Int. j. inf. tecnol. (June 2018) 10(2):211–216
https://doi.org/10.1007/s41870-018-0090-7

ORIGINAL RESEARCH

Classification of Protein Kinase B using discrete wavelet transform

Shruti Jain[1]

Received: 16 March 2017 / Accepted: 12 January 2018 / Published online: 24 January 2018

Abstract In this paper a CAD system was designed for the classification of Protein Kinase B (PKB) using ten different discrete wavelet transforms and SSVM and SVM classifier. A set of different images has been collected from which data is divided into training and testing data set. The PKB is categorized into two classes called absent or present. The highest overall classification accuracy of 80% was obtained with biorthogonal: bior 4.4 wavelet transforms and daubechies: db6 wavelet transforms using SSVM classifier.

Keywords CAD design · Protein Kinase B · Discrete wavelet texture transform · Overall classification accuracy

1 Introduction

PKB/Akt used in the cell promotion and impediment of cell death [1–4]. There are three major types of input proteins which lead to cell death/survival: epidermal growth factor (EGF) [5–7] and Insulin [8–13] were the survival proteins while tumor necrosis factor (TNF) [14–17] is the protein which leads cell death. Insulin and EGF are the main proteins which activates PKB. There are mainly three different types of AkT: Akt1, Akt2 and Akt3. Phosphorylation of Thr308 and Ser473 activates AkT1 by PDK1. AkT1 and AkT2 were presumed in signal transduction of insulin in adipose tissue, liver skeleton muscle, while AkT3 is not activated by insulin. Figure 1 shows the different steps of PI3K that leads to cell survival/death [2]. As shown in Fig. 1, there are many proteins which on activation/deactivation with PKB leads to cell death/survival like BAD, p53, NFκB are survival proteins while FKHR is an apoptotic protein. There are different pathway which can lead to cell death/survival using EGF/Insulin taking PI3K, PKB as main components. First pathway is EGF/Insulin → PI3K → PKB → BAD (survival pathway). If any of the protein is absent in this path then final result is cell death. Likewise second pathway: EGF/Insulin → PI3K → PKB → p53 (survival pathway), third pathway: EGF/Insulin → PI3K → PKB → NFkB (survival pathway), fourth pathway: EGF/Insulin → PI3K → PKB → FKHR (death), fifth pathway: Insulin → PI3K → mammalian target of rapamycin (mTOR) → IRS (survival pathway). Fifth pathway also leads to protein synthesis.

A CAD system was designed using different discrete wavelet filters and different classification techniques.

2 Materials and methods

The work has been carried out on images available from Weiss [3]. The data consists of different images of pAkt: phospho-Akt, AkT, ptAkt: phospho-to-total Akt. A CAD system was designed which can be used as a second opinion tool for radiologists. Figure 2 shows the flow chart of the CAD system designed for PKB.

For CAD system different steps were taken into consideration: data collection, pre-processing, feature selection, feature extraction, partitioning and classification.

Firstly the data was collected from Weiss [3]. Analysis was done in such a manner so as to remove the redundant/duplicate data. If the data is numeric/continuous the regression analysis was done but if the data is categorical

---

✉ Shruti Jain
jain.shruti15@gmail.com

[1] Jaypee University of Information and Technology, Solan, Waknaghat, HP, India

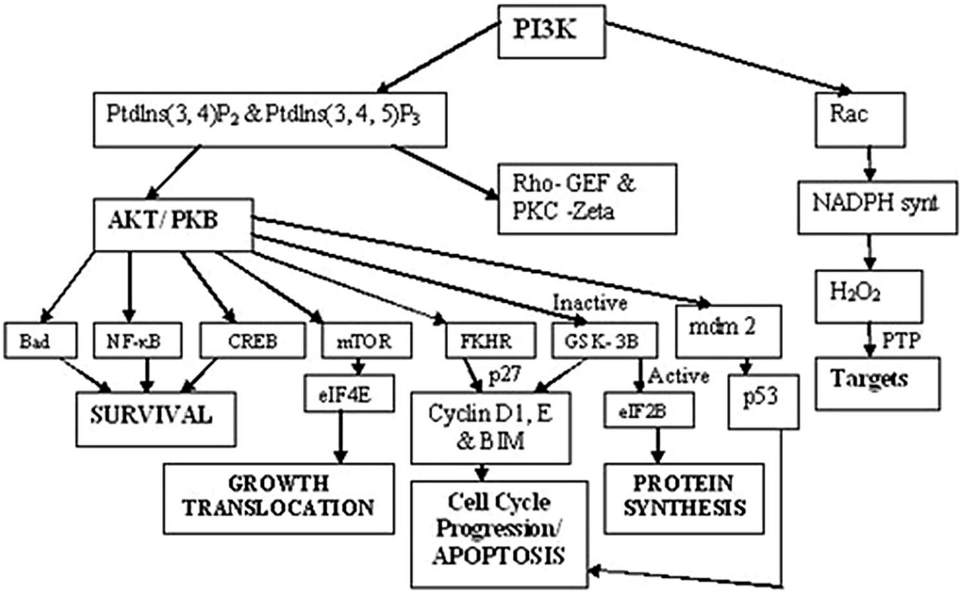

Fig. 1 Illustration of signal communication network triggered by PI3K [2]

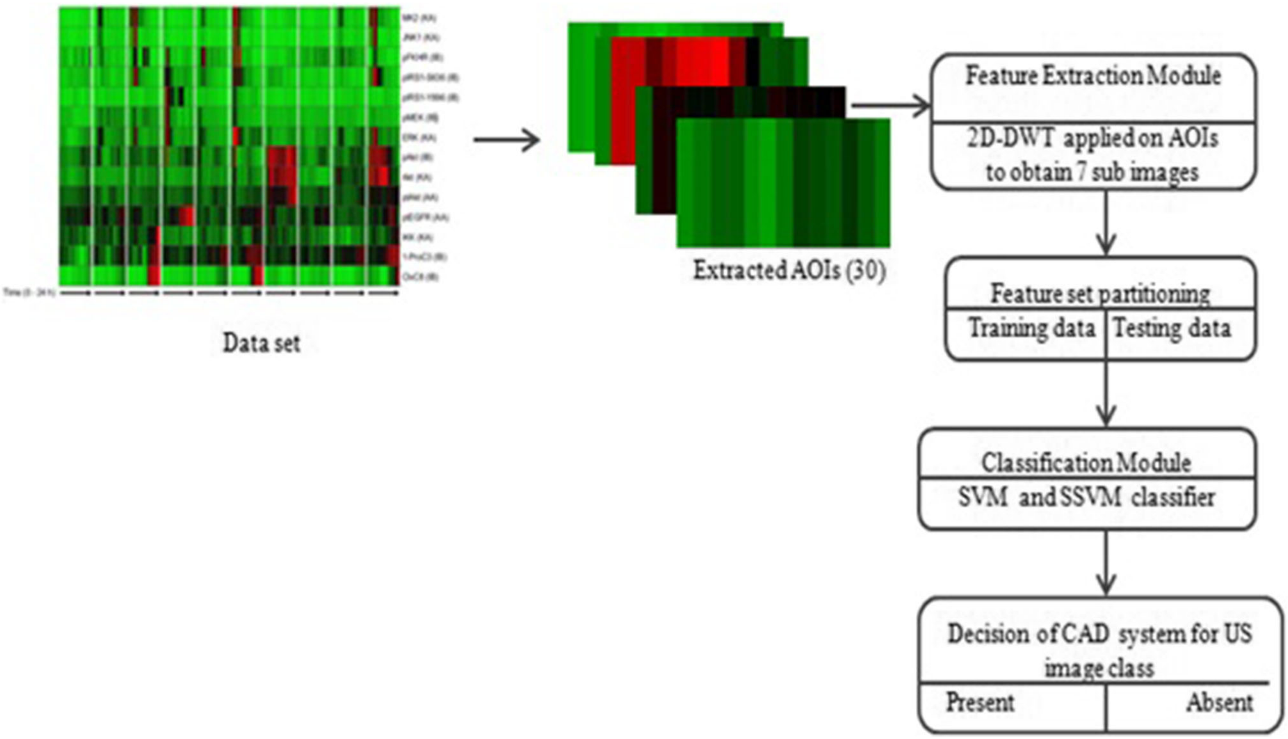

Fig. 2 Proposed CAD system for Protein Kinase B

than Chi square test or different data mining techniques was applied.

The pre processing technique was applied which consists of cleaning, transformation etc. We clean the data by filling out the missing values, remove noisy data, remove outliers and resolve whether there was any inconsistency in the data or fill out the missing values. There are different approaches to remove the noise from data like binning method, clustering or regression. The data can be transformed by normalization (min–max, z-normalization, by decimal scaling, euclidean, forbenius), aggregation or generalization (hierarchy climbing).

Feature extraction module: There are two types of feature extraction techniques: shape based and texture based [18–27] as explained in Fig. 3. In this paper we have used transform domain method which is a part of texture based

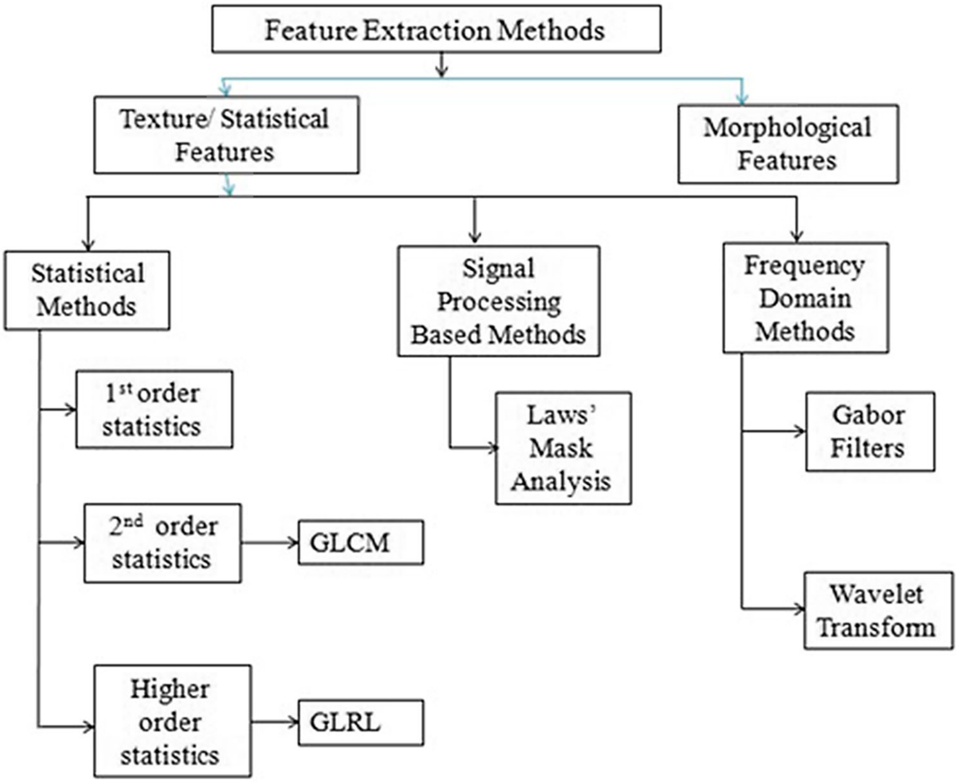

Fig. 3 Block diagram representation of different feature extraction methods

approach. Discrete wavelet transform (DWT) [28], is one of the part of transform domain method which is used in this paper.

Any signal which we have to decompose is first passed through low pass filter (LPF) and high pass filter (HPF). Approximation (A) decomposition is obtained after LPF and Details (D) decomposition is obtained after HPF. Down sampling of columns has been carried out which gives cA and cD decomposition. Later, down sampling of rows were carried out which yields approximation decomposition (CA) and horizontal decomposition (CH) after decomposing cA while vertical decomposition (CV) and diagonal decomposition (CD) after decomposing cD. It means we will get four decomposition values: CA1, CH1, CV1, and CD1. After decomposing CA1 we get again four parts: CA2, CH2, CV2 and CD2 as shown in Fig. 4. There are different types of wavelet transforms. In this paper we are using ten wavelet transforms such as Haar wavelet: db1, Coiflets wavelet: coif1 and coif2, Symlets wavelet: sym3 and sym5, Biorthogonal wavelet: bior3.1, bior3.3 and bior4.4 and Daubechies wavelet: db4 and db6 filters. For each wavelet transforms we have calculated seven texture feature vectors (TFV). Seven TFV's are mean, standard deviation (SD), energy, mean + SD, SD + energy, energy + mean, mean + SD + energy.

$$\text{Mean } \mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}, \tag{2}$$

$$\text{Energy } E = \int_{-\infty}^{\infty} |x(t)|^2 dt, \tag{3}$$

where x_i is the observed value, n is the sample size.

Figure 5 shows the two level decomposition of one image that how we have decomposed all the images using different wavelet transforms. Figure 6 shows the de-noised image. This also represents the horizontal, diagonal and vertical detailed images.

After feature extraction, partitioning of data is done where data is partitioned into testing and training. There were different approaches for partitioning: holdout (single, repeated), cross validation, boot strap, sampling (re-sampling, stratified sampling) and three way data splits.

Last step is Feature classification module: Classification can be done by two processes supervised learning and unsupervised learning. Basically classification helps in finding the numerical properties of the images and arranges those values into different classes. For classification we divide data into two parts: testing and training set. There are different types of classifiers. In this paper we are using SVM and SSVM classifier. For the execution of SVM classifier LibSVM library has been used [29]. For calculating SSVM classifier, a library was added which was

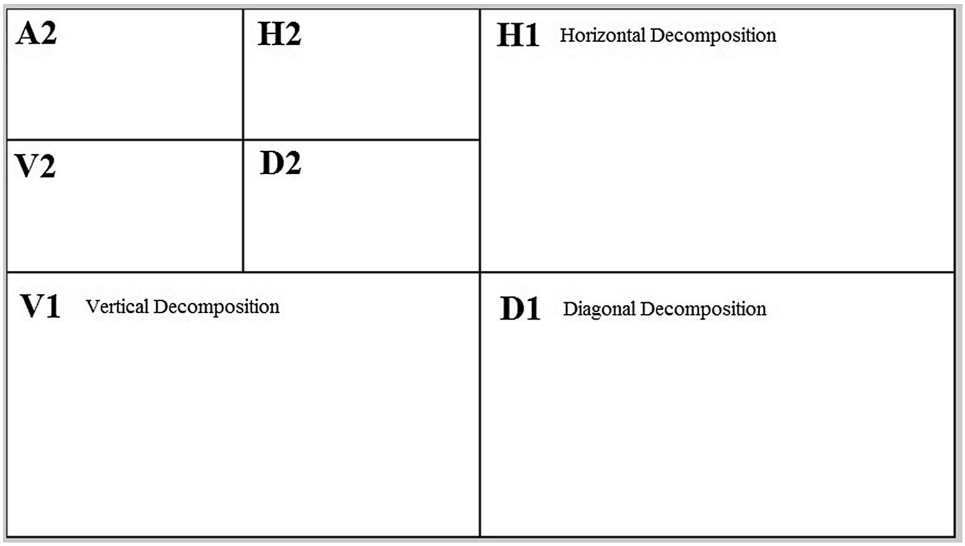

Fig. 4 Two level decomposition

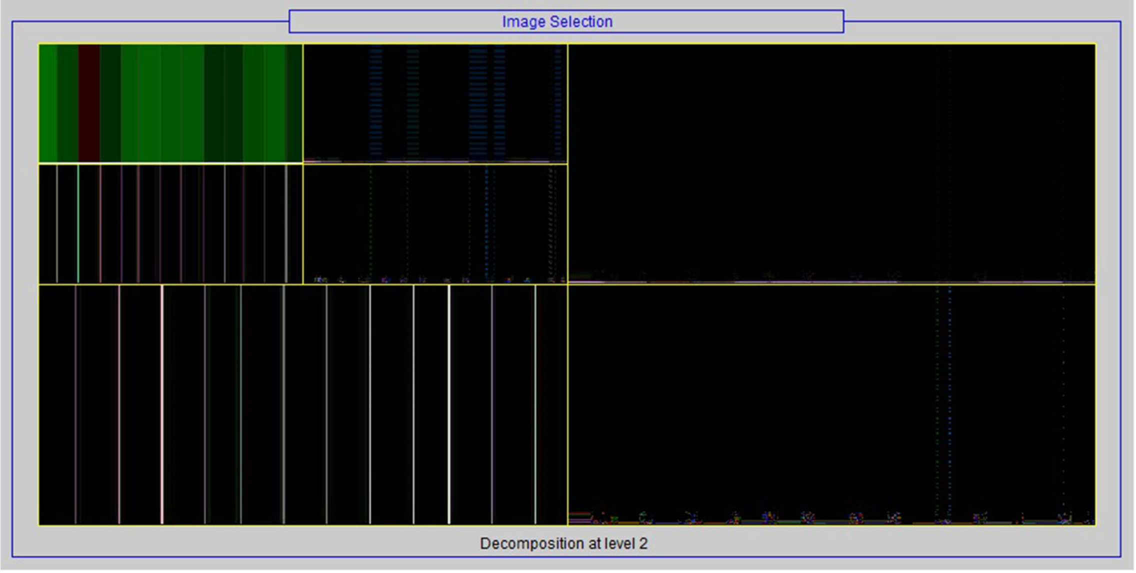

Fig. 5 Two level decomposition of our image

developed by Laboratory of Data Science and Machine Intelligence, Taiwan [30].

3 Results and discussion

A CAD system has been proposed, where different experiments has been carried out to obtain the classification performance which discriminates PKB into cell death/survival using different DWT with SSVM and SVM classifier. The results were depicted in Table 1. For simulation we have used MATLAB 14.7 software. Algorithm for wavelet transform:

Step 1: Initially different images of AkT, pAkt, and ptAkt were collected from Weiss [3]. Each image consists of sub thirteen sections which explains the time from 0 to 24 h in the interval of 0, 5 min, 15 min, 30 min and 1 h, 1 h 30 min, 2 h, 4 h, 8 h, 12 h, 16 h, 20 h, and 24 h.
Step 2: Colored images are converted to gray scale images.
Step 3: Computation of first level decomposition (CA1, CH1, CV1, and CD1) for different wavelet transforms using symmetrical mode.
Step 4: Computation of second level decomposition (CA2, CH2, CV2, CD2) for different wavelet transforms using symmetrical mode.

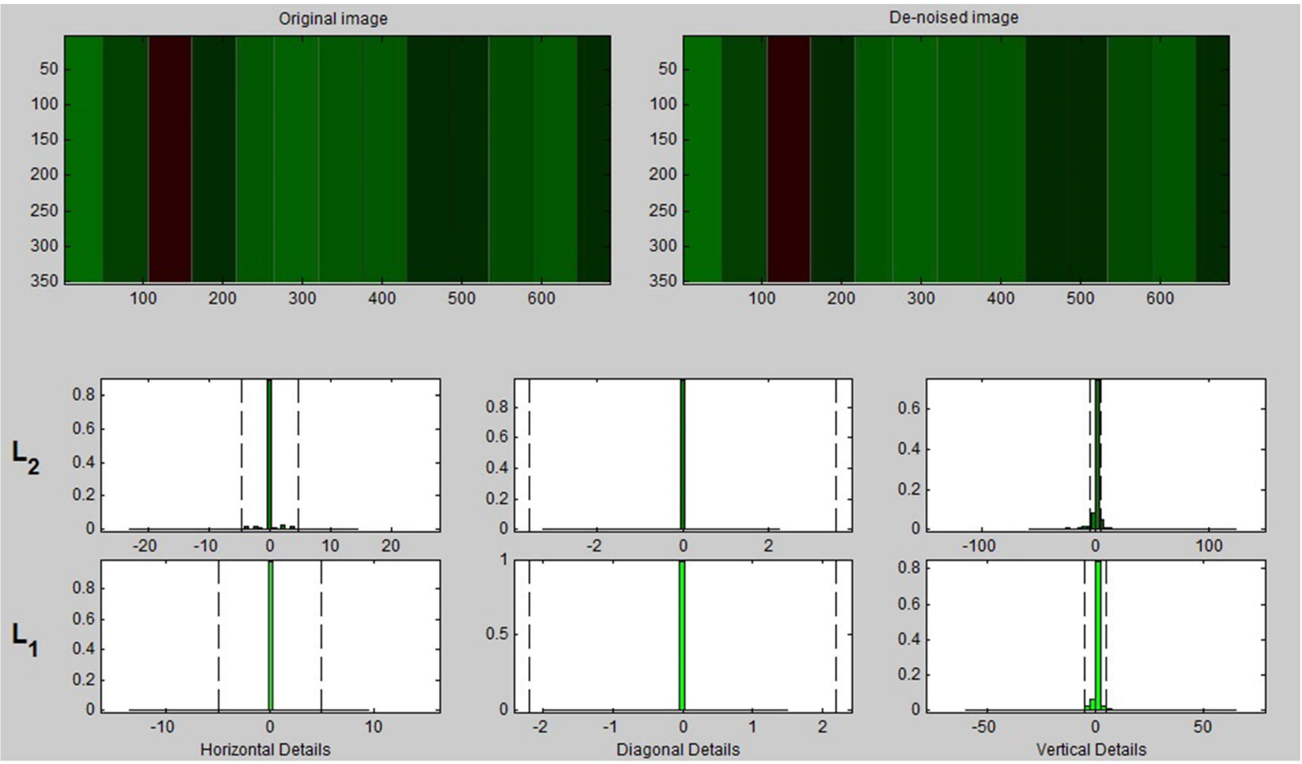

Fig. 6 Denoised Image and its horizontal, vertical and diagonal detailed images

Table 1 Classification performance for the seven TFVs using SSVM/SVM classifier

TFV	Length of TFV	Maximum accuracy (%)	Wavelet filter	Classifier	Minimum accuracy (%)	Wavelet filter	Classifier
TFV1	7	66.67	bior 4.4, bior 3.1, bior 3.3, coif 1, coif 2, db4, db6, sym3, sym5	SSVM	40	bior 4.4	SVM
TFV2	3	73.33	bior 4.4, bior 3.1, bior 3.3, coif 1, coif 2, db4, db6, sym3, sym5	SSVM	53.33	db1, sym5	SVM
TFV3	6	73.33	bior 4.4, bior 3.1, bior 3.3, coif 1, coif 2, db4, db6, sym3, sym5	SSVM	40	sym5	SVM
TFV4	4	73.333	bior 3.3, coif 1, coif 2, db4, db6, sym3, sym5	SSVM	53.33	bior 3.3, sym5	SVM
TFV5	4	66.67	bior 4.4, coif 1, coif 2, db6, sym3, sym5	SSVM	40	db1	SVM
TFV6	4	66.67	bior 4.4, bior 3.3, coif 1, coif 2, db4, db6, sym3	SSVM	40	db1	SVM
TFV7	3	80	Bior 4.4, db6	SSVM	33.33	db1	SVM

Step 5: Different features were calculated for different decomposed levels using forbenius normalization. Forbenius normalization of an m by n matrix X is define as

$$\|X\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{i,j}\right|^2}. \tag{4}$$

The length of the first feature is seven which generates: CH1, CV1, CD1, CA2, CH2, CV2, and CD2, second feature is three: CH1, CV1, CD1, third feature is six: CH1, CV1, CD1, CH2, CV2, and CD2, fourth feature is four: CH1, CV1, CD1, and CD2, fifth feature is four: CH1, CV1, CD1, and CA2, sixth feature is four: CA2, CH2, CV2, and CD2 and seventh feature is three: CH2, CV2, and CD2.

From Table 1, it was observed that maximum accuracy of 80% was obtained using bior 4.4 wavelet transform and db6 wavelet transform for seventh TFV with SSVM classifier. The lowest accuracy (33.33%)was achieved by haar wavelet transform for seventh TFV which consists of length of six features using SVM classifier. A confusion

Table 2 Classification performance for the best TFVs using SSVM classifier

TFV	CM		OCA (%)	Precision
	Present	Absent		
TFV7				
Present	50	20	80	77.78%
Absent	10	70		

matrix (CM) of the best TFV is shown in Table 2. Table also shows overall classification accuracy (OCA) for present and absent using SSVM classifier.

4 Conclusions

This paper explains a CAD system which helps in classification of PKB using different discrete wavelet transform. From different experiments it was concluded that, for characterization of PKB, bior 4.4 & db6 wavelet transforms yields the best results. These wavelet transforms gives the maximum OCA of 80%. No paper till yet reported such type of work. In future author will try for more DWT and compare the results. Furthermore author can also try for more texture features.

References

1. Jain S, Chauhan DS (2015) 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. pp 81–88
2. Jain S (2012) 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
3. Weiss R (2001) Cellular computation and communications using engineered genetic regulatory networks. PhD Dissertation, MIT
4. Libermann TA, Razon TA, Bartal AD et al (1984) Expression of epidermal growth factor receptors in human brain tumors. Cancer Res 44:753–760
5. Normanno N, De Luca A, Bianco C et al (2006) Epidermal growth factor receptor (EGFR) signaling. Cancer Gene 366:2–16
6. Ullrich A, Schlessinger J (1990) Signal transduction by receptors with tyrosine kinase activity. Cell 61:203–211
7. Jain S, Chauhan DS (2015) mathematical analysis of receptors for survival proteins. Int J Pharma Bio Sci 6(3):164–176
8. Lizcano JM, Alessi DR (2002) The insulin signalling pathway. Curr Biol 12:236–238
9. White MF (2003) Insulin signaling in health and disease. Science 302:1710–1711
10. Jain S, Naik PK, Bhooshan SV (2011) Mathematical modeling deciphering balance between cell survival and cell death using insulin. Netw Biol 1(1):46–58
11. Jain S, Naik PK, Bhooshan SV (2010) A system model for cell death/survival using SPICE and ladder logic. Digest J Nanomater Biostruct 5(1):57–66
12. Jain S, Naik PK (2012) System modeling of cell survival and cell death: a deterministic model using Fuzzy System. Int J Pharma BioSci 3(4):358–373
13. Jain S (2015) Mathematical analysis and probability density function of FKHR pathway for cell survival/death. In: Proceedings of the control system and power electronics—CSPE, Bangalore. pp 84–93
14. Gaudet S, Kevin JA, John AG et al (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Manuscript M500158-MCP200, 2005
15. Kevin JA, John AG, Suzanne G et al (2005) A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science 310:1646–1653
16. Thoma B, Grell M, Pfizenmaier K, Scheurich P (1990) Identification of a 60-kD tumor necrosis factor (TNF) receptor as the major signal transducing component in TNF responses. J Exp Med 172:1019–1023
17. Jain S, Naik PK, Bhooshan SV (2011) Mathematical modeling deciphering balance between cell survival and cell death using tumor necrosis factor α. Res J Pharm Biol Chem Sci 2(3):574–583
18. Raja BK, Madheswaran M, Thyagarajah K (2010) Texture pattern analysis of kidney tissues for disorder identification and classification using dominant Gabor wavelet. Mach Vision Appl 21(3):287–300
19. Rana S, Jain S, Virmani J (2016) Classification of kidney Lesions using gabor wavelet texture features. In: Proceeding of the 10th INDIACom 3rd 2016 international conference on computing for sustainable global development. pp 2528–2532
20. Bhusri S, Jain S, Virmani J (2016) Classification of breast Lesions based on Laws' feature extraction techniques. In: Proceeding of the 10th INDIACom 3rd 2016 international conference on computing for sustainable global development. pp 2523–2527
21. Subramanya MB, Kumar V, Mukherjee S, Saini M (2014) SVM-Based CAC system for B-mode kidney ultrasound images. J Digit Imaging Soc Imaging Inf Med 28(4):448–458
22. Bhusri S, Jain S, Virmani J (2016) Breast Lesions classification using the amalagation of morphological and texture features. Int J Pharma BioSci 7(2):617–624
23. Rana S, Jain S, Virmani J (2016) SVM-based characterization of focal kidney Lesions from B-mode ultrasound images. Res J Pharm Biol Chem Sci 7(4):837–846
24. Bhusri S, Jain S, Virmani J (2016) Classification of breast lesions using the difference of statistical features. Res J Pharm Biol Chem Sci 7(4):1365–1372
25. Jain S, Naik PK, Bhooshan SV (2011) Nonlinear modeling of cell survival/death using artificial neural network. 2011. In: The proceedings of international conference on computational intelligence and communication networks, Gwalior, India. pp 565–568
26. Jain S, Naik PK, Bhooshan SV (2010) Petri net implementation of cell signaling for cell death. Int J Pharma Bio Sci 1(2):1–18
27. Virmani J, Kumar V, Kalra N, Khandelwal N (2011) Prediction of cirrhosis from liver ultrasound B-mode images based on Laws' masks analysis. In: The proceedings of IEEE international conference on image information processing, ICIIP-2011, Waknaghat, HP, India. pp 1–5
28. Virmani J, Kumar V, Kalra N, Khandelwal N (2013) SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors. J Digit Imaging 26:530–543
29. LIBSVM (2016) http://www.csie.ntu.edu.tw/~cjlin/libsvm. Accessed 15 Jan 2016
30. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2:1–43