

LIST OF FIGURES

Figure No.	Description	Page No.
1	Central dogma of gene expression.	4
2	Image of mutated protein where protein is coloured gray and the side chains of both wild type and new mutant residues are shown and coloured green and red respectively.	5
3	Image of protein-protein interaction network (Zhang et al., 2011).	7
4	Depicts the interfaces between three chains of the protein glutathione s transferase. The PDB code is 1gwc. There are two interface between chains AB and BC. Chains A and C are not close enough to form interface. In the BC interface, the magenta is the contacting residues and cyan is the nearby residues. In the AB interface, red residues are the interacting residues and the yellow residues near the red residues are the neighbouring (nearby) residues. The side chains of the interacting residues are also displayed in the figure.	13
5	The above image shows how side chains of protein interact with each other to result in a phenotype. If this arrangement gets disturbed then it leads to disease.	15
6	The machine learning process showing learning and prediction as its two phases.	18
7	Process of supervised learning.	19
8	Construction of ensemble of trees in random forest algorithm.	20
9	Manual view of Random Forest.	21
10	Process of unsupervised learning.	22
11	Output of hierarchical clustering algorithm showing nested clusters (left) and dendogram (right).	23
12	Workflow of the present study.	30
13	Screenshot of Interactome3D web service.	31
14	The image depicts the PDB ID for the complex between NCK1 and RASA1. The PDB structure is also shown along with the chains interacting.	32
15	Screenshot of 2P2Iinspector tool.	33
16	Screenshot of Cygwin GUI.	34
17	Weka experimenter to select the classifier algorithm.	35
18	Random Forest sheet showing number of trees as 30.	36
19	Weka explorer to generate model based on classifier algorithm.	36
20	Weka explorer to cross validate the dataset.	37
21	K-fold cross validation, one subset is used for training the model generated by rest of subsets as train set. The action is repeated in such a way that each subset becomes a test set at least once. The average of all is the final model.	41
22	Bar graphs showing the proteins and their interactions with color	41

	indications of experimental structures, models, and no structural information.	
23	Interaction network with the mapped structures. Colour legend is provided on the top left corner.	42
24	Bar graph for negative set.	42
25	Structural interaction network for negative set.	43
26	Bar graph for unknown set.	43
27	Structural Interaction network for unknown polymorphic proteins	44
28	Biological Assembly Image for 1RY7. Crystal Structure of the 3 Ig form of FGFR3c in complex with FGF1. Protein chains are colored from the N-terminal to the C-terminal using a rainbow (spectral) color gradient.	44
29	Figure 29: Screenshot of 2P2Inspector displaying interface Properties.	45
30	Jmol visualization of the interface showing interface residues.	50
31	Bar graph with different classifiers and the statistical values.	51
32	Cross validation results of model generated using Weka.	52
33	ROC curve for Random Forest_30.	54

LIST OF TABLES

Table No.	Description	Page No.
1	OMIM Morbid Map.	28
2	An excel sheet representing details of human structural interaction network.	29
3	List of all diseases associated with cardiovascular class of disorders along with the proteins and the interacting partners. NA – protein id was not available. NI – structural interactions were not found.	39
4	Interface parameters for positive set. ASA- Accessible Surface Area, GV- Gap Volume, GVI- Gap Volume Index, %CR- charged residues, HB- hydrogen bonds, Sec stru- Secondary structure, SB – number of salt bridges, DB- number of disulphide bonds.	45
5	Interface parameters for negative set. ASA- Accessible Surface Area, GV- Gap Volume, GVI- Gap Volume Index, %CR- charged residues, HB- hydrogen bonds, Sec stru- Secondary structure, SB – number of salt bridges, DB- number of disulphide bonds.	47
6	Interface properties for test set. ASA- Accessible Surface Area, GV- Gap Volume, GVI- Gap Volume Index, %CR- charged residues, HB- hydrogen bonds, Sec stru- Secondary structure, SB – number of salt bridges, DB- number of disulphide bonds.	48
7	Comparisons between different classifiers.	49
8	Proteins predicted as potentially associated with cardiovascular disorders and their associations described.	55

LIST OF ABBREVIATIONS

%CV	% Charged Residue
3D	3 Dimensional
ASA	Accessible Surface Area
CSV	Comma Separated Value
CVD	Cardiovascular disorder
DB	Disulfide Bonds
DIPS	Database of Interacting Proteins
DNA	Deoxyribonucleic acid
GV	Gap Volume
GVI	Gap Volume Index
HB	Hydrogen Bonds
HPRD	Human Protein Reference Database
Htt	Huntington
MIPS	Mammalian Protein-Protein Interaction Database
NCBI	National Centre for Biotechnology Information
OMIM	Online Mendelian Inheritance in Man
PDB	Protein Data Bank
Pfam	Protein family database
PPI	Protein-Protein Interaction
ROC	Receiver Operator Characteristic
Rmsd	Root mean square deviation
SB	Salt Bridges
Sec Stru	Secondary Structure
UniProt KB	UniProt Knowledgebase
VMD	Visual Molecular Dyanamics