International Journal of Biological & Pharmaceutical Research

e- ISSN 0976 - 3651
Print ISSN 2229 - 7480

www.ijbpr.com

Research Article

COMPUTATIONAL EVALUATION OF DIFFERENT MARKER PROTEINS USING MACHINE LEARNING

Shruti Jain*

Jaypee University of Information Technology. Waknaghat**,** Waknaghat Teh Kandaghat, Distt. Solan - 173 234, Himachal Pradesh, India.

ABSTRACT

In this paper we have discussed the mean (observed and predicted), standard deviation, sum of squared error, absolute. mean error using different classifiers like k-NN and SVM of seven marker proteins (AkT, EGFR, ERK, IRS, MK2, JNK and FKHR) which occur due to the combination of TNF, EGF and Insulin. For k-NN we have used three different methods i.e. Chebyshev, Cityblock and Euclidean while for SVM we have used linear, polynomial, RBF and Sigmoid method by Type 1 and Type 2 approaches. Results using Euclidean method of k-NN classifier and RBF method of SVM classifier were good. In this paper we have also discussed the training, test & validation perfection, training algorithm, hidden & output activation function using different approaches of ANN of different marker proteins. Different training algorithms like BFGS and RBFT were used. We have used different types of activation functions like gaussian, exponential, logistic, tanh etc which was used as hidden activation and output activation. Results are the best in all cases for MLP instead of RBF.

Key Words: *Salvadora persica*, Genotoxicity, Natural products, Chromosomal aberrations, Sister chromatid exchanges.

Access this article online		
Home page: http://ijbpr.com/ **DOI:** http://dx.doi.org/10.21276/ijbpr.2017.8.4.7		**Quick Response code**
Received:25.08.17	**Revised**:12.09.17	**Accepted**:15.10.17

Corresponding Author

Shruti Jain
Jaypee University of Information Technology. Waknaghat**,** Waknaghat Teh Kandaghat, Distt. Solan - 173 234, Himachal Pradesh, India.

Email:- jain.shruti15@gmail.com

INTRODUCTION

Communication triggered (Amandeep *et al.*, 2017; Bhusri S *et al.*, 2016; Bhusri S *et al.*, 2016) for cell death and cell survival is by three different input proteins: pro apoptotic protein : tumor necrosis factor (TNF) (Bhusri S, Jain S, 2017; Gaudet S *et al.*, 2005; Jain K, 2012; Jain N and Naik PK, 2012; Jain S, 2016) and survival protein : epidermal growth factor (EGF) and insulin (Jain S, 2016; Jain S, 2012; Jain S, 2015; Jain S, 2010). In this paper we have discussed different classifier techniques : *k*-nearest neighbor (*k*NN) classifier (Jain S, 2011; Jain S, 2010; Jain S, 2002; Jain S, 2009) , support vector machine (SVM) classified (Jain S, 2017; Jain S, 2015; Jain S, 2016), and artificial neural network (ANN) classifier (Normanno N *et al.*, 2006; Rana S *et al.*, 2016; Rana S *et al.*, 2014)for seven different marker proteins i.e. AkT (Sharma S *et al.*, 2017; Sharma S *et al.*, 2016) two receptor proteins : Epidermal growth factor receptor (Thoma B *et al.*, 1990), Insulin receptor substrate (IRS) [16], MAPK proteins (Weiss R, 2001). ERK, JNK , Mitogen- kinase 2 and cell death protein : Forkhead transcription factor (FKHR) (White MF, 2003) shown in fig 1.

Basically AkT promotes cell survival & glycogen synthesis. Akt can lead to cell survival by making different proteins absent i.e FKHR, Bad and caspase-9. And it also leads to cell survival by making

different proteins present i.e. Bad, NF-kB , CREB. The EGF binds with its receptor i.e EGFR and Insulin binds to the receptor leads to IRS which further binds to Src homology 2 (SH2) leading three pathways : MAPK pathway, AkT pathway and JAK/ STAT pathway.There are three steps involved in processing : Data, learning and modeling/ classification. In this paper we have used three different types of classification techniques : *k*NN classifier, SVM classifier , and ANN classifier. In this paper different approaches were used for calculating different values

- *k* -NN classifier in which Chebyshev, Cityblock and Euclidean method was used
- SVM classifier in which linear, polynomial, RBF and Sigmoid method by Type/Tier 1and Type/Tier 2 approach was used

ANN classifier in which training, test & validation perfection, different hidden & output activation function and training algorithm was used.

MATERIALS AND METHODS

For Biomedical Image Processing we have different steps: data collection, data pre-processing, ROI extraction, feature selection, feature extraction and feature classification. Classification is one of the main aspects for the analysis of data set. In this paper we are stressing on different classification techniques: k-NN, SVM and ANN.

The experimental data of cell survival or cell death for different marker proteins was taken from Gaudet, Janes treated with ten cytokine combinations of different input proteins. For each marker protein, the signal values were normalized (1: red; 0.5: black; 0: green) to the maximum value obtained for that signal and an excel data was prepared for ten combinations for 0-24 hrs . Fig 2 shows the main image which was used in this paper for the 10 treatments. We have applied *k*-NN, SVM and NN classifier on all marker proteins i.e. AkT, EGFR, ERK, IRS, MK2, JNK and FKHR.

k Nearest Neighbor (*k*NN)

A distance/ metric function are a function which explains the distance between different objects/ elements of a set. This distance plays an important role in clustering technique. There are various methods for calculation of distance between clusters of various methods for clustering. Clustering plays important role in Data mining or we can say clustering is a division of objects in different groups. Clusters are forms in such a way so that objects of same group are similar while in other groups are dissimilar. There are two types of models: parametric and non- parametric.

Parametric model is that which can differentiated by a hooked set of parameters while a non parametric model is that which cannot differentiated by a hooked set of parameters. Non parametric models are also known as Instance Based Learning. The *k*-NN is a non- parametric method which can be used for classification, regression, cross validation, distance metric, and distance weighing, *k*-NN predictions. The value of *k* can be adjusted which is known as cross validation. If the value of *k* is low than we can over fit the curve while if the value of *k* is too high than we can under fit the curve. *k*-NN is also known as Lazy Learning, Instance / Case/ Memory/ Example Based Reasoning (Bhusri S and Jain S, 2017).

Let's define a function (*u*, *v*), where *u* & *v* are two elements or we can say that *u* and *v* are the query point and a case sample, respectively. The value of distance function is a real positive value which is defined by Cartesian product of *u* and *v* for a set *D* We have three different axioms :

The *identity axiom* which explains that distance between *u* and *v* is equal i.e. $u = v$

The *triangle axiom* which explains that the addition of distance between *u* and *v* and distance between *v* and *w* must be greater than and equal to distance between *u* and *w*.

The *symmetry axiom* states that distance between *u* and *v must* be equal to distance between *v* and *u*.

The main methods for measuring the distance are Euclidean distance, Chebyshev distance, City block/ Manhattan distance, Minkowski distance function.

Euclidean distance (ED): It is general method used for measuring the distance. In this method we will use the square root of the square difference between the coordinates (*u*, *v*) of the elements.

$$D(u,v) = \sqrt{\sum_{i=1}^{m} (u_i - v_i)^2} \quad (1)$$

City block distance / Manhattan distance (CBD): This method calculates the absolute difference between the coordinates (*u*, *v*) of the elements.

$$D(u,v) = \sum_{i=1}^{m} |u_i - v_i| \quad (2)$$

Chebyshev distance /Maximum value distance (CD): This method calculates absolute magnitude of the difference between coordinates of two elements and finally maximum value is considered.

$$D(u,v) = \max |u_i - v_i| \quad (3)$$

Minkowski distance (MD): It is represented by equation 4. If the value of $x = 2$ in equation 4, then the formula is same as that of Euclidean distance. If $x = 1$, then the formula is same as that of city block distance. CD is a special case of MD if we replace $x = \infty$.

$$D(u,v) = \left(\sum_{i=1}^{m} \left| u_i - v_i \right|^x \right)^{1/x} \quad (4)$$

This paper represents the simulation of *k*-NN algorithm using ED, CD and CBD method. For all the marker proteins stated above has : observed mean is 0.504329, observed standard deviation (S.D) is 0.021662 using kNN. We have also calculated predicted mean (PM), predicted standard deviation (PSD), sum of squared error (SSE), error mean (EM), error standard deviation (ESD), absolute error mean (AEM), standard deviation ratio (SDR), correlation (COR) for different marker proteins. Table 1 to Table 7 shows the all the parameters for *AkT, EGFR, ERK, MK2, JNK, IRS, FKHR* respectively.

Support Vector Machine (SVM)

SVM is a supervised model which used for regression analysis, recognize patterns, analyzing data, and classification. SVM have advantages i.e. it has different types of kernel functions which is used as decision functions, high dimensional spaces, versatile, can be used where number of dimensions is large than number of samples.

For given training data (x_i, x_j) for with $x \in \Re^d$ and $y_i \in (1, -1)$, to explain classifier $f(x)$ as

$$f(x_i) = \begin{cases} \geq 0 & \text{then } y_i = +1 \\ < 0 & \text{then } y_i = -1 \end{cases} \quad (5)$$

SVM classifier is represented using a function

$f(x, w, b) = (w^T x + b)$ (6)

where x : test data, w : weight vector, and b : bias

For plus plane the value $f(x, w, b) = w^T x + b = +1$, for minus plane the value $f(x, w, b) = -1$ and for separately hyper plane line the value $f(x, w, b) = 0$ or we can say distance between $(w^T x + b) = +1$ or -1.

Kernels

There are different kernel functions like linear, RBF (gaussian), polynomial and sigmoid. Training a SVM with linear kernel is faster than other kernels because with linear kernel only *C*- regression parameter optimization is done while with others γ parameter optimization was used. Using kernel functions we can write eq.6 as

$$y_i \left(w^T \phi(x) + b \right) \geq 1 - \xi_i \quad for\ i = 1, 2, \ldots\ldots l \ and\ \xi_i \geq 0 \quad (7)$$

where $\phi(x)$ is the mapping from input space to feature space. Multiplication of two mapping function gives kernel function i.e.

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \quad (8)$$

where

$$K(x_i, x_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\text{For Linear} \quad K(x_i, x_j) = x_i . x_j \quad (9)$$

$$\text{For Polynomial} \quad K(x_i, x_j) = \left(\gamma x_i . x_j + C \right)^d \quad (10)$$

$$\text{For RBF} \quad K(x_i, x_j) = \exp\left(-\gamma \left| x_i - x_j \right|^2 \right) \quad (11)$$

$$\text{For Sigmoid} \quad K(x_i, x_j) = \tanh\left(\gamma x_i . x_j + C \right) \quad (12)$$

For all the marker proteins stated above has : observed mean is 0.504329, observed standard deviation (S.D) is 0.021662 using kNN. We have also calculated predicted mean (PM), predicted standard deviation (PSD), sum of squared error (SSE), error mean (EM), error standard deviation (ESD), absolute error mean (AEM), standard deviation ratio (SDR), correlation (COR) for linear, polynomial, RBF and sigmoid function for type/tier 1and type/tier 2. Table 8 to Table 14 shows the all the parameters for AkT, EGFR, ERK, MK2, JNK, IRS, FKHR respectively.

Artificial Neural Network (ANN)

ANN classifier is designed for different marker proteins which predicts whether cell will survive or die. Broydan Fletcher Goldfarb Shanno (BFGS) and Radial Basis function transform (RBFT) was used as different training algorithm. BFGS have good results for non-smooth optimization and it uses the first and second derivatives of the functions. It is one of the quasi Newton methods. BFGS is divided into two i.e. L- BFGS and BFGS –B. For large number of variables in which limited memory is used than L-BFGS was considered while BFGS – B is considered where less number of variables. There are different Activation Functions shown in Table 15 which were used in ANN.In this paper we have used exponential, tanh, logistic, identity, gaussian as hidden activation functions and exponential, logistic, tanh and identity as output activation function. In the figures some activation functions are represented as for Exponential : 1, Logistic: 2, Tanh : 3, Identity : 4,

Gaussian : 5, RBFT : 6. We have simulated the ANN model using STATISTICA data miner software. Fig 3 shows the training perfection (TrP), test perfection (TeP) & validation perfection (VaP), training algorithm (TaA), hidden activation function (HaF) & output activation function (OaF) of *EGFR* using ANN classifier for Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF). Fig 4 to Fig 7 shows the parameters of *ERK* , *MK2*, *JNK*, *IRS* using *ANN* classifier. MLP is trained by back propagation algorithm. MLP basically consists of input layer, hidden layer, output layer and artificial neurons/ bias. Basically bias is a neuron where its activation function is always set as 1. The two important characteristics of MLP are there non linear processing elements and massive interconnectivity. Massive interconnectivity signifies that any given layer links to all the blocks/ elements of nest layer.

RBF is a feed forward technique which means neurons are organized in a layer. It consists of three layers : input, hidden and output layer using non linear function. Further it transfers to output layer which is linear. It can be expressed as RBF *x-y-z* i.e *x* represents input layer, *y* represents hidden layer and *z* represents output layer.

Fig 1. Marker Proteins due to combination of input proteins that leads to cell death/ survival.

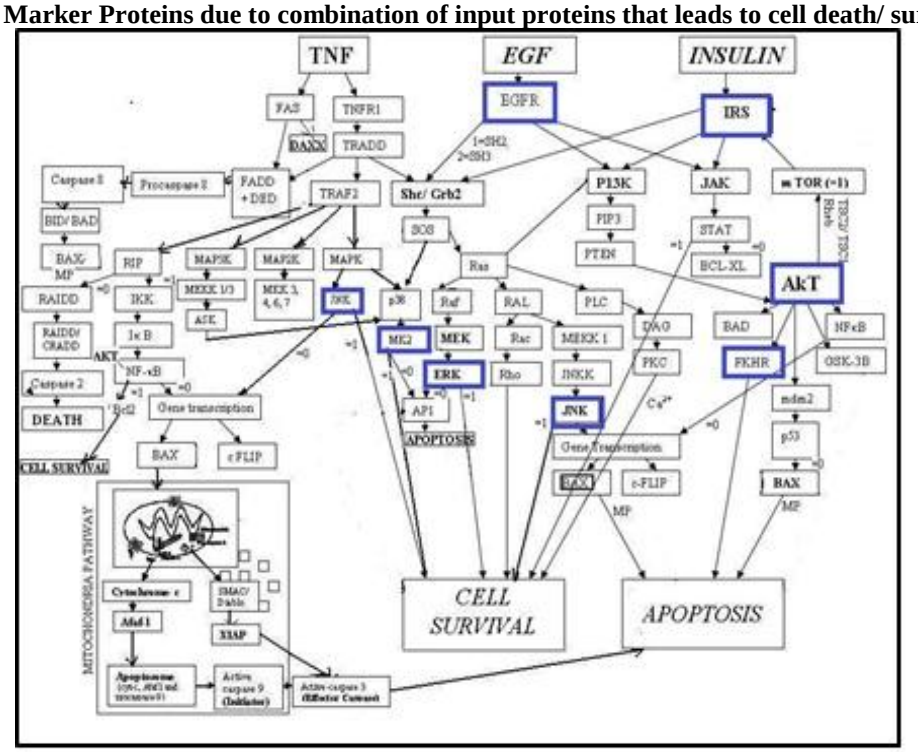

Fig 2. Heat map showing the level of the 11 marker proteins and the cell death response with respect to the treatments of TNF, EGF and Insulin with 10 cytokine combinations

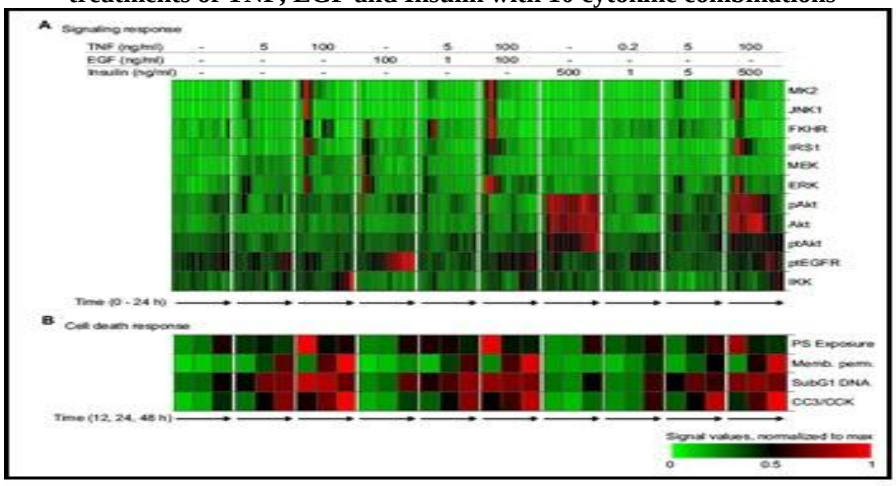

Fig 3. ANN parameters for EGFR using ten different concentrations

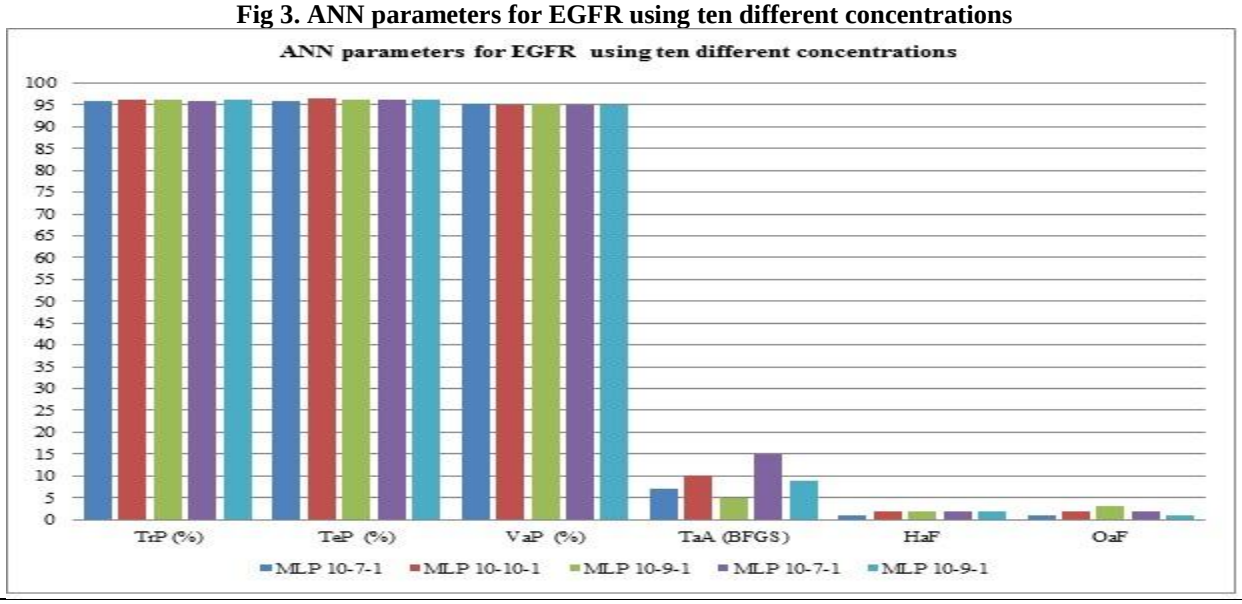

Fig 4. ANN parameters for ERK using ten different concentrations

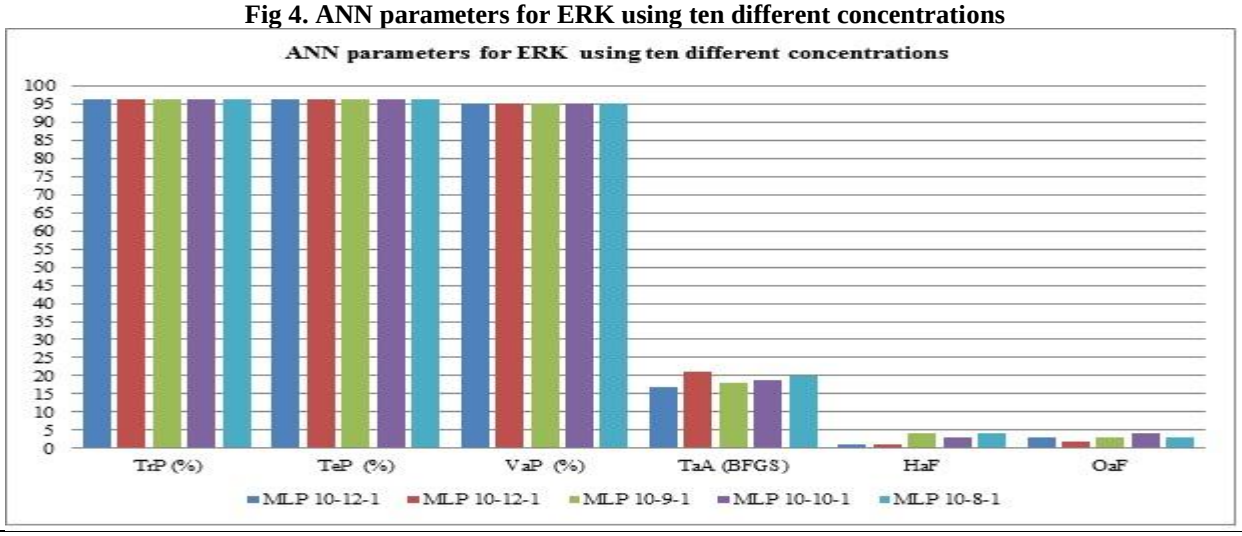

Fig 5. ANN parameters for MK2 using ten different concentrations

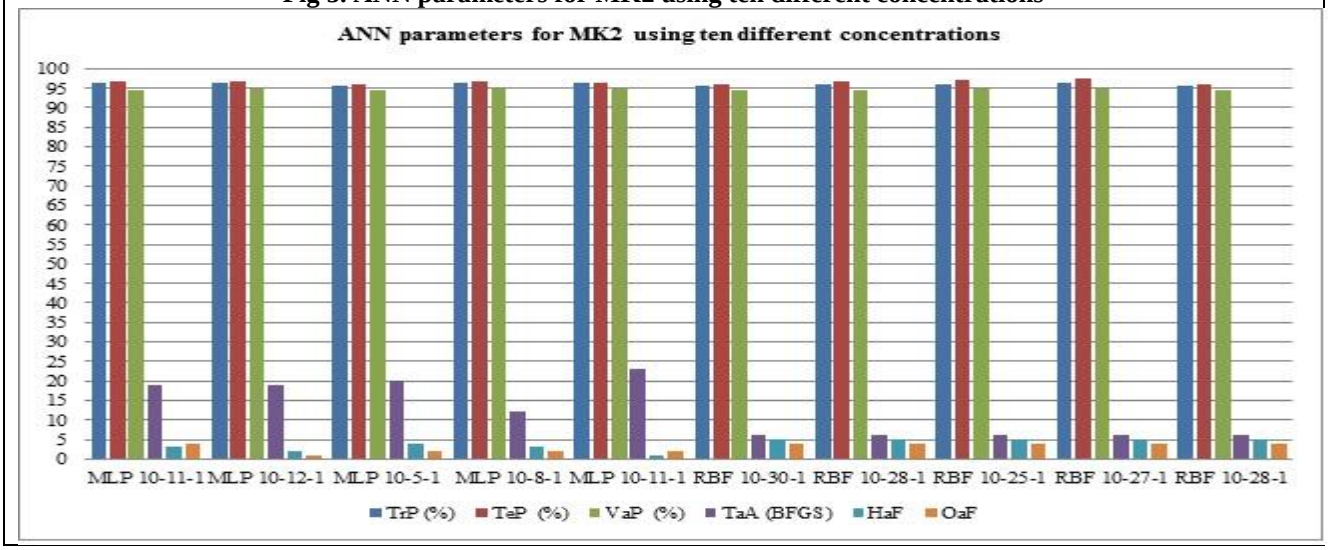

Fig 6. ANN parameters for JNK using ten different concentrations

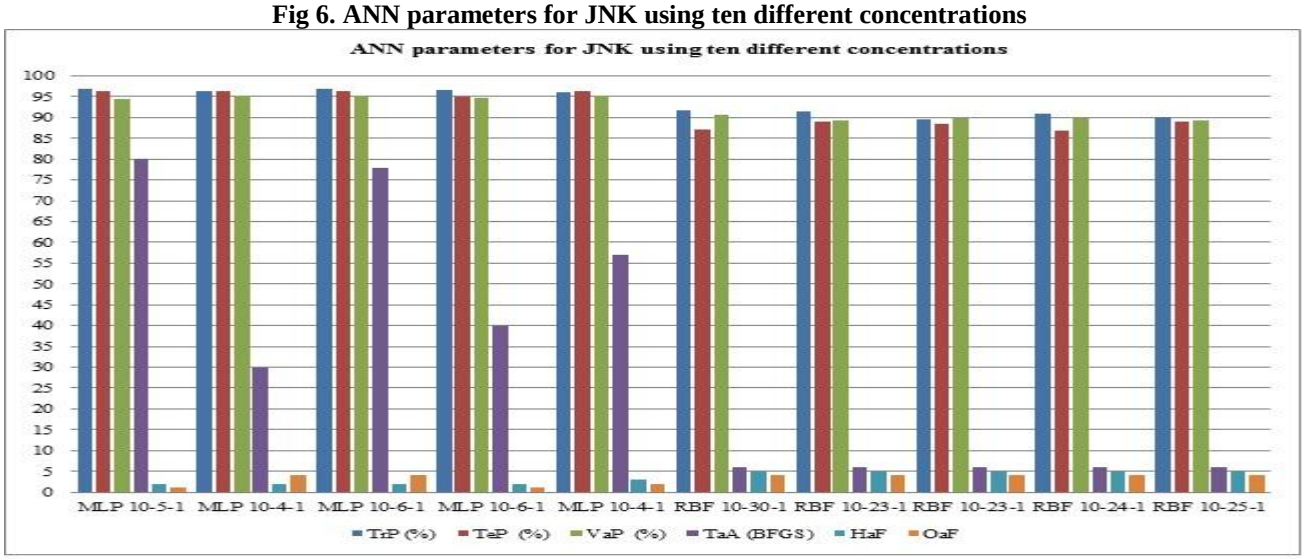

Fig 7. ANN parameters for IRS using ten different concentrations

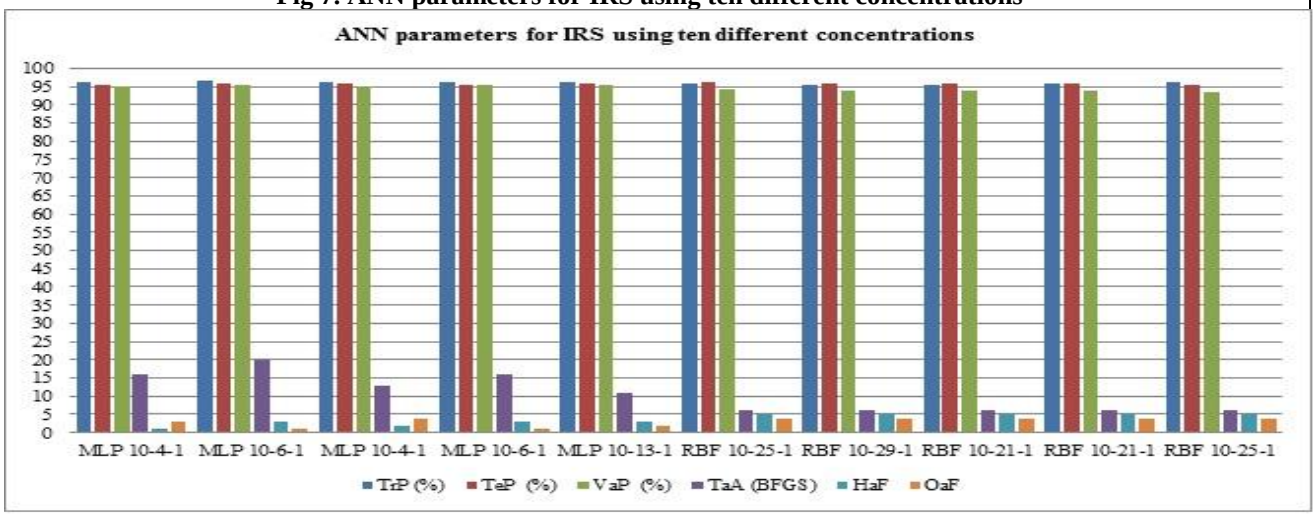

Table 1. *k*-NN classifier using different functions for AkT

	CD	CBD	ED
PM	0.505477	0.504550	0.505183
PSD	0.020344	0.020139	0.019677
SSE	0.000060	0.000070	0.000076
EM	-0.001148	-0.000221	-0.000854
ESD	0.007726	0.008430	0.008731
AEM	0.006319	0.006632	0.006956
SDR	0.356648	0.389171	0.403052
COR	0.934252	0.921205	0.915203

Table 2. *k*-NN classifier using different functions for EGFR.

	CD	CBD	ED
PM	0.504747	0.505558	0.505549
PSD	0.021035	0.021563	0.021542
SSE	0.000065	0.000067	0.000072
EM	-0.000418	-0.001229	-0.001220
ESD	0.008104	0.008133	0.008460

AEM	0.006636	0.006738	0.006851
SDR	0.374117	0.375433	0.390525
COR	0.928363	0.929213	0.923335

Table 3. *k*-NN classifier using different functions for ERK.

	CD	**CBD**	**ED**
PM	0.504152	0.505263	0.504707
PSD	0.020877	0.021351	0.020829
SSE	0.000064	0.000074	0.000067
EM	0.000177	-0.000934	-0.000378
ESD	0.008059	0.008608	0.008204
AEM	0.006809	0.007120	0.006917
SDR	0.372044	0.397368	0.378745
COR	0.928871	0.920005	0.926177

Table 4. k-NN classifier using different functions for MK2

	CD	**CBD**	**ED**
PM	0.503792	0.505188	0.505101
PSD	0.020553	0.020640	0.020514
SSE	0.000050	0.000059	0.000052
EM	0.000537	-0.000859	-0.000772
ESD	0.007127	0.007675	0.007185
AEM	0.005456	0.005993	0.005155
SDR	0.329004	0.354309	0.331676
COR	0.944339	0.935293	0.943400

Table 5. *k*-NN classifier using different functions for JNK.

	CD	**CBD**	**ED**
PM	0.506297	0.504285	0.505110
PSD	0.021535	0.021728	0.021851
SSE	0.000099	0.000063	0.000087
EM	-0.001968	0.000044	-0.000781
ESD	0.009840	0.007997	0.009382
AEM	0.007065	0.006146	0.006894
SDR	0.454263	0.369196	0.433131
COR	0.896231	0.932061	0.907048

Table 5. *k*-NN classifier using different functions for IRS.

	CD	**CBD**	**ED**
PM	0.505795	0.505245	0.505548
PSD	0.020757	0.020414	0.021191
SSE	0.000062	0.000052	0.000053
EM	-0.001466	-0.000916	-0.001219
ESD	0.007787	0.007190	0.007191
AEM	0.006282	0.005953	0.005986
SDR	0.359488	0.331913	0.331943
COR	0.933478	0.943310	0.943924

Table 6. *k*-NN classifier using different functions for FKHR

	CD	**CBD**	**ED**
PM	0.504667	0.504581	0.504063
PSD	0.020335	0.020854	0.020859
SSE	0.000059	0.000056	0.000053

EM	-0.000338	-0.000252	0.000266
ESD	0.007712	0.007511	0.007319
AEM	0.006360	0.006170	0.005913
SDR	0.356002	0.346730	0.337865
COR	0.934495	0.938282	0.941440

Table 7. SVM approach using different functions for AkT

	Linear		**Polynomial**		**RBF**		**Sigmoid**	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.505338	0.505214	0.504565	0.504681	0.504158	0.504255	0.504509	0.504514
PSD	0.025252	0.022969	0.019992	0.019912	0.020784	0.020471	0.020829	0.020681
SSE	0.000059	0.000079	0.000045	0.000044	0.000036	0.000036	0.000059	0.000062
EM	-0.001009	-0.000885	-0.000235	-0.000352	0.000171	0.000074	-0.000180	-0.000185
ESD	0.007655	0.008925	0.006726	0.006696	0.006067	0.006013	0.007741	0.007893
AEM	0.005983	0.007161	0.005468	0.005463	0.004992	0.005013	0.006265	0.006317
SDR	0.353398	0.412032	0.310517	0.309107	0.280055	0.277599	0.357339	0.364365
COR	0.958210	0.921661	0.950983	0.951578	0.959984	0.960827	0.934370	0.931545

Table 8. SVM approach using different functions for EGFR.

	Linear		**Polynomial**		**RBF**		**Sigmoid**	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.505541	0.503937	0.504352	0.504422	0.504452	0.506221	0.503994	0.503994
PSD	0.019024	0.025190	0.019342	0.019368	0.020186	0.020129	0.019939	0.019939
SSE	0.000076	0.000076	0.000043	0.000043	0.000042	0.000049	0.000046	0.000046
EM	-0.001212	0.000392	-0.000023	-0.000092	-0.000123	-0.001892	0.000335	0.000335
ESD	0.008712	0.008792	0.006601	0.006624	0.006510	0.006766	0.006811	0.006811
AEM	0.007248	0.006945	0.005467	0.005493	0.005381	0.005926	0.005438	0.005438
SDR	0.402201	0.405861	0.304727	0.305791	0.300538	0.312368	0.314442	0.314442
COR	0.916345	0.940581	0.954425	0.953979	0.954027	0.950193	0.949728	0.949728

Table 9. SVM approach using different functions for ERK.

	Linear		**Polynomial**		**RBF**		**Sigmoid**	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.506275	0.503034	0.504482	0.504536	0.504766	0.504734	0.504082	0.505530
PSD	0.018771	0.022385	0.019700	0.019768	0.020443	0.020765	0.020465	0.019418
SSE	0.000058	0.000061	0.000041	0.000041	0.000036	0.000039	0.000039	0.000044
EM	-0.001946	0.001295	-0.000153	-0.000207	-0.000437	-0.000405	0.000248	-0.001201
ESD	0.007393	0.007769	0.006438	0.006461	0.006052	0.006301	0.006264	0.006578
AEM	0.006384	0.006140	0.004865	0.004897	0.004848	0.004979	0.005026	0.005562
SDR	0.341281	0.358650	0.297201	0.298248	0.279405	0.290897	0.289157	0.303654
COR	0.943072	0.938303	0.955946	0.955452	0.960317	0.956756	0.957365	0.954555

Table 10. SVM approach using different functions for MK2.

	Linear		**Polynomial**		**RBF**		**Sigmoid**	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.503586	0.503991	0.504171	0.504324	0.503968	0.504027	0.505164	0.504334
PSD	0.018907	0.022926	0.019756	0.019714	0.019770	0.020369	0.023435	0.021678
SSE	0.000052	0.000071	0.000043	0.000043	0.000041	0.000038	0.000105	0.000069
EM	0.000743	0.000338	0.000158	0.000005	0.000361	0.000302	-0.000835	-0.000005
ESD	0.007247	0.008472	0.006590	0.006581	0.006403	0.006235	0.010305	0.008384
AEM	0.005957	0.007145	0.005393	0.005420	0.005315	0.005202	0.008345	0.006952
SDR	0.334561	0.391085	0.304212	0.303821	0.295574	0.287847	0.475702	0.387023
COR	0.945144	0.929351	0.953508	0.953728	0.956317	0.957836	0.898512	0.925163

Table 11. SVM approach using different functions for JNK.

	Linear		Polynomial		RBF		Sigmoid	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.504956	0.506052	0.505286	0.505102	0.505096	0.504927	0.506826	0.506153
PSD	0.016294	0.020965	0.018718	0.018468	0.018741	0.018618	0.019556	0.017971
SSE	0.000127	0.000118	0.000082	0.000084	0.000077	0.000076	0.000239	0.000193
EM	-0.000627	-0.001723	-0.000957	-0.000773	-0.000767	-0.000598	-0.002497	-0.001824
ESD	0.011342	0.010775	0.009076	0.009169	0.008784	0.008773	0.015367	0.013853
AEM	0.009076	0.008885	0.007035	0.007120	0.007012	0.006974	0.011894	0.011299
SDR	0.523604	0.497415	0.418967	0.423285	0.405500	0.404982	0.709394	0.639512
COR	0.858579	0.872712	0.909117	0.907670	0.915479	0.916074	0.726515	0.771011

Table 12. SVM approach using different functions for IRS.

	Linear		Polynomial		RBF		Sigmoid	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.504631	0.504896	0.504835	0.504809	0.504444	0.504501	0.504924	0.504941
PSD	0.018093	0.024086	0.019795	0.019863	0.020394	0.021025	0.019976	0.020280
SSE	0.000043	0.000046	0.000037	0.000036	0.000036	0.000039	0.000061	0.000055
EM	-0.000302	-0.000567	-0.000506	-0.000480	-0.000115	-0.000172	-0.000595	-0.000612
ESD	0.006625	0.006788	0.006075	0.006050	0.006006	0.006296	0.007858	0.007463
AEM	0.005501	0.005197	0.004968	0.004941	0.004835	0.005115	0.006248	0.005963
SDR	0.305855	0.313341	0.280431	0.279307	0.277266	0.290629	0.362778	0.344536
COR	0.960246	0.961478	0.961034	0.961223	0.960991	0.956933	0.931926	0.938777

Table 13. SVM approach using different functions for FKHR.

	Linear		Polynomial		RBF		Sigmoid	
	Tier-1	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**	**Tier-1**	**Tier-2**
PM	0.505276	0.504947	0.504920	0.504915	0.504564	0.504242	0.505210	0.505588
PSD	0.018702	0.020740	0.019800	0.019960	0.019836	0.019914	0.019957	0.020476
SSE	0.000044	0.000043	0.000037	0.000036	0.000033	0.000032	0.000035	0.000036
EM	-0.000947	-0.000618	-0.000591	-0.000586	-0.000235	0.000087	-0.000881	-0.001259
ESD	0.006577	0.006534	0.006068	0.006024	0.005740	0.005710	0.005850	0.005933
AEM	0.005471	0.005364	0.004976	0.004963	0.004678	0.004644	0.004789	0.004941
SDR	0.303638	0.301634	0.280138	0.278111	0.264989	0.263606	0.270051	0.273881
COR	0.957417	0.953432	0.961112	0.961379	0.965538	0.965747	0.963783	0.961907

Table 14. Different activation functions and their range & expressions.

S.No	Activation function	Expression	Range
1	Identity	x (The input to this activation level is directly the output of same level)	$(-\infty, +\infty)$
2	Logistic	$1 / 1 - e^{-1}$	(0,1)
3	Hyperbolic	$\left(e^{x} - e^{-x}\right) / \left(e^{x} + e^{-x}\right)$	(-1, +1)
4	Exponential	e^{-x}	$(0, +\infty)$
5	Gaussian	combination of radial synaptic function and negative exponential activation function	
6	Softmax	$e^{x} / \sum_{i} e^{x_i}$	(0, +1)

7	Unit sum	$x \Big/ \sum_i x_i$	(0, +1)
8	Square root	$\sqrt{x}$	(0, + ∞)
9	Sine	$\sin(x)$	[0, +1]
10	Ramp	$\text{Ramp Function} : \begin{cases} -1, & x \le -1 \\ x, & -1 < x < +1 \\ +1, & x \ge +1 \end{cases}$	[-1,+1]
11	Step	$\text{Step Function} : \begin{cases} 0, & x < 0 \\ +1, & x \ge 0 \end{cases}$	1 or 0

CONCLUSION

In this paper we have studied seven marker proteins which lead to cell survival / death using different inputs. We have applied *k*-NN, SVM and ANN classifier on all the marker proteins. *k*-NN was done by three different methods i.e. Chebyshev, Cityblock and Euclidean while SVM was done by linear, polynomial, RBF and Sigmoid method by Type 1 and Type 2 approach. In both cases we have calculated mean (observed and predicted) & S.D., Sum of squared error, abs. mean error, S.D. ratio and Correlation for every classifier. The training, test & validation perfection, training algorithm, hidden & output activation function of different marker proteins were also calculated. The results with Euclidean method of *k*-NN classifier and RBF method of SVM classifier are the best for every marker protein. The results with MLP are the best with training test and validation perfection for all the marker proteins is greater than 95% with BFGS and RBFT as training algorithm and logistic, identity, exponential function as hidden activation and output activation.

ACKNOWLEDGEMENT

Nil

CONFLICT OF INTEREST

No interest

REFERENCES

Amandeep Jain S, Bhusri S. CAD for Two Class Classification of Lung Cancer using Statistical Features. *International Journal of Pharmaceutical Sciences Review and Research*. 2017; 45(2): 34.

Bhusri S, Jain S and Virmani J. Breast Lesions Classification using the Amalagation of morphological and texture features. *International Journal of Pharma and BioSciences*. 2016; 7(2): 617-24.

Bhusri S, Jain S, Virmani J. Breast Lesions Classification using the Amalagation of morphological and texture features. *International Journal of Pharma and BioSciences*. 2016; 7(2): 617-624.

Bhusri S, Jain S. Classification of Breast Lesions Using Texture Ratio Vector Technique, March 1st - 3rd , *In –Proc* of the 11th INDIACom: 4th 2017 International Conference on Computing for Sustainable Global Development, BVICAM, New Delhi. 2017; 6289-6293.

Gaudet S, Kevin JA, John AN, *et al.* A compendium of signals and responses trigerred by prodeath and prosurvival cytokines. *Manuscrip*, M500158-MCP200. 2005.

Jain K. Communication of signals and responses leading to cell survival / cell death using Engineered Regulatory Networks. PhD Dissertation, Jaypee University of Information Technology, Solan, Himachal Pradesh, India. 2012.

Jain N and Naik PK. Computational Modeling of Cell Survival using VHDL, *BVICAM's International Journal of Information Technology (BIJIT).* 2012; 2(1): 47-51.

Jain S .Compedium model using frequency / cumulative distribution function for receptors of survival proteins: Epidermal growth factor and insulin. *Network Biology*. 2016; 6(4). 101-110.

Jain S and Naik PK. System Modeling of cell survival and cell death: A deterministic model using Fuzzy System. *International Journal of Pharma and BioSciences.* 2012; 3(4): 358-73.

Jain S and Chauhan DS. Linear and Non Linear Modeling of Protein Kinase B/ AkT. *In*: Proceeding of the International Conference on Information and Communication Technology for Sustainable Development, Ahmedabad, India. 2015; 81-88.

Jain S, Bhooshan SV, Naik PK. Computational Modeling of Cell Survival/Death Using BiCMOS. *International Journal of Computer Theory and Engineering*. 2010; 2(4): 478-81.

Jain S, Naik PK and Bhooshan SV. Mathematical modeling deciphering balance between cell survival and cell death using Tumor Necrosis Factor α. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*.2005; 2(3): 574-83.

Jain S, Naik PK and Bhooshan SV. Mathematical modeling deciphering balance between cell survival and cell death using insulin. *Network Biology*. 2011; 1(1): 46-58.

Jain S, Naik PK and Bhooshan SV. Petri net Implementation of Cell Signaling for Cell Death. *International Journal of Pharma and Bio Sciences*. 2010; 1(2); 1-18.

Jain S, Naik PK, Sharma R. A Computational Modeling of Apoptosis Signaling using VHDL and MATLAB Simulator. *International Journal of Information and Communication Technology*. 2002; 2(1-2): 7-12.

Jain S, Naik PK, Sharma R. A Computational Modeling of cell survival/ death using VHDL and MATLAB Simulator. *Digest Journal of Nanomaterials and Biostructures*. 2009; 4(4): 863- 79.

Jain S. Implementation of Marker Proteins Using Standardised Effect. *Journal of Global Pharma Technology*. 2017; 9(5): 22-27.

Jain S. Mathematical Analysis and Probability Density Function of FKHR pathway for Cell Survival /Death. *In*: Proceedings of the Control System and Power Electronics, Bangalore. 2015; 84-93.

Jain S. Mathematical Analysis using Frequency and Cumulative Distribution functions for Mitogenic Pathway. *Research Journal of Pharmaceutical, Biological and Chemical Sciences.* 2016; 7(3): 262-72.

Jain S. Regression analysis on different mitogenic pathways. *Network Biology*. 2016; 6(2): 40-46.

Kevin JK, John AG, Suzanne G, Peter SK, Douglas LA, Michael YB. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science*. 2015; 310: 1646-53.

Normanno N, DeLuca A, Bianco C, *et al.* Epidermal growth factor receptor (EGFR) signaling. *Cancer Gene*. 2006; 366: 2–16.

Rana S, Jain S and Virmani J. SVM-Based Characterization of Focal Kidney Lesions from B-Mode Ultrasound Images. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*. 2016; 7(4): 837- 46.

Rana S, Jain S, Virmani J. Classification of Kidney Lesions using Gabor Wavelet Texture Features. *In: Proc* of the 10th India Com 3rd 2016 International Conference on Computing for Sustainable Global Development, 2528-2532.

Rana S, Jain S, Virmani J. SVM-Based Characterization of Focal Kidney Lesions from B-Mode Ultrasound Images. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*. 2016; 7(4): 837- 46.

Sharma S, Jain S, Bhusri S. Classification of Breast Lesions using Gabor Wavelet Filter for Three Classes *In –Proc* of the 11th India Com- 4th 2017International Conference on "Computing for Sustainable Global Development, BVICAM, New Delhi. 2017; 6282-6284.

Sharma S, Jain S, Bhusri S. Two Class Classification of Breast Lesions using Statistical and Transform Domain features. *Journal of Global Pharma Technology.* 2017; 9(6): 87.

Thoma B, Grell M, Pfizenmaier K and Scheurich P. Identification of a 60-kD tumor necrosis factor (TNF) receptor as the major signal transducing component in TNF responses. *J Exp Med*. 1990; 127: 1019-23.

Weiss R. Cellular computation and communications using engineered genetic regulatory networks. PhD Dissertation, MIT, 2001.

White MF. Insulin Signaling in Health and Disease. *Science.* 2003; 139: 1710–11.

Cite this article:
Shruti Jain. Computational evaluation of different marker proteins using machine learning.. *International Journal of Biological & Pharmaceutical Research*. 2017; 8(4):173-183. DOI: **http://dx.doi.org/10.21276/ijbpr.2017.8.4.7**