

# Index

## A

- ALOGPS program, 195
- Antiretroviral drugs
  - clinical trials in, 133
  - HIV susceptibility to, 129–130
- Aquatic ecosystem, ANN models
  - bacteria, 66–67
  - crustacea, 71–74
  - fish, 75–77
  - insects, 74
  - protozoa, 67–71
- Artificial intelligence (AI)
  - categories of, 85–86
  - tools
    - expert system approach, 1–2
    - self-organizing maps, 2
- Artificial neural network ensemble (ANNE), 193
- Artificial neural networks (ANNs). *See also* Three-layer feedforward ANN model
  - applications of analytical chemistry
    - 13 C-NMR spectral simulation, 113–114
    - chromatographic parameter modeling, 100–105
    - electrophoresis, 105–112
    - heparanase inhibitors, 96–100
    - HETP derivatives, 93
    - mass spectral simulation, 116, 119–121
    - peptide mapping, 105–106
  - sulfonamides, 94–95
  - architecture of, 86
  - architecture of biological neurons, 16–17
  - Hopfield net, 21–22
  - Kohonen maps, 22–23
  - learning process
    - classifications and error correction methods, 19
    - nearest neighbour learning approach, 20
    - network topology
      - layers, 18–19
    - types of connections, 19
  - node character, 17–18
  - perceptron, 21
  - QSAR, 26
  - QSAR model
    - computational methods, 142–143
    - data preprocessing, 151
  - supervised learning, 144–150
  - supervised learning algorithms, 87–91
  - unsupervised learning algorithms, 91–92
- Associative neural network (ASNN)
  - data sharing and model accuracy, 197–198
  - Euclidian space, 192, 200
  - global and local models for, 190
  - GM and bias correction
    - ANNE and CCNN in, 192–193
  - computer resources for, 193–194
  - LIBRARY correction
    - ALOGPS program in, 195
    - drug design, 195
    - Gauss function prediction, 196–197
  - mean absolute errors (MAE), 198–199
  - lipophilicity and aqueous solubility, 199–200
  - log  $P$  and log  $S$  space, 200
  - mathematical formulation of, 190–192
  - property-based similarity, 199–201
  - software programs based on, 201
  - types of memory, 189
- Autoassociative artificial neural networks (AAANNs), 7
- Automatic relevance determination (ARD), 35, 39, 171

**B**

- Back-propagation algorithm (BPA), pesticide toxicity, 66
  - Chironomus riparius*, 74
  - Oncorhynchus mykiss*, 77
  - Tetrahymena pyriformis*, 68–69
- Back-propagation artificial neural networks (BPANNs), 90–91
- Backpropagation neural network (BPNN), 9, 26, 143–144
- Bayes' theorem, 29–31, 148
- Bayesian regularization, 149
- Bayesian regularized artificial neural network (BRANN), 69
  - advantage of, 26–27
    - automatic relevance determination (ARD), 39
    - nonlinear QSAR method, 36
    - relative independence of, 37
  - robustness and reproducibility, 37–38
  - back propagation, 30–31
- Bayesian inference
  - conditional probability, 29
  - microscopic hematuria (MH) diagnosis, 30
  - implementation difficulties, 34
  - incorporation of Bayes' theorem, 31–33
  - linear regression analysis, 26–28, 36
  - multiple linear regression (MLR) analysis, 34
  - overfitting problems, 33–34
  - QSAR models interpretation, 35
  - regression and regularization, 27–29
  - software, 40
- Bio-basis function, peptide classification, 167–169
- Biological neural network, 16–17
- Blood-brain barrier (BBB) permeability, 37–39

**C**

- C5 algorithm. *See* Data mining algorithms
- C5 decision tree, 244–247
- Capillary zone electrophoresis (CZE), 106–107
- Cascade correlation neural networks (CCNNs), 143, 151, 192–193
- ChemNet neural network, 156
- Chironomus riparius*, ANN insect model, 74
- Chromatographic parameter modeling, 100–105
- Comparative molecular surface analysis (CoMSA), 154
- Counterpropagation artificial neural network (CPANN) models, 71, 76

- Counterpropagation neural networks
  - advantage of, 46
  - predictive model construction, 54–56
  - property prediction, 51
  - reliability of prediction ability, 56–58
  - reversibility of, 58–59
  - supervised learning strategy, 50–51

**D**

- Daphnia magna*, ANN arthropoda model, 71–72
- Data mining
  - and neural networks, 236
  - definition, 235
  - drugs classification, 245
  - eukaryotic DNA, splice junction identification
    - C5 decision tree, 244–245
  - MLPs architecture, 244
  - HIV-1 protease inhibition, 246–247
  - molecular dynamics analysis, 245–246
  - NK1 receptors substance binding, 243–244
  - SUBSET algorithm, 237
  - TREPAN algorithm, 238–239
- Dayhoff matrix, peptide classification, 168
- Decision-making systems, ANN
  - HIV model system
    - ANN predictions, 133–136
    - antiretroviral drugs in, 129
    - drug-resistance mutations, 138
    - effective antiretroviral regimen, 133
    - materials used in, 130–131
    - phenotyping and genotyping, 130
    - real-life clinical data, 131–132
    - oncology, 128–129
- DNA Databank of Japan (DDBJ), 212–213
- 3D-QSAR studies, ANN, 154–155

**E**

- Electronic noses, 3, 5
- Electrophoretic mobility, CZE, 106–107
- Error correction methods, 19

**F**

- Fathead minnow. *See* *Pimephales promelas*
- Feedback neural networks
  - architecture of, 88–89, 121
  - Hopfield net, 21–22
  - Kohonen maps, 22–23
  - network topology, 18–20

- Feedforward neural networks
- architecture, 225–226
  - architecture of, 88–89
  - balanced training, 226
  - BPANN, 90–91, 94
  - data labeling, 220–221
  - data set
    - cross-training in, 218–219
    - number of free parameters (NFP), 218, 225
    - protein data bank and, 217–218
  - sequence-unique sets, 219–220
  - different layers of, 211–212
  - layers of, 121
  - network topology, 18–20
  - perceptron, 20–21
  - performance estimate
    - cross-validation and, 228
    - Q* measures, 229–230
    - receiver operating characteristic (ROC), 229
    - true positive rate and false positive rate, 228–229
  - validation/test sets for, 226–228
  - protein sequence input
    - amino acid type, 221–223
    - neighborhood aa information, 223–224
    - PSI-BLAST, 224–225
    - sparse encoding, 222–223
    - secondary structure prediction, 214–216
- Fuzzy prediction model, 176
- G**
- Gammarus fasciatus*, 72
- Global regression or classifier learning method (GM), ASNN, 190
- H**
- HAART (highly active antiretroviral therapy).  
See Antiretroviral drugs
- Hadamard transformation, 46, 52
- Hidden-layer neurodes, 27, 36–38
- Highly active antiretroviral therapy (HAART), 175–176
- HIV model system
- ANN predictions
    - antiretroviral therapy in, 134–135
    - size and diversity, 133–134
  - vs. rules-based predictions, 135–136
  - antiretroviral drugs in, 129
  - drug-resistance mutations, 138
  - materials used in, 130–131
  - methods
    - effective antiretroviral regimen, 133
    - real-life clinical data, 131–132
    - phenotyping and genotyping, 130
- HIV-1 protease
- knowledge extraction, 246–247
  - TREPAN algorithm, 243
- HIV-1 virus, 164, 174–176
- Hopfield net, 21–22
- 1-[2-Hydroxyethoxy)-methyl]-6-(phenylthio)thymine (HETP) derivatives, 93
- I**
- Infrared spectra interpretation model
- artificial neural networks (ANNs), 52
  - counterpropagation neural networks
    - architecture of, 54–55
    - prediction ability, 56
    - reliability of prediction, 56–58
    - reversibility of, 58–59
  - structural fragment prediction, 55
  - Kohonen neural network, 53–54
  - training set selection, 54
- Interpretative models, 35
- Inverse probability law, 29
- IR spectra
- Hadamard transformation, 53
  - Kohonen neural network, mapping, 53–54
  - organic compounds, data collection of, 52
  - simulation, 58–59
- K**
- Kohonen neural networks
- architecture of, 91
  - feedback network, 22–23
  - information and memory storage, 46–47
  - multidimensional data reduction, 4, 7
  - properties of, 49–50
  - topology of, 49
  - training
    - amount of correction, 48
    - competitive learning, 47
    - epochs, 48–49
    - unsupervised learning and basic architecture, 47
- Kohonen self-organizing maps, 4
- Kolmogorov's superposition theorem, 144
- Kovats retention index, chromatography, 100

**L**

Levenberg-Marquardt (L-M) algorithm, 98, 122  
 LIBRARY mode, 195–196  
 Lignins, spectral simulation, 113–114  
 Linear regression analysis, 26–28, 36  
 Local learning method (LM), ASNN, 190

**M**

M-of-N rule representation, 238  
 Matlab neural networks, 239  
 Matthews correlation coefficient (MCC), 177  
 Microtox™, 66  
 Midge larvae. *See Chironomus riparius*, ANN insect model  
 Minimum description length (MDL) principle, QSAR model, 148–149  
 Molecular dynamics trajectory analysis  
   knowledge extraction, 245–246  
   TREPAN algorithm, 242–243  
 Multilayer perceptrons (MLPs). *See also* Three-layer feedforward ANN model  
   architecture of, 244  
   decision tree extraction, 239  
   rule extraction, 237–238  
   training and testing sets, 243  
 Multiple linear regression (MLR) analysis, 34

**N**

Nearest neighbor learning approach, 21  
 NETLAB, 239  
 Neural device for direct structure-property correlation (NDDSPC), 143, 156  
 Neural networks. *See also* Counterpropagation neural networks; Kohonen neural networks  
   analytical chemistry  
     applications of, 6  
     olive oil assessment, 4–5  
   sensor technology, 5–6  
   applications  
     biological systems, 9–10  
     chromatography, 8–9  
     growing cell structure (GCS) networks, 8  
   spectroscopy, 8  
   characteristics and drawbacks of, 3  
   predictive models  
     construction of, 54–55  
     interpretative approaches, 35  
   principle of, 46  
   property prediction

  autoassociative artificial neural networks (AAANNs), 7  
   QSAR and QSPR models, 6–7  
 Neural networks (NN). *See* Artificial neural networks (ANN)  
 Neuron model, 87  
 NK1 receptors  
   knowledge extraction, 243–244  
   TREPAN algorithm, 240–241  
 Nonredundant databases, for proteins, 213–214

**O**

Ockham's Razor principle, 79  
*Oncorhynchus mykiss*, 77  
 Orthogonal peptide machine (OPM), 173–174  
 Orthogonal vector machine (OVM), 169  
 Overfitting, ANN models, 147–149  
 Overtraining, ANN models  
   phenomenon, 148–149  
   prevention of  
     early stopping procedure, 151  
     regularization, 149

**P**

PAPNET, ANN-based systems, 128  
 Pattern recognition algorithms, 5  
 PCA. *See* Principal components analysis  
 PCA sensor array studies, 6–7  
 Peptide classification, bioinformatics  
   Bayesian statistics, 171  
   disorder segments, 165  
   drug resistance, HIV  
     genotypic method for, 175  
     immunologic dynamics of, 175–176  
   RT and protease database, 176  
   feature extraction methods  
     bio-basis function, 167–168  
     frequency estimation, 167  
   orthogonal vector, 166  
   functional site and flanking residues, 164  
 issues and model validation  
   blind test and, 179–180  
   cross-validation and jackknife method, 179  
   protein-oriented validation, 180  
   resampling method, 178–179  
   type-I model, 181–182  
   type-I vs. type-II model, 183  
 type-II models, 182–183  
 orthogonal peptide machine, 173–174  
 protease cleavage sites and, 164–165

relevance peptide machine, 171–173  
 support peptide machine, 169–171  
 Peptide mapping, 105–106  
 Perceptron, 21. *See also* Feedforward neural network  
*Photobacterium phosphoreum*. *See* *Vibrio fischeri*, ANN model  
*Pimephales promelas*, 75–77  
 PLS model, 78–79  
   gammarid toxicity, 72–73  
   *Tetrahymena* toxicity, 69–70  
 Principal components analysis (PCA), 69  
 Principle of economy. *See* Ockham's Razor principle  
 Probabilistic artificial neural network (PANN) models, 71–72, 75–76  
 Protein data bank (PDB), 209, 217–218  
 Proteins  
   characterization of, 208  
   databases for, 212–214  
   functional analysis of, 209–210  
   homology and, 210–211  
   neural networks and  
     amino acid type, 221–223  
     data labeling, 220–221  
     overfitting avoidance, 218–219  
     performance estimate, 226–230  
     protein sequences selection, 217–218  
     PSI-BLAST, 224–225  
     secondary structure prediction, 215–216  
   sequence-unique sets, 219–220  
   structural determination of, 208–209

## Q

QSAR model building  
   Back-propagation neural networks, 143–144  
   descriptors  
     molecular field, 154–155  
     selection procedures, 150–151  
     substituent, 152–153  
     substructural, 153–154  
   superstructural, 154  
   drawbacks of, 150  
   global approximation of, 145  
   interpretation of, 146–147  
   PCA, 151  
   supervised learning  
     generalization, 147  
     MDL principle, 148–149  
     regression analysis, 144–146  
   regularization, 149

tasks performed, 144  
 vector-and graph-based, 155–156  
 VLANN, 154  
 Quantitative structure-activity relationship (QSAR). *See also* Artificial neural networks (ANNs), QSAR model; Three-layer feedforward ANN model  
   advantages of nonlinear method, 36  
   ANN used in, 26, 30  
   descriptors, molecules characteristics, 7  
   heparanase inhibitors, 96–100  
   HETP derivatives, 93  
   interpretation of neural net, 35  
   molecular properties, 6  
   sulfonamides, 94–95  
 Quantitative structure-property relationships (QSPR), 6–7  
 Quasi-Newton algorithm (QNA), 68

## R

Radial basis function artificial neural network (RBFANN), 70  
 Rainbow trout. *See* *Oncorhynchus mykiss*  
 Recursive cascade correlation neural network (RCCNN), 143, 156  
 Regression model, 175–176  
 Relevance peptide machine (RPM), 171–173  
 Relevance vector machine (RVM), 169  
 Resistance factor, HIV, 176

## S

Self-organizing maps (SOMs), 92. *See also* Unsupervised artificial neural networks  
   definition, 46  
   feedback networks, 19, 22–23  
   olive oil composition, 4  
   uses of, 7–8  
 Self-training artificial neural network (STANN), 102, 104–105  
 Sensitivity analysis, 35  
 Sensor array techniques, 6  
 Supervised artificial neural networks, 93, 121.  
   *See also* Three-layer feedforward ANN model  
     back-propagation learning rule, 91  
     schematic representation, 87–88  
 Supervised learning  
   ANN QSAR model, 144  
   biological studies, 19, 50–51  
   data preprocessing, 151

- Supervised learning (*cont.*)  
 descriptor  
   molecular field, 154–155  
   selection, 150–151  
   substituent, 152–153  
   substructural, 153–154  
 superstructural, 154  
 ensembles, 150  
 generalization, 147  
 MDL principle, 148–149  
 regression analysis, 144–146  
 regularization, 149  
 VLANN, 154  
 Support peptide machine (SPM), 169–171  
 Support vector machine (SVM), 169
- T**  
 Terrestrial ecosystem, ANN model, 77–78  
*Tetrahymena pyriformis*, ANN model, 68–71  
 TETRATOX database, 68  
 Three-layer feedforward ANN model  
   bacteria, 66–67  
   *Chironomus . riparius*, 74  
   gammarid toxicity, 73–74  
   honeybees, 77–78  
   *Pimephales promelas*, 75–77  
   *Tetrahymena pyriformis*, 68–71  
 Trained neural networks  
   decision trees and rules extraction  
     decompositional and pedagogical  
       approaches, 237  
     extraction benefits, 236  
     TREPAN algorithm, 238–239  
 Treatment change episode (TCE), 132  
 TrEMBL databases, 212–214  
 TREPAN algorithm  
   bioinformatics applications, 238, 247  
   biological and chemical problems  
     druglike molecules classification, 242  
     eukaryotic DNA sequences, splice  
       junction identification, 241–242  
     HIV-1 protease inhibition, 243  
     molecular dynamics simulations,  
       242–243  
   NK1 receptors substance binding,  
     240–241  
   cheminformatics applications, 247  
   classifier, 239  
   data distribution, 238  
   limitation of decision tree methods,  
     239–240
- U**  
 Unsupervised artificial neural  
   networks, 121  
   Kohonen network, 91  
   schematic representation, 87–88  
   self-organizing maps, 92  
 Unsupervised learning, 19, 22, 47
- V**  
*Vibrio fischeri*, ANN model, 66, 75, 80  
 Volume learning algorithm neural network  
   (VLANN), 143, 154–155
- W**  
 Waterflea. *See Daphnia magna*, ANN  
   arthropoda model
- X**  
 Xanthones, 114–116