

皮肤病易发性的空气环境因素相关性 智能分析模型及应用

王伟华

(北京交通大学医院医学外科,北京 100044)

摘要: 皮肤病的发生和外界诱发因素有一定相关性,其中空气质量因素是重要因素之一。通过对一定时间段内的门诊病例分析,以痤疮为代表划分了门诊病例中皮肤病易发性的程度等级,结合同期的空气质量指标的变化情况,应用人工神经网络方法,建立了分析痤疮皮肤病和空气质量指标相关性的预测模型,探讨了皮肤病风险预测的新的途径。

关键词: 皮肤病;易发性;空气环境因素;神经网络方法;分析模型

中图分类号:R751

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2021.01.008

文章编号:1006-1959(2021)01-0027-04

The Intelligent Analysis Model and Application of the Correlation of Air Environmental Factors on the Susceptibility of Skin Diseases

WANG Wei-hua

(Medical Surgery,Beijing Jiaotong University Hospital,Beijing 100044,China)

Abstract: The occurrence of skin diseases has a certain correlation with external inducing factors, among which air quality factors are one of the important factors. Through the analysis of outpatient cases in a certain period of time, acne is used as the representative to classify the degree of susceptibility to skin diseases in outpatient cases, combined with the changes in air quality indicators during the same period, and the artificial neural network method is used to establish an analysis of acne skin The prediction model of the correlation between disease and air quality indicators explored new ways of skin disease risk prediction.

Key words: Skin disease;Susceptibility;Air environmental factors;Neural network method;Analysis model

因外界环境因素的影响不同,各类型皮肤病的发病风险也差异较大^[1,2],另外,空气环境因素对皮肤病的诱发作用本身也不易准确衡量,各相关因素的作用程度不明确,对于这类复杂相关性问题的处理需要能够考虑复杂因素的非线性综合分析方法。人工神经网络方法是从信息处理角度对人脑神经网络进行抽象,建立模型,对复杂多因素问题进行综合分析的人工智能方法,近年来,人工神经网络的研究工作不断深入,已经取得了很大的进展,在自动控制、生物、医学、经济等领域已成功地解决了许多现代计算机难以解决的实际问题,表现出了良好的智能特性^[3-6]。本研究应用人工神经网络方法,建立了分析痤疮皮肤病和空气质量指标相关性的预测模型,旨在探讨皮肤病风险预测的新的途径。

1 神经网络方法概述

人工神经网络是以工程技术手段来模拟人脑神经网络的结构和特征的系统。利用人工神经网络可以构成各种不同拓扑结构的神经网络,是生物神经网络的一种模拟和近似。神经网络的主要连接形式主要有前馈型和反馈型神经网络。常用的前馈型有感知器神经网络、BP神经网络,常用的反馈型有Hopfield网络。BP神经网络,即误差反向传播算法,是一种按误差逆传播算法训练的多层前馈网络,也是目前应用最广泛的神经网络模型之一。BP神经网络模型拓扑结构包括输入层、隐含层和输出层,其中

隐含层可以是一层也可以是多层,见图 1。

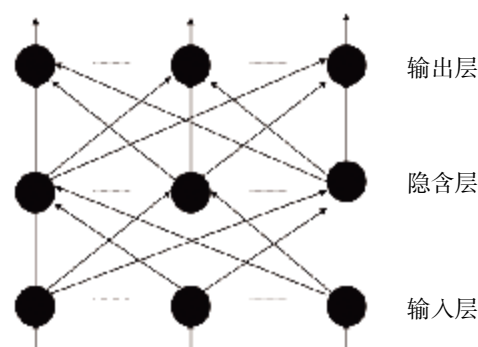


图 1 三层神经网络结构图

任何从输入到输出的连续映射函数都可以用一个三层的非线性网络实现 BP 算法由数据流的前向计算(正向传播)和误差信号的反向传播两个过程构成。正向传播时,传播方向为输入层→隐含层→输出层,每层神经元的状态只影响下一层神经元。若在输出层得不到期望的输出,则转向误差信号的反向传播流程。通过这两个过程的交替进行,在权向量空间执行误差函数梯度下降策略,动态迭代搜索一组权向量,使网络误差函数达到最小值,从而完成信息提取和记忆过程。误差反向传播的过程是权值学习的过程网络权值根据不同的训练模式进行更新。常用的有在线模式和批处理模式。在线模式中的训练样本是逐个处理的,而批处理模式的所有训练样本是成批处理的。人工神经网络分析模型的建立过程,需要确定的参数有:初始权值、阈值、学习率参数、动量参数、隐层节点数等。通过对既有案例数据的学习训

作者简介:王伟华(1968.7-),女,江苏徐州人,本科,副主任医师,主要从事皮肤外科的临床工作

练,实现对各参数取值的确定,BP 学习训练过程如下:①初始化网络,对网络参数及各权系数进行赋值,权值系数取随机数;②输入训练样本,通过个节点间的连接情况正向逐层计算各层的值,并与真实值相比较,得到网络的输入误差;③依据误差反向传播规则,按照梯度下降法调整各层之间的权系数,使整个神经网络的连接权值向误差减小的方向转化;④重复②和③,直到预测误差满足条件或是训练次数达到规定次数。

神经网络算法对于不同的网络需用不同的参数,参数的选择需要根据已知的测试数据进行训练测试,称这个调试参数的过程为试错法。神经网络试错:神经网络需要设置的参数比较多,初始权值、学习率、动量系数、隐层节点数需要依据具体问题具体设置,要分类数目越多,参数调整难度越大。调整过程中需不断观察误差和正确率的变化。无论调整哪个参数,遵循的原则为:如果调小该参数,正确率增大则继续调小该参数,到达某一临界值如果继续调小参数正确率减小,则在该临界值附近寻找最优的该参数的取值^[7-11]。

2 神经网络分析模型

评价预测基本思路,首先用具体时间段的调查数据作为典型单元,将可能影响发病率的的各种气

象因素予以恰当的量化,作为输入层结点的输入值,并依据危险性程度的不同将患病可能性划分为不同等级,量化后作为输出层结点的期望输出。用这些已知样本对网络进行训练,直至网络总误差达到精度要求,即认为网络已掌握了各要素与期望输出之间的关系。然后即可通过神经网络的联想记忆功能直接输出预测结果,对其它时间段的患病风险进行评价预测。

2.1 典型样本数据 以皮肤科门诊初诊患者共 1597 例为研究对象。其中男 872 例,女 725 例为对象进行了皮肤病患病情况的调查,年龄 17~24 岁,平均 19.8 岁。由皮肤科医师进行体检调查后填写表格。内容包括年龄,性别,出生地,卫生习惯,饮食偏好,皮肤性质,皮肤患病情况,诊断标准依据《现代皮肤性病学》,见表 1。

在接诊的患者中调查了 1597 病例,共发现皮肤病 37 种、2374 例次,部分病例有两种或两种以上皮肤病。按临床皮肤病学分类统计常见皮肤病的发生并率为 25.60%。各病种的病例中,以痤疮的发病率最高,患者人数最多,以痤疮为例,分析了疾病的发生和环境空气质量的相关性。痤疮病例在调查期间的时间分布如表 2。

表 1 门诊常见皮肤病调查统计(n,%)

皮肤病类型	男性	占男性患者皮肤病率	女性	占女性患者皮肤病率	合计	占皮肤病总患病率
痤疮	413	28.10	259	29.70	672	34.30
浅部真菌病	119	10.96	78	8.94	197	10.01
荨麻疹	87	8.02	84	9.63	171	8.74
过敏性皮炎	85	7.83	78	8.97	163	8.33
溢脂性皮炎	81	7.45	74	8.48	155	7.92
神经性皮炎	56	5.16	62	7.11	118	6.03
疣	53	4.92	50	5.73	103	5.25
单纯疱疹	31	2.86	37	4.24	68	3.47
接触性皮炎	33	3.04	25	2.86	58	2.96
毛囊炎	32	1.64	24	2.75	56	2.86
湿疹	19	1.75	16	1.84	35	3.48
鸡眼	31	2.95	23	2.63	54	2.76
玫瑰糠疹	19	2.49	21	2.41	40	2.04
银屑病	9	0.83	17	2.05	26	1.33
白癜风	8	0.73	7	0.80	15	0.77
鱼鳞病	5	0.46	4	0.46	9	0.50
疥疮	4	0.37	13	1.49	17	0.87

2.2 同期空气质量数据 根据中国气象台中国空气质量在线监测分析平台公布数据,2016 年 9 月~2018 年 12 月的北京市空气质量指标数据见表 3。

2.3 预测模型 选取空气质量的主要控制指标 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 CO 、 NO_2 、 O_3 作为影响皮肤病发生的外部空

气质量因素的控制指标,分析其对皮肤病发生率的可能影响。选取 $PM_{2.5}$ 、 SO_2 、 CO 、 NO_2 作为主要影响因素。建立预测模型的基本思路为:将可能影响痤疮皮肤病发生的各空气质量指标主要因素 $PM_{2.5}$ 、 SO_2 、 CO 、 NO_2 作为网络输入层各结点的输入值,针对调查分

表 2 痤疮病例随时间变化情况 ($n, \%$)

时间(月份-年度)	9-16	10-16	11-16	12-16	1-17	2-17	3-17	4-17	5-17	6-17	7-17	8-17	9-17	10-17	11-17
病例数	37	34	12	17	23	37	31	15	19	27	22	11	32	33	16
比率	5.5	5.1	1.8	2.5	3.4	5.5	4.6	2.2	2.8	4.0	3.3	1.6	4.8	4.9	2.4

时间(月份-年度)	12-17	1-18	2-18	3-18	4-18	5-18	6-18	7-18	8-18	9-18	10-18	11-18	12-18
病例数	17	23	36	30	21	19	21	26	13	31	21	13	15
比率	2.5	3.4	5.4	4.5	3.1	2.8	3.1	3.9	1.9	4.6	3.1	1.9	2.2

表 3 北京空气质量监测数据

月份-年度	质量等级	PM _{2.5}	PM ₁₀	SO ₂	CO	NO ₂	O ₃
9-16	良	55	73	3	0.76	43	104
10-16	轻度污染	85	101	4	1.155	56	47
11-16	轻度污染	100	134	11	1.797	68	28
12-16	中度污染	133	159	17	2.51	80	28
1-17	轻度污染	116	131	18	2.203	65	41
2-17	良	71	84	18	1.318	53	65
3-17	良	63	84	11	0.913	50	86
4-17	良	53	104	7	0.693	48	101
5-17	轻度污染	48	104	7	0.674	37	161
6-17	轻度污染	42	75	6	0.783	38	184
7-17	轻度污染	52	71	3	0.894	34	174
8-17	良	38	56	3	0.803	35	131
9-17	良	58	100	4	0.913	49	123
10-17	良	57	63	3	0.923	46	47
11-17	良	46	73	5	0.903	49	44
12-17	良	44	68	8	1.09	49	25
1-18	良	34	65	7	0.768	39	48
2-18	良	50	72	9	0.786	31	70
3-18	轻度污染	82	104	9	1.003	53	81
4-18	良	59	108	6	0.673	38	115
5-18	良	45	85	4	0.648	36	134
6-18	轻度污染	43	67	4	0.7	32	168
7-18	良	44	48	3	0.787	24	135
8-18	良	31	46	3	0.745	26	133
9-18	良	28	45	3	0.543	34	93
10-18	良	42	59	4	0.668	45	70
11-18	轻度污染	71	104	7	1.07	60	37
12-18	良	38	73	7	0.835	46	40

注:中国空气质量在线监测分析平台公布数据

析时间区间,重点皮肤病痤疮的发生比率,以调查区间的总的病例数为基础,针对 28 个月总病例数 672 例的情况,将痤疮的发生分为高易发(30 例以上/月)、较高易发(20~30 例/月)、中等易发(10~20 例/月)、低易发(10 例以下/月)4 个等级。并将其作为输出层各结点的期望输出。用这些已知病例实例作为已知样本对网络进行训练,直至网络已掌握数据间的复杂非线性关系为止。然后用待预测的时间段的空气质量指标资料作为预测样本,输入给已学习好

的网络,网络便可通过其联想记忆功能直接输出预测结果。

输出层变量的取值原则采用二值法,皮肤病发生率程度最靠近的等级取 1,其余取 0,即低易发为(1,0,0,0),中等易发为(0,1,0,0),较高易发为(0,0,1,0),高易发为(0,0,0,1)。中间隐含层取 5 个节点,用于提高网络的联想能力。运用上述网络,用表 3 所给出的空气质量数据实例数据对网络进行学习训练,迭代 10000 次后精度满足要求,达到实际输

出与期望输出的误差平方和小于 1.0×10^{-4} 。网络训练成功,得到各个节点之间的连接权值和节点阈值,见表4、表5,从而建立了一个基于实测空气质量数据实例的痤疮皮肤病易发性的判定神经网络模型。

2.4 模型应用 应用上述网络模型,根据表6所给出的目标时段的空气质量样本,对新的日期的空气质量情

表4 输入层与隐含层节点间连接权值与隐含层阈值

输入层节点	隐含层节点				
	1	2	3	4	5
1	-10.26	2.71	-16.88	11.64	4.32
2	2.85	-14.28	7.29	-12.68	11.58
3	9.19	7.81	8.62	1.44	-21.42
4	-31.41	12.68	4.33	-4.38	4.29
隐含层阈值	3.49	2.66	-20.19	15.88	-0.19

况对痤疮皮肤病发生的易发性风险进行了分析,结果表明,基于痤疮皮肤病实例的神经网络判定模型,能够较好地处理影响痤疮皮肤病的各主要因素和易发程度之间的复杂非线性关系,对痤疮皮肤病受空气质量影响的易发程度的判定具有参考价值。

表5 隐含层与输出层节点间连接权值与输出层阈值

隐含层节点	输出层节点			
	1	2	3	4
1	7.69	-16.42	17.88	-9.72
2	-9.23	-8.41	10.68	-8.77
3	-1.99	-6.11	-9.66	15.88
4	13.33	-8.29	-21.01	18.55
5	-0.99	14.72	-14.11	-12.91
输出层阈值	-12.45	-4.89	4.81	-3.21

表6 检测样本分析结果

日期	PM _{2.5}	SO ₂	CO	NO ₂	判定输出	误差	预测程度	易发风险程度
2020-07	41	3	0.671	21	0.000	0.055	0.055	低易发
					0.000	0.039	0.039	
					0.000	0.016	0.016	
					1.000	0.908	-0.905	
2020-08	24	3	0.614	23	0.000	0.049	0.049	低易发
					0.000	0.018	0.018	
					0.000	0.039	0.039	
					1.000	0.940	-0.099	

3 结论

综合考虑可能对皮肤病发生有一定外界影响作用的空气质量指标,建立神经网络预测模型,应用于受空气质量影响的皮肤病易发性程度预测,通过大量实际观测数据的学习、训练,使网络掌握皮肤病易发性程度与各个影响因素之间的映射关系,考虑的影响因素越全面,学习的样本越多,网络记忆的特性也就越多,从而使得网络的预测精度不断提高。神经网络模型的建立和应用可为皮肤病预防提供新的借鉴。

参考文献:

- [1]刘显彤,石春蕊,董继元,等.空气细颗粒物PM_{2.5}浓度与皮炎湿疹类皮肤病日门诊人次的相关性研究[J].临床皮肤科杂志,2018,47(9):559-563.
- [2]甄雅,刘刘玮.环境空气污染与皮肤健康[J].中华皮肤科杂志,2015,48(1):67-70.
- [3]周开利,唐耀红.神经网络模型及其MATLAB仿真程序设计[M].北京:清华大学出版社,2005.
- [4]韩力群.人工神经网络理论设计及应用[M].2版.北京:化学

工业出版社,2007.

- [5]冀军,杜超,钟小钢,等.基于机器学习算法的原发性高血压并发冠心病患病风险研究[J].解放军医学杂志,2020,45(7):735-741.
- [6]曹云,卢毅,陈建新,等.基于机器学习的胃食管反流病中医智能辨证模型的应用[J].北京中医药大学学报,2019,42(10):869-874.
- [7]崔秀明.深度神经网络在阿尔兹海默病预测中的应用分析[J].曲阜师范大学学报,2020,46(1):56-67.
- [8]高凡,高克强,贺川江,等.基于神经网络芯片的阿尔兹海默病体外病理模型及其实时动态分析[J].生物医学工程学杂志,2019,36(6):893-910.
- [9]罗悦,杨威龙,周紫仪,等.基于BP神经网络的咳嗽疾病归经研究[J].时珍国医国药,2019,30(12):3048-3049.
- [10]排孜丽耶·尤山塔依,严传波,木拉提·哈米提等.基于神经网络的新疆高发哈萨克族食管癌医学图像分型研究[J].科技通报,2019,35(1):71-81.
- [11]唐奇伶,刘志梅,符玲玲,等.基于多层级融合网络的糖尿病检测[J].中南民族大学学报,2020,39(4):383-389.

收稿日期:2020-08-09;修回日期:2020-09-25

编辑/宋伟