BIOMEDICAL ENGINEERING FUNDAMENTALS

BIOMEDICAL ENGINEERING FUNDAMENTALS

Myer Kutz Editor

Third Edition

New York Chicago San Francisco Athens London
Madrid Mexico City Milan New Delhi
Singapore Sydney Toronto

Library of Congress Control Number: 2020952464

Biomedical Engineering Fundamentals, Third Edition

1 2 3 4 5 6 7 8 9 CCD 26 25 24 23 22 21

ISBN 978-1-260-13626-5
MHID 1-260-13626-4

This book is printed on acid-free paper.

Sponsoring Editor
Robin Najar

Editing Supervisor
Stephen M. Smith

Production Supervisor
Pamela A. Pelton

Acquisitions Coordinator
Elizabeth M. Houde

Project Manager
Jyoti Shaw, MPS Limited

Copy Editor
Mohammad Taiyab Khan, MPS Limited

Proofreader
Nikhil Roshan

Indexer
Edwin Durbin

Art Director, Cover
Jeff Weeks

Composition
MPS Limited

For Arlene, forever

ABOUT THE EDITOR

MYER KUTZ is founder and president of Myer Kutz Associates, Inc., a publishing and information services consulting firm.

CONTENTS

CONTRIBUTORS

Crystal Anderson-Cunanan *Boston Scientific Corporation, Marlborough, Massachusetts* (Chap. 18)

Ronald E. Barr *University of Texas, Austin, Texas* (Chap. 9)

Christopher Batich *University of Florida, Gainesville, Florida* (Chap. 15)

Ravi V. Bellamkonda *Georgia Institute of Technology/Emory University, Atlanta, Georgia* (Chap. 21)

Anthony J. Brammer *University of Connecticut Health, Farmington, Connecticut* (Chap. 13)

Thomas S. Buchanan *University of Delaware, Newark, Delaware* (Chaps. 8, 14)

Hsiang-Han Chen *University of Minnesota, Minneapolis, Minnesota* (Chap. 3)

Vladimir Cherkassky *University of Minnesota, Minneapolis, Minnesota* (Chap. 3)

Shmuel Einav *Tel Aviv University, Tel Aviv, Israel* (Chap. 5)

David Elad *Tel Aviv University, Tel Aviv, Israel* (Chaps. 5, 7)

Michele J. Grimm *Michigan State University, East Lansing, Michigan* (Chap. 20)

Paul Grimshaw *University of Adelaide, Adelaide, Australia* (Chap. 11)

James B. Grotberg *University of Michigan, Ann Arbor, Michigan* (Chap. 6)

Pinchas Halpern *Tel Aviv Medical Center, Tel Aviv, Israel, and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel* (Chap. 7)

Michael N. Helmus *Medical Devices, Drug Delivery, and Nanotechnology, Worcester, Massachusetts* (Chap. 18)

William Hersh *Oregon Health & Science University, Portland, Oregon* (Chap. 2)

Arif Iftekhar *University of Minnesota, Minneapolis, Minnesota* (Chap. 16)

Dentcho Ivanov *Microelectronics Fabrication Center, New Jersey Institute of Technology, Newark, New Jersey* (Chap. 25)

Teena James *Department of Biomedical Engineering and Microelectronics Research Center, New Jersey Institute of Technology, Newark, New Jersey* (Chap. 25)

Chris Jones *Independent Researcher, Adelaide, Australia* (Chap. 11)

Tony M. Keaveny *University of California, San Francisco, California, and University of California, Berkeley, California* (Chap. 10)

David H. Kohn *University of Michigan, Ann Arbor, Michigan* (Chap. 17)

Patrick Leamy *LifeLink Foundation, Tampa, Florida* (Chap. 15)

Hyunjung Lee *Georgia Institute of Technology, Atlanta, Georgia* (Chap. 21)

Merilyn Lock *University of South Australia, Adelaide, Australia* (Chap. 11)

Eric J. Madsen *University of Michigan, Ann Arbor, Michigan* (Chap. 17)

Kurt T. Manal *University of Delaware, Newark, Delaware* (Chap. 8)

Manu Sebastian Mannoor *Department of Biomedical Engineering and Microelectronics Research Center, New Jersey Institute of Technology, Newark, New Jersey* (Chap. 25)

Nancy J. Meilander *National Institute of Standards and Technology, Gaithersburg, Maryland* (Chap. 21)

Elise F. Morgan *University of California, Berkeley, California* (Chap. 10)

Jit Muthuswamy *Arizona State University, Tempe, Arizona* (Chaps. 23, 26)

Roger J. Narayan *University of North Carolina and North Carolina State University, Raleigh, North Carolina* (Chap. 24)

Michael D. Nowak *University of Hartford, West Hartford, Connecticut* (Chap. 12)

Govind Pandey *King George Medical University, Lucknow, India* (Chap. 24)

Prem C. Pandey *Banaras Hindu University, Varanasi, India* (Chap. 24)

Marcus G. Pandy *University of Melbourne, Melbourne, Australia* (Chap. 9)

Donald R. Peterson *Northern Illinois University, DeKalb, Illinois* (Chap. 13)

Bonnie Pierson *University of North Carolina and North Carolina State University, Raleigh, North Carolina* (Chap. 24)

Shrirang V. Ranade *Genentech Inc., a member of the Roche Group, South San Francisco, California* (Chap. 18)

Anat Ratnovsky *Afeka College of Engineering, Tel Aviv, Israel* (Chap. 7)

Narender P. Reddy *University of Akron, Akron, Ohio* (Chap. 1)

Qi Shao *University of Delaware, Newark, Delaware* (Chap. 14)

Bruce C. Towe *Arizona State University, Tempe, Arizona* (Chap. 22)

Bradley Jay Willenberg *University of Central Florida, Orlando, Florida* (Chap. 15)

Oscar C. Yeh *University of California, Berkeley, California* (Chap. 10)

Roya Zandparsa *Tufts University School of Dental Medicine, Boston, Massachusetts* (Chap. 19)

Liang Zhu *University of Maryland, Baltimore County, Baltimore, Maryland* (Chap. 4)

PREFACE

The purpose of this handbook is to provide engineers, physicians, and other medical professionals with fundamental information that will help them use engineering sensibilities and methodologies to invent, design, properly operate, and evaluate medical instruments, devices, and machines. This third edition is an expanded and updated version of Vol. 1, *Fundamentals*, of the *Biomedical Engineering and Design Handbook,* Second Edition, published in the summer of 2009, and follows the same outline. Again, there are four major parts: Biomedical Systems Analysis, which now has three chapters instead of one; Biomechanics of the Human Body, with 11 chapters in both editions; Biomaterials, with seven chapters in both editions; and Bioelectronics, with five chapters in both editions.

Of the 26 chapters in the third edition—two chapters more than in the second edition—four chapters are entirely new and nine have been updated. Two second edition chapters have been dropped: Biodynamics: A Lagrangian Approach and Biomedical Signal Processing. I should point out that the first volume of the second edition constituted a substantial revision of the corresponding parts of the first edition. So the overall revision from first to third editions has been extensive, in keeping with the growth of knowledge in the biomedical engineering discipline.

The four new chapters in this edition cover the following important topics, which add value for readers: Biomedical Informatics (Chap. 2, contributed by William Hersh), which deals with the efficient storage, acquisition, and use of information in healthcare; Machine Learning for Biomedical Engineering (Chap. 3, contributed by Vladimir Cherkassky and Hsiang-Han Chen), which involves methodological issues and assumptions important for applying machine learning to biomedical data so that limitations of predictive data-analytic models can be understood properly, available data can be modeled mathematically, and the right questions can be asked of the data; Biomechanics of the Foot and Knee (Chap. 11, contributed by Paul Grimshaw, Chris Jones, and Merilyn Lock), fundamental knowledge which underlies, for example, treatment of injury to the anterior cruciate ligament; and Neural Interfaces (Chap. 26, contributed by Jit Muthuswamy), which can be implanted so that electrical activity from single neurons in the brain and peripheral nervous system can be recorded.

As noted above, nine chapters from the second edition have been updated for this edition. Comments regarding those updates are presented in the following paragraphs.

Liang Zhu, contributor of Chap. 4, Heat Transfer Applications in Biological Systems, writes: "There are quite a lot of updates due to my teaching of bioheat transfer and my research in the past 10 years. In addition to the major updates, including some figures, many new references after 2007 were added to the manuscript." Among the sections updated are those dealing with the Pennes bioheat equation, bioheat transfer modeling, temperature, thermal properties and blood flow measurements, hyperthermia, and thermal damage assessment.

Marcus G. Pandy, co-contributor with Ronald E. Barr, of Chap. 9, Biomechanics of the Musculoskelatal System, writes: "I have removed the text that appeared in the very last section of the previous version and replaced this with a description of running biomechanics, which I believe the reader will find interesting plus it follows nicely from the description of walking biomechanics appearing immediately above. I also have included two new figures."

In Chap. 12, Finite-Element Analysis, contributor Michael D. Nowak has added text on coronary artery disease, vascular assist devices, and aortic valve replacement, plus two complex figures.

Donald R. Peterson, co-contributor with Anthony J. Brammer, of Chap. 13, Vibration, Mechanical Shock, and Impact, writes: "The chapter has significant updates as some of the

second edition version is obsolete now." Among the sections with major updates are Biodynamic Models in the section on Models and Human Surrogates; Occurrence of Health Effects and Injury in the section on Countermeasures; Protection against Hand-Transmitted Vibration in the section on Countermeasures.

In Chap. 15, Polymeric Biomaterials, the contributors, led by Christopher Batich, added an extensive new section on biocompatibility and tissue engineering by Bradley Jay Willenberg, as well as a major new section on sterilization by Patrick Leamy. Part of the section on tissue scaffolds has been rewritten.

David H. Kohn, co-contributor of Chap. 17, Bioceramics, writes: "The main changes to the chapter included an expansion of the inorganic/organic hybrid section, especially discussion of peptide conjugation to calcium phosphates and nanotechnology."

Michael N. Helmus, co-contributor of Chap. 18, Cardiovascular Biomaterials, wrote to me with a lengthy list of changes, including updates to tables that were in the second edition; addition of new tables, together with discussion; a new section, Special Risks Associated with the Use of Materials of Animal Origin; plus five new discussions on other topics.

Michele J. Grimm, contributor of Chap. 20, Orthopedic Biomaterials, writes: "Sections on mechanobiology were added to the discussions of bone, cartilage, and ligament/tendon. References were updated where key new information was available (e.g., tantalum, zirconia, calcium sulfate injectables, wear particles, metal-on-metal hip implants, and others). A section on biodegradable metal was added."

Chapter 24, Biosensors (lead contributor, Roger J. Narayan), was updated with several changes to the text and the addition of two figures.

Although the majority of this handbook's contributors work in the United States, there is a noteworthy international component. Contributors of two chapters, Biomechanics of the Musculoskelatal System, which was updated, and Biomechanics of the Foot and Knee, which is new, are located in Australia. Two chapters, carried over unchanged from the second edition, were written by contributors located in Israel. A co-contributor of the chapter on Vibration, Mechanical Shock, and Impact is based part-time in Canada, and contributors of updates to the chapter on Biosensors are located in India. Most contributors work in academia, with a smattering of them in industry and government. I am immensely grateful to all of the contributors—those who contributed new and updated chapters, as well as those whose chapters from the second edition have been carried over unchanged into this one. I know how difficult it is to find the time to do any kind of writing, particularly a scholarly chapter on a technical subject, in these days when professionals in academia and industry are so hard-pressed. Thank you all.

Thanks also to my wonderful wife, Arlene, whose love is the thing I cherish most in all the world.

Myer Kutz
Delmar, New York

P · A · R · T · 1

BIOMEDICAL SYSTEMS ANALYSIS

CHAPTER 1
MODELING OF BIOMEDICAL SYSTEMS

Narender P. Reddy
University of Akron, Akron, Ohio

Models are conceptual constructions which allow formulation and testing of hypotheses. A mathematical model attempts to duplicate the quantitative behavior of the system. Mathematical models are used in today's scientific and technological world due to the ease with which they can be used to analyze real systems. The most prominent value of a model is its ability to predict as yet unknown properties of the system. The major advantage of a mathematical or computer model is that the model parameters can be easily altered and the system performance can be simulated. Mathematical models allow the study of subsystems in isolation from the parent system. Model studies are often inexpensive and less time consuming than corresponding experimental studies. A model can also be used as a powerful educational tool since it permits idealization of processes. Models of physiological systems often aid in the specification of design criteria for the design of procedures aimed at alleviating pathological conditions. Mathematical models are useful in the design of medical devices. Mathematical model simulations are first conducted in the evaluation of the medical devices before conducting expensive animal testing and clinical trials. Models are often useful in the prescription of patient protocols for the use of medical devices. Pharmacokinetic models have been extensively used in the design of drugs and drug therapies.

There are two types of modeling approaches: the black box approach and the building block approach. In the black box approach, a mathematical model is formulated based on the input-output characteristic of the system without consideration of the internal functioning of the system. Neural network models and autoregressive models are some examples of the black box approach. In the building block approach, models are derived by applying the fundamental laws (governing physical laws) and constitutive relations to the subsystems. These laws together with physical constraints are used to integrate the models of subsystems into an overall mathematical model of the system. The building block approach is used when the processes of the system are understood. However, if the system processes are unknown or too complex, then the black box approach is used. With the building block approach, models can be derived at the microscopic or at the macroscopic levels. Microscopic models are spatially distributed and macroscopic models are spatially lumped and are rather

global. The microscopic modeling often leads to partial differential equations, whereas the macroscopic or global modeling leads to a set of ordinary differential equations. For example, the microscopic approach can be used to derive the velocity profile for blood flow in an artery; the global or macroscopic approach is needed to study the overall behavior of the circulatory system including the flow through arteries, capillaries, and the heart. Models can also be classified into continuous time models and models lumped in time domain. While the continuous time modeling leads to a set of differential equations, the models lumped in time are based on the analysis of discrete events in time and may lead to difference equations or sometimes into difference-differential equations. Random walk models and queuing theory models are some examples of discrete time models. Nerve firing in the central nervous system can be modeled using such discrete time event theories. Models can be classified into deterministic and stochastic models. For example, in deterministic modeling, we could describe the rate of change of volume of an arterial compartment to be equal to rate of flow in minus the rate of flow out of the compartment. However, in the stochastic approach, we look at the probability of increase in the volume of the compartment in an interval to be dependent on the probability of transition of a volume of fluid from the previous compartment and the probability of transition of a volume of fluid from the compartment to the next compartment. While the deterministic approach gives the means or average values, the stochastic approach yields means, variances, and covariances. The stochastic approach may be useful in describing the cellular dynamics, cell proliferations, etc. However, in this chapter, we will consider only the deterministic modeling at the macroscopic level.

The real world is complex, nonlinear, nonhomogeneous, often discontinuous, anisotropic, multilayered, multidimensional, etc. The system of interest is isolated from the rest of the world using a boundary. The system is then conceptually reduced to that of a mathematical model using a set of simplifying assumptions. Therefore, the model results have significant limitations and are valid only in the regimes where the assumptions are valid.

1.1 COMPARTMENTAL MODELS

Compartment models are lumped models. The concept of a compartmental model assumes that the system can be divided into a number of homogeneous well-mixed components called compartments. Various characteristics of the system are determined by the movement of material from one compartment to the other. Compartment models have been used to describe blood flow distribution to various organs, population dynamics, cellular dynamics, distribution of chemical species (hormones and metabolites) in various organs, temperature distribution, etc.

Physiological systems (e.g., cardiovascular system) are regulated by humoral mediators and can be artificially controlled using drugs. For instance, the blood pressure depends on vascular resistance. The vascular resistance in turn can be controlled by vasodilators. The principle of mass balance can be used to construct simple compartment models of drug distribution. Figure 1.1 shows a general multicompartmental (24-compartment) model of drug distribution in the human body. The rate of increase of mass of a drug in a compartment is equal to the rate of mass flowing into the compartment minus the rate of mass leaving the compartment, minus the rate of consumption of the drug due to chemical reaction in the compartment. In the model shown in Fig. 1.1, the lungs are represented by three compartments: Compartment 3 represents the blood vessels (capillaries, etc.) of the lung, the interstitial fluids of the lung are represented by compartment 4, and the intracellular components of the lung are represented by compartment 5. Each other organ (e.g., kidneys) is represented by two compartments consisting of the blood vessels (intravascular) and the tissue (extravascular consisting of interstitial and intracellular components together). Let us consider the model equations for a few compartments.

For compartment 3 (lung capillaries),

$$V_3 dC_3/dt = Q_2 C_2 - Q_3 C_3 - K_{3\text{-}4} A_{3\text{-}4} (C_3 - C_4) \tag{1.1}$$

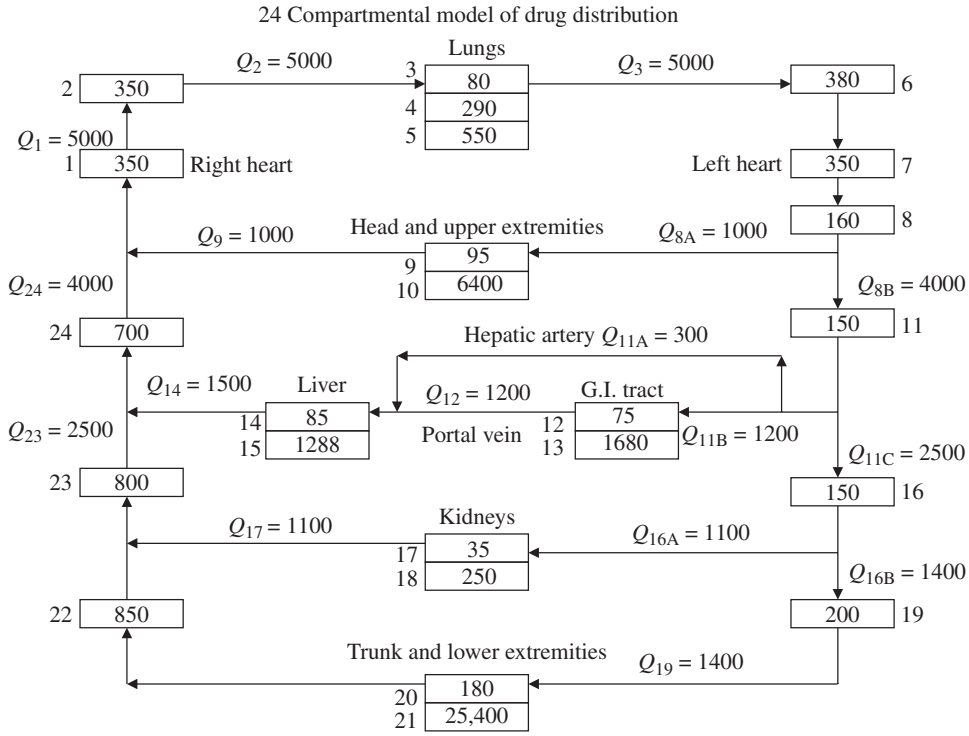

FIGURE 1.1 A generalized multicompartment (24) model of the human body to analyze drug distribution in the body. The numbers in the compartments represent volumes in milliliters. The numbers on the lines are flow rates in mL/min.

Q_2C_2 is the rate of mass flowing into compartment 3 from compartment 2, and Q_3C_3 is the rate of mass flowing out of compartment 3 into compartment 6. In addition, there is the interface mass transfer (diffusion) from capillaries into the interstitial spaces. This is represented by the last term. $K_{3\text{-}4}$ is the diffusional permeability of lung capillary. The diffusional permeability depends on capillary pore size, the number of pores per unit area, the diffusion coefficient for the drug molecule, the ratio of the diameter of the drug molecule, and the pore diameter. This permeability is different from the hydraulic permeability. $A_{3\text{-}4}$ is the lung capillary (interface) surface area. Mass is equal to volume times concentration. The change in volume occurs over a longer duration when compared to the changes in concentration. Consequently, volumes are assumed to be constant.

For the interstitial compartment,

$$V_4 dC_4/dt = K_{3\text{-}4}A_{3\text{-}4}(C_3 - C_4) - K_{4\text{-}5}A_{4\text{-}5}(C_4 - C_5) \tag{1.2}$$

For the intracellular compartment,

$$V_5 dC_5/dt = K_{4\text{-}5}A_{4\text{-}5}(C_4 - C_5) - M_R \tag{1.3}$$

where M_R is the rate of metabolic consumption of the drug. This could be a constant at high concentrations and a function of concentration at low concentrations. Recently, Kim et al. (2007) have developed a whole body glucose homeostasis during exercise and studied the effect of hormonal control.

Simple one compartmental models can be used for the prescription of treatment protocols for dialysis using an artificial kidney device. While the blood urea nitrogen (BUN) concentration in the

normal individual is usually 15 mg% (mg% = milligrams of the substance per 100 mL of blood), the BUN in uremic patients could reach 50 mg%. The purpose of the dialysis is to bring the BUN level closer to the normal. In the artificial kidney, blood flows on one side of the dialyzer membrane and dialysate fluid flows on the other side. Mass transfer across the dialyzer membrane occurs by diffusion due to concentration difference across the membrane. Dialysate fluid consists of a makeup solution consisting of saline, ions, and the essential nutrients so as to maintain zero concentration difference for these essential materials across the membrane. However, during the dialysis, some hormones also diffuse out of the dialyzer membrane along with the urea molecule. Too rapid dialysis often leads to depression in the individual due to the rapid loss of hormones. On the other hand, too slow dialysis may lead to unreasonable time required at the hospital. Simple modeling can be used to calculate the treatment protocols of mass coming into the body from the dialyzer, plus the metabolic production rate. When the patient is *not* on dialysis, the concentration of urea would increase linearly if the metabolic production rate is constant or will increase exponentially if the metabolic production rate is a linear function of the concentration (first-order reaction). When the patient is *on* dialysis, the concentration would decrease exponentially. This way, the treatment protocol can be prescribed after simulating different *on* and *off* times (e.g., turn on the dialyzer for 4 hours every 3 days) to bring the BUN under control. In the chapter on artificial kidney devices, a simple one compartmental model is used to compute the patient protocol.

Compartmental models are often used in the analysis of thermal interactions. Simon and Reddy (1994) formulated a mathematical model of the infant-incubator dynamics. Neonates who are born preterm often do not have the maturity for thermal regulation and do not have enough metabolic heat production. Moreover, these infants have a large surface area to volume ratio. Since these preterm babies cannot regulate heat, they are often kept in an incubator until they reach thermal maturity. The incubator is usually a forced convection heating system with hot air flowing over the infant. Incubators are usually designed to provide a choice of air control or the skin control. In air control, the temperature probe is placed in the incubator air space and the incubator air temperature is controlled. In the skin control operation, the temperature sensor is placed on the skin and infant's skin temperature is controlled. Simon et al. (1994) used a five-compartmental model (Fig. 1.2) to compare the adequacy of air control and skin control on the core temperature of the infant. They considered the infant's core, infant's skin, incubator air, mattress, and the incubator wall to be four separate well-mixed compartments.

The rate of change of energy in each compartment is equal to the net heat transfer via conduction, convection, radiation, evaporation, and the sensible heat loss. There is a convective heat loss from the infant's core to the skin via the blood flow to the skin. There is also conductive heat transfer from the core to the skin. The infant is breathing incubator air, drawing in dry cold air at the incubator air temperature and exhaling humidified hot air at body temperature. There is heat transfer associated with heating the air from incubator air temperature to the infant's body (core) temperature. In addition, there is a convective heat transfer from the incubator air to the skin. This heat transfer is forced convection when the hot air is blowing into the incubator space and free convection when the heater manifolds are closed. Moreover, there is an evaporative heat loss from the skin to the incubator air. This is enhanced in premature infants as their skin may not be mature. Also, there is a conductive heat transfer from the back surface of the skin to the mattress. Also, exposed skin may radiate to the incubator wall. The incubator air is receiving hot air (convective heat transfer) from the hot air blower when the blower is in the *on* position. There is convective heat transfer from the incubator air to the incubator wall and to the mattress. In addition, there is metabolic heat production in the core. The energy balance for each compartment can be expressed as

$$mCp(dT/dt) = \Sigma Q_{\text{in}} - \Sigma Q_{\text{out}} + G \tag{1.4}$$

where m is the mass of the compartment, T is the temperature, t is the time, Q is the heat transfer rate, and G is the metabolic heat production rate. G is nonzero for the core and zero for all other compartments. G is low in low-birth-weight and significantly premature babies. Simon et al. (1994) investigated infant-incubator dynamics in normal, low birth weight, and different degrees of prematurity under skin and air control. Recently, Reddy et al. (2008) used the lumped compartmental

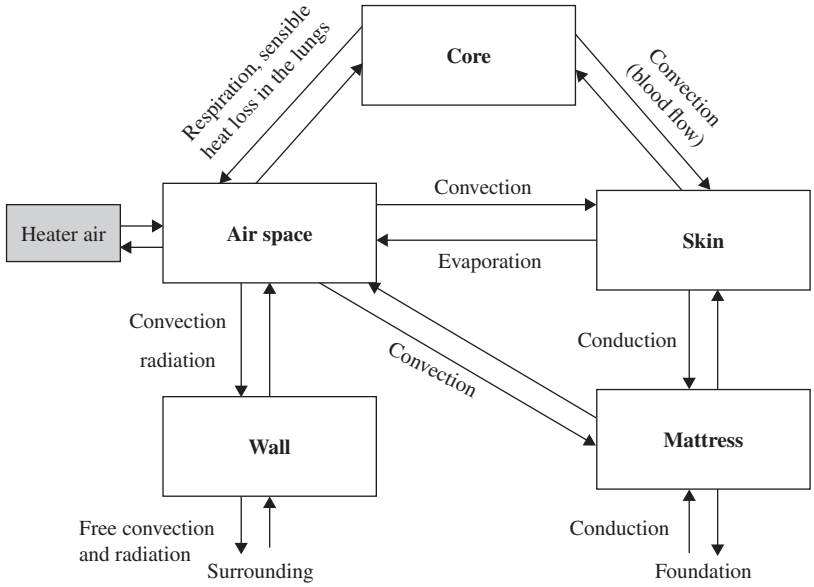

FIGURE 1.2 A lumped parameter model of the infant-incubator dynamics used by Simon et al. (1994) to simulate the effect of various control modes in a convectively heated infant incubator. Infant's core and skin are modeled as two separate compartments. The incubator air space, the incubator wall, and the mattress are treated as three compartments. Heat interactions occur between the core (infant's lungs) and the incubator air space through breathing. Skin-core heat interactions are predominantly due to blood flow to the skin. Heat transfer between the infant's skin and the incubator air is due to conduction and convection. Heat transfer from the skin to the mattress is via conduction, and heat transfer to the wall is via radiation from skin and convection from the air.

model of Simon et al. (1994) to evaluate the efficacy of air control, skin, control, and fuzzy logic control which incorporates both skin and air temperatures.

Compartmental models have been used to model particle dynamics. The growing number of cases of lung diseases, related to the accumulation of inhaled nonsoluble particles, has become a major problem in the urban population. Sturm (2007) has developed a simple multicompartment model for the clearance of nonsoluble particles from the tracheobronchial system (Fig. 1.3). While most of the particles are rapidly transported toward the pharynx by the beating celia, the particles caught in between celia in the highly viscous gel layer (compartment 1) may enter the low viscous sol layer (compartment 2) via diffusion. From the sol layer, they could enter the epithelium (compartment 5) and eventually enter the regional lymph node (compartment 6) or enter the blood circulation. Alternatively, they could be captured by the macrophages (compartment 4) in any of these layers and could reach the regional lymph node or the blood circulation (compartment 6) or the gastrointestinal tract (GIT; compartment 3). Macrophages could release phagocytosed particles into any of these layers. In addition, the particles could defuse among all three layers (gel, sol, and epithelium) in both directions. Sturum (2007) has derived model equations based on the diffusion of particles and other modes of transport.

1.2 ELECTRICAL ANALOG MODELS OF CIRCULATION

Electric analog models are a class of lumped models and are often used to simulate flow through the network of blood vessels. These models are useful in assessing the overall performance of a system or a subsystem. Integration of the fluid momentum equation (longitudinal direction, in cylindrical

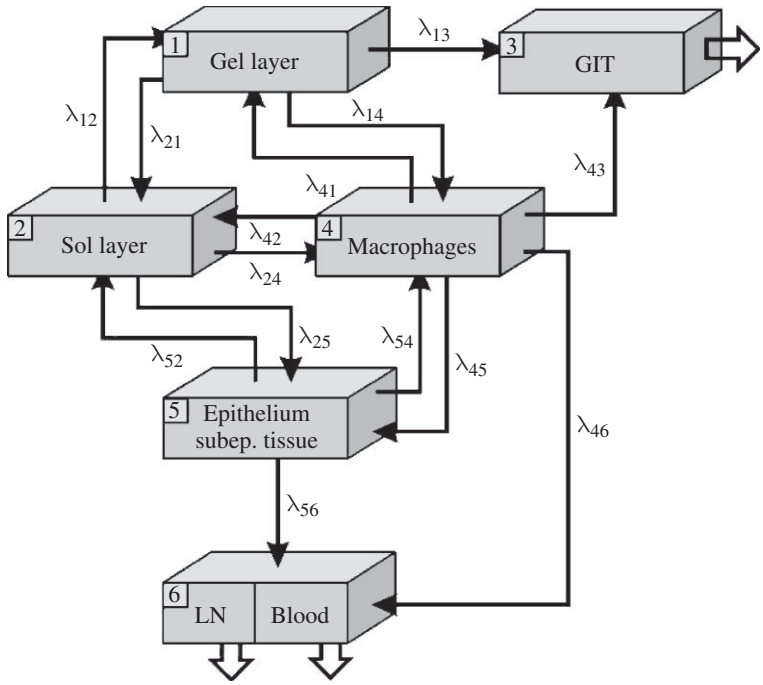

FIGURE 1.3 A multicompartmental model for the clearance of inhaled insoluble particles from the lung. [*Reproduced with permission from Sturm (2007).*]

coordinates) across the cross section results in the following expression (Reddy, 1986; Reddy and Kesavan, 1989):

$$\rho dQ/dt = \pi a^2 \, \Delta P/\ell 2a\tau_w \tag{1.5}$$

where ρ is the fluid density, Q is the flow rate, a is the wall radius, P is the pressure, ℓ is the length, and τ_w is the fluid shear stress at the wall. If we assume that the wall shear stress can be expressed using quasisteady analysis, then the wall shear stress can be estimated by $\tau_w = 4\mu Q/a^3$. Upon substituting for the wall stress and rearranging, the results are

$$[\rho\ell/(\pi a^2)]dQ/dt = \Delta P - [8\mu\ell/(\pi a^4)]Q \tag{1.6}$$

The above equation can be rewritten as

$$LdQ/dt = \Delta P - RQ \tag{1.7}$$

where $L = \rho\ell/(\pi a^2)$ and $R = 8\mu\ell/(\pi a^4)$.

It can be easily observed that flow rate Q is analogous to electrical current i, and ΔP is analogous to the electrical potential drop (voltage) ΔE. In the above equation, L is the inductance (inertance) and R is the resistance to flow. Therefore, Eq. (1.5) can be rewritten as

$$Ldi/dt = \Delta \overline{E} - Ri \tag{1.8}$$

Fluid continuity equation, when integrated across the cross section, can be expressed as

$$dV/dt = \Delta Q = Q_{\text{in}} - Q_{\text{out}} \tag{1.9}$$

where V is the volume. However, volume is a function of pressure. Momentum balance for the vessel wall can be expressed as

$$P = P_{ext} + (h/a_0)\sigma \tag{1.10}$$

where P_{ext} is the external pressure on the outside of the vessel wall, h is the wall thickness, and σ is the hoop stress in the wall. The hoop stress is a function of wall radius a and modulus of elasticity E of the wall, and can be expressed as

$$\sigma = (E/2)[(a/a_0)^2 - 1] \tag{1.11}$$

where a_0 is the unstretched radius. Since the length of the segment does not change, the above equation can be expressed as

$$\sigma = (E/2)[(V/V_0) - 1] \tag{1.12}$$

where V is the volume of the vessel segment and V_0 is the unstretched volume. Equations (1.10), (1.11), and (1.12) can be combined as

$$dV/dt = CdP/dt \tag{1.13}$$

where

$$C = (2V_0\, a_0/hE) \tag{1.14}$$

C is often referred to as the compliance or capacitance.

Substituting Eq. (1.13) in Eq. (1.9) results in

$$CdP/dt = Q_{in} - Q_{out} \tag{1.15}$$

Equation (1.15) can be expressed in terms of an electrical equivalent as follows:

$$\overline{E} = (1/C)\int i dt \tag{1.16}$$

Equations (1.7) and (1.16) can be used to simulate either a segment of a blood vessel or the entire blood vessel itself. In small blood vessels, the inductance L is very low when compared to the resistance term R, and therefore, the inductance term can be neglected in small arteries, arterioles, and capillaries. Since there is no oscillation of pressure in the capillaries, the inductance term can be neglected in vessels downstream of the capillary including venules, veins, vena cava, etc. (Chu and Reddy, 1992).

An electrical analog model of the circulation in the leg is illustrated in Fig. 1.4. Let us consider the flow from the femoral artery into the small leg arteries. There is no inductance in small leg arteries, and there is only the resistance. Since the small arteries are distensible, they have capacitance (compliance). The muscular pressure (P_{MP}) acts as the external pressure on the majority of small leg arteries. Consequently, P_{MP} is used as the reference pressure across the capacitor. The arterioles do not have inductance, but have a variable resistance which is controlled by neurogenic and metabolic factors. In this model, the precapillary sphincters and the capillaries are lumped together. Since the capillaries are rather rigid, they do not have any capacitance (compliance), but the combined resistance of the sphincters and capillaries is variable subject to metabolic control. For instance, precapillary sphincters dilate in the presence of lactic acid and other end products of metabolism. Venules have resistance and a variable capacitance. This capacitance is subject to neurogenic control since the diameter of the venule is under neurogenic control. From the venules, the flow goes into leg small veins which have a resistance and a variable capacitance subject to neurogenic control. In addition, the venules have valves which only permit unidirectional flow. These valves can be modeled as diodes. Again, the reference pressure for the capacitor is the muscle pressure P_{MP}. It is well known that the blood flow in the legs is aided by the muscle pump which is essentially the external pressure oscillations on the blood vessel wall due to periodic skeletal muscle contractions during walking, etc. The muscle pump is absent in bedridden patients. Extremity pumps are used on such patients to enhance blood flow to the legs. These extremity pumps provide a periodic a graded sequential external compression of the leg. The electrical analog model shown in Fig. 1.4 can be easily modified to simulate the effect of these extremity pumps.

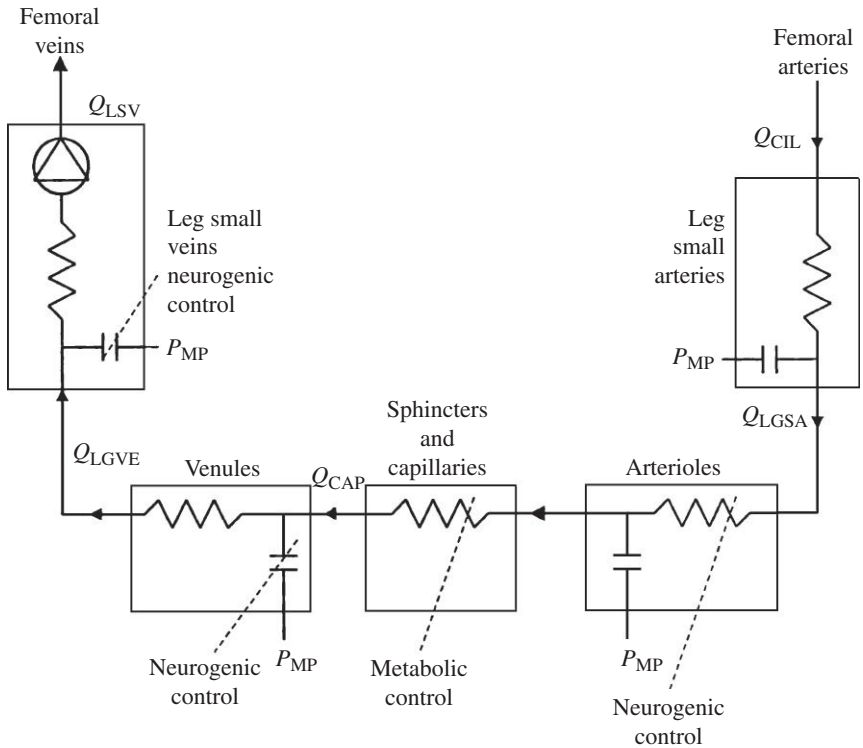

FIGURE 1.4 Electrical analog model of the circulation of the leg P_{MP} is the muscle pump which exerts a periodic external pressure on the blood vessels, Q is the flow rate, Q_{LGSA} is the flow through the leg small arteries, Q_{CAP} is the flow rate through the capillary, and Q_{LGVE} is the flow through the leg small veins. The elasticity is simulated with capacitance. The nonlinear capacitance of the leg small veins and the nonlinear resistance of arterioles and venules are under neurogenic control. The resistance of precapillary sphincters and capillaries is subject to metabolic control. The valves in the veins are simulated using diodes which permit only the unidirectional flow.

An electric analog model of pulmonary circulation is shown in Fig. 1.5. The flow is considered from node to node where the pressure is defined. The model equations for flow from compartment 1 (right ventricle) to the pulmonary arteries can be expressed by

$$L(dQ_1/dt) = P_1 - P_2 - R_1Q_1 \tag{1.17}$$

The pressure in compartment 2 can be expressed as

$$P_2 - P_{i\text{th}} = (1/C_1)\int(Q_1 - Q_2)dt \tag{1.18}$$

where $P_{i\text{th}}$ is the intrathoracic pressure, which is pressure acting on the outside of the pulmonary vessels. Similarly,

$$P_2 - P_3 = R_2Q_2 \tag{1.19}$$

$$P_3 - P_{i\text{th}} = (1/C_2)\int(Q_2 - Q_3)dt \tag{1.20}$$

$$P_3 - P_5 = R_3Q_3 \tag{1.21}$$

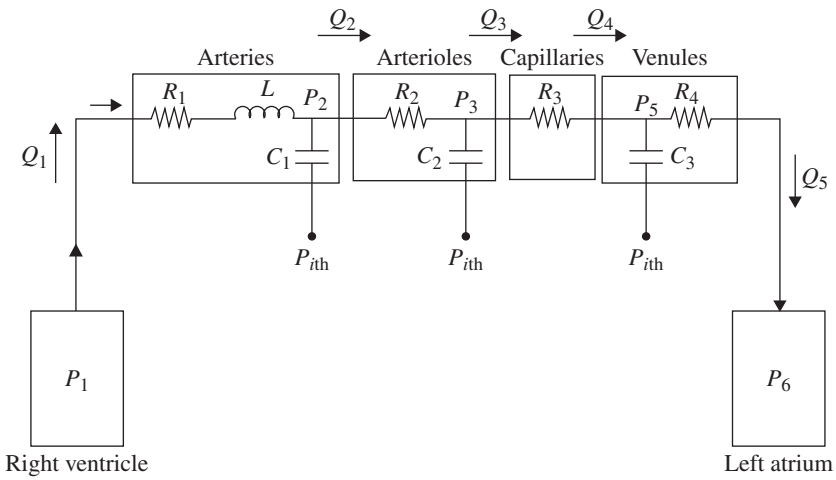

FIGURE 1.5 A model of pulmonary circulation. P_{ith} is the intrathoracic pressure which is the external pressure on the pulmonary blood vessels.

$$P_5 - P_{ith} = (1/C_3)\int(Q_4 - Q_5)dt \tag{1.22}$$

$$Q_3 = Q_4 \tag{1.23}$$

$$P_5 - P_6 = R_4 Q_5 \tag{1.24}$$

The capacitance is due to distensibility of the vessel. The capillaries are stiffer and less distensible, and therefore have minimal capacitance.

Electrical analog models have been used in the study of cardiovascular, pulmonary, intestinal, and urinary system dynamics. Recently, Barnea and Gillon (2001) have used an electrical analog model to simulate flow through the urethra. Their model consisted of a simple L, R, C circuit with a variable capacitor. The time varying capacitor simulated the time-dependent relaxation of the urethra. They used two types of resistance: a constant resistance to simulate Poiseouille-type viscous pressure drop and a flow-dependent resistance to simulate Bernoulli-type pressure loss. With real-time pressure-flow data sets, Barnea and Gillon (2001) have used the model to estimate urethral resistance and changes in urethral compliance during voiding, and have suggested that the urethral elastance (inverse of compliance) estimated by the model provides a new diagnostic tool. Ventricular and atrial pumping can be modeled using similar techniques. The actual pump (pressure source) can be modeled as a variable capacitor. Figure 1.6 shows a model of the left heart with a multisegment representation of the ventricle (Rideout, 1991). Kerckhoffs et al. (2007) have coupled an electrical analog model of systemic circulation with a finite element model of cardiac ventricular mechanics.

1.3 MECHANICAL MODELS

Mechanical models consisting of combinations of springs and dashpots are very popular in numerous disciplines. Spring dashpot models have been used to model the mechanical behavior of viscoelastic materials and can be used to represent the one dimensional behavior of tissue and other biological materials. In a linear spring, the force is proportional to the change in length or the strain.

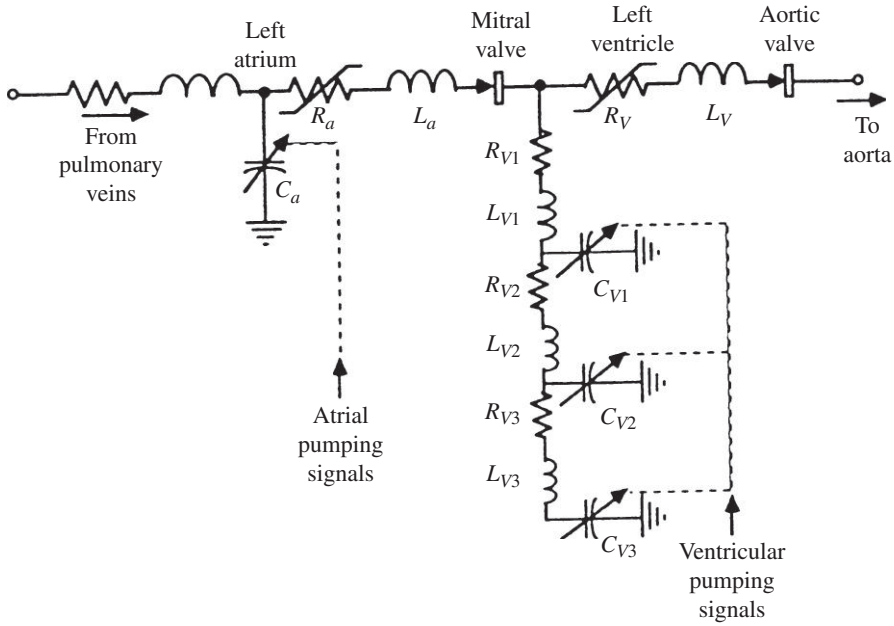

FIGURE 1.6 Electrical analog model to simulate atrial and ventricular pumping. Variable capacitances simulate the muscle contractions, and the filling and emptying through the ventricle can be simulated by a series of inductance and resistance elements. [*Adapted from Rideout (1991).*]

On the other hand, the force in a dashpot is proportional to the rate of change in strain. Consider a mass supported by a spring and a dashpot in parallel. Let a force F be acting on the mass. Force in a dashpot is $b(dX/dt)$ since the force in a fluid depends on strain rate. Here, b is a constant. The force in the spring is given by kX, where k is the spring constant.

Application of Newton's law results in

$$m(d^2X/dt^2) + b(dX/dt) + kX = F \tag{1.25}$$

where X is the elongation or change in length with respect to the steady-state value, b is the constant of the dashpot, and k is the spring constant.

It should be pointed out that the above mechanical equation is similar to the following electrical equation:

$$L(di/dt) + Ri + (1/C)\int i dt = E \tag{1.26}$$

where L is the inductance, R is the resistance, i is the current, and E is the voltage. This equation can be expressed in terms of the charge q instead of the current as

$$L(d^2q/dt^2) + R(dq/dt) + (1/C)q = E \tag{1.27}$$

Equations (1.25), (1.26), and (1.27) are similar. Therefore, mass is analogous to the inductance, the dashpot is analogous to the resistor, and the spring is analogous to the capacitor. The spring and the capacitor are storage units, whereas the dashpot and the resistor are the dissipaters of energy. The charge is analogous to the deformation or elongation, the current is similar to the velocity, and force is analogous to the voltage. Therefore, any electrical system can be modeled using mechanical analogs and any mechanical system can be modeled using electrical analogs.

Lumped mechanical models have been used to analyze the impact dynamics. Generally, muscle is represented by a combination of a spring and a dashpot, whereas a ligament is modeled using a spring.

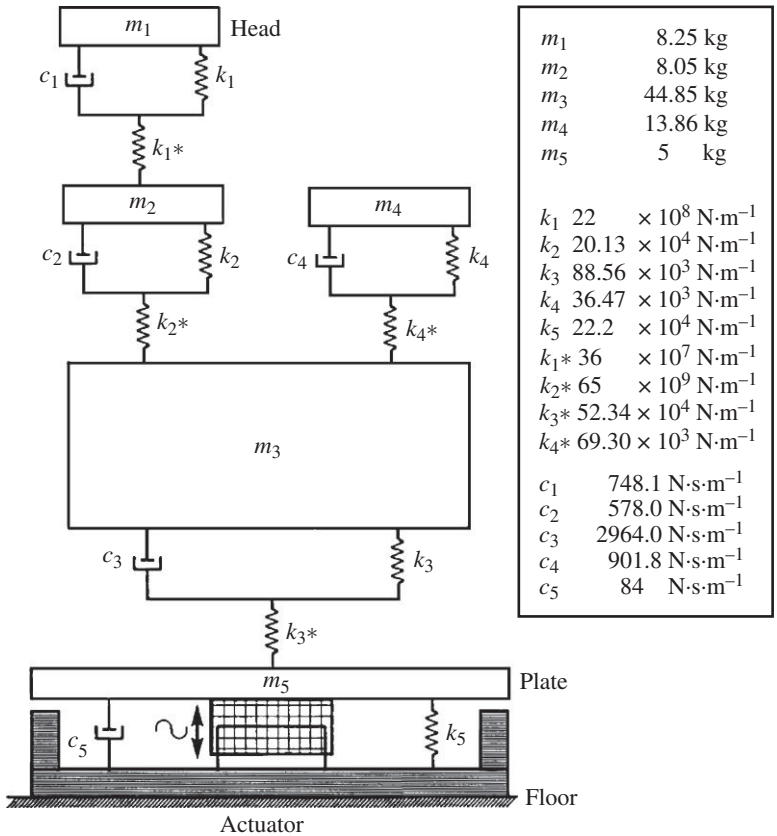

FIGURE 1.7 A lumped mechanical analog model for the analysis of vibration in relaxed standing human. [*Reproduced with permission from Fritton et al. (1997).*]

Human body vibrations can be analyzed using similar lumped models. Fritton et al. (1997) developed a lumped parameter model (Fig. 1.7) to analyze head vibration and vibration transmissibility in a standing individual. The model results are in good agreement with the experimental results. Such models are useful in the design of automobile seat cushion, motorcycle helmet design, etc.

1.4 MODELS WITH MEMORY AND MODELS WITH TIME DELAY

Time delay and memory processes occur in several biomedical disciplines. An example of such an application occurs in modeling of the immune system (Reddy and Krouskop, 1978). In cellular immune response, lymphocytes are sensitized to a foreign material and have memory. The immune response is significantly enhanced if the similar material is reintroduced after certain lag time. Another example could be the application to stress-induced bone remodeling. Modeling of the nervous system would involve time delays and memory. Similar hereditary functions are used to describe the material responses of viscoelastic materials. The effect of environmental pollutants can be modeled using such hereditary functions. Stress-induced bone remodeling involves time lags between the actual application of stress and actual new bone formation, and also involves

stress/strain histories. To illustrate the modeling of the effects of memory and time delay, let us consider a model to predict the number of engineers in the United States. Then we will consider a model of cell-mediated immunity which has similar delays and memory functions.

1.4.1 A Model to Predict the Number of Engineers in the United States

An easy-to-understand example of a deterministic model with time delay and memory is a model to predict the number of biomedical engineers in the United States at any given time. Let us restrict our analysis to a single discipline such as biomedical engineering. Let E be the number of engineers (biomedical) at any given time. The time rate of change of the number of engineers at any given time in the United States can be expressed as

$$dE/dt = G + I - R - L - M \tag{1.28}$$

where G represents the number of graduates entering the profession (graduating from an engineering program) per unit time, I represents the number of engineers immigrating into the United States per unit time, R represents the number of engineers retiring per unit time, L represents the number of engineers leaving the profession per unit time (e.g., leaving the profession to become doctors, lawyers, managers, etc.), and M represents the number of engineers dying (before retirement) per unit time.

In Eq. (1.28), we have lumped the entire United States into a single region (a well-stirred compartment) with homogeneous distribution. In addition, we have not made any discrimination with regard to age, sex, or professional level. We have considered the entire pool as a well-stirred homogeneous compartment. In reality, there is a continuous distribution of ages. Even with this global analysis with a lumped model, we could consider the age distribution with a series of compartments with each compartment representing engineers within a particular age group. Moreover, we have assumed that all engineering graduates enter the workforce. A percentage of them go to graduate school and enter the workforce at a later time.

The number of graduates entering the profession is a function of the number of students entering the engineering school 4 years before:

$$G(t) = k_1 S(t - 4) \tag{1.29}$$

where $S(t)$ is the number of students entering the engineering school per unit time. The number of students entering the engineering school depends on the demand for the engineering profession over a period of years, that is, on the demand history.

The number of engineers immigrating into the United States per unit time depends on two factors: demand history in the United States for engineers and the number of visas that can be issued per unit time. Assuming that immigration visa policy is also dependent on demand history, we can assume that I is dependent on demand history. Here we have assumed that immigrants from all foreign countries are lumped into a single compartment. In reality, each country should be placed in a separate compartment and intercompartmental diffusion should be studied.

The number of engineers retiring per unit time is proportional to the number of engineers in the profession at the time:

$$R(t) = k_2 E(t) \tag{1.30}$$

The number of engineers leaving the profession depends on various factors: the demand for the engineering profession at that time and demand for various other professions at that time as well as on several personal factors. For the purpose of this analysis, let us assume that the number of engineers leaving the profession in a time interval is proportional to the number of individuals in the profession at that time:

$$L(t) = k_3 E(t) \tag{1.31}$$

The number of engineers dying (before retirement) per unit time is proportional to the number of engineers at that time:

$$M(t) = k_4 E(t) \tag{1.32}$$

The demand for engineers at any given time is proportional to the number of jobs available at that time ($J(t)$) and is inversely proportional to the number of engineers available at that time:

$$D(t) = kJ(t)/E(t) \tag{1.33}$$

The number of jobs available depends on various factors such as government spending for R&D projects, economic growth, sales of medical products, number of hospitals, etc. Let us assume in this case (biomedical engineering) that the number of jobs is directly proportional to the sales of medical products (p), directly proportional to government spending for health care R&D (e), and directly proportional to the number of new medical product company startups (i):

$$J(t) = (k_6 e + k_7 c + k_8 i + k_9 + kp) \tag{1.34}$$

Although we assumed that the number of jobs at the present time is dependent on $e(t)$, $c(t)$, $h(t)$, $i(t)$, and $p(t)$, in reality the number of jobs at present may depend on previous values of these parameters, or on the history of these parameters.

Let us now analyze the demand history. This history depends on the memory function. Let us assume that the effect of demand existing at a time decays exponentially (exponentially decaying memory). The net effect of demands from time = 0 to t can be expressed as

$$H_1(t) = {}_{\tau=0} I^{\tau=t}\{D(\tau)\exp[-k_{10}(t-\tau)]\}d\tau \tag{1.35}$$

The number of students entering the engineering school per unit time is

$$S(t) = k_{11} H_1(t) \tag{1.36}$$

Immigration rate can similarly be expressed as

$$I(t) = k_{12} H_2(t) \tag{1.37}$$

where

$$H_2(t) = {}_{\tau=0} I^{\tau=t}\{D(\tau)\exp[-k_{13}(t-\tau)]\}d\tau \tag{1.38}$$

H_1 and H_2 are called hereditary functions. Instead of an exponential decay of memory, we could have a sinusoidal or some other functional form of memory decay, depending on the physical situation.

$$dE/dt = k_1 k_{10} H_1(t-4) + k_{11} H_2(t) - (k_2 + k_3 + k_4)E(t) \tag{1.39}$$

In this analysis, making various assumptions, we have formulated a lumped parameter deterministic model to predict the number of engineers (biomedical) present in the United States at any given time. If we want to know the geographical distribution, we can take two approaches. We can divide the entire United States into a number of compartments (e.g., northeast, east, west, etc.) and study the intercompartmental diffusion. Alternatively, we can make E a continuous variable in space and time I (x, y, t) and account for spatial diffusion.

1.4.2 Modeling the Cell-Mediated Immunity in Homograft Rejection

In cell-mediated immunity, lymphocytes in the tissue become sensitized to the target (graft) cells and travel to the regional lymph nodes where they initiate an immunological response by increasing the production of immunocompetent lymphocytes. The newly produced lymphocytes are then transported

into the blood stream via the thoracic duct. Lymphocytes recirculate from the blood stream through the tissue and return to the blood stream via the lymphatic system. When foreign cells are introduced into the tissue, blood lymphocytes migrate into the tissue at an increased rate and bring about the destruction of the target cells. Lymphocytes have memory and they exhibit an increased secondary response, e.g., if after the rejection of the first graft, a second graft is introduced into the host, the second graft is rejected much faster. A similar situation occurs in delayed hypersensitivity, which is another cell-mediated reaction. In this analysis, let us assume that blood and tissue are well-stirred compartments and that the newly produced lymphocytes are introduced into the blood compartment (Reddy and Krouskop, 1978).

For sensitization to occur, a lymphocyte has to come in contact with a target cell. The number of lymphocytes becoming sensitized at any given time ($L_s(t)$) is a function of the number of lymphocytes in the tissue ($L_T(t)$) and the number of target (foreign) cells ($g(t)$)

$$L_s(t) = C_1 L_T(t) g(t) \tag{1.40}$$

Certain lymphocytes, upon encountering target cells, are transformed into memory cells. The memory cell formation depends upon the number of lymphocytes in the tissue and the number of target cells. The number of memory cells formed at any time (t) may thus be expressed as

$$L_{\text{ms}}(t) = C_1 L_T(t) g(t) \tag{1.41}$$

Sensitized lymphocytes stimulate the production of immunocompetent lymphocytes and the effect of each sensitized cell lasts for a given period of time. For the purpose of the present analysis, it is assumed that the effect of each sensitized lymphocyte decays exponentially over a period of time. The production rate of blood lymphocytes at any time (t) due to the primary response $(dL_B/dt)_{\text{prim}}$ would then be equal to the sum of the residual effect of all the lymphocytes sensitized between time 0 and time $t - \Phi_1$, where Φ_1 is the time lag between sensitization and production of the lymphocytes.

The number of lymphocytes produced due to primary response between time t and time $(t - \Phi_1)$ would be

$$L_B(t) - L_B(t - \Delta t) = C_3\{L_S(t - \Phi_1)\Delta t + L_S(t - \Phi_1 - \Delta t)\Delta t + L_S(t - \Phi_1 - 2\Delta t)e^{-K1\Delta t}\Delta t$$

Due to lymphocytes sensitized at $t - \Phi_1$ | Due to lymphocytes sensitized at $t - \Phi_1 - \Delta t$ | Due to lymphocytes sensitized at $t - 2\Phi_{1-} - \Delta t$

$$+ L_S(t - \Phi_1 - r\Delta t)e^{-K1\, r\Delta t}\Delta t + \cdots\} \tag{1.42}$$

Due to lymphocytes sensitized at $t - \Phi_{1-} = r\Delta t$

$$= C_3 3 L_S(t - \Phi_1 - r\Delta t)e^{-K1\, r\Delta t}\Delta t \tag{1.43}$$

Dividing by Δt and taking the limits as $\Delta t \,\Pi\, 0$, the left-hand side becomes a derivative and the right-hand side can be represented as an integral in terms of the hereditary function

$$(dL_B(t)/dt)_{\text{primary}} = C_3 \int_0^{t - \Phi 1} L_S(\tau) e^{-K1(t - \Phi_1 - \tau)} d\tau \tag{1.44}$$

Substituting for L_S in terms of L_T,

$$(dL_B(t)/dt)_{\text{primary}} = k_2 \int_0^{t - \Phi 1} L_T(\tau) e^{-K1(t - \Phi_1 - \tau)} d\tau \tag{1.45}$$

For the secondary response to appear, a memory cell must encounter a target cell, and therefore the secondary response depends upon the number of memory cells and the number of target cells. Similar to (Eq. 1.45), Reddy and Krouskop (1978) expressed the secondary response in terms of a hereditary function

$$(dL_B(t)/dt)_{\text{secondary}} = k_3 \int_0^{t - \Phi 2} L_T(\tau) g(\tau) e^{-K4(t - \Phi_2 - \tau)} g(t - \Phi_3) d\tau \tag{1.46}$$

In developing the above equation, it is assumed that the effect of a memory cell also decays exponentially over a period of time. Thus the production rate of blood lymphocytes at time (t) due to secondary response $(dL_B/dt)_{\text{second}}$ is due to the sum of the residual effects of all the memory cells formed between 0 and time $t - \phi_2$, where ϕ_2 is the time lag between memory cell formation and the appearance of the secondary response.

The net rate in change of blood lymphocytes may then be described as

$$dL_B/dt = \underbrace{k_2 \int_0^{t-\Phi_1} L_T(\tau) e^{-K1(t-\Phi_1-\tau)} d\tau}_{\text{Due to primary response}} + \underbrace{k_3 \int_0^{t-\Phi_2} L_T(\tau) g(\tau) e^{-K4(t-\Phi_2-\tau)} g(t-\Phi_3) d\tau}_{\text{Due to secondary response}}$$

$$\underbrace{+ K_5 L_T}_{\text{Recirculation}} \quad \underbrace{- K_6 L_B}_{\text{Death in the blood}} - K_7 L_B \quad \underbrace{- K_8 L_B g}_{\substack{\text{Migration into tissue due to} \\ \text{target cell presence}}} \tag{1.47}$$

The rates of change of tissue lymphocytes and the number of target cells can be described using similar mass balance

$$dL_T/dt = \underbrace{K_8 L_B g}_{\text{Increased migration}} \quad \underbrace{- K_5 L_T + K_6 L_B}_{\text{Recirculation}} \quad \underbrace{- K_9 L_T g}_{\text{Loss due to target cell destruction}} \tag{1.48}$$

$$dg/dt = (dg/dt)_{\text{input}} - k_{10} L_T g \tag{1.49}$$

These equations were simulated by Reddy and Krouskop. Figure 1.8 shows the production rate of lymphocytes and the number of target cells when the target cells were introduced on day 0 and again on day 4. Figure 1.9 shows the production rate of lymphocytes and the number of target cells present in the tissue when the target cells were introduced continuously.

1.5 ARTIFICIAL NEURAL NETWORK MODELS

Neural network models represent the black box type of model. These models are used where the precise functioning of the system is not understood but the sample input-output data are known. Neural networks represent a new generation of information processing systems that are artificially (virtually) constructed to make use of the organizational structure of the neuronal information processing in the brain. A neural network consists of several interconnecting neurons also called as nodes. These nodes are organized into an input layer, one or more hidden layers, and an output layer. The number of input layer nodes in the input layer depends on the number of input variables. The number of nodes in the output layer is determined by the number of output parameters. Each input parameter is represented by a node in the input layer, and the each output parameter is represented by a node in the output layer. The number of nodes in the hidden layer could be variable. Each node in a feed-forward neural network is connected to every node in the next level of nodes. That means each input node is connected to all the nodes in the hidden layer neurons. Let us, for simplicity, consider only one hidden layer. Now, each node in the hidden layer is connected to all the nodes in the output layer. Figure 1.10 shows a network with four output nodes and five input nodes with four hidden layer nodes. The connection strengths are determined by the weights.

Let us assume that W_{ij} represents the weight of the connection from the jth node in the hidden layer to the ith node in the output layer, and let us assume that $w_{j,k}$ represents the weight of connection from the kth input node to the jth node in the hidden layer. Let X_k represents the value of kth input node.

The sum of weighted inputs to the jth node in the hidden layer is

$$I_j = \Sigma w_{j,k} X_k \tag{1.50}$$

In other words, $I_1 = w_{1,1}X_1 + w_{1,2}X_2 + w_{1,3}X_3 + w_{1,4}X_4$, where X_1, X_2, X_3, and X_4 are the values of the four input parameters.

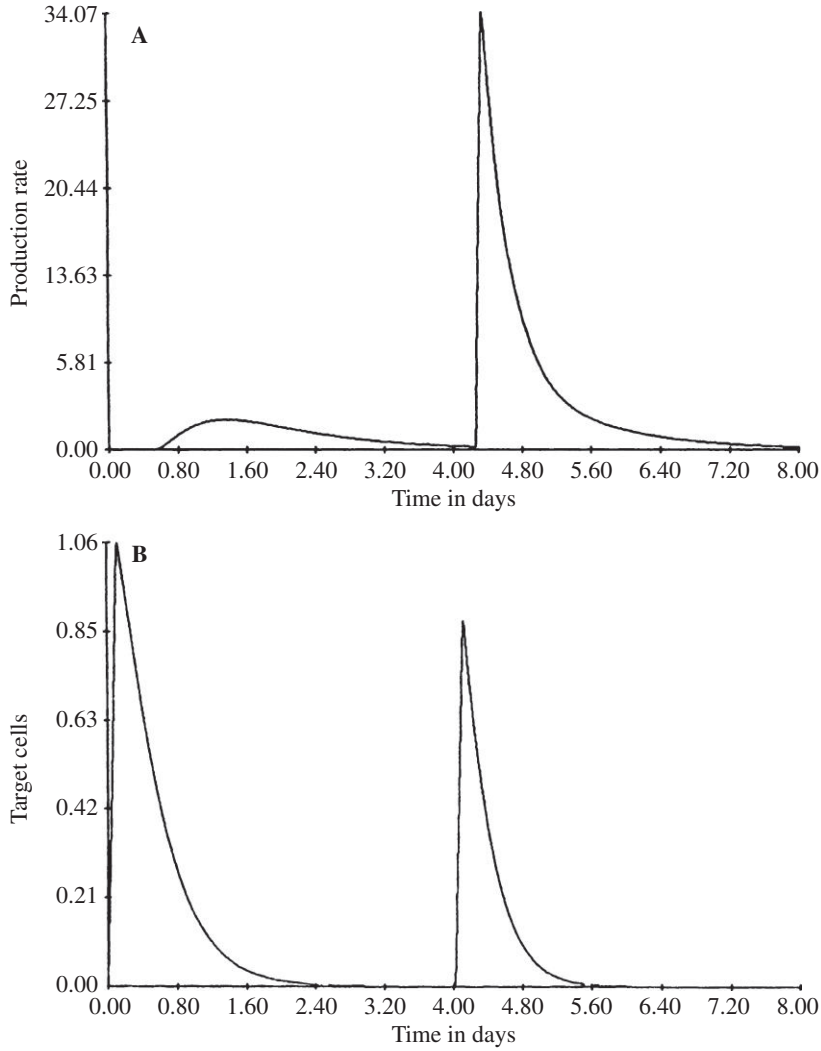

FIGURE 1.8 The simulation results of the production rate of lymphocytes (A), and the number of target cells or foreign cells (B) plotted as a function of time in days. In the simulation, the target cells were introduced on day 0 and again on day 4. [*Reddy and Krouskop (1978).*] Note the increased secondary response.

The output of a hidden layer neuron is a function of its input

$$H_j = f(I_j) \tag{1.51}$$

This function f is called the activation function. An example of this function is the sigmoid function

$$H_j = k[2/(1 + \exp(-aI_j + B) - 1] \tag{1.52}$$

where k, a, and B are constants. B is called the bias. B can be a zero. In general, any monotone, nondecreasing differentiable signal function can be used as the activation function.

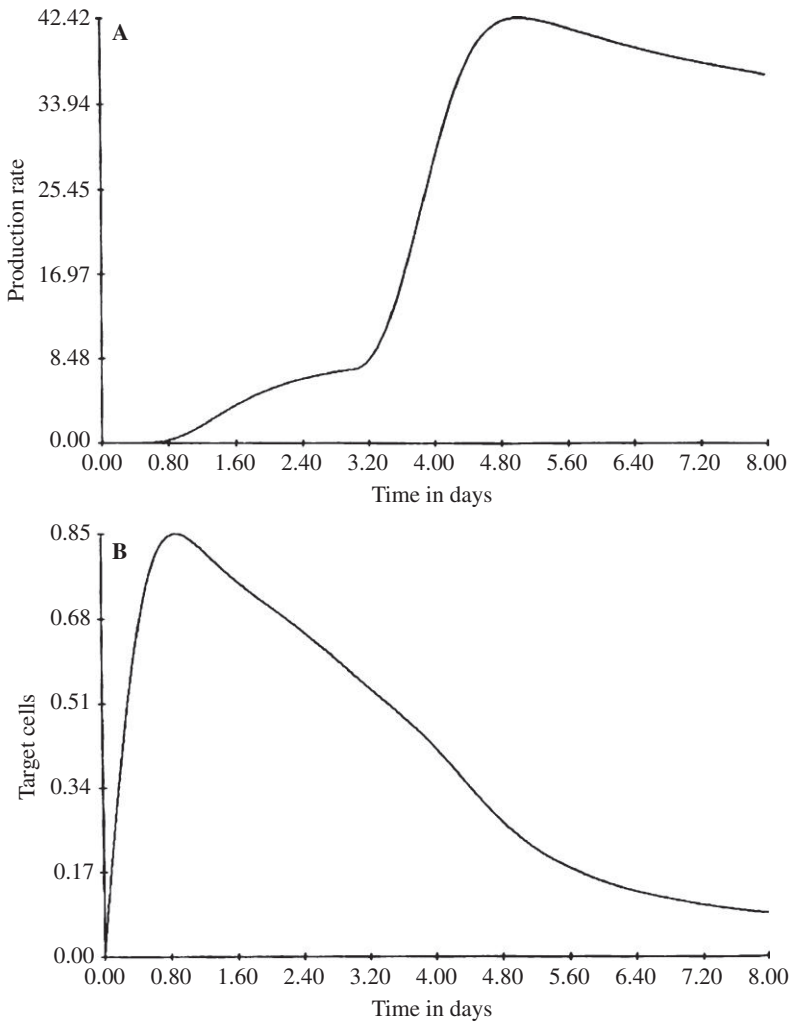

FIGURE 1.9 The simulation results of the production rate of lymphocytes (A), and the number of target cells or foreign cells (B) when the target cells were continuously introduced. [*Reddy and Krouskop (1978).*]

The input G_i to the ith node in the output layer is the sum of its weighted inputs.

$$G_i = \Sigma W_{i,j} H_j \tag{1.53}$$

The output of the node in the output layer is some function of the input node.

$$Y_i = F(G_i) \tag{1.54}$$

The activation function F of the output neurons can be any monotone, nondecreasing differentiable function. Sigmoid or logistic functions are usually used.

If the weights $w_{j,k}$, and $W_{i,j}$ are all known, then given the input X_k, the output Y_i of the system can be calculated. The weights are determined through a training algorithm using the sample input-output data.

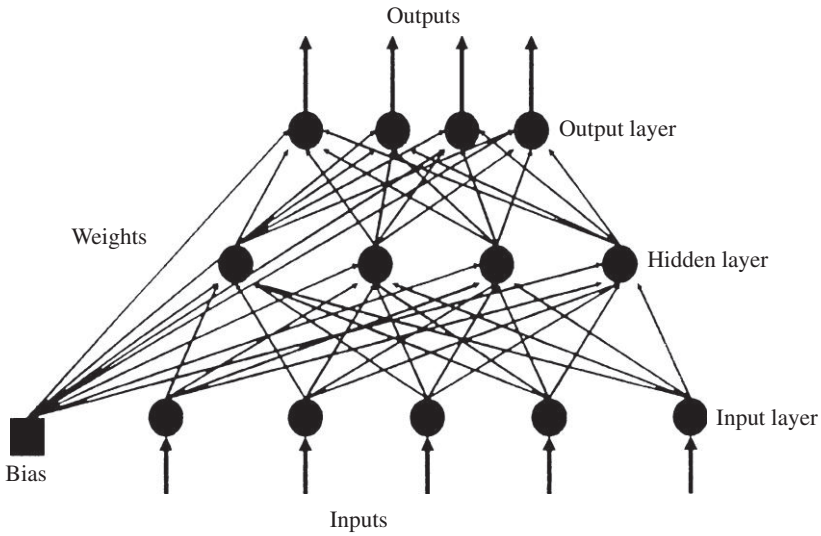

FIGURE 1.10 A neural network model consists of several input layer neurons (nodes), one or more neurons in the output layer, and one or more layers of hidden layer neurons each consisting of several neurons. Each neuron in the input layer corresponds to an input parameter, and each neuron in the output layer corresponds to an output parameter. Each neuron in a layer is connected to each of the neurons in the next level. In this example only one hidden layer is used. Each of the input neurons is connected with each neuron in the hidden layer and each neuron in the hidden layer is connected to each neuron in the output layer. The connection strengths are represented by weights.

There are several training techniques and the most popular technique is the back propagation technique. Let us assume that for a set of sample inputs X_k, we know the actual outputs $d_{i.}$ Initially, we do not know the weights, but we could have a random initial guess of the weights $w_{j,k}$ and $W_{i,j}$. As an example, we could define all weights initially to be $w_{j,k} = W_{i,j} = 0.2$ or 0.5. Using the above equations along with the sample input vector $\mathbf{X}_k$, we can calculate the output of the system Y_i. Of course, this calculated value is going to be different from the actual output (vector if there is more than one output node) value d_i, corresponding to the input vector $\mathbf{X}_k$. The error is the difference between the calculated output value and the actual value. There are various algorithms to iteratively calculate the weights, each time changing the weights as a function of the error. The most popular of these is the gradient descent technique.

The sum of the error in the mth iteration is defined as

$$e_i^m = d_i - y_i^m = d_i - F(\Sigma W_{i,j}^m H_j) = d_i - F(\Sigma W_{i,j}^m f(\Sigma w_{j,k}^m X_k) \tag{1.55}$$

The instantaneous summed squared error at an iteration m, corresponding to the sample data set n, can be calculated as

$$E_n^m = (1/2)\ \Sigma(e_i^m)^2 \tag{1.56}$$

The total error E at each iteration, for all the sample data pairs (input-output), can be calculated as the sum of the errors E_n for the individual sample data.

Adjusting the weights for each iteration for connections between the hidden layer neurons and the output layer neurons $W_{i,j}$ can be calculated as

$$W_{i,j}^{m+1} = W_{i,j}^m - \eta(\delta\ E^m/\delta W_{i,j}) \tag{1.57}$$

where η is the learning rate.

The error gradient can be expressed as

$$(\delta E^m/\delta W_{i,j}) = (\delta E^m/\delta F)(dF/dW_{i,j}) \tag{1.58}$$

For a sigmoid function (F), it turns out that the differential is a simple function of the sigmoid as follows:

$$dF = b(1 - F)F \tag{1.59}$$

where b is a constant. Thus,

$$(dF/dW_{i,j}) = b(1 - F(W_{i,j}))F(W_{i,j}) \tag{1.60}$$

For adjusting the weights for connections between the input and the hidden layer neurons, the error is back propagated by calculating the partial derivative of the error E with respect to the weights $w_{j,k}$ similarly (Haykin, 1999).

The whole process of calculating the weights using the sample data sets is called the training process. There is a neural network package in the MATLAB which can be easily used in the training process. There are several algorithms in the package, including the back propagation, modified back propagation, etc. which the user can choose in the MATLAB software. Once the weights are calculated using MATLAB or any other software, it becomes a matter of obtaining the output vector for a given input vector using matrix multiplications. The most important aspect of a neural network is that it should be tested with data not used in the training process.

Neural networks have been used for classification and control. For instance, Reddy et al. (1995) used neural networks to classify the degree of the disease in dysphagic patients using noninvasive measurements (of throat acceleration, swallow suction, pressure, etc.) obtained from dysphagic patients during swallowing. These measurements were the inputs to the network and the outputs were normal, mild, moderate, and severe. Neural network performance depends on the sample data, initial weights, etc. Reddy et al. (1995) trained several networks with various initial conditions and activation functions. Based on some initial testing with known data, they recruited the best five networks into a committee. A majority opinion of the committee was used as the final decision. For classification problems, Reddy and Buch (2003) and Das et al. (2001) obtained better results with committee of neural networks (Fig. 1.11) when compared to a single network, and the majority opinion of the committee was in agreement with clinical or actual classification.

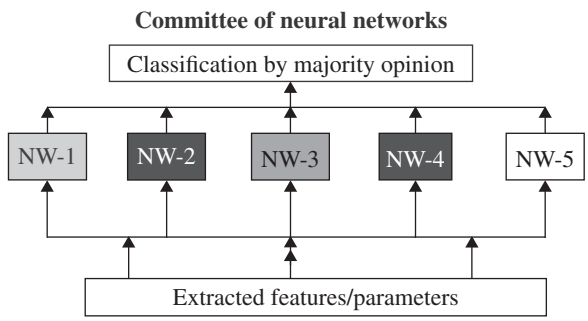

FIGURE 1.11 The committee of neural networks. Each of the input parameters is simultaneously fed to several networks working in parallel. Each network is different from the others in terms of initial training weights or the activation function (transfer function at the nodes). A majority opinion of the member networks provides the final decision of the committee. This committee of networks simulates the parliamentary process, and emulates a group of physicians making the decision. [*Reddy and Buch (2000).*]

1.6 FUZZY LOGIC

Most real world systems include some element of uncertainty and cannot be accurately modeled using crisp logic. Moreover, some of these systems require modeling of parameters like human experience, intuition, etc., which involve various degrees of conceptual uncertainties and vagueness. Several of these applications require fuzzy definition of boundaries and fuzzy classes, whose membership values are in the form of degree of membership, rather than in the form of true or false. In crisp logic, a parameter can belong to only one class. However, in fuzzy logic, a parameter could belong to more than one class at the same time. Let us assume that we want to classify the age of a person into two classes: young and old. In crisp logic, the individual belongs to either old or young, as the crisp logic requires a clear boundary between the two classes. In fuzzy logic, the individual can belong to both classes at the same time. Figure 1.12 provides a general comparison of a crisp and a fuzzy variable and its membership to subsets. The variable x is divided into two subsets "young" and "old." For example, if $x = 40$, the crisp classification would be "young." On the other hand, the fuzzy classification would be (0.7, 0.3), indicating that the individual belongs to both classes (70 percent young and 30 percent old). The individual's membership to the subset "young" would be 0.7, and his membership to subset "old" would be 0.3. The function which defines the boundaries of the domains (or subsets) is called "membership function" and the values 0.7 and 0.3 are called the membership values.

Overall scheme of the fuzzy logic system is shown in Fig. 1.13. In the fuzzy logic, each measured parameter is fuzzified by calculating the membership values to various subsets using predefined membership functions. The membership values for the parameters are then sent to a predefined rule base to provide a fuzzy output. The fuzzy output is then defuzzified using a defuzzification scheme. Usually, the centroid defuzzification is used to come up with a crisp output. The first step in designing a fuzzy logic system is to first define the number of subsets, and the membership functions which define the subset domains. The membership functions can be linear, triangular, trapezoidal, or sigmoidal, or can be of irregular geometry. Fuzzy logic can be used for classification (Suryanarayanan et al., 1995; Steimann and Adlassnig, 1998; Sproule et al., 2002; Sakaguchi et al., 2004; Mangiameli et al., 2004) and control problems (Suryanarayan and Reddy, 1997; Kutava et al., 2006). Examples of both of these are presented below.

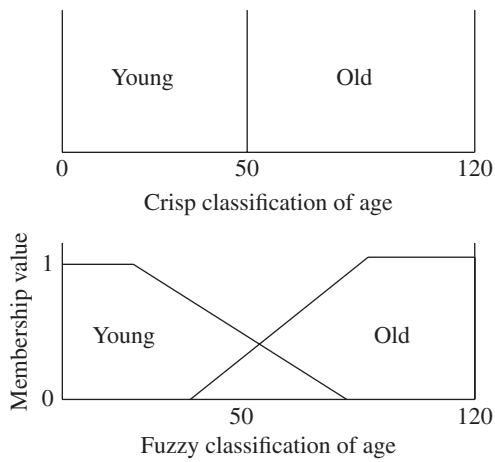

FIGURE 1.12 A comparison of crisp logic and fuzzy logic. In crisp logic a variable (age of an individual in this example) belongs to a single subdomain. In fuzzy logic, a variable can belong to a number of subdomains with varying degree of membership value. The domain boundaries are defined by membership functions.

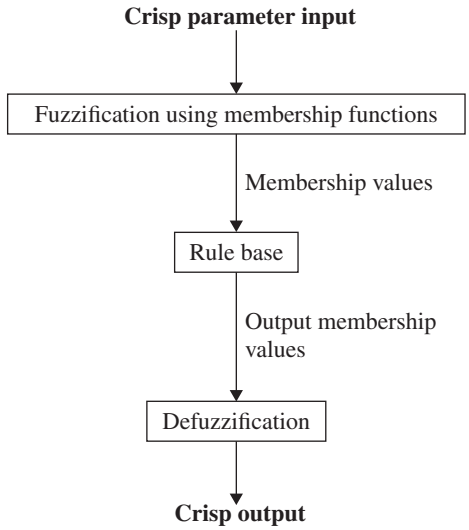

FIGURE 1.13 The overall scheme of fuzzy logic. The measured crisp parameters are first fuzzified with the aid of membership functions to obtain the membership value to various subdomains. The membership values are then subjected to a predefined rule base. The output of the rule base is in the form of output membership values to various subdomains. This fuzzy output is then defuzzified to obtain a crisp output.

1.6.1 Fuzzy Classification of Risk for Aspiration in Dysphagic Patients

Suryanarayanan et al. (1995) developed a fuzzy logic diagnosis system to classify the dysphagic patient into "normal, mild, moderate, and severe dysphagia" based on several parameters measured from the dysphagic subject. Dysphagia denotes dysfunction of the swallowing mechanism and presents a major problem in the rehabilitation of stroke and head injury patients. Dysfunction of the pharyngeal phase of swallow can lead to aspiration, chocking, and even death. Consequently, the assessment of risk for aspiration is important from a clinical point of view. Reddy et al. (1990, 1991, 1994) have identified and developed instrumentation and techniques to noninvasively quantify various biomechanical parameters that characterize the dysphagic patient and clinically evaluated the technique by correlating with the videofluorography examination (Reddy et al., 2000). For the assessment of the pharyngeal phase, they have placed an ultraminiature accelerometer at the throat at the level of thyroid cartilage and asked the patient to elicit a swallow (Reddy et al., 1991, 1994). Swallowing in normal subjects gave rise to a characteristic acceleration pattern which was distorted or absent in dysphagic individuals. In addition to the acceleration measurements, they measured swallow suction pressure (with a catheter placed toward the posterior aspect of the tongue), and the number of attempts to swallow before eliciting a swallow response, etc. Suryanarayanan et al. (1995) fuzzified these measurements by defining membership functions for each of these parameters (magnitude of acceleration, swallow pressure, and number of attempts to swallow) which defined the four subdomains (severe risk, moderate risk, mild risk for aspiration, and normal) for each of these parameters (Fig. 1.14). Membership functions were constructed using the function

$$\mu = 1/(\exp(\alpha x + \beta) \tag{1.61}$$

where μ is the membership value, x is the measured parameter, and α and β are constants. The slope of the sigmoid function can be changed by changing the value of α.

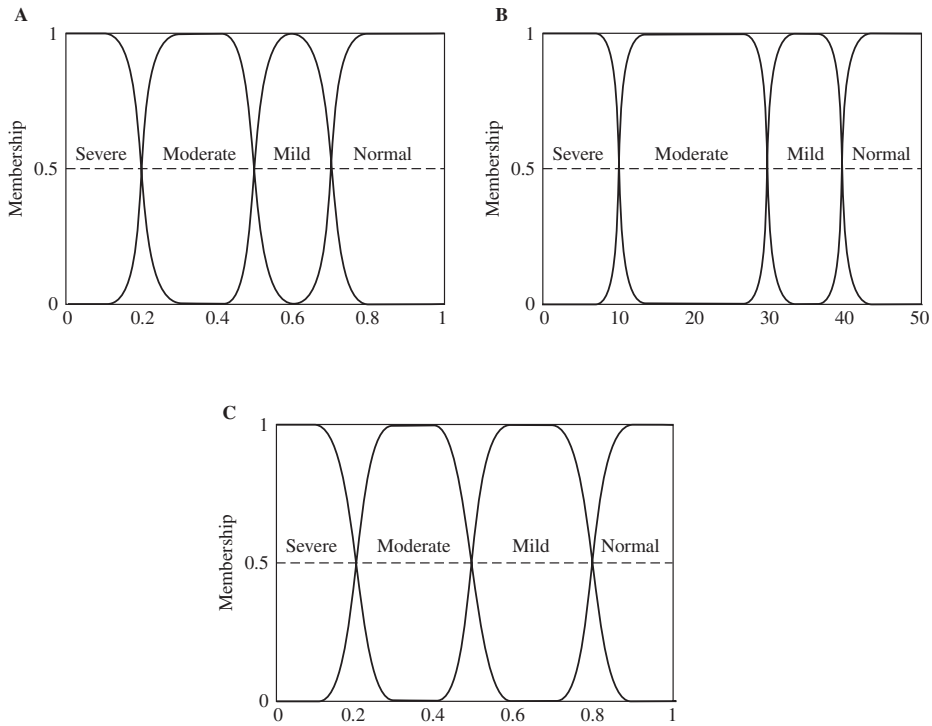

FIGURE 1.14 The membership functions are used to compute the membership values corresponding to the measured parameters: (A) membership functions for the acceleration magnitude (peak-to-peak value); (B) membership functions for the peak swallow pressure; and (C) the measured number of attempts to swallow. The variable number of attempts to swallow n, was transformed by a linear transfer function given by defining a new variable $f = (8 - n)/7$ and membership functions defined on the new variable f. [*Reproduced with permission from Suryanarayanan et al. (1995).*]

The membership values are represented as $\mu_{i,j}$, where i represents the parameter and j represents the subset. Corresponding to each parameter a, we have a vector of membership values given by

$$\mu_a(X) = [\mu_{a,1}(x), \mu_{a,2}(x), \mu_{a,3}(x), \mu_{a,4}(x)] \tag{1.62}$$

where a is the parameter, $\mu_{a,j}$ is the membership to the jth subset corresponding to parameter a. For instance, $\mu_{a,1}$ refers to membership value related to severe distortion in the acceleration magnitude, $\mu_{a,2}$ refers to moderate distortion in acceleration magnitude, $\mu_{a,3}$ refers to mild distortion in acceleration magnitude, and $\mu_{a,4}$ refers to normal acceleration magnitude. Similarly, the elements in $\mu_{b,1}$, $\mu_{b,2}$, $\mu_{b,3,}$ and $\mu_{b,4}$ refer to severe distortion in the swallow pressure magnitude, moderate distortion in swallow pressure magnitude, mild distortion in swallow pressure magnitude, and normal swallow pressure.

For each patient, the measured parameter values are fuzzified to obtain the four membership values for each of the three parameters. These values are then fed to a rule base R. The function R is defined as a relation between the parameters defined. A typical relation between input and output parameters is the IF-THEN rule. The output is given by the rule set R acting upon the membership values computed for the measured parameters and is represented as

$$\Phi = R \cdot [(\mu_a(x), \mu_b(y), \mu_c(z)] \tag{1.63}$$

where subscript a refers to the parameter "acceleration magnitude" and x is the corresponding value, subscript b refers to the parameter "swallow pressure magnitude" and y is the corresponding value of this parameter, and subscript c refers to the parameter "number of attempts to swallow" and z is the corresponding value of this parameter. The rule base can be in the form of a lookup table or in

the form of a minimum or maximum operation. Since the dysphagia classification involves estimating the severity of the disease, Suryanarayanan et al. (1995) used the maximum rule.

Using the maximum rule, the output membership value corresponding to the severe risk (Φ_1) can be calculated as

$$\Phi_1 = [\mu_{a,1}(x) V \mu_{b,1}(y) V \mu_{c,1}(z)] \tag{1.64}$$

The output membership value corresponding to the moderate risk can be expressed as

$$\Phi_2 = [(\mu_{a,2}(x) V \mu_{b,2}(y) V \mu_{c,2}(z)] \tag{1.65}$$

The output membership value corresponding to the mild risk can be expressed as

$$\Phi_3 = [(\mu_{a,3}(x) V \mu_{b,3}(y) V \mu_{c,3}(z)] \tag{1.66}$$

The output membership value corresponding to the normal risk can be expressed as

$$\Phi_4 = [(\mu_{a,4}(x) V \mu_{b,4}(y) V \mu_{c,4}(z)] \tag{1.67}$$

where V indicates the maximum operator.

The output can be defuzzified using certroid defuzzification scheme to obtain a crisp value C.

$$C = \left[\frac{\sum_{j=1}^{j=4} j\Phi_j}{\sum_{j=1}^{j=4} \Phi_j} \right] \tag{1.68}$$

This crisp output gives a continuous value. In the present case, C has a value between 1 and 4 where 1 represents severe risk for dysphagia, 2 represents moderate risk for dysphagia, 3 represents mild risk for dysphagia, and 1 represents normal. Figure 1.15 compares the classification made by the fuzzy logic diagnosis system and the classification made by the clinician. There was complete agreement between the fuzzy logic system and the clinician classification. In four cases, the fuzzy logic system overestimated the risk by half a category. It should be noted that the clinician classification itself is subjective based on qualitative observations.

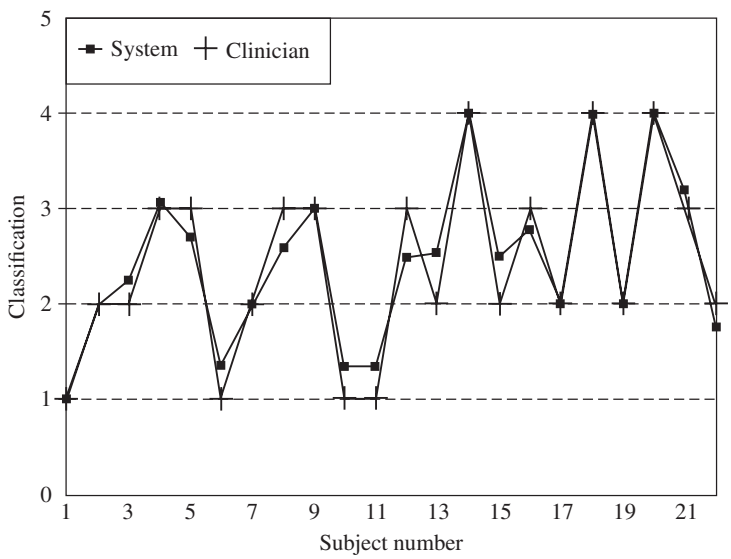

FIGURE 1.15 A comparison of the classification made by the fuzzy logic system and the clinician. [*Reproduced with permission from Suryanarayanan et al. (1995).*]

1.6.2 Fuzzy Mechanics

Virtual reality (VR) is gaining importance in every discipline, including medicine and surgery. VR is a computer-generated pseudo space that looks, feels, hears, and smells real, and fully immerses the subject. Potential applications of VR include medical education and training, patient education, VR-enhanced rehabilitation exercises, VR-induced biofeedback therapy and teletherapy, VR-aided emergency medicine, VR surgical simulations, etc. Surgical simulations in VR environment can aid the surgeon in planning and determining the optimal surgical procedure for a given patient, and also can aid in medical education. Song and Reddy (1995) have demonstrated the proof of concept for tissue cutting in VR environment. They have developed a technique for cutting in VR using interactive moving node finite element models controlled by user-exerted forces on instrumented pseudocutting tool held by the user. In Song and Reddy's system, the user (surgeon) holds the instrumented wand (instrumented pseudotool) and manipulates the tool. The forces exerted by the wand (on a table) together with the orientation and location of the wand are measured and fed to a finite element model of the tissue and the deformations are calculated to update the model geometry. Cutting takes place if the force exerted is larger than the critical force (computed from local tissue properties) at the node. Otherwise, tissue simply deforms depending on the applied force. However, finite element models require significant amount of computational time and may not be suitable for cutting large amount of tissue. Recently, Kutuva et al. (2006) developed a fuzzy logic system for cutting simulation in VR environment. They fuzzified the force applied by the operator and the local stiffness of the tissue. The membership functions for the force and stiffness are shown in Fig. 1.16. They have developed

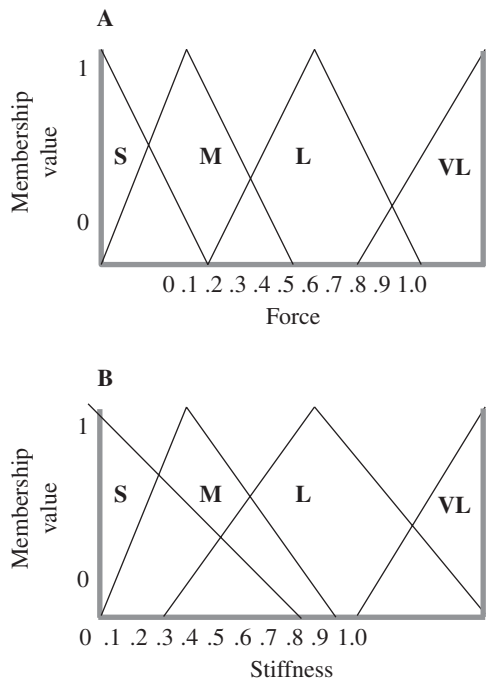

FIGURE 1.16 Fuzzy membership functions for (A) force exerted on the pseudohand held by the user, and (B) the local stiffness of the tissue. Fuzzy membership functions describe small, medium, large, and very large subdomains. [*Reproduced with permission from Kutava et al. (2006).*]

TABLE 1.1 Fuzzy-rules lookup table

Stiffness \ Force	Small	Medium	Large	Very large
Small	No cut	Small	Medium	Large
Medium	No cut	No cut	Small	Medium
Large	No cut	No cut	No cut	Small
Very large	No cut	No cut	No cut	No cut

The force subdomains are in the horizontal column and the stiffness domains are described in the vertical column. The corresponding output subdomain is given in the cells.

a lookup table (IF-THEN rules) to compute the cutting depth. Table 1.1 shows the lookup table. For instance, if the normalized force is 0.4, the membership values are (0, 0.5, 0.3, 0). The membership value for small is zero, for medium is 0.5, for large is 0.3, and zero for very large. If the normalized stiffness at the cutting location is 0.32, the membership values for stiffness are (0.4, 0.8, 0.4, 0). The stiffness membership value is 0.4 for small, 0.8 for medium, 0.4 for large, and 0 for very large. The contribution to the output domain small cut is calculated from the lookup table by adding all the possibilities for small cut. Small cut is possible if the force is medium (0.5) and the stiffness is small (0.4) which results in 0.4 times 0.5 which is 0.2; additional possibility is if force is large (0.3) and stiffness is medium (0.8) which results in $0.3 \times 0.8 = 0.24$ for small cut. Another possibility is if force is very large (0) and stiffness is large (0.4) which results in 0. Now, the membership value for the output subdomain small cut is calculated by adding all these possibilities: $0.2 + 0.24 + 0 = 0.44$. Similar calculation for medium cut results in $0.3 \times 0.4 = 0.12$. The membership value for no cut results in $0.5\ (0.8 + 0.4) + 0.3 \times 0.4 = 72$. The output membership values for this example are (0.72, 0.44, 0.12, 0). The fuzzy output of the cutting depth is then defuzzified using the centoid defuzzification scheme to calculate a crisp value for the cutting depth. In the above example, the crisp value is calculated as

$$C = \frac{0.75 \times 1 + 0.44 \times 2 + 0.12 \times 3 + 0 \times 4}{0.72 + 0.44 + 0.12 + 0} \tag{1.69}$$

At each epoch, the force exerted by the user on the pseudocutting tool is measured, fuzzified, and the cutting depth is calculated. The virtual cutting tool is then advanced to the depth computed by the fuzzy logic system. The node(s) along the cutting path are released, and the display is updated with cut view. Then, the user's input force exerted on the pseudocutting tool is measured, defuzzified, and the cutting is performed by advancing the tool to the new location by the amount of cutting depth. The procedure is repeated as long as the pseudocutting tool is in the cutting space. The procedure is continued until the pseudocutting tool is withdrawn out of the virtual tissue space. Figure 1.17 provides a demonstration of fuzzy-logic-based tissue cutting in VR environment using two-dimensional models.

1.7 MODEL VALIDATION

This chapter discussed the art of modeling with few examples. Regardless of the type of model developed, a mathematical model should be validated with experimental results. Validation becomes very important in the black box type of models such as the neural network models. Moreover, the model results are valid only to certain regimes where the model assumptions are valid. Sometimes, any model can be fit to a particular data by adjusting the parameter values. Moreover, the techniques of parameter estimation were not presented in this chapter. In addition, the presentation was limited to lumped parameter analysis or macroscopic modeling.

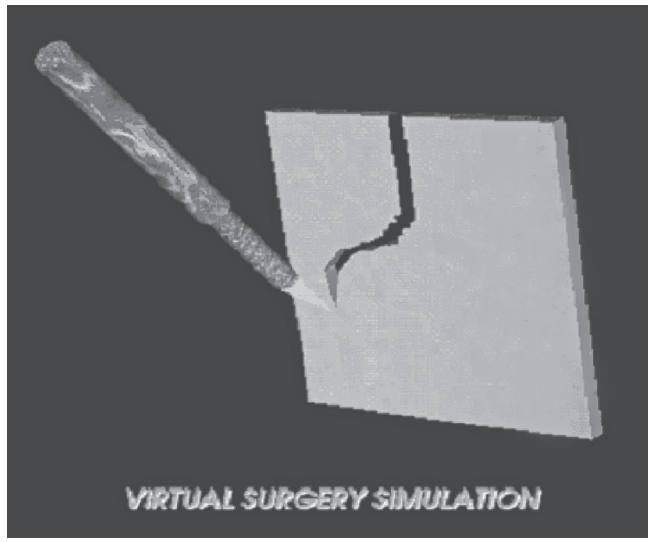

FIGURE 1.17 A demonstration of the fuzzy logic cutting. [*Reproduced with permission from Kutava et al. (2006).*]

REFERENCES

Barnea, O., and Gillon, G., (2001), Model-based estimation of male urethral resistance and elasticity using pressure-flow data. *Comput. Biol. Med.* **31:**27–40.

Chu, T. M., and Reddy, N. P. (1992), A lumped parameter mathematical model of the splanchnic circulation. *J. Biomech. Eng.* **114:**222–226.

Das, A., Reddy, N. P., and Narayanan, J. (2001), Hybrid fuzzy logic committee neural networks for recognition of acceleration signals due to swallowing. *Comput. Methods Programs Biomed.* **64**:87–99.

Fritton, J. C., Rubin, C. T., Quin, Y., and McLeod, K. J. (1997), Whole body vibration in the skeleton: development of a resonance-based testing device. *Ann. Biomed. Eng.* **25**:831–839.

Haykin, S. (1999), *Neural Networks: A Comprehensive Foundation,* 2nd ed., Prentice Hall, Upper Saddle River, N.J.

Kerckhoffs, R. C., Neal, M. L., Gu, Q., Bassingthwaighte, J. B., Omens, J. H., and McCulloch, A. D. (2007), Coupling of a 3D finite element model of cardiac ventricular mechanics to lumped system models of the systemic and pulmonic circulation. *Ann. Biomed. Eng.* **35**:1–18.

Kim, J., Saidel, G. M., and Cabrera, M. E. (2007), Multiscale computational model of fuel homeostasis during exercise: effect of hormonal control. *Ann. Biomed. Eng.* **35**:69–90.

Kutuva, S., Reddy, N. P., Xiao, Y., Gao, X., Hariharan, S. I., and Kulkarni, S. (2006), A novel and fast virtual surgical system using fuzzy logic. *Proceedings of IADIS Multi Conference on Computer Graphics and Visualization 2006* (Nian-Shing Chen and Pedro Isaias, eds.) IADIS Press, pp. 277–281.

Mangiameli, P., West, D., and Rampal, R. (2004), Model selection for medical decision support systems. *Decis. Support Syst.* **3**:247–259.

Reddy, N. P. (1986), Lymph circulation: physiology, pharmacology and biomechanics. *CRC Crit. Rev. Biomed. Eng.* **14**:45–91.

Reddy, N. P., and Krouskop, T. A. (1978), Modeling of the cell mediated immunity in homograft rejection. *Int. J. Biomed. Comp*. **9:**327–340.

Reddy, N. P., and Kesavan, S. K. (1989), Low Reynolds number liquid propulsion in contracting tubular segments connected through valves. *Math. Comput. Modeling.* **12:**839–844.

Reddy, N. P., Costerella, B. R., Grotz, R. C., and Canilang, E. P. (1990), Biomechanical measurements to characterize the oral phase of dysphagia. *IEEE Trans. Biomed. Eng.* **37**:392–397.

Reddy, N. P., Canilang, E. P., Casterline, J., Rane, M. B., Joshi, A. M., Thomas, R., Candadai, R. (1991), Noninvasive acceleration measurements to characterize the pharyngeal phase of swallowing. *J. Biomed. Eng.* **13**:379–383.

Reddy, N. P., Thomas, R., Canilang, E. P., and Casterline, J. (1994), Toward classification of dysphagic patients using biomechanical measurements. *J. Rehabil. Res. Dev.* **31**:335–344.

Reddy, N. P., Prabhu, D., Palreddy, S., Gupta, V., Suryanarayanan, S., and Canilang, E. P. (1995), Redundant neural networks for medical diagnosis: diagnosis of dysphagia, In *Intelligent Engineering Systems through Artificial Neural Networks: Vol. 5 Fuzzy Logic and Evolutionary Programming* (C. Dagli, A. Akay, C. Philips, B. Fernadez, J. Ghosh, eds.) ASME Press, N.Y., pp. 699–704.

Reddy, N. P., Katakam, A., Gupta, V., Unnikrishnan, R., Narayanan, J., and Canilang, E. P. (2000), Measurements of acceleration during videofluorographic evaluation of dysphagic patients. *Med. Eng. Phys.* **22**:405–412.

Reddy, N. P., and Buch, O. (2003), Speaker verification using committee neural networks. *Comput. Methods Programs Biomed.* **72:**109–115.

Reddy, N. P., Mathur, G., and Hariharan, H. I. (2008), Toward fuzzy logic control of infant incubators. (in press)

Rideout, V. C. (1991), *Mathematical and Computer Modeling of Physiological Systems.* Prentice Hall, Englewood Cliffs, N.J.

Sakaguchi, S., Takifuji, K., Arita, S., and Yamaeu, H. (2004), Development of an early diagnostic system using fuzzy theory for postoperative infections in patients with gastric cancer. *Diagn. Surg.* **21**:210–214.

Simon, B. N., Reddy, N. P., and Kantak, A. (1994), A theoretical model of infant incubator dynamics. *J. Biomech. Eng.* **116**:263–269.

Song, G. J., and Reddy, N. P. (1995), Tissue cutting in virtual environments, In *Interactive Technology and the New Paradigm for Healthcare* (R. M. Satava, K. Morgan, H. B. Seiburg, R. Mattheus, and J. P. Cristensen, eds.), IOP Press and Ohmsha, Amstardam, pp. 359–364.

Sproule, B. A., Naranjo, C. A., and Tuksen, I. B. (2002), Fuzzy pharmacology: theory and applications. *Trends in Pharmacol. Sci.* **23:**412–417.

Steimann, F., and Adlassnig K. P. (1998), Fuzzy medical diagnosis. In *Handbook of Fuzzy Computation*. IOP Press, Oxford, pp. 1–14.

Sturm, R. (2007), A computer model for the clearance of insoluble particles from the tracheobronchial tree of the human lung. *Comput. Biol. Med.* **37**:680–690.

Suryanarayanan, S., Reddy, N. P., and Canilang, E. P. (1995), A fuzzy logic diagnosis system for classification of pharyngeal dysphagia. *Int. J. Biomed. Comput.* **38:**207–215.

Suryanarayanan, S., and Reddy, N. P. (1997), EMG based interface for position tracking and control in VR environments and teleoperation. *PRESENCE: teleoperators and virtual environ*. **6:**282–291.

CHAPTER 2
BIOMEDICAL INFORMATICS

William Hersh
Oregon Health & Science University, Portland, Oregon

2.1 INTRODUCTION

The modern healthcare system, and professionals and patients within it, spends a substantial portion of its time collecting, analyzing, and using data and information, not only during patient care but also for research, quality measurement and improvement, and public health. The discipline devoted to the efficient storage, acquisition, and use of information in healthcare is called *biomedical and health informatics* (BMHI, or often biomedical informatics for short) [1]. The term *informatics* itself usually refers to the confluence of people, information, and technology [2] (see Fig. 2.1). While information technology (IT) infrastructure (i.e., the networks, devices, and software) is essential for effective application of informatics, the larger goal is the benefit that information provides, in the case of BMHI, to healthcare and optimal health of individuals and populations.

2.2 AREAS WITHIN BIOMEDICAL AND HEALTH INFORMATICS

Although informatics is a relatively new discipline compared to others in medicine, it has accumulated a history over a half-century that has evolved with advances in IT [3]. The term *informatics* has been attributed to Dreyfus from 1962 [4]. The word was used initially mostly throughout Europe—starting in France as the term *informatique* and then later around the rest of Europe—all through the 1960s. In Europe, the focus of the term has been on computing issues related to information use, which is somewhat different from how we use informatics now, especially in the context of BMHI. The term *medical informatics* is purported to first have been used in 1974 [5]. There are two books that have been published by the late Dr. Morris Collen—an initial volume in 1995 [6] and then an updated volume that he finished shortly before his death in 2015 [3]. At present, the term *informatics*

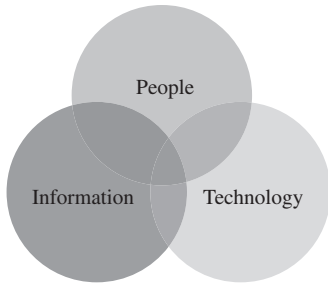

FIGURE 2.1 Venn diagram of essential components of informatics. (*Adapted from Ref. 2.*)

is most substantially used in the biomedical arena, although there are fields such as legal informatics, chemoinformatics, social informatics, and so forth.

As shown in Fig. 2.1, modern usage of informatics focuses on it as the confluence of people, information, and technology. One adage about the field is that it is more about information than technology. Friedman has defined the "fundamental theorem" of informatics, which states that informatics is more about using technology to help people perform their work better than about building systems to mimic or replace human expertise [7]. He has also described what informatics is (information sciences applied in a biomedical application domain with the aim of helping people) and is not (any use of IT in healthcare) [8].

The various areas within BMHI are depicted in Fig. 2.2. Sometimes narrower words appear in front of the term *informatics*. Other areas of BMHI include:

- *Bioinformatics*—the application of informatics in cellular and molecular biology, often with a focus on genomics [9].
- *Imaging informatics*—informatics with a focus on imaging, including the use of systems to store and retrieve images across all types of informatics [10].
- *Clinical informatics*—informatics applied in healthcare settings [11].
- The application of informatics focused on specific healthcare disciplines, such as nursing (*nursing informatics*) [12], dentistry (*dental informatics*), pathology (*pathology informatics*), etc. [13].
- *Consumer health informatics*—the field devoted to informatics from a consumer view, often with a focus on mobile health [14].

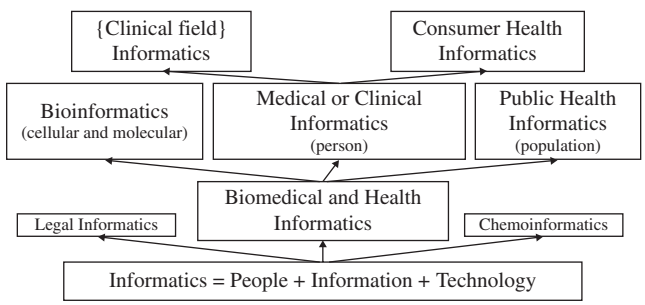

FIGURE 2.2 Biomedical and health informatics and areas within it. (*Adapted from Ref. 2.*)

- *Clinical research informatics*—the use of informatics to facilitate clinical research, with increasing emphasis on *translational research* that aims to accelerate research findings into clinical practice [15].
- *Public health informatics*—the application of informatics in areas of public health, including surveillance, reporting, and health promotion [16].

We can also view the continuum of bioinformatics, clinical informatics, and public health informatics as a spectrum of informatics applied to cellular and molecular processes to people to populations, although there is considerable overlap, as exemplified through the use of genomics technologies to investigate public health outbreaks. Some use the term *health informatics* to describe the field, which Shortliffe and Cimino denote as the combination of clinical informatics and public health informatics [17].

2.3 MAJOR APPLICATIONS OF BIOMEDICAL AND HEALTH INFORMATICS

There are many applications of BMHI that vary based on the focus of the subarea of the field. One application that has relevance to most areas of BMHI is the *electronic health record* (EHR), which serves several key functions to improve care delivery, not only in documenting data and information of care delivery but also in providing access to other participants in the system, most importantly the patient. But there are many other important applications as well.

2.3.1 Electronic Health Record

One of the most central applications of clinical informatics is the EHR. In the past, the term *electronic medical record* (EMR) was more commonly used, but EHR implies a broader and more longitudinal collection of information about the patient. There is also increasing use of the term *personal health record* (PHR). This usually refers to the patient-controlled aspect of the health record and may or may not be *tethered* to one or more EHRs from healthcare delivery organizations. Growing numbers of health systems have adopted the Open Notes approach that provides patients access to clinical notes along with a substantial amount of other data [18–20].

The EHR is not meant to be a mere replacement for the paper-based record, rather it should ideally serve as a tool to transform and improve healthcare delivery. One major component of the EHR is *clinical decision support* (CDS), which allows detection of errors and adverse events and can facilitate improved care delivery and quality [21]. The most critical time for intervention is when the physician is entering patient orders, so the optimal time to make CDS readily available is within the functioning of *computerized physician/provider order entry* (CPOE). Related to CPOE is *electronic prescribing* (e-prescribing), which focuses more narrowly on the electronic ordering of medications. The major categories of CDS include:

- *Information display*—showing general or patient-specific information in context of the current clinical situation.
- *Reminder systems*—reminding clinicians to perform actions, such as preventive measures, when they are due.
- *Alerts or notifications*—alerting to critical clinical situations, e.g., interacting drugs or abnormal lab values, that may negatively impact patient safety and health outcomes.
- *Clinical practice guidelines*—guiding treatment to promote consistent care based on best evidence.

An exemplar EHR has been the *Veterans Health Information Systems and Technology Architecture* (VistA) system, which is used in 1800 locations across the world including all U.S. Department of Veterans Affairs medical centers as well as several national health systems. Figure 2.3 shows the

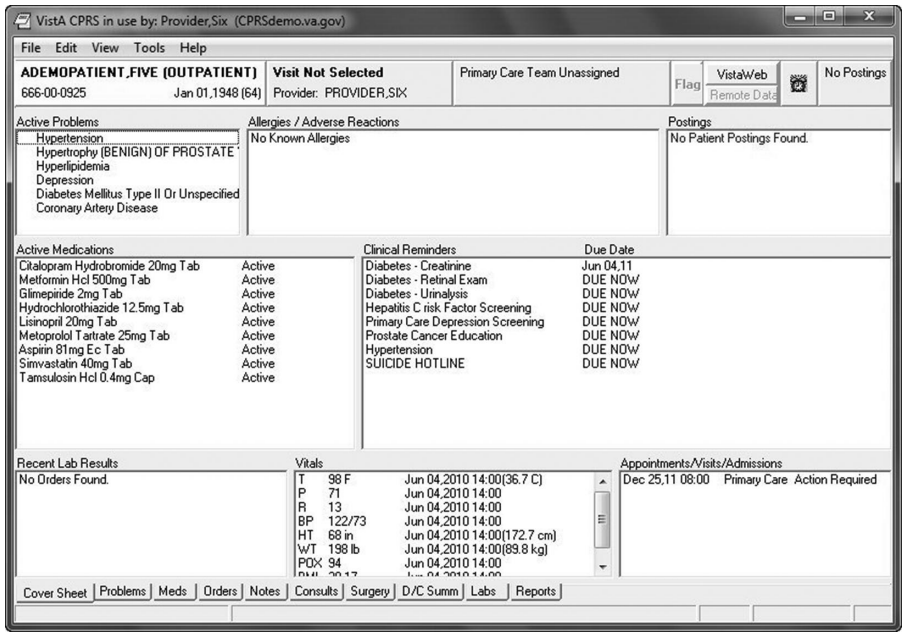

FIGURE 2.3 Cover page of the Veterans Health Information Systems and Technology Architecture (VistA).

cover page of the VistA EHR, which provides an overview of the patient, including their active problems and medications as well as recent results. This page also shows an example of CDS, listing relevant clinical reminders. The tabs at the bottom of the screen allow the user to drill down into more details on specific aspects of the patient's care, such as medications and laboratory results. Many of these screens feature additional CDS, such as indicating drug–drug interactions. Current Department of Veterans Affairs plans include replacing VistA with a commercial EHR over the next decade to better align with the commercial EHR system being adopted by the Department of Defense so that the care of military personnel can be seamlessly integrated as they transition to veteran status.

As the use of EHRs has grown, it has become apparent that information does not seamlessly flow among healthcare providers, and across different healthcare organizations. This has led to growing advocacy for *health information exchange* (HIE), which is the exchange of health information for patient care across traditional business boundaries in healthcare [22]. Even many healthcare organizations that have exemplary HIT systems have difficulty providing their patient information to other entities where the patient may receive care. An increasingly mobile population demands "data following the patient" as patients move in, out, and across healthcare systems.

One of the impediments to HIE has been suboptimal *interoperability* of EHR systems, with systems unable to seamlessly exchange data. Optimal interoperability requires adoption and adherence to *standards* to define data structures and formats. Although many standards exist for exchange of information and consistent use of terminology, they have not been consistently applied for a variety of reasons [23]. The major categories of standards include:

- *Identifiers*—of patients, clinicians, health plans, insurance companies, etc.
- *Transactions*—eligibility, enrollment, payments, etc.
- *Message exchange*—transmission of data, images, documents, etc.
- *Terminology*—standard descriptions of diagnoses, tests, treatments, etc.

The longest-standing and most widely used messaging standard in healthcare is Health Level 7 (HL7) Version 2. However, there are no formal standard terminology requirements for names of diagnoses, tests, and treatments in HL7 Version 2 as well as a number of other technological limitations, which limits its accomplishing of interoperability. These limitations have led to a new messaging standard, the *Fast Healthcare Interoperability Resources* (FHIR) [24]. A key element of FHIR is Resources, which provide structured and standardized modeling of all data components of healthcare, from patients to observations to medications [25, 26]. A complementary standard is *Substitutable Medical Apps, Reusable Technology* (SMART), which provides a standardized platform for building EHR and PHR "apps" (which can be Web-based or run on mobile devices). These two have been married to form *SMART on FHIR*, which aims to provide new forms of interaction based on standardized data on top of the EHR [27].

The presence of standardized interoperable data and systems not only leads to improved direct care of patients but also enables *reuse* (also called *secondary use*) of clinical data, where data from clinical settings is used for other applications, such as quality measurement and improvement, clinical and translational research, and public health [28].

2.3.2 Beyond the EHR

BMHI is not limited to EHRs. Another vital component for optimal patient care is access to information and knowledge. The field devoted to indexing and retrieval of knowledge-based information is called *information retrieval* (IR) or *search* [29]. IR systems are ubiquitous in modern life, and a large majority of people search for health-related information using Google or other general search engines [30]. In addition, virtually all healthcare professionals search the scientific literature, including electronic journals, books, and other websites, in their professional practice [31].

Searching is a requisite skill in the practice of *evidence-based medicine* (EBM), a process that includes the proper phrasing of clinical questions, seeking the best evidence to answer such questions, critically appraising what was retrieved, and applying such evidence to patient care. One recent textbook of EBM notes, "Searching for current best evidence in the medical literature has become a central skill in clinical practice. On average, clinicians have 5 to 8 questions about individual patients per daily shift … Some now even consider that 'the use of search engines is as essential as the stethoscope'" [32]. The importance of search goes beyond EBM, as clinicians must have skills in finding high-quality information for use by professionals and patients alike.

Additional important applications of clinical informatics are telemedicine and telehealth [33]. *Telemedicine* is the delivery of healthcare when the participants are separated by time and/or distance, while *telehealth* has a larger aspect of all telecommunications applications devoted to health. As with informatics, the "tele-" terms sometimes reflect medical specialties in which they are applied, e.g., teleradiology and telepathology. A variety of practice models embracing telehealth have now emerged, including e-ICU and telestroke services which are commonly employed to deliver expertise to a broader population. The uptake of telemedicine and telehealth increased substantially at the start of the COVID-19 pandemic.

There are also a number of related terms used that refer to informatics in specific contexts. The term *eHealth* usually refers to health-related applications that use information and communications technology [34]. The term *mHealth* refers to informatics applied to mobile devices such as smartphones [35]. The definition of *digital health* varies, but has been defined to include genomics [36].

Other applications of BMHI focus on bioinformatics. This work began with the sequencing of the human genome but has now spread to other "omes and omics." Not only have next-generation sequencing technologies led to rapidly lowering costs [37], but other biomolecular technologies have produced many applications relevant not only to genomics but also to other omes and omics [38]. These include:

- *Gene expression and transcriptomics* to measure expression of DNA and RNA.
- *Proteomics* to elucidate structure and function of proteins.
- *Microbiome*—microorganisms in other living organisms.
- *Mapping phenotype to genotype*—closing the circle of genomics to clinical data.

A rich source of data and tools, especially but not limited to bioinformatics, comes from the National Center for Biotechnology Information (NCBI) of the U.S. National Library of Medicine (NLM) [39]. A larger catalog of data and systems is published annually in the database issue of the journal, *Nucleic Acids Research* [40].

Another example of the various subareas of BMHI converging is in *precision medicine,* where diagnosis and therapy are driven by identifying genetic, environmental, and lifestyle factors related to disease [41]. Pharmacogenomics is a subset of precision medicine that studies how genetics affect a person's response to particular drugs [42]. The U.S. government has recently committed a substantial investment in research around precision medicine in the AllOfUs initiative [43].

2.4 HISTORY OF BIOMEDICAL AND HEALTH INFORMATICS

The scientific origin of the field of informatics is often attributed to the paper by Ledley and Lusted (1959) [44], which is one of the most widely cited papers in the field. This paper aimed to model and understand physician reasoning through:

- *Symbolic logic*—representing concepts such as patient findings, tests, and diagnoses.
- *Probability*—likelihood of outcomes (e.g., diagnosis) based on concepts (symbols).
- *Value theory*—complexity of values going into medical decision making.

This paper led to early attempts at computer-based decision making in medicine and led to the first era of *artificial intelligence* (AI). The focus of this era was on hand-crafted knowledge bases with algorithms to provide "artificial" intelligence. One early application of applying AI included a mathematical model for diagnosing congenital heart disease, predicting diagnosis with the highest conditional probability given a set of symptoms [45]. Another approach was to use "problem-knowledge couplers" that aimed to connect patient findings and diagnoses [46].

This was followed by the emergence of "expert systems," which were computer programs that aimed to mimic human expertise. One line of work focused on rule-based expert systems, including the PhD dissertation of Shortliffe [47] and subsequent work [48]. Another early AI approach applied scoring algorithms. INTERNIST-1 [49] and DxPlain [50] used disease profiles and scoring aiming to cover all of general internal medicine. These early systems, based on computer technology of the time that included time-share networks and command-line user interfaces, were slow and time-consuming to use. They also did not provide much value in everyday clinical practice, leading one early researcher to write a paper on the "Demise of the Greek Oracle" approach to AI and advocating that systems instead aim to provide supportive rather than independent role [51]. This coincided with the general decline of interest in the first era of AI and led to the more focused type of CDS that emerged in the 1990s [21].

The 1960s also saw the first work that led to early EHR systems. Most of these systems were primitive by today's standards but set BMHI on the path to learning about implanting EHRs in clinical settings. Some well-known early systems included:

- *COmputer STored Ambulatory Record (COSTAR)*—from Massachusetts General Hospital [52], based on the MUMPS programming language, which is still used in many modern EHR systems [53].
- *Health Evaluation through Logical Processing (HELP)*—from LDS Hospital in Utah [54].
- *The Medical Record (TMR)*—from Duke University [55].
- *Regenstrief Medical Record System (RMRS)*—from Indiana University [56], which led to development of Gopher [57].
- El Camino Hospital in California [58].
- *Veterans Health Information Systems and Technology Architecture (VistA) and Computerized Patient Record System (CPRS)*—from the U.S. Veteran's Health Administration [59].

A critical early leader in BMHI was the NLM [60]. One of the first and still one of the most important applications was the ability to search bibliographic databases. This started with the paper volumes of *Index Medicus* that were used before the widespread availability of systems such as command-line ELHILL [61], PC-based Grateful Med [62], and Web-based PubMed that is now used ubiquitously. The NLM has also developed other information resources for clinicians, researchers, and consumers. It has also been a leader in terminology development and standardization [63] as well as a funder of research and training. The NLM was led for over 30 years by Donald Lindberg, MD, who recently passed the torch to Patricia Brennan, PhD, RN [64].

Another important historical event for BMHI was the *Human Genome Project* to sequence human genome. The project started in 1988 and in 2001, NIH-based project published "first draft" [65] simultaneously with private effort from Craig Venter of Celera Genomics [66]. The project was declared to be "completed" in 2003 [67]. The sequencing of more humans and other organisms has increased understanding of genomic variation and complexity and led to continued advances in bioinformatics.

Also important to BMHI was the thought leadership of a series of reports published by the Institute of Medicine (IOM; now National Academy of Medicine):

- *The Computer-Based Patient Record*—paper records illegible, inefficient, and error-prone; computer-based record vital to modern healthcare [68].
- *For the Record: Protecting Electronic Health Information*—benefits of electronic health information compromised by inadequate protection; informed HIPAA legislation [69].
- *Networking Health*—value of computer networks and the Internet for health [70].
- *To Err is Human*—medical errors are common and a systems problem [71].
- *Crossing the Quality Chasm*—set of aims and rules for high-quality twenty-first century healthcare [72].

Another important concept to emerge from the IOM was that of the *learning health system*—must measure provision and outcomes of care to know what works [73]. Components of learning health system included transparency of data and information; rewarding of outcomes and value, not volume; and errors promptly identified and corrected [74]. The IOM also noted that although health IT systems may improve healthcare, they may also introduce error and cause harm if not designed and applied properly [75].

The importance of clinical informatics in healthcare delivery began to emerge in the latter part of the twentieth century. A series of seminal reports from the National Academy of Medicine (NAM), known at the time as the Institute of Medicine (IOM), documented significant problems in healthcare delivery and led to proposed solutions based on best information technology (IT) and evidence supporting its use. The first NAM report documented the harms resulting from incomplete and illegible paper-based medical records [68]. Probably the most high profile of these reports focused on errors in hospitals estimated to result in up to 96,000 deaths per year [71]. Another NAM report described deficiencies in the quality of healthcare as a "chasm" between known, evidence-based best practices and their actual use in the healthcare system. Constraints on exploiting the revolution in information technology were named as one of the underlying reasons for inadequate quality of care, and increasing the use of information technology was cited as a means of improving quality of care [72].

A number of studies supported the conclusions of these reports. In 1995, Bates et al. documented error rates of 6.5 adverse drug events (ADEs) per 100 hospitalized patients [76]. Quality problems were quantified more clearly in 2003 by McGlynn et al., who assessed the records of 6259 patients in 12 metropolitan areas and found that only 54.9 percent of care delivered was consistent with evidence-based known best practices [77]. Paper-based medical information was associated with clinical decisions being made with incomplete information, as Smith et al. showed that information was missing and impacted up to 44 percent of patients in primary care settings [78].

Additionally, there was emerging evidence for the value of health IT. In 1993, Tierney et al. documented that CPOE in hospitals was associated with a 12.7 percent decrease in total charges and 0.9 days shorter length of stay [79]. Shortly afterwards, Bates and colleagues showed that CPOE

reduced serious medication errors by 55 percent, with adverse drug events reduced by 17 percent [80]. Other studies showed CPOE led to a reduction in redundant laboratory tests [81] and increased prescribing of equally efficacious but less costly medications [82]. Much of this work culminated in a systematic review of 257 studies of health IT documenting its association with increased adherence to guideline-based care, enhanced surveillance and monitoring, and decreased medical errors [83].

Modeling studies were also being published demonstrating return on investment for EHRs as well as HIE, the exchange of information across the boundaries of healthcare organizations [84]. Johnston and colleagues assessed the potential benefit of CPOE in ambulatory settings and noted savings of up to $28,000 per practice per year, although most of the savings went to laboratories and insurance companies, and not the physician practices making the investment [85]. Another modeling study by Hillestad et al. applied results of known research in an attempt to scale to entire U.S. healthcare system, finding that HIE could potentially result in savings of $81 billion per year to the system and a reduction of 200,000 ADEs per year [86].

Another important happening for BMHI was the *American Recovery and Reinvestment Act* (ARRA), the economic stimulus package passed by Congress in early 2019. Within ARRA was the *Health Information Technology for Economic and Clinical Health* (HITECH) Act, which allocated around $30 billion for incentives for EHR adoption in the United States. This led to the so-called "meaningful use" program that required adherence to criteria to qualify for incentive payments. While the program overall substantially increased EHR adoption in the United States up to 96 percent in hospitals and 85 percent among office-based physicians [87], there were also criticisms of inadequate attention paid to usability and interoperability of systems [88].

Even though not limited to the work of physicians, clinical informatics has been recognized as a medical subspecialty [11] and has been defined by the Accreditation Council for Graduate Medical Education (ACGME) as the field that "transforms healthcare by analyzing, designing, implementing, and evaluating information and communication systems to improve patient care, enhance access to care, advance individual and population health outcomes, and strengthen the clinician-patient relationship" [89]. Since 2013, physicians who have worked in the field or completed a fellowship in informatics and have a primary board certification in their specialty have been eligible to become additionally board certified in clinical informatics. That subspecialty certification is available to physician specialists who are certified by any of the 24 member boards of the American Board of Medical Specialties, which endorses the broad clinical relevance of expertise in clinical informatics [11]. Since the first certification exam was offered in 2013, over 2000 physicians have become board-certified.

2.5 DATA SCIENCE AND ANALYTICS

One of the promises of the growing critical mass of clinical data accumulating in the EHR is secondary use (or reuse) of the data for other purposes, such as quality improvement, operations management, and clinical research [28]. There has also been substantial growth in other kinds of health-related data, most notably through efforts to sequence genomes and other biological structures and functions. This has led to the application of the related areas of data science and data analytics to biomedicine and health.

Data science is "the science of learning from data; it studies the methods involved in the analysis and processing of data and proposes technology to improve methods in an evidence-based manner" [90]. *Data analytics* may be viewed as the practical application of data science, i.e., "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions" [91].

Hospitals and other healthcare organizations are generating an exploding amount of data. Clinical data takes a variety of forms, from structured (e.g., images and lab results) to unstructured (e.g., textual notes including clinical narratives, reports, and other types of documents). Additionally, healthcare organizations capture and generate data as a byproduct of the care delivery process. This can include billing, quality, management, and other financial data which is increasingly important in

the optimization of healthcare delivery. Kaiser-Permanente estimated in 2013 that its data store for its 9+ million members exceeded 30 petabytes of data [92].

These two areas have been enabled by both the massive increase in the collection and availability of data, along with the development of new techniques to make predictions and recommend actions from that data. Another facilitator of this advance has been the capabilities of *machine learning,* the subarea of computer science defined as the ability of computer programs to learn without being explicitly programmed [93] or the use of computers to optimize a performance criterion using example data or past experience [94]. Many of the recent successes in machine learning come from the use of neural networks that, when used in deep layers, called *deep learning,* have led to impressive results [95].

As noted above, one of the reasons for the growth and success of data analytics and deep learning has been the development of *Big Data,* which refers to the large and ever-increasing quantities of data that adhere to the following attributes [96]:

- *Volume*—ever-increasing amounts.
- *Velocity*—quickly generated.
- *Variety*—many different types.
- *Variability*—variation in amounts, generation, and types.

Bates et al. have elucidated a number of use cases where big data methods might lead to improved patient outcomes [97]:

- *High-cost patients*—looking for ways to intervene early.
- *Readmissions*—preventing.
- *Triage*—providing appropriate level of care.
- *Decompensation*—alerting when patient's condition worsens.
- *Adverse events*—raising awareness.
- *Treatment optimization*—especially for diseases affecting multiple organ systems.

Adams and Klein authored a primer on data analytics in healthcare that defines different levels of the application of analytics and describes their attributes [98]. They note three levels of analytics, each with increasing functionality and value:

1. *Descriptive*—standard types of reporting that describe current situations and problems, e.g., reports of patients with certain diagnoses or outlier test results.
2. *Predictive*—simulation and modeling techniques that identify trends and portend outcomes of actions taken, e.g., lists of patients who may be at risk for poor outcomes or repeated admissions to the hospital.
3. *Prescriptive*—optimizing clinical, financial, and other outcomes, e.g., recommendations for patients to maintain health or to prevent poor outcomes.

Much work is focusing now on *predictive analytics,* especially in clinical settings attempting to optimize health and financial outcomes, including in clinical practice [99].

Deep learning algorithms have been developed for many tasks, especially those in physician specialties that use imaging. A sampling of these studies includes:

- *Radiology*—diagnosis comparable to radiologists for pneumonia [100], tuberculosis [101], and intracranial hemorrhage [102].
- *Dermatology*—detecting skin cancer from images [103–105].
- *Ophthalmology*—detecting diabetic retinopathy from fundal images [106, 107], diagnosing plus disease [108], and predicting cardiovascular risk factors from retinal fundus photographs [109].
- *Pathology*—classifying various forms of cancer from histopathology images [110, 111] and detecting lymph node metastases [112].

- *Cardiology*—cardiac arrhythmia detection [113] and interpreting electrocardiograms [114] comparable to cardiologists.
- *Gastroenterology*—assessing endocytoscopic images for diagnose-and-leave strategy for diminutive, nonneoplastic, rectosigmoid polyps [115].

Success has not been limited to images, however, and deep learning algorithms have also been able to process data in EHRs to predict:

- Several dozen diseases [116].
- Length of stay, mortality, readmission, and diagnosis at two large medical centers [117].
- Prognosis in palliative care [118].
- Thirty-day readmission in heart failure [119].
- Patient mortality from coronary artery disease more accurately than traditional cardiovascular risk models [120].
- Early risk of chronic kidney disease in patients with diabetes [121].
- Many pediatric diagnoses at a major referral center [122].
- Clinical outcomes in rheumatoid arthritis [123].

The world's growing base of scientific literature is another source of data, and can be linked with EHR and other patient data to improve outcomes of care. One approach to this problem that has generated attention is the IBM Watson project, which was made famous by winning at the TV game show Jeopardy! [124]. One of the areas where IBM and its partners have been applying Watson is in the healthcare arena [125].

A more peripheral but related term is *business intelligence,* which in healthcare refers to the "processes and technologies used to obtain timely, valuable insights into business and clinical data" [98]. A major motivator for data-driven decision making in healthcare is the move from volume-driven (e.g., fee for service) to value-driven (where health systems and providers share risk) reimbursement [126].

A concern for more intensive use of data is that data generated in the routine care of patients may be limited for analytical purposes [127]. For example, such data may be inaccurate or incomplete. It may be transformed in ways that undermine its meaning (e.g., coding for billing priorities). For example, services or diagnoses that are highly reimbursed may be coded more reliably than other entities. It may exhibit the well-known statistical phenomenon of *censoring*, i.e., the first instance of disease in record may not be when it was first manifested (left censoring) or the data source may not cover a sufficient time interval to reflect the full course of disease (right censoring). Data may incompletely adhere to well-known standards, which makes combining it from different sources more difficult.

2.6 BENEFITS AND CHALLENGES OF INFORMATICS

With the massive adoption of EHRs in the United States driven by the HITECH Act, focused research on current systems and their impact on medical practice demonstrates a dichotomy between specific benefits and general dissatisfaction. Following an original systematic review in 2006 [83], three subsequent reviews using a similar methodology published in 2009 [128], 2011 [129], and 2014 [130] have shown persistent benefits.

At the same time, a growing number of problems have been identified related to the EHRs in clinical practice. One of these is excess focus on the computer over the patient [131, 132]. Another is the demise of traditional communications during care, such as radiology rounds [133]. There are also problems with documentation losing the patient's story through use of documentation templates that replace the narrative with elements such as checkboxes [134]. While this structuring of data assists with the use of data for other purposes, it loses the nuance of the patient's and clinician's narrative.

Finally, there is the problem of inappropriate use of "copy and paste," which may propagate errors and lead to uncertainty as to which provider rendered specific observations and recommendations (attribution) [135].

Another challenge is the increased time EHRs require of physicians and other clinicians, which takes time away from direct care of patients. Several recent time-motion studies have found that physicians spend up to half their workday interacting with the computer [136–138]. This has been shown to be one of the major causes of the growing epidemic of physician burnout [139]. It is important to note that studies of physician time over several decades, even in the pre-EHR era, have shown that physicians spend up to half of their time in "indirect care" not in the presence of patients, performing tasks such as documentation; in communication with patients, other physicians, and other clinical staff; and in transit [140, 141]. One reason for excess time requirements for the EHR in the United States is the increased billing and regulatory burden, as evidenced by the fact that physician notes in U.S. EHRs are longer, sometimes several-fold, than notes of physicians in other countries where EHRs are used [142]. There is no question that U.S. physicians spend too much time entering data into the EHR, but it must also be determined what is the optimal time devoted to documentation to enter data that informs the system to enable improved care.

Although EHRs have been touted to improve patient safety, there are also growing concerns that some aspects of their use may introduce new safety problems [75]. Two recent high-profile mishaps included CDS leading to massive overdosing of a common antibiotic [143] and poor communication between clinicians resulting in accidental discharge of a patient infected by the Ebola virus [144]. There are also growing concerns over the security of health information [145]. The year 2015 saw several massive security breaches, leading to exposure of records of over 100 million Americans [146–148]. The black market value of a medical record has been estimated to be 10 times that of a credit card number due to its containing larger quantities and more sensitive information [149].

It is clear that we must continue to improve EHRs and leverage their benefits to improve healthcare delivery. Two high-profile reports from several professional organizations have issued white papers specifying improvements in the EHR [150] and patient documentation [151]. The American Medical Association has laid out a set of principles for improved usability and interoperability [152]:

- Enhance physicians' ability to provide high-quality patient care.
- Support team-based care.
- Promote care coordination.
- Offer product modularity and configurability.
- Reduce cognitive workload.
- Promote data liquidity.
- Facilitate digital and mobile patient engagement.
- Expedite user input into product design and postimplementation feedback.

A recent report from the Pew Charitable Trusts, the American Medical Association, and MedStar Health focused on methods for improving EHR safety [153]. The report noted seven usability or safety issues where efforts should be focused:

1. *Data entry:* EHR data entry is difficult or not possible given the clinicians' work process, preventing the clinician from appropriately entering desired information.
2. *Alerting:* EHR alerts or other feedbacks from the system are inadequate because they are absent, incorrect, or ambiguous.
3. *Interoperability:* Interoperability is inadequate within components of the same EHR or from the EHR to other systems, hindering the communication of information.
4. *Visual display:* EHR display of information is confusing, cluttered, or inaccurate, resulting in clinicians having difficulty interpreting information.
5. *Availability of information:* EHR availability of clinically relevant information is hindered because information is entered or stored in the wrong location or is otherwise inaccessible.