



Trends in Bioinformatics

ISSN 1994-7941

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

SVM Model to Predict Human Death Domain Protein-Protein Interactions Based on Amino Acid Composition

¹Prakash Arun Nemade and ²Kamal Raj Pardasani

¹Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal, 462003, Madhya Pradesh, India

²Department of Mathematics, Maulana Azad National Institute of Technology, Bhopal, 462003, Madhya Pradesh, India

Corresponding Author: Prakash Arun Nemade, Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal, 462003, Madhya Pradesh, India Tel: +91-9993711667

ABSTRACT

Protein-Protein Interactions (PPIs) play crucial role in regulation of virtually all biological processes in any living system such as DNA transcription, replication, metabolic cycles and signaling cascades. The PPIs also play an important role in the complex process of cell death which occurs via apoptosis and necrosis in eukaryotic cells. The PPIs detection via high throughput experimental methods are time consuming, expensive and are generating huge amount of PPIs data. Therefore, there is need to develop computational methods to efficiently and accurately predict PPIs. This study attempts to develop computational model for predicting human death domain PPIs. First, the protein primary sequences are encoded into descriptors based on amino acid composition of proteins which are monomers of protein. Then, the support vector machine and sequential minimal optimization of WEKA tool is employed to classify interacting and non interacting protein pairs. The various kernel functions were evaluated to build the model and it is observed that libSVM with linear kernel is found to be the best on the basis of performance measures. The validation has been performed by 10 fold cross validation technique. The optimum model gives us the accuracy of 76.47% in predicting human death domain protein-protein interactions. Such models can be useful in providing PPI information of death domain proteins which can be useful in understanding the molecular mechanisms involved in death of cells taking place due to ageing, programmed cell death and various diseases.

Key words: Protein-protein interactions, support vector machine, sequential minimal optimization, cell death, apoptosis, death domain, death effector domain, caspase recruitment domain, pyrin domain, caspases

INTRODUCTION

Understanding protein functions is fundamental to understand various biological processes. The Protein-Protein Interactions (PPIs) are responsible for cellular processes such as DNA transcription, replication, metabolic cycles and signaling cascades. The study of the PPIs provides crucial insights into molecular mechanisms involved in the cell and their related disease processes. The PPIs also play an important role in the complex process of cell death which is extremely important process of the eukaryotic cells. There are millions of cells which are getting created by human body every day and same amount dies as well by programmed cell death known as

apoptosis. Approximately, 10 million cells per day undergo apoptosis in a healthy adult human (Curtin and Cotter, 2002). Cell death is classified by its morphological appearance as apoptosis (Kerr *et al.*, 1972) necrosis (Golstein and Kroemer, 2007) autophagy (Kroemer *et al.*, 2008) and cornification (Melino, 2001). Apoptosis is an orderly cellular suicide program which is critical for the normal development and maintaining homeostasis of a multicellular organism. Cell undergo lots of alterations during apoptosis like chromatin condenses, cell loses their attachment to the surrounding and shrink, cell membrane starts blebbing and engulfment by resident phagocytes (Kerr *et al.*, 1972; Kroemer *et al.*, 2008). Necrosis involves cytoplasmic swelling (oncosis) gain in cell volume, swelling of organelles, plasma membrane rupture and subsequent loss of intracellular content (Golstein and Kroemer, 2007; Proskuryakov *et al.*, 2003). Autophagy serves as a cell survival mechanism (Levine and Yuan, 2005) and is related to numerous physiological and pathological processes (Mehrpour *et al.*, 2010). Cornification leads to the formation of the outermost skin barrier i.e., the cornified layer (keratinization) as well as to the formation of hair and nails. Epidermal keratinocytes undergo a unique form of terminal differentiation known as cornification (Candi *et al.*, 2005; Eckhart *et al.*, 2013).

Apoptosis and inflammation (Coussens and Werb, 2002) are associated with many human diseases and they have crucial importance in myriad of physiological and pathological processes. Failure to regulate apoptosis negatively is associated with degenerative disease and failure to regulate apoptosis positively is associated with cancer and autoimmune disease. Thus, too much or too little cell death can have catastrophic consequences (White, 1993). Upon receiving signals to trigger apoptosis or inflammation signals, the assembly of caspase activating complexes occur via the DD superfamily (Bratton *et al.*, 2001). The death domain superfamily is possessed with apoptotic or non apoptotic functions and therefore their functional role is classified as death related and death unrelated. The death domain superfamily is one of the largest class of protein interaction modules which consists of Death Domain (DD) subfamily, the Death Effector Domain (DED) subfamily, the CAspase Recruitment Domain (CARD) subfamily and the PYrin (PYD) domain subfamily (Reed *et al.*, 2004). They are instrumental in apoptosis, inflammation, necrosis and immune cell signaling pathways. These domains play important roles in the assembly and activation of apoptotic and inflammatory complexes. They have major role in human diseases and disorders such as cancers, neuro-degenerative diseases and immunological disorders, thus their study have great biological and clinical importance and therefore, DD superfamily has emerged as a promising target for therapeutic intervention (Rieux-Laucat *et al.*, 2003).

There are various experimental attempts reported in the literature for the study of PPIs. Oncley *et al.* (1952) studied PPIs between human lipoprotein and globulin. Waugh (1954) studied the basic principles of PPIs. Chothia and Janin (1975) studied the principles of protein-protein recognition and reported that the hydrophobicity is major factor stabilizing protein-protein association. Phizicky and Fields (1995) studied the methods for detection and analysis of PPI. Jones and Thornton (1996) reviewed the principles of PPIs. There are many experimental techniques which have been introduced to identify the PPIs such as yeast two-hybrid (Y2H) (Fields and Song, 1989; Ito *et al.*, 2001), mass spectrometry (Ho *et al.*, 2002), co-immunoprecipitation (CoIP) (Gavin *et al.*, 2006), Tandem Affinity Purification (TAP) (Gavin *et al.*, 2002) and other high throughput experimental techniques. The experimental approaches are very expensive and time consuming. Also, the results obtained from experimental methods suffer from both false positive, false negative and may contain missing values (Von Mering *et al.*, 2002). Therefore, there is a need for the development of fast computational approaches for the study of

PPIs which can efficiently and accurately predict the PPIs. Attempts are reported in the literature for Insilco studies of interactions among different types of proteins. Marcotte *et al.* (1999) proposed computational method for inferring protein interactions from genome sequences. Bock and Gough (2001) studied the PPIs based solely on primary structure of protein and associated physicochemical properties. Valencia and Pazos (2002) proposed the computational method for PPI prediction based on sequence and genomic information.

There are few methods which derive information directly from amino acid sequences for PPIs prediction. These methods are sequence based methods which are based on encoding sequence features of primary sequence and selection of machine learning algorithms. The experimental results shows that information of amino acid sequences alone is sufficient to predict the PPIs (Yang *et al.*, 2010). Among them, Martin *et al.* (2005) encoded the sequence information for a protein pair by a product of signatures. Shen *et al.* (2007) proposed the conjoint triad frequency of continuous subsequences of three residues. Guo *et al.* (2008) used auto-correlation values of seven different physicochemical scales of protein sequences as protein interaction predictors. From the literature survey it is evident that no attempt is reported for the study of human death domain protein-protein interactions by any computational approach.

In recent days, with the rapid development of high throughput sequencing technologies, the sequence data is growing at a faster rate. Therefore, efficient and accurate prediction of PPIs directly from amino acid sequences is one of the challenges. In the present study, support vector machine model is proposed to predict the human death domain protein-protein interactions based upon amino acid composition of human death domain proteins. First, the protein primary sequences are encoded into descriptors based on amino acid composition. Then, the support vector machine and sequential minimal optimization framework of WEKA software is employed to classify interacting and non interacting protein pairs. The sequence information of both positive and negative data set of PPIs is used in the proposed model to classify the pairs of potentially interacting proteins. Various kernel functions are evaluated to propose the optimum SVM model. Various evaluation metrics like sensitivity, specificity, precision, accuracy, AUROC etc., have been computed. The validation is performed by 10 fold cross validation technique. The results demonstrate that the proposed approach has performed well in terms of evaluation metrics, has achieved the good accuracy and they are in agreement with available experimental data.

MATERIALS AND METHODS

Construction of dataset: The focus of this study is human death domain proteins. Therefore, the data is collected from Death Domain (DD) database (Kwon *et al.*, 2011) which is maintained at www.deathdomain.org website. The database provides comprehensive information on Protein-Protein Interactions (PPIs) of the death domain superfamily and the database was created by manually curating hundreds of peer reviewed studies that were published in the literature. This is an experimental data produced by relevant analytical methods. The major proteins of whole data set includes Apaf-1 (Apoptosis protease activating factor), ASC (Apoptosis associated Spec-like protein containing a CARD), Bcl-10 (B-cell lymphoma/leukemia 10), BinCARD (Bcl-10 interacting CARD protein), CARD (Caspase recruitment domain containing protein), Caspase protein, CIAP (Baculoviral IAP repeat containing protein), MAVS (Mitochondrial antiviral signaling protein), NLRP (NACHT, LRR and PYD domain containing protein), NOD (Nucleotide binding oligomerization domain protein), RAIDD (Death domain containing protein CRADD), RIG1 (Probable ATP dependent RNA helicase), RIPK (Receptor interacting Ser/Thr protein kinase),

ANK3 (Ankyrin 3), DAPK (Death associated protein kinase), DR (Tumor necrosis factor receptor superfamily member), IRAK (Interleukin receptor associated kinase), FADD (Fas-associated death domain), DEDD (Death effector domain containing protein), POP (Pyrin domain containing protein) etc. Sixty eight death domain PPI pairs are studied in this work. The redundancy in the data set is identified and removed to obtain non redundant data set. This non redundant data set contains 68 pairs which are used to compute the feature vectors for input to the SVM framework. This final dataset includes equal number of positive and negative training data.

Feature encoding: Feature encoding is an important step. To extract feature vectors from protein primary sequences, in which the important information content of proteins accountable for interaction purpose is encoded, is one of the most important computational challenges, while using machine learning framework to predict the PPIs. Therefore, before prediction of PPI every protein pair is to be represented as feature vectors. The features correspond to the amino acid composition reflects the amino acid characteristics. Amino acid composition reflects the nature of protein and will form the feature set. The complete set of these features are sufficient enough to decide upon its interaction. Amino acid composition is suggested to have a weight attribute of numerical values and its involvement gives importance to the nature of protein composition, likelihood of affinity and binding towards interacting partner. Also, weightage to the compositional residue frequency in terms of number of times of its occurrence, project the over presentation of certain residues which are actually responsible for protein-protein interaction process.

Machine learning techniques such as SVM requires a fixed length of input data for training. Since different proteins have different sequence length, the protein pairs are presented with unequal length vectors of varied features. Therefore, the sequences should be first converted into fixed size feature vectors (Shah *et al.*, 2008). In order to simulate the interaction prediction problem in the framework of learning algorithms, there is an essential need of a suitable encoding of the protein information in some vector space. The features are created in terms of descriptors of residues which drastically cuts off the data input to the classifier. The descriptors also convert variable lengths of protein primary sequence to the homogeneous matrices. As such the heterogeneous input data should be first converted into homogeneous fixed size data matrix for the input to the SVM framework. Every protein sequence is represented by a vector space consisting of features of amino acids, the PPI pair is characterized by concatenating the two vector spaces of two individual proteins of protein pair. These features are then used in conjunction with Support Vector Machine (SVM) framework to predict the interaction between the proteins (Zaki *et al.*, 2006).

There are 20 amino acids, each is represented by amino acid composition. Each protein pair in death domain dataset is represented by binary feature vector corresponding to the features of protein pair. The computed features can be used directly within a Support Vector Machine (SVM) framework.

Consider protein pair $[P_i, Q_i]$ which corresponds to each data point $Z_i, Z_i \in \mathbb{R}^N$ (N dimensional euclidean space). Compute the frequency of 20 amino acids in each protein P_i (Eq. 1) and Q_i (Eq. 2) by generating 20 dimensional vectors for each P_i and Q_i . The X_i represents the frequency of 20 amino acids in a protein P_i . Each dimension X_i is the frequency of particular amino acid in P_i and Q_i .

$$P_i = (X_1, X_2, X_3, \dots, X_{20}) \quad (1)$$

$$Q_i = (X_1, X_2, X_3, ..., X_{20}) \quad (2)$$

Each protein pair is represented as binary feature vector S_{ij} (Eq.5) which is obtained from concatenation of individual feature vectors S_i (Eq. 3) and S_j (Eq. 4) associated with protein pair $[P_i, Q_i]$, respectively. S_i and S_j are computed as follows:

$$S_i = [(X_1, X_2, X_3, ..., X_{20})]_{20} \quad (3)$$

$$S_j = [(X_1, X_2, X_3, ..., X_{20})]_{20} \quad (4)$$

Binary feature vector S_{ij} would be obtained as:

$$S_{ij} = S_i \oplus S_j \quad (5)$$

where, $\oplus$ is a concatenation operator.

Therefore, S_{ij} is the data matrix for support vector machine framework consisting of both positive and negative protein interaction pairs. Finally, S_i and S_j consist of total of 20 descriptors values each i.e., 20 dimensional feature vector has been built to represent each protein sequence. The representation of protein interaction pair is formed by concatenating descriptors of two protein sequences in a protein pair leading to a total of 40 descriptors value i.e., 40 dimensional feature vector in S_{ij} . The protein interaction is predicted via binary classifier. Each data point Z_i or protein pair $[P_i, Q_i]$ is associated with a binary class variable, Y_i (Eq. 6) which has two values +1 and -1:

$$Y_i \in [+1, -1] \quad (6)$$

where, +1 represents the interacting death domain protein-protein interaction pair. On the other hand, -1 represents the non-interacting death domain protein-protein interaction pair.

Support Vector Machines (SVM): Support Vector Machine (SVM) is machine learning powerful state of art algorithm to study and analyze the biological data and well suited for this problem (Cristianini and Taylor, 2000). The SVM is based on statistical learning theory given by Vapnik (1998). It is a supervised learning algorithm for classification and regression problems. It has improved generalization performance over other techniques in real world problems. The training in SVM seeks a global optimized solution and avoids over fitting. Therefore, it has ability to deal with large number of features (Cortes and Vapnik, 1995). First, SVM maps the original data into high dimensional feature space through linear or non linear mapping function which is based on the selection of the kernel function. Then, within the feature space it seeks an optimized linear division i.e., a separating hyperplane shall be constructed by viewing an input data as two sets of vectors which separates the data into two classes. The SVM aims to find the maximum margin hyperplane to separate two classes of patterns. It is based on structural risk minimization principle of statistics theory (Cortes and Vapnik, 1995).

WEKA package: WEKA is a collection of data mining algorithms for solving data mining tasks (Holmes *et al.*, 1994). It contains tools for data pre-processing, classification, regression, clustering,

association rules and visualization (Hall *et al.*, 2009). The support vector machine classifiers SMO (Platt, 1999) and libSVM (Chang and Lin, 2011) with various kernel function of WEKA have been used for classification of death domain protein-protein interaction prediction in the present study.

Ten-fold cross validation and performance evaluation: For validation purpose in this study the 10 fold cross validation technique is used. The data set is randomly partitioned into 10 equal sized bins. The 9 bins are picked 10 times to train the models and remaining bin is used to test them each time leaving out a different bin. The performance of prediction models for two class problem is typically evaluated using confusion matrix. There are many measures for two class classifiers such as sensitivity (SN, Eq. 7), specificity (SP, Eq. 8), precision (PE, Eq. 9), accuracy (ACC, Eq. 10) and Mathew Correlation Coefficient (MCC, Eq. 11), F-measure (Eq. 12), Area Under Receiver Operating Curve (AUROC) (Ferri *et al.*, 2009). These measures have also been employed in the present study for the assessment of models and their expressions are given below:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (9)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (10)$$

$$\text{MCC} = \frac{(TP)(TN) - (FP)(FN)}{\sqrt{[TP + FP][TP + FN][TN + FP][TN + FN]}} \times 100 \quad (11)$$

$$\text{F-measure} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

where, TP, TN, FP and FN are True Positives, True Negatives, False Positives and False Negatives, respectively. If the outcome from the prediction is true and actual value is also true then it is called as True Positive (TP) which is the number of true PPIs that are predicted correctly. If the outcome from the prediction is negative and actual value is also negative then it is called as True Negative (TN) which is the number of true non interacting pairs that are predicted correctly. If the outcome from the prediction is positive and actual value is negative then it is called as False Positive (FP) which is the number of true non interacting pairs that are predicted to be PPIs. If the outcome from prediction is negative and actual value is positive then it is called as False Negative (FN) which is the number of true PPIs that are predicted to be non interacting pairs.

There is tradeoff between precision and recall and therefore combination of both is required in a single efficiency measure known as F-measure which considers both precision and recall equally important (Witten and Frank, 2005). It is important to note that ROC analysis provides important information about the classifiers performance. Therefore, for drawing ROC curves True Positive Rate (TPR) and False Positive Rate (FPR) are required. The TPR defines how many correct positive results occur among all the positive samples available. The FPR defines how many incorrect positive results occur among all negative samples available. The ROC curve is defined by FPR and TPR as x and y-axes, respectively which depicts relative trade-offs between true positive and false positive. Each prediction result or instance of a confusion matrix represents one point in the ROC curve (Hanley and McNeil, 1982). If the prediction method yields points in the upper left corner or coordinate (0,1) of ROC space, it represents 100% sensitivity (no false negatives) and 100% specificity (no false positives). The coordinate point (0,1) is called the perfect classification. A completely random guess would give a point along a diagonal line (the line of no discrimination) from the left bottom to the top right corners. The diagonal divides the ROC space in two parts. Points above the diagonal represent good classification results while points below the diagonal line are considered as poor results. The closer the apex of the curve toward the upper left corner, the greater is the discriminatory ability of the classifier (i.e., the TPR is high and FPR is low). Mostly the machine learning community uses the AUROC statistic for model comparisons. Huang and Ling (2005) suggested that AUROC is a better measure than accuracy when comparing the classifiers.

RESULTS AND DISCUSSION

The data of human death domain protein-protein interaction is collected from the death domain database (Kwon *et al.*, 2011). The death domain protein-protein interaction pairs were obtained from the death domain database. The redundancy in the data set is identified and removed to obtain non redundant data set. This non redundant data set contains 68 pairs which are then used to compute the 40 dimensional feature vector for input to SVM framework. The problem of human death domain protein-protein interaction prediction is a two class prediction problem making it as a binary classification problem in which the outcome of classifier is labeled as positive (interacting) or negative (non-interacting). The data is transformed into feature vectors to prepare input in appropriate format required by SVM framework. A total of 40 dimensional feature vector is obtained after concatenation of both protein sequence features of protein pairs. Two different classifiers namely (Sequential Minimal Optimization) SMO and libSVM which have been implemented in WEKA are used to propose the models. In SMO, four kernels namely Normalized Polykernel, Polykernel, Puk and RBF were employed and in libSVM two kernels namely Linear and Polynomial were employed leading to total of six support vector machine models. Out of these six, an optimal model is identified. The values of performance measures like sensitivity/recall, specificity, precision, accuracy, MCC, F-measure, kappa statistic, ROC area were computed for each of the kernels employed and are presented in Table 1. Also, the ROC curves have been plotted for each of the six kernels and presented in Fig. 1a-f. Figure 2 shows the plot of evaluation criterions for each of the classifier built before standardization of the feature vectors.

From Table 1 and Fig. 1a-f, it can be observed that among SMO with four different kernels, SMO with Puk kernel has given the maximum accuracy of 67.64%, along with sensitivity of 64.70%, specificity of 70.58%, precision of 68.75%, MCC of 0.3535, F-measure of 0.6666, kappa statistic of 0.3529 and AUROC of 0.686 (Fig. 1c). Among libSVM with two different kernels, it is found that

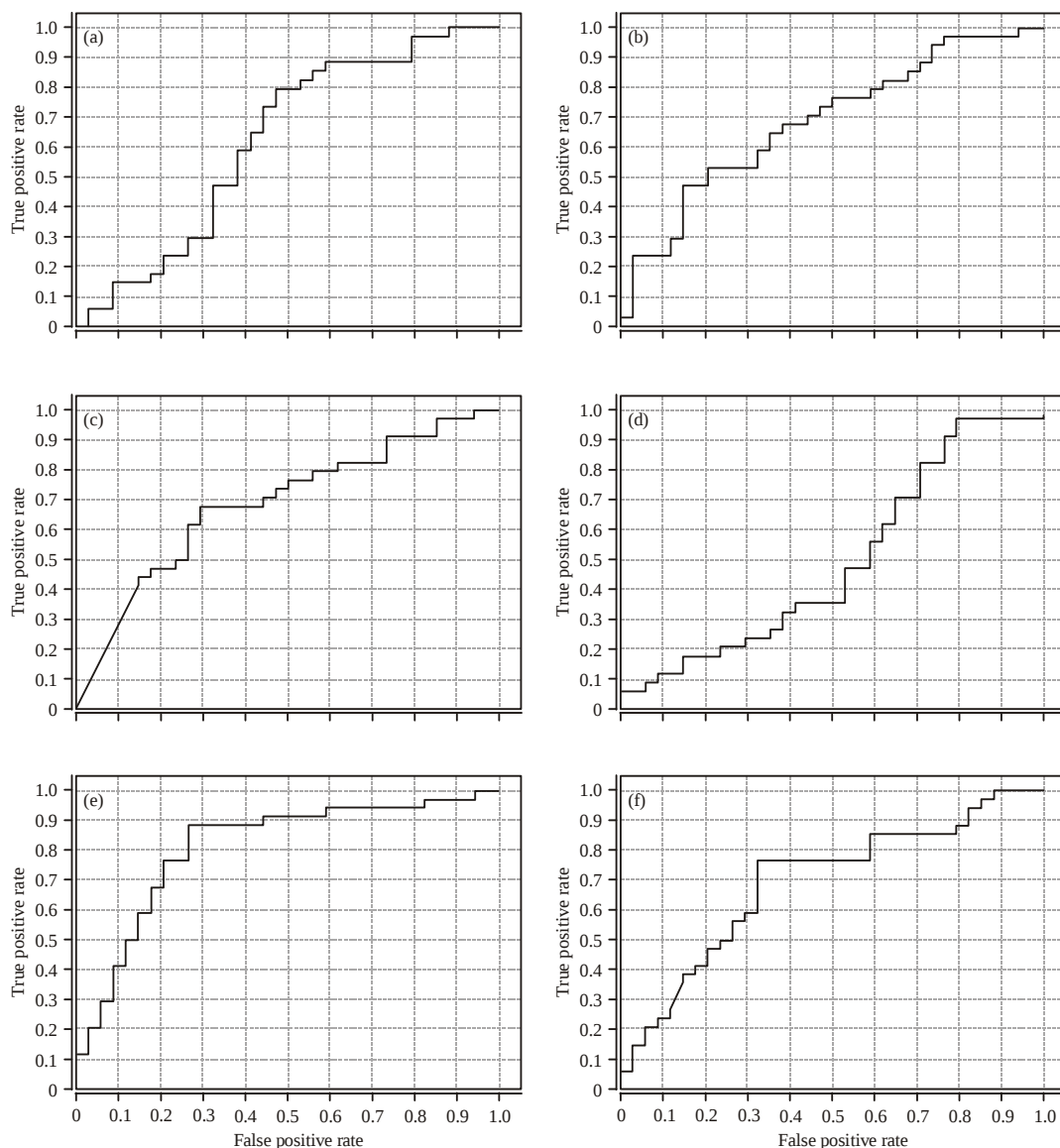


Fig. 1(a-f): ROC Curve of Classifiers (a) SMO, Kernel-normalized polykernel, AUROC = 0.619, (b) SMO, Kernel-polykernel, AUROC = 0.687, (c) SMO, Kernel-Puk, AUROC = 0.686, (d) SMO, Kernel-RBF, AUROC = 0.492, (e) LibSVM, Kernel-linear, AUROC = 0.811 and (f) LibSVM, Kernel-polynomial, AUROC = 0.692

Table 1: SVM based prediction performances of different kernels for the discriminating between interacting and non interacting death domain protein pairs

Classifier	Kernel	Sensitivity/ recall (%)	Specificity (%)	Precision (%)	Accuracy (%)	MCC	F-measure	Kappa statistic	ROC area
SMO	Normalized polykernel	61.76	58.82	60.00	60.29	0.2059	0.6086	0.2059	0.619
SMO	Polykernel	67.64	61.76	63.88	64.70	0.2946	0.6571	0.2941	0.687
SMO	Puk	64.70	70.58	68.75	67.64	0.3535	0.6666	0.3529	0.686
SMO	RBF	35.29	47.05	40.00	41.17	-0.1777	0.3750	-0.1765	0.492
LibSVM	Linear	73.52	79.41	78.12	76.47	0.5303	0.7575	0.5294	0.811
LibSVM	Polynomial	35.29	85.29	70.58	60.29	0.2377	0.4705	0.2059	0.692

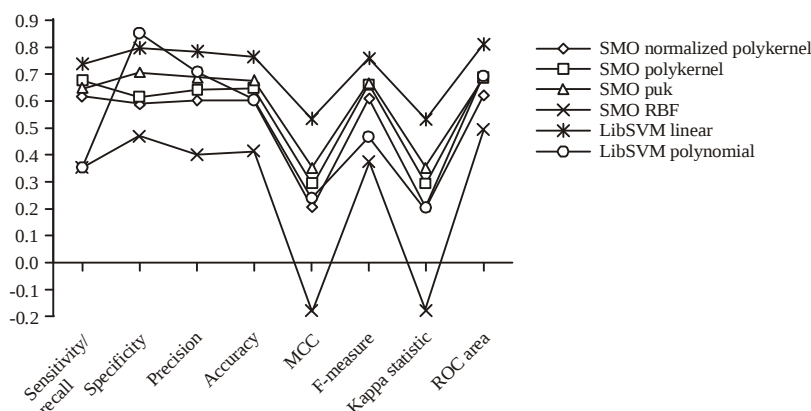


Fig. 2: Performance of different measures of classifiers

libSVM with Linear kernel has given the maximum accuracy of 76.47%, along with sensitivity of 73.52%, specificity of 79.41%, precision of 78.12%, MCC of 0.5303, F-measure of 0.7575, kappa statistic of 0.5294 and AUROC of 0.811 (Fig. 1e). After applying different classifiers with different kernels, it is observed that the most efficient model created is libSVM with Linear kernel which has given the maximum accuracy of 76.47% and achieved highest MCC of 0.5303, F-measure of 0.7575 with AUROC of 0.811 (Fig. 1e) using 10 fold cross validation technique.

Since no model is reported in the literature for prediction of human death domain protein-protein interaction and therefore, no existing theoretical results are available for comparison. However, the results obtained in this study are in agreement with the available experimental data of protein-protein interaction of death domain proteins (Kwon *et al.*, 2011).

CONCLUSION

A computational approach for constructing SVM model is proposed and successfully employed for prediction of human death domain protein-protein interactions based on amino acid composition of protein. On the basis of performance measures it is concluded that libSVM with linear kernel is optimal model for the prediction of human death domain protein-protein interactions. The accuracy of model is also good. The results obtained are in agreement with the available experimental data and can be useful in understanding the signaling network which is mediated by death domain superfamily. Also, the information generated can be useful in getting crucial insights into molecular mechanisms of their actions, cellular processes and related disease processes providing the basis for new therapeutic approaches. Such models can be developed further to generate interaction sites in death domain proteins which can serve as a potential site for drug designing. Also, these models can be useful to generate information for understanding complex biological networks and evolution.

ACKNOWLEDGMENTS

Authors are thankful to the Bioinformatics Infrastructure Facility (BIF) of Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India and Ministry of Human Resources and Development, Government of India.

REFERENCES

- Bock, J.R. and D.A. Gough, 2001. Predicting protein-protein interactions from primary structure. *Bioinformatics*, 17: 455-460.
- Bratton, S.B., G. Walker, S.M. Srinivasula, X.M. Sun, M. Butterworth, E.S. Alnemri and G.M. Cohen, 2001. Recruitment, activation and retention of caspases 9 and 3 by Apaf 1 apoptosome and associated XIAP complexes. *EMBO J.*, 20: 998-1009.
- Candi, E., R. Schmid and G. Melino, 2005. The cornified envelope: A model of cell death in the skin. *Nat. Rev. Mol. Cell Biol.*, 6: 328-340.
- Chang, C.C. and C.J. Lin, 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, Vol. 2. 10.1145/1961189.1961199
- Chothia, C. and J. Janin, 1975. Principles of protein-protein recognition. *Nature*, 256: 705-708.
- Cortes, C. and V. Vapnik, 1995. Support-vector networks. *Mach. Learn.*, 20: 273-297.
- Coussens, L.M. and Z. Werb, 2002. Inflammation and cancer. *Nature*, 420: 860-867.
- Cristianini, N. and J.S. Taylor, 2000. *An Introduction to Support Vector Machines and Other Kernel based Learning Methods*. Cambridge University Press, Cambridge, UK., ISBN-13: 9780521780193, Pages: 189.
- Curtin, J.F. and T.G. Cotter, 2002. Apoptosis: Historical perspectives. *Essays Biochem.*, 39: 1-10.
- Eckhart, L., S. Lippens, E. Tschachler and W. Declercq, 2013. Cell death by cornification. *Biochimica et Biophysica Acta (BBA)-Mol. Cell Res.*, 1833: 3471-3480.
- Ferri, C., J. Hernandez-Orallo and R. Modroiu, 2009. An experimental comparison of performance measures for classification. *Pattern Recognition Lett.*, 30: 27-38.
- Fields, S. and O. Song, 1989. A novel genetic system to detect protein-protein interactions. *Nature*, 340: 245-246.
- Gavin, A.C., M. Bosche, R. Krause, P. Grandi and M. Marzioch *et al.*, 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415: 141-147.
- Gavin, A.C., P. Aloy, P. Grandi, R. Krause and M. Boesche *et al.*, 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440: 631-636.
- Golstein, P. and G. Kroemer, 2007. Cell death by necrosis: Towards a molecular definition. *Trends Biochem. Sci.*, 32: 37-43.
- Guo, Y., L. Yu, Z. Wen and M. Li, 2008. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic Acids Res.*, 36: 3025-3030.
- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, 2009. The WEKA data mining software: An update. *SIGKDD Explorations Newslett*, 11: 10-18.
- Hanley, J.A. and B.J. McNeil, 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve *Radiology*, 143: 29-36.
- Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader and L. Moore *et al.*, 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415: 180-183.
- Holmes, G., A. Donkin and I.H. Witten, 1994. WEKA: A machine learning workbench. *Proceedings of the 2nd Australian and New Zealand Conference on Intelligent Information Systems*, November 29-December 2, 1994, IEEE Computer Society Press, pp: 357-361.
- Huang, J. and C.X. Ling, 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowledge Data Eng.*, 17: 299-310.
- Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.*, 98: 4569-4574.

- Jones, S. and J.M. Thornton, 1996. Principles of protein-protein interactions. *Proc. Nat. Acad. Sci.*, 93: 13-20.
- Kerr, J.F., A.H. Wyllie and A.R. Currie, 1972. Apoptosis: A basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer*, 26: 239-257.
- Kroemer, G., L. Galluzzi, J.M. Vicencio, O. Kepp, E. Tasdemir and M.C. Maiuri, 2008. To die or not to die: That is the autophagic question. *Curr. Mol. Med.*, 8: 78-91.
- Kwon, D., J.H. Yoon, S.Y. Shin, T.H. Jang and H.G. Kim *et al.*, 2011. A comprehensive manually curated protein-protein interaction database for the Death Domain superfamily. *Nucleic Acids Res.* 10.1093/nar/gkr1149
- Levine, B. and J. Yuan, 2005. Autophagy in cell death: An innocent convict? *J. Clin. Investigat.*, 115: 2679-2688.
- Marcotte, E.M., M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates and D. Eisenberg, 1999. Detecting protein function and protein-protein interactions from genome Science, 285: 751-753.
- Martin, S., D. Roe and J.L. Faulon, 2005. Predicting protein-protein interactions using signature products. *Bioinformatics*, 21: 218-226.
- Mehrpour, M., A. Esclatine, I. Beau and P. Codogno, 2010. Autophagy in health and disease. 1. Regulation and significance of autophagy: An overview. *Am. J. Physiol. Cell Physiol.*, 298: C776-C785.
- Melino, G., 2001. The Sirens' song. *Nature*, 412: 23-23.
- Oncley, J.L., E. Ellenbogen, D. Gitlin and F.R.N. Gurd, 1952. Protein-protein interactions. *J. Phys. Chem.*, 56: 85-92.
- Phizicky, E.M. and S. Fields, 1995. Protein-protein interactions: Methods for detection and analysis. *Microbiol. Rev.*, 59: 94-123.
- Platt, J.C., 1999. Fast Training of Support Vector Machines Using Sequential Minimal Optimization Support Vector Learning. MIT Press, Cambridge, MA., USA., ISBN:0-262-19416-3, pp: 185-208.
- Proskuryakov, S.Y., A.G. Konoplyannikov and V.L. Gabai, 2003. Necrosis: A specific form of programmed cell death? *Exp. Cell Res.*, 283: 1-16.
- Reed, J.C., K.S. Doctor and A. Godzik, 2004. The domains of apoptosis: A genomics perspective. *Sci. Signaling* 10.1126/stke.2392004re9
- Rieux-Laucat, F., A. Fischer and F. Le Deist, 2003. Cell-death signaling and human disease. *Curr. Opinion Immunol.*, 15: 325-331.
- Shah, A.R., C.S. Oehmen and B.J. Webb-Robertson, 2008. SVM-HUSTLE-an iterative semi-supervised machine learning approach for pairwise protein remote homology detection. *Bioinformatics*, 24: 783-790.
- Shen, J., J. Zhang, X. Luo, W. Zhu and K. Yu *et al.*, 2007. Predicting protein-protein interactions based only on sequences information. *Proc. Nat. Acad. Sci.*, 104: 4337-4341.
- Valencia, A. and F. Pazos, 2002. Computational methods for the prediction of protein interactions. *Curr. Opinion Structural Biol.*, 12: 368-373.
- Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York, USA.
- Von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields and P. Bork, 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417: 399-403.
- Waugh, D.F., 1954. Protein-protein interactions. *Adv. Protein Chem.*, 9: 325-437.

- White, E., 1993. Death-defying acts: A meeting review on apoptosis. *Genes Dev.*, 7: 2277-2284.
- Witten, I.H. and E. Frank, 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edn., Morgan Kaufmann Publishers, Massachusetts, USA., ISBN-13: 9780080477022, Pages: 560.
- Yang, L., J.F. Xia and J. Gui, 2010. Prediction of protein-protein interactions from protein sequence using local descriptors. *Protein Peptide Lett.*, 17: 1085-1090.
- Zaki, N.M., S. Deris and H. Alashwal, 2006. Protein-protein interaction detection based on substring sensitivity measure. *Int. J. Biomed. Sci.*, 1: 148-154.