



医学信息学

高 岚 主编



科学出版社

www.sciencepress.com

医学信息学

高岚 主编

科学出版社

北京

内 容 简 介

医学信息学作为一门发展较快的新兴交叉学科,如今已有 30 多年的发展历史。为使中国的广大读者对该学科有一个较全面的了解,中国协和医科大学、北京大学医学部、四川大学华西医学部、浙江大学生物医学工程系、上海交通大学医学院等单位的专家根据教高函[2002]17 号文编写了本书。

本书既有对国际上医学信息学发展和应用最新成果的介绍,也有对我国医学信息学的发展现状、问题和对策的详细分析。在内容安排上,除对医学信息学的学科性质、任务、研究对象、研究方法和应用成果做概要介绍外,着重从基础层面对该学科中的共性议题,以及从应用层面对该学科应用最活跃的领域进行详细的分析,最后对医学信息学的未来做了展望。

本书的读者对象为医科院校各专业和综合大学生物医学工程系、信息管理系和相关专业学生,医药卫生领域内的广大医、教、研人员以及医疗卫生事业各级管理者,生物医学信息事业的从业人员,此外,对计算机在医学领域应用有兴趣的 IT 工作者也可参考。

图书在版编目(CIP)数据

医学信息学/高岚主编. —北京:科学出版社,2007
ISBN 978-7-03-018938-7

I. 医… II. 高… III. 医学:信息学 IV. R-05

中国版本图书馆 CIP 数据核字(2007)第 065828 号

责任编辑:康 蕾 王 霞/ 责任校对:陈丽珠

责任印制:刘士平 / 封面设计:黄 超

版权所有,违者必究。未经本社许可,数字图书馆不得使用

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2007年6月第 一 版 开本:787×1092 1/16

2007年6月第一次印刷 印张:31 1/2

印数:1—2 000 字数:753 000

定价:79.00 元

(如有印装质量问题,我社负责调换〈科印〉)

《医学信息学》编委

主 编 高 岚

副主编 代 涛

编 委 (按姓氏笔画为序)

王继中 北京市计算中心

尹 源 北京大学医学部

田新平 中国协和医科大学

代 涛 中国协和医科大学

许培扬 中国协和医科大学

李包罗 中国协和医科大学

李秉严 四川大学华西医学部

姚志洪 上海交通大学医学院

高 岚 中国协和医科大学

曹德贤 中国协和医科大学

楼正国 浙江大学

前 言

医学信息学(medical informatics)作为一门发展较快的新兴交叉学科,迄今已有 30 年的发展历史。一般认为,医学信息学是运用信息、计算机、认知科学等的理论、技术和方法,研究医学信息的性质和运动规律,进行医学信息的采集、加工、存储、传输、分析、利用,乃至医学知识的表述,为临床决策提供支持,以改进医疗卫生保健服务,提高医疗卫生保健质量,为广大居民提供经济、优质、高效和安全的卫生保健。

由于医学信息学的发展与以计算机、网络和远程通信技术为代表的信息技术的飞速发展密不可分,所以,医学信息学在发达国家的科研、教学活动中非常活跃,成果颇丰,其技术方法在日常卫生保健实践中也得到广泛应用。20 世纪 80 年代以来,随着中国信息基础设施的不断加强,医学信息学在我国也得到一定的发展。

为向我国的广大读者,尤其是从事卫生保健服务及其科研、教学的人员进一步介绍医学信息学的知识和发展,我们根据教育部教高函[2002]17 号文组织国内相关专家编写了这本《医学信息学》,作为教育部高等教育“十五”国家级教材。全书共分 12 章。第一章为绪论,介绍医学信息学的由来与发展,医学信息学的方法学和主要应用领域,以及国外医学信息学教育和培训的一些情况;第二章介绍医学数据分类、编码系统和医学信息标准;第三章介绍数据库概念与数据库管理技术;第四章介绍信息系统的分析与设计,重点介绍信息系统的基本概念,特别是信息系统的开发方法、系统分析与系统设计技术;第五章介绍医院信息系统;第六章介绍生物信号处理;第七章介绍电子病历;第八章介绍医学成像系统;第九章介绍医学决策方法和临床决策支持系统;第十章介绍医学信息资源检索和利用;第十一章介绍远程医学;第十二章展望了医学信息学的未来。以期通过这样的安排,能让读者对医学信息学的全貌以及各领域的发展有一个清晰的了解。在内容介绍上,既考虑科学性、新颖性,更考虑实用性;既介绍国外的最新发展,也用相当篇幅介绍中国医学信息学的发展和应用现状、面临的问题和应采取的对策。

本书的读者对象是医学院校医疗、护理、卫生、口腔、儿科、中医等专业,以及综合大学生物工程和相关专业的学生,医药卫生领域内的广大医、教、研人员,医疗卫生事业各层次的管理者,对计算机在医学领域应用有兴趣的 IT 工作者也可参考。

由于医学信息学还是一门发展中的学科,涉及医学、计算机软硬件、网络通信、决策科学等各部门学科知识,如何吸取国外的成熟经验编出一本适合既定读者对象的教材,对我们来说是一项新的任务。由于水平有限,疏漏以至错误在所难免,恳请广大读者提出宝贵意见,以便我们今后改进。

高 岚

2007 年 5 月于北京

目 录

前言

第一章 绪论	1
第一节 医学信息学的由来与发展	1
第二节 医学信息学方法和主要应用领域	4
第三节 国外医学信息学的教育和培训	7
第二章 医学数据分类、编码系统和医学信息标准	11
第一节 信息标准化的定义与特点	11
第二节 数据分类与编码的基本原则和方法	15
第三节 医学数据的类型	23
第四节 一些重要的国际医学信息标准	24
第五节 国内医学信息标准研究与制定工作进展	37
第三章 数据库概念与数据库管理技术	44
第一节 概述	44
第二节 数据库的若干基本概念	45
第三节 数据库的类型	48
第四节 数据库系统基础架构与数据库管理系统	53
第五节 SQL 查询语言和数据库应用系统设计要点	57
第六节 关系型数据库的局限性	59
第七节 后关系型数据库的兴起	60
第四章 信息系统的分析与设计	74
第一节 信息系统的基本概念	74
第二节 信息系统分析与设计	80
第三节 医学信息系统建设的特殊问题	109
第五章 医院信息系统	115
第一节 医院信息系统概述	116
第二节 医院信息系统的技术基础	129
第三节 医院信息系统的设计与实现	140
第六章 生物信号处理	147
第一节 生物信号概述	147
第二节 计算机化的心电描记法	150
第三节 病人监护系统中的生物信号处理	159
第七章 基于计算机的病历系统——电子病历	179
第一节 基于计算机的病历与纸质病历的区别	179

第二节	基于计算机的病历的功能构成	181
第三节	基于计算机病历的基本问题	184
第四节	基于计算机的病历面临的挑战	188
第八章	医学成像系统	195
第一节	医学成像基础	195
第二节	数字医学图像通信标准(DICOM)	201
第三节	医学图像传输和存储系统(PACS)	209
第四节	远程放射学	226
第九章	医学决策方法和临床决策支持系统	237
第一节	概率性的医学推断	237
第二节	贝叶斯定理和预测值	246
第三节	期望值决策(expected-value decision-making)	252
第四节	决策后的行动	258
第五节	临床决策支持系统	260
第十章	医学信息资源检索和利用	272
第一节	基本概念和知识	272
第二节	《医学主题词表》及其使用	277
第三节	PubMed 资源及检索	281
第四节	中国生物医学文献数据库	291
第五节	循证医学信息资源检索和利用	300
第六节	国家科技图书文献中心网络服务系统	321
第七节	中国高等教育文献保障系统	327
第八节	其他医学信息资源简介	332
第九节	著名搜索引擎和网站	395
第十一章	远程医学	416
第一节	远程医学概述	416
第二节	远程医学技术	427
第三节	远程医疗应用	443
第四节	远程医学中的法律与道德问题	471
第十二章	医学信息学的未来	478
第一节	生物信息学和医学信息学的协同和合作	478
第二节	对个性化卫生保健提供的支持	482
附录		487
附录 1	HL7 数据类型表概貌	487
附录 2	国家科技图书文献中心面向全国开通的国外网络版期刊	490

第一章

绪论

第一节 医学信息学的由来与发展

一、信息、数据与知识

(一) 信息概念的演变和定义

人类社会已经跨入 21 世纪,进入与农业文明和工业文明时代所不同的信息时代。在信息时代,人类正以前所未有的规模大量产生信息,广泛使用信息,不断创造新的知识。

信息与人类的生产、生活息息相关,因而是当代社会使用最多、最广的词汇之一。然而,对“什么是信息”迄今却尚未有公认的定义。在古代,人们认为信息就是消息,就像在英文中有时将“information”(信息)和“message”(消息)混用一样。只是到了近现代,人们才开始把信息作为科学研究的对象,试图描述信息的特点,给出信息的概念和定义。例如,通信专家申农在定量测定通信系统中的信息时,把信息界定为“用来减少随机不确定性的东西”;计算机和通信技术出现后,信息被看成“数据”,并在计算机和通信科学的许多基础理论中得到广泛应用;第二次世界大战后,随着科技信息服务业的兴起,又出现了信息是“决策所需要的知识”的说法;在互联网飞速发展的今天,人们又将信息看成“网络上传输的一切数据、符号、信号、资料”等。

有人曾经统计过,迄今有关信息的概念不下百种。之所以有这么多种对信息概念的不同解释,与不同的社会发展时期、不同的约束条件有关。

如果不考虑各种约束条件,多数人会同意信息是“一种事物存在的方式和运动状态的表现形式”。这是最普遍、最广义的信息概念,有人将其称之为本体论层次的信息。在这个意义上,信息可与物质和能量并驾齐驱。如果考虑到信息的产生、获取、利用等都离不开人这个主体,那就产生了认识论层次上的信息的概念:主体所感知或认识的事物存在的一种方式

和运动状态。

在本体论层次上,信息在人类出现以前就客观存在;而在认识论层次上则不同,没有主体就不能认识信息。由于人具有感知、理解和判断能力,所以认识论层次上的信息比本体论层次上的信息概念具有更广泛的内涵。事实上,人只有感知了事物的存在方式和它的运动状态,理解了其含义,明确了其效用后,才真正掌握了该事物的信息,因而才能做出正确的决策。所以有学者说,既考虑到事物的存在方式和它的运动状态、又考虑其内在含义和使用价值的信息称为“全信息”,而把只考虑其中的形式因素的信息称为“语法信息”,考虑其内容因素的信息称为“语义信息”,考虑其效用因素的信息称为“语用信息”。

(二) 信息、数据和知识的联系与区别

人们在研究“信息”时,往往离不开“数据”和“知识”这两个概念。“数据”是散在的、无关的,或按一定规律排列组合的事实、数字或符号。数据是潜在的信息。而知识是与用户的能力和经验丰富相结合并用于解决问题或产生新知识的信息。由此我们可以说,数据是信息的原料,而信息则是知识的原料。

本书所讨论的医学信息既包括医疗卫生保健信息或临床信息,又包括医药科技文献信息两大部分。

二、医学信息学的发展历史

(一) 医学信息名词及概念的由来

医学信息学(medical informatics)从20世纪70年代兴起,经历了30多年的发展历程,逐渐形成为一门计算机、信息等科学和医学科学交叉的新兴学科。在20世纪50年代,有关电子数字计算机在医学上的应用的参考文献最早出现在生物物理、生物工程和生物医学电子学的出版物上。随后,一些新名词,如医学计算、医学计算机科学、医学自动数据处理、医学信息科学、医学软件工程等陆续出现在各种出版物上,不过通常交叉使用,甚至出现了“医学计算机科学”与“医学信息科学”互用的情况,将处理对象的信息与如何处理信息的计算机混为一谈。“medical informatics”作为一个专业词汇首次出现,是在1974年出版的信息处理国际联盟医学信息学论文丛书第一卷《卫生保健人员信息学教育》(IFIP Medical Informatics Monograph Series, Volume 1, Education in Informatics of Health Personnel)中。

(二) 关于医学信息学的定义

最早的医学信息学的定义出现在1977年在东京召开的医学信息学第三次世界大会[the third world conference on medical informatics(MEDINFO 80)]上。当时对医学信息学下的定义是:“医学信息学是将计算机技术应用到医学的所有领域——医疗保健、医学教育和医学研究”。当然也有人强调,医学信息学绝不只是计算机在医学上的应用,因为有些研究者把医学计算作为一门学科而不是一种工具。而美国测验和材料协会建立的医学信息学分委会1985年建议,应将医学信息学定义为“在卫生保健和医学研究、教育、实践各领域的计算机和信息科学、工程和技术”。

自20世纪80年代以来,由于以计算机、微电子和现代通信技术为代表的信息技术的飞速发展,不但促进了包括医学在内的生命科学的发展,同时也推动了生物信息学、医学信息学等学科的发展。人们对什么是生物信息学和医学信息学也有了更多的共识。医学信息学是运用信息、计算机、认知科学等的理论、技术和方法,研究医学信息的性质和运动规律,进行医学信息的采集、加工、存储、传输、分析、利用以至于医学知识的表述,为临床决策提供支持,以改进医疗卫生保健服务,提高医疗卫生保健质量,为广大居民提供经济、优质、高效和安全的卫生保健。

而对于生物信息学,美国国立医学图书馆编制的医学主题词表(MeSH)中采用的是计算生物学(computational biology)这一主题词,在其范畴注释中是这样解释的:生物信息学是“有关采集和操作生物学数据的技术开发,并使用这些数据进行生物学发现和预测的一门生物学科。该学科包括应用于分子生物学的所有计算方法和理论,及解决生物学问题(包括操作模型和数据集)的基于计算机技术的各个领域”。

由于有些生物信息学技术、方法及成果已经或将要应用到卫生保健工作中,所以生物信息学和医学信息学在研究领域和研究目的方面会有愈来愈多的共同语言,所以现在又出现了生物医学信息学这一提法。对此,人们既可理解这是与生物信息学和医学信息学不同的学科,也可理解为将生物信息学技术、方法及成果应用到卫生保健工作中。对前一种看法需要慎重,因与对生物信息学和医学信息学的定义、研究对象、研究方法以及应用领域已经有相当多的不同观点,对生物医学信息学的本质还有待更深入的探讨。

三、我国医学信息学的一些发展情况

随着改革开放,特别是在20世纪80年代末90年代初,我国信息基础设施建设的加强,我国医学信息学也得到了一定的发展。我国医学信息学的应用最早在医院进行。计算机在医院的应用始于20世纪的70年代,但真正在医院信息管理上的应用则是在1984年微型机在全国推广以后。其发展大约经历了三个阶段:①孤立应用阶段,如药品、器械和库房的管理;②部门或方面级的应用阶段,如收费系统、医疗统计管理等;③涉及医院各主要业务活动的集成医院信息系统阶段,也有的称临床信息应用阶段。这第三个阶段是目前正在发展的阶段,全国估计有不足5%的医院处于这一阶段。

在公共卫生方面,20世纪80年代以来,很多单病监测系统开始发展,建立了全国疾病监测点监测系统,开始了长期综合疾病监测工作,以后又陆续开展了非传染病监测和其他公共卫生监测。

在一些单位特别是高校、部队开展了医学信息学的基础和应用研究,包括生物信号处理、人工智能、医学成像研究、电子病历、远程医学教育和服务等。经过最近若干年的努力,我国在医学信息资源(主要是科技文献资源)的检索和利用上与国外的差距总体上已大大缩小。近一两年来,有一些高校已开设医学信息学专业课程,进行专门人才的培养。在相关学术活动方面,1981年成立了中国电子学会医药信息学分会CMIA(China Medical Informatics Association),其成员主要是医药卫生领域内从事计算机工作的专业人士,学术活动的主要内容是交流计算机技术和信息科学在医药卫生领域中的应用成果和经验。1994年成立了中华医学会医学信息学会,成员主要由我国医学情报(信息)图书馆的专业人员构成,活动的重点侧重于医学信息(特别是文献信息)检索和利用的科研和实践。

尽管有以上这些方面的进展,但医疗卫生领域内信息基础设施建设与国内其他部门和行业相比还有较大的差距,更不要说与发达国家的差距了。主要表现在相当长的时期内,在政府层面上对医疗卫生信息化建设认识不到位,缺乏整体规划;经费投入不足,系统投资渠道分散,信息系统整体效益难以发挥;卫生机构部门现有的管理体制制约了信息资源的开发和共享;卫生信息标准体系建设滞后等。在2003年SARS流行时,对突发性公共卫生

事件的应急反应能力严重不足使以上问题暴露无疑。

为应对存在的问题,根据国务院的要求,2003年4月由卫生部信息化工作领导小组制订了《全国卫生信息化发展规划纲要2003—2010年》,同年卫生部及时制订了《国家公共卫生信息系统建设方案》,我国法定传染病监测与突发公共卫生事件监测报告信息系统随后于2004年1月开通。为加强卫生信息化标准制订工作,卫生部信息化工作领导小组办公室加强了对卫生信息标准化专业委员会开展卫生信息标准研究、维护等工作的大力支持,分别委托相关单位开展《国家卫生信息标准基础框架》、《我国医院信息标准最小数据集》、《我国公共卫生信息标准最小数据集》和《社区卫生标准》的制订和开发工作。相信以上这些工作进展会大大促进医学信息学在我国医疗卫生领域的应用和发展。

如上所述,医学信息学是一门交叉学科,它的发展、应用需要跨学科、跨部门、跨行业、跨地区的协作,这就要求国家(或其他权威部门或组织)从宏观上加强管理、规划和协调,就学科建设、研究重点、应用领域和队伍培训等方面制定统一规划,加强标准和法规建设。因此,从这个意义上说,卫生部目前抓的几项工作只能算作一个良好的开端,国家层次上的管理决策部门、相关科研机构 and 学术团体、IT行业等在我国医学信息学的学科发展及应用上都还任重道远。

第二节 医学信息学方法和主要应用领域

一、医学信息学的常用研究方法

医学信息学是一门多学科交叉的新兴学科,它的研究方法既有各门学科通用的一般方法,更有医学信息学特有的研究方法,尽管有些方法还不太成熟、不太完善。由于医学信息学可看作是信息学向医学渗透的产物,医学信息系统既涉及人复杂的生命系统(由同样复杂的若干子系统组成),又涉及计算机通信网络这一系统,所以医学信息学面对的是多信道、多用户网络、多个通信终端的庞大复杂系统。在这些系统中,信息的产生、获取、加工、存储、使用等是十分复杂的。不仅要研究信源(发出信息的来源)发出多少信息,信宿(信息接受者)接受到多少信息,信道(传输信息的媒介)上的信息流量,更要研究信息的语义和信息的效用(价值)。兹将业界目前认可的常用研究方法扼要列举如下:

(1) 医学信息的采集、加工、传输、存储、分析和利用:就中文医学信息而言,主要包括汉字信息处理和汉语信息处理两个方面,前者涉及编码问题,后者涉及词法(包括词的切分)、句法、语义、语境的处理等。

(2) 计算机和网络技术:包括计算机软硬件系统和应用系统,因特网协议标准、局域网和互联网、网络管理和网络安全技术等。

(3) 信号处理和医学成像技术:包括随机信号的提取、分析、变换、滤波、检测、估计与识别,数字图像的采集、存储、检索、表达和像素关系,图像变换,图像增强、恢复、重建,图像分类、切割,以及分子影像成像技术。

(4) 人工智能:包括搜索技术,知识表示和推断,机器学习(判别分析、特点抽取、错误估计、聚类分析)等。

(5) 医学决策分析方法:包括决策树、对策论、敏感性分析等。

(6) 数据安全:首先是在计算机网络环境中保持数据的机密性、完整性和确证性的问题。数据安全中密码技术是关键(包括密码算法、密钥管理、数字签名、身份认证、安全协议等)。

二、医学信息学的应用领域

医学信息学的应用领域非常广泛,涵盖从分子水平——基因组分析、基因芯片和药物开发等;细胞水平——生物信号处理和分析;器官水平——图像分析和处理,如磁共振及CT;个人水平——电子病历;机构水平——医院信息系统;社会群体水平——公共卫生信息系统等。应用范围遍及医院管理和临床应用、远程医学、医学研究应用、医学知识库应用、医学教育、医学信息检索、医学决策支持、公共卫生、疾病预警监视、防范生物恐怖袭击等诸多领域。以下是国外的一些具体应用实例。

(一) 一种访问公共卫生数据的因特网应用

访问流行病学数据对公共卫生事业是极其重要的。然而,一般情况下绝大多数已公布的数据都是过时的,由于保密、不同的用户需求、成本、安全考虑以及其他原因,现有的数据库又不能访问。尽管现在有了因特网这个工具,但数据仍然不及时,并经常是过时印刷本报告的Web版本,而且没有操作数据的能力。针对这种情况,美国华盛顿州卫生局的流行病学、卫生统计和公共卫生实验室开发了基于Web应用的联机EpiQMS(epidemiologic query and mapping system, 流行病学查询和作图系统, <http://EpiQMS.doh.wa.gov>)。该系统可以对州或地区的公共卫生数据(死亡、出生、癌症登记、住院、性传播疾病、损伤、突发事件应对、环境执法)进行研究、统计和地理分析,以便进行卫生评估或监察支持。该系统特别强调事件发生在“何处”,在不同地理等级(县、州和不同邮编地区等)的卫生评估会按年龄、性别的发生率进行展示。对小数量问题的发生率,运用经验贝叶斯平滑法进行调整后也可获得。信息提供方式有图表、表格和地图。提供对象分为三个层次(三种安全水平):一般群众、公共卫生和医疗保健工作者以及流行病学家和卫生官员。

该系统开发的原则是尊重隐私,信息准确、及时和容易使用,高速存取,对发生率数据的计算和表示标准化,有帮助各层次用户的辅导员能力,容易维护,具有便携性,容易在新地点安装,用户不花钱或少花钱,有处理小数量问题的能力,可处理许多不同类型的数据,可按地理分布对数据作图,以及具有不同的访问水平。

该数据库对100种死亡原因按年龄、性别、种族、年份、地理进行分解,现有1 036 800条记录。如果用户提出有另外的死亡原因要加到数据库中,可根据新的国际疾病分类法用SAS程序脱机运行。在保护隐私方面采取了许多措施,如数据库结构不允许测定记录水平的信息,也不能用信息的反向工程推导某人的身份等。该系统在因特网地址上不使用个人档案,用户查询的数据库是一种累积档案数据库,经标引后可高速访问。当地软件安装对用户不收费。美国其他州也可进行非正式查询。

(二) 对心电图数据的获取和分析的标记语言

心电图是临床上使用最频繁的一种常规检查方法。但是,心电图数据的获取、存储和分析使用的是不同的格式和软件平台。数据只有在开放式的情况下共享和解释,其益处才能充分利用。这就需要研究跨平台的解决方案,以支持生物医学培训、决策和远程医学的应用。

英国的 Haiying Wang 等人研究了一种用于心电图数据获取和分析的标记语言(ecgML),取得了很好的结果。基于 XML(extensible markup language,可扩展的标记语言)技术的优越性,ecgML 对表示和交换心电图数据,有能力提供一种独立于系统、应用和格式的解决方案。该方案与 FDA 提出的表示已采集的时间序列数据的 XML 词汇表有很大的不同。ecgML 的目的是拟订一种开放透明的表示、交换和挖掘心电图数据的方法。因此,ecgML 不仅由可用于在心电图数据中进行知识发现的基本成分构成,同时它还遵循将内容和表示信息分开的原则。

ecgML 将能够把心电图数据无缝整合到电子病历和临床实践指南中。该方案能够支持不同的心电图获取和可视化设备之间的数据交换。它也许还能够使用异质软件平台和应用进行数据挖掘。在 ecgML 一条记录中含有的数据和元数据,也许对改进心电图应用中的模式识别有用。它同样能帮助实施自动决策支持模型,如基于案例的推理。

(三) 计算机辅助诊断(computer aided diagnosis, CAD)

计算机辅助诊断作为一种帮助放射科医生进行图像诊断的基本工具,近来已转变为一种更复杂的应用系统。图像处理功能和模式识别程序通常需有一套数据存储在一个数据库系统中,作为一套标准或数据仓库。在诊断过程中由医生进行的操作(提醒注意)要有一个友好的用户图形界面。要将图形处理、数据库结构和图形用户界面整合在一起,就要设计多层 CAD 系统。图像处理方法和数据可视化(或用户友好的图形界面)会影响 CAD 的有效性。

波兰的 Ewa Pietka 等为对骨龄评估方法进行评价,设计构建了计算机辅助诊断系统信息学基础设施。该系统设计为三层:第一层为数据库管理系统,第二层为应用层,最后一层为表示层。数据库管理系统负责数据存档和分配。应用层包括管理应用服务器、WWW 服务器和一个 CAD 工作站,共三个部分。应用层的任务包括数据库访问(文本和图像),由图形用户界面授予的数据表示,以及由设计的动态网页所允许的对整个 CAD 系统的远程访问。表示层通过一个网页取景器允许对系统进行方便的远程访问。在用户端,网页取景器则作为访问整个系统的一个界面。

计算机辅助诊断在检测和诊断乳腺癌方面同样发挥了不可替代的作用。乳腺 X 线摄影是早期检测乳腺癌的最佳筛查方法。乳腺放射线获取技术能使乳腺癌早期征在乳房 X 线照片上更明显。但乳腺 X 线检查的精确度则取决于乳腺 X 线图像的质量和放射科医生对这些图像的解释能力。使用对某一图像的计算机化的分析输出,可以帮助放射科医生检测和诊断乳腺癌,并且可能改进对乳腺癌图像的解释和后续的卫生保健。

计算机医学诊断应用于从三个不同方面改进了分析:首先,通过使用定量特征而不是直

观插入提高了客观性。这些特征是自动而不是手工提取的,这样一方面提高了分析的速度,同时使测量不受外部条件(光线)、测量的不精确和错误(刻度错误)等的影响。第二,规定了一种透明算法,使决策过程在特征分析中标准化。这又再次提高了客观性并解决了可复制性的问题。最后,这样可使一套图像和描述图像的特征能够被采集并被认可为一种医学标准。

(四) 生物芯片技术在临床工作中的应用

生物芯片(biochip)的概念源自计算机芯片,是指采用光导原位合成或微量点样等方法,将大量生物大分子如核酸片段、多肽分子甚至组织切片、细胞等生物样品有序地固化于支持物(如玻片、硅片、聚丙烯酰胺凝胶、尼龙膜等载体)的表面,组成密集二维分子排列,然后与已标记的待测生物样品中的靶分子杂交,通过特定的仪器如激光共聚焦扫描或电荷偶联摄像机(CCD)对杂交信号的强度进行快速、并行、高效的检测分析,从而判断样品中靶分子的数量。生物芯片具有高通量、微型化和自动化的特点,生物芯片发展的最终目标是从样品制备、生化反应到检测的整个分析过程集成化,以获得所谓的微型全分析系统或称缩微芯片实验室。生物芯片在临床疾病诊断、新药研发、环境监测等领域有着广泛的应用前景。目前应用最广泛的生物芯片是基因芯片(gene chip),DNA芯片(DNA chip)和DNA微阵列(DNA microarray)。虽然生物芯片技术的特点是高通量,但实际上临床应用发展的趋势是低密度生物芯片技术。

日本的 Akito Daiba 等人报告,他们设计和证实了一种具有惟一参照 RNA 的低密度 cDNA 微阵列,作为一种模型,用模式识别分析来预测干扰素对丙型肝炎病人的效率。为确保微阵列数据精确和前后一致,他们从每一个基因(mRNA)的三个点上采集数据并使用三种不同的扫描条件。为提高数据采集精度,对个人要使用参照对比。但因丙型肝炎病人炎症损害遍及全肝,所以对丙肝病人使用参照对比很困难。于是他们就使用 RNA 作为人工参照。由于高密度微阵列所产生的数据量太大,难以分析和理解其含义。他们就选了 296 个基因用作低密度微阵列系统。他们还研究了一种新的算法,以取代标准的层级聚类方法。该法可将研究对象区分成无效和持续有效/一过性有效两组,预测准确率 100% ($P < 0.005$)。

第三节 国外医学信息学的教育和培训

医学信息学是一门发展中的交叉学科,在国外,根据受训对象的不同,正规教育设有硕士和博士两种课程;同时作为继续教育的一项内容,也可以根据对象情况及培训时间长短做出不同的培训安排。各项课程设置的目的、要求,学习的课程及学时,以及考核评估等在网

一、对在职医师的培训

以美国爱荷华大学 Cedar Rapids 教育基金会 2002 年培训家庭医生(family physicians)

方案为例,该方案课程重点在循证医学、医学信息学原理、行为科学、病人教育和计算机技能,其具体要求是:

(1) 计算机基本知识和工具:计算机/操作系统技能(计算机概念,包括计算机如何工作、计算机硬件知识、如何使用和维护笔记本电脑、如何处理常见问题),计算机日常管理(安装和删除程序、定制自己的工作环境等),日常办公室应用(文字处理、制表、准备学术报告稿、日程安排、个人数据库——文件操作/诊断经验),通信技术(收发电子邮件、利用电子邮件处理病人请求,了解因特网基础和设备及软件需要,会用浏览器及搜索引擎),故障检修,如何使用技术支持等。

(2) 处理/评估信息:建立在证据基础上的评估和处理(对循证医学的认识——什么是最佳证据,有用性公式,处理不确定性等),使循证医学与临床经验相一致。评估诊断试验,对获得的治疗信息进行评估,基于地区人口的评估,临床决策支持等。

(3) 存储信息以便恰当使用:信息的性质,文献提供的原则,面向对象的记录,分类、编码,特定文献提供,卫生信息的隐私和安全。

(4) 采集和检索信息:数据库原理,检索和查询原理,使用计算机化病人信息,病人病历结构,自动提示,不同数据来源(实验室、X线和电子病历)的整合,无障碍采集/检索,有效医学文献和指南的检索,及时盘存信息。

(5) 信息交流:交流和学习原则,对病人的交流,个性化的病人教育,将交流工具和行为科学整合,医生—病人—计算机关系,设计医疗保健系统(质量改进、混沌系统、结果评估、有效性),与第三方交流。

(6) 高级和选修题目:如何购买计算机,编程和编程语言,规划信息资源,知识基础和人工智能,医学信息学中的争论和医学信息学研究。

二、研究生教育

研究生教育是以在该学科领域培养专门人才为目标,在学科设置和内容深度上有更严格的要求。现以美国斯坦福大学医学院从2004年秋季开始的生物医学信息学(由广义的生物信息学和临床信息学组成)研究生培养规划为例,摘要介绍培养目标、课程重点、相关要求等。

斯坦福生物医学信息学的硕士生培养分学术型(academic)、专业型(professional)和本科与硕士学位学习同时进行三种方式:①学术型的学生须用两年时间完成学业。第一年通过课堂作业和研究项目要掌握基本概念和工具,在第一年和第二年要用一半时间参加研究工作,要求学员必须有由两名指导老师批准的书面研究报告。该学科毕业生应能在生物医学信息学的基础和应用研究上从事创造性的工作。②专业型的硕士学位课程是为正在从事专业工作的人士设计的。这些人已在某一学科受过高级培训,希望获得交叉学科的技能。这种人可采取兼职的方式通过网上就读学位,只要能在五年内完成学业就行。③允许大学本科生在完成大学学士学位课程时,同时在本系或其他系进行硕士学位的学习,不过只有斯坦福的本科生才能申请就读该学位。

从事硕士学位学习的学生,可以将重点放在生物医学研究中信息管理和分析的任何一

个方面,通过培训他们既可使用信息技术处理、分析和理解生物医学数据,也可开发新的工具,去利用信息改进卫生保健。具体研究领域包括:决策支持、集成工作站、知识获取、语言输入、笔输式计算、医学病历、计算生物学、生物序列分析、生物三维结构表达、药物遗传学、药物基因组学、基因组学、合作技术(collaborative technologies)、基因网络的表述、生物医学信息和文献检索、医学成像、不确定事件下的推理、可控生物医学词汇表、技术评估和卫生服务研究等。

培训课程要求从以下三个方面进行:①用于表述知识和开发模型的信息学方法;②实现这些表述和模型的计算机科学;③必要的专门化生物医学知识,以便找出重要问题并对这些问题施加影响。由于技术的进步,产生了新的伦理问题,这在课程设计中也要有所反映。课程的指导方针是在下列五个领域内分别提出各自的核心要求。

(1) 生物医学信息学(15个单元):要求掌握生物信息学的基本方法,可选课程有:①生物医学信息学入门、系统和应用;②计算生物学的表达和算法;③生物医学信息学研究项目。

(2) 概率、统计学和决策科学(9个单元,学生要在概率、贝叶斯统计、决策分析、机器学习和实验设计技术五门中至少选三门学习):要求掌握概率理论、统计推断和卫生保健中成本、风险和利益的分析。概率可选的课程有:①概率分析;②影响图和概率网络。贝叶斯统计可选的课程有:①随机模型的建立;②数据分析;③计算遗传学统计方法;④现代应用统计学;⑤人工智能概率模型;⑥统计学习及模式分类等。决策分析可选决策分析概论或决策分析I和决策分析II。

(3) 计算机科学(9个单元):可选:①重要编程和机器结构知识;②计算机系统和汇编语言;③编程概论。

(4) 生物医学知识(9个单元)。

(5) 伦理和社会政策(6个单元):可选 ①卫生保健系统和政策概论;②卫生保健成本、风险和利益分析;③计算机、伦理和社会责任;④美国卫生保健政治经济学等。

以上共48个单元。此外,研究生还须另选修6个单元,以达到修满总计54个单元的要求。

对打算在生物医学信息学领域从事独立研究工作,并在生物信息学、临床信息学或成像信息学有应用经验的人员,可申请就读博士学位。对这些人员除要完成以上硕士学位的学习课程外,还要由生物医学信息学交叉学科委员会提出另外的要求,除与其他博士生培养相一致的要求外(作为辅导老师承担教学、提出科研计划、参加科研工作和完成毕业论文等),要制定和完成同一的学习计划(包括核心课程、口试以及在硕士生培养规划外的要求),还要选修至少两门高级课程。

(高岚)

阅读思考题 1

1. 简述信息、数据和知识三种概念之间的联系和区别。
2. 简述医学信息学和生物信息学的定义及其相互联系与区别。

3. 试说医学信息学的应用实例。

参考文献

刘东等. 1995. 信息工作理论与实践. 北京: 科学技术文献出版社, 1~18

马费成等. 2000. 信息管理学基础. 武汉: 武汉大学出版社, 1~16

Akito Daiba et al. 2004. A low-density cDNA microarray with a unique reference RNA; pattern recognition analysis for IFN efficacy prediction to HCV as a model, *Biochem Biophys Res Commun.* 19: 315(4): 1088~1096

A medical informatics curriculum for the 21 st century A curriculum developed by the faculty and residents of CRMEF. <http://www.crmeff.org/curriculum>

Collen MF. 1986. Origins of medical informatics, *The Western Journal of Medicine* 145:778~785

Collen MF. 1994. The origins of informatics, *J Am Med Inform Assoc.* 1(2):91~107

Ewa Pietka et al. 2005. Computerized medical imaging and graphics. 29(2~3):157~169

Guide to biomedical informatics curriculum structure. <http://smi.stanford.edu/academics/bioinformatics-curriculum.html>

Haiying Wang et al. 2003. A markup language for electrocardiogram data acquisition and analysis (ecgML), *BMC Medical Informatics and Decision Making.* 3:4

Hoskins, Richard E et al. 2002. EpiQMS: An internet application for access to public health data for citizens, providers, and public health investigators. *J Public Health Manag Pract.* 8(3):30~36

(续前)

1 医学信息学