



Database of Support Vector Machine Based Tools and Servers Used in Computational Biology

Muheet A Saifi*

Department of Zoology, College of Science, King Saud University
Riyadh-11451, Kingdom of Saudi Arabia.
msaifi@ksu.edu.sa

Abstract: Though there are a lot of tools and servers available to solve a biological issue but a comprehensive and extensive search is prerequisite to find out an appropriate tool or server. Even after this tedious brain storming effort, the huge amount of data leads to a great possibility of redundancy in the obtained work stations. Therefore, compilation of all the related resources together is the most demanding necessity of time. Out of multiple approaches, support vector machine (SVM) has great significance in the development of biological tools and web servers for prediction of protein structure-function, genome annotation, structural and functional genomics, comparative proteomics and genomics, immuno-informatics and phylogenetic analysis. In this study we are reporting 110 SVM based tools and web servers which will exponentially accelerate the research work in biological science.

(Saifi, MA. **Database of Support Vector Machine Based Tools and Servers Used in Computational Biology.** *Lif e Sci J* 2021;18(9):63-73] ISSN 1097-8135 (print); ISSN 2372-613X (online) <http://www.lifesciencesite.com>. 8. doi:[10.7537/marslsj180921.08](https://doi.org/10.7537/marslsj180921.08).

Keywords: Epitope; Mitochondrial protein; Poly (A) tail; Sub-cellular localization; SV

1.Introduction:

In recent years tremendous amount of biological data has been produced through different expensive projects one of them is genome projects of model organisms (Baetu, 2012). After this it become very tedious job to our scientific community to how to analyze and store the data (www.brown.edu/Research/CCMB/). This results the emergence of computational biology. Computational biology is the study of computational theories, concepts, mathematical modelling, and theoretical modelling, web technologies, database development, web server, tools development, artificial intelligence and support vector machine (SVM) in the solving of biological problems (Handl et al, 2007; Taboureau et al, 2007; Hogeweg, 2011; Benson et al, 2007 and Piro and Cunto 2012). SVM plays a very important role in the solving of biological problems by developing different types of tools and web servers. Initially, support vector machines (SVMs) were proposed for solving binary classification problems by Vapnik (Ahmed et al, 2011). SVM draws an optimal hyperplane in a high dimensional feature space that defines a boundary that maximizes the margin between data samples in two classes, therefore giving a better generalization property. The first generation of SVMs was only designed for binary

classification. However, most real-life diagnostic tasks are not binary, and solving multiclass classification problems is much harder than solving binary ones (Li et al 2012 and Rifkin et al 2003). Fortunately, several algorithms have been proposed for extending binary SVMs to multiclass classification. These algorithms are grouped into two types. One is by constructing and combining several binary SVM classifiers. One-versus-all (OVA) and one-versus-one (OVO) SVMs are two typical methods in this type. The other type, called 'all-together', is to directly solve one optimization problem which takes all classes into consideration (Shi & Zheng, 2011; Crammer & Singer, 2011 and Li et al 2011). A large number of problem domains of importance like optical character character recognition, face detection (Ahmed & Raghva 2011), gene selection (Guyon et al, 2002), bankruptcy prediction (Min and Lee, 2005), object recognition, combustion engine detection, function estimation, text categorization, chaotic system, handwritten digit recognition, and database marketing falls into this domain. SVMs are not very sensitive to the curse of dimensionality, and they are well-suited to work with high dimensional data, such as microarray gene expression data (Furey et al 2005). On some applications it has obtained higher accuracy than neural networks. SVM has

also been applied to solve the biological problems (Yeang et al, 2001).

2. Methods:

For the present study we used a lot of available academic search engines as BMC journal, Genomics journal, JSTOR, Oxford journals, PubMed, Science direct, Scirus etc. After this comprehensive and extensive search we got so many tools and web servers. We sorted out the obtained data and presented in a tabular form.

3. Results:

SVM provides easy way solution of biological problems by predictions, identification, classification and recognition approaches. Different types of tools and web servers helpful for a computational biologist are summarized in Table 1.

Prediction:

We got so many tools and web servers for structural and functional annotation of proteomics and genomics data like transmembrane, beta haipin, proteasome cleavage, TAP Binders, CTL epitopes, MHC class II alleles binding peptides, MHC class II binders, Gene Expression, siRNA, Aminoacyl tRNA Synthetases, nuclear receptors, Subcellular localization, Bacterial Toxins, mitochondrial proteins, secretory or non-secretory protein, Oxygen Binding Proteins, voltage gated ion channels, Neurotoxins, Virulence factors, Information molecule, Cellular process and Metabolism molecule, allergenic proteins and Mapping of IgE Epitopes, Plasmodium Secretory and Infected Erythrocyte Associated Proteins, physical or functional interactions between protein molecules, chloroplast proteins, cancer lectins, antibacterial peptides in a protein sequence, antibacterial peptide prediction, kinase protein, inhibition constant, Cytochrome P450 Isoform, functional classification of a protein from its primary sequence, caspase substrates cleavage sites, families and subfamilies of G-protein coupled receptors, constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences, pre-microRNAs, long disordered regions, protein folding kinetic order and rate, DNA and RNA binding sites in amino acid sequences, allergenicity and allergic cross-reactivity in proteins, MicroRNA target, snoRNAs, caspase substrate cleavage sites, GPCR-G-protein coupling, bioluminescent proteins, protein fold recognition, SNPs, RNA-binding residues, MHC class I binding peptides, disordered proteins, virulent proteins in bacterial pathogens, recognition of cis- elements, Malarial adhesins and adhesin-like proteins, protein-protein interactions and

mammalian secreted proteins.

Classification:

Table 1 showed the tools for the classification of sodium, potassium, calcium and chloride ion channels from their primary amino acid sequences.

Identification:

Novel cyclins, protein remote homologs, cytokine-receptor interactions, TIM-barrel proteins and protein kinase-specific phosphorylation sites can be identified by using the tools and servers mentioned in Table 1.

Recognition:

We can also recognize cis-elements and cytokine super family by using the software mentioned in the Table 1.

4. Conclusion:

We know very well that wherever computer science is involved it speeds up the rate of work. SVM is a very good example of this. So many tools and web servers are developed by this but a lot of effort needed to optimize it. Add more parameters at the time of development and take huge amount of data so that it provides us a holistic picture.

Acknowledgment: This study was supported by King Saud University, Deanship of Scientific Research, College of Science, Research Centre.

Corresponding Author:

Dr. Muheet Alam Saifi
Department of Zoology
King Saud University
Riyadh-11451, KSA
E-mail: msaifi@ksu.edu.sa

References:

- [1]. Baetu TM. Genes after the human genome project, Stud. Hist. Philos. Biol. Biomed. Sci. 2012, 43 191-201.
- [2]. <http://www.brown.edu/Research/CCMB/undergraduate.htm>
- [3]. Handl J, Kell DB, Knowles J. Multiobjective optimization in bioinformatics and computational biology, IEEE/ACM Trans Comput Biol Bioinform 2007, 4: 279 – 292.
- [4]. Taboureau O, Baell JB, Fernández-Recio J, Villoutreix BO. Established and emerging trends in computational drug discovery

- in the structural genomics era. *Chem. Biol.* 2012, 19: 29-41.
- [5]. Hogeweg P. The Roots of Bioinformatics in Theoretical Biology, *PLoS Comput. Biol.* 2011, 7 e1002021.
- [6]. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank, *Nucleic Acids Res. (Database issue)* 2008, 36: D25–D30. doi:10.1093/nar/gkm929.
- [7]. Piro RM, Di Cunto F. Computational approaches to disease-gene prediction: rationale, classification and successes, *FEBS J.* 2012, 279: 678-96.
- [8]. Ahmed F, Benedito VA, Zhao PX. Mining functional elements in messenger RNAs: overview, challenges, and perspectives. *Front. Plant Sci.* 2011, 2: doi:10.3389/fpls.2011.00084.
- [9]. Li X, Peng S, Chen J, Lü B, Zhang H, Lai M. SVM-T-RFE: A novel gene selection Algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles, *Biochem. Biophys. Res. Commun.* 2012, doi:10.1016/j.bbrc.2012.01.087
- [10]. Rifkin R, Mukherjee S, Tamayo P, Ramaswamy S, Yeang C, Angelo M, Reich M, Poggio T, Lander ES, Golub TR, Mesirov JP. An analytical method for multiclass molecular cancer classification. *SIAM Rev.* 2003, 45:706–723.
- [11]. Shi, B. Zhang M. Semi-supervised learning improves gene expression-based prediction of cancer recurrence, *Bioinformatics* 2011, 27: 3017-3023.
- [12]. Crammer K, Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines, *J. Mach. Learn. Res.* 2001, 2: 265-292.
- [13]. Li L, Rakitsch B, Borgwardt K. ccSVM: correcting Support Vector Machines for confounding factors in biological data classification, *Bioinformatics* 2011, 27: i342-i348.
- [14]. Ahmed F, Raghava GP. Designing of highly effective complementary and mismatch siRNAs for silencing a gene, *PLoS One* 2011, 6: e23443.
- [15]. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines, *Machine Learning* 2002, 46: 389-422.
- [16]. Min JE, Lee YC. Bankruptcy prediction using optimal choice of kernel function parameters, *Expert Syst. Appl.* 2005, 28: 603-614.
- [17]. Furey TS, Cristianini N, Duffy N, Bednarski DB, Schummer M, Haussler D. Support Vector Machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics* 2000: 906-914.
- [18]. Yeang CH, Ramaswamy S, Tamayo P, Mukherjee S, R. Rifkin RM, Angelo M, Reich M, Lander E, Mesirov J, and Golub T. Molecular classification of multiple tumor types, *Bioinformatics* 2001, 17: 31.

CTKPred	Prediction and classification of the cytokine superfamily.	http://bioinfo.tsinghua.edu.cn/~huangni/CTKPre d/
CTLPred	Prediction of CTL epitopes crucial in subunit vaccine design.	http://www.imtech.res.in/raghava/ctlpred/
CyclinPred	Prediction method to identify novel cyclins using various features of proteins.	http://bioinfo.icgeb.res.in/cyclinpred/
CytoSVM	For identification of cytokine-receptor interactions.	http://bioinf.xmu.edu.cn/software/cytosvm/cytos vm.php
Dachisi	Dipeptide Composition Based Method For Recognition & Classification of histones	http://bio.dfc.harvard.edu/dachis/
DescFold	Protein fold recognition.	http://202.112.170.199/DescFold/index.html
DesiRm	Short interfering RNA (siRNA) has become a major tool in basic sciences for functional gene knockdown and in molecular medicine to suppress aberrant gene expression.	http://www.imtech.res.in/raghava/desirm/
DMKPred	Drug molecules for kinase protein	http://www.imtech.res.in/raghava/dmkpred/
DPROT	Prediction of disordered proteins using evolutionary information.	http://www.imtech.res.in/raghava/dprot/
Dr. PIAS	an integrative system for assessing the druggability of protein-protein interactions	http://asp.gridasp.net/drpias/index.php
DROP	Domain linker predictor trained with optimal features selected by random forest.	http://tuat.ac.jp/~domserv/DROP.html
ECGPred	Prediction of Gene Expression from its Nucleotide Composition.	http://www.imtech.res.in/raghava/ecgpred/
ESLpred	Subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST	http://www.imtech.res.in/raghava/eslpred/
ESLPred2	Improved prediction of subcellular localization of eukaryotic protiens.	http://www.imtech.res.in/raghava/eslpred2/
FaaPred	Prediction method for fungal adhesins and adhesin-like proteins.	http://bioinfo.icgeb.res.in/faap

FungalRV	Adhesin prediction and immunoinformatics portal for human fungal pathogens	http://fungalrv.igib.res.in
GNBSL	Integrative system to predict the subcellular location for Gram-negative bacteria proteins.	http://166.111.24.5/webtools/GNBSL/index.htm
GPCRpred	Prediction of families and subfamilies of G-protein coupled receptors.	http://www.imtech.res.in/raghava/gpcrpred/info.html
GPMiner	Integrated system for mining combinatorial cis-regulatory elements in mammalian gene group	http://GPMiner.mbc.nctu.edu.tw/
GRIFFIN	Hidden Markov model and SVM based system for predicting GPCR-G- protein coupling selectivity.	http://griffin.cbrc.jp/
GSTpred	Web-server specially trained for the Glutathione S-transferase protein.	http://www.imtech.res.in/raghava/gstpred/
HemeBIND	Heme binding residue prediction by combining structural and sequence information.	http://mleg.cse.sc.edu/hemeBIND/
HLA_Affi	To predict the binding affinity of peptides to MHC-I. The method takes into consideration the amino acid sequence and the physio-chemical properties of proteins.	http://www.imtech.res.in/raghava/hla_affi/
HLA-DR4Pred	SVM and ANN based HLA-DRB1*0401(MHC class II alleles) binding peptides prediction method.	http://www.imtech.res.in/raghava/hladr4pred/
HotSprint icaars	Database of computational hot spots in protein interfaces. Identification & Classification of aminoacyl tRNA synthetases.	http://prism.cccb.ku.edu.tr/hotsprint http://www.imtech.res.in/raghava/icaars/
I-Mutant2.0	Predicting stability changes upon mutation from the protein sequence or structure.	http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi
K-Fold	Prediction of the protein folding kinetic order and rate.	http://gpcr.biocomp.unibo.it/cgi/predictors/K-Fold/K-Fold.cgi
KiDoQ	Prediction of inhibition constant using docking and QSAR.	http://crdd.osdd.net/raghava/kidoq/

LGEpred	Correlation analysis and prediction of genes expression from amino acid sequence of proteins.	http://www.imtech.res.in/raghava/lgepred/
LOCSVMPSI	Subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST.	http://Bioinformatics.ustc.edu.cn/LOCSVMPSI/LOCSVMPSI.php
MAAP	Malarial adhesins and adhesin-like proteins predictor.	http://maap.igib.res.in
MemHyb	Predicting membrane protein types by hybridizing SAAC and PSSM.	http://111.68.99.218/MemHyb-SVM
Mem-PHybrid	Hybrid features based Prediction system for Classifying Membrane Protein Types	http://111.68.99.218/Mem-Phybrid
MetaPred	Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule.	http://crdd.osdd.net/raghava/metapred/
MHC2Pred	Prediction of promiscuous MHC class II binders.	http://www.imtech.res.in/raghava/mhc2pred/
MiPred	Classification of real and pseudo microRNA precursors using random forest prediction model with combined features.	http://www.bioinf.seu.edu.cn/miRNA/
miRD	Prediction of novel pre-microRNAs with high accuracy through boosting and SVM.	http://mcg.ustc.edu.cn/rpg/mird/mird.php
miREE	MiRNA recognition element ensembles.	http://didattica-online.polito.it/eda/miREE/
miRFam	Automatic miRNA classification method based on n-grams and a multiclass SVM	http://admis.fudan.edu.cn/projects/miRFam.htm
MiRPara	For prediction of most probable microRNA coding regions in genome scale sequences	http://www.whioiv.ac.cn/bioinformatics/mirpara
MiRTif	MicroRNA target interaction filter.	http://bsal.ym.edu.tw/mirtif
MitPred	Prediction of proteins which are destined to localize in mitochondria in yeast and animals particularly.	http://www.imtech.res.in/raghava/mitpred/

NClassG+	Classifier for non-classically secreted Gram-positive bacterial proteins	http://www.biolisi.unal.edu.co/web-servers/nclassgpositive/
NRpred	Prediction and classification of nuclear receptors on the basis of amino acid composition or dipeptide composition.	http://www.imtech.res.in/raghava/nrpred/
NTXpred	Prediction of neurotoxins and its source and probable function from primary amino acid sequence using SVM based on composition and PSI- Blast.	http://www.imtech.res.in/raghava/ntxpred/
OSCAR	One-class SVM for accurate recognition of cis-elements.	http://bioinfo.au.tsinghua.edu.cn/oscar
Oxy-Pred	Prediction and classification of oxygen binding proteins.	http://www.imtech.res.in/raghava/oxypred/
pathClass	R-package for integration of pathway knowledge for biomarker discovery.	http://cran.r-project.org .
Pcleavage	Prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences.	http://bioinformatics.uams.edu/mirror/pcleavage
penalizedSVM	R-package for feature selection SVM classification.	http://cran.r-project.org/
PFMpred	Prediction of mitochondrial proteins of malaria parasite Plasmodium falciparum.	http://www.imtech.res.in/raghava/pfmpred/
PhosTryp	Phosphorylation site predictor specific for parasitic protozoa of the family trypanosomatidae	http://phostryp.bio.uniroma2.it
PiRaNhA	Prediction of RNA-binding residues in protein sequences.	http://www.bioinformatics.sussex.ac.uk/PIRANHA .
POIMs	Positional oligomer importance matrices—understanding support vector machine-based signal detectors.	http://www.fml.tuebingen.mpg.de/raetsch/projects/POIM .

PolyApred	Prediction of polyadenylation signal (PAS) in human DNA sequence.	http://www.imtech.res.in/raghava/polyapred/
POODLE-L	Predicting long disordered regions.	http://mbs.cbrc.jp/poodle/poodle-l.html
POPI	Predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties.	http://iclab.life.nctu.edu.tw/POPI
POPISK	T-cell reactivity prediction using support vector machines and string kernels	http://iclab.life.nctu.edu.tw/POPISK
PRED_PPI	Predicting protein-protein interactions based on sequence data with probability assignment.	http://cic.scu.edu.cn/bioinformatics/predict_ppi/default.html .
PRINTR	Prediction of RNA binding sites in proteins.	http://210.42.106.80/printr/ .
ProPrInt	Predicts physical or functional interactions between protein molecules.	http://www.imtech.res.in/raghava/proprint/
Protein solubility	Sequence based prediction and experimental verification.	http://webclu.bio.wzw.tum.de:8080/proso
PSEApred	Prediction of Plasmodium secretory and infected erythrocyte associated proteins.	http://www.imtech.res.in/raghava/pseapred/
PSEApred2	Predicting secretory proteins of malaria parasite, which is based on SVM module, uses dipeptide composition with PSSM profiles obtained from PSI-BLAST.	http://www.imtech.res.in/raghava/pseapred2/
pSLIP	Protein subcellular localization prediction using multiple physicochemical properties.	http://pslip.bii.a-star.edu.sg/
PSLpred	Predicts subcellular localization of prokaryotic proteins	http://www.imtech.res.in/raghava/pslpred/
PTIGS-IdIt	Species identification by DNA sequences of the psbA-trnH intergenic spacer region	http://psba-trnh-plantidit.dnsalias.org
RF-DYMHC	Detecting the yeast meiotic recombination hotspots and cold spots by random forest model using gapped dinucleotide composition features.	http://www.bioinf.seu.edu.cn/
RNASnoop	Efficient target prediction for H/ACA snoRNAs	http://www.tbi.univie.ac.at/~htafer/RNASnoop

Rslpred	A SVM Based method for Subcellular localization prediction of Rice Proteins.	http://www.imtech.res.in/raghava/rslpred/
SCYPPred	Predictor of SNPs for Human Cytochrome P450.	http://snppred.sjtu.edu.cn
SecretP	A new method for predicting mammalian secreted proteins.	http://cic.scu.edu.cn/bioinformatics/secretp/index.htm .
SecretPV2	Identifying bacterial secreted proteins by fusing new features into Chou's pseudo-amino acid composition.	http://cic.scu.edu.cn/bioinformatics/secretPV2/index.htm .
SRTpred	SVM-based method for the classification of protein sequence as secretory or non-secretory protein.	http://www.imtech.res.in/raghava/srtpred/
SSPred	Prediction of identification and classification of proteins involved in bacterial secretion systems.	http://www.bioinformatics.org/sspred
SVM-Fold	Discriminative multi-class protein fold and superfamily recognition.	http://svm-fold.c2b2.columbia.edu .
SVM-HUSTLE	Iterative semi-supervised machine learning approach for pairwise protein remote homology detection.	http://www.sysbio.org/sysbio/networkbio/svm_hustle
svmPRAT	SVM-based Protein Residue Annotation Toolkit	http://www.cs.gmu.edu/~mlbio/svmprat
SVM-Prot	Functional classification of a protein from its primary sequence.	http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi
T3SEdb	Data warehousing of virulence effectors secreted by the bacterial Type III Secretion System	http://effectors.bic.nus.edu.sg/T3SEdb http://www.imtech.res.in/raghava/tapped/
TAPPred	Cascade SVM Based method for predicting binding affinity of TAP Binders	
TargetMiner	MicroRNA target prediction with systematic identification of tissue-specific negative examples	www.isical.ac.in/~bioinfo_miu
TBBpred	Transmembrane β -barrel prediction server.	http://www.imtech.res.in/raghava/tbbpred/

TBpred	Predicts four subcellular localization (cytoplasmic, integral membrane, secretory and membrane attached by lipid anchor) of mycobacterial proteins.	www.imtech.res.in/raghava/tbpred/
--------	---	--

TEclass	Automated classification of unknown eukaryotic transposable elements.	http://www.compgen.uni-muenster.de/teclass
TESTLoc	Protein subcellular localization prediction from EST data.	http://megasun.bch.umontreal.ca/~shenyq/TESTLoc/TESTLoc.html .
TIM-Finder	Identification of TIM-barrel proteins	http://202.112.170.199/TIM-Finder/
TSSub	Eukaryotic protein subcellular localization by extracting features from profiles.	http://166.111.24.5/webtools/TSSub/index.html .
VGChan	Prediction of voltage gated ion channels and classify them into ion channels from primary amino acid sequence using SVM method based on composition , PSI-Blast and HMM .	http://www.imtech.res.in/raghava/vgchan/
VICMpred	Prediction of virulence factors, Information molecule, Cellular process and metabolism molecule in the bacterial proteins.	http://www.imtech.res.in/raghava/vicmpred/
VirulentPred	Virulent proteins in bacterial pathogens.	http://bioinfo.icgeb.res.in/virulent/

5/4/2021