

# Machine Learning: An Effective Technique in Bio-Medical Signal Analysis and Classification

Mihir Narayan Mohanty<sup>1</sup> and Hemanta Kumar Palo<sup>2</sup>

<sup>1,2</sup>Department of Electronics and Communication Engineering, ITER, Siksha 'O' Anusandhan  
(Deemed to be University) Khandagiri, Bhubaneswar, Odisha, India  
researchmihir16@gmail.com, hemantapalo@soauniversity.ac.in

## Abstract

*Advancement in the field of digital signal processing and modern machine learning (ML) approaches has witnessed substantial growth in biomedical engineering. The diagnostic power of these machines has grown manifolds mainly due to the exploration of effective and discriminate feature spaces that remain crucial for pattern recognition. It has enhanced the ability of machine learners to model the complex patterns accurately and make them adaptable to new task domains with explanation/experience learning approaches. Many vivid application domains including the artificial intelligent systems and robotics with critical and innovative thinking are going to rely on effective ML systems for efficiency and optimization. This has made the Artificial Neural Networks (NN) an emerging field of research and motivates the authors to classify the MIT-BIH arrhythmia data as abnormal or normal using different ANN models. Finally, the results have been validated with that of the colon cancer gene data.*

DOI: <https://doi.org/10.30991/IJMLNCE.2017v01i01.001>

## Keywords

*Bio-Medical Signal, Machine Learning Algorithm, Classification, Support Vector Machine, Wavelet Transform, Neural Network.*

## 1. Introduction

As a subfield of advanced computer engineering, the ML must learn the desired patterns and need not require to be programmed explicitly. It has found to be successful in infeasible and difficult computing tasks where effective design and explicit programming are essential ingredients of the systems to be efficiently performing. It relies on either related or overlapped computational statistics that emphasizes on prediction-decision making using computers. The efficiency of any ML system depends on robust mathematical algorithms, and optimization techniques that form the general basis for its theory and application framework. Occasionally it misidentified as data mining although the latter emphasizes primarily on exploratory data analysis and is considered unsupervised learning [1-3].

Based on the system feedback or the learning signal and its nature, an ML system can be categorized into supervised, unsupervised and Reinforcement learning [1].A brief on these learning techniques has been explained below.

- 1) *Supervised learning*: In this type of learning, a teacher is assigned to feed the computers with input examples and the corresponding outputs. The aim is to map the inputs to the designated outputs with the help of some general rule [4].
- 2) *Unsupervised learning*: In this case the learning algorithm is not assigned with any labels thus; the structure has to be decided by the network itself. It aims to set its own goal such as choice of the hidden patterns by itself and achieves its end result using the concept known as the feature learning.
- 3) *Reinforcement learning*: The learning method interacts with a dynamic environment to achieve the desired goal. These environments can be playing games or driving a vehicle. Correspondingly, it provides the requisite feedback such as the rewards or the punishments during its navigations in the problem domain.

Semi-supervised learning exists between the unsupervised and supervised algorithm in which the supervisor feeds the incomplete details of the training data and often with some or many missing target data. In this case, sometimes the whole problem sets are available during the learning time using a principle known as Transduction. However some target parts may not be available.

The ML task domain may be summarized as deep learning (DL) or shallow learning. While the DL has many hidden layers and many hidden neurons in each hidden layers, the shallow learning has a single a single hidden layer and is computationally [3, 5]. Classification, of patterns desires the division of inputs among different classes and assignment of the unknown features into the desired class using ML approach. Mostly the approach followed in supervised in nature. One such example of classification is Spam filtering in which the received e-mails are divided into either spam or not spam classes. Similar to classification, another supervised approach is the regression analysis where in the outputs are assumed continuous in nature.

Most of the unsupervised task domain follows clustering approach. In this case, different unknown groups are formed and described by a cluster centre. A particular feature is assigned to a particular group based on some distance measure and its closeness to the corresponding cluster centre. Density estimation in another approach in ML which distributes the input features in some designated space whereas a mapping of high-dimensional space into a low-dimensional space is general approach followed in feature reduction or selection methods. In this regard the Topic modeling ensures a program appearing in a specific list of human language documents and aims to figure out the documents that describe similar topics [4].

The major objective of ML algorithms is the error minimization corresponding to a set of unknown data which significantly defers from that of classical statistical approaches. It aims to quantify complex algorithms and keeps a limit on the upper and lower error bounds [2]. Judicious ML application can deliver a significant value to a host of organization by presenting unknown and hidden knowledge available in the data stream. The work has been accomplished using the two phases. In first phase, a supervised approach to generate the desire disease model using electrocardiogram (ECG) data [6] has been performed. In the next phase, the initial model has been verified for the colon cancer data using the chosen classifiers.

## 2. Support Vector Machine (SVM) Classification Model

Many ML algorithms have been successfully employed for diagnosis of disease, its detection and classification detection [7-12]. Among these, the Decision tree learns a follows a set of rules accurately and compares the recognition accuracy with that of other algorithms [4]. The Naive Bayes algorithm obeys a posterior probability using Bayes theorem. For a particular task domain, it computes different classes' conditional probabilities to choose the target class having highest posterior. Rapid development in information technology has made the ECG analysis more informative for the diagnosis and monitoring of different fatal diseases [13-14]. Due to availability of vital information in these signals during physiological investigation of human being, evaluation of health status using these signals has become increasingly popular.

A SVM platform has been used in this work to create support vectors for further classification of the chosen diseases. The choice of the classifier is made due to the following attributes. First, for larger samples, the classifier remains more effective. Second, it is capable to handle large feature space having the desired theoretical bounds with good network generalization [14-16]. Third, the network needs a few parameters for tuning, easy to implement and optimize using the convex quadratic cost function. Fourth, it ensures a globally unique solution [2]. Finally, it follows statistical learning theory and is based on the structural risk minimization principle. Mostly it has shown systematic and attractive results to separate the linear and non-linear class boundaries [14]. The classifier builds a hyper-plane that acts as the decision surface to create a separation margin between negative and positive examples for optimization. Nevertheless, support vectors are defined are emphasizing on the informative training data thus, considered sparse in comparison with the training data [5, 9].

Consider an  $N$ -dimensional hyperspace with  $X_M$  being a set of  $M$  labeled data points and is given by

$$X_M = [(s_1, k_1), \dots, (s_M, k_M)] \in (S \times K)^M \quad (1)$$

Where  $s_i \in S$ ,  $S$  denotes an input space where as  $k_i \in K$ ,  $K = \{-1, +1\}$  denotes the label space.

To formulate the problem, we need to design the function that obeys which is able to predict using the input . Generally, it is not possible to partition with the linear decision boundary although it is made linearly separable by transforming it into a higher or equal dimensional feature space using Cover's Theorem. The non-linear problem is thus reduced to estimating the optimal *hyperplane* that can separate the two classes. In the transformed domain, the *hyperplane* or the feature space can be parameterized using weights and biases as:

$$\sum_{i=1}^n (s_i) + b = 0 \quad (2)$$

Where  $n$  denotes the number of input features and is the mapping not required to be estimated explicitly rather an inner product kernel is defined to compute the optimal *hyperplane* which is expressed as [9]

$$\langle \vartheta(s_j), \vartheta(s_j) \rangle = (s_j, s_j) \quad (3)$$

This paper uses the Cohen's group of time-frequency kernels and has been discussed briefly. Generally, use of *hyperplanes* or linear function cannot make the input patterns perfectly separable even after transformation. To alleviate the issue, this work uses the soft boundaries to build an optimal *hyperplane* so as to minimize the error in classification.

For the training set the weights and biases are updated to optimal values so as to fulfill the constraint as given by

$$k_j (w^T s_j + b) \geq 1 - \delta, \text{ for } j = 1, 2, \dots, M \quad (4)$$

Where  $\delta$  denotes the slack variable. for all values of  $i$  in such a manner that the and attempts to minimize the cost function and is expressed as

$$\theta(w, \delta) = \frac{1}{2} w^T w + p \sum_{j=1}^M \delta_j \quad (5)$$

where signifies the user-specified positive parameter.

Using duality, the formulation of this optimization problem can be made as

$$\text{maximize } (\alpha) = \sum_{j=1}^M -\frac{1}{2} \sum_{j=1}^M \alpha_j \alpha_j k_j k_j s_j^T s_j \quad (6)$$

Subject to

$$\sum_{j=1}^M \alpha_j k_j = 0, 0 \leq \alpha_j \leq p \text{ for } j = 1, 2, \dots, M \quad (7)$$

For linear classification problem, maps input a feature space of equal or higher dimension. In this case, the dual variable optimized values are used to compute the optimum weights as given by

$$w_o = \sum_{j=1}^M \alpha_{o,j} k_j \vartheta(s_j) \quad (8)$$

where indicates the initial value of the optimal weight vector and corresponds to the dual variable optimum vector. In the feature space, the decision function is expressed as

$$o(x) = \sum_{j=1}^N \alpha_j k_j L(x, s) + b \quad (9)$$

Where  $L(x, s)$  denotes the kernel [9].

The recognition is based on,

If  $o(s) > 0$ , is in state 1

$o(s) < 0$ , is in state 1

### 3. Feature Extraction Using Wavelet Transform

The wavelet transform uses the multi-resolution capability to decompose a signal into detailed and approximate coefficients using filtering. Both the time and frequency representation can be possible due the scaling property. It has made the wavelet transform suitable for analysis of the non-stationary signals such as speech and ECG signals [17-20]. This work is confined to ECG signal which remains noisy and non-stationary using the wavelet transform. The aim is to remove the noise as well as to locate the desired peaks accurately. The ECG signal is initially decomposed into low and high resolution corresponding to low and high frequency components. Using two orthogonal by:

$$\{ \xi_{j+1} (t-2^{j+1}l)_{eY} \} \text{ of } G_{j+1} \quad (10)$$

$$\{ \xi_{i+1} (t-2^{j+1}l)_{eY} \} \text{ of } H_{j+1} \quad (11)$$

where,  $H_{j+1}$  and  $G_{j+1}$  represent the detailed and the approximation coefficients respectively.

### 4. Results

The accuracy has been estimated and compared using the reduced feature sets with that of the results obtained earlier. It has also been compared to the existing feature/attribute reduction techniques as shown in Table 1. In every cluster, it has tabulated the top five genes corresponding to the Colon Cancer-6 datasets.

Clusters	Gene Accession Number				
	Rank1	Rank2	Rank3	Rank4	Rank5
1	X52734	X87159	M90684	M16029	T62635

2	U30498	U19261	R35885	X51346	R77780
3	R80427	H13292	R73660	X16354	D31887
4	M64110	R21427	L42611	X63629	D21205
5	H72110	L38696	T68848	H40560	T48386
6	L06328	L16510	J04046	T64941	T41432

Table 1: Each cluster with the top five genes corresponding to the Colon Cancer -6dataset

The detection of a fatally ill patient has been graphically shown in Figure 1 using dB2 family. Similarly, this has been shown for a normal person using dB 4 wavelet family in Figure 2.

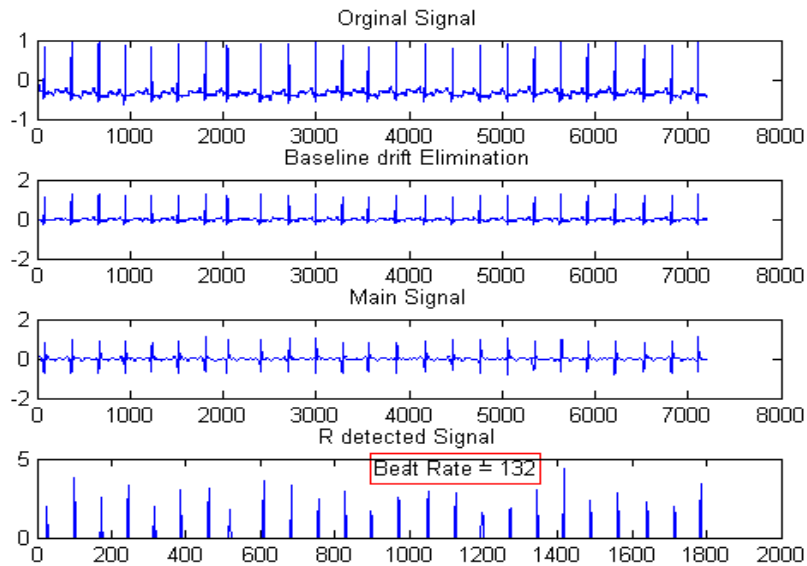


Figure 1. The detection of a fatally ill patient using db2

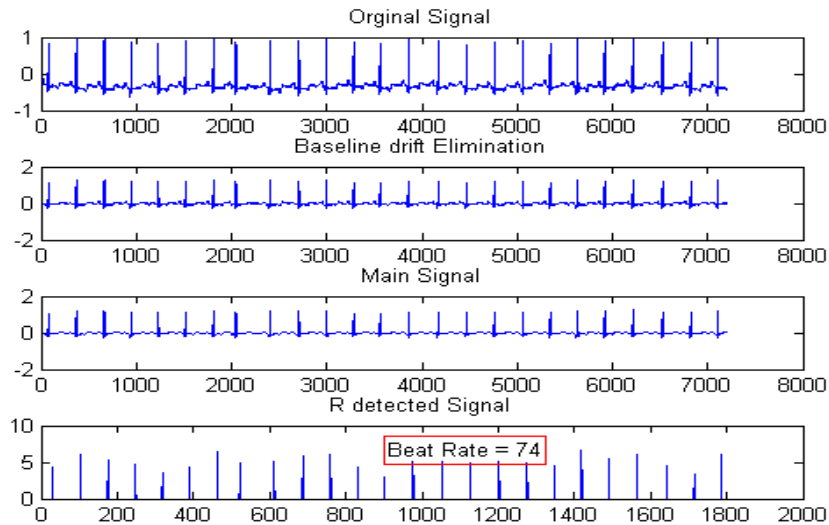


Figure 2. The detection of a normal human being with db4

A comparison is made with a number of NN based algorithms corresponding to the Colon Cancer-6 data in Table 2. As observed, the SVM has outperformed all the chosen classifiers with a highest average recognition of 97.97 % and 95.6% with respect to the Colon cancer and ECGdata respectively. However, the decision tree has shown better performance as compared to the NN as our result suggests.

For Colon Cancer

<b>Classifiers</b>	<b>Colon Cancer</b>	<b>ECG Signal</b>
Decision Tree	90.2	91.3
Neural Network	89.9	90
SVM	97.87	95.6

Table 2: The accuracy percentage of different classification algorithms

## 5. Conclusions

In this work, it has been concluded that overlapping segments provides better accuracy than the non-overlapping segments used to train the machine learners as observed in past studies. Among the chosen set of classifiers, the SVM has shown better result as compared to others. When the two datasets have been compared, it is found that the ECG has shown better accuracy as compared to the Colon cancer data. Finally, it can be inferred that the work is able to represent the desired disease accurately using the multi-resolution capability of wavelet analysis. Use of other ML algorithms including the deep learning may provide new future direction in this field.

## 6. References

- [1] T. Mitchell, (1997), "Machine Learning", McGraw Hill. p. 2. ISBN 0-07-042807-7.
- [2] V. Vapnik, (1998), "Statistical Learning Theory", Wiley, New York.
- [3] Russell, Stuart & Norvig, Peter (2003), "Artificial Intelligence: A Modern Approach", 2nd ed., Prentice Hall. ISBN 978-0137903955.
- [4] R. Duda, P. Hart, & D. Stork, (2001), "Pattern classification", Wiley, New York.
- [5] Mihir Narayan Mohanty, and Aurobinda Routray, (2015), "Machine Learning Approach for Emotional Speech Classification", Springer International Publishing Switzerland 2015, SEMCCO 2014, LNCS 8947, Chap. 43, pp. 1–12.
- [6] M. Lichman, (2013), "UCI machine learning repository", [Online].

- [7] Mihir Narayan Mohanty, &A. Routray, (2012), “Estimation of Autocorrelation Space for Classification of Bio-medical Signals”, SEMCCO, Book Chapter in Book title Swarm, Evolutionary, and Memetic Computing, LNCS 7677, pp. 697–704, © Springer-Verlag Berlin Heidelberg.
- [8] D. Du, K. Li, X. Li, & M Fei, (2014), “A novel forward gene selection algorithm for microarray data”, Neurocomputing, No.133, pp. 446-458.
- [9] Mihir N. Mohanty, A. Routray, &P. Kabisatpathy, (2010), “Optimization of features using Evolutionary Algorithm for EEG Signal Classification”, Int. J. Computational Vision and Robotics, Inder Science Pub., Vol.1, No.3, pp-297-310.
- [10] B. Liao, Y. Jiang, W. Liang, W. Zhu, L. Cai, & Z. Cao, (2014), “Gene selection using locality sensitive Laplacian score”, IEEE/ACM Trans. Comput. Biol. Bioinform. Vol. 11, pp.1146–1156.
- [11] Piyush ,Mundra A , & J. C. Rajapakse, (2016), “Gene and sample selection using T-score with sample selection”, Journal of Biomedical Informatics. Vol. 59, pp.31–41.
- [12] P. Addison, (2005), “Wavelet transforms and the ECG: a review”, Physiological measurement, Vol.26, No. 5, p.155.
- [13] K. Polat, B. Akdemir, &S. Gune, (2008), “Computer aided diagnosis of ECG data on the least square support vector machine”, Digital Signal Processing, Vol.18, pp.25–32.
- [14] J Cao, L. Zhang, B. Wang, L. Fi, & J. Yang, (2015), “A fast gene selection method for mulit-cancer classification using multiple support vector description”, Journal of Biomedical Informatics, Vol. 53, pp. 381–389.
- [15] D. Liu, H. Qiann, G. Dai, & Z. Zhang, (2013) “An iterative SVM approach to feature selection and classification in high-dimensional datasets”, Pattern Recognition, Vol.46, pp.2531–2537.
- [16] M. Tong, K. H. Liu, X. Chungui, & W Ju, (2013), “An ensemble of SVM classifiers based on gene pairs”, Computers in Biology and Medicine, Vol.43, pp.729–737.
- [17] H. K. Palo, M. Chandra, & M. N. Mohanty, (2016), “Efficient feature combination techniques for emotional speech classification”, International journal of speech technology, Vol. 19, No.1, pp.135-150.
- [18] A. Al-Fahoum, &I.Howitt, (1999), “Combined wavelet transformation and radial basis neural networks for classifying life- threatening cardiac arrhythmias”, Medical & biological engineering & computing, Vol.37, No.5, pp.566--573.
- [19] H. K. Palo, M. Chandra, &M. N. Mohanty, (2018), “Recognition of Human Speech Emotion Using Variants of Mel-Frequency Cepstral Coefficients”, In Advances in Systems, Control and Automation, Lecture Notes in Electrical Engineering, Springer, Singapore, Vol. 442, pp. 491-498.
- [20] B. Chandra, B. Naresh,& K. V. Babu, (2014), “Classification of gene expression data using Spiking Wavelet Radial Basis Neural Network”, Expert Systems with Applications. Vol. 41, pp. 1326-1330.