Wrapper based Feature Selection for Virtual Colonoscopy Classification

B. Ratnaraju*, G. N. Swamy1 and K. PadmaRaju3

1ECE BITS, Vizag, Visakhapatnam - 530048, Andhra Pradesh, India; ratnarajubonasu@gmail.com
2VRSEC, Vijayawada - 520007, Andhra Pradesh, India; gavini_s@yahoo.com
3JNTUK, Kakinada - 533004, Andhra Pradesh, India; padmaraju_k@yahoo.com

Abstract
One of the primary causes for deaths due to cancer all over the world is CRC or Colorectal cancer. In order to diagnose and treat disorders of the colon Colonoscopy is used widely. If it is performed properly it is usually safe and correct and tolerated well. The feature selection's focus is the selection of subset variables taken from within the input to describe efficiently the data of input simultaneously bringing down the effects of variables that are not relevant as well as that of noise and still continue to give good results of prediction. An algorithm based on the socio-political change which was named ICA or the Imperialist Competitive Algorithm was introduced. K-NN or the nearest neighbor K was a classifier that maintained a minimal distance which was Euclidean between the feature of query vector and all the data in the nature of prototype training. Here the ICA was used for the selection of features in the detection of colon cancer and its treatment. Results have proved that this method is far superior to other methods in its metrics of performance.

Keywords: Colonoscopy, Feature Selection, Imperialist Competitive Algorithm (ICA), K-Nearest Neighbor (K-NN) Classifier

1. Introduction
A research body that is increasing has shown that the microenvironment of cancer is very different and complex to the extent that they can be seen as immunologic organs that are distinct. CRC or Colorectal Cancer as mentioned is one of the primary reasons for deaths related to cancer1. Many systematic methods for pathological condition diagnosis have led to a high rate of early detection of CRC patients thereby reducing mortality rates. For diagnosing pathological conditions might contribute to a high detection rate of patients at early stages of CRC, leading to a reduction in mortality rates. Sigmoidoscopy and a blood test known as fecal occult test are currently the methods of screening that has reduced mortality due to CRC3. But these also have their own limitations; the fecal occult test of blood has low sensitivity and flexible sigmoidoscopy on the other hand is not a comfortable technique for patients. 19-9 (CA19-9) a carbohydrate antigen and CEA or Carcino Embryonic Antigen are widely used as markers for detecting various types of cancer like liver, stomach, pancreatic, colon etc., But these markers have low sensitivity for detection mainly when the disease is at its early stages. So diagnostic markers that are CRC specific markers have to be developed and sensitive, non-invasive and rapid screening have to be ensured.

Patients suffering from IBD or inflammatory bowel diseases, CD or Crohn's disease, UC or Ulcerative Colitis have a higher rate of risk of CRC3. 18% was the magnitude of risk at 30 years post the initial diagnosis that was in UC3. As a consequence of this the guidelines for a professional society make recommendations that colonoscopies for surveillance have to ideally start between 8 and 10 years post diagnosis with a repeat after every 2 or 3 years on people with colonic CD or UC. In the literature that is recent there are conflicts based on secular changes pertaining to risks of IBD or with CRC patients. Most of the studies that come from cohorts based on population show

*Author for correspondence
that IBD and CRC incidence has reduced. Many reasons exist for this kind of a temporal reduction like wide usage of maintenance and its treatment, increase in frequency of colectomy, increase in uptake in surveillance colonoscopy in practice in allowing identification of dysplastic lesions before CRC develops.

In the general population, screening Colonoscopy screening in generic population has been connected with a reduction in the mortality rate of colorectal cancer. Those with negative colonoscopy are under a risk of further lesions especially on the period of follow up that can extend up to 10 years after the first examination. However, data that can prove whether this kind of protective effect really exists among IBD patients that have not undergone examination. In fact, this is a critical clinical question as the molecular pathogenesis of cancer is different from sporadic carcinoma. Therefore, the surveillance colonoscopies and their benefits can never by extrapolated by comparing the normal population with IBD patients. Further, some more factors can affect the potential estimates. The decline in the risk of CRC patients that have IBD can also bring down the benefits of regular examinations of surveillance. There is also an existing controversy that the best method to perform the surveillance examinations as well as the yield of biopsies that are randomly taken. The individuals that are diagnosed with IBD when their age is very less have to examine colonoscopies change the risk of the resultant patient population after CRC.

Mucosa visualization of the large intestine and the distal terminal ileum can normally be made possible in colonoscopy. Colon cancer incidence can be reduced by removal of Polyps. Colonoscopy is a very popular method for the evolution of the colon in many adults that have symptoms of large bowel, results that are abnormal when a radiographic study is made, deficiency of iron or anemia, CRC screening positive results, post polypectomy, surveillance of post cancer resection and surveillance after diagnosis in cases of bowel diseases that are anti-inflammatory.

A feature is an individual measurable property that is individualistic in nature for observing a process is a feature. When a set of these features are used a classification can be performed by any algorithm of machine learning. In the years of the past machine learning applications or recognition of patterns and their domain have expanded manifold. Many new techniques have been included to address the issue of reduction in variables that are not relevant or redundant which can hinder the performance of various tasks. Selection of feature through elimination can be of great help in understanding data, bringing down requirement of computation as well as the problems in dimensionality and improving of performance of predictors. The methods of selection of features can be through three different models which are 1. Filter method 2. Wrapper method and 3. Embedded method. The first approach designs an independent measure for a given algorithm. So the features that precisely indicate the original data are used. But this method does not consider the inter feature interactions. In this approach the following methods based on correlation for selection of feature are present, the t-test, mutual information, information gain, and finally methods based on entropy. The models known as Wrapper models normally concentrated on the improvisation of the accuracy of classification of problems of patterns and their recognition in order to bring about filter models that are better. But the wrapper approach is quite lengthy. The embedded techniques make used of an algorithm that is inductive representing the classifier as well as the feature selector.

Classification is perhaps the most common problems that are taken up for study in Statistics, databases and machine learning. Many such algorithms have been proposed in the past, but recently cancer classification has been gaining prominence among researchers. Different machine learning systems that are stochastic are being applied for the treatment of cancer. They are hierarchical clustering. SVM or support vector machines, k-nearest neighbors, self-organizing maps and Bayesian networks.

Some ideal benefits in using k-NN or k-Nearest Neighbor is robust training data to noisy ones which is effective in case of large data and very fast in case of small training sets. This does not need knowledge of the structure of data that is contained in the training set or retraining along with the training present when new patterns of training are included.

2. Literature Review

Rad et al. made a selection of a feature set best suited for classifying the varieties of rice on the basis of images of huge samples by means of an algorithm called imperialist competition. This was a new as well as an evolutionary method of optimization which derived its inspiration form competition in imperialistic situations. Results proved that feature sets chosen by this method gave better
performance of classification in comparison to genetic technique of algorithm.

Dienstmann et al. made a focus on molecular, gene and pathologic expression of characterizations in the early stages of cancer of colon; there were new insights into the biomarkers that made predictions and prognostication insights thereby helping in the definition of proper adjuvant treatments for patients in clinical practice of routine nature.

Moraru et al. further presented another different approach to analyze colon cancer in ultrasonography. A method known as speckle suppression was presented. A wavelet transform known as Daubechies was used for the shift invariance and its property and additional information on the plane of the complex wavelet domain in comparison with the real domain. This method gave good results and proved the image usefulness and their techniques of processing the diagnosis of imaging. The ROI had a variance known as local echogenicity which was used for the comparison of the distribution of the local echogenicity inside the image acquired. This image was further analyzed with histograms that could explain images of gray-level. This is very valuable and important information to help discriminate tumors.

Anantha Krishnan et al. analyzed data from a total of 6823 people with IBD in which 2764 had colonoscopy recently and 4059 without. They were followed for 3 years in hospitals in Boston. The result of this was a CRC diagnosis. An examination of the number of patients that underwent colonoscopy from within 36 months before diagnosis or also when a follow up period ended was made. They excluded ones performed within 6 months before diagnosis. A process Multivariate logistic regression was done. Colonoscopy (within 36 months) was connected with a low incidence of patients with CRC also with IBD and a much lower rate of mortality of those with CRC.

Sears& Garrett discussed the insights of micro biota contributions to explore a variety of framework for the evaluation of the role of microbes in causing cancer. Some findings were pointed out on CRC-potentiating species and knowledge gaps. Lastly, the microbial metabolism and its role were examined in relation to xenobiotics, bile acids and diet in both etiology and also in therapeutics of the CRC.

Siegel et al. made a provision for the statistics of cancer of colon which made an inclusion of the most recent data on incidence of this cancer, rate of survival and mortality. The first data was procured from the Epidemiology of the National Cancer Institute's Surveillance. The final result and the Epidemiology and finally the central cancer registry of the Association of North America. The death rates of Colorectal cancer has come down by 2% in the 90’s and further to 3% recently. The progress made in bringing this down and the decrease in death rates can further be brought down by improving access and usage of screening as well as standard treatment.

Dehal et al. further examined the connection between T2DM or Type2 Diabetes Mellitus and the rate of survival among patients that also had CRC and analyze if this varies according to insulin treatment, duration of insulin use and sed. Out of 2278 people diagnosed with rectal cancer or colon cancer from 1992 to 2007 a study of incidence of cancer was made. Those who had CRC along with T2DM had a very high mortality risk than those with only CRC without T2DM.

3. Methodology

3.1 Daubechies Wavelet Transform

The wavelet transforms of Daubechies have been defined just like the wavelet transform of Haar was done by calculating the averages of running and its difference through scalar products along with signals for scaling them the difference between them only being how the signals of scaling are actually defined. This type of wavelet has a very balanced response to frequency but they are non-linear. These wavelets make use of overlapping windows, to ensure that the coefficient of high frequency can reflect all changes in frequency. This makes the Daubechies wavelets more useful in the removal of noise and compression when processing audio signals.

3.2 Gabor Filter

Texture features are extracted by Gabor filters in the retrieval of medical images. These filters are used for identification and detection of image or texture. This can extract features that are edge like in different scales and also orientations where different locations of images are observed. These features resemble those that are procured by the visual system of humans. More specifically, texture segmentation, analysis and recognition are done by 2D Gabor filters. They are sinusoidal functions that are complex that are handled by a rotated Gaussian. The function of the Gabor is shown in the spatial domain equation (1):
\[ g(x, y) = g_n(x', y')e^{(2\pi jfx)} \]

In which the rotated co-ordinates \((x', y')\) are
\[(x', y') = (x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi) \]

And \(g_n\) denotes the Gaussian functions.

For the extraction of textural features an image that is textured into an input and later transformed through a bank of Gabor filters and then a Gaussian smoothing filter in cascades. A feature vector can be constructed from the images and their statistics of outputs that comes from the filter bank branches. In case a filter bank consisting of \(N\) number of branches is taken for extraction of features a Gabor filter denoted by \(G_i(i = 1, 2, \ldots, N)\) which is complex, the filter output also contains numbers that are complex. Therefore in case of a magnitude operator that is non-linear, the modulus of all the complex numbers at the point of output is computed that is at a post Gaussian stage is \(P_i\) \((i = 1, 2, \ldots, N)\). Which is the Gaussian post filter \((I)\) and the output required is the needed is the feature image, where the output statistics provides texture features. Let, \(W_jk\) denote \((2m+1)\times(2n+1)\) in the neighbourhood of \((j,k)\)th pixel of the output image \(I\). Feature vector for \((j,k)\)th pixel (given by \(F_k\), needed for classification pixel wise) of an input image \(I\) which is formed in statistics of second order statistics of the local pixel features \((j,k)\). The construction of feature vector and of its length \(2xN\) is given in equation (3):

\[
F_i(2r-1) = \mu[W,(j,k)] \text{ and } F_i(2r) = \sigma[W,(j,k)]; r = 1, 2, \ldots, N
\]

In which, \(\mu(\bullet)\) and \(\sigma(\bullet)\) denote the operation of computing mean and standard deviation.

### 3.3 Imperialist Competitive Algorithm

ICA or Imperialist Competitive Algorithm, which is derived from social evolution of humans, becomes a component of the theory swarm intelligence. ICA, is a meta-heuristic technique which imitates the social behavior of the imperialistic competition between the countries in solving problems of optimization. The initial version of the algorithm has been used in solving optimization problems continuously\(^{22}\). In its similarity to other algorithms that are evolutionary ICA with its first random population known as countries. Every one of this has an \(n\)-dimensional problem of optimization which is denoted as a \(1\times n\) array as shown by equation (4):

\[ Country = [p_1, p_2, \ldots, p_n] \]

From among those countries that possess ideal function values of cost imperialists are chosen. The first few empires are formed by colony distribution between imperialists based in proportion on the power that has been normalized by the imperialists \((P_i)\) as depicted in equation (5):

\[ P_n = \frac{C_n}{\sum_{i=1}^{n} C_i} \]

In which \(C_n\) is cost of nth imperialist, where the imperialist’s normalized cost \((C_n)\) is defined by means of equation (6):

\[ C_n = c_n - \max(c_i) \]

Once the distribution of colonies between imperialists is completed closing of the empires corresponding to them takes place. The result of this is that the entire power of the empire is nothing but the imperialist country’s power along with an additional power percentage of their respective colonies. Now the empires get into the process of increasing their power by means of bringing closer to them the other empires. And the result of this is that there is a slow decrease in the weak empire’s power thereby an increase in the stronger empire’s power takes place automatically. The manner in which the strongest empire competes with the weakest colony belonging to the weakest of the empires is known as imperialistic competition. As the power of an empire increases it begins to attract more and more weak colonies thereby rendering the weak empires unable to see success in their competition with others and not able to improve the condition of their power which slowly gets eliminated completely based on some algorithm iteration. The final result of this is the totality of power of all the strong empires gradually goes up and weak empires completely collapse. After this takes place the colonies belonging to that empire is distributed to the remaining empires. Finally, all the colonies gradually become one state belonging to one empire for the entire world and making all the other countries its colonies\(^{13}\).

**Algorithm for ICA:**

- Initialization of the Empires
- Moving of the colonies to the imperialist relevant to them
- Verifying the presence of a colony where the cost is higher than the imperialist
Conduct an exchange between the imperialist and the colony positions
Otherwise total cost of empires to be computed
Picking of the colony that is the weakest from among the weakest empire and hand it over to the empire that is most likely to process it
Check if there is an empire without colonies
Elimination of the empire without a colony
Otherwise stop and check if condition is satisfied
Completed. If not re start from step 2.

3.4 Classification Algorithm

3.4.1 K-Nearest Neighbor (KNN) Algorithm

A method for the classification of objects on the basis of training examples that are closest to the feature space is known as k-NN or k-nearest neighbor algorithm [21] k-NN is that learning that is instance based also known as lazy learning in which the functions are locally approximated and until classification is complete all types of computation are deferred. One of the simplest types of algorithms of machine learning is the k-nearest neighbor algorithm which was originally proposed by Cover and Heart in the year 1968. This is an object that is identified by the neighbors’ majority of votes by an object that is assigned to the most common class from among its neighbors that are k-nearest where k is a small integer that is positive. When k= 1, the assigning of the object is made to the commonest value that is the nearest examples of training to k. All ties are broken randomly. The nearest neighbor algorithm of k that has been used in classification of neighborhood as a value of prediction of the value of the query instance that is new. From among a set of those objects the neighbors are pulled out where the classification is correctly available. Normally any object is classified in terms of the distance that is has from that of its neighbors that is with the objective of ideally being assigned to the relevant class that is the commonest among the nearest neighbors of k distance. The key word for this algorithm is distance. Every single object here is duly represented by vectors of a feature space that is multidimensional where the Euclidean distance can be used for the purpose of calculating the actual distance that exists between both the positions of the vectors.

Classification of k Nearest Neighbor takes place on finding a group of k objects by the algorithm in a set of training that is actually the closest in terms of input and the classification of the input which is based on the class and its majority in that group. The prime elements that are needed for the approach are 1. A labeled objects set 2. Metrics and number of k’s distance. The algorithm for k-nearest neighbor and its classification is as in figure 1.

3.4.2 Naive Bayes (NB)

A naive bayes classifier is based on Aa bayes theorem based classifier of naive bayes is capable of getting an efficient performance in classification. The probabilities of condition are \( P(x_j \mid c_i) \) and the probabilities prior to it are \( P(c_i) \). \( P(c_j) \) which is calculated by the counting of the samples of training and then by the dividing of the result of the count by using the set size of training. The naive bayes classifier is shown in equation (7):

\[
F_i(x) = P(x_j \mid c_i)P(c_i)
\]

In which \( X = (x_1,x_2,\ldots,x_N) \) denotes the feature vector, \( c_j \) denotes class labels, \( j=1,2,\ldots,N. \) This model is a simpler version of the model known as Bayesian probability theory. The naive Bayes classifier is on the assumption that features do not depend on it but operate on it. One attribute’s probability does not affect that of the other. It performs well on the classification but however it shows that errors happen owing to three factors which are bias, noise in training data and variance. The noise in training data can be reduced only by choosing good data for training. This bias has a huge impact on the variance which is small while training data is grouped.

4. Results and Discussion

Table 1 shows the summary of results for Classification accuracy, positive predictive for positive and negative sentiment, sensitivity for positive and negative sentiment and F measure for defect and no defects. Figure 2 to 5 shows the same.

| Input: D with k sets of training objects x, y and test object x’y’ |
| Process: compute distance \( d \) \( x', x \) between test object and every object. |
| Select \( D_k \subset D \) set of k closest training objects to test object. |
| Output: \( y' = \arg\max_c \sum_{x \in D_k} I(x,y) = y_i \) |
| Where \( v \) is the class label, \( I \) is a indicator function. |

Figure 1. Algorithm for K-Nearest Neighbor (KNN).
Table 1. Summary of Results

<table>
<thead>
<tr>
<th></th>
<th>GA based feature selection, Naive Bayes</th>
<th>GA based feature selection - KNN</th>
<th>Imperialist competitive algorithm - Naive Bayes</th>
<th>Imperialist competitive algorithm - KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification Accuracy</td>
<td>0.9545</td>
<td>0.9425</td>
<td>0.9715</td>
<td>0.9677</td>
</tr>
<tr>
<td>Positive Predictive Value for</td>
<td>0.9831</td>
<td>0.981</td>
<td>0.9845</td>
<td>0.9826</td>
</tr>
<tr>
<td>Positive sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Predictive Value for</td>
<td>0.9141</td>
<td>0.8903</td>
<td>0.9519</td>
<td>0.9453</td>
</tr>
<tr>
<td>Negative sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity for Positive</td>
<td>0.9417</td>
<td>0.9238</td>
<td>0.9686</td>
<td>0.9641</td>
</tr>
<tr>
<td>sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity for Negative</td>
<td>0.9746</td>
<td>0.9718</td>
<td>0.9761</td>
<td>0.9732</td>
</tr>
<tr>
<td>sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F measure for defect</td>
<td>0.9434</td>
<td>0.9293</td>
<td>0.9638</td>
<td>0.959</td>
</tr>
<tr>
<td>F measure for no defect</td>
<td>0.962</td>
<td>0.9515</td>
<td>0.9765</td>
<td>0.9733</td>
</tr>
</tbody>
</table>

Figure 2. Classification accuracy.

Figure 3. Positive predictive value for positive and negative sentiment.

It is observed from Table 1 and Figure 2 that the classification accuracy of imperialist competitive algorithm with Naive Bayes performs better by 1.77% than GA based feature selection with Naive Bayes, by 3.03% than GA based feature selection with KNN and by 0.39% than imperialist competitive algorithm with KNN.

It is observed from Table 1 and Figure 3 that the positive predictive value for positive sentiment of imperialist competitive algorithm with Naive Bayes performs better by 0.31% than GA based feature selection with KNN and by 0.39% than imperialist competitive algorithm with KNN.
0.14% than GA based feature selection with Naive Bayes, by 0.36% than GA based feature selection with KNN and by 0.19% than imperialist competitive algorithm with KNN. Over all positive predictive value for positive sentiment performs better than negative sentiments.

It is observed from table 1 and figure 4 that the sensitivity for negative sentiment of imperialist competitive algorithm with naive bayes performs better by 0.15% than GA based feature selection with Naive Bayes, by 0.44% than GA based feature selection with KNN and by 0.29% than imperialist competitive algorithm with KNN. Over all sensitivity for negative sentiment performs better than negative sentiments.

It is observed from Table 1 and Figure 5 that the F measure for no defect of imperialist competitive algorithm with naive bayes performs better by 1.49% than GA based feature selection with Naive Bayes, by 2.59% than GA based feature selection with KNN and by 0.33% than imperialist competitive algorithm with KNN. Over all F measure for no defect performs better than with defect.

5. Conclusion

CRC or Colorectal Cancer is a very common one which caused huge malignancies and health issues the entire world over. The mortality from this has come down in the last twenty years as detection is early and good treatments are available. A general searching method known as ICA from the imperialistic competition from the interactions that are geo-political between countries. This algorithm is extremely sensitive to the data’s local structure. One of the algorithms which have proved to be extremely simple and easy to practice. Results have proved that the accuracy of the classification of the competitive accuracy of the imperialist which has the naive bayes that can perform much better 1.77% than that of a feature selection based on GA with the KNN and further by 0.39% in comparison with the imperialist algorithm with the KNN. The predictive values that are positive its sensitivity and the proposed method’s F measures that can give better performance.

6. References

17. Elfouly FH, Mahmoud MI, Dessouky MI, Deyab S. Comparison between Haar and Daubechies wavelet