A Handy Diagnostic Tool for Preventive Therapy based on Prior Estimation of Metastasis

Madhura Gangaiah[1], Divya Patil[2], Sangeetha D R[3], Dr. Andhe Pallavi[4], Dr. Sachin Suresh Jadhav (MD)[5], Dr. Jayanthi Thumsi (MS)[6]

[1]Assistant Professor, IT Department, RNSIT, Bangalore, India

[2, 3]PG Student, IT Department, RNSIT, Bangalore, India

[4]Prof. & HOD, IT Department, RNSIT, Bangalore, India

[5, 6]BGS Global Hospital, Bangalore, India

Abstract: *In this paper fuzzy arithmetic is used to develop a prediction model for the first site of breast cancer metastasis. Membership functions for the overly expressed protein biomarkers are developed. Three first sites of metastasis i.e., bone, liver and lung are chosen. A fuzzy relation between these metastases and the protein biomarkers is computed. A neural network is trained with these fuzzy relations to predict the first site of metastasis. The trained network is tested for different combination of proteins for the prediction of metastasis. These fuzzy relations were defuzzified using Lambda cut sets to predict the first site of metastasis. These results were compared to that obtained with the neuro-fuzzy model. The different protein biomarkers chosen were COX2, HER2, ER, PgR, CK5, CK18, GATA3 and E-cadherin.*

Keywords: Protein biomarkers, Metastasis, Artificial neural network (ANN), Defuzzification.

1. Introduction

Prediction of cancer based on biomarkers such as diet, age, proteins genetic factors etc [10] is prevalent. In our paper protein biomarkers are used for the prediction of metastasis of the cancer where in very little research is found.

Protein structures corresponding to a metastasis is identified. These overly expressed proteins need to be immobilized by using inhibitors. Many surveys [2,3] have been carried out on cancer patients. In our paper these survey data are mined for certain patterns common to cancer patients exhibiting metastasis. Finland survey [1] reported on the statistics of presence/ absence of proteins in a set of patients exhibiting with or without single/ multiple metastasis. It is proposed to develop a suitable mining algorithm based on the protein biomarkers for the prediction of metastasis. The goal in our project is to extract information from different metastasis survey data sets and use it to predict the first site of metastasis and its probability of occurrence and hence design the therapy for each individual . Fuzzy information deals with knowledge that is ambiguous or imprecise/ uncertain and is suitable to be used in medical context for representing most of the survey data.

In the first stage of our project the inbuilt MATLAB tool command ANFIS is used to model the survey data. This neuro fuzzy structure is trained with our data sets and tested for accuracy [4]. ANFIS provides different models such as Subtractive and Fuzzy C-Mean (FCM). The models generated using subtractive clustering usually are more accurate than those generated using FCM algorithm [5]. FCM requires a training algorithm to accurately generate models. In contrast subtractive clustering doesn't need prior training algorithms to generate the models. FCM has inconsistent problem that is, FCM executed at different

times yields different results as the algorithm will choose an arbitrary matrix 'v' each time. On the other hand subtractive algorithm produces consistent results. C-Mean Algorithm (CMA) [7] is easy and its convergence speed is very fast but it is sensitive to the initialization condition and it has different clustering result for different initialization values. Clustering method based on genetic algorithms can solve the problems of initialization sensitivity of CMA and has a lot of chance to get the optimal solution. To overcome the limitation in symbolic ANN architectures, the fuzzy interference system is widely adopted.

Protein structures having a bearing on metastasis are identified and tabulated in table 1[8,11]. As a part of therapy for prevention of metastasis these overly expressed proteins need to be immobilized by using inhibitors. The inhibitor structures are also shown in table 1. These inhibitors when prescribed may have a more adverse effect on the various organs as compared to the influence of corresponding proteins on metastasis. Hence their prescription must be minimized. As there are more than one protein causing a particular metastasis it is advantageous to use inhibitors which can immobilize multiple proteins instead of a single protein. Hence combination of effective protein clusters for particular metastasis is to be identified, for which we are using fuzzy clustering in this paper.
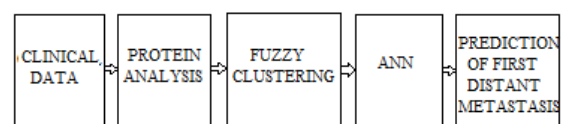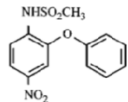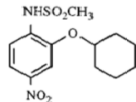
2. Case Flow
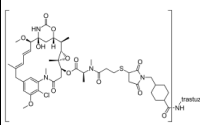
Figure 1: Block diagram approach of our project

Clinical data of breast cancer patients available at various sites [1] is mined for our project. From this data set, we are concentrating on the different protein levels present in the tumor cell for each type of metastasis. The clustering of these protein for each metastasis site is carried out.

Table 1: Protein structures and their inhibitors

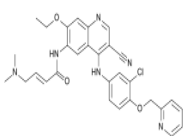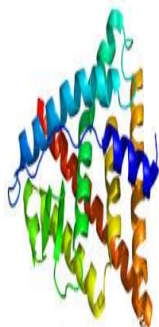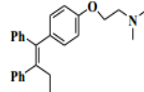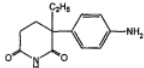| Proteins	Structure	Features	Inhibitors	Disadvantage
COX2[8]		Cyclo oxygenase also known as prostaglandin endoperoxide synthase (PTGS) is responsible for formation of biological mediators called prostanoids including prostaglandin, prostacyclin and thrompoxane	1.Nimesulide 2.NS398	COX-2 inhibitors are associated with a moderately increased risk of vascular event and high-dose regimens of some traditional diclofenac and ibuprofen are associated with a similar increase in risk of vascular events. Combining low dose aspirin with COX-2 inhibitors causes increased damage to the gastric mucosa.
HER2[8,11]		HER2 is a member of the human epidermal growth factor receptor family. Amplification or over expressed oncogene plays an important role in the development and progression of certain aggressive types of breast cancer. This protein is an important biomarker and target of therapy for approx. 30% of breast cancer patients	1.Trastuzumab emtansine 2. Neratinib	Trastuzumab is effective only in cancers where HER2 is over-expressed. An important effect of trastuzumab binding to HER2 causes increase in p27, which halts cell proliferation. Pertuzumab inhibits dimerization of HER2 and HER3 receptors whose combination with trastuzumab was approved by FDA.
ER[8]		Estrogen receptors are a group of proteins found inside cells. These receptors are activated by the hormone estrogen. Once activated by estrogen, the ER is able to translocate into the nucleus and bind to DNA to regulate the activity of different genes	1.Tamoxifen 2.Aromatase	Tamoxifen treatment negatively affected verbal memory and planning skill. Depression, Women taking tamoxifen and paroxetine had at least a 24 percent greater risk of mortality than women taking tamoxifen and any other antidepressant.

3. Fuzzy Clustering

The characteristics of fuzzy clustering are [9]:
- Fuzzy clustering which is also referred as soft clustering has data elements which can be grouped into clusters.
- These data elements may belong to more than one cluster.
- The association of the data points to each cluster is defined by membership level.
- These membership levels indicates the strength of association between the data elements and particular cluster

3.1 Allotment of Membership Function

Fuzzy clustering of proteins versus first metastatic site is carried out. Membership functions developed are[1]

$$\mu_{COX2} = \frac{0.435}{bone} + \frac{0.222}{liver} + \frac{0.296}{lung}$$

$$\mu_{HER2} = \frac{0.439}{bone} + \frac{0.273}{liver} + \frac{0.273}{lung}$$

$$\mu_{ER} = \frac{0.599}{bone} + \frac{0.218}{liver} + \frac{0.141}{lung}$$

In our project the protein biomarker ER is present in the tumor cell of both bone and liver metastasis. Its membership level is different for bone and liver metastasis. It is allocated membership function of 0.599 and 0.218 for bone and liver metastasis respectively. In the same way fuzzy membership functions have been created for 5 or 6 protein biomarkers.
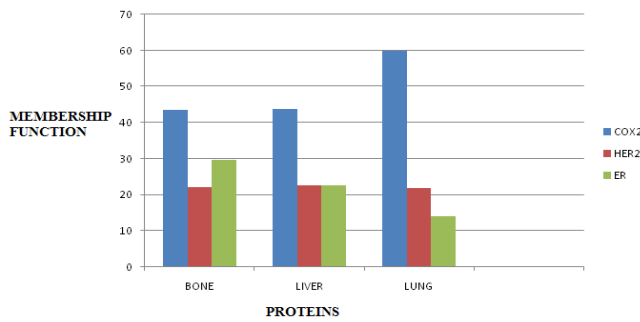
Paper ID: SUB151115

446

Figure 2: Bar graph

3.3. Algorithm

1. Form the Input matrix 't' in MATLAB for cluster heads of different metastasis sites using the clinical data obtained from breast cancer patients.

$$t = \begin{bmatrix} 0.435 & 0.439 & 0.599 & 1 \\ 0.222 & 0.273 & 0.218 & 0 \\ 0.296 & 0.273 & 0.141 & 0 \end{bmatrix}$$

Figure 3: Input matrix

In the 't' matrix the first three columns represent the protein concentration (inputs), last column represents the output. The three columns represent the proteins COX2, HER2 and ER respectively. The three rows correspond to the cluster head for bone, liver and lung respectively. In the example of 't' shown in Fig. the first row's last column has the entry '1' which represents presence of the bone cancer and the zero in the other 2 rows last column indicates absence of the liver and lung metastasis. This matrix is used in the training of the ANFIS model to recognize the bone metastasis in the absence of liver and lung metastasis.

2. The network is trained with the corresponding training matrices for different metastasis. During training the training error is observed and the corresponding changes in the input parameters of the ANFIS model are made to reduce these errors.

3. Once the result is satisfactory, performance of the obtained FIS is tested. The neuro fuzzy model developed is next tested for obtaining the best combination of proteins which predict the required first metastatic site with more accuracy. The test matrix with different combination of presence of proteins created as shown below. The neuro fuzzy structure is trained and tested with matrix 'r' shown in fig 4

$$r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 4: Training matrix

In the above matrix (r) columns corresponds to COX2, HER2 and ER respectively. An entry of '1' indicates the presence of the protein and a '0' its absence. For better prediction the one and zero in the 'r' matrix can be replaced with the actual concentrations of the proteins for each individual patient.

4. By using the command "evalfis" in MATLAB the output obtained for the matrix r shown in Fig 5. is

$$0 = \begin{bmatrix} 0.0632 \\ 0.055 \\ 0.1571 \end{bmatrix}$$

Figure 5: Output matrix

5. Resulting ANFIS structure for bone is shown in Fig 6 This ANFIS structure has five layers which are input, input membership, rules, output membership and output. It uses 27 rules.
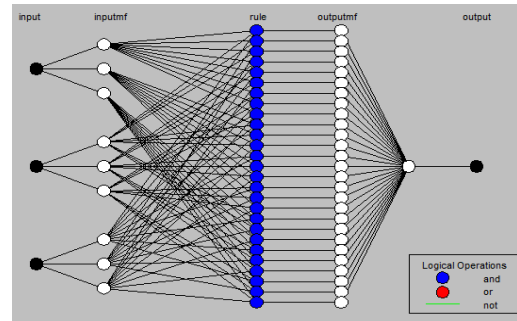

Figure 6: ANFIS Structure

6. Inference
From the vector 'O' the following is inferred. 15.7% of patients with only ER had the first distant metastasis in the bone. 5.5% of patients with only HER2 had bone as the first distant metastasis and 6.32% of Patients with only COX2 had the bone metastasis. Among these three proteins, HER2 has least effect on bone metastasis. Though the ANFIS is optimised to give good results, it is bulky, this can be applied only to limited number of proteins and needs MATLAB command. Thus this computation can be carried out using fuzzy arithmetic

4. Fuzzy Arithmetic

This procedure consists of four steps:

Step 1: The value for proteins corresponding to first distant metastasis are computed by taking ratio of number of patients corresponding to particular organ to number of first metastatic site. These computed values are tabulated in table 2.

Step 2: The value of 'x' is computed for normalization. This 'x' is obtained by taking a ratio of number of breast cancer with protein expression to the number of breast cancers examined. These computed values are shown in table 3

Step 3: Normalization is done by multiplying the table 1 values with 'x' from table 3. The normalized values are shown in table 4.

Step 4: Maximum and minimum values for each column from table 3 are chosen. A lambda value is chosen which is multiplied with maximum value to get a lambda cut set. When the cancer is in second or third stage, the lambda value chosen should be less as the chances of metastasis is more in advanced stages. Lesser the value of lambda results

Paper ID: SUB151115

447

in the presence of more number of proteins causing metastasis and hence more inhibitors are required to immobilize these proteins.

The value of lambda chosen here is 0.33 for which value of lambda cut set is 0.123 which is then compared with the each value in column of table 3. If the corresponding value is greater than lambda cut set level, then this value is defuzzified to 1 else 0. This is tabulated in the table 4.

Table 2: The value of proteins corresponding to first distant metastasis

	BONE	LIVER	LUNG	SKIN	BRAIN	OTHERS
COX2	0.313	0.16	0.2133	0.0733	0.0466	0.1933
HER2	0.318	0.1978	0.1978	0.098	0.01	0.1758
ER	0.4427	0.1614	0.1041	0.088	0.0156	0.1875
PgR	0.4335	0.1678	0.0909	0.0769	0.0069	0.2237
EGFR	0.25	0.1718	0.2343	0.093	0.04687	0.2031
CK5	0.3	0.1	0.233	0.0333	0.1333	0.2
Nestin	0.37	0.111	0.1481	0.074	0.111	0.1851
Prominin1	0.26	0.130	0.1739	0.043	0.1304	0.2608
SMA	0.3076	0.1538	0.1538	0	0.1538	0.2307
SNAI1	0.435	0.14	0.11	0.085	0.025	0.205
SNAI2	0.407	0.1604	0.1604	0.037	0.037	0.1975
CK18	0.3877	0.1802	0.1326	0.074	0.034	0.1904
Ecadherin	0.401	0.1776	0.1312	0.0617	0.0368	0.196
GATA3	0.421	0.1818	0.1	0.0765	0.023	0.196

Table 3: Normalization value 'x' for different proteins

PROTEINS	X
COX2	0.4655
HER2	0.2812
ER	0.612
PgR	0.46
EGFR	0.1836
CK5	0.1014
Nestin	0.0923
Prominin1	0.0833
SMA	0.041
SNAI1	0.0677
SNAI2	0.2657
CK18	0.9684
E- cadherin	0.8957
GATA3	0.6875

Table 4: Normalized protein value

	BONE	LIVER	LUNG	SKIN	BRAIN	OTHERS
COX2	0.1443	0.07448	0.09929	0.034	0.0216	0.0899
HER2	0.089	0.05562	0.05562	0.0028	0.0028	0.04934
ER	0.2709	0.0987	0.0637	0.0538	0.0095	0.1147
PgR	0.1997	0.077	0.0419	0.0354	0.00318	0.1031
EGFR	0.0459	0.0318	0.043	0.017	0.0086	0.037
CK5	0.03	0.01014	0.0236	0.00337	0.0134	0.02028
Nestin	0.0344	0.0102	0.01366	0.0068	0.0102	0.017
Prominin1	0.0216	0.0108	0.1448	0.00358	0.0108	0.0217
SMA	0.0126	0.0063	0.0063	0	0.0063	0.00945
SNAI1	0.294	0.0947	0.0744	0.0575	0.0169	0.1387
SNAI2	0.108	0.0426	0.0426	0.0098	0.0098	0.052
CK18	0.375	0.1745	0.128	0.071	0.0329	0.1843
Ecadherin	0.359	0.159	0.1175	0.055	0.0275	0.1755
GATA3	0.29	0.125	0.0689	0.0527	0.0158	0.135

Table 5: Defuzzification results

	BONE	LIVER	LUNG
COX2	1	0	0
HER2	0	0	0
ER	1	0	0
PgR	1	0	0
EGFR	0	0	0
CK5	0	0	0
Nestin	0	0	0
Prominin1	0	0	1
SMA	0	0	0
SNAI1	1	0	0
SNAI2	0	0	0
CK18	1	1	1
Ecadherin	1	1	1
GATA3	1	1	0

Inference: Proteins CK18 and E-Cadherin are expressed in bone, liver and lung metastasis. Hence a single inhibitor which binds with both these proteins has to be selected to prevent these metastasis. This reduces the number of prescriptions given to an individual patient resulting in a tailor made drug delivery. Protein GATA3 is expressed in both liver and lung metastasis. Proteins such as SNAI1, COX2, ER and PgR are expressed in bone metastasis and prominin-1 is expressed only in lung metastasis. The combination of proteins is represented pictorially by the Venn diagram as shown in Fig. 7. Hence conclusions of the proteins which have an impact on the development of a particular metastasis can be drawn.
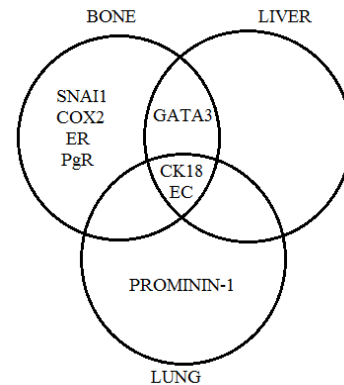

Figure 7: Venn diagram

5. Conclusion

In our paper we developed a prediction model for the first site of breast cancer metastasis. Membership functions for overly expressed proteins are developed. The Trained neural network is tested with different inputs. These results are found to be restricted for limited number of proteins. Hence fuzzy arithmetic is used to get the results for more number of proteins using lambda cut sets. These results were compared to that obtained from the survey paper's inference and were found to be very useful for the practitioner to propose the individual therapy

6. Future work

Clinical trials based on the above results if carried out and the resulting results incorporated in the input data sets used

in the above algorithm, still better prediction models can be developed.

References

[1] "Breast cancer biological subtypes and protein expression predict for the preferential distant metastasis sites " a nationwide cohort study Harri Sihto, Johan Lundin, Mikael Lundin, Tiina Lehtimäki, Ari Ristimäki, Kaija Holli, Liisa Sailas, Vesa Kataja, Taina Turpenniemi-Hujanen, Jorma Isola, Päivi Heikkilä and Heikki Joensuu. Sihto et al. Breast cancer research 2011.

[2] "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications" Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL. Proc Natl Acad Sci 2001, 98:10869-10874.

[3] "Subtypes of breast cancer show preferential site of relapse". Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, Foekens JA, Martens JW. Cancer Res 2008, 68:3108-3114.

[4] "Fuzzy Versus Quantitative Association Rules" Heikki Koivo 2000

[5] "Fuzzy Clustering Technique for Numerical and Categorical dataset" Revati Raman Dewangan, Lokesh Kumar Sharma, Ajaya Kumar Akasapu Dept. of Computer Science and Engg. , CSVTU Bhilai(CG), Rungta College of Engineering and Tech. Bhilai(C.G.),India Dept. of IT and MCA, CSVTU Bhilai(CG), Rungta College of Engineering and Tech. Bhilai(C.G.),India Dept. of Computer Science and Engg. , CSVTU Bhilai(CG), Rungta College of Engineering and Tech. Bhilai(C.G.),India NCICT 2010 Special Issue

[6] "An adaptive resource-allocating network for automated detection, segmentation, and classification of breast cancer nuclei topic area image processing and recognition" Kyoung-Mi Lee , Street, W.N. Dept. of Comput. Sci., Duksung Women"s Univ., Seoul, South Korea Neural Networks, IEEE Transactions on (Volume:14 , Issue: 3) 2003

[7] "A Generalized Fuzzy Clustering Regularization Model With Optimality Tests and Model Complexity Analysis" Jian Yu and Miin-Shen Yang ieee transactions on fuzzy systems, vol. 15, no. 5, october 2007.

[8] "Small Molecule Tyrosine Kinase Inhibitors of ErbB2/HER2/Neu in the Treatment of Aggressive Breast Cancer" Richard L. Schroeder 1, Cheryl L. Stevens 2 and Jayalakshmi Sridhar Molecules 2014,

[9] "Combining Fuzzy Clustering and Fuzzy Inferencing in Information Retrieval" Donald H. Kraj, Jianhua Chen and Andreja Mikulcic Department of Computer Science Louisiana State University Baton Rouge. 2000 IEEE.

[10] "Occurrence of cancer at multiple sites: Towards distinguishing multigenesis from metastasis" Wei-Kang Zhang, Chun Zhang, Jing J Zhang and Shi V Liu. Published in: 11th april 2008 © 2008 Zhang et al; licensee BioMed Central Ltd.

[11] Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein" Aertgeerts, K.; Skene, R.; Yano, J.; Sang, B.C.; Zou, H.; Snell, G.; Jennings, A.; Iwamoto, K.; Habuka, N.; Hirokawa, A.; et al. J. Biol. Chem. 2011, 286, 18756–18765.

Paper ID: SUB151115

449