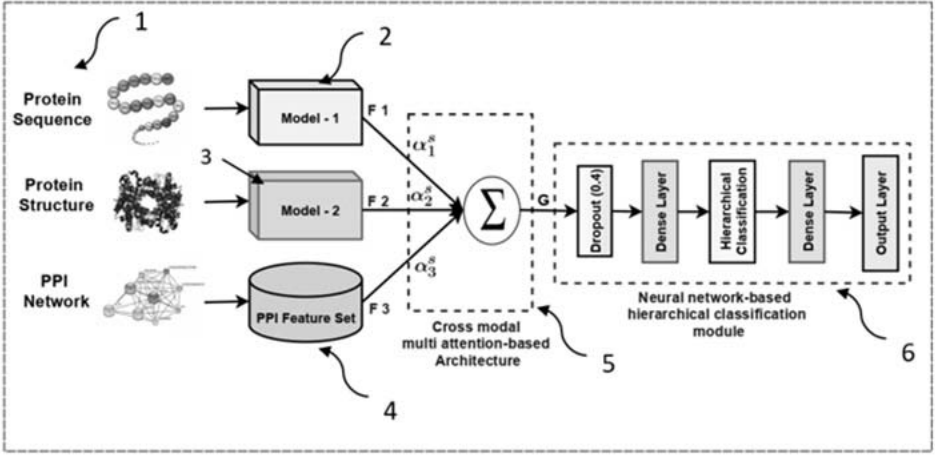
1
2
3
100
Protein Sequence
Protein Structure
PPI Network
Model - 1
Model - 2
PPI Feature Set
F 1
F 2
F 3
α_1^s
α_2^s
α_3^s
Σ
G
Dropout (0.4)
Dense Layer
Hierarchical Classification
Dense Layer
Output Layer
Cross modal multi attention-based Architecture
Neural network-based hierarchical classification module
4
5
6

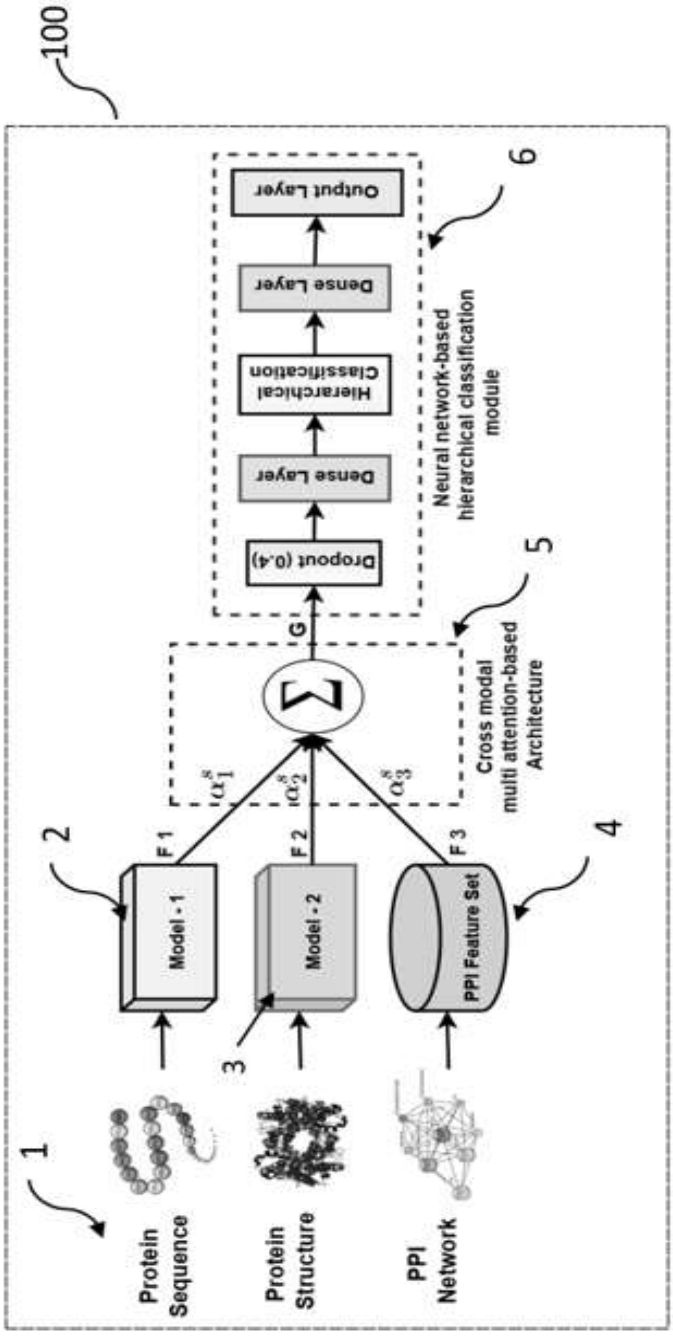

Figure 1

SHRUTI KAUSHIK
Agent for the Applicant
IN/PA-1324

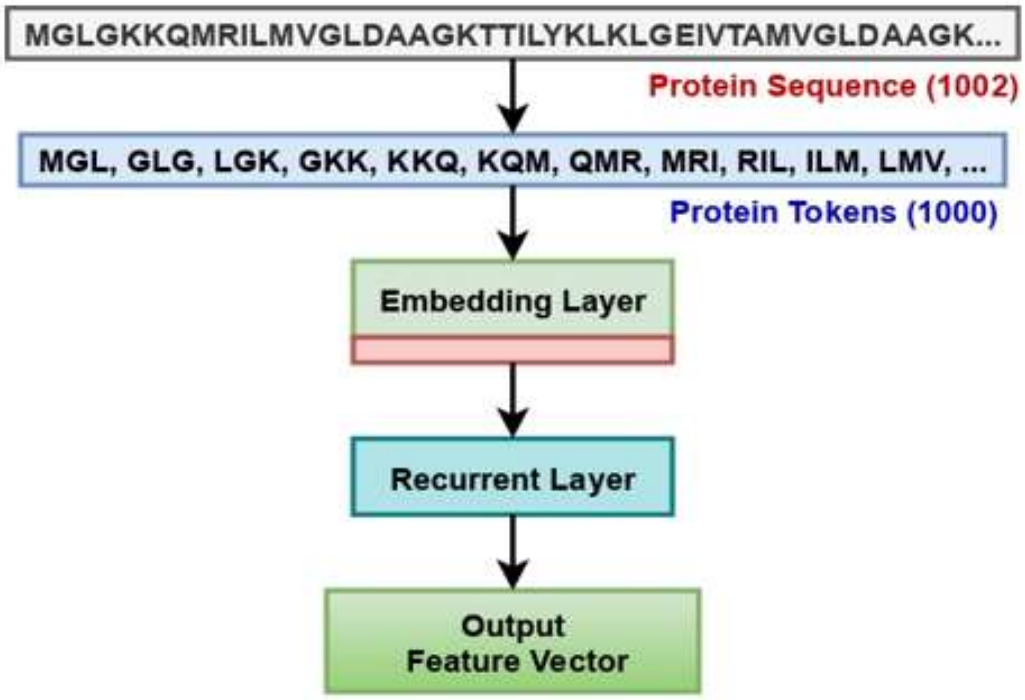

Figure 2

SHRUTI KAUSHIK
Agent for the Applicant
IN/PA-1324

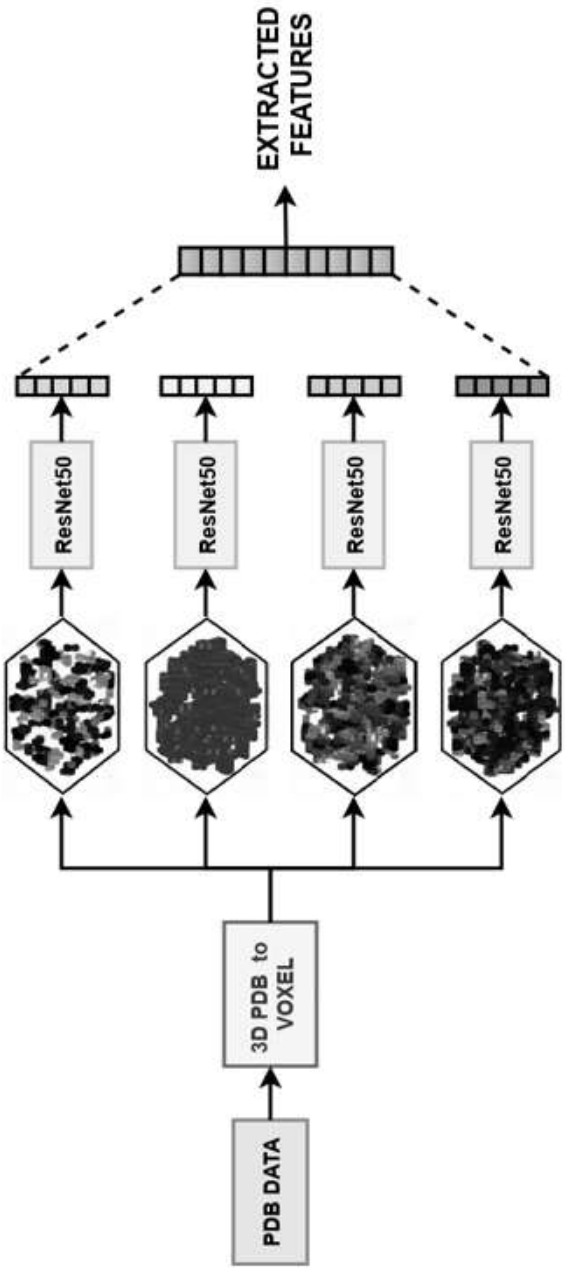

Figure 3

SHRUTI KAUSHIK
Agent for the Applicant
IN/PA-1324

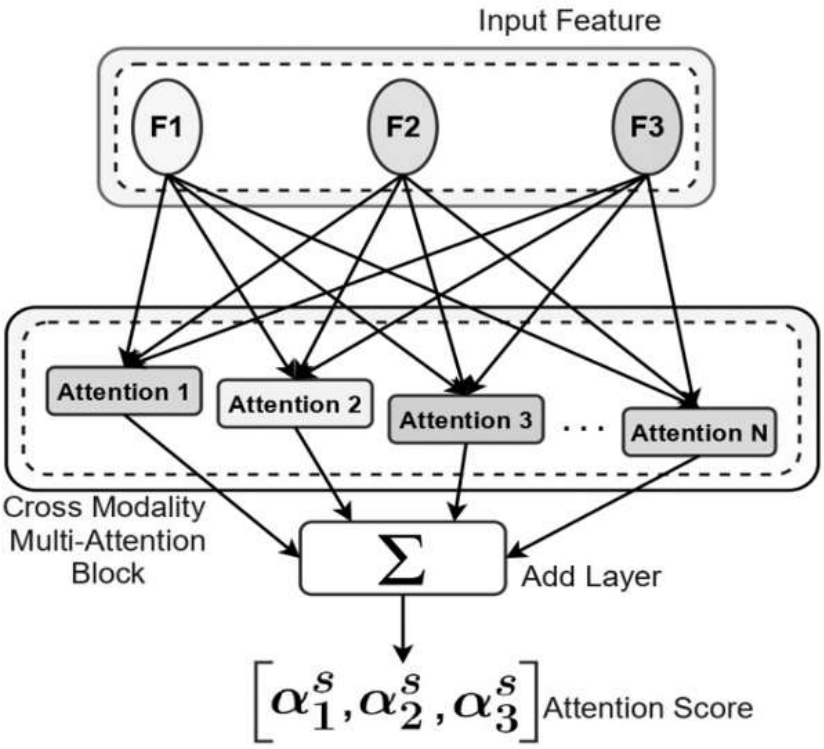

Figure 4

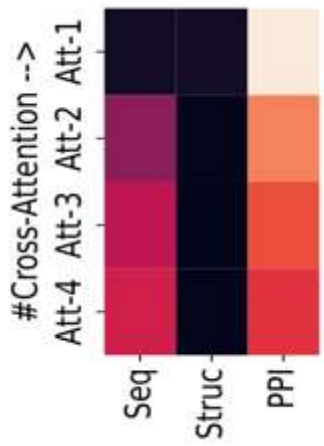

Figure 5 (a)

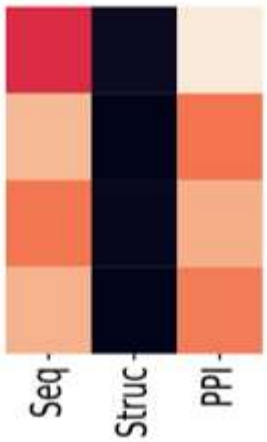

Figure 5 (b)

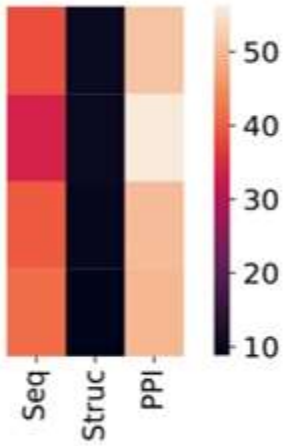

Figure 5 (c)



SHRUTI KAUSHIK
Agent for the Applicant
IN/PA-1324

FORM 2

THE PATENTS ACT, 1970

(39 of 1970)

&

THE PATENT RULES, 2003

COMPLETE SPECIFICATION

[See Section 10 and Rule 13]

TITLE:

"A CROSS MODAL MULTI ATTENTION MECHANISM BASED SYSTEM AND METHOD FOR PROTEIN FUNCTION PREDICTION"

APPLICANT:

NATIONAL INSTITUTE OF TECHNOLOGY, PATNA
An Indian Autonomous body
having address at
Patna-800005, Bihar, India

PREAMBLE TO THE DESCRIPTION:

The following specification particularly describes the invention and the manner in which it is to be performed:

FIELD OF THE INVENTION

The present invention relates to the protein function detection system and method. More particularly, the present invention relates to a cross modal multi attention mechanism based system and method for protein function prediction that works by capturing a relative importance of feature-sets that belong to various modalities with lower number of trainable parameters.

BACKGROUND OF THE INVENTION

Proteins are complex biomolecules that perform a wide range of functions in the body. Many proteins act as enzymes, which catalyse biochemical reactions in the body. Enzymes help to speed up chemical reactions that would otherwise occur too slowly to support life. Some proteins provide structural support to cells and tissues. For example, collagen is a protein that gives strength and elasticity to connective tissues such as skin, tendons, and cartilage. Some proteins act as carriers and transport molecules across membranes. For example, haemoglobin is a protein that carries oxygen in the blood. Many proteins act as signalling molecules that transmit signals between cells. Hormones, such as insulin, are examples of proteins that regulate physiological processes by signalling to target cells. Some proteins act as antibodies that recognize and bind to foreign substances, such as bacteria and viruses, to help defend the body against infection. Overall, proteins are essential for many biological processes and play a critical role in maintaining the proper functioning of the body.

Protein function prediction is important because it provides valuable insights into the biological processes that underlie various diseases and help to identify potential drug targets. Knowing the function of a protein helps researchers to identify the potential drug targets. If a protein is involved in a disease process, then targeting that protein with a drug could be an effective way to treat the disease. Protein function prediction is used to identify disease-specific

biomarkers that are used for early diagnosis or disease monitoring. Proteins are involved in many biological processes, and knowing their functions help researchers to understand these processes. This knowledge helps to identify new pathways or mechanisms that are targeted for therapeutic interventions.

Protein function prediction is used to compare the functions of proteins across different organisms. This provides insights into the evolution of biological processes and help to identify conserved pathways or functions. Overall, protein function prediction is an important tool for understanding the biological processes that underlie disease and for developing new therapeutics to treat these diseases.

WO2017011779A1 discloses an accelerated conformational sampling method for predicting target peptide and protein structures comprising a process of determining energy minimized synthetic templates using a simple system for modelling individual molecular bonds within the subject peptide or protein. Use of these synthetic templates greatly reduces the computational resources necessary for optimally determining structural features of the target peptide or protein. The cited disclosure also provides methods for rapid and efficient analysis of the effect of mutations on target peptides and proteins. However, this invention fails to preserve the modality relative contribution towards the protein function prediction, due to which this invention does not offer an efficient solution.

WO2012093869A2 discloses a method for predicting the location of a target protein in a cell under predetermined conditions by inputting static characteristics of an individual protein or gene, protein interaction information with a neighbouring protein, an expression profile of a protein or a gene, and the like. By employing the method of the present invention, the location of a target protein in a cell can be effectively predicted under specific conditions by inputting an already known target protein and predetermined conditions, and reversely, conditions such as disease stages can be ascertained through the

location of the target protein within a cell. In addition, the citation provides a method for predicting a biological process or molecular function of a protein for each condition by inputting the biological process or molecular function of the target protein using the method of the present invention. However, this invention fails to preserve the modality relative contribution towards the protein function prediction.

All the conventional multi-modal schemes in the state of art, either use the protein representation directly or their orthogonal features, and they simply concatenate these features instead of capturing any relative information amongst the modality, they give equal weightage to each representation, which is suitable to use for protein function prediction task.

Due to aforementioned drawbacks, there is a need to introduce a cross modal multi attention mechanism based system and method for protein function prediction that shows better performance while keeping the number of trainable parameters small enough to be deployed on a modern desktop computer.

OBJECT OF THE INVENTION

The main object of the invention is to provide a system and method for protein function prediction based cross modal multi attention mechanism.

Another object of the invention is to provide a system and method that utilizes robust and highly efficient deep learning based cross modal multi attention architecture, which captures the relative importance of feature-sets belonging to different modalities.

Yet another object of the invention is to provide a system and method that is capable of extending the application of attention mechanism from inter-modality to multi-modality paradigm.

Still another object of the present invention is to provide a system and method that offers an attention mechanism to capture the relative importance/contribution of each modality applied to a multi-modal architecture that improves upon state-of-the-art results while keeping the number of trainable parameters small enough to be deployed on a modern desktop computer.

SUMMARY OF THE INVENTION

The present invention relates to a system and method for protein function prediction based cross modal multi attention mechanism that captures a relative importance of each feature, and utilizes recurrent neural networks in multi-modal setup to achieve improved accuracy with much fewer trainable parameters than other contemporary state-of-the-art methods.

In an embodiment, the present invention provides a cross modal multi attention mechanism based system for protein function prediction, comprising of: an input module, a plurality of data processing module, wherein said input module receives a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator; said plurality of data processing module include a feature vector generator module, a feature extraction module, a protein-protein interaction feature set module, cross modal multi attention module and a neural network based hierarchical classification module; said feature vector generator module extracts a set of protein tokens from said protein sequence data and analyzes said set of protein tokens to extract a feature vector via an embedding operation and a recurrent operation; said feature extraction module extracts a set of features from a volumetric representation of said protein structure data; said protein-protein interaction feature-set module annotates a protein to an optimal function for producing a protein-protein interaction feature-set; said cross modal multi attention module processes said feature vector, set of

feature-set to capture a relative attention distribution value; and said neural network based hierarchical classification module for classifying a function of said protein sample by processing said relative attention distribution value.

In another embodiment, the present invention provides a cross modal multi attention mechanism based method for protein function prediction comprising steps of: a) receiving a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator via an input module, b) extracting a set of protein tokens from said protein sequence data and analyzing said set of protein tokens to extract a feature vector via an embedding operation, recurrent operation and a feature vector generator module, c) extracting a set of features from a volumetric representation of said protein structure data via a feature extraction module, d) adding a protein-protein interaction feature set with two others said feature set of protein sequence and volumetric representation of protein structure, e) processing said feature vector, set of features and said protein-protein interaction feature set to capture a relative attention distribution value via a cross modal multi attention module, classifying a function of said protein sample by processing said relative attention distribution value via a neural network based hierarchical classification module.

The above objects and advantages of the present invention will become apparent from the hereinafter set forth brief description of the drawings, detailed description of the invention, and claims appended herewith.

BRIEF DESCRIPTION OF THE DRAWING

An understanding of the cross modal multi attention mechanism based system and method for protein function prediction of the present invention may be obtained by reference to the following drawings:

Figure 1 is a block diagram of cross modal multi attention mechanism based system and method for protein function prediction according to an embodiment of the present invention.

Figure 2 is a flow chart of working of feature vector generator module in the cross modal multi attention mechanism based system according to an embodiment of the present invention.

Figure 3 is a flow chart of working of feature extraction module in the cross modal multi attention mechanism based system according to an embodiment of the present invention.

Figure 4 is a flow chart of working of cross modal multi attention module in the cross modal multi attention mechanism based system according to an embodiment of the present invention.

Figure 5(a), **5(b)** and **5(c)** are graphical representations of comparisons of each modality contribution on varying number of attention layer for all the sub-ontology datasets according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will now be described hereinafter with reference to the accompanying drawings in which a preferred embodiment of the invention is shown. This invention may, however, be embodied in many different forms and should not be construed as being limited to the embodiment set forth herein. Rather, the embodiment is provided so that this disclosure will be thorough, and will fully convey the scope of the invention to those skilled in the art.

Many aspects of the invention can be better understood with references made to the drawings below. The components in the drawings are not necessarily drawn to scale. Instead, emphasis is placed upon clearly illustrating the

components of the present invention. Moreover, like reference numerals designate corresponding parts through the several views in the drawings. Before explaining at least one embodiment of the invention, it is to be understood that the embodiments of the invention are not limited in their application to the details of construction and to the arrangement of the components set forth in the following description or illustrated in the drawings. The embodiments of the invention are capable of being practiced and carried out in various ways. In addition, the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting.

The present invention relates to a system and method for protein function prediction based cross modal multi attention mechanism that captures a relative importance of each feature, and utilizes recurrent neural networks in multi-modal setup to achieve improved accuracy with much fewer trainable parameters than other contemporary state-of-the-art methods so as the present invention is comfortably deployed on a modern desktop hardware.

In an embodiment, the present invention provides a cross modal multi attention mechanism based system for protein function prediction, comprising: an input module, a plurality of data processing module, wherein said input module receives a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator; said plurality of data processing module include a feature vector generator module, a feature extraction module, a protein-protein interaction feature set generator module, a cross modal multi attention module and a neural network based hierarchical classification module; said feature vector generator module extracts a set of protein tokens from said protein sequence data and analyzes said set of protein tokens to extract a feature vector via an embedding operation and a recurrent operation; said feature extraction module extracts a set of features from a volumetric representation of said protein structure data; said protein-protein interaction feature-set generator

module annotates a protein to an optimal function for producing a protein-protein interaction feature set; said cross modal multi attention module processes said feature vector, set of features and said protein-protein interaction feature-set to capture a relative attention distribution value; and said neural network based hierarchical classification module for classifying a function of said protein sample by processing said relative attention distribution value.

In another embodiment, the present invention provides a cross modal multi attention mechanism based method for protein function prediction comprising steps of: a) receiving a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator via an input module, b) extracting a set of protein tokens from said protein sequence data and analyzing said set of protein tokens to extract a feature vector via an embedding operation, recurrent operation and a feature vector generator module, c) extracting a set of features from a volumetric representation of said protein structure data via a feature extraction module, d) adding a protein-protein interaction feature set with two others said feature set of protein sequence and volumetric representation of protein structure, e) processing said feature vector, set of features and said protein-protein interaction feature-set to capture a relative attention distribution value via a cross modal multi attention module, classifying a function of said protein sample by processing said relative attention distribution value via a neural network based hierarchical classification module.

Figure 1 is a block diagram of cross modal multi attention mechanism based system **(100)** and method for protein function prediction according to an embodiment of the present invention. The cross modal multi attention mechanism based system **(100)** for protein function prediction, comprising of: an input module **(1)**; a plurality of data processing module, wherein said input module **(1)** receives a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator;

said plurality of data processing module include a feature vector generator module **(2)**, a feature extraction module **(3)**, a protein-protein interaction feature-set module **(4)**, cross modal multi attention module **(5)** and a neural network based hierarchical classification module **(6)**; said feature vector generator module **(2)** extracts a set of protein tokens from said protein sequence data and analyzes said set of protein tokens to extract a feature vector via an embedding operation and a recurrent operation; said feature extraction module **(3)** extracts a set of features from a volumetric representation of said protein structure data; said protein-protein interaction feature-set module **(4)** annotates a protein to an optimal function by adding a protein-protein interaction feature-set; said cross modal multi attention module **(5)** processes said feature vector, set of features and said protein-protein interaction feature-set to capture a relative attention distribution value; and said neural network based hierarchical classification module **(6)** for classifying a function of said protein sample by processing said relative attention distribution value.

As shown in **Figure 1**, the present invention involves three main entities, namely Model - 1, Model - 2, and cross-modal multi attention module. RNNs's Bi-GRU has been used in Model – 1 for learning the features from the protein sequence. Resnet-50 is applied on protein structure in Model – 2 for learning the features. These two sets of features extracted through deep learning models, along with the protein interaction features, are fed to the cross-modal multi-attention module. This module is a key contribution of this invention, which is able to capture the relative contribution of the three modalities using multi-attention module for the PFP task. Further, the learned features are passed to the downstream network neural network-based hierarchical classification module.

Figure 2 is a flow chart of working of feature vector generator module **(2)** in the cross modal multi attention mechanism based system **(100)** according to

an embodiment of the present invention. In order to efficiently build feature representation of protein sequences with a fixed length, the present invention deployed a bi-directional gated recurrent unit (Bi-GRU) architecture that learns long and short-range relationships between the protein words in both forward and backward directions. It takes input as a protein sequence and outputs (F-1) the corresponding feature vector, which is eventually fed to the cross modal multi attention module **(5)**.

Here, FASTA-formatted sequences were retrieved from the Uniprot database. Sequences vary in length from 2 to 35,213. Due to this high variation in length and the computational constraints, only sequences shorter than 1003 were used (as was also done in DeepGO).

Information flows through as follows:

Protein Tokens: Here, input protein sequences are segmented into substrings of size k, called k-mers (protein words), consider k = 3 to transform a sequence to words. Since, the maximum length of a protein sequence is restricted to 1003, a maximum of 1000 tokens are generated. Total 8000 distinct protein words form a dictionary as a result of various combinations of 20 amino acids.

Embedding Layer: The embedding layers use a neural network to transform protein words into meaningful embedding of size 128, resulting in a lookup table of dimensions 1000*128 corresponding to the maximum number of 1000 tokens. To prevent overfitting, a dropout layer represented in a light red box in **Figure 2** is added with a 0.2 dropout probability.

Recurrent Layer: This layer takes input from the dropout layer and learns the short-and-long interactions between protein words in both forward-and-backward directions. To achieve this, popular Bi-LSTM is applied, due to fewer gates (update and reset gate), which reduces network complexity and also beneficial for smaller corpus. These gates are formulated as (below Equations (1) and (2)):

$$\text{Update gate: } Z_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \textbf{ -(1)}$$

$$\text{Reset gate: } R_t = \sigma(W_j X_t + U_j h_{t-1} + b_j) \textbf{ -(2)}$$

where, W_i, W_j, U_i, and U_j denote weight matrices, with subscript terms i and j represent update gate and reset gate respectively, b_i and b_j are the bias vectors, and X_t denotes the current information.

Output Feature Vector: The output of the recurrent layer, is further used as the final feature vector represented as F-1, which are further used for classification or as input to the downstream network (cross modal multi attention module **(5)**).

Figure 3 is a flow chart of working of feature extraction module **(3)** in the cross modal multi attention mechanism based system **(100)** according to an embodiment of the present invention. Here, CNNs-based ResNet50 model has been used for learning the features from the volumetric representation of the protein 3D-structure. This model comprises of four parallel ResNet50 architectures for each of the four volumetric representations. The outputs from the four ResNet50 models are concatenated and eventually fed to the cross modal multi attention module, whose output is represented as F-2 as shown in **Figure 3**. The 3D structure of a protein depicts how its constituent amino acids come together to form the protein structure. It is an important source of information that are used for protein function prediction and to learn features from a protein structure, it is transformed into a volumetric representation, on which the ResNet50 architecture is employed to learn features from the volumetric representation as shown in **Figure 3**.

Each raw protein 3D structure downloaded from protein data bank is mapped into a grid structure in order to extract features. In this regard, a binary volumetric shape with volume elements called voxel is used. A voxel is seen as a 3D structure cube (V) with fixed grid length (L). These two parameters act as tradeoff to accommodate desired number of proteins irrespective to the sizes of

their structures. Four 3D volumetric representation of proteins, namely, binary (RB), hydropathy (RH), isoelectric (RI), and charge (RC), are extracted. These four feature representations are fed to the ResNet50 architecture.

ResNet50 architecture is applied on four 3D-structure representation individually. For each 3D volumetric representation, a 256-dimensional output is produced by the ResNet50 model. The outputs from each representation are weighted and concatenated using a two-layer neural network with 1024 (256*4) and 256 neurons in input and output layer respectively. Finally, 256 dimensional representations are fed to the downstream network (cross modal multi attention module **(5)**).

Protein-protein interaction (PPI) network information is also an important source to annotate the protein to its correct function(s). PPI data for multiple species are be accessed form the StringDB database. Confidence score and orthological relation of the proteins were extracted from the EggNOG database, mainly used for discarding similar proteins from the experimental evidence in this work. Further, proteins having less than 300 connections in the PPI graph were also discarded to retain only the significant proteins. The final dataset used for generation of graph embedding has 11.5 Billion edges and 8 Million proteins. Graph embedding vectors of length 256 were generated for each protein sequence using the Node2Vec approach. These embedding for each protein are passed to the downstream network (cross modal multi attention module) as shown in **Figure 1**.

Figure 4 is a flow chart of working of cross modal multi attention module **(5)** in the cross modal multi attention mechanism based system **(100)** according to an embodiment of the present invention.

The cross modal multi attention module **(5)** captures the relative importance amongst the different modalities while deploying much fewer trainable parameters than the contemporary state-of-the-art methods, which capture

relative importance amongst the three modalities towards predicting the function(s) performed by the proteins. It does so by deploying attention module. The cross modal multi attention module takes features as inputs from different modalities and applies cross modal multi attention on these modalities, resulting in relative attention output and score for each modality. These attention outputs are passed to the downstream network.

Input Feature: Feature vectors of each modality: F-1 represents the output feature vectors of the sequence data obtained from the RNNs-based architecture, F-2 represents the output feature vector of the 3D structure obtained from the CNNs-based architecture, and F-3 represents the feature vector for the PPI dataset. Each feature vector is of size 256. Each modality's features are fed to the downstream network (cross modal multi attention module).

Cross modal multi attention module: This module consists of multiple attention module-based architecture to capture relative attention distribution (scores) amongst the modalities. The steps of attention module are described next.

The hidden state h_f, corresponding to the feature-set of the fth modality, is passed through single layer perceptron to obtain the hidden representation O_f as depicted in Equation (3):

$$O_f = \tanh(W_a h_f + b_a) - \textbf{(3)}$$

where, $W_a \; and \; b_a$ represents the trainable weight matrix and bias, respectively.

The relative attention score α_f (refer to Equation (4)) for the feature-set of a modality is calculated using the softmax function:

$$\alpha_f = \frac{\exp(o_f)}{\sum_{f-1}^{F} \exp(o_f)} \text{ - (4)}$$

where, F is the number of features (modality).

Add Layer: The relative attention scores represented for each feature (modality) from the different cross modality attention-based module are combined together to obtain the final attention score α_f^g for each feature (modality) (refer to Equation (5)):

$$\alpha_f^g = \sum_{n=1}^{N} \alpha_f[n] - \textbf{(5)}$$

where, $\alpha_f[n]$ is the attention value for feature f and N is the number of cross-modal attention module.

The final vector G is computed (refer to Equation (6)) by aggregating the attention score α_f^g with the corresponding hidden states h_f, where $f \in \{1,2,3,\dots,F\}$.

$$G = \sum_{f=1}^{F} \alpha_f^s h_f - \textbf{(6)}$$

The final vector G is passed to the neural network-based hierarchical classification module as shown in **Figure 1**.

The present invention extends the conventional attention mechanism from intra-feature to inter-feature level. Instead of capturing the relative importance amongst the words in a sentence or elements in a vector, it captures the relative importance amongst the different modalities or feature-sets while deploying much fewer trainable parameters than the contemporary state-of-the-art methods. This provides relative attention to each modality to learn the importance by itself. The present invention is the first to capture the short-and-long range interaction amongst the protein words in a protein sequence in a multi-modal setup. The bidirectional nature of the recurrent layer helps to

learn the interactions from both forward and backward directions. This helps the model to have local as well as global alignment information of the protein words.

The outcomes on protein datasets were compared with the state-of-the-art contemporary methods. The present invention shows increment in the results than the state-of-the-art contemporary methods. Improvement of +3.31%, +6.9%, and +4.81% in F_{max} score for MF, CC, and BP sub-ontologies, respectively over the state-of-the-art MultiPredGO method [R1: S. J. Giri, P. Dutta, P. Halani and S. Saha, "MultiPredGO: deep multi-modal protein function prediction by amalgamating protein structure, sequence, and interaction information," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, p. 1832–1838, 2020].

EXAMPLE 1

Experimental Analysis

Figure 5(a), **5(b)** and **5(c)** are graphical representations of comparisons of each modality contribution on varying number of attention layer for all the sub-ontology datasets according to an embodiment of the present invention.

Relative importance comparison on different number of attention layer: The modality contribution of the present invention is compared with the other existing schemes. Here, to perform the comparison, cross modal multi attention mechanism with different attention layers is applied. The variation of relative contribution of each modality on varying the number of attention layer for all the sub-ontologies datasets are shown in **Figure 5(a)**, **5(b)** and **5(c)**. Here, brighter is the color, more is the contribution of that modality towards protein function prediction task.

The comparison on the relative contribution amongst the different modality scheme with other state-of-the-art scheme are shown in **Table 1**. Here, proteins interaction modality contributes more than the other two modalities,

further, 3D structure have the lowest contributions than others modalities. Whereas, state-of-the-art scheme does not have any relative contribution of the modality, and is shown with value 1.0, for equal contribution. [R2: D. Piovesan, M. Giollo, E. Leonardi, C. Ferrari and S. C. E. Tosatto, "INGA: protein function prediction combining interaction networks, domain assignments and sequence similarity," Nucleic acids research, vol. 43, p. W134–W140, 2015. R3: M. Kulmanov, M. A. Khan and R. Hoehndorf, "DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier," Bioinformatics, vol. 34, p. 660–668, 2018. R4: F. Zhang, H. Song, M. Zeng, Y. Li, L. Kurgan and M. Li, "DeepFunc: a deep learning framework for accurate prediction of protein functions from protein sequences and interactions," *Proteomics,* vol. 19, p. 1900019, 2019. R5: Y. Cai, J. Wang and L. Deng, "SDN2GO: an integrated deep learning model for protein function prediction," *Frontiers in bioengineering and biotechnology,* vol. 8, p. 391, 2020.]

Table 1: Overall relative contribution of each modality (NA- Not Applicable)

Scheme	Protein Sequence	3D Structure	Protein-protein Interaction	Domain Information
IGNA [R2]	1.0	NA	1.0	1.0
DeepGO [R3]	1.0	NA	1.0	NA
DeepFunc [R4]	1.0	NA	1.0	NA
SDN2GO [R5]	1.0	NA	1.0	1.0
MultiPredGO [R1]	1.0	1.0	1.0	NA
Present Invention	0.3833	0.089	0.5266	NA

Better feature learning in multi-modal setup using Recurrent Neural Networks: Here, to perform the comparison, RNNs (Bi-GRU) are used on place of CNN for learning the sequential information. The comparison between both the CNNs and RNNs based end-to-end scheme is shown in **Table 2**. The present invention shows better accuracy than the state-of-the-art MultiPredGO [9] scheme as shown in **Table 2**.

Table 2: Comparison of contribution of RNNs and CNN in the state-of-the-art scheme

Sub-Ontology Dataset	Scheme	F_{max}	AvgPre	AvgRec	AUC	MCC
BP	MultiPredGO [R1]	0.3278	0.3257	0.3439	0.8169	0.2865
	RNNs-MultiPredGO	0.3578	0.3552	0.3605	0.8708	0.3684
CC	MultiPredGO [R1]	0.5366	0.5684	0.5184	0.9102	0.5192
	RNNs-MultiPredGO	0.5880	0.5888	0.5873	0.9521	0.5488
MF	MultiPredGO [R1]	0.3671	0.3630	0.3685	0.8505	0.3123
	RNNs-MultiPredGO	0.3771	0.4201	0.3207	0.9102	0.3739

Overall better Accuracy: Here, to perform comparison, the present invention was evaluated on different sub-ontology datasets namely biological process (BP), cellular components (CC), and molecular function (MF). The overall accuracy was compared amongst the state-of-the-art schemes with the proposed invention on different sub-ontology datasets. **Table 3**, **4**, and **5** show

the comparison of the different multi-modal scheme with the present invention for BP, CC and MF sub-ontology datasets. **Table 6** shows the trainable parameter comparison of the existing state-of-the-art multi-modal schemes with the present invention. These table show that the proposed scheme have performed best amongst the state-of-the-art schemes and have the highest overall accuracy while deploying much fewer trainable parameters than the other methods.

Table 3: Comparison amongst different state-of-the-art scheme and the present invention on BP sub-ontology dataset

S. No.	Scheme	F_{max}	AvgPre	AvgRec	AUC	MCC
1	IGNA [R2]	0.1971	0.1576	0.2632	0.6034	0.1270
2	DeepGO [R3]	0.3332	0.3136	0.3528	0.8161	0.2805
3	MultiPredGO [R1]	0.3278	0.3257	0.3439	0.8169	0.2829
4	Present Invention	0.3759	0.3991	0.3551	0.8840	0.3973

Table 4: Comparison amongst different state-of-the-art scheme and the present invention on CC sub-ontology datasets

S. No.	Scheme	F_{max}	AvgPre	AvgRec	AUC	MCC
1	IGNA [R2]	0.2230	0.2300	0.2160	0.7700	0.1410
2	DeepGO [R3]	0.4767	0.4748	0.4866	0.8620	0.4240
3	MultiPredGO [R1]	0.5366	0.5684	0.5184	0.9102	0.5192
4	Present Invention	0.6056	0.6494	0.5674	0.9555	0.5628

Table 5: Comparison amongst different state-of-the-art scheme and the present invention MF sub-ontology dataset

S. No.	Scheme	Fmax	AvgPre	AvgRec	AUC	MCC
1	IGNA [R2]	0.2010	0.1583	0.2750	0.6402	0.1547
2	DeepGO [R3]	0.2605	0.2672	0.2572	0.7632	0.2433
3	MultiPredGO [R1]	0.3671	0.3630	0.3685	0.8505	0.3123
4	Present Invention	0.4002	0.5115	0.3286	0.9109	0.3832

Table 6: Comparison amongst different state-of-the-art scheme and the present invention

S. No.	Scheme	Trainable parameters (Millions)
1	DeepGO [R3]	3.385
2	MultiPredGO [R1]	4.086
3	Present Invention	2.796

The present invention produces accuracy scores of 37-60% for F_{max} and 88-95% for AUPR, while using 17-35% fewer trainable parameters than contemporary methods, which is best amongst the state-of-the-art methods.

Therefore, the present invention provides a cross modal multi attention mechanism based system **(100)** and method for protein function, which preserve the modality relative contribution towards the protein function prediction task while deploying much fewer trainable parameters than the contemporary state-of-the-art methods. The present invention is tested on three representation or modality for protein function prediction task. Moreover, it is able to be adopted to N-number of modality for any other task

and to provide better relative information of such modalities, to ensure the relative effect of the representation, which ensure the utility of a modal in comparison to the rest of other modality in multi-modal setup. The evaluated results using three representations for protein function prediction in comparison to other state-of-the-scheme shows the better performance analysis in terms of different performance evaluation metric.

Many modifications and other embodiments of the invention set forth herein will readily occur to one skilled in the art to which the invention pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the invention is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

The foregoing description of embodiments of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention. The embodiments were chosen and described in order to explain the principles of the invention and its practical application to enable one skilled in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated.

CLAIMS

We claim:

1. A cross modal multi attention mechanism based system **(100)** for protein function prediction, comprising:

an input module **(1)**; and

a plurality of data processing module;

wherein:

said input module **(1)** receives a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator;

said plurality of data processing module include a feature vector generator module **(2)**, a feature extraction module **(3)**, a protein-protein interaction feature-set module **(4)**, a cross modal multi attention module **(5)** and a neural network based hierarchical classification module **(6)**;

said feature vector generator module **(2)** extracts a set of protein tokens from said protein sequence data and analyzes said set of protein tokens to extract a feature vector via an embedding operation and a recurrent operation;

said feature extraction module **(3)** extracts a set of features from a volumetric representation of said protein structure data;

said protein-protein interaction feature-set module **(4)** annotates a protein to an optimal function via a protein-protein interaction feature-set;

said cross modal multi attention module **(5)** processes said feature vector, set of features and said protein-protein interaction feature set via a cross modal multi attention protocol to capture a relative attention distribution value; and

said neural network based hierarchical classification module **(6)** for classifying a function of said protein sample by processing said relative attention distribution value.

2. The cross modal multi attention mechanism based system **(100)** for protein function prediction as claimed in claim 1, wherein said input module **(1)** include scanner and keyboard.

3. The cross modal multi attention mechanism based system **(100)** for protein function prediction as claimed in claim 1, wherein said plurality of data processing module includes but is not limited to data processors like Arduino Uno.

4. The cross modal multi attention mechanism based system **(100)** for protein function prediction as claimed in claim 1, wherein said feature vector generator module **(2)** deploys a bi-directional gated recurrent unit that processes a long and a short range relationship between said protein tokens in both forward and backward directions.

5. A cross modal multi attention mechanism based method for protein function prediction comprising steps of:

a) receiving a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator via an input module;

b) extracting a set of protein tokens from said protein sequence data and analyzing said set of protein tokens to extract a feature vector via an

embedding operation, recurrent operation and a feature vector generator module;

c) extracting a set of features from a volumetric representation of said protein structure data via a feature extraction module;

d) annotating a protein to an optimal function for producing a protein-protein interaction feature set via said protein-protein interaction feature set generator module; and

e) processing said feature vector, set of features and said protein-protein interaction feature set to capture a relative attention distribution value via a cross modal multi attention module, classifying a function of said protein sample by processing said relative attention distribution value via a neural network based hierarchical classification module.

6. The cross modal multi attention mechanism based method for protein function prediction as claimed in claim 5, wherein said method comprise a step of deploying a bi-directional gated recurrent unit that processes a long and short range relationships between said protein tokens in both forward and backward directions.

7. The cross modal multi attention mechanism based method for protein function prediction as claimed in claim 5, wherein the said cross modal multi attention mechanism based method achieves an accuracy to 88 to 95% while utilizing 17 to 35% lesser trainable parameters.

Dated this 24th day of April, 2023



SHRUTI KAUSHIK
of **PATENTWIRE**
Agent for the Applicant
[IN/PA 1324]

ABSTRACT

"A CROSS MODAL MULTI ATTENTION MECHANISM BASED SYSTEM AND METHOD FOR PROTEIN FUNCTION PREDICTION"

The present invention relates to a cross modal multi attention mechanism based system **(100)** for protein function prediction, in which the input module **(1)** receives a protein sequence data, a protein structure data and a protein-protein interaction network data of a protein sample from an operator, the plurality of data processing module include a feature vector generator module **(2)**, a feature extraction module **(3)**, a protein-protein interaction feature set module **(4)**, cross modal multi attention module **(5)** and a neural network based hierarchical classification module **(6)** for classifying a function of said protein sample by processing said relative attention distribution value while deploying much fewer trainable parameters than the other methods.

Figure 1 on sheet no. 1 of the drawings may accompany the abstract when published.