

References

- [1] L. Han *et al.*, “Recent progress in the application of machine learning approach for predicting protein functional class independent of sequence similarity.,” *Proteomics*, vol. 6, pp. 4023–4037, Mar. 2006.
- [2] A. Al-Shahib, R. Breitling, and D. R. Gilbert, “Predicting protein function by machine learning on amino acid sequences – a critical evaluation,” *BMC Genomics*, vol. 8, p. 78, Mar. 2007.
- [3] A. Brazma and J. Vilo, “Gene expression data analysis,” *FEBS Lett.*, vol. 480, no. 1, pp. 17–24, Aug. 2000.
- [4] D. V. Nguyen, A. B. Arpat, N. Wang, and R. J. Carroll, “DNA microarray experiments: biological and technological aspects,” *Biometrics*, vol. 58, no. 4, pp. 701–717, Dec. 2002.
- [5] M. Kanehisa and S. Goto, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.
- [6] M. H. Serres and M. Riley, “MultiFun, a multifunctional classification scheme for Escherichia coli K-12 gene products,” *Microb. Comp. Genomics*, vol. 5, no. 4, pp. 205–222, 2000.
- [7] A. Ruepp *et al.*, “The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes,” *Nucleic Acids Res.*, vol. 32, no. 18, pp. 5539–5545, 2004.
- [8] “The Gene Ontology project in 2008,” *Nucleic Acids Res.*, vol. 36, no. Database issue, pp. D440–D444, Jan. 2008.
- [9] M. Ashburner *et al.*, “Gene ontology: tool for the unification of biology. The Gene Ontology Consortium,” *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.
- [10] P. Bork, T. Dandekar, Y. Diaz-Lazcoz, F. Eisenhaber, M. Huynen, and Y. Yuan, “Predicting function: from genes to genomes and back,” *J. Mol. Biol.*, vol. 283, no. 4, pp. 707–725, Nov. 1998.
- [11] I. M. Keseler *et al.*, “EcoCyc: a comprehensive database resource for Escherichia coli,” *Nucleic Acids Res.*, vol. 33, no. Database issue, pp. D334–337, Jan. 2005.
- [12] D. H. Haft, J. D. Selengut, and O. White, “The TIGRFAMs database of protein families,” *Nucleic Acids Res.*, vol. 31, no. 1, pp. 371–373, Jan. 2003.
- [13] H. W. Mewes *et al.*, “MIPS: a database for genomes and protein sequences,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 31–34, Jan. 2002.
- [14] “Gene Ontology Consortium.” [Online]. Available: <http://geneontology.org/>. [Accessed: 27-Mar-2016].
- [15] V. G. Krishnan and D. R. Westhead, “A comparative study of machine-learning methods to predict the effects of single nucleotide polymorphisms on protein function,” *Bioinformatics*, vol. 19, pp. 2199–2209, May 2003.
- [16] A. A. Freitas, D. C. Wieser, and R. Apweiler, “On the importance of comprehensible classification models for protein function prediction,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 7, no. 1, pp. 172–182, Jan. 2010.
- [17] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic Local Alignment Search Tool,” *J.Mol.Biol.*, pp. 403–410, Oct. 1990.
- [18] W. R. Pearson and D. J. Lipman, “Improved tools for biological sequence comparison,” *Proc. Natl. Acad. Sci. USA*, vol. 85, pp. 2444–2448, Apr. 1988.
- [19] T. Smith and M. Wayerman, “Identification of common molecular subsequences,” *J. Mol. Biol.*, pp. 195–197, 1981.
- [20] A. Bairoch, P. Bucher, and K. Hofmann, “The PROSITE database, its status,” *Nucleic Acids Research*, vol. 25, pp. 217–221, Oct. 1996.

- [21]T. . Attwood, M. . Beck, A. . Bleasby, and D. . Parry-Smith, "PRINTS- a database of protein motif fingerprints," *Nucleic Acids Research*, vol. 22, pp. 3590–3596, 1994.
- [22]"Sequence Motifs: Highly Predictive Features of Protein Function," in *Feature Extraction*, vol. 207, Springer Berlin Heidelberg, 2006, pp. 625–645.
- [23]S. F. Altschul *et al.*, "Grapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, pp. 3389–3402, Jul. 1997.
- [24]L. Holm and C. Sander, "Protein Structure Comparison by Alignment of Distance Matrices," *J. Mol. Biol.*, pp. 123–138, 1993.
- [25]J. C. Whisstock and A. M. Lesk, "Prediction of protein function from protein sequence and structure," *Quarterly Reviews of Biophysics*, vol. 36, no. 3, pp. 307–340, Aug. 2003.
- [26]D. Wang and B. Larder, "Enhanced Prediction of Lopinavir Resistance from Genotype by Use of Artificial Neural Networks," *The Journal of Infectious Diseases*, Mar. 2003.
- [27]L. J. Lancashire, C. Lemetre, and G. R. Ball, "An introduction to artificial neural networks in bioinformatics-application to complex microarray and mass spectrometry datasets in cancer studies.," *Briefings in Bioinformatics*, Feb. 2009.
- [28]M.-H. Chae, F. Krull, S. Korenzen, and E.-W. Knapp, "Predicting protein complex geometrics with a neural network," *Proteins*, Oct. 2009.
- [29]N. Donges, "Pros and Cons of Neural Networks," *Towards Data Science*, 17-Apr-2018. [Online]. Available: <https://towardsdatascience.com/hype-disadvantages-of-neural-networks-6af04904ba5b>. [Accessed: 10-Jun-2018].
- [30]joshiblog, "Advantages and disadvantages of hidden markov model," 10:38:33 UTC.
- [31]V. De Fonzo, F. Aluffi-Pentini, and V. Parisi, "Hidden Markov Models in Bioinformatics," *Current Bioinformatics*, pp. 49–61, 2007.
- [32]W. Majoros, M. Pertea, and S. Salzberg, "Efficient implementation of a generalized pair hidden Markov model for comparative gene finding.," *Bioinformatics*, pp. 1782–1788, February 2005.
- [33]Z. Yang, "Biological applications of support vector machines," *Breif Bioinform*, pp. 328–338, Dec. 2004.
- [34]L. Schietgat, C. Vens, J. Struyf, H. Blockeel, D. Koccev, and S. Dzeroski, "Predicting gene function using hierarchical multi-label decision tree ensembles," *BMC Bioinformatics*, vol. 11, Jan. 2010.
- [35]C. Vens, J. Struyf, L. Schietgat, S. Dzeroski, and H. Blockeel, "Decision trees for hierarchical multi-label classification," vol. 73, no. 2, pp. 185–214, Nov. 2008.
- [36]S. L. Wong *et al.*, "Combining biological networks to predict genetic interactions," *PNAS*, Sep. 2004.
- [37]D. Che, Q. Liu, K. Rasheed, and X. Tao, "Decision tree and ensemble learning algorithms with their applications in bioinformatics," *Adv Exp Med Biol.*, 2011.
- [38]Y. Huang and Y. Li, "Prediction of protein subcellular locations using fuzzy k-NN method," *Bioinformatics*, vol. 20, no. 1, Apr. 2003.
- [39]P. Horton *et al.*, "WoLF PSORT: protein localization predictor," *Nucl. Acids Res.*, vol. 35, Apr. 2007.
- [40]L. Lan, N. Djuric, Y. Guo, and S. Vucetic, "MS-kNN: protein function prediction by integrating multiple data sources," *BMC Bioinformatics*, vol. 14, Jul. 2010.
- [41]S. M. Thede, "An Introduction to Genetic Algorithms," *Journal of Computing Sciences in Colleges*, vol. 20 Issue 1, pp. 115–123, Oct. 2004.
- [42]D. J. Sobajic, *Neural Network Computing for the Electric Power Industry: Proceedings of the 1992 Inns Summer Workshop*. Psychology Press, 2013.
- [43]C.-F. Tsai, W. Eberle, and C.-Y. Chu, "Genetic algorithms in feature and instance selection," *Knowl.-Based Syst.*, vol. 39, pp. 240–247, Feb. 2013.
- [44]A. S. Perera and W. Perrizo, "Gene Function Prediction.," in *ResearchGate*, 2009, pp. 26–31.

- [45] M. Middlemiss and G. Dick, "Design and Application of Hybrid Intelligent Systems," A. Abraham, M. Köppen, and K. Franke, Eds. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2003, pp. 519–527.
- [46] J. Brownlee, "Supervised and Unsupervised Machine Learning Algorithms," *Machine Learning Mastery*, 16-Mar-2016. .
- [47] Z. Yao and W. L. Ruzzo, "A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data," *BMC Bioinformatics*, vol. 7, no. Suppl 1, p. S11, Mar. 2006.
- [48] "Data Mining Concepts." [Online]. Available: https://docs.oracle.com/cd/B19306_01/datamine.102/b14339/3predictive.htm. [Accessed: 07-May-2016].
- [49] S. M. Thede, "An Introduction to Genetic Algorithms," *J Comput Sci Coll*, vol. 20, no. 1, pp. 115–123, Oct. 2004.
- [50] "Han and Kamber: Data Mining---Concepts and Techniques, 2nd ed., Morgan Kaufmann, 2006." [Online]. Available: <http://web.engr.illinois.edu/~hanj/bk2/slidesindex.htm>. [Accessed: 11-Mar-2017].
- [51] J. L. DeRisi, V. R. Iyer, and P. O. Brown, "Exploring the metabolic and genetic control of gene expression on a genomic scale," *Science*, vol. 278, no. 5338, pp. 680–686, Oct. 1997.
- [52] C. M. Perou *et al.*, "Molecular portraits of human breast tumours," *Nature*, vol. 406, no. 6797, pp. 747–752, Aug. 2000.
- [53] S. Chu *et al.*, "The transcriptional program of sporulation in budding yeast," *Science*, vol. 282, no. 5389, pp. 699–705, Oct. 1998.
- [54] S. Friedland, A. Niknejad, M. Kaveh, and H. Zare, "An Algorithm for Missing Value Estimation for DNA Microarray Data," in *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings, 2006*, vol. 2, pp. II–II.
- [55] A. Gelman and J. Hill, "Missing-data imputation," in *Data Analysis Using Regression and Multilevel Hierarchical Models*, Cambridge University Press, 2006.
- [56] A. Mockus, "Missing Data in Software Engineering," in *Guide to Advanced Empirical Software Engineering*, F. Shull, J. Singer, and D. I. K. Sjøberg, Eds. Springer London, 2008, pp. 185–200.
- [57] E. Acuña and C. Rodríguez, "The Treatment of Missing Values and its Effect on Classifier Accuracy," in *Classification, Clustering, and Data Mining Applications*, D. D. Banks, D. F. R. McMorris, D. P. Arabie, and P. D. W. Gaul, Eds. Springer Berlin Heidelberg, 2004, pp. 639–647.
- [58] J. Kaiser, "Algorithm for Missing Values Imputation in Categorical Data with Use of Association Rules," *ArXiv12111799 Cs*, Nov. 2012.
- [59] J. C. Wayman and P. D. *Multiple Imputation for Missing Data: What is It and How Can I Use It*. 2003.
- [60] A. P. Gasch, M. Huang, S. Metzner, D. Botstein, S. J. Elledge, and P. O. Brown, "Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p," *Mol. Biol. Cell*, vol. 12, no. 10, pp. 2987–3003, Oct. 2001.