Office of the Controller General of Patents, Designs & Trade Marks
Department of Industrial Policy & Promotion,
Ministry of Commerce & Industry,
Government of India

(http://ipindia.nic.in/index.htm)

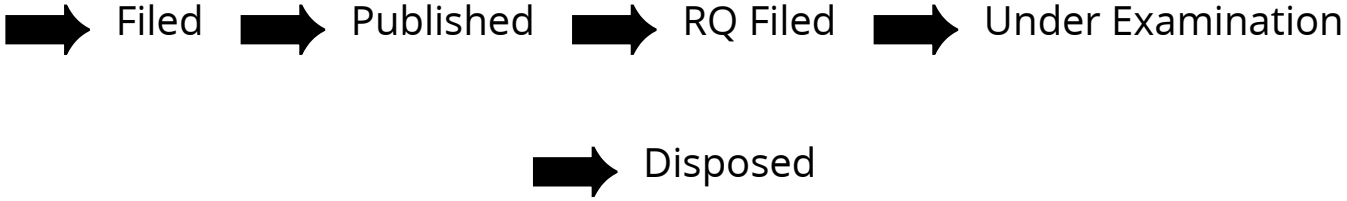
(http://ipindia.nic.in/index.htm)

Application Details	
APPLICATION NUMBER	202121027473
APPLICATION TYPE	ORDINARY APPLICATION
DATE OF FILING	19/06/2021
APPLICANT NAME	1 . Dr. Ishan Y. Pandya 2 . Dr Renuka Sharma 3 . Dr. Navneet Joshi 4 . Dr. Mandheer Kaur 5 . Dr. Muntazir Mushtaq 6 . Dr. Aruna sharma 7 . Dr. Amandeep Kaur 8 . Dr. Neera Batra 9 . Dr. Sonali Goyal 10 . Dr. Pavithra G 11 . Dr. Shashi Raj K., 12 . Prof. Trupti Shripad Tagare
TITLE OF INVENTION	A HYBRID DEEP NEURAL NETWORK SYSTEM AND METHOD FOR GENES EXPRESSION CLASSIFICATION
FIELD OF INVENTION	BIO-CHEMISTRY
E-MAIL (As Per Record)	mail@ideas2ipr.com
ADDITIONAL-EMAIL (As Per Record)	mail@ideas2ipr.com
E-MAIL (UPDATED Online)	
PRIORITY DATE	
REQUEST FOR EXAMINATION DATE	--
PUBLICATION DATE (U/S 11A)	23/07/2021

Application Status

APPLICATION STATUS	Awaiting Request for Examination

			View Documents



In case of any discrepancy in status, kindly contact ipo-helpdesk@nic.in

FORM 2
THE PATENTS ACT 1970
(39 of 1970)
&
THE PATENTS RULES, 2003
COMPLETE SPECIFICATION
(See section 10 and rule 13)

1. TITLE OF THE INVENTION

A HYBRID DEEP NEURAL NETWORK SYSTEM AND METHOD FOR GENES EXPRESSION CLASSIFICATION

2. APPLICANT

(i) NAME : Dr. Ishan Y. Pandya
(ii) NATIONALITY : IN
(iii) ADDRESS : Academic Director, iClonexon Science Institute Gandhinagar-Gujarat.

(i) NAME : Dr Renuka Sharma
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor, RISE (RESEARCH INNOVATION, SPONSORED PROJECTS AND ENTREPRENEURSHIP).Chandigarh group of colleges

(i) NAME : Dr. Navneet Joshi
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor Department of BioSciences. Mody University of Science and Technology, Lakshmangarh, Pin: 332311, District: Sikar, Rajasthan, India.

(i) NAME : Dr. Mandheer Kaur
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor Chandigarh group of colleges landran mohali ,140307

(i) NAME : Dr. Muntazir Mushtaq
(ii) NATIONALITY : IN
(iii) ADDRESS : Senior Researcher, ICAR-National Bureau of Plant Genetic Resources, Division of germplasm Evaluation, 110012.

(i) NAME : Dr. Aruna sharma
(ii) NATIONALITY : IN
(iii) ADDRESS : Ph.D. (Virologist) Contact Address: 26A laxmi vihar colony, near maliwara chowk ghaziabad Uttar Pradesh 201001.

(i) NAME : Dr. Amandeep Kaur
(ii) NATIONALITY : IN
(iii) ADDRESS : Associate Professor, Computer Science & Engineering Department, M.M.University Sadopur Ambala, Haryana.

(i) NAME : Dr. Neera Batra
(ii) NATIONALITY : IN
(iii) ADDRESS : Associate Professor, MMEC, Mullana, Maharishi Markandeshwar (Deemed to be University), Mullana.

(i) NAME : Dr. Sonali Goyal
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor , MMEC MULLANA, Maharishi Markandeshwar (Deemed to be University), Mullana.

(i) NAME : Dr. Pavithra G
(ii) NATIONALITY : IN
(iii) ADDRESS : Associate Professor, Electronics & Communication Engg Dept. (ECE), Dayananda Sagar College of Engg. (DSCE), Block No. 17, Room No. 17205, Kumaraswamy Layout, Shavigemalleshwara Hills, Bangalore- 560078, Karnataka, India.

(i) NAME : Dr. Shashi Raj K.,
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor, Electronics & Communication Engg Dept. (ECE), Dayananda Sagar College of Engg. (DSCE), Kumaraswamy Layout, Shavigemalleshwara Hills, Bangalore- 560078, Karnataka, India.

(i) NAME : Prof. Trupti Shripad Tagare
(ii) NATIONALITY : IN
(iii) ADDRESS : Assistant Professor, Electronics & Communication Engg Dept. (ECE), Dayananda Sagar College of Engg. (DSCE), Kumaraswamy Layout, Shavigemalleshwara Hills, Bangalore- 560078,Karnataka, India.

2. PREAMBLE TO THE DESCRIPTION

COMPLETE

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 a hybrid deep neural network system and method for genes expression classification.

BACKGROUND OF THE INVENTION

In medical field genes are one of the main units used for analysis. It is the physical and functional unit of heredity that consists of specific sequence of nucleotides. These are contained in chromosomes and a human body consists of 20,000 to 25,000 genes located on 46 chromosomes (23 pairs). Chromosomes are coded for specific proteins sometimes they can be RNA molecule. Genes are expressed by genes expression, these expressions are formed during transcription from DNA into RNA, and then it translated into protein. The functional proteins production rate in the cell is regulated at many stages of genes expression not only at the primary transcription level but also at the post transcription levels. Gene's expression are considered to be a series of sequential steps during protein modifications from transcription to post-translation. Generally genes expressions are considered to be high dimensional and to perform classification in high dimensional data is very difficult.

So to perform classification on high dimensional dataset, data preprocessing is very much needed. In these types of data, preprocessing is done in terms of dimensionality reduction. Dimensionality reduction is done through feature selection and feature extraction approach. Feature selection is regularly named as variable selection, variable subset selection and attribute selection. It is the way toward decreasing info highlights to the most enlightening ones for use in model development. Feature selection ought to be recognized from feature extraction, although both of them are utilized to decrease the quantity of features in a dataset. Feature selection incorporates and rejects the characteristics that are available in the information without evolving them.

Feature selection in big data is viewed as an answer for selecting the most useful highlights that could bolster the advancement of vigorous and precise machine learning models. There are a few procedures in information examination. The more up to date algorithms on dimensionality decrease are asymptotically superior to the past calculations.

In the view of the forgoing discussion, it is clearly portrayed that there is a need to have a hybrid deep neural network system and method for genes expression classification.

SUMMARY OF THE INVENTION

The present disclosure seeks to provide a system and method based on classification approach using hybridization of deep neural network named as hybrid deep neural network in which convolutional neural networks are combined with recurrent neural networks to perform genes expression classification.

In an embodiment, a hybrid deep neural network system for genes expression classification is disclosed. The system includes a pre-processing unit for picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file. The system further includes a gene selection unit for obtaining available gene expression and annotation information. The system further includes a transcription start site (TSS) localization unit for considering TSS start location for genes with positive genomic strand and negative genomic strand for end location. The system further includes a feature window created for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity. The system further includes a quantization unit for evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools. The system further includes a gene expression binarization unit for calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

In an embodiment, forecast is made taking into account the fact that a set of 56 problems of binary classification and one for each type of cell, wherein each gene has 100 containers, and each container has 5 histone modification signals, so each data sample consists of a 100 * 5 matrix.

In an embodiment, the system further comprises a memory cell consisting of a gateway, a self-healing neuron, forgetting gates and exit gates for obtaining longer sequential contextual information from the data.

In an embodiment, the system further comprises a first two convolution layers to distinguish between clusters of quality information of 100 × 5 cross-sectional information on 32 object maps with a filter size of 3 × 3 in steps of 3.

In an embodiment, before submitting the output data, Convolution layers in a repeating layer, component maps are connected along the histone axis to enhance the effects of modifications histones, keeping container nodes unchanged.

In an embodiment, in the 32-node LSTM layer, abandonment exercises are associated with the entrance gate and repeating relationships to facilitate retooling.

In an embodiment, a rectified linear block (RELU) and an abandoned regularizer at each level, and function of activating the sigmoid colon at the productivity level is initiated for a consistent final forecast.

In another embodiment, a hybrid deep neural network method for genes expression classification is disclosed. The method includes picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file. The method further includes obtaining available gene expression and annotation information using a gene selection. The method further includes considering TSS start location for genes with positive genomic strand and negative genomic strand for end location through a transcription start site (TSS) localization. The method further includes creating a feature window for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity. The method further includes evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools. The method further includes calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low using a gene expression binarization, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

In an embodiment, quantization of the effects to gene expression from histone modifications, wherein the five central histone modifications are H3K4me3, H3K4me1, H3K36me3, H3K9me3 and H3K27me3.

An object of the present disclosure is to develop hybridizing deep neural network for genes expression classification.

Another object of the present disclosure is to predict genes expression signal from histone modifications.

Another object of the present disclosure is to predict genes expression level based on five core histones modification signals.

Yet another object of the present invention is to develop expeditious and cost-effective hybrid deep neural network method for genes expression classification.

To further clarify advantages and features of the present disclosure, a more particular description of the invention will be rendered by reference to specific embodiments thereof, which is illustrated in the appended drawings. It is appreciated that these drawings depict only typical embodiments of the invention and are therefore not to be considered limiting of its scope. The invention will be described and explained with additional specificity and detail with the accompanying drawings.

BRIEF DESCRIPTION OF FIGURES

These and other features, aspects, and advantages of the present disclosure will become better understood when the following detailed description is read with reference to the accompanying drawings in which like characters represent like parts throughout the drawings, wherein:

Figure 1 illustrates a block diagram of a hybrid deep neural network method for genes expression classification in accordance with an embodiment of the present disclosure;

Figure 2 illustrates a flow chart of a hybrid deep neural network method for genes expression classification in accordance with an embodiment of the present disclosure; and

Figure 3 illustrates an architecture of hybrid neural network in accordance with an embodiment of the present disclosure.

Further, skilled artisans will appreciate that elements in the drawings are illustrated for simplicity and may not have necessarily been drawn to scale. For example, the flow charts illustrate the method in terms of the most prominent steps involved to help to improve understanding of aspects of the present disclosure. Furthermore, in terms of the construction of the device, one or more components of the device may have been represented in the drawings by conventional symbols, and the drawings may show only those specific details that are pertinent to understanding the embodiments of the present disclosure so as not to obscure the drawings with details that will be readily apparent to those of ordinary skill in the art having benefit of the description herein.

DETAILED DESCRIPTION:

For the purpose of promoting an understanding of the principles of the invention, reference will now be made to the embodiment illustrated in the drawings and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the invention is thereby intended, such alterations and further modifications in the illustrated system, and such further

applications of the principles of the invention as illustrated therein being contemplated as would normally occur to one skilled in the art to which the invention relates.

It will be understood by those skilled in the art that the foregoing general description and the following detailed description are exemplary and explanatory of the invention and are not intended to be restrictive thereof.

Reference throughout this specification to “an aspect”, “another aspect” or similar language means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Thus, appearances of the phrase “in an embodiment”, “in another embodiment” and similar language throughout this specification may, but do not necessarily, all refer to the same embodiment.

The terms "comprises", "comprising", or any other variations thereof, are intended to cover a non-exclusive inclusion, such that a process or method that comprises a list of steps does not include only those steps but may include other steps not expressly listed or inherent to such process or method. Similarly, one or more devices or sub-systems or elements or structures or components proceeded by "comprises...a" does not, without more constraints, preclude the existence of other devices or other sub-systems or other elements or other structures or other components or additional devices or additional sub-systems or additional elements or additional structures or additional components.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The system, methods, and examples provided herein are illustrative only and not intended to be limiting.

Embodiments of the present disclosure will be described below in detail with reference to the accompanying drawings.

Referring to **Figure 1**, a block diagram of a hybrid deep neural network method for genes expression classification is illustrated in accordance with an embodiment of the present disclosure. The system 100 includes a pre-processing unit 102 for picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file.

In an embodiment, a gene selection unit 104 is configured for obtaining available gene expression and annotation information. In an embodiment, a transcription start site (TSS) localization

unit 106 is connected with the gene selection unit 104 for considering TSS start location for genes with positive genomic strand and negative genomic strand for end location.

In an embodiment, a feature window 108 is created for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity. In an embodiment, a quantization unit 110 is attached with the feature window 108 for evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools.

In an embodiment, a gene expression binarization unit 112 is employed for calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

In an embodiment, forecast is made taking into account the fact that a set of 56 problems of binary classification and one for each type of cell, wherein each gene has 100 containers, and each container has 5 histone modification signals, so each data sample consists of a 100 * 5 matrix.

In an embodiment, the system further comprises a memory cell consisting of a gateway, a self-healing neuron, forgetting gates and exit gates for obtaining longer sequential contextual information from the data.

In an embodiment, the system further comprises a first two convolution layers to distinguish between clusters of quality information of 100 × 5 cross-sectional information on 32 object maps with a filter size of 3 × 3 in steps of 3.

In an embodiment, before submitting the output data, Convolution layers in a repeating layer, component maps are connected along the histone axis to enhance the effects of modifications histones, keeping container nodes unchanged.

In an embodiment, in the 32-node LSTM layer, abandonment exercises are associated with the entrance gate and repeating relationships to facilitate retooling.

In an embodiment, a rectified linear block (RELU) and an abandoned regularizer at each level, and function of activating the sigmoid colon at the productivity level is initiated for a consistent final forecast.

Figure 2 illustrates a flow chart of a hybrid deep neural network method for genes expression classification in accordance with an embodiment of the present disclosure. At step 202, the method 200 includes picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file.

At step 204, the method 200 includes obtaining available gene expression and annotation information using a gene selection. At step 206, the method 200 includes considering TSS start location for genes with positive genomic strand and negative genomic strand for end location through a transcription start site (TSS) localization.

At step 208, the method 200 includes creating a feature window 108 for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity. At step 210, the method 200 includes evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools.

At step 212, the method 200 includes calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low using a gene expression binarization, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

In an embodiment, quantization of the effects to gene expression from histone modifications, wherein the five central histone modifications are H3K4me3, H3K4me1, H3K36me3, H3K9me3 and H3K27me3.

Figure 3 illustrates an architecture of hybrid neural network in accordance with an embodiment of the present disclosure. A hybrid combination of convolutional and recurrent neural network i.e. HDNN or Hybrid Deep Neural Network is proposed for predicting genes expression level based on five core histones modification signals. It overshadowed the weakness of CNN and RNN by preserving and integrating their strength. The whole approach is based on same dataset i.e. Roadmap Epigenomics Mapping Consortium database (REMC) which is used in Deepchrome method.

The basic gene annotation is used from the latest GENCODE release for the exploration of hidden relation formed between genes expression and histone modification signals. The data is downloaded from REMC database which contains Histone modifications files and RNA-seq quantification of 56 cell types. To generate the input features there are five steps which are as follows:

Gene Selection - Initially, we pick the items with feature type "gene" from essential gene annotation file. There is a gene expression value corresponds to 56 cell types and it has filter out the values that are not included in the RNA-seq file. The available gene expression and annotation information are taken from the selected 19356 genes.

Transcription start site (TSS) localization - Normally in case of gene annotation file, TSS considered to be start location for genes with genomic strand "+" and genomic strand "-" for end location.

Construction of feature window – In this work, we took 10,000 base pairs riding on the TSS of each gene and divided 10,000 base pairs into 100 containers or bins, and each of these containers or bins contains 100 base pairs. This approach is made to create a feature window 108 for each gene. It is very necessary to gather more information about genetic regulation, because it is possible to cover a large area around TSS, and this largely supports continuity. A quantitative assessment of the effects on gene expression was made using the modification of histones; we selected five main modifications of histones: H3K4me3, H3K4me1, H3K36me3, H3K9me3 and H3K27me3.

Quantization of the effects to gene expression from histone modifications: Quantification of the effect on histone gene expression: the five central histone modifications that we choose are H3K4me3, H3K4me1, H3K36me3, H3K9me3 and H3K27me3. To evaluate the measurement of the histone modification signal in each window of 100 base pairs, we apply the intersection of characteristics to each histone modification file using BEDTools. To obtain the final input characteristics of each gene, we minimally relate the number of readings of five modifications of the central histone.

Gene expression binarization: The median of all gene expression values for a particular cell type in the RNAseq file is calculated as a threshold value. If the observed value exceeds the threshold value, in this case we consider that the gene is highly expressed, otherwise it is considered low. The forecast is made taking into account the fact that a set of 56 problems of binary classification and one for each type of cell. Each gene has 100 containers, and each container has 5 histone modification signals, so each data sample consists of a 100 * 5 matrix.

The implemented HDNN technology is a merger of CNN and RNN. Previous studies have shown that CNNs can detect patterns in binary directions (along the bin axis and along the histone axis in this work) using convolution filters, but many times they could not get longer sequential contextual information from the data. Long Short Term Memory (LSTM) is an exceptional recurrent neural system proposed by [19], which can display long-distance conditions in information, combining the experience of additional information approaches in a measurement sequence. To defeat

this CNN flaw, use a memory cell consisting of a gateway, a self-healing neuron, forgetting gates and exit gates. Recently, applications based on the consolidation of CNN and RNN have appeared which have achieved excellent results. In this work, we combine two methodologies in our HDNN model, which incorporates its qualities, into our genomic information. Below is the architecture of the work, consisting of six levels, in which two are convolutional levels, followed by an LSTM layer, two dense layers and one output layer.

The first two convolution layers are used to distinguish between clusters of quality information of 100 × 5 cross-sectional information on 32 object maps with a filter size of 3 × 3 in steps of 3. Before submitting the output data, Convolution layers in a repeating layer, component maps were connected along the histone axis to enhance the effects of modifications histones, keeping container nodes unchanged. In the 32-node LSTM layer, abandonment exercises are associated with the entrance gate and repeating relationships to facilitate retooling.

At the moment and in the foreseeable future, there are two levels that are completely connected with 100 and 20 nodes independently of each other. We carry out the possibility of initiating a rectified linear block (RELU) and an abandoned regularizer at each level, as well as the function of activating the sigmoid colon at the productivity level for a consistent final forecast.

The drawings and the forgoing description give examples of embodiments. Those skilled in the art will appreciate that one or more of the described elements may well be combined into a single functional element. Alternatively, certain elements may be split into multiple functional elements. Elements from one embodiment may be added to another embodiment. For example, orders of processes described herein may be changed and are not limited to the manner described herein. Moreover, the actions of any flow diagram need not be implemented in the order shown; nor do all of the acts necessarily need to be performed. Also, those acts that are not dependent on other acts may be performed in parallel with the other acts. The scope of embodiments is by no means limited by these specific examples. Numerous variations, whether explicitly given in the specification or not, such as differences in structure, dimension, and use of material, are possible. The scope of embodiments is at least as broad as given by the following claims.

Benefits, other advantages, and solutions to problems have been described above with regard to specific embodiments. However, the benefits, advantages, solutions to problems, and any component(s) that may cause any benefit, advantage, or solution to occur or become more pronounced are not to be construed as a critical, required, or essential feature or component of any or all the claims.

WE CLAIM:

1. A hybrid deep neural network system for genes expression classification, the system comprises:

a pre-processing unit for picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file;

a gene selection unit for obtaining available gene expression and annotation information;

a transcription start site (TSS) localization unit for considering TSS start location for genes with positive genomic strand and negative genomic strand for end location;

a feature window created for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity;

a quantization unit for evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools; and

a gene expression binarization unit for calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

2. The system as claimed in claim 1, wherein forecast is made taking into account the fact that a set of 56 problems of binary classification and one for each type of cell, wherein each gene has 100 containers, and each container has 5 histone modification signals, so each data sample consists of a 100 * 5 matrix.

3. The system as claimed in claim 1, comprises a memory cell consisting of a gateway, a self-healing neuron, forgetting gates and exit gates for obtaining longer sequential contextual information from the data.

4. The system as claimed in claim 1, comprises a first two convolution layers to distinguish between clusters of quality information of 100 × 5 cross-sectional information on 32 object maps with a filter size of 3 × 3 in steps of 3.

5. The system as claimed in claim 1, wherein before submitting the output data, Convolution layers in a repeating layer, component maps are connected along the histone axis to enhance the effects of modifications histones, keeping container nodes unchanged.

6. The system as claimed in claim 1, wherein in the 32-node LSTM layer, abandonment exercises are associated with the entrance gate and repeating relationships to facilitate retooling.

7. The system as claimed in claim 2, wherein a rectified linear block (RELU) and an abandoned regularizer at each level, and function of activating the sigmoid colon at the productivity level is initiated for a consistent final forecast.

8. A hybrid deep neural network method for genes expression classification, the method comprises:

picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file;

obtaining available gene expression and annotation information using a gene selection;

considering TSS start location for genes with positive genomic strand and negative genomic strand for end location through a transcription start site (TSS) localization;

creating a feature window for gathering more information about genetic regulation to cover a large area around TSS, and this largely supports continuity;

evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools; and

calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low using a gene expression binarization, wherein if an observed value exceeds the threshold value, then the gene is considered as highly expressed, otherwise low expressed.

9. The method as claimed in claim 8, wherein quantization of the effects to gene expression from histone modifications, wherein the five central histone modifications are H3K4me3, H3K4me1, H3K36me3, H3K9me3 and H3K27me3.

Dated this 18th day of June, 2021

Rajat

RAJAT MALHOTRA

Agent for the Applicant [IN/PA-1775]

Ideas2IPR

ABSTRACT

A HYBRID DEEP NEURAL NETWORK SYSTEM AND METHOD FOR GENES EXPRESSION CLASSIFICATION

The method comprises picking up items with gene feature type for filtering out irrelevant values in an RNA-seq file; obtaining available gene expression and annotation information using gene selection; considering TSS start location for genes with positive genomic strand and negative genomic strand for end location through a TSS localization; creating a feature window for gathering more information about genetic regulation to cover large area around TSS, and this largely supports continuity; evaluating measurement of histone modification signal in each window of 100 base pairs by applying intersection of characteristics to each histone modification file using BEDTools; and calculating median of all gene expression values for a particular cell type in the RNAseq file as a threshold value for classifying expression into high and low using a gene expression binarization, wherein if an observed value exceeds the threshold value then the gene is considered as highly expressed, otherwise low expressed.

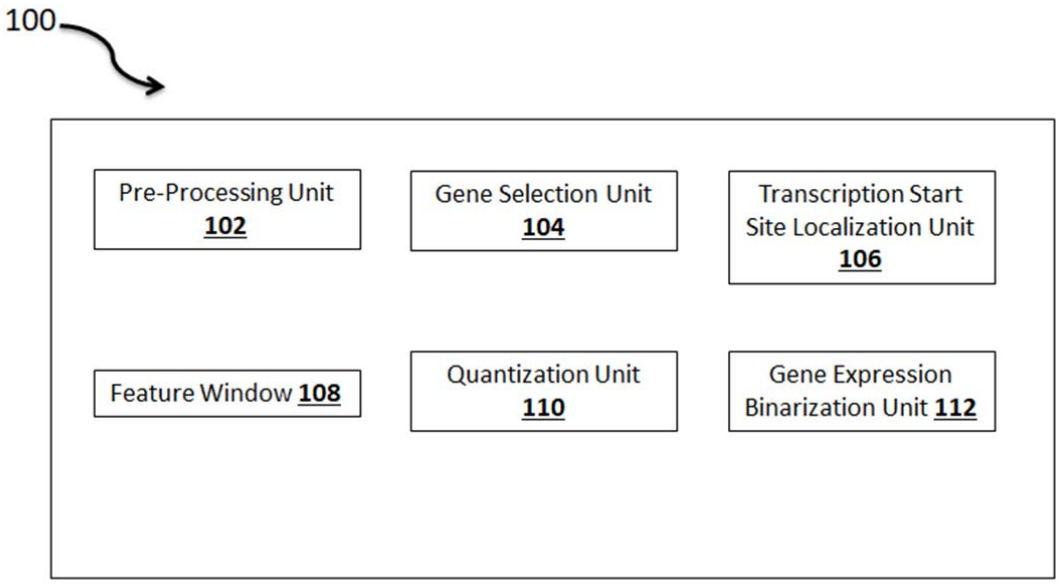


Figure 1