

LIST OF SYMBOLS

k	Number of nearest neighbors
d	Euclidian distance
H_1	Hydrophobicity
H_2	Hydrophilicity
M	Side chain mass
n	Number of training data sets
$mtry$	Size of random subset of features
n_{tree}	Number of trees in the random forest
$X_1, X_2, \dots, X_{1497}$	Sequence derived properties
c_k	Class variable
S	Feature subset
m	membership coefficient of fuzzy k-NN
Σ	Summation

LIST OF ABBREVIATIONS

AAC	Amino acid composition
ACC	Accuracy
ANN	Artificial neural network
CART	Classification And Regression Tree
CF	Correlation factors
CTD	Composition, transition and distribution of properties
DC	Dipeptide composition
DNA	Deoxyribonucleic acid
FCBF	Fast correlation based filter
GPCRs	G-protein coupled receptors
k-NN	k-nearest neighbors.
MCC	Matthew's correlation coefficient
MLP	Multi-layer perceptron
MRMR	Minimum redundancy and maximum relevancy
NCBI	National center for biotechnology information
PAAC	Pseudo amino acid composition
PDB	Protein data bank
RNA	Ribonucleic acid
ROC	Receiver operating characteristics
SD	Sequence order descriptors
SVM	Support vector machine
SVM-RFE	Support vector machine based recursive feature elimination

LIST OF KEYWORDS

Bioinformatics
Enzymes
Receptors
Ion Channels
Protein function prediction
G-protein coupled receptors
Sequence derived properties
Random forest
Rotation random forest
Weighted k-nearest neighbor
Fuzzy k- nearest neighbor
Voltage gated ion channels
Ligand gated ion channels
Feature selection
Prediction
Pattern classification
Minimum redundancy and maximum relevancy