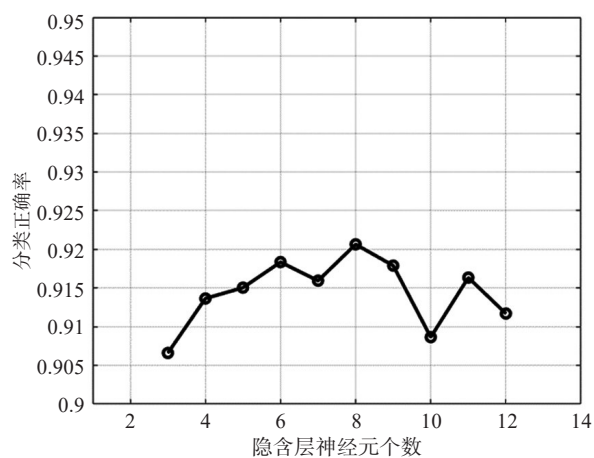


图3 不同隐含层神经元个数的分类正确率  
Fig.3 Classification accuracy with different number of neurons in hidden layers

### 2.3 测试样本的分类

训练任务完成后,将 400 个测试样本输入训练好的 BP 神经网络,重复之前的 100 次实验,并与实际的甲状腺异常病例进行比较,平均准确率达到 91.43%,标准差为 3.01%。

## 3 讨论与结论

疾病预测是从上个世纪 60 年代发展起来的一门学科,目的是为了病人和医生能够提前了解疾病的状况,并且为治疗疾病争取更多的时间。目前存在的预测方法有很多种,比如应用较多的时间序列

预测法、回归预测法、Markov 预测法等,但是各种方法都存在一定的优势和不足之处。随着社会的发展和医疗设备的现代化和智能化,人们对疾病预测学的研究也在不断的深入和扩展,人们开始尝试从不同的角度结合多种数学模型进行疾病预测。

传统的疾病预测对心脏病、高血压等疾病的预测研究较多<sup>[8-16]</sup>,但是针对甲状腺疾病尤其是儿童甲状腺疾病的预测研究较少。文献[17]对甲状腺特异性自身免疫性疾病-Graves 病(简称 GD)展开研究,以单核苷酸多态性作为标记,通过计算 20 种已知的 GD 药敏变量的加权遗传风险评分(记作 wGRS)来预测普通人群中 GD 风险,该 wGRS 预测模型虽然稳定和方便,但预测性能不高。文献[18]对新生儿的甲状腺功能和常见疾病展开研究,认为孕妇甲状腺的状况会影响胎儿的甲状腺功能,孕妇甲状腺的功能障碍可能会对胎儿和新生儿产生重大影响,还有一些新生儿因素会影响甲状腺功能。文献[19]使用非线性混合效应方法在 Graves 病患者中建立甲状腺吸收和消除放射性碘的人群生物动力学模型。

本研究根据 8~11 岁体检儿童的一些体检数据,构建儿童甲状腺疾病的 BP 神经网络预测模型。实验仿真结果说明,该模型误差率较小,准确度达到 91.43%,对儿童甲状腺疾病的预测效果较好。

### 【参考文献】

[1] 朱全美. 试析甲状腺结节在体检人群中的检出情况和相关因素[J]. 中国卫生标准管理, 2017, 8(22): 6-8.  
ZHU Q M. Analysis of the detection and related factors of thyroid nodules in the physical examination population[J]. China Health Standards Management, 2017, 8(22): 6-8.

[2] 陆俐. 体检人群甲状腺超声检查的临床价值[J]. 临床医药文献杂志, 2019, 6(11): 131-134.  
LU L. Clinical value of thyroid ultrasonography in physical examination population[J]. Journal of Clinical Medicine Literature, 2019, 6(11): 131-134.

[3] 蒋宁一. 甲状腺-形似蝴蝶实为盾甲[J]. 养生大世界, 2017, 9(9): 52-53.  
JIANG N Y. Thyroid-A shield shaped like a butterfly[J]. Health World, 2017, 9(9): 52-53.

[4] 章晓庆. 基于深度学习的多疾病风险预测模型研究[D]. 郑州: 郑州大学, 2019.  
ZHANG X Q. The study on multi-disease risk prediction model based on deep learning[D]. Zhengzhou: Zhengzhou University, 2019.

[5] 刘灿, 黄俊, 胡丹, 等. 基于慢性病预测的老年人健康监护软件设计与实现[J]. 信息通信, 2019, 2(2): 93-96.  
LIU C, HUANG J, HU D, et al. Elderly health monitoring software based on chronic disease prediction design and implementation[J]. Information & Communications, 2019, 2(2): 93-96.

[6] 中华人民共和国卫生部. WS276-2007 地方性甲状腺肿诊断标准[S]. 北京: 人民卫生出版社, 2007.



- Ministry of Health PR CHINA. WS276-2007 diagnostic criteria for endemic goiter [S]. Beijing: People's Medical Publishing House, 2007.
- [7] 陈金宏, 吴海云, 何耀, 等. 基于BP神经网络的老年男性保健人群缺血性心脑血管病预测模型研究[J]. 第三军医大学学报, 2011, 33(8): 797-799.
- CHEN J H, WU H Y, HE Y, et al. Construction and evaluation of predictive model for ischemic cardiovascular diseases of senior men based on BP neural network [J]. Journal of Third Military Medical University, 2011, 33(8): 797-799.
- [8] LIANG Y R, LI Q, CHEN P S, et al. Comparative study of back propagation artificial neural networks and logistic regression model in predicting poor prognosis after acute ischemic stroke[J]. Open Med, 2019, 14(1): 324-330.
- [9] 李万莎. 应用人工神经网络分析冠心病和脑梗塞相关因素及临床意义[D]. 北京: 北京协和医学院, 2011.
- LI W S. Prediction of coronary heart disease and cerebral infraction and evaluation of relative risk factors using artificial neural network [D]. Beijing: Chinese Academy of Medical Sciences & Peking Union Medical College, 2011.
- [10] 朱悦. SVM在冠心病分类预测中的应用研究[D]. 广州: 华南理工大学, 2013.
- ZHU Y. Application study of SVM in prediction of coronary heart disease [D]. Guangzhou: South China University of Technology, 2013.
- [11] 鹿显涛. 基于BP神经网络的心脏病预测研究与实现[D]. 长春: 吉林大学, 2012.
- PANG X T. Research and implementation on heart disease prediction based on BP neural network [D]. Changchun: Jilin University, 2012.
- [12] 万琦, 王威, 黄薇, 等. 基于武汉市某社区居民电子健康档案的高血压预测模型[J]. 现代预防医学, 2018, 45(6): 1030-1033.
- WAN Q, WANG W, HUANG W, et al. Prediction model of hypertension based on electronic health records of a community resident in Wuhan [J]. Modern Preventive Medicine, 2018, 45(6): 1030-1033.
- [13] 温宗良, 岳桂华, 杨靖, 等. 基于共轭梯度算法的BP神经网络在高血压证候诊断中的应用[J]. 山东中医药大学学报, 2012, 36(3): 183-184.
- WEN Z L, YUE G H, YANG J, et al. Application of BP neural network based on conjugate gradient algorithm in diagnosis of hypertension syndrome [J]. Journal of Shandong University of Traditional Chinese Medicine, 2012, 36(3): 183-184.
- [14] 岳桂华, 温宗良, 杨靖, 等. 基于BP神经网络的高血压病中医证候与分级关系建模研究[J]. 中国中医基础医学杂志, 2012, 18(3): 326-327.
- YUE G H, WEN Z L, YANG J, et al. Research on relationship between the TCM syndrome and grade of hypertension [J]. Chinese Journal of Basic Medicine in Traditional Chinese Medicine, 2012, 18(3): 326-327.
- [15] 王智, 刘薇, 兰燕. 基于BP神经网络的脑血管疾病院内感染风险预测模型研究[J]. 重庆医学, 2017, 46(12): 1691-1693.
- WANG Z, LIU W, LAN Y. Study on risk prediction model of nosocomial infection in cerebrovascular disease based on BP neural network [J]. Chongqing Medicine, 2017, 46(12): 1691-1693.
- [16] 张文茜, 苏海霞, 尚磊, 等. 基于BP神经网络和RBF神经网络预测老年痴呆症疾病进展的对比研究[J]. 现代生物医学进展, 2017, 17(4): 738-741.
- ZHANG W Q, SU H X, SHANG L, et al. A comparative study on prediction of Alzheimer's disease progression based on BP and RBF neural network [J]. Progress in Modern Biomedicine, 2017, 17(4): 738-741.
- [17] MA Y R, ZHAO S X, LU L, et al. A weighted genetic risk score using known susceptibility variants to predict graves disease risk [J]. J Clin Endocrinol Metab, 2019, 104(6): 2121-2130.
- [18] CHAPMAN A K, FARMER Z J. Neonatal thyroid function and disorders [J]. Clin Obstet Gynecol, 2019, 62(2): 373-387.
- [19] PEREZ J M, SALAS A O. Improving individualised dosimetry in radioiodine therapy for hyperthyroidism using population biokinetic modelling [J]. Phys Med, 2019, 62: 33-40.
- (编辑: 陈丽霞)