



By using MADALINE Learning with Back Propagation and Keras to Predict the Protein Secondary Structure

Shivani Agarwal, Pankaj Agarwal

Abstract: Understanding of intermediate protein structure prediction serves as a crucial component to find the function of residues of amino acid. In this paper, focus on the intermediate protein structure by using feed forward and feedback method and enhancing the concept of sliding window. Prediction of secondary structure is a very cosmic problem of bioinformatics. This can be reduced by predicting or unfold the protein structures if it is unfolded so that can give the great results in medical sciences. Our main motive is to improve the accuracy of secondary structures and minimize the error. Experimentally, use the Multilayer ADALINE network for learning and KERAS TENSORFLOW use for train the weight matrix and sigmoid function for calculating the resultant with back propagation. Resultant of this paper results provides more prominent results as compare to already existing methods. Those improve the accuracy of secondary structure prediction.

Keywords- Residue, Multilayer ADALINE learning, KERAS TENSORFLOW, Back propagation, Improve accuracy, Sliding Window

I. INTRODUCTION

Computational Secondary structure prediction (CSSP) is a method that can be used to generate the intermediate structure of proteins, those helps to fulfill the diagnosis of medical sciences. Protein folding is the very immense problem of bioinformatics, this problem can be minimized by uncover the middle structure of protein rather than 3-D structure from its initial structure. Protein secondary structure provides to facilitate in better drug designing. Before the secondary or intermediate structure prediction of proteins, 3-D structure is resolute from the primary protein sequences but this structure is very massive. Due to this structure designing of any drug generates a very huge problem. By introducing the intermediate structure the protein folding problem can be resolved. The minimization of the problem of protein folding provides rapidly growth in the area of bioinformatics and medical drug designing. If identification of intermediate structure as compare to tertiary structure so this can be enhanced the learning of medical sciences. The precision of existing protein intermediate construction guess scheme is evaluate in tabloid benchmarks such as Live Bench and CLC Workbench.

So that in this proposed work define a prediction model that model uses trained weight matrix from tensor flow GPU's and learn the network by using Multilayer ADALINE learning. It completely focused on how to increase the accuracy of secondary structures.

II. RELATED WORK

Primary structure represents in the form of alphabets and intermediate protein structure in the form of alpha-Helix, beta sheets, and turn or coils. Many literature presents for solve this problem, Nitin Bhardwaj et al, aims to focus on binding DNA-residues from that predicts the secondary protein sequences after refinement in depending the distances. Lee et al, resolves by developing new computational biology and bioinformatics methods and new database servers. John Moullet et al, defines innovative schemes for envisage three-dimensional drop-line led to a two-fold enhancement in contact precision. Akosua Busia et al proposes a ensemble method and a chain of deep neural network with convolution model. In this predict 8 class structure rather than 3 class by using Q8 accuracy method. Gokmen-atlag, used a tensor flow approach and snake make workflow that train the large databases and calculate the high grade performance. Shivani et al used an algorithm that can be used computational models to reduce error of intermediate protein structure prediction. Haris Hasic et al, intend a innovative hybrid scheme that depends on the convention of numerous neural networks through the employ of a compromise task and measure up to our approach with other proficient manner. Sandhya Parasnath et al, presents a parallel programming approach with accelerating prediction by using GPU. Rashid et al, prompts a new loom for structure factors of proteins to train and an exploration that guide to improve it.

So it is conclude from this literature that need the eminent exactness tolerate the employ of the prophecy in protein folding problem, taxonomy of structural motifs, and enhancement of sequence alignments.

III. METHODOLOGY

Feed-forward networks are the distinct category of neural network that is very popular. The attractiveness of this particular method is that it dilutes the problem of many recognized fields of bio-informatics; this trouble is one of them. It originates from the actuality that they have been functionally flourishing to a extensive range of information progression assignment in miscellaneous fields.

Revised Manuscript Received on December 30, 2019.

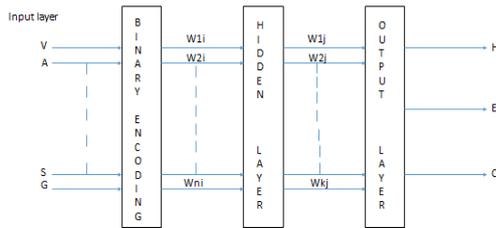
* Correspondence Author

Shivani Agarwal, Assistant professor, Computer Science & Engineering Department, IMS Engineering College, Ghaziabad, India (email:kasishivani@gmail.com)

Pankaj Agarwal, Head, Computer Science & Engineering Department, IMS Engineering College, Ghaziabad, India (email:pankaj7877@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Intermediate structure of protein prediction is one of them. Multi-layered feed forward network mold is taught with the help of middle layers and by varying sliding window size to determine optimal window size giving highest accuracy. In this figure, three layers of hidden neurons, input layer and output layer .The connectivity of this layered structure connected to each layer to another layer, in this structure hidden layers are varied according to 1200 for first layer,600 for second layer and 100 for third layer and other problems these are varied as per the problem.



ARCHITECTURE OF SECONDARY STRUCTURE PREDICTION OF PROTEIN

Fig 1:-layered structure of Primary, Hidden and Output

The primary structure of Amino Acid is the prime concern for the intermediate structure prediction. So that, sequences of primary structures in the form of alphabets like A, C, D...20 amino acids used in primary sequences are collected and rest 6 alphabets are silent. These alphabets related to secondary structures from PDB (protein data bank) that is in the form of H, E, and C. For example MVLSEGEWQLVLHVWAK VLSEGEWQLVLHVWAK VLSEGEWQLVLHVWAKVE, this is the primary sequence corresponding to secondary sequence each residue relates with secondary like HHEEEEECCCC.... CCHHHH... CCCCCEE ...HHHHHEEEEE. So the key intent of this research work is to predict intermediate structure of protein from primary structures.

A Encoding

In this model, envisage the intermediate structure of proteins from its initial sequences. The neural network takes the protein sequence as the inputs and it undergoes calculations and then predicts the output class. First we can collect data as input and output sequences from PDB (protein data bank), these output sequences use only for testing. We have 3 output classes (H, E, C) and the 20 input typeset (A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y). After that Number of inputs to the network is decided by the window size, the size of sliding window is 17. we do encode of primary sequences by using binary encoding, each residue defines as 20+1(extra) bit data so complete window size is 357 as inputs Each character of the sequence is being binary encoded. There are 21 characters possible in the sequence so each character is represented as 1 at their respective position and rest of all as 0. So each character representation length is 21.

Table 1:-Encoding of primary sequences

A	10000000000000000000
C	01000000000000000000
D	00100000000000000000
E	00010000000000000000

F	00001000000000000000
G	00000100000000000000
H	00000010000000000000
I	00000001000000000000
K	00000000100000000000
L	00000000010000000000
M	00000000001000000000
N	00000000000100000000
P	00000000000010000000
Q	00000000000001000000
R	00000000000000100000
S	0000000000000000100000
T	00000000000000000010000
V	000000000000000000001000
W	0000000000000000000000100
Y	00000000000000000000000010

B Model structure and Weight training

In this prophecy model we have used three hidden layers, first middle layer has 1200 neurons, second middle layer has 600 neurons and the third middle layer has 100 neurons. In this we use sigmoid activation function. Processing starts from input layer to hidden layer1 to hidden layer 2 to hidden layer 3 and then output. After the feed forward processing the error correction factor is being calculated and then back propagated for the calculation of the error correction factor at different hidden layers and change in weights. Then the final weights are being updated with the change in weights. In this, we use the Keras Machine Learning Library which basically uses the Tensor Flow for the processing. Tensor Flow is being provided by the Google for the Machine learning solutions. Different packages like Numpy, Pickel have been used to design the program. The Model is being designed as:

- 1200 at hidden layer1
- 600 at hidden layer2
- 100 at hidden layer3
- at the output layer

The training of the network is being lead using the Tensor Flow which used to do the processing using threads.

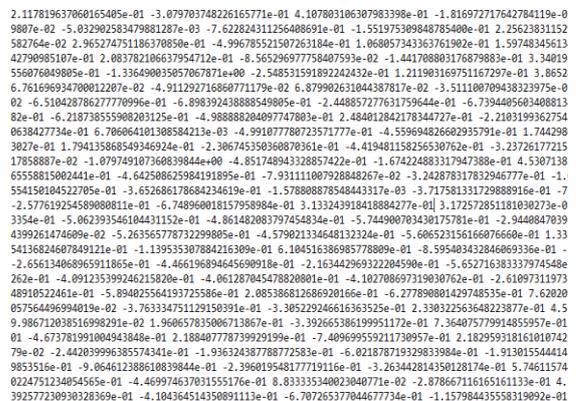


Fig 2:-Weight trained by using Keras from input to hidden (357-12000)



]-3.846540749073028564e-01 -1.471607387065887451e-01 -1.767558753490447990e-01 -4.502636492252349854e-01 -6
43018e-01 -1.41615364700558014e-02 1.143437027931213379e-01 7.091961055994033813e-02 8.996157348155975342e
e-01 -1.328594386577606201e-01 2.26905488967895078e-01 -3.28055024147833914e-01 1.077416837215423584e-01
565139771e-01 -1.262962669134140015e-01 -5.54412417137809753e-02 2.643782198429107666e-01 -6.45444765668698
6025066375732e-01 -1.624810864925384521e-01 -1.788490461246490479e-01 3.725984692573547363e-02 3.2033005356
51552526132774353e-02 -6.056356002582740704e-02 -1.652380079030990601e-01 -3.978739976802934570e-01 -3.66
5e-01 1.90182944419860840e-01 -2.777939848681818085e-02 4.56604081002473831e-03 -1.233481019735336304e-01
51e-01 -2.859788239002227783e-01 -7.002321630716323853e-02 -1.789635278833969116e-01 2.0612317940338054e-
0041594e-01 -3.11193887889385223e-02 2.079661376774311066e-02 -2.153613567352294922e-01 -1.509227305650711
42279853e-01 -6.860302388668060303e-02 2.29062587022187121e-01 1.858824640512466431e-01 -1.079471781849861
95905113220215e-01 5.180670693516731262e-02 -1.208408520860671997e-01 -1.416535377502441406e-01 -4.02255654
57307624359131e-01 -3.020272776484489441e-02 -2.547311484013690186e-01 4.074330674545280080e-01 5.59900940
12769323304360707e-02 1.659294217824935913e-02 -2.631448209285736084e-01 -7.994205504655830013e-02 -9.01830
2 -2.494541257615957788e-01 -2.070099115371704102e-01 9.808233380317687980e-02 -1.091980934143066406e-01 -1
0 -6.4765871512899398004e-03 -8.883760124444961540e-02 4.802381483912467957e-02 -1.516155451536178589e-01 4
0e-01 2.049382925833569336e-01 -3.89961113929748535e-01 7.6116167008078008054e-02 -1.154404357075691223e-01
-04 -9.095114469528198242e-02 -4.59968062550586652e-01 7.125113159418106079e-02 -1.307120472192764202e-01
-01 1.582590407679195956e-01 3.474141359329223633e-01 4.46953564882278442e-01 -9.69149991869926452e-02 -6
50238e-04 -1.092285794019699097e-01 -1.09168671071529384e-01 -3.766109794378200640e-02 -1.7882373929023742
841e-01 -5.64277888613939285e-03 -7.031619548797607422e-02 -3.18421453237535693e-01 1.59406941384070721e
6e-01 9.339723736047744751e-02 -2.505130171775818781e-01 4.014546424159466919e-02 -1.82416334782681274e-01
05e-01 9.2929270734786987e-01 -8.273929357528686523e-02 -8.655322715640068054e-03 4.065061318055560303e-0
e-01 -3.89944238780871582e-01 -9.741436467843532562e-03 -2.435516119083295890e-01 -1.159665063023567200e-0
893e-05 -5.530630550112915839e-01 1.225474402380446050e-01 3.137162700295448303e-02 -7.420564442873001099e-
7 7.435446977615356445e-02 -2.69197374522906494e-01 -1.935736685991287231e-01 -2.443008376224517822e-01 -1.
8707618.154740929603576660e-01 -2.58635640144381481445e-01 -1.015774067658306230e-01 -2.84076958549026275e-
26e-02 9.993721544742584229e-02 2.141423076391220093e-01 -3.196657821536064140e-02 3.310988843441009521e-01

Fig 3:- Wight trained by using Keras from hidden to hidden (1200-600)

]-1.115092709660530090e-01 -1.012464389204978943e-01 -6.435373425483703613e-02 1.474127232435379028e-01 -1.865
261908710002899170e-01 1.360289603471755981e-01 6.944581121206283569e-02 1.028298493474721909e-02 -9.0345777
7.17503529649734497e-02 2.058215570449829102e-02 -4.002393037080764771e-02 -1.343817561864852905e-01 -1.492
68776156082435608e-02 2.288932502269744873e-01 2.703226543962955475e-02 6.273482739925384521e-02 1.392496667
870872490e-02 -1.194480061531066095e-01 -1.50333946312615210e-01 1.690215319395065308e-01 2.86826170593855
-1.3550008033835571e-01 -1.204859279096126556e-02 1.466082781553268433e-01 -8.3681434392907715e-02 7.1
61619496455200e-01 6.428872793912887573e-02 -7.39917516095352173e-01 2.10975423455238342e-01 7.5079917907
3062782876e-02 7.19729736447342896e-02 -8.644952625036239624e-02 -1.42110779047012329e-01 1.1107756942510
5e-01 -2.54196450154113770e-01 4.651799425482749939e-02 1.425513327121734619e-01 2.83760040998458623e-01 -
41179704661377e-01 1.182092726230621338e-01 1.384226325899362564e-02 4.213167726993560791e-01 -2.335948558
16e-01 -2.052759975194931030e-01 -1.27165228128432275e-01 -3.654119744896888733e-02 -2.507688105106353760e-
427856e-01 1.13009689879417194e-01 7.839534431695938110e-02 5.33329620957345720e-02 -1.8275083974080379181e
69849e-02 -2.662726328125000000e-02 -2.124220430850982666e-02 -2.313606292009353638e-01 -1.91578537252723266
3210e-02 3.809597406704587936e-02 -4.838115349411964417e-02 5.222194641828536987e-02 -4.244453832507133484e-
960e-01 -1.071515679359436035e-02 4.682549235091934204e-02 -3.07607797684478760e-02 2.207894221375882626e-0
5e-01 8.65227720621075439e-02 -2.956833541393280092e-01 -2.0993408558052062980e-01 -6.162724271416664124e-0
25220e-01 -1.8190521001811579590e-01 -2.138800621032714044e-01 5.489327013492584229e-02 -1.477221250534057617
0766016e-01 3.585356101139369964e-02 2.007784508168697357e-02 1.230138074606657820e-01 -1.09877899289131164
02 2.002777954638004303e-02 2.786623127758502960e-02 1.440096199512481689e-01 -3.284054249525070190e-02 -3.1
06230801343917847e-01 1.5308766067028084565e-01 -1.743616163730621338e-01 -6.635121256113052368e-02 5.7795364
552703806e-01 6.059813071970713400e-01 3.915357962250709534e-02 2.044175416231155396e-01 1.691787093077702358
-02 -1.4180381858433496e-01 -8.121126882374286515e-03 -7.343175262212752396e-02 8.981700241565704340e-02 7.
4.0276635430280381165e-02 -4.168535768985748291e-02 -3.62037358002469254e-03 1.182854697108260738e-01 1.796
e-01 -1.224475502967834473e-01 -2.857937812805175781e-01 -1.097980793565511703e-02 -1.360928518636703491e-01

Fig 4: Wight trained by using Keras from hidden to output (1200-600)

IV. PROPOSED ALGORITHM

1. Input size: for a given sequence to be predicted, a window size of |w| is considered. Input layer is presented with input of size |w|* 21 where 21 corresponds to 20 amino acids + one extra unknown residue. (here window size is 17)
2. The network takes as input and output sequences; those are collected from protein data bank (PDB) prediction for each position in the sequence.

Input Sequence

A	C	D	I	K	T	L	E	Y
H	P	R	P	S	U	T	E	H

Output Sequence

H	H	H	H	E	E	C	T
C	C	H	H	H	H	E	E

3. Each input data residue encoded in to binary form for example:-

A	10000000000000000000
C	01000000000000000000

D	00100000000000000000
E	00010000000000000000

4. They expect type will be either 'H'- α -helix element, 'E'- β - beta strand or '-' representing turn/coil or non prediction. The process begins with taking a large input set of training patterns of defined size. Secondary structure for each of the patterns is predicted using some efficient existing prediction tool. Accordingly the auto generated pattern ID, the pattern itself & its associated structure is recorded in a database relation. Input Layer established the input in this form X_i ($i=1$ to $i=n$), n is the value of sliding window and y_{inj} ($j=1$ to $j=4$) assesses as:

$$y_{inj} = W_{oj} + \sum_{i=1}^n x_i \cdot W_{ij} \dots [1]$$

The production from hidden layer neurons is considered as

$$y_j = f(y_{inj}) \dots [2], \text{ Where } f(x)$$

$$\text{Sigmoid function} = (1 - e^{-x}) / (1 + e^{x})$$

5. Network training begins by presenting the analyzed patterns to the network one by one. An initial weight matrix initialed randomly is considered and weight training has done by Keras machine learning libraries.
6. Based on some processing done at the hidden layer, the production expected from hidden layer neurons The productive z_k is calculated using the equations.

$$Z_{ink} = W_{ok} + \sum_{j=1}^p Z_j \cdot V_{jk} \dots [3]$$

$$Z_k = f(Z_{ink}) \dots [4]$$

7. Based on some processing done at the hidden layer, the crop of the system is compared with the intermediate structure of the pattern and calculating the error value if there is any. The error is feedback to the hidden layer and accordingly the weight matrix is adjusted. This process is iterated until the error factor falls within the accepted range.

$$E = (t - z)^2 \dots [5]$$

$$\delta_k = (t_k - z_k) f'(z_{ink}) \dots [6]$$

$$f'(z_{ink}) = \left(\frac{\lambda}{2}\right) [(1 + f(z_{ink})) / (1 - f(z_{ink}))] \dots [7]$$

8. Accordingly the modified weight matrix expressed in the form of an array or string is stored by using Keras database. This process is repeated for all the stored patterns. This accomplishes the network training process.
9. Finally a naïve pattern is presented to the network and using some efficient search technique the most suitable pattern is found within the database prepared during the training step and its corresponding weight matrix is extracted. Suitability is defined in terms of degree of similarity as test process.



10. For assessment of observed and authentic protein intermediate structure for the experiment records, we estimate the figure of similarities for the predicted protein secondary structures and calculated the efficiency for three hidden states for each test data as $Q^* = (N \text{ predicted} / N \text{ observed}) * 100$

C. Finally, the accurate input result depends on the family of proteins to which it belongs. Many methods lie to calculate the performance of these for predict the intermediate structure of proteins. The most common method is a simple success rate, Q^* , which is the proportion of suitably predicted scrums on three types of secondary structures:, tiny proportion of initial structure acceptably calculate as alpha-helix, beta-strand, and turns

D. Datasets

Secondary structures are generated by folding the initial sequences in the outline of Helix, β -sheets, turns or coils. These structures resolved the important problem of bioinformatics. Some sequences are freely accessible in Worldwide Protein Data Bank. These are freely accessible on the Internet via the websites of its member organization like RCSB. In this paper collect the primary and secondary sequences from PDB bank. There are 153836 formations in Protein Data Bank.

Table 2: DATA SET OF PRIMARY AND SECONDARY SEQUENCE

V. RESULTS AND DISCUSSION

Diverse computational methods are present in literature for secondary structure prediction but this proposed method provides the optimize results. We are recognizing that this proposed algorithm increasing accuracy. Accuracy of NN is expected to be comparable to other popular secondary structure tools provided a significant number of patterns are trained and stored. To get more significant results from molecular biology point of view, it is felt that training patterns needs to be longer is size probably 15-20 characters long. Currently program is being tested on patterns of size 17 .in this size we received highest accuracy as compare to other window sizes.

Table 3: distinction of Accuracy with diverse frame size

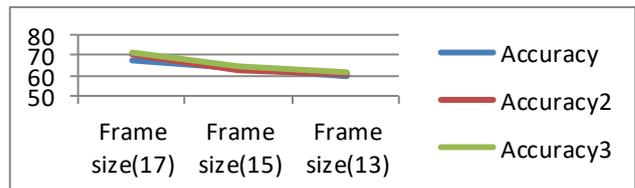
Frame Size	Accuracy
13	60-63%
15	63-65%
17	70-75%

In this paper, find the maximum accuracy by using 17th frame size rather than other size of frames, check the accuracy with different frames and with different sequence ids for single frame size:-

Table 4: Distinction sequence ids with window size obtained diverse accuracies.

Single convolution matrix	Sequence Id	Window size	Layer 1 window	Layer 2 window	Layer 3 window	Accuracy
17*21	1	17	357*1200	1200*600	600*100	68.0 %
17*21	2	17	357*1200	1200*600	600*100	70.0 %
17*21	3	17	357*1200	1200*600	600*100	71.0 %
17*21	4	17	357*1200	1200*600	600*100	69.0 %
15*21	1	15	351*1200	1200*600	600*100	65.0 %
15*21	2	15	351*1200	1200*600	600*100	64.0 %
13*21	1	13	273*1200	1200*600	600*100	62.0 %
13*21	2	13	273*1200	1200*600	600*100	61.0 %

Results shown the calculated accuracies at different frame sizes:-



VI. CONCLUSION

The precise discovery of protein intermediate formation is obligatory to detect syndrome and deficit exactly [1]. The implementation of the above proposed algorithm, the accuracy should come out to be approximately 70%. This gives the conclusion that the multi layered feed forward and feedback network can be used for the prediction of intermediate structure of protein with an accuracy of approximately 70% whereas the highest accuracy with this method till date is 70%.



In future we predict the protein 3-D construction and improve accuracy.

22. Principles of Soft Computing by S.N. Sivanandam and S.N. Deepa.

REFERENCES

- Gokmen atlag,"Tensor flow based deep learning model and snakemakeworkflow for peptide protein binding prediction, bioRxiv, doi: <http://dx.doi.org/10.1101/410928>, Sep. 6, 2018.
- Shivani agrawal, Vijander Singh,pankaj agarwal, Asha Rani, "Prediction of Secondary Structure of Proteins using Sliding windowand Back-propagation algorithm", SPRINGER SIGMA2018, Advances in Intelligent Systems and Computing book series (AISC, volume 698),2018.
- Leonardo de Lim, Corrêa, Bruno Borguesan, Mathias J.Krause, Márcio Dorn, Three-dimensional protein structure prediction based on memetic algorithms, Computers & Operations Research, Vol. 91, March 2018, pp. 160-177
- John Moul,t,"Critical assessment of methods of protein structure prediction (CASP)—Round XII", wiley,DOI: 10.1002/prot.25415, 27 October 2017.
- Haris Hasic ,,"A hybrid method for prediction of protein secondary structure based on multiple artificial neural networks", IEEE, 22-26 May, 2017.[DOI 1]
- Ali Abdulhafidh Ibrahim, "Using Neural Networks to Predict Secondary Structure for Protein Folding", scientific research, JAN2017, [DOI 3]
- JyotshnaDongardive,"Reaching optimized parameter set: protein secondary structure prediction using neural network", ACM,August 2017,[DOI 2].
- Sheng Wang,"Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields",Scientific Reports volume6, Article number: 18962 (2016)
- Rashid, et.al,"Protein secondary structure prediction using a small training set (compact model) combined with a Complex-valued neural network approach,BMC Bioinformatics(2016) [DOI 4].
- Sandhya Parasnath, Dubey N, Gopalakrishna Kini, M.Sathish Kumar, S.Balaji, Ab initio protein structure prediction using GPU computing,Perspectives in Science, Vol. 8, September 2016, pp. 645-647.
- Akosua Busia,"Protein Secondary Structure Prediction Using Deep Multi-scale Convolutional Neural Networks and Next-Step Conditioning", arXiv:1611.01503v1 [cs.LG] 4 Nov 2016.
- Alfonso E.Márquez-Chamorro, Gualberto Asencio-Cortés, Cosme E.Santiesteban-Toca, Jesús S.Aguilar-Ruiz, Soft computing methods for the prediction of protein tertiary structures: A survey, Applied Soft Computing, Vol. 35, October 2015, pp. 398-410.
- Bruno Borguesan, Mariel Barbachane Silva, Bruno Grisci, MarioInostroza-Ponta, Márcio, Dorn, APL: An angle probability list to improve knowledge-based metaheuristics for the three-dimensional protein structure prediction, Computational Biology and Chemistry, Vol. 59, Part A, December 2015, pp.142-157.
- Shivani Agarwal,et.al," Prediction of Secondary Structure of Protein Using Support Vector Machine" IJCA,2014. [DOI 7]
- Márcio Dorn, Mariel Barbachane, Silva Luciana, S.Buriola, Luis C.Lamb, Review Three-dimensional protein structure prediction: Methods and computational strategies, Computational Biology and Chemistry, Vol. 53, Part B, December 2014, pp. 251-276.
- Liang Kong, Lichao Zhang, Jinfeng Lv, Accurate prediction of protein structural classes by incorporating predicted secondary structure information into the general form of Chou's pseudo amino acid composition, Journal of Theoretical Biology, Vol. 344, 7 March 2014, pp. 12-18.
- Shuyan Ding, Shoujiang Yan, Shuhua Qi, Yan Li, Yuhua Yao, A protein structural classes prediction method based on PSI-BLAST profile,Journal of Theoretical Biology, Vol. 353, 21 July 2014, pp. 19-23.
- Asha Rani, Vijander Singh, J.R.P. Gupta, Development of soft sensor for neural network based control of distillation column, ISA Transactions, 52, 3, 438-449, 2013.
- Nitin Bhardwaj,"Residue-level prediction of DNA-binding sites and its application on DNA-binding protein predictions", Bioinformatics Program, Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA., February 2007.
- Lee Whitmore," Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases,"Wiley, <https://doi.org/10.1002/bip.20853>, September 2007.
- Bissan Al-Lazikani, Joonjung, Zhixin Xiang, Barry Honig, Review: Protein Structure Prediction, Current Opinion In Chemical Biology, Vol. 5, Issue 1, 1 February 2001, pp. 51-56.

AUTHORS PROFILE



Shivani Agarwal, currently functioning in IMS Engineering College,Ghaziabad as Assistant professor in Computer Science and engineering department of IMSEC college,Ghaziabad. She has a teaching experience of more than 11 years in engineering colleges.she is Pursuing PhD from Uttrakhad Technical University,Dehradun. She is having numerous research papers in presumed International Journals and Conferences. There are above 8 publication in international journals and international conferences to her glory. Her Specialization regions & research field includes Soft computing, Computer organization, Machine Learning and Digital Logic Design. Basically her research focused on bioinformatics problems like Prediction of intermediate structure prediction by using various machine learning techniques.



Dr. Pankaj Agarwal, currently operational in IMS Engineering college deputed as HOD & Professor in CSE department Ghaziabad..He accomplished his Ph.D in Computer Science from Jamia Millia Islamia Central University, Delhi in 2009. He has a teaching experience of more than 15 years in engineering colleges and research experience of more than 8 years. There are above 35 publications in international journals and conferences to his glory. His proficiency fields & research includes Algorithms in Computational Biology, Soft computing Techniques, Parallel Algorithms, Machine Learning using R & Python. Basically his research focused on problem of bioinformatics solved by using various machine learning techniques.