Classification of RR-Interval and Blood Pressure for Different Postures using KNN Algorithm

Indu Saini, Dilbag Singh and Arun Khosla

Dr B R Ambedkar National Institute of Technology Jalandhar,
Jalandhar - 144011, India
indu.saini1@gmail.com, drdilbag@gmail.com, khoslaak@nitj.ac.in

Abstract

Power spectral analysis of the heart rate and blood pressure variations has commonly used to provide indices of autonomic cardiovascular modulation, but the effect of changing posture from lying to standing on these variations and the interpretation of their power spectra is still largely in dispute. It was due to the reason that till now no study was made yet that clearly outlines the variations in terms of RR-interval and blood pressure series from lying to standing position. Thus the aim of this paper lies in the application of classifying the subjects based on their RR-intervals, systolic and diastolic blood pressure series, prior to spectral analysis, at two different physical activity related postures. In this paper K-Nearest Neighbor algorithm has been proposed as a classifier for classifying the subjects based on lying and standing postures. Here we also studied the classification accuracy achievable with a KNN classifier using three different methods (i) Euclidean (ii) City block and (iii) Correlation of calculating the nearest distance in order to propose the optimal one. Further an attempt has been made to evaluate each of these methods for five different values of K=1, 3, 5, 7 and 9 in order to propose the best fit value of K for classifying the subjects. After performing the comparative analysis between these three methods of distance metrics and for different values of K, it is found that K=1 is the best choice out of 3, 5, 7 and 9 and Correlation has been emerged as one of the optimal method for computing the nearest distance with highest classification accuracy of 98.60 % with K=1 for lying and 99.95 % with K=1 for standing postures.

Keywords: *Classifier, Nearest Neighbor, Distance Metrics, Postures*

1. Introduction

In recent years growing number of reports has shown that the analysis of cardiovascular variability is a valid non-invasive tool capable of providing adequate information on autonomic modulation of the sinoatrial node in normal subjects and in patients with variety of cardiac and non-cardiac diseases. The period of heart beat is not constant and changes over the time. These variations in heart rates and respective periods are called heart rate variability (HRV) and respective beat-to-beat fluctuations of blood pressure, is known as blood pressure variability (BPV) [1].

The HRV, for most part, is the reflection of underlying BPV operating by the way of baroreflex. If the causal BP variations are not taken into considerations, conclusion based on HRV alone may be spurious. Therefore the BPV and its interaction with HRV have also been taken into account for understanding cardiovascular modulation [2]. Further, although the analysis of variability in heart rate (HR) and blood pressure (BP) was proven to be useful in understanding the cardiovascular regulation, of the subjects rested in lying position, in a range of conditions including heart failure, hypertension, diabetes etc [3]. But the interpretation of their power spectra when any subject is under some physical activity is still not totally resolved [4].

It is generally believed that in the lying position the parasympathetic modulation is dominant; and causes stronger high frequency fluctuations. In contrast, decreased parasympathetic function occurs in standing position [4]. But this fact is not well established, as many times contradictory results have left the clinical researchers skeptical about the prognostic or diagnostic information provided by the low and high frequency fluctuations with variable postures and there exists no clear consensus on how to estimate HRV and BPV in daily physical activities [5]. A number of studies have been performed to investigate the differential modulation of sympathetic and parasympathetic nervous system outflow during postural change by estimating HRV using various methods [2, 5-8] but no clear cut understanding has been made yet. This was due to the reason that it is not so far clear that how the RR-interval series and blood pressure vary in response to the postures and whether there is any significant change in terms of these variations with posture that has a direct impact on the HRV and BPV characteristics. Thus without the thorough understanding of this concept prior to spectral analysis it is not possible to completely decipher the hidden dynamics of cardiovascular signal. In the earlier studies the researchers used the same analysis technique whether linear or nonlinear to explore these changed dynamics for all the subjects participated in the study. But it is not true that RR-intervals and BP series vary in the same sense for all the subjects and reflects the changes in Power Spectral Analysis (PSA) in the same way. Thus firstly the changes in terms of RR-intervals and BP series are classified according to postures and then select as per the classification accuracy the analysis technique. Thus an effort is made to speculate how fast the RR-interval series and BP series changes in response to different postures by observing their changed dynamics using a KNN classifier.

The KNN classifier is a simple and popular classification algorithm that has so far not received much attention in the domain of biomedical signals. The idea underlying the KNN method is to assign new unclassified examples to the class to which the majority of its k nearest neighbors belongs [9]. One advantage of the KNN method over many other supervised learning methods like support vector machine, decision tree, neural network etc [9] is that it can easily deal with problems in which the number of classes is bigger than two. Furthermore, the KNN method allows adding examples to the training dataset without retraining the classifier. Clearly, the ability to deal with multiple classes as well as the ability to update classifiers online is important for biomedical applications. Moreover, in the KNN based classification model, the good selection of K and the type of distance metrics used computing nearest distance has a bigger effect on categorization results. Their choice becomes more critical when KNN classifier is used in biomedical applications prior to signal processing [10-11]. After an exhaustive literature survey on KNN based classification problems it has been observed that there is an inconsistency in the selection of the value of k and distance metric used for classification.. Unfortunately, an adequate justification about which value of K and which method for computing the nearest distance is most suitable for any KNN based application is usually lacking and most of the studies were silent on this approach. Thus most importantly there is an urgent need to optimize the value of K and type of distance metric used in KNN algorithm. In this paper, in order to find the optimal value of K and distance metric a comparative analysis has been performed between K=1, 3, 5, 7, and 9 and between Euclidean (EU), City block (CB) and Correlation (CO) distance metrics. Such comparative performance evaluation has not been yet made in the literature and no such systematic experiments were published earlier which could infer that how and in what sense the different values of K and distance metrics affects the classification accuracy of KNN classifier. In this work our main stress is on the application of KNN algorithm and on exploring the optimal value of K and distance metric for the obtaining the highest accuracy in classifying the subjects based upon their postures but not on selecting the analysis technique as per its classification accuracy. Moreover, the study which we made here is to analyze the combined effect of HR and BP fluctuations and hence allows a wider insight into the role of varying postures in cardiovascular control, in classifying the subjects.

2. Methodology

The study was performed on ten sets of lying and ten sets of standing postures of RR-interval and BP obtained from standard Eurobavar database available on internet (http://www.cbidongnocchi.it/glossary/eurobavar.html) (subject no. 1 to 10) and ECG data of standard lead-II, recorded in our own laboratory, using BIOPAC MP100 system.

2.1 Overview of K-Nearest Neighbors Method

The KNN is one kind of prospective statistical classification algorithms (Supervised Learning). It is a method for classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification [12].

The KNN algorithm consists of two phases: Training phase and Classification phase. The training examples are vectors in a multidimensional feature space, each with a class label and the training phase of the algorithm consists only of storing the feature vectors (library of reference vectors) and class labels of the training samples. In the classification phase, K is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the K training samples nearest to that query point. In other words the KNN method compares the query point or an input feature vector with a library of reference vectors and the query point is identified to be of the class of library feature vector to which it has the closest distance. This way of categorizing query points based on their distance to points in a training data set is a simple yet effective way of classifying new points [13].

In addition, the accuracy of the KNN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. Thus for an efficient KNN based classification the prerequisites are (i) the input data should be noise free and (ii) the selected feature space should be coherent. In this regard much research effort has been put into, in selecting the ectopic free ECG records and in selecting or scaling the most appropriate features to improve classification. In this paper we select three features for classifying the subjects based upon their postures: RR-interval series, Systolic blood pressure and diastolic blood pressure.

2.2 Selection of Parameter K and Distance Metric

The best choice of K depends upon the data; generally, larger values of K reduce the effect of noise on the classification, but make boundaries between classes less distinct. The special case where the class is predicted to be the class of the closest training sample (i.e. when K = 1) is called the nearest neighbor algorithm. In binary (two class) classification problems, it is helpful to choose K to be an odd number as this avoids tied votes. Thus, the value of K is defined in such a way that it produces the highest correct classification rate [14].

Further the classification accuracy of KNN is also largely dependent upon the distance metrics used for computing the nearest distance. Thus in this paper in order to find the optimal distance metric variety of methods which has been used are: Euclidean distance, City Block and correlation.

2.3 Distance Metrics

Given an *mx*-by-*n* data matrix *X*, which is treated as *mx* (1-by-*n*) row vectors x_1, x_2, ..., x_{mx}, and *my*-by-*n* data matrix *Y*, which is treated as *my* (1-by-*n*) row vectors y_1, y_2, ...,y_{my}, the various distances between the vector x_s and y_t are defined as follows:

Euclidean (EU) distance metric: This is the most usual, "natural" and intuitive way of computing a distance between two objects. It examines the root of square differences between coordinates of a pair of objects and is defined using equation (1) [13, 14, 15]

$$d_{st} = \sqrt{\sum_{j=1}^{n} \left(x_{sj} - y_{tj}\right)^2} \tag{1}$$

City Block (CB) distance metric: It is based on Taxicab geometry, first considered by Hermann Minkowski in the 19th century, is a form of geometry in which the usual metric of Euclidean geometry is replaced by a new metric in which the distance between two points is the sum of the absolute differences of their coordinates defined using equation (2)

$$d_{st} = \sum_{j=1}^{n} \left| x_{sj} - y_{tj} \right| \tag{2}$$

The CB distance is also known as Manhattan distance, boxcar distance, absolute value distance. It represents distance between points in a city road grid. While the EU corresponds to the length of the shortest path between two points (i.e. "as the crow flies"), the CB distance refers to the sum of distances along each dimension (i.e. "walking round the block").

Correlation (CO) distance metric: It is one minus the sample correlation between points (treated as sequences of values) and is defined using equation (3)

$$d_{st} = 1 - \frac{\left(x_s - \overline{x_s}\right)\left(y_t - \overline{y_t}\right)'}{\sqrt{\left(x_s - \overline{x_s}\right)\left(x_s - \overline{x_s}\right)'}\sqrt{\left(y_t - \overline{y_t}\right)\left(y_t - \overline{y_t}\right)'}} \tag{3}$$

where

$$\overline{x_s} = \frac{1}{n}\sum_j x_{sj} \text{ and } \overline{y_t} = \frac{1}{n}\sum_j y_{sj}$$

3. Results and Discussion

The aim of the study which we performed here is two folds: firstly to visualize the effect of varying postures on heart rate and blood pressure series using a KNN classifier and secondly to propose the optimal value of K and type of distance metric for achieving the highest possible classification accuracy. In this paper a comparative analysis has been performed between different values of K and between different distance metrics by selecting (i) RR-interval (ii) Systolic blood pressure and (iii) Diastolic blood pressure records of 15 subjects under lying and standing positions as features for the KNN classifier. Here we first trained the KNN classier by presenting 10 lying and 10 standing data files of 10 subjects, each consisting of three types of features with a feature vector space of 17000 samples each. In the classification phase 5 testing data files of the subjects from I to V under lying and standing postures, different from the training phase, were given to the KNN classifier in a random manner. The results obtained in terms of percentage classification accuracy, for K=1, 3, 5, 7 and 9 using EU, CB, and CO methods of computing the nearest distance, for the subjects in lying and standing postures are shown in Table 1 and 2 respectively. After verifying the results shown in table 1 for lying posture, it is found that the averaged classification accuracy obtained using CO distance metric for K=1 is found to be 98.60 % in comparison to EU (94.71%) and CB (94.34%) methods. The superiority of CO method is established over the other two distance metrics in terms of achieving the highest classification accuracy for all the five subjects. Similarly the averaged classification rate obtained with CO method is highest i.e. 98.02%, 97.40%, 97.84%, 97.84% for K=3, 5, 7 and 9 respectively than EU and CB methods for lying posture. Further, the superiority of the CO method is demonstrated graphically by plotting the averaged classification accuracy obtained using EU, CB and CO distance metrics for K=1,3,5,7 and 9 in Figure 1. After the visual analysis it is found that the averaged classification accuracy obtained using CO distance metric, shown by a line with triangles on it in Figure 1, is better than other two metrics, showing by a line with circles on it (obtained using EU distance metric) and by a line with squares on it (obtained using CB distance metric) respectively.

Table 1. Comparative Analysis between k = 1, 3, 5, 7 and 9 and between EU, CB and CO Distance Metrics for the Subjects under Lying Posture

Subject No.	K=1			K=3			K=5			K=7			K=9		
	EU	CB	CO	EU	CB	CO	EU	CB	CO	EU	CB	CO	EU	CB	CO
I	99.54	99.34	99.71	99.28	99.14	99.57	99.43	99.28	99.65	99.28	99.28	99.62	99.28	99.28	99.66
II	99.78	99.08	99.84	98.40	98.39	99.05	99.62	98.42	98.80	98.82	98.56	99.81	99.72	98.76	99.79
III	74.26	73.56	93.47	73.30	73.21	92.69	74.24	74.08	90.78	75.30	75.30	91.82	75.14	76.80	91.65
IV	100	100	100	97.3	98.69	99.90	97.39	98.20	98.76	97.22	98.30	99.12	97.15	98.30	99.45
V	100	100	100	96.20	96.56	98.91	94.75	95.29	99.02	94.75	95.66	98.84	94.93	95.11	98.66
AVG.	**94.71**	**94.34**	**98.60**	**92.89**	**93.19**	**98.02**	**93.08**	**93.05**	**97.40**	**93.07**	**93.42**	**97.84**	**93.24**	**93.65**	**97.84**

EU: Eucledian distance metric, CB: City block distance metric, CO: Correlation distance metric

Table 2. Comparative Analysis between k = 1, 3, 5, 7 and 9 and between EU, CB and CO Distance Metrics for the Subjects under Standing Posture

Subject No.	K=1			K=3			K=5			K=7			K=9		
	EU	CB	CO	EU	CB	CO	EU	CB	CO	EU	CB	CO	EU	CB	CO
I	95.65	95.97	100	95.56	95.79	99.77	95.45	95.79	100	94.19	94.77	99.77	94.54	94.88	99.88
II	99.23	99.56	100	99.19	99.66	99.82	99.09	100	100	100	100	100	99.20	99.55	100
III	99.85	99.83	99.98	99.57	99.72	99.88	99.72	99.72	99.89	99.72	99.57	99.88	99.72	99.58	99.80
IV	100	100	100	99.73	99.73	99.96	99.73	99.73	100	99.73	99.73	99.96	99.73	99.73	100
V	100	100	100	91.17	90.87	95.97	89.22	89.07	95.12	88.92	87.87	96.08	87.87	87.57	96.23
AVG.	**98.94**	**99.07**	**99.99**	**97.04**	**97.15**	**99.08**	**96.64**	**96.86**	**99.00**	**96.51**	**96.38**	**99.13**	**96.21**	**96.26**	**99.18**

EU: Eucledian distance metric, CB: City block distance metric, CO: Correlation distance metric

Further the results showing the classification rate for the subjects under standing posture are given in Table 2 for five values of K and three distance metrics. These results demonstrate that the classification accuracy obtained for K=1 using CO as distance metrics is better than that obtained for K=3, 5, 7 and 9 using EU and CB as distance metrics. In addition, this optimality of the value of K=1 and CO as distance metric is clearly outlined by observing a line with triangles on it in Figure 2. Moreover, this study is also performed on the ECG records obtained in our own laboratory settings from 10 healthy volunteers, and the same trends in the observations are obtained.

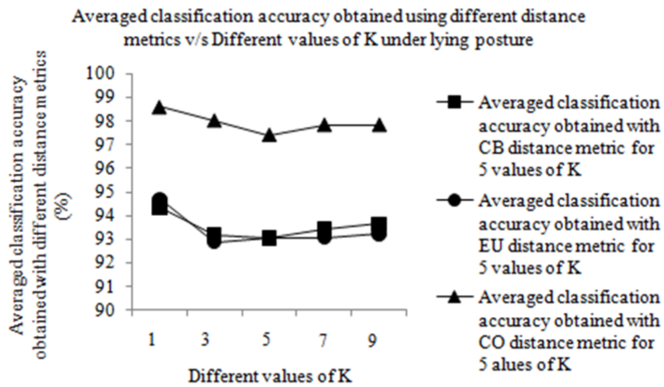

Figure 1. Plot of Averaged Classification Accuracy Obtained using Euclidean (EU), City Block (CB), Correlation (CO) Distance Metrics for 5 Different Values of K for the Subjects under Lying Posture

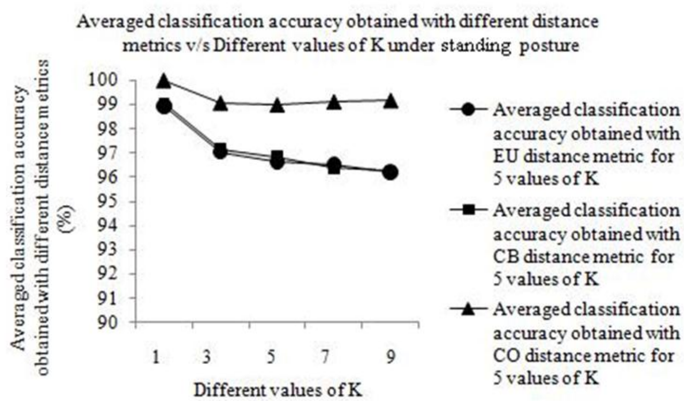

Figure 2. Plot of Averaged Classification Accuracy Obtained using Euclidean (EU), City Block (CB), Correlation (CO) Distance Metrics for 5 Different Values of K for the Subjects under Standing Posture

The purpose of the study which we performed here is to present an accurate method for classifying the subjects based upon their postures using RR-intervals, systolic blood pressure and diastolic blood pressure as classification parameters, for obtaining better quantification of autonomic nervous system. It is found that KNN algorithm for K=1 with CO as distance metric has performed well in classifying fast autonomic nervous system adaptations in changing postures in comparison to other variants of K and distance metrics.

4. Conclusion

The present paper lays considerable emphasis on (1) the algorithmic considerations of KNN classifier for 5 different values of K and three different methods of calculating the nearest distance (2) their methodological aspects related to signal processing and (3) their capabilities of classifying subjects based upon their using RR-intervals and blood pressure series as features leading to the ability to precisely discriminate between pathologies and efficient spectral analysis. Although there are several postures and physical activities in daily life, but we have used only lying and standing postures as activity indicators for heart rate and blood pressure series. In this paper KNN algorithm has been used for understanding the role of RR-interval, systolic and diastolic blood pressure in lying and standing postures. Further, for the accurate classification of the subjects based upon the postures by using KNN classifier an optimal value of K=1 and correlation as a distance metric has been proposed in this paper. The efficacy of the proposed value of K and distance metric has been demonstrated by performing the comparative analysis between (i) K=1, 3, 5, 7 and 9 and between Euclidian, City block and Correlation distance metrics. The proposed value of K and distance metric for KNN based classification based upon postural changes prior to spectral analysis could be the start for classification rate based analysis and also aimed at the prediction of several of episodes of heart rate and blood pressure series. In future more rigorous investigations are needed to study the role of postures in autonomic nervous system modulation.

References

[1] G. G. Bertson, J. T. Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone and M. W. Vann Der Molen, “Heart rate variability: Origins, methods and interpretive caveats,” *Psychophysiology*, vol. 34, **(1997)**, pp. 623-648.

[2] Deboer, R.W., Karemaker, J.M., and Strackee, J., “Relationship between Short-Term Blood-Pressure Fluctuations and Heart Rate Variability in Resting Subjects 1: A Spectral Analysis Approach,” *Medical and Biological Engineering and Computing*, vol. 23, **(1985)**, pp. 352-358.

[3] Task Force of the European Society of Cardiology and North American Society of Pacing and Electrophysiology, “Heart rate variability – standards of measurement, physiological interpretation and clinical use,” *European Heart Journal*, vol. 17, **(1996)**, pp. 354-381.

[4] H. L. Chan, C. H. Lin, and Y. L. Ko, “Segmentation of HRV in different physical activities,” *IEEE Conference 2003 on Computers in Cardiology*, vol. 30, **(2006)**, pp. 133-139.

[5] Rajendra Acharya U., N. Kannathal, Lee Mei Hua, and Leong Mei Yi, “Study of heart rate variability signals at sitting and lying postures,” *Journal of Bodywork and Movement Therapies*, vol. 9, **(2005)**, pp. 134-141.

[6] Hsiao-Lung Chan, Shih-Chin Fang, Yu-Lin Ko, Ming-An Lin, Hui-Hsun Huang, and Chun-Hsien Lin, “Heart rate variability characterization in daily physical activities using wavelet analysis and multilayer fuzzy activity clustering,” *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, **(2006)** January, pp.133-139.

[7] Allan Kardec Barros and Noboru Ohnishi, “Heart Instantaneous Frequency (HIF): An Alternative Approach to Extract Heart Rate Variability,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 8, **(2001)** August, pp. 850-855.

[8] Younhee Choi, Seok-Bum Ko, and Ying Sun, “Effect of postural changes on baroreflex sensitivity: A study on the EUROBAVAR data set,” *IEEE CCECE/CCGEL*, Ottawa, **(2006)** May, pp. 110-114.

[9] Ashkan Yazdani, Touradj Ebrahimi, Ulrich Hoffmann, “Classification of EEG Signals using Dempster Shafer Theory and a K-Nearest Neighbor Classifier,” *Proceedings of the 4^{th} International IEEE EMBS Conference on Neural engineering, Antalya, Turkey*, **(2009)** April 29-May 2, pp. 327-330.

[10] Gang Zheng and Guochao Cao, “A modified K-NN algorithm for holter waveform classification based on kernel function,” *IEEE Computer Society-Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, **(2008)**, pp. 343-346.

[11] Jawad H Alkhateeb, Fouad Khelifi, Jianmin Jiang, Stan S Ipson, “A New Approach for Off-Line Handwritten Arabic Word Recognition using KNN Classifier,” *IEEE International Conference on Signal and Image Processing Applications,* **(2009)**, pp.191-194.

[12] Qingmiao Wang and Shiguang Ju, “A mixed classifier based on combination of HMM and KNN,” *Fourth International Conference on Natural Computation*, **(2008)** October 18-20, pp. 38-42.

[13] http://en.wikipedia.org/wiki/k-nearest_neighbor_algorithm

[14] Tsang-Long Pao, Yun-Maw Cheng, Jun-Heng Yeh, Yu-Te Chen, Chen-Yu Pai, and Yao-Wei Tsai, "Comparison between weighted D-KNN and other classifiers for Music Emotion Recognition," *The 3rd International Conference on Innovative Computing Information and Control (ICICIC'08),* **(2008)**.

[15] S Karimifard, A. Ahmadian, M. Khoshnevisan, M. S. Nambakhsh, "Morphological heart arrhythmia detection using Hermitian basis functions and KNN classifier," *Proceedings of the 28th IEEE EMBS Annual International Conference*, New York City, USA, **(2006)** August30-Sept. 3, pp.1367-1370.

Authors

Indu Saini was born in Ferozepur, India, in 1971. She received her B.Tech degree in Electronics & Communication Engineering from Guru Nanak Dev University, India in 1994 and then obtained her M.Tech(by Research) degree in Electronics & Communication Engineering from National Institute of Technology Jalandhar. Presently, she is pursuing her Ph.D. in the area of Biomedical Signal Processing from Dr B R Ambedkar National Institute of Technology Jalandhar, where she is also serving as Assistant Professor in Electronics & Communication Engineering Department since 2002.

Dilbag Singh was born in Adyana (Panipat), on February 26, 1969. He received the B.E. (Hons.) degree in Electrical Engg from Punjab Engineering College, Chandigarh, and M.E. degree in Control & Guidance from the University of Roorkee, and the Ph.D. degree in Engineering from the Indian Institute of Technology Roorkee in 1991, 1993, and 2004 respectively. The Ph.D. degree thesis was developed at the Instrumentation & Signal Processing Laboratory of the Electrical Engineering Department under the direction of Vinod Kumar, S. C. Saxena and K. K. Deepak (AIIMS). After a brief stint at Goodyear India Limited, Faridabad, in September 1994 he joined the Department of Instrumentation and Control Engineering at of the Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, and since 2005 has been Associate Professor of Instrumentation & Control Engineering there, where he is teaching UG/PG courses related to biomedical signal processing and instrumentation. His professional research interests are in signal processing, in particular applied to biomedical applications.

Arun Khosla received his PhD degree from Indraprastha University, Delhi in the field of Information Technology. He is presently working as Associate Professor in the Department of Electronics and Communication EngineeringNational Institute of Technology, Jalandhar. India. His areas of interest are fuzzy modeling, biologically inspired computing and high performance computing and their applications to engineering and business. He is a reviewer for various IEEE and other National and International conferences and journals. He also serves on the editorial board of International Journal of Swarm Intelligence Research. He has conducted a number of tutorials in the domain of soft computing at various national & international conferences.