A New Way To Prevent Colorectal Cancer Using Supervised Learning Technique

Balaji Vicharapu\textsuperscript{a}, Anuradha Chint\textsuperscript{b}, S.R. Chandra Murty Patnala\textsuperscript{c}

\textsuperscript{a}Research Scholar, Department of CSE, Acharya Nagarjuna University, A.P, India
\textsuperscript{b}Assistant professor, Department of CSE, V RSiddhartha Engineering College, India
\textsuperscript{c}v.balaji.anu@gmail.com

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: The colorectal cancer prompts to more number of death as of late. The diagnosis of colorectal cancer as early is protected to treat the patient. To distinguish and treat this type of cancer, Colonoscopy is applied ordinarily. Several risk prediction models for colorectal cancer have been created and approved in various populations but colon cancer effecting the young adults. In this research, we projected a Supervised Learning Technique for detecting colorectal cancer in high dimensional information. One of the most important and very popular tool for performing the machine learning tasks that includes novelty detection, classification or regression is Support vector machine (SVM). Training the SVM requires large quantity of quadratic programming. Due to memory constraints conventional methods are not directly applied. To overcome these inadequacies, we introduced, Least Square (LS), Particle Swarm Optimization (PSO), Quadratic Programming and Quantum-behave PSO methods for training SVM. To corroborate the competence and proficiency of our predictable system, it is developed in open source called NCSS Software. The acquired outcomes of these approaches are verified on a CCG1.11 colorectal dataset and related with the particular resolution model.

Keywords: Colorectal Cancer, Machine Learning, Support Vector Machine, Particle Swarm Optimization, CCG 1.11 and Classification Accuracy

1. Introduction

Now a days, cancer deaths is a very dangerous out of all, only 9.6 M peoples are died due to the cancer diseases worldwide in 2018, whatever the reason/ distortion it is. In twenty five years, cancer deaths are decreased by 27 percent in the United States, but this rate is not acceptable. In 2019, more than 6, 00, 000 cancer deaths are predictable and 1.7M or more new cancer cases are recorded with diagnosis. "Cancer is a group of diseases in which cells in the body grow, change, and multiply out of control" [1]. In Pattern recognition domain, cancer detection is a very significant research area. This research paper implementing an automatic diagnostic system and classifies cancer patients by building a linear optimal classifier using support vector machine for colorectal cancer. Here four models are used for training SVM such as Quantum-behave PSO, Least Square (LS), Particle Swarm Optimization (PSO), Quadratic Programming methods and also calculated the classification accuracy. Now a day's usage of classification in medical diagnosis system gradually increases. The most important factors in diagnosis system are patient’s evaluation data and experts decisions. Though, different AI techniques and classifications systems, we can minimize the classification errors those are garnered due to lack of qualified persons and also provide examination of medical information in short time and more exhaustive way. Fig1 illustrates the different steps used in classification design system. As it is outward from the remarks indicators, these steps are dependent. On the opposite, they're depending and interconnected, on the consequences, one may go-back to restructure previous phases in an effort to improve the complete overall performance.
The remaining of this research work is structured as shadows. In Segment 2, the literature work relate to this field is summarized. Segment 3 examines the projected model called supervised learning system. Then, Segment 4 and 5 designates the Research Methodology in detail and compared experimental outcomes with other prototypes. In final Segment 6, summary and forthcoming work is described.

2. Literature Review

In medical field, the integration and advancement of technology is rapidly increasing. Various innovative methodologies have been introduced that are helpful for identification of diseases, providing clinical trial research, radiology, drug discovery, manufacturing, personalized treatment, epidemic outbreak predictions, radiotherapy and health records etc. Various types of cancers can be detected and characterized using amount of CAD arrangements, especially it is intended/ utilized for detecting the breast tumour diseases. It is also a significant tool in the interpretation of mammographic process and support for radiologists to come into a definite conclusion. In clinic, now CAD system is utilized as second reader for recognition of breast cancer and for malignant and benign lesions classification under the advancement by many research groups. For predicting the breast cancer, many innovative techniques haven evolved in the modern days with the advancement of technology. The literature work relate to this field is summarized as follows:

Many research works from the previous studies on diagnosis and prediction of diseases is based machine learning methods for cancer recognition. Machine learning techniques includes KNN, decision trees, SVM, Bayesian classification etc. out of these classifiers KNN procedure is repeated utilized, since its adaptability and simplicity in implementation and it leads to efficient and accurateness outcomes. According to various surveys shows that KNN is most commonly used machine learning method. Liu et.al projected a prototype for cancer recognition using machine learning algorithm. Author work utilized the logistic regression model for performing the classification operation on standard breast cancer databases. Two main features called perimeter and texture are selected and accurateness of projected classifier is 96.5%. Zerhouni et.al projected a prototype called Breast Cancer CAD that is based on Deep Neural networks and joint variable selection. For predicting the recurrence cut-off value, authors collect the data from Belfort hospital at France and it is named as Wisconsin Breast Cancer Database. Projected methodology is also smeared to minimize the no of response variables. The presentation of novel method increases and generates efficient and accurateness results using deep learning networks.

Bellaachia et.al projected a novel method that uses a combination of classifiers like C4.5 decision tree, the back-propagated, Naïve Bayes and neural network algorithms for breast cancer. Author uses SEER database that consists of 482,052 records and 16 attributes and this database is taken as model one due huge quantity of patient and a moderate no of attributes. Out of these projected classifiers C4.5 decision tree algorithm gives the better performance when compared to remaining classifiers with an accurateness of 86.7%. A new methodology for breast cancer diagnosis was projected by Xiao et.al by combining a deep research method based machine learning feature mining processes, auto encoding method with optimal methodology for extracting the key features and information, SVM model for recognising new features into malignant tumors and benign . The projected method is tested using important breast cancer database called Wisconsin Diagnostic. Finally Experimental outcomes displays enhanced the presentation of classification and providing a capable method to breast cancer diagnosis.

Many Researchers in past years and forthcoming which are purposes to perceive the most important structures that are obliging in benevolent cancer and forecasting malignant. And also helpful for selecting the specific prototypes and selection of hyper parameters. The main aim and objective of all researchers is to generate high accurateness outcomes in less computational time.

3. Supervised Learning System

SVM method is widely used for classification, density estimation and regression analysis. The SVM is an accepted discriminative classifier due to its outstanding features, high accuracy and brilliant empirical throughput.
The thought of SVM is to build a "hyper-plane" as the assessment plane in such a manner that that the edge of division between negative and positive samples maximize as shown in figure 2. They have been effectively applied to lots of dissimilar applications, such as text classification, speaker verification, image categorization, and bioinformatics. SVM are based on the instinctive thought of maximizing the edge of division between two challenging classes, where the border is clear as the distance between the choice hyper plane and the neighbouring training. It has been bare to be linked to minimalizingahigher bound on the interpretation fault.

For direct isolatable training pair of 2 classes, the particular verdict "hyper-plane"in multi-dimensional element trajectory gi(z) is known in the subsequentequivalence:

\[ g_i(z) = W_i^T \cdot Y + w_i0 = 0 \] ……………………………………(1)

Anywhere \( g_i(z) = \) Outcome feature trajectory
\[ W_i^T = [w_1, w_2, \ldots, w_n] \] T = weight vector
n = Total attributes
\( w_i0 = \) a scalar verge / bias weight
\( z = \) I/p feature trajectory

![Figure 2: Maximum Margin Separations for simple classification task](image)

The verdict "hyper-plane" then the subsequent is suitable.

\[ g_i(z_1) = W_i^T z_1 + w_i0 = 0 \] …………………………………………(2)
\[ g_j(z_2) = W_j^T z_2 + w_j0 = 0 \] …………………………………………(3)

Subtract two equations will provide the following series of equations:

\[ g_i(z_1) - g_j(z_2) = 0 \] \[ \iff \] \[ W_i^T z_1 - W_j^T z_2 = 0 \]
\[ g_{ij}(z_1, z_2) = 0 \] \[ \iff \] \[ W_i^T z_1 - W_j^T z_2 = 0 \]

i.e. \( g_{ij}(z_1, z_2) = 0 \) \[ \implies \] \[ W_i^T(z_1 - z_2) = 0 \] …………………………………………(4)

Where \( (z_1 - z_2) \) is a trajectory equivalent to the choice border and is intended for from \( z_1 \) to \( z_2 \). Since the point creation is 0, the path for \( W^T \) must be vertical to choice border. So, at all point that lies on top of the separating hyper plane.

\[ w_1 z_1 + w_2 z_2 + w_i0 = k > 0 \] …………………………………………(5)

Likewise, at every point that deception underneath the sorting out "hyper-plane" fulfils i.e. for each sphere positioned underneath the choice border, we can display that

\[ w_1 z_1 + w_2 z_2 + w_i0 = 0 \] …………………………………………(6)
If we tag the class +1 as squares and class -1 as circles, then we can work out the class tag \( Z \) for all test example Q1.

The values can be attuned so that the hyper planes important the edges of the border can be transcribed as

\[
H_1: \quad w_1z_1+w_2z_2 + w_0 \geq 1, \text{ for } Z_i = +1
\]

\[
H_2: \quad w_1z_1+w_2z_2 + w_0 \leq -1, \text{ for } Z_i = -1
\]

Some tuples falls on or on top, any tuple that falls on or below \( H_2 \) belongs to class -1, and \( H_1 \) fits to class +1. Connecting the two discriminations of equivalences and we get

\[
Z_i (w_1z_i+w_2z_2 + w_0) \geq 1, \text{ for all } i
\]

The edge can be computed by subtracting the statement 2 from the statement 1. This is equal with

\[
\frac{1}{||w||} + \frac{1}{||w||} = \frac{2}{||w||}
\]

Necessary that

\[
wiTz + wio \geq 1, \quad \forall z \in w1
\]

\[
wiTz + wio \leq 1, \quad \forall z \in w2
\]

Process the parameters \( w, wio \) of the "hyper-plane":

So, diminish \( J(w,wio) = \frac{1}{2}||w||^2 \)……………………………………...(10)

\[
Z_i (w_1Tz_i + wio) \geq 1, \quad i=1, 2, \ldots, N
\]

The Karush-Kuhn-Tucker expresses that the minimalize of above equivalences needs to legitimize argument:

\[
\frac{\partial}{\partial w} L(w, wio, \lambda) = \begin{cases} 0 & \text{for category } Z_1 \\ 0 & \text{for category } Z_2 \\
\end{cases}
\]

where \( \lambda_i \geq 0 \)

\[
\sum_{i=1}^{N} \lambda_i[Y_i(w_1Ty_i + wio) – 1] = 0
\]

Integrate the equations (5.16), (5.17) and (5.18), we get

\[
W = \sum_{i=1}^{N} \lambda_i Z_i \quad \text{and} \quad \sum_{i=1}^{N} \lambda_i = 0
\]

A novel technique based on the SVM classification for PQ disturbances. It is experimental that SVM properly classifies PQ disturbances. The projected method using SVM generates over all classification rate of 99.1%. Hence the technique can be used for classification of PQ disturbances.

If two divisions are in "non-linear case", equations (10) and (11) are no longer suitable and have dissimilar procedures. The training feature vector depends on the subsequent groups:

Trajectories that drop outer the sphere and are properly categorized. These trajectories agree with the restrictions

\[
Z_i (w_1Tzi + wio) \geq 1, \quad i=1, 2, \ldots, N
\]

Trajectories lessening inside the sphere and are properly categorized. These are the arguments placed in quadrangles of the "hyper-plane" and they accomplish the discrimination

\[
0 \leq Z_i (w_1Tzi + wio) < 1
\]

Trajectories that mis-classified. They are together with this by spheres and submit the discrimination

\[
Z_i (w_1Tzi + wio) < 0
\]

Total 3 circumstances can be preserved under a single type of restrictions by starting a novel objective function \( \Omega \) is given by

\[
Z_i (w1Tyi + wio) \geq 1 - \Omega_i
\]

For category Z-1: \( \Omega_i = 0 \) for category Z-2: \( 0 \leq \Omega_i < 1 \)
for category 3: $\varnothing_i \geq 1$

The variables $\varnothing_i$ is called as slack variables. The goal now is to make the margin as giant as probable but at the similar period to retain the number of arguments with $\varnothing > 0$ as slight as conceivable. This equals to adopting to minimize the "cost-function"

$$J(w, w_0, \varnothing) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} l(\varnothing_i)$$

Anywhere $\varnothing$ is the trajectory of the constraints $\varnothing_i$ and

$$\varnothing_i = \begin{cases} 1, & \text{if } \varnothing_i > 0; \\ 0, & \text{if } \varnothing_i = 0; \end{cases}$$

The constraint $C$ is an optimistic constant that manage the comparative effect of the challenging positions. Optimization issue can be resolved by minimalizing Lagrange utility.

$$L(w, w_0, \varnothing, \lambda, \mu) = \frac{1}{2} ||w||^2_2 + C \sum_{i=1}^{N} \varnothing_i \sum_{i=1}^{N} \varnothing_i \mu_i$$

$$= \sum_{i=1}^{N} \lambda_i [Y_i (w_i^T z_i + w_0) - 1]$$

The correspondent Karush-Kuhn-Tucker circumstances that the minimize of above equivalences has to satisfy are

$$\frac{\partial L}{\partial w} = 0 \text{ or } w = \sum_{i=1}^{N} \lambda_i z_i$$

$$\frac{\partial L}{\partial w_0} = 0 \text{ or } \sum_{i=1}^{N} \lambda_i z_i = 0$$

$$\frac{\partial L}{\partial \mu_i} = 0 \text{ or } \sum_{i=1}^{N} C - \mu_i \lambda_i = 0 \text{ or } \mu_i \geq 0, \lambda_i \geq 0 \text{ or } \mu_i \lambda_i = 0, \mu_i \geq 0, \lambda_i \geq 0 \text{ or } \mu_i \lambda_i = 0, \mu_i \geq 0, \lambda_i \geq 0$$

In non-linear circumstance, SVM compare the input trajectories $y$ into a lofty feature space through some non-linear comparing. In this work, the following algorithms are used to solve optimization and non-convex optimization issue.

Algorithm: SVM learning algorithm with optimal parts

Input: $\{(O^{(r)}, W^{(r)}_e)\}$ C and accuracy $\varepsilon$;

Initialize $\alpha$ and empty constant set: $W_r \leftarrow \emptyset$

repeat

for $r = 1$ ...... $R$ do

$$W^{(r)} \leftarrow \text{arg}_w \max_{\alpha} [L(W, W^{(r)}_e + \alpha^T O^{(r)}, W)]$$

if $\alpha^T [O^{(r)} W^{(r)}_e] < L(W^{(r)}_e, W^{(r)}_e) - E_r \cdot \varepsilon$ then

/* put it in constraint set */

$$W_r \leftarrow W_r \cup W^{(r)}$$

$$\langle \alpha, E_r \rangle \leftarrow \min_{\frac{1}{2}} \frac{1}{2} \alpha^T \alpha + \frac{C}{R} E_r = \frac{1}{2} E_r$$

end if

end for

end repeat
\[ W^{(1)} \in W_1 : \alpha^T [ \Phi(O^{(1)}, W_{ref}^{(1)}, \Phi(O^{(1)}), W^{(1)} \geq L(W_{ref}^{(1)}, W^{(1)}) - E_1 \]

\[ W^{(R)} \in W_R : \alpha^T [ \Phi(O^{(R)}, W_{ref}^{(R)}, \Phi(O^{(R)}), W^{(R)} \geq L(W_{ref}^{(R)}, W^{(R)}) - E_R \]

until no \( W_r \) has changed during iteration.

return \( \alpha \).

Algorithm: SVM learning algorithm with non-convex optimal segmentation

1. fixing \( \alpha \), optimize the reference segmentation \( \theta^{(r)} \) for every training pair

\[ \theta^{(r)} = \arg\max_{\alpha} \alpha^T \Phi(O^{(r)}, W_{ref}^{(r)}; \theta^{(r)}), \forall r \]

2. Fixing \( \theta^{(r)} \), optimize \( \alpha \) by minimizing the subsequent curving higher bound using the cutting plane procedure.

\[ \frac{1}{2} \alpha^T \Phi^{(r)} - \alpha^T \Phi^{(r)}; \theta^{(r)} = \frac{1}{2} \sum_{r=1}^{R} \alpha^T \Phi^{(r)}; L(W_{ref}^{(r)}, W^{(r)}) \]

3. Repeat Step 1 until converge;

return;

4. **SVM Training Methods**

For construction of SVM classifiers, different techniques are examined. In order to determine the optimal value of nonnegative multipliers, four different methodologies (i.e. SVM training methods) are used. These methodologies include: i). Least Square Method ii). Particle Swarm Optimization iii). Quadratic programming iv). Quantum behaved PSO

4.1. Particle swarm optimization

In PSO, searching operation is performed via swarm of particles and updates can take place iteration to iteration. For obtaining the optimal solution, particles are moved from previous position called pbest and hbest position in swarm. One has

\[ q_{bes} (j, t) = \arg\min_{p=1, \ldots, t} [g(Q_j(p))], j \in \{1, 2, \ldots, MQ\}, \]

\[ h_{bes} (t) = \arg\min_{j=1, \ldots, MQ} [g(Q_j(t))], j \in \{1, 2, \ldots, MQ\}, t \]

Where \( j \) indicates the particle index, MQ the whole quantity of particles, \( t \) the current redundancy, Q the position and \( g \) the fitness function. The position Qand velocity U of particles are rationalised by the subsequent equivalences:

\[ U_j (t+1) = \omega U_j (t) + d_1 c_1 (q_{bes} (j, t) - Q_j (t)) + d_2 c_2 (h_{bes} (t) - Q_j (t)), \ldots \]

\[ Q_j (t+1) = Q_j (t) + U_j (t+1), \ldots \]

Where \( U \) denotes the velocity, \( \omega \) is the inertia weight used to balance the global search and local utilisation, \( d_1 \) and \( d_2 \) are optimistic constant factors called acceleration coefficients, and \( r_1 \) and \( r_2 \) are consistently scattered irregular factors inside range \([0, 1]\). It is common to fix a higher head for the speed factor. Speed packing was used as an approach to bound particles floating out of the investigation space. The 1st portion of formulation (14), known as inertia, signifies the preceding velocity, which delivers the essential motion for particles to travel through the exploration space. The 2nd portion, known as the reasoning constituent, signifies the separate particle sophisticated of every particle. It emboldens the particles to transference to their own best places originate so far. The 3rd portion, the collaboration constituent, signifies the concerted consequence of the particles to discovery the global optimum solution. The pseudo code representation of PSO procedure is shown below:

Stage 1. Introduction

For every particle \( j = 1, \ldots, R \), do

(a) Adjust the particle’s location with a consistently scattering as \( Q(0) \sim U(BL, BU) \).
A New Way To Prevent Colorectal Cancer Using Supervised Learning Technique

Where BL and BU signify the inferior and higher boundaries of the exploration space

(b) Adjust best to its firstplace: \( q_{\text{best}}(i, 0) = Q_j(0). \)

(c) Adjust best to the nominal charge of the swarm: \( g_{\text{best}}(0) = \arg\min \{ Q_j(0) \}. \)

(d) Adjust speed: \( U_j \sim (-|BU - BL|, |BU - BL|). \)

Stage 2. Replication until a endconditions is met for every particle \( j = 1, \ldots, U, \) do

(a) \( \)Update random amounts: \( c_1, c_2 \sim (0, 1). \)

(b) Update particle’s speed. See formulation (2).

(c) Update particle’s location. See formulation (3).

(d) Adjust speed: \( U_j \sim (-|BU - BL|, |BU - BL|). \)

Stage 3. Output \( h(t) \) that grips the best originate solution.

4.2 Least Square Method

A classification problem is deliberated as binary, taking a group of training vectors (D) belonging to 2 separate classes.

\[ D = \{(x^1, y^1), \ldots, (x^l, y^l)\}, x \in \mathbb{R}^n, y \in \{-1, +1\} \] \hspace{1cm} (16)

Where \( x \in \mathbb{R}^n \) is a multi-dimensional information vector, with each example having a place with both of two classes marked \( y \in \{-1, +1\} \), and \( l \) is the quantity of preparing information. This examination utilizes \( d, c, \phi, \beta, H \) as info boundaries. So \( x = [d, c, \beta, \phi, r, H] \), inside the current context of categorizing the position of the gradient, the 2 classes labelled +1 and −1 may mean stableslope and failedslope. The Support Vector Machine (SVM) approach targets building a classifier of the structure:

\[ Y(x) = \text{sign} \left[ \sum_{k=1}^{N} \alpha_k y_k k(x, x_k) + b \right] \] \hspace{1cm} (17)

Where \( \alpha_k \) are positive real constants, \( b \) is that the scalar edge, \( N \) is that the quantity of the informational index and \( k(x, x_k) \) is the Kernelfunction. For the instance of two classes, one assumes:

\[ w^T \phi(x_k) + b \geq 1, \text{if } y_k = +1 \text{ (stableslope)}, \]
\[ w^T \phi(x_k) + b \leq 1, \text{if } y_k = -1 \text{ (failed slope)} \] \hspace{1cm} (18)

Where \( w \) is a flexible weight vector, \( T \) is the translation and \( \phi(\cdot) \) is the component map that maps the input space into a better dimensional space, which is equivalent to:

\[ y_k (w^T \phi(x_k) + b) \geq 1, k = 1, \ldots, N \] \hspace{1cm} (19)

As indicated by the structural risk minimization principle, the possibility bound is minimized by formulating the subsequent optimization problem:

\[ \text{Minimize: } \frac{1}{2} w^T w + \gamma/2 \sum_{k=1}^{N} e_k^2 \]
\[ \text{Subjected to: } y_k (w^T \phi(x_k) + b) \geq 1 - e_k, \quad k = 1, \ldots, N \] \hspace{1cm} (20)

Where \( \gamma \) is the regularization parameter, deciding the compromise between the fitting mistake minimization and smoothness, and \( e_k \) is the error variable. This optimization problem (Eq. (20)) is solved by Lagrange multipliers and its solution is given by:

\[ Y(x) = \text{sign} \left[ \sum_{k=1}^{N} \alpha_k y_k k(x, x_k) + b \right] \] \hspace{1cm} (21)

Where \( \text{sign} \) (that is the signum function, It gives +1 (stableslope) if the component is \( \geq 0 \), and −1 (failed slope) if it is less than zero.)
4.3 Quantum-behave PSO

Heisenberg, de Broglie, Bohn, Schrödinger, and Bohr are the main findings in the twentieth century for the development of quantum mechanics. Their research enforced the researchers to reconsider the applicability of traditional mechanics and the classical sympathetic of the surroundings of warning signs of microscopic substances. As in step with traditional PSO, a particle is indicated via its region trajectory \( y_i \) and speed trajectory \( u_i \), which define the trajectory of the atom. The atom transfers along a determined trajectory subsequent Newtonian mechanics. Though if we deliberate quantum mechanics, then the time period trajectory is pointless, due to the fact \( y_i \) and \( u_i \) of an atom cannot be determined concurrently according to the uncertainty principle. Consequently, if separate particles in a PSO machine have quantum behavior, the enactment of PSO will be distant from that of conventional PSO. In the critical model of a PSO, the nation of a particle is represented through wave feature \( \Psi(y,t) \), in place of location and speed. The dynamic behavior of the atom is appreciably divergent from that of the atom in conventional PSO systems. In this attitude, the likelihood of the particle's seeming in region \( y_i \) from threat density feature \( |\Psi(y,t)|^2 \), shape of which depends on the potential arena the atom lies in equation (1) shown under:

\[
y_i = y_i + u_i \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldot
calculate a direction, $d_k$, to transfer to the next point $x_{k+1} = x_k + d_k$ such that the new repeat is practicable in $W_k$ and the objective function is minimalized at $x_k + d_k$. Since $x_k$ is recognized at the current stage, it will be observed as a continuous vector and the unidentified vector is only $d_k$. The problem is specified as:

$$\min_{d_k} f(d_k) = 1/2 (x_k + d_k)^T Q(x_k + d_k) + c^T (x_k + d_k) \quad \ldots \quad (29)$$

Subject to:

$$a_i^T (x_k + d_k) = b_i, \; i \in W_k \quad \ldots \quad (30)$$

Expanding the new objective function we have:

$$f(d_k) = 1/2 x_k^T Q x_k + 1/2 d_k^T Q d_k + x_k^T Q d_k + c^T x_k + c^T d_k \quad \ldots \quad (31)$$

The term $1/2 x_k^T Q x_k + c^T x_k$ is constant for a given $x_k$, thus it can be removed from the objective function without changing the solution.

We denote:

$$g_k = Q x_k + c \quad \ldots \quad (32)$$

and the function to be minimized becomes:

$$f(d_k) = 1/2 d_k^T Q d_k + (Q x_k + c^T) d_k = 1/2 d_k^T Q d_k + 1/2 g_k^T d_k \quad \ldots \quad (33)$$

Note that $Q$ is symmetric, thus $Q = Q^T$. Because $x_k$ is a feasible point within the working set $W_k$, the equivalence restriction:

$$a_i^T x_k = b_i, \; i \in W_k \quad \ldots \quad (8)$$

is satisfied. From (31) and (29) we get the equivalence restriction of the new QP sub-problem. It will be expressed as:

$$\min_{d_k} 1/2 d_k^T Q d_k + g_k^T d_k \quad \ldots \quad (34)$$

Subject to:

$$a_i^T d_k = 0, \; i \in W_k \quad \ldots \quad (35)$$

We may continue in a way related to the one applied for equality-constrained QP problems. For evaluation, we try to best one technique from training procedures stated such as Subset selection processes, Iterative processes, Exploiting alternative SVM constructions.

5. **Proposed Methodology**

The following Figure 3 depicts the proposed methodology for Colorectal Cancer Diagnosis Model. With this model, we can pre-process the data using scaling operation and processed data can be divided into two datasets: testing and training. SVM classifier is build using these training data and validation of each classifier is done using two important parameters: Sensitivity and Specificity in distinctive "cancer patients" from non-cancer controls. Different combination of features are used for building SVM classifies in order to reach the SVM Classifier to its maximum value. Cross validation methodology is utilized for calculating the classification accuracy and the parameter like generalization error is evaluated using validation dataset. For construction of SVM classifiers, here we used four different methodologies (i.e. SVM training methods) such as: i). Least Square Method ii). Particle Swarm Optimization iii). Quadratic programming iv). Quantum behaved PSO.

A New Way To Prevent Colorectal Cancer Using Supervised Learning Technique
Figure 3: Proposed Methodology of Colorectal Cancer Model

The experiments are done on the Colorectal Cancer CCG 1.11 dataset from the UCI [12]. It is 1yr consistent relative subsistence proportion for adults. A cumulative pointer for 1yr subsistence for all types of cancers in adults above 15. The probability estimation of subsistence from cancer alone is known as relative subsistence. It is definite as the proportion of the perceived subsistence and the subsistence that would have been predictable if the cancer patients had practiced the identical circumstantial humanity by sex and age as the common populace. The outcomes of the four approaches were equated and sample CCG1.11 database is exposed in table 1:

| Year of diagnosis | Period of coverage | Breakdown | Level | Level Description | Indicator value | Precision |
|-------------------|--------------------|-----------|-------|-------------------|----------------|-----------|
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00C | NHS Darlington CCG | 68.5 | 1.94 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00D | NHS Durham Dales, Easington and Sedgefield CCG | 67.38 | 2.8 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00F | NHS Gateshead CCG | 69.65 | 4.06 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00G | NHS Newcastle North and East CCG | 70.34 | 0 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00H | NHS Newcastle West CCG | 69.91 | 3.26 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00J | NHS North Durham CCG | 68 | 4.61 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00K | NHS Hartlepool and Stockton-On-Tees CCG | 68.53 | 3.59 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00L | NHS Northumberland CCG | 69.64 | 6.04 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00M | NHS South Tees CCG | 70.37 | 5.05 |
| 2011              | Diagnosis: 1/1 to 31/12/2011 Followed up until 31/12/2012 | CCG 00N | NHS South Tyneside CCG | 69.87 | 3.8 |

The outcomes of the four techniques were tested and equated with the above dataset called Colorectal Cancer CCG1.11.

6. Results And Discussion

In this segment, the efficiency of four SVN training methods are evaluated and compared. The objective of this comparison is two or more supervised learning techniques were evaluated alongside by considering the
A New Way To Prevent Colorectal Cancer Using Supervised Learning Technique

performance of SVM classifier (i.e. trained with PSO and Quantum) into perception. To corroborate the competence and proficiency of our predictable system, it is developed in open source called NCSS Software. In order to evaluate the efficiency of the projected technique, several parameters/measures were used. These parameters includes Error rate, negative and positive predictive values, confusion matrix, classification accuracy, specificity, sensitivity and distributed ROC curves. These measures are distributed curves (figure 4), analysis of specificity (Figure 5) and sensitivity (Figure 6), Error rate (Figure 7), classification accuracy (Figure 8), negative and positive predictive value (Table 3) and confusion matrix in Table 4.

![Figure 4: Distributed Curves of Colorectal Cancer CCG1.11 Dataset](image)

![Table 2: Confusion Matrix of Colorectal Cancer CCG1.11 Dataset.](table)

|       | C1       | C2       | C3       | C4       | C5       | C785     |
|-------|----------|----------|----------|----------|----------|----------|
| C1    | 1.000000 | 0.934786 | 0.831511 | 0.800765 | 0.781757 | 0.662829 |
| C2    | 0.934786 | 1.000000 | 0.892806 | 0.852014 | 0.841799 | 0.687042 |
| C3    | 0.831511 | 0.892806 | 1.000000 | 0.914945 | 0.834729 | 0.666610 |
| C4    | 0.800765 | 0.852014 | 0.914945 | 1.000000 | 0.917275 | 0.667555 |
| C5    | 0.781757 | 0.841799 | 0.834729 | 0.917275 | 1.000000 | 0.666786 |
| C785  | 0.662829 | 0.687042 | 0.666610 | 0.667555 | 0.666786 | 1.000000 |
Table 3: Comparison of SVM Training Methods with Different Parameters.

| S.N | Parameter/ Training Method | PSO  | Q-B PSO | QP  | LSM  |
|-----|----------------------------|------|---------|-----|------|
| 1   | Specificity                | 15.996| 15.686 | 14.620 | 15.308 |
| 2   | Sensitivity                | 16.185| 16.357 | 15.480 | 15.618 |
| 3   | Error rate                 | 04.666| 04.987 | 11.029 | 15.635 |
| 4   | PPV                        | 15.876| 15.600 | 15.102 | 15.618 |
| 5   | NPV                        | 16.271| 16.409 | 15.067 | 15.308 |
| 6   | Accuracy                   | 16.082| 16.013 | 15.084 | 15.635 |

Figure 5: Specificity value on Colorectal Cancer CCG1.11 Dataset. The PSO Training Method shows the highest accuracy.

Figure 6: Sensitivity value on Colorectal Cancer CCG1.11 Dataset. The QPSO Training Method shows the highest accuracy.
A New Way To Prevent Colorectal Cancer Using Supervised Learning Technique

Figure 7: Error rate on Colorectal Cancer CCG1.11 Dataset. The PSO shows the lowest error.

Figure 8: Correction rate value on Colorectal Cancer CCG1.11 Dataset. The PSO Training Method shows the highest accuracy.

Here, we can conclude that classifier outcomes from training the SVM with Particle Swarm Optimization shows improved performance i.e. it shows best area under the curve. From ROC curve: i) the upper point (1, 1) represents positive classification and the point (0, 1) indicates perfect classification. ii) The lower point (0, 0) signifies no positive classification, such type of classifier obligates no false positive errors. The classifiers which are appearing LHS of ROC curve make the positive classification, which means make some false positive errors and low true positive values also. The classifiers which are appearing RHS of ROC curve make the positive classification weak evidence, which means make high false positive errors and correctly classifies all positives.

6. Conclusion

Colorectal cancer recognition is exact sizeable within the subject of clinical field in addition to Bioinformatics. The diagnosis of colorectal cancer as early is safe to deal with the affected person. To perceive and deal with this form of most cancers, Colonoscopy is implemented commonly. Several danger prediction models for colorectal cancer have been developed and validated in different populations but colon cancer effecting the young adults. In this research, we projected a Supervised Learning Technique for detecting colorectal cancer in high dimensional data. One of the important and very popular tool for performing the machine learning tasks that includes novelty detection, classification or regression is Support vector machine (SVM). Training the SVM requires large quantity of quadratic programming. Due to memory constraints conventional methods are not directly applied. To overcome these inadequacies, we introduced, Least Square (LS), Particle Swarm Optimization (PSO), Quadratic Programming and Quantum-behave PSO methods for training SVM. To corroborate the competence and
The proficiency of our predictable system, it is developed in open source called NCSS Software. The acquired outcomes of these approaches are verified on a CCG1.11 Colorectal dataset and the classifier outcomes show that improved performance from training the SVM with Particle Swarm Optimization.

References

West, D., Mangiameli, P., Rampal, R. and West, V., 2005. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. European Journal of Operational Research, 162(2), pp.532-551.

Liu, L., 2018, May. Research on logistic regression algorithm of breast cancer diagnose data by machine learning. In 2018 International Conference on Robots & Intelligent System (ICRIS) (pp. 157-160). IEEE.

Yassin, N.I., Omran, S., El Houby, E.M. and Allam, H., 2018. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Computer methods and programs in biomedicine, 156, pp.25-45.

Smith, J.J., Deane, N.G., Wu, F., Merchant, N.B., Zhang, B., Jiang, A., Lu, P., Johnson, J.C., Schmidt, C., Bailey, C.E. and Eschrich, S., 2010. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology, 138(3), pp.958-968.

Janghel, R.R., Tiwari, R. and Shukla, A., 2010."Breast cancer diagnosis using artificial neural network models", in Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on. IEEE, pp. 89–94.

Exarchos KP, Goletsis Y, Fotiadis DI. Multiparametric decision support system for the prediction of oral cancer reoccurrence. IEEE Transactions on Information Technology in Biomedicine. 2011 Aug 18;15(6):1127-34.

Rajkumar G. Intelligent Pattern Mining and Data Clustering for Pattern Cluster Analysis using Cancer Data. International journal of Engineering Science and Technology. 2010;2(12):7459-69.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P, and Poggio, T., 2001. Multiclass cancer diagnosis using tumor gene expression signatures. Proceedings of the National Academy of Sciences, 98(26), pp.15149-15154.

Radan L, Ben-Haim S, Bar-Shalom R, Guralnik L, Israel O. The role of FDG-PET/CT in suspected recurrence of breast cancer. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2006 Dec 1;107(11):2545-51.

Kang DD, Sibille E, Kaminski N, Tseng GC. MetaQC: objective quality control and inclusion/exclusion criteria for genomic meta-analysis. Nucleic acids research. 2012 Jan 1;40(2):e15-e15.

Van Gerven M, Bohte S. Artificial neural networks as models of neural information processing. Frontiers in Computational Neuroscience. 2017 Dec 19;11:114.

Blake C. UCI repository of machine learning databases. http://www. ics. uci. edu/~ mlearn/MLRepository. html. 1998