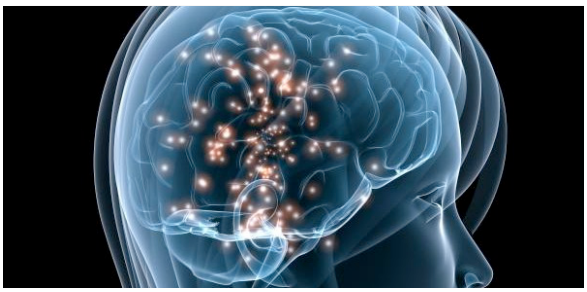
Artificial Neural Networks in Protein Secondary Structure Prediction: A Critical Review of Present and Future Applications



BIOMEDIN 231: Computational Molecular Biology

Professor and Instructor:
Doug Brutlag and Dan Davison

Student:
Anthony C. Chang, MD, MBA, MPH
AChang007@aol.com

Due date: June 11th, 2014

Abstract

Today, as a result of information sharing and rapid sequencing, bioinformatics is deluged with data ("big data") and faces the daunting multi-dimensional challenges of sequence, structure, expression, and pathway analyses. The explicit understanding of protein secondary structure and beyond can yield great benefit to understanding of human diseases and development of therapeutic drugs and enzymes in the ensuing years. The myriad of machine learning methodologies have their respective relative strengths in various bioinformatics disciplines: hidden Markov model for sequence and profile alignment; support vector machine for protein fold recognition; Bayesian networks and gene regulatory network; and artificial neural networks (ANN) and protein secondary structure prediction.

Due to the relatively high capability of ANN to elucidate complex patterns, classify massive data, and make accurate predictions in large complicated amino acid/ protein data sets, ANN has become an essential methodology in the current era of artificial intelligence in computational molecular biology problems such as sequence encoding and output interpretation, sequence correlations, DNA and RNA nucleotide sequences, and protein structure prediction. While there are numerous studies that compared ANN with at least one other machine learning methodology, there is no canonical approach in predicting protein secondary structure.

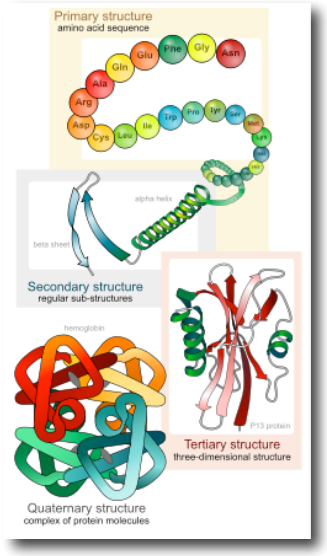
It is possible that the more ideal solution to protein secondary structure prediction is an advanced ensemble strategy of selected individual machine learning tools (beyond the traditional ensemble methods such as bagging, boosting, and random forest). Novel ANN-inspired methodologies and strategies can be implemented to provide predictions for higher levels of protein structures (tertiary and quaternary) so that protein function can be understood and drug/enzyme therapy can be designed in the future.

A new type of thinking is essential if mankind is to survive and move to higher levels.

Albert Einstein

Introduction

It is almost 25 years since Hunter's sentinel paper *Artificial Intelligence and Molecular Biology* appeared in *AI Magazine* ([1]). Today, as a result of information sharing and rapid sequencing, bioinformatics is deluged with data ("big data") and faces the daunting multi-dimensional challenges of sequence, structure, expression, and pathway analyses ([2]).



Protein structure (see figure at left) consists of amino acid sequence (primary structure) leading to sub-structures of α-helices, β-sheets, and random coils as a result of hydrogen bonds (secondary structure). These secondary structures, if certain attractions are present, can lead to a three-dimensional structure (tertiary structure) and even complexes of protein molecules (quaternary structure) ([3]).

With exponential growth of sequencing data, it is even more vital and urgent now to couple computational molecular biology and artificial intelligence strategies to elucidate protein structure and function. **Protein secondary structure** was first predicted by a Bayesian classifier machine learning method in 1978 with the investigators using X-ray crystallography data as training data set ([4]).
The explicit understanding of protein secondary structure and beyond can yield great benefit to understanding of

human diseases and development of therapeutic drugs and enzymes in the ensuing years.

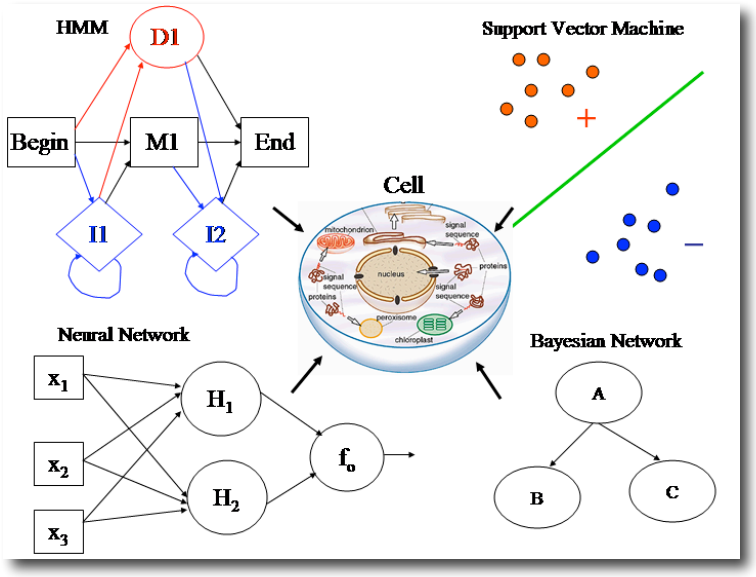
This manuscript will focus primarily on current and future application of the machine learning methodology of **artificial neural network** on **protein secondary structure prediction**.

Background: Machine Learning and Artificial Neural Network

Machine Learning

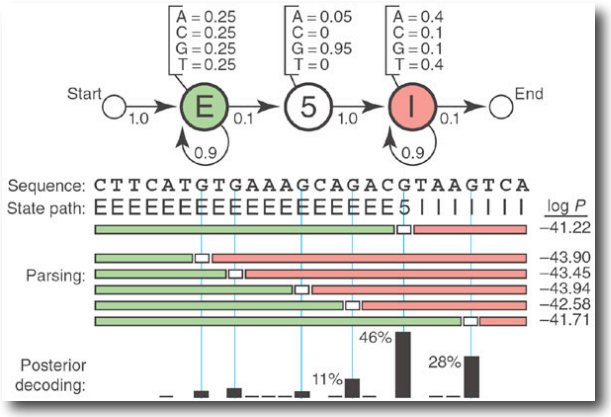
A myriad of **artificial intelligence** methodologies along with massive parallel computing are being applied to meet this daunting challenge of protein secondary structure prediction. Standard statistical techniques such as generalized linear models and discriminant analysis have limitations when there are highly nonlinear and complex interactions. In the current era of bioinformatics, **machine learning**, or the design and development of algorithms that allow the computers to "learn", enables computer programming to improve performance with biological data sets ([5])([6]).

The myriad of machine learning methodologies have their respective relative strengths in various bioinformatics disciplines (see figure below): **hidden Markov model** for sequence and profile alignment; **support vector machine** for protein fold recognition; **Bayesian networks** and gene regulatory network; and **artificial neural networks** and protein secondary structure prediction ([7]).



In predicting protein secondary structure, the more commonly used machine learning techniques have had a relatively long investigative history and include ([8]):

Hidden Markov models (HMM):



This machine learning methodology, "hidden" as states are unobserved or not visible, uses a probabilistic model and is considered the simplest **dynamic Bayesian network**. It is used widely in speech and handwriting recognition.
The figure to the left shows a HMM for 5' splice site recognition with three states labeled "E" for exon, "5" for 5SS, and "I" for Intron ([9]).

Support vector machines (SVM):

A supervised learning model, SVM are associated with learning algorithms and classification and regression analysis in its construction of a **hyperplane**; in other words, this technique extend to patterns that are not

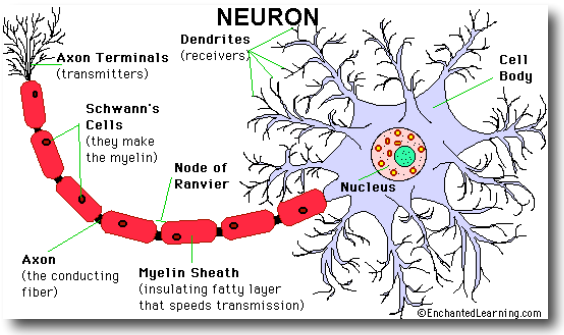
linearly separable and is identified as a **kernel** function. These machines are widely used in bioinformatics due to its capability to handle high dimension data, its flexibility in modeling diverse types of data, and its high accuracy.

Bayesian networks:

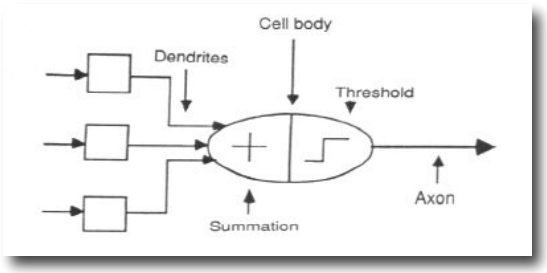
Also called Bayes nets or belief network, this machine learning methodology is a **probabilistic graphical model** (hybrid of probability theory and graph theory) that represents a set of random variables (as nodes on the graph) with their conditional dependencies (as edges between nodes). Bayesian networks are easier to comprehend than ANN and can be represented by a **directed acyclic graph** (DAG).

Artificial Neural Networks

Architecture



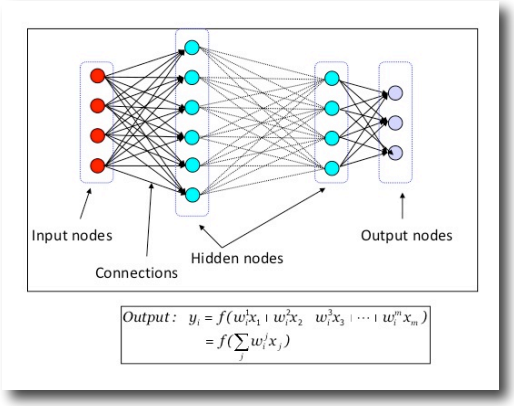
Artificial neural networks (ANN), with both statistical (linear regression and discriminant analysis) and artificial intelligence roots, are information processing units that that are modeled after the brain and its 100 billion neurons (see figure at right). In a neuron, the distal and proximal dendrites receive signals and communicate to the cell body, which in turn communicates with other neurons via its axon and its terminals.



Similarly, an ANN receives **inputs** (dendrites) that are processed with influence by weights to become **outputs** (axon)(see figure at left). The neurons or **nodes** interconnect with

informational flows (unidirectional or bidirectional) at various weights or strengths.

The simplest architecture is the **perceptron**, which consists of 2 layers (input and output layers) that are separated by a linear discrimination function ([10]). In a **multi-layer perceptron (MLP)** model (see figure at right), there are three layers: the input nodes, the hidden nodes layer, and the output nodes.



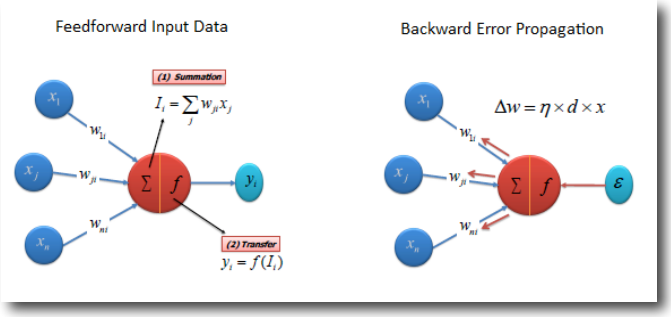
Other ANN types include: learning vector quantification, radial basis function network, Hopfield networks, and Kohonen self-organizing maps (SOMs) ([11]).

Learning/ Training

In a **feed-forward** neural network architecture, a unit will receive input from several nodes or neurons belonging to another layer. These highly interconnected neurons therefore form an infrastructure (similar to the biological central nervous system) that is capable of learning by successfully perform pattern recognition and classification tasks.

Training of the ANN is a process in which learning occurs from representative data and the knowledge is applied to the new situation. This training or learning process occurs by arranging the algorithms so that the weights of the ANN are adjusted to lead to the final desired output. The learning in neural networks can be **supervised** (such as the multilayer perceptron that trained with sets of input data) or **unsupervised** (such as the Kohonen self-organizing maps which learn by finding patterns). Neural networks can also perform both regression and classification.

The ANN learning process consists of both a forward and a backward propagation process. The **forward propagation** process involves presenting data into the ANN whereas the important **backward propagation** algorithm (see figure below) determines the values of the weights for the nodes during a training phase. This latter process is accomplished by directing the errors for input values backwards so that corrections for the weights can be made to minimize the error of actual and desired output data. A **recurrent neural network** is a series of feed-forward neural networks sharing the same weights and is good for time series data. ANN can therefore **extract patterns** or **detect trends** from complicated and imprecise data sets.



Application of ANN to bioinformatics needs the following strategy ([12]): extraction of features from **molecular sequences** to serve as training/prediction data; **preprocessing** that consists of feature selection and encoding into vectors of real numbers; **neural network** for training or prediction; **postprocessing** that consists of output encoding from the neural network; and finally the myriad of **applications** (such as sequence analysis, gene expression data analysis, or protein structure prediction).

Strengths and Limitations of ANN

The following are some distinct strengths of ANN ([13]):

1) **Adaptive learning**: ANN can learn new tasks with relatively small amount of training data (even non-linear data albeit with some difficulty);
2) **Self-organization**: ANN can organize its own data to achieve pattern recognition;
3) **Real-time operation**: ANN computations can be performed in parallel and therefore are relatively efficient;
4) **Computationally powerful**: ANN can predict complex biological patterns with training; and
5) **Fault tolerance**: ANN can retain performance with destruction of parts of the infrastructure.

The following are some distinct limitations of ANN:

1) **Suboptimal speed**: If the number of neurons are high, ANN can be computationally slow and challenging;

2) **Overfitting tendency**: ANN can overfit especially if the training data is small and thus cannot generalize well to the unseen data;
3) **Design difficulty**: ANN is challenging to design and train especially with complex problems;
4) **Lack of transparency**: ANN is not as easy to comprehend as other machine learning systems (unless it is a single layer perceptron); and
5) **Data pre-processing**: ANN performs better if there is data normalization as part of pre-processing.

Artificial Neural Network: Current Applications in Protein Structure Prediction

Due to the relatively high capability of ANN to elucidate complex patterns, classify massive data, and make accurate predictions in large complicated amino acid/ protein data sets, ANN has become an essential methodology in the current era of artificial intelligence in computational molecular biology problems such as sequence encoding and output interpretation, sequence correlations, DNA and RNA nucleotide sequences, and protein structure prediction ([14])([15]). Recent improvements in accuracy by using **statistical context-based scores (SCORPION)** ([16]) as well as incorporating tertiary structure information with the **ROSETTA *de novo* tertiary structure prediction method** ([17]) demonstrate continual improvements in ANN approach to protein structure prediction.

Early ANN Models

ANN, with its ability to learn complex functions from large data sets, can be ideally used for protein structure prediction and has been doing so with increasing accuracy, particularly after it has been used to predict the number of structural domains from protein structure. ANN is used to predict secondary structure by creating a data set with input sequences with output secondary structures and encoding both the input and output to the ANN.

First in the ANN application in molecular biology was the work of Qian, who used the **feed-forward perceptron model** to predict secondary globular protein structure with a success rate of 64% on three types of secondary structure ([18]). The input layer has 273 units, the output layer three sigmoidal units with the hidden layer of 40 sigmoidal units. Following the advent of BLAST as the basic local alignment tool, Rost (see figure next page) used evolutionary information contained in multiple sequence alignments as inputs to the neural network called the **PHD prediction server** and was able to increase the overall

accuracy to above 72% ([19]). The potential overfitting problem from the Qian model was addressed with **ensemble averages** achieved by training various networks ([20]).

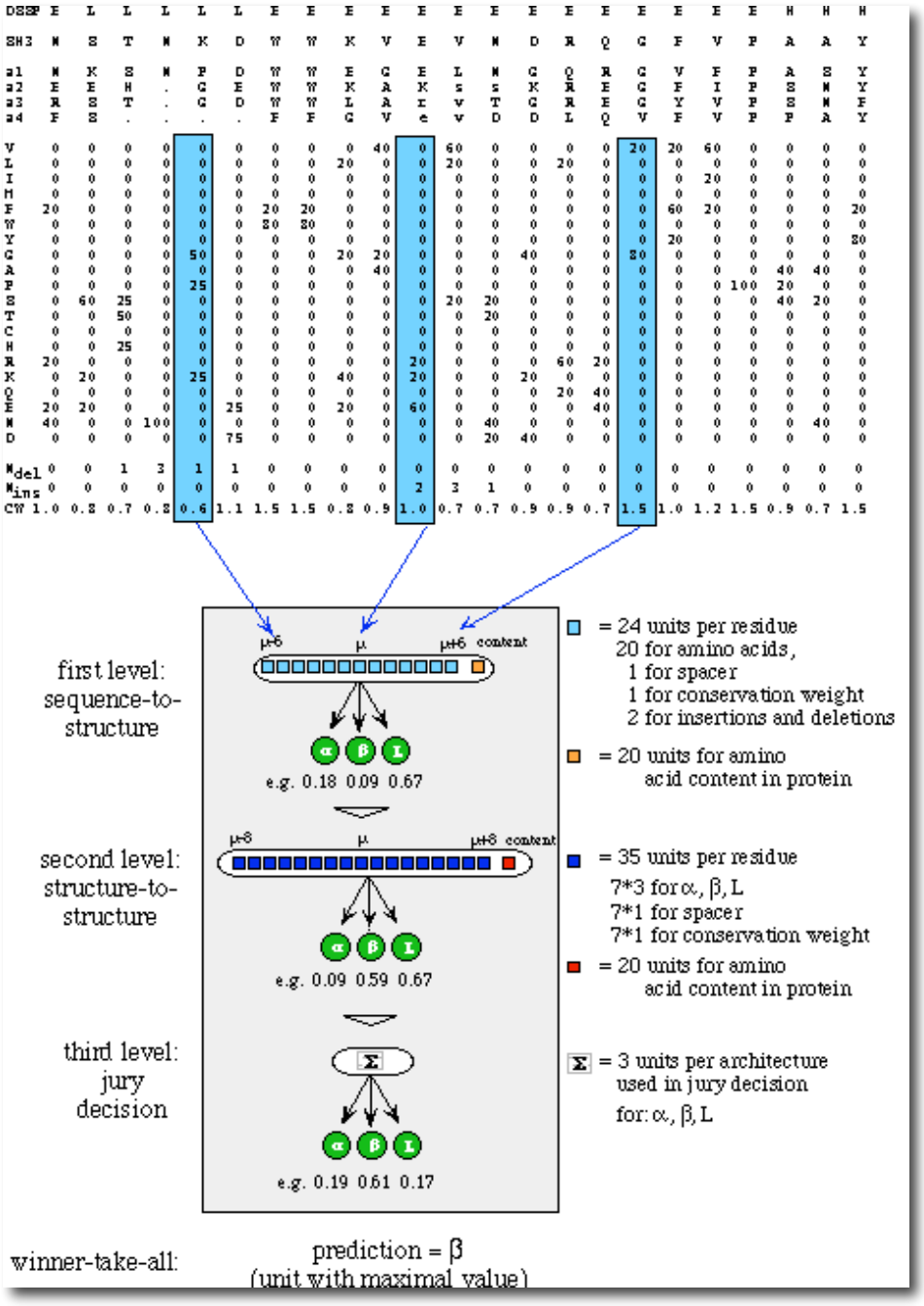
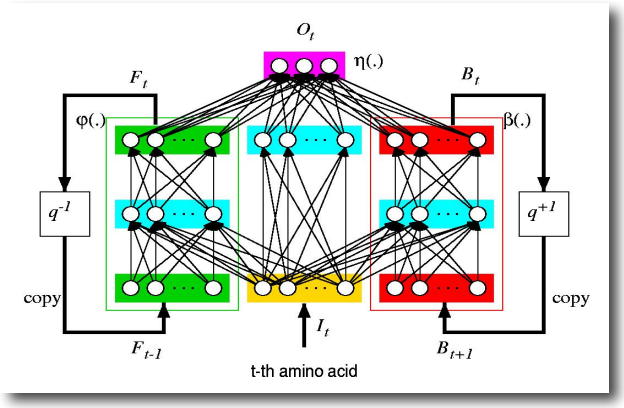


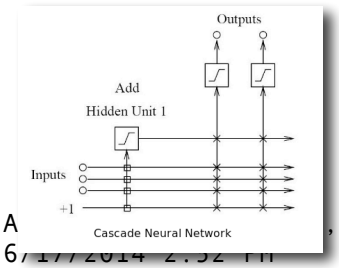
Figure. Network architecture (PHD). A profile-based neural network system for protein secondary structure prediction. The **multiple alignment** is seen at the top with a profile of amino acid occurrences compiled. Then the alignment is fed into the **neural network**, which consists of 3 layers: 2 network layers and an additional layer for averaging over the independently trained networks (www.rostlab.org).



Recently, Baldi also used a large-scale **bidirectional recursive ANN** model (see figure at left where O_t is the output layer) for his work ([21]) on secondary structure with wide range of data structures of variable sizes, formats (sequences, trees, lattices, etc.), and dimensionality (2D, 3D, etc.); this is an advantage over feedforward neural networks that can only process input data of a fixed size. Cheng (from the Baldi group) later used a further **modified bidirectional recursive ANN** model in predicting disordered region with a true positive rate of 93% and a false positive rate of 5% ([22]).

Other ANN-related Models

There are several noteworthy ANN-related algorithms during this era of ANN-inspired protein secondary structure prediction projects. Jones reported the **position-specific scoring matrices approach (PSIPRED)** that was demonstrated to have Q3 score (percentage of correctly classified residues but does not account for over-prediction) of between 76.5% to 78.3% ([23]). Petersen achieved an even higher overall performance of 77-80% for three-state prediction (helix, strand, and coil) using 800 neural network predictions and a set of 126 protein chains with **output expansion** to provide prediction concomitantly for neighboring residues ([24]). Cuff subsequently showed application of **multiple sequence alignment profiles (JNET)** where training with different representations of the same alignment data resulted in an average Q3 score of 84%, higher than the PHD program ([25]).



Finally, Wood used an alternative **cascade-correlation architecture** ([26])(see figure at left), which is a constructive supervised learning

algorithm (with no hidden units at the beginning) that was demonstrated to have a faster speed over the other back-propagation ANN algorithms while matching results (75.6% vs 75.7% for Q3).

Artificial Neural Network vs Other Machine Learning Methodologies in Protein Structure Prediction

The performance of present methods in protein structure prediction is assessed in the **Critical Assessment of Techniques for Protein Structure Prediction** (CASP) ([27]). While there are numerous studies that compared ANN with at least one other machine learning methodology, there is no canonical approach in predicting protein secondary structure.

ANN and HMM

HMM as a probabilistic model provides relevant information related to the sequence-structure relationship but overall the accuracy has been below the other machine learning techniques.

In comparison with three other methodologies including statistical approach, nearest neighbor method, and HMM, ANN approach proved to the most successful amongst the four methodologies ([28]). In another study on ubiquitin protein structure prediction, however, showed that the SVM model was superior to both ANN and HMM (86.8% for SVM vs 83.4% for ANN and 72.1% for HMM) ([29]).

ANN and SVM

In protein structure prediction, SVM has been demonstrated to be superior
in predicting the location of turns ([30]). Another potential advantage for SVM is the requirement for a relatively small training set as to avoid overfitting of the data ([31]).

A study that directly compared ANN with SVM demonstrated that the former proved to have much better accuracy and

take much less training and computation time (with SVM also requiring much larger memory and powerful processor) ([32]). Contrary to the aforementioned study, a current demonstrated that SVM outperformed ANN with an overall accuracy of 89.3% in identification of lipid-binding proteins (LPBs) from non-LBPs ([33]).

ANN and Bayesian Network

The first machine learning technique in protein structure prediction was partly based on Bayesian statistics ([34]). Bayesian network works well over large databases, an advantage over heuristic methods and compared to ANN, Bayesian networks are less opaque ([35]). There is a paucity of published literature comparing ANN directly with Bayesian network in protein structure predictions.

ANN and Other Machine Learning Methods

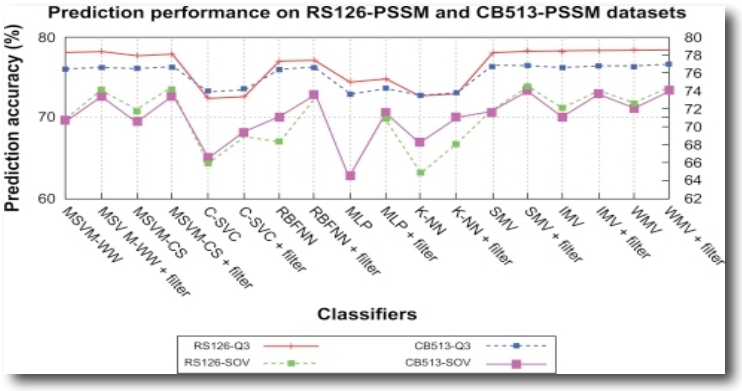
Yi used a neural network and **nearest-neighbor method** for predicting secondary structure of proteins and demonstrated that the nearest-neighbor method had an overall three-state accuracy of 72%, higher than that of neural network ([36]). In addition, Simas's group also described that **nonlinear dimensional reduction** in protein secondary structure prediction yielded similar results compared to ANN ([37]).

Ensemble Strategy

It is possible that the more ideal solution to protein secondary structure prediction is an advanced **ensemble strategy** of selected individual machine learning tools (beyond the traditional ensemble methods such as bagging, boosting, and random forest)([38]).

One example of this type of strategy is the **SCRATCH prediction server** ([39]). This server software suite is based on a combination of recursive ANN and other machine learning techniques (SVMs and Bayesian networks along with graph matching algorithms). The three class per amino acid accuracy of this tool was about 77%.

Another example is the **YASPIN software** ([40]), which utilizes a combination of separately-trained ANN (315 input and 7 output units) with HMM to make its protein secondary structure predictions. This hybrid method uses a single-layer ANN for predicting the secondary structures in a 7-state local structure scheme and then optimizes the output using a hidden Markov model (thus a "hidden neural network"). The overall prediction accuracy (compared to other prediction methods) was the highest with the 3-state per-residue prediction accuracy measure of 77.05% while predicting at a faster speed.



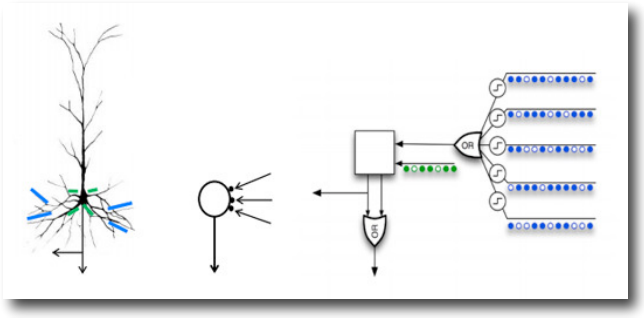
Recently, Bouziane explored various **ensembles** of machine learning techniques and proposed a system to yield significant performance gains ([41]). The study (see figure at left) showed that three-combination schemes yielded higher performance (y-axis) in benchmark data sets compared to individual classifiers.

Artificial Neural Network: Future Applications in Protein Structure Prediction

Novel ANN-inspired methodologies and strategies can be implemented to provide predictions for higher levels of protein structures (tertiary and quaternary) so that protein function can be understood and drug/enzyme therapy can be designed in the future.

Hierarchical Temporal Memory

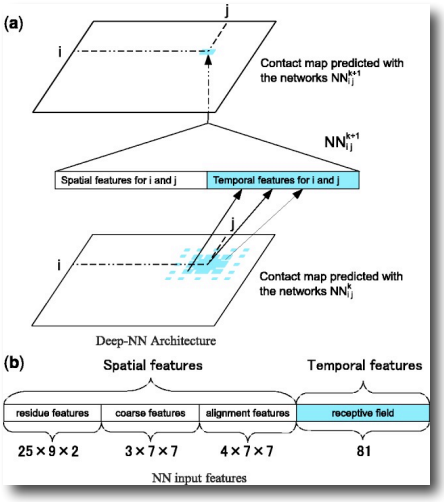
Jeff Hawkins of the Redwood Center for Theoretical Neuroscience proposed a modified model of the traditional ANN: a **hierarchical temporal memory (HTM) model** with **sparsity distributed representations** and **cortical learning algorithms** to create a memory-prediction framework that can find patterns in noisy data ([42]).



The figure to the left shows a biological and artificial neuron in ANN on the left and middle, respectively. At the figure's right is the HTM cell which is more biologically realistic than the ANN neuron. It is in essence a hybrid of the two neurons.

The HTM neuron has a **proximal dendrite** (in green) with linear summation and feed forward connections, **distal dendrites** (in blue) with dozens of regions and threshold coincidence detectors as well as connections to other cells in layer, **synapses** with thousands on distal dendrites and hundreds on proximal dendrites, and an **output** that is active (fast or burst) or predictive (slow). This concept is already utilized in a ANN-inspired **machine intelligence software Grok** (Numenta, Inc.) that is designed for anomaly detection but has not yet explored possibilities in bioinformatics or medicine ([43]).

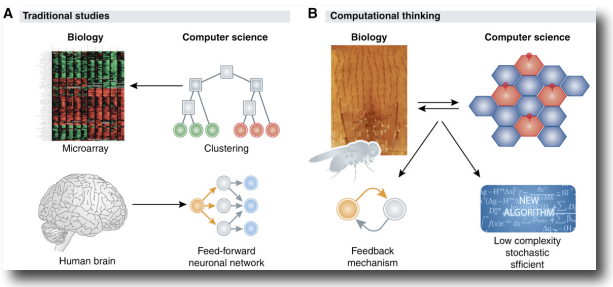
Deep Neural Network (DNN)



Deep learning, or deep machine learning, involves deep architectures for learning higher level of representations from data ([44]). Recent work on the use of deep architecture in prediction of protein contact maps by the Baldi group demonstrates early success with the deep-NN modular architecture consists of a 3D stack of neural networks with identical architecture but different weights ([45])(see figure at left). The bottom of the figure illustrates the input feature vector of each level of neural network.

Neural networks and deep learning, together, is providing solutions for problems in areas such as image recognition and natural language processing. Two such approaches include **deep belief nets** and **stacked auto-encoders**. Recent advances in semi-supervised manifold learning and evolutionary programming approaches in deep learning may accelerate the use of this strategy. These higher level abstractions may be ideally suited for more advanced protein structure predictions that involve tertiary and quaternary structures in three-dimensional configurations.

Biological Computational Thinking

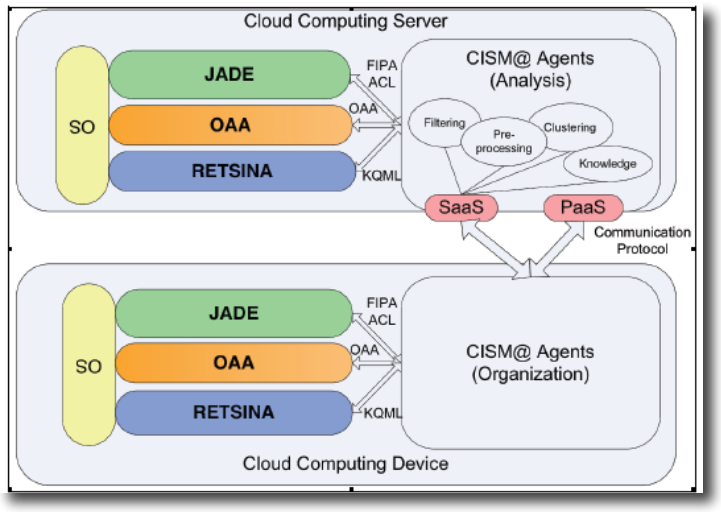


One future methodology will combine biology with computer science and can result in a **synergistic feedback mechanism** similar to nature's physiological central

nervous system feedback loop (see figure below) ([46]). In other words, in the design of algorithms about molecular biology processes is no longer unidirectional, and the feedback can lead to improve the design of the algorithms.

Cloud Computing Integrated on Service-Oriented Multi-Agent (CISM)

The protein structure prediction realm can be more collaborative if the results of the prediction investigations are gathered in a cloud-computing model. This virtualization of the protein structure prediction investigations at the international level can accelerate the progress as some of the work will no longer need to be duplicated. In addition, there are additional synergies that can result from a more intimate level of collaboration. There is preliminary work on using ANN concepts in cloud computing platform for bioinformatics ([47]).



The **service-oriented architecture** (SOA) and agent frameworks can provide this artificial intelligence-inspired network (see figure at left). An advanced version of this concept is the so-called **cloud computing integrated on service-oriented multi-agent** or **CISM**, which creates a dynamic and distributed research milieu with inter-operability ([48]). CISM will have four components: 1) platform as a service (PaaS); 2) agent platform for massive data analysis; 3) software as a service (SaaS); and 4) communication protocol. This framework has already been adapted for the microarray environment and can also be implemented for the protein structure prediction realm.

Conclusion

The accuracy of ANN in protein secondary structure prediction has steadily increased in the past several decades from implementation of elements such as evolutionary information, ensemble averaging, output expansion, and position-specific scoring matrices. With the exponential rise in genomic sequencing capability and big data, however, there is now an increasing **sequence-structure gap** in molecular biology ([49]). Due to the heterogeneity of the study constructs, it is difficult to demonstrate clear superiority of any of the key machine learning methodologies. Advanced ANN-inspired models, however, are continually being explored for even higher accuracy in protein structure prediction.

The optimal ANN strategy in protein secondary structure prediction could well be a **ensemble approach** with utilization of a myriad of techniques to include most if not all aspects of machine learning with ANN being an essential component ([50]). In addition, innovative strategies such as hierarchical temporal memory, deep neural networks, computational biology thinking, and cloud computing can all add to the ANN portfolio. Lastly, novel applications of other artificial intelligence techniques such as graph theory, data visualization, and biomedical framework, along with *in silico* simulation and even DNA computing, will give rise to the eventual vision of routine tertiary and quaternary protein structure prediction and perhaps the nascent field of "**biomolecular intelligence**", where molecular biology and artificial intelligence converge and become one discipline.

References

[1] Hunter L. Artificial Intelligence and Molecular Biology. *AI Magazine* 1990; 11(5): 27-36.

[2] Velentini G. Computational Intelligence and Machine Learning in Bioinformatics. *Artificial Intelligence in Medicine* 2009; 45: 91-96.

[3] Lee D et al. Predicting Protein Function from Sequence and Structure. *Nature Reviews Molecular Cell Biology* 2007; 8: 995-1005.

[4] Garnier J et al. Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. *Journal of Molecular Biology* 1978; 120: 97-120.

[5] Keedwell E and Narayanan A. *Intelligent Bioinformatics*, John Wiley and Sons, West Sussex, England, 2005.

[6] Larranaga P et al. Machine Learning in Bioinformatics. *Briefings in Bioinformatics* 2005; 7(1): 86-112.

[7] Cheng, JL. Machine Learning Methods for Bioinformatics. Course at University of Missouri.

[8] Inza I et al. Machine Learning: An Indispensable Tool in Bioinformatics in Matthiesen R (ed) *Bioinformatics Methods in Clinical Research*, Springer Science, 2010.

[9] Eddy SR. What is a Hidden Markov Model? *Nature Biotechnology* 2004; 22: 1315-1316.

[10] Rosenblatt F. The Perceptron- *A Perceiving and Recognizing Automaton*. Report 85-460-1, Cornell Aeronautical Laboratory, 1957.

[11] Kohonen T. A Simple Paradigm for the Self-Organized Formation of Structured Feature Maps in Amari S (ed) *Competition and Cooperation in Neural Nets*, 1982.

[12] Keedwell E and Narayanan A. Neural Networks in *Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems*. John Wiley and Sons, Chichester, UK, 2005.

[13] Stergiou C and Siganos D. *Neural Networks*. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/re[prt.html

[14] Wu CH et al. Neural Networks. Encyclopedia of Life Sciences 2010; 1-5.

[15] Baldi P and Brunak S. Neural Networks: Applications in *Bioinformatics: The Machine Learning Approach*, MIT Press, 2001.

[16] Yaseen A and Li Y. Context-based Features Enhance Protein Secondary Structure Prediction Accuracy. *Journal of Chemical Information and Modeling* 2014; 54(3): 992-1002.

[17] Meiler J and Baker D. Coupled Prediction of Protein Secondary and Tertiary Structure. *Proceedings of the National Science Society* 2003; 100(21): 12105-12110.

[18] Qian N and Sejnowski TJ. Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *Journal of Molecular Biology* 1988; 202(4): 865-884.

[19] Rost B and Sander C. Combining Evolutionary Information and Neural Networks to Predict Protein Secondary Structure. *Proteins: Structure, Function, and Genetics* 1994; 19: 55-72.

[20] Riis SK and Krogh A. Improving Prediction of Protein Secondary Structure Using Structured Neural Networks and Multiple Sequence Alignments. *Journal of Computational Biology* 1996; 3: 163-183.

[21] Baldi P and Plllastri G. The Principled Design of Large-Scale Recursive Neural Network Architectures-DAG-RNNs and the Protein Structure Prediction Problem. *Journal of Machine Learning Research* 2003; 4: 575-602.

[22] Cheng JL et al. Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining and Knowledge Discovery* 2005; 11:213-222.

[23] Jones DT. Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *Journal Molecular Biology* 1999; 292: 195-202.

[24] Petersen TN et al. Prediction of Protein Secondary Structure at 80% Accuracy. *Proteins: Stucture, Function, and Bioinformatics* 2000; 41(1): 17-20.

[25] Cuff JA and Barton GJ. Application of Multiple Sequence Alignment Profiles to Improve Protein Secondary Structure Prediction. *Proteins* 2000; 40: 502-511.

[26] Wood MJ and Hirst JD. Predicting Protein Secondary Structure by Cascade-Correlation Neural Networks. *Bioinformatics* 2004; 20: 419-420.

[27] Moult J et al. Critical Assessment of Methods of Protein Structure Prediction (CASP)- Round X. *Proteins: Structure, Function, and Bioinformatics* 2004; 82(S2): 1-6.

[28] Bordoloi H and Sarma KK. Protein Structure Prediction Using Artificial Neural Network. *International Journal of Computer*

Applications on Electronics, Information, and Communication Engineering 2011; 3: 22-25.

[29] Jaiswal K. Prediction of Uniquitin Proteins Using Artificial Neural Networks, Hidden Markov Models, and Support Vector Machines. *In Slico Biology* 2007; 7(6): 559-568.

[30] Pham TH et al. Support Vector Machines for Prediction and Analysis of Beta and Gamma-Turns in Proteins. *Journal of Bioinformatics and Computational Biology* 2005; 3(2): 343-358.

[31] Zhang Q et al. Improved Method for Predicting Beta-Turn Using Support Vector Machine. *Bioinformatics* 2005; 21(10): 2370-2374.

[32] Johal AK and Singh R. Protein Secondary Structure Prediction Using Improved Support Vector Machine and Neural Networks. *International Journal of Engineering and Computer Science* 2014; 3(1): 3593-3597.

[33] Bakhtiarizadeh MR et al. Neural Network and SVM Classifiers Accurately Predict Lipid Binding Proteins, Irrespective of Sequence Homology. *Journal of Theoretical Biology* 2014; 10:213-222.

[34] Garnier J et al. Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. *Journal of Molecular Biology* 1978; 120: 97-120.

[35] Thompson MJ and Goldstein RA. Predicting Protein Secondary Structure with Probabilistic Schemata of Evolutionary Derived Information. *Protein Science* 1997; 6(9): 1963-1975.

[36] Salamov AA and Solovyev VV. Prediction of Protein Secondary Structure by Combining Nearest-Neighbor Algorithms and Multiple Sequence Alignments. *Journal of Molecular Biology* 1995; 247: 11-15.

[37] Simas GM et al. Dimensional Reduction in Protein Secondary Structure Prediction- Nonlinear Method Improvements in *Innovations in Hybrid Intelligent Systems,* Springer-Verlag, 2007. p.425-432.

[38] Sommer C and Gerlich DW. Machine Learning in Cell Biology- Teaching Computers to Recognize Phenotypes. *Journal of Cell Science* 2013; 126: 5529-5539.

[39] Cheng J et al. SCRATCH: A Protein Structure and Structural Feature Prediction Server. *Nucleic Acid Research* 2005; 33:W72-76.

[40] Lin K et al. A Simple and Fast Secondary Structure Prediction Method Using Hidden Neural Network. *Bioinformatics* 2005; 21(2): 152-159.

[41] Bouziane H et al. Profiles and Majority Voting-based Ensemble Method for Protein Secondary Structure Prediction. *Evolutionary Bioinformatics* 2011; 7: 171-189.

[42] Hawkins J and Blakeslee S. *On Intelligence*. Times Books, New York, 2004.

[43] Personal communication with Jeff Hawkins, June 4, 2014.

[44] Arel I et al. Deep Machine Learning- A New Frontier in Artificial Intelligence Research. *IEEE Computational Intelligence* 2010; 13-18.

[45] DiLena P et al. Deep Architecture for Protein Contact Map Prediction. *Bioinformatics* 2012; 28(19): 2449-2457.

[46] Navlakha S and Bar-Joseph Z. Algorithms in Nature: The Convergence of Systems Biology and Computational Thinking. *Molecular Systems Biology* 2011; 7:546-556.

[47] Ikram AA. Neural Network-based Cloud Computing Platform for Bioinformatics. *Systems, Applications, and Technology Conference* 2013.

[48] Bajo J et al. Cloud Computing in Bioinformatics. *Computing and Artificial Intelligence* 2010; 79: 147-155.

[49] Cozzetto D and Tramontano A. Advances and Pitfalls in Protein Structure Prediction. *Current Protein and Peptide Science* 2008; 9(6): 567-577.

[50] Ezziane Z. Applications of Artificial Intelligence in Bioinformatics: A Review. *Expert Systems with Applications* 2006; 30:2-10.