

Article

Effective Features to Classify Ovarian Cancer Data in Internet of Medical Things

Mohamed Elhoseny^{2,#}, Gui-Bin Bian^{3,#}, SK. Lakshmanaprabu⁴, K. Shankar⁵, Amit Kumar Singh⁶ and Wanqing Wu^{1,*}

¹ CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology-Chinese Academy of Sciences (SIAT-CAS), Shenzhen, China; wq.wu@siat.ac.cn

² Faculty of Computers and Information, Mansoura University, Egypt; Mohamed_elhoseny@mans.edu.eg

³ Institute of Automation, Chinese Academy of Sciences, Beijing, China; guibin.bian@ia.ac.cn

⁴ Department of Electronics and Instrumentation Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India; prabusk.leo@gmail.com

⁵ School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, India; shankarcrypto@gmail.com

⁶ National Institute of Technology, Patna, India; amit_245singh@yahoo.com

* Correspondence: wq.wu@siat.ac.cn; Tel.: +86-755-8639-2090

Abstract: Ovarian Cancer (OC) is a type of cancer that affects ovaries in women, and is difficult to detect at initial stage due to which it remains as one of the leading causes of cancer death. The ovarian cancer data generated from the Internet of Medical Things (IoMT) was used and a novel approach was proposed for distinguishing the ovarian cancer by utilizing Self Organizing Maps (SOM) and Optimal Recurrent Neural Networks (ORNN). SOM algorithm was utilized for better feature subset selection and was also utilized for separating profitable, understood and intriguing data from huge measures of medical data. In supervised learning techniques, the SOM-based feature selection seems to be a tougher challenge because of the absence of class labels that would guide the search for relevant information to the classifier model. The classification approach can identify ovarian cancer data as benign/malignant. The ovarian cancer detection process can be improved by optimizing the weights of RNN structure using Adaptive Harmony Search Optimization (AHSO). The proposed model in this study can be used to detect cancer at early stages with high accuracy and low Root Mean Square Error (RMSE).

Keywords: Ovarian Cancer; Features Classification; Self-Organizing Map; Optimal Neural Networks; Adaptive Harmony Search Optimization; Internet of Things

1. Introduction

IoT alludes to the following collection of internet data which contains trillions of hubs communicating to various articles from minimal widespread sensor gadgets and handhelds to vast web servers and processer groups. Ovarian carcinoma has the most astounding death rate and acts as the main source of death due to gynecologic complications primarily occur because of late

diagnosis notwithstanding the repeated ovarian cancer occurrence in spite of protection through treatment [1]. The sensational distinction in cure between patients with nearby disease and those with inaccessible disease (15%~25%)[2] emphatically demonstrates the requirement for an non-intrusive yet viable test that is pertinent to a huge population predicted to be affected with ovarian cancer in its initial times itself[3]. Despite the fact that ovarian cancer is just the 6th most common female cancer in Singapore, it remains as one of the main sources of death from gynecological malignancies [4].

Thus the hazard level of cancer gets expanded through women's lifetime and it results in few issues, for example, the women with ovarian cancer could never give birth to kids and it makes an issue in the menopause stage amid their lifetime [5]. Thus, the precise ID, solid cancer analyses and the treatment process ought to be enhanced in order to overcome these [6] challenges for the women. At one point in IoMT, the assessment of ovarian cancer got improved by the huge measure of the dataset on the grounds that it has different potential data [7]. In any case, early recognition of ovarian cancer is significant. It is recorded that 13~21% of the women who underwent surgery were observed to have malignant tumors [8]. Notwithstanding every one of these endeavors, early identification of ovarian cancer remains an essential issue. There has been ongoing research in order to recognize novel cancer markers with the help of proteomic design in serum [9].

The exact reason or causal factor for ovarian cancer is obscure, which hampers the emphasis on early recognition of ovarian cancer. The association between diagnosis and survival in ovarian cancer led to reasoning and actions to enhance the results acquired during recognition at initial stages [10]. This detail utilized to settle on more astute choice and furthermore it expends less cost. There are five reasons how big data approach helps in cancer identification process [11]. Among the several unsupervised learning models, SOM Neural Network system was considered to classify the image pixel by its similarity of a particular pattern to the zone in the same class [12]. As a rule, extremely-assorted data, for example, classification of natural items, compound dynamic data, or even clinical parameters can be handled mostly in a similar way [13]. The neural system regularly faces setback on account of two noteworthy reasons, one of which is moderate training methodology which is known as back-propagation over-fitting of training data [14]. The input data set contains ion intensity levels at particular mass/charge esteems for every patient and the objective arrangement of IoMT concludes whether patients are affected with cancer or not.

In view of this same investigation, pathologists additionally observed and compared the varieties of gene articulations in this particular ovarian cancer and the levels of gene articulation in various neurotic stages and conditions [15]. Organic as well as statistical computing can be utilized to characterize the gene articulations for such correlations, especially during dynamic states, when the tumor advances from statistical computing ovarian cancer tissue. From statistical and machine learning section, various classification techniques were used in the cancer classification process, but it has some non-trivial task problem [16]. The gene articulation information is altogether different from any information which these techniques handled already. The supervised learning utilizes the training data set of the classified information to develop a classifier, which can be utilized to classify new data. In two cases, the big data applications exhibit the developing number of features and the developing volume of the input information [17-21].

The manuscript is organized as follows. Section 2 discusses the survey of different literature works about OC classification whereas section 3 depicts the basics of the study and section 4 extravagantly discussed the current proposed approach. Section 5 demonstrates the result analysis of the proposed model followed by the conclusion of the study.

The contribution of work is to identify or classify OC by inventive features and this feature determination process SOM was used in classifier process. Here the optimal classifier (ORNN) was considered in addition and the reason behind this is the proposed classifier acts as the fundamental property of the diversity among the base classifiers to have a decent-performing ensemble and furthermore the adaptive function is introduced to enhance the updating strategy of AHSO.

1.1. Literature Review

In 2018, John Lewis Etter et al. [18] proposed to retrieve the user data from Familial Ovarian Cancer Registry of 2,636 families with different instances of ovarian cancer. The researchers methodically looked at relative frequencies of ovarian cancer among the relatives of men with testicular and non-testicular cancers. Thirty-one families with instances of both ovarian and testicular cancer were distinguished. Among men with cancer, those with testicular cancer probably have a mother with ovarian cancer than those with non-testicular cancers (OR = 3.32, $p = 0.004$). Zero-fatherly grandmas of men with testicular cancer had ovarian cancer.

In 2017, BeantKaur et al. [19] presented an algorithm which was utilized to feature the extraction strategy with the help of SIFT algorithm. For any question, there are numerous features and intriguing focuses on the protest that can be extricated to give and feature a depiction of the question. The hereditary algorithm was used to upgrade the removed feature with the assistance of fitness function. The fitness function relies on three parameters i.e., each feature; add up to features and classification mistake rate. The location of the ovarian cancer and its stages were discovered utilizing a convolution neural system. The accuracy accomplished with CNN classifier was 98.8% and with SVM, it was 85.01%. The execution parameters utilized were sensitivity, specificity and accuracy.

Error-guided Artificial Bee Colony (EABC) algorithm was proposed by Aditya Singh et al. in 2017 [20]. EABC was assessed for true issue of neural system training and EABC-trained neural system was utilized to identify ovarian cancer from mass spectrometric data collected from blood tests. From the outcomes, it was inferred that EABC displays a significantly better execution in finding the enhanced neuron weights of the neural system when compared with other training algorithms.

In 2017, Fatma Ben Aissa et al [21] proposed another unsupervised strategy in view of the Self-Organizing Map (SOM) for features extraction. This technique comprised of two fundamental advances such as extraction of sub-locales of a picture as indicated by their purposes of intrigue and utilization of it for diverse perspectives of a picture, for example, shading, surface, and shape. At one point, both were consolidated to have a long last «Multi-View» vector trademark. The proposed technique was assessed using three picture classification datasets such as Cloud, Coil100 and CIFAR-10.

In 2018, Yasodha et al [22] designed a novel approach for recognizing ovarian cancer using joined Self Organizing Maps Immune Clonal Selection (SOMICs) along with Grammatical Evolution Neural Networks (GENN). SOMICs algorithm was utilized for better feature selection and also to expel noteworthy, verifiable and intriguing data from enormous measure of therapeutic information whereas GENN was utilized for the classification procedure. The test happened to have exhibited the examination of the proposed procedure and other characterization systems using three distinct classifiers, for instance, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Feed Forward Neural Network (FFNN). The combined SOMICs and GENN procedure yielded promising results on classification as well as feature selection precision for ovarian cancer dataset with 98.23% classification accuracy and 0.0021% MSE.

In 2016, Pengfei Liu et al [23] demonstrated multitask learning system to mutually learn differently-related assignments. In light of the intermittent neural system, three unique instruments were proposed to offer data for displaying content to undertake particular and shared layers. The whole system was prepared together on each one of these undertakings. The analyses on four benchmark content classification assignments demonstrated that the proposed models can enhance the execution of an undertaking with the assistance of other related errands.

In 2013, Ibtissem, B. et al [24] introduced and explained the detailed process of Harmony Search (HS) algorithm for making an unsupervised clustering of the considered input data. From that point, the researchers assessed the execution of this apparatus by examining the outcomes acquired. These outcomes demonstrated that the legitimacy file decides the suitable number of classes consequently that characterize the data. The investigation, acknowledged with a few legitimacy records, enabled us to locate the best legitimacy list to assess the execution and strength of HS algorithm. The acquired results with this HS demonstrated viability and execution in the image clustering for a given issue.

In 2017, Mario Antunes et al [30] proposed the semantic methodologies for setting affiliation and stretching out the study's unsupervised model to learn word classifications normally. The study answer was assessed against 'Mill operator Charles dataset' and 'IoT semantic dataset removed from a standard IoT organizes 'which accomplished a relationship value of 0.63. Likewise, non-negative cross section factorization can be utilized to find dormant semantic information in distributional profiles and increment exactness.

1.2. Background of the Study

When comparing cancers, ovarian cancer is not to be considered as ordinary single disease rather it includes subtypes with various roots, different risk factors, different genetic mutations, different biological behaviors and different prognoses, and much information remains to be unearthed about it in IoMT. At last, less accentuation has been set on the research work that mainly focused to enhance therapeutic mediations by subtype or on the best way to diminish the morbidity of ovarian cancers. There has been little emphasis given towards understanding survivorship issues and the steady care needs of women with ovarian cancer including IoMT of the physical symptoms of treatment (counting both beginning and chronic continuous treatments) and tending to the psychosocial impacts of determination and treatment. Under different circumstances, the ovarian cancer was characterized through the utilization of various techniques in the research study. Major ovarian carcinoma subtypes are named on the basis of how intently the tumor cells take over the ordinary cells that line the diverse organs in the female genitourinary tract in IoMT. A classification system in view of tumor histogenes is would be a legitimate option even though the histogenes is of some ovarian tumors is questionable.

2. Materials and Methods

Nowadays, cancer is the leading research zone in medical field. Correct predictions of different tumor types possess high value in conceding better treatment and reducing destructiveness in patients' health. The current imaginative proposed work aims at ovarian cancer identification process and for the most part, all the medical data were gathered from health informatics which has huge data stored up. This OC database in informatics has huge data gathered from various sources and available for utilization in the detection and process-identification of ovarian cancer.

The dataset comprises of different quantities of features and has missing data and noise for better OC location. A few features (attributes from the database) were chosen by utilizing SOM unsupervised learning procedure which is used to find a collection of structures in a dataset. In any case, the data set possess immense number of data which is hard to pre-process. So the feature subset must be selected before handling the features. From the chosen features to supervised learning approach, it was understood that RNN classifier was utilized to recognize OC disease as malignant or benign. The proposed model should be improved for performance so that the ACS optimization can be utilized to optimize the weights in RNN structure and from the optimal structure, the data can be classified. Some valuable data can be picked from classification procedure which remains as the reason for qualities that work as a gathering in settling the cancerous tissues. This system is graphically portrayed in figure 1, and this feature selection and classification methods strategy are discussed in detail below.

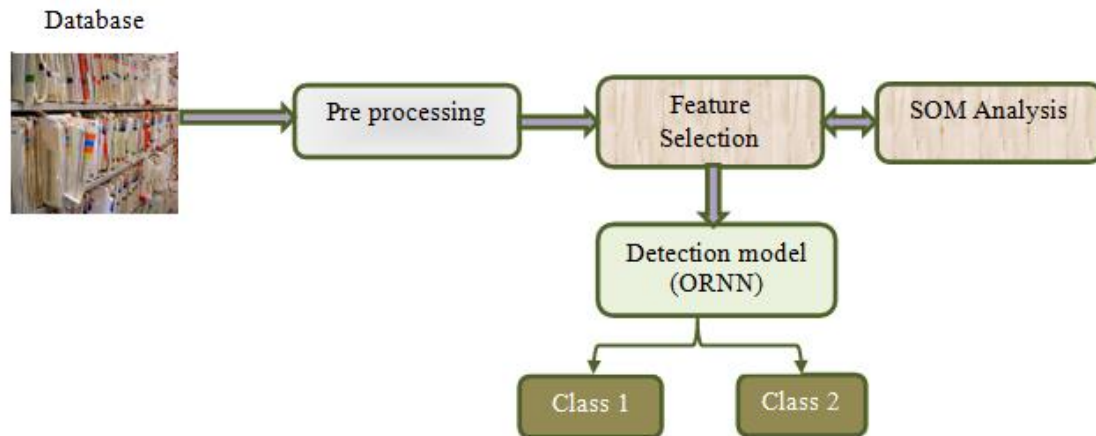


Figure 1. Block diagram for cancer detection

2.1. IoMT Analytics for Health Informatics

Medical data is a tremendous collection of information which is typically communicated via utilizing the speed, veracity, volume, assortment and value terms. In this manner, enormous data approach is utilized to store the health informatics that can be utilized for disease detection and the treatment via IoT. Suitable utilization of huge data for social care has a colossal guarantee and could be connected to various phases of research in cardiology. For example, huge scale populace health administration, cardiovascular disease risk prediction, accurate medicine that utilizes genomic data, and clinical choice support through machine-learning algorithms. Likewise, the current proposed work additionally aimed at examining the huge data in health informatics so as to furnish the distinctive assortment of data with ease and high calibre.]

2.2. Missing values removal

During pre-processing, a large number of records from the medical dataset were incorporated into one social database table and the records which did not coordinate with the format were erased from the record set. In the event that any estimations of the data type or idea drop out of the scope of the normal value, those records were erased. The wrong and missing data from the records were also erased. Evacuation of missing attributes [25] or occasions that expel the cases and have no (or) missing characteristic qualities were likewise finished.

2.3. Feature subset selection

Feature selection systems depend on a selection of factors and do not adjust their unique space. Interestingly, feature extraction strategies search for a mapping either straight or nonlinear to the first feature space into an anticipated space and for the most part with bring down measurement and more viable in depicting the feature. The critical issue of feature selection is to provide the factors which give the 'best' homogenous clustering. Accordingly, the proposed heuristic measure comprises of selecting the most vital set of variables. Feature selection algorithm tends to the issue of selecting input factors that give the best model speaking to the issue and SOM procedures were utilized.

2.3.1. Self-Organizing Maps

SOM presented in the literature [26] was utilized for representation and clustering high-dimensional databases. Perception, as projected on SOM strategies, was used with deformable discrete cross section to make an interpretation of the data similitude into unique

connections. The execution of the SOM was assessed in feature extraction by utilizing the time arrangement created from known examples along with sensitivity tests under different tuneable parameter decisions to choose qualities. In SOM strategy, the discussed learning is an unsupervised realization where the system does not use the class membership for test training rather utilizes the data in a group of neurons to change the neighborhood parameter. The imperative issue of feature selection is to supply the factors which provide the ‘best’ homogenous clustering. In this manner, the heuristic measure comprises of selecting the essential set of variables which give the ‘best’ map collected from the SOM algorithm. For choosing the features, initially random values were selected weights and the data was found based on this calcite distance factor of all processing elements followed by the updating of the weights.

The topological ordering protection implies that the neighbours’ test points in the input space must be appointed to a similar unit or to neighbours units. This can be accomplished by presenting the area imperative in the k-clustering function characterized as follows:

$$F_{(p,s)} = \sum \sum s_{ij} * [h_j * d^2(x_i, s_i)] \quad (1)$$

Where d and s represents segment matrix and matrix whose rows are the clusters, h_{ij} represents the estimation of the area work between the units j .

$$s = \arg \min_i \{ \| h_i - x_i \| \} \quad (2)$$

This algorithm was carried on the technique in light of K-means clustering method. The input vectors instated for the training process, at that point, look at the separation between every vector as is figured by utilizing the Euclidean Distance (ED).

2.3.2. Euclidean distance (ED)

It is the most widely recognized metric for estimating two vectors and the value is given by the square root of the summation of the squares of the distinctiveness between vectors part.

$$ED = \sqrt{\sum_{i=0}^n (I_{fi} - M_{fi})^2} \quad (3)$$

where I_{fi} and M_{fi} are the feature vectors of input and mean features in qualities individually. If the Euclidian Distance is less between two vectors of features, then it implies that the most plausible coordination is there and it will be chosen.

Steps:

- (1) Recognize the attribute data from OC database.
- (2) Assign the underlying weights for every quality in the data.
- (3) Randomly select k focuses on centroids and selects the number of clusters.
- (4) For each point, allocate the point to the closest cluster.
- (5) Calculate comparability between weights-based clustering in data by ED as per condition (1).
- (6) Mapping each comparable data into single clusters in view of ED.
- (7) Pick the minimum distance clusters for feature selection process.

At that point, the maximum dissimilarity features or data are wiped out and the optimized features were chosen as discussed above about the SOM algorithm. The selected subsets must be pre-handled for expelling noise from the data set which will be useful amid the classification for acquiring the most noteworthy classification accuracy. The ‘t’ of features is recognized in light of the positioning of execution retrieved from actualizing the classifier on every single discovered subset.

2.4. OC Detection Process

Detection of cancer through gene articulation data is notorious to grasp the keys for demonstrating the fundamental inconveniences encountered during cancer investigation and drug identification. The neural network was optimized by recurrent function based on context HS algorithm. In light of the feature information ^[27], ORNN classifier was considered to distinguish a tumor, malignant or benign. These models could be adjusted for a reason that it changes over clinical occasion successions and related time-stamped data into pathways pertinent to early recognition of disease. At present, in the field of cancer electronic system that utilizes ultrasound images, a large portion of the work focuses on detection as well as classification. The future scope of this work is to support the recognition of abnormal tissue with the help of resolution ultrasound imaging devices.

2.4.1. Recurrent Neural Network (RNN)

Neural network is the proposed algorithm in which there will be neural network for each stage. The code takes the favored image and actualizes it through neural systems one by one and each time results in a score. RNN is an extended traditional model of feed-forward neural network. An input succession was assumed with the hidden vector combination ^[28] and an output vector. This classification procedure considered 'input' as 'feature subset' and its structure is illustrated in figure 2.

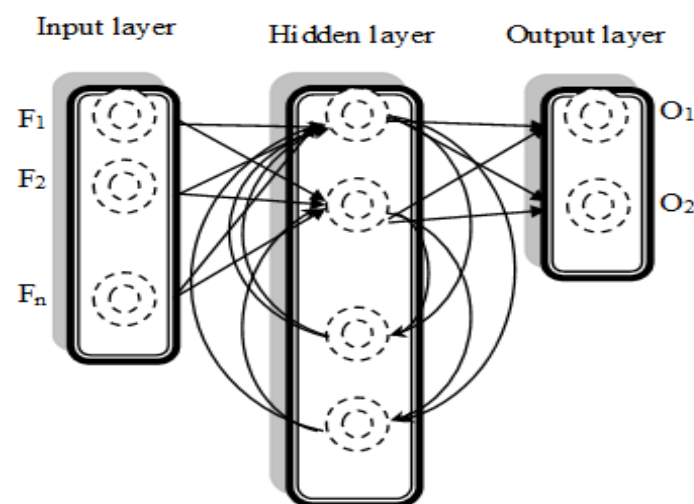


Fig 2: RNN Structure

Unlike feed-forward neural systems, RNN has cyclic affiliations which make it extreme for modeling successions. RNN performs effectively in typical networks yet when network size becomes better, the computation turns out to be more intricate and this way makes RNN more inclined towards focusing local minima. It consumes more memory at the time of experimentation because of additional association weights.

2.4.2 Adaptive Harmony Search Optimization (AHSO)

Every artist (musician) is similar to every choice variable; the gathering of people's aesthetics are undifferentiated from the objective function ^[29]; the pitch scope of a melodic instrument relates to the value scope of the decision factors; the artists' act of spontaneities can be compared to both local and global search plans; melodic harmony at certain times is compared to an answer vector with certain emphasis. For decreasing error rate of RNN, adaptive function was utilized and here the study considered crossover rate to improve the function of HS approach.

2.5. Optimal RNN for Classification

While initialization can undeniably affect the achievement of the strategy, it by and large can utmost postpone the issue of gradient explosion during optimization. Any local search strategy is characteristically attached to some thought of geometry over the inquiry space. RNN was created to settle linear and nonlinear constrained optimization issues in the most recent decade. Contrasted with conventional numerical optimization algorithms, neural network approach has few potential benefits progressively and the flowchart of AHSO is illustrated in the figure 3.

Step 1:

Initialization: Assume input features as input sequence, hidden layer sequence and finally the output layer sequence vector values. Then the optimization parameters such as Harmony Memory Size (HMS) or the number of solution vectors in the harmony memory; Harmony Memory considering Rate (HMCR), HMCR, Pitch Adjusting Rate (PAR) are to be initialized.

Step 2:

Weight assignment: Epoch, the weights, and biases are introduced with CS and then those weights are assigned to RNN. The weights in NN were evaluated. In the following cycle, AHS is updated with weights with the best possible solution and CS continues its search for the best weights until either the last cycle/epoch of the network is attained or the MSE is accomplished.

$$D = \sigma(w_{xh}x_t + w_{hh}h_t + b) \quad (3)$$

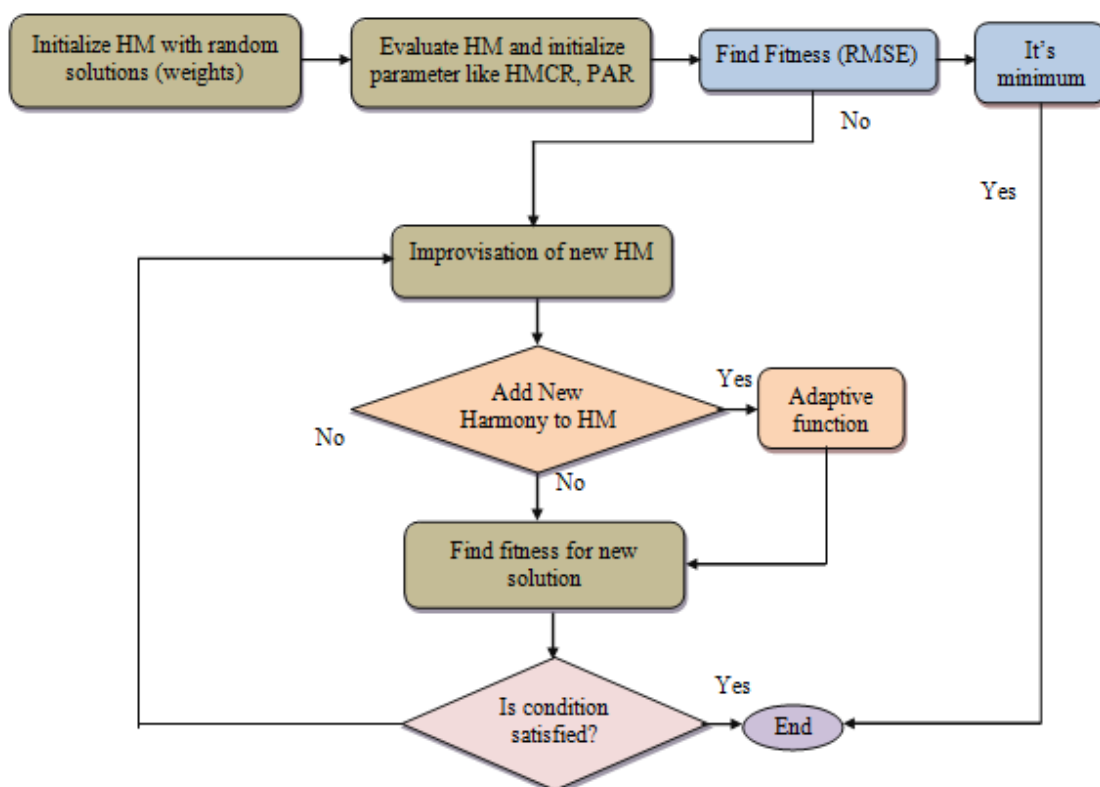


Figure 3: Flowchart for AHSO

Step 3:

Objective function: Estimate the fitness value of each and every harmony result with the help of equation (5) and then compute the finest result values.

$$RMSE = \sqrt{\frac{1}{N} \sum_j (A_j - P_j)^2} \quad (5)$$

In the Above equation σ as a nonlinearity function, w weight matrix, and b is a bias, N is the total number of prediction, A_j and P_j and are the initial and anticipated data separately. Since each hidden layer has its own weights and activations, they act like self-ruling. By stating so, the objective is to recognize the association between dynamic data sources. Here, the weights and bias of these layers are exceptional.

Step 4:

Improvise a new harmony: a new harmony vector was made in perspective of three norms, for instance, memory thought, pitch change, and random selection. The HMCR is the rate of opting one incentive than other old qualities stowed in the HM. However, $(1-HMCR)$ is the rate of arbitrarily picking one value from the probable range of values.

$$w_i' = \begin{cases} w_i' \in \{w_i^1, w_i^2, \dots, w_i^{HMS}\} \text{ with prob } HMCR \\ w_i' \in w \text{ with prob } (1-HMCR) \end{cases} \quad (6)$$

For instance, an HMCR value of 0.90 assigns that the HS algorithm will show the decision variable value from truly stored qualities in the HM with 90% probability or from the total conceivable range together with (100~90) % probability. These operations use the PAR parameter i.e., the rate of pitch modification as below:

$$w_i' = \begin{cases} \text{Adjusting pitch with probability} \\ \text{Doing nothing with probability } (1-PAR) \end{cases} \quad (7)$$

The value of (1-PAR) groups the rate of being idle. When the pitch alteration decision for is YES w_i' , is substituted as follow:

$$w_i' = w_i' \pm rand \times bw \quad (8)$$

where bw indicates an arbitrary distance bandwidth, $rand$ represents a random number and its selection is based on the velocity updation of the crossover rate parameters.

Step 5:

Update harmony memory: for each and every novel value of harmony, the value of the objective function is considered. When the new harmony vector is improved than the worst harmony in HM, the new harmony is encompassed in the HM whereas the prevailing worst harmony is left out from the HM.

Step 6:

Adaptive Function (crossover): there is no crossover function between two input weight solutions. The crossover operation yields the superlative 'z' chromosomes. With at the 'n' point crossover, 'n' cut points are arbitrarily selected within the strings while the n-1 segments between the 'n' cut points of the two parents are swapped. The function of genetic operator is to create the best offspring to produce a new population of the adaptive process.

$$Adaptive_crossover = \begin{cases} P_1(f_{max} - f_x) / (f_{max} - f_{avg}), f_x \geq f_{avg} \\ P_3, f_x \leq f_{avg} \end{cases} \quad (9)$$

Here f_{min} , f_{max} and f_{avg} are the maximum, minimum and average fitness of populations. Then $p1$ and $p2$ range between (0, 1). The optimization procedure-based optimization of weights occurs in RNN to compute Root Mean Square Error (RMSE) function of distinguished data. Thus every one of these layers carries on autonomously and cannot be joined together.

To yield a proper classification result, the weights of the RNN must be upgraded for some execution measure. Supervised learning (i. e., altering the association weights with the end goal that there is an increase in the correspondence between the network outputs and desired target data) gives a reasonable method for adjusting the weights and accordingly, the dynamic behavior of the RNN. To perform the classification effectively, reasonable

progression speaking to the pertinent attributes of the time arrangement must be developed in RNN. These assembled models were assessed and there is a need to distinguish the accessible data into training and testing data. The training data is utilized to fabricate the model, and the test data is used to assess it.

3. Results

The proposed classification was actualized in the working stage of MATLAB 2016 with the system setup, i5 processors with 4GB RAM.

Table 1. Performance measures

| | |
|---------------------------------------|---|
| Root Mean square (RMS) | $RMS = 1 - \sqrt{\frac{\sum_{i=1}^m (T_i - P_i)^2}{N}}$ |
| Mean Absolute Error Percentage (MAPE) | $MAPE = 1 - \left(\frac{\sum_{i=1}^m (T_i - P_i)^2}{N} \times 100 \right)$ |
| Sensitivity | $Sen = \frac{TP}{TP + FP}$ |
| Specificity | $Spc = \frac{TN}{TN + FP}$ |
| Accuracy | $Acc = \frac{TP + TN}{TP + TN + FP + FN}$ |

The OC detection proposed model result was assessed with the help of few measurements like RMSE, MAPE, accuracy, sensitivity, and specificity. These values were compared with existing classification methodologies and database investigated in this area (Table 1).

3.1. Database description

The objective of this examination is to recognize proteomic patterns in serum that distinguish ovarian cancer from non-cancer. This examination is critical to women who have a high danger of ovarian cancer because of familial or individual history of cancer. The proteomic spectra were created by mass spectroscopy and the data set used here is 6-19-02, which incorporates 91 controls (Normal) and 162 ovarian cancers. The raw spectral data of each example contains the relative adequacy of the force at each molecular mass/charge (M/Z) personality.

The data compared two sorts of patients, i.e. cancer and control patients. Cancer patients are the patients who have cancer and control patients are typical patients i.e., non-cancerous. The input set of data contained particle force levels at a particular mass/charge esteems for every patient and the objective set of data characterized whether patients have cancer or not. Figure 4 shows that since the errors are squared before, the mean value was found whereas RMSE gives a moderately high weight to huge errors. RMSE value can be contrasted with actual and SOM-based chosen features (proposed) with different preliminaries.

Figure.5 demonstrates the error rate (RMSE and MAPE) for various optimizations in RNN ideal weight process outcomes and the strategy, for example, Error-guided Artificial Bee Colony (EABC), Genetic Algorithm (GA), HS and AHSO. The classification error was limited in the event in such a way that the limit between classified classes was

indistinguishable to the Bayesian limit. Trial results on simulated information represented the productivity of the current study's way to deal with minimizing the MAPE through classification of what likewise guarantees its proficiency in OC detection. Despite the fact that MAPE is fundamentally the same as the RMSE, it is less touchy to substantial forecast errors. The least error rate 0.085 of optimization in AHSO differed with other optimizations. This optimal RNN technique introduced the best anticipating execution, yet every one of the three techniques showed surprising outcomes and affirmed that they are sufficient enough for time prediction.

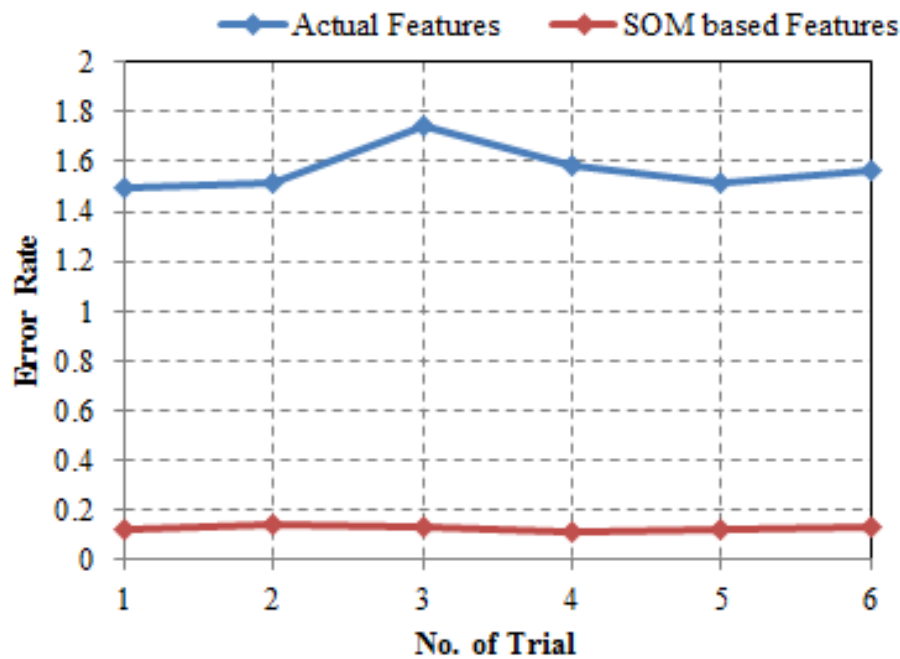


Figure 4. Trial vs RMSE

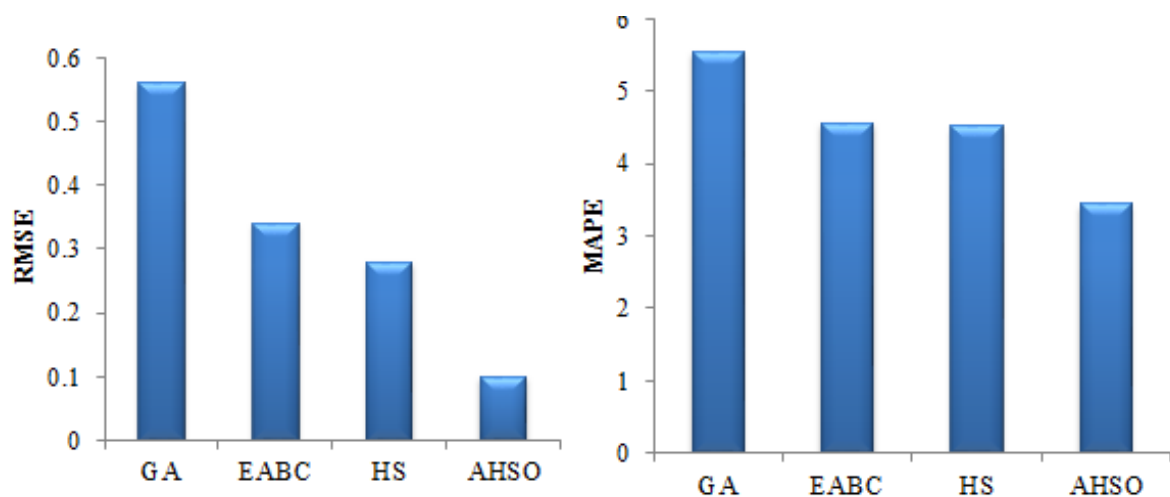


Figure 5. Error Rate vs Optimization

Table 2. OC detection analysis for proposed classifier (ODNN)

(a) 90% Training -10% Testing

| Validation's | Accuracy | Sensitivity | Specificity | RMSE | MAPE |
|--------------|----------|-------------|-------------|------|------|
|--------------|----------|-------------|-------------|------|------|

| | | | | | |
|--------------|-------|-------|-------|-------|------|
| Validation 1 | 96.23 | 89.56 | 82.22 | 0.015 | 2.58 |
| Validation 2 | 94.55 | 84.5 | 89.56 | 0.08 | 3.56 |
| Validation 3 | 96.23 | 86.22 | 85.5 | 0.028 | 2.28 |
| Validation 4 | 94.5 | 91.2 | 90.2 | 0.087 | 0.89 |

(b) 80% Trinaiang-20% testing

| Validation's | Accuracy | Sensitivity | Specificity | RMSE | MAPE |
|--------------|----------|-------------|-------------|-------|------|
| Validation 1 | 92.22 | 81.22 | 79.89 | 0.078 | 3.25 |
| Validation 2 | 89.23 | 80.2 | 81.22 | 0.099 | 4.56 |
| Validation 3 | 92.22 | 84.56 | 80.2 | 0.09 | 4.59 |
| Validation 4 | 92.8 | 85.5 | 85 | 0.11 | 0.98 |

(c) 70% Trinaiang-20% testing

| Validation's | Accuracy | Sensitivity | Specificity | RMSE | MAPE |
|--------------|----------|-------------|-------------|-------|-------|
| Validation 1 | 78.2 | 68.8 | 66.23 | 0.11 | 4.56 |
| Validation 2 | 76.2 | 82.22 | 72.2 | 0.089 | 3.78 |
| Validation 3 | 74.22 | 75.55 | 79.22 | 0.28 | 4.596 |
| Validation 4 | 80.2 | 79.2 | 80.2 | 0.22 | 3.563 |

Table 3. Overall OC detection Results

| Testing Data (%) | Sensitivity | Specificity | Accuracy | RMSE | MAPE |
|------------------|-------------|-------------|----------|-------|------|
| 20 | 70.89 | 82.22 | 95.22 | 0.015 | 2.35 |
| 40 | 72.56 | 88.5 | 96.22 | 0.011 | 3.48 |
| 60 | 82.22 | 84.2 | 92.22 | 0.03 | 2.48 |
| 100 | 85.2 | 85.2 | 96.33 | 0.02 | 3.55 |

For each validation, random experiments were repeated more than 80times independently. In the classification process, the computation of average value remains essential. The table 2, 3 and 4 discussed the classification performance of the proposed model; (a) shows the analysis of data to be 90% training and 10% testing data whereas the performance was 92% to 97% in accuracy with similarity in other measures. Utilizing this approach, a classifier with 80% accuracy in segregating patients with ovarian cancer from patients with favorable infection and healthy controls from a blinded test set was produced.

Each one of the tables depicted the execution of ODNN classifier for OC recognition. All these tables were compared to prove 80% training and 20% testing data as better performances.

Initially, the information weights were analyzed and the lowest weight was eliminated because the lowest rank data does not provide any valuable information during classification. A specificity of 80% and sensitivity of 88.6% were obtained. RMSE and MAPE were also found to be minimum in this process only and generally if taken for maximum training data, the testing results were found to be best. The information utilizing statistical minutes preprocessing got preferable exactness over those without preprocessing and those in reference. In this way, the utility of the strategy was proposed in proteomics information investigation. From the reduced MSE, the proposed system classified the immense data collected with most elevated sensitivity as well as specificity rate.

Convergence of different optimizations is shown in the figure 6. A minimum RMSE was attained in AHSO with minimum iterations. The graph fundamentally tackles the AHSO process and offers minimum fitness with maximum iteration. It is visible from the graph that the minimum classification is realized in AHSO technique i.e., 0.2397 in the initial iteration. The convergence diagram starts with cycle fitness value is high in view of the objective function of the algorithm which, at that point, minimizes the error values. The chart successfully outfits the optimal fitness value delineating the stunning results of the GA, HSO. Table 5 shows OC detection result for the proposed classifier. Here, the performance measures such as sensitivity, specificity, accuracy, and RMSE were evaluated for testing data 20, 40, 60, and 100%. The proposed classifier detected cancer with an optimal value of 85.2% sensitivity, 85.2% specificity, 96.22 %accuracy and 0.02 as RMSE.

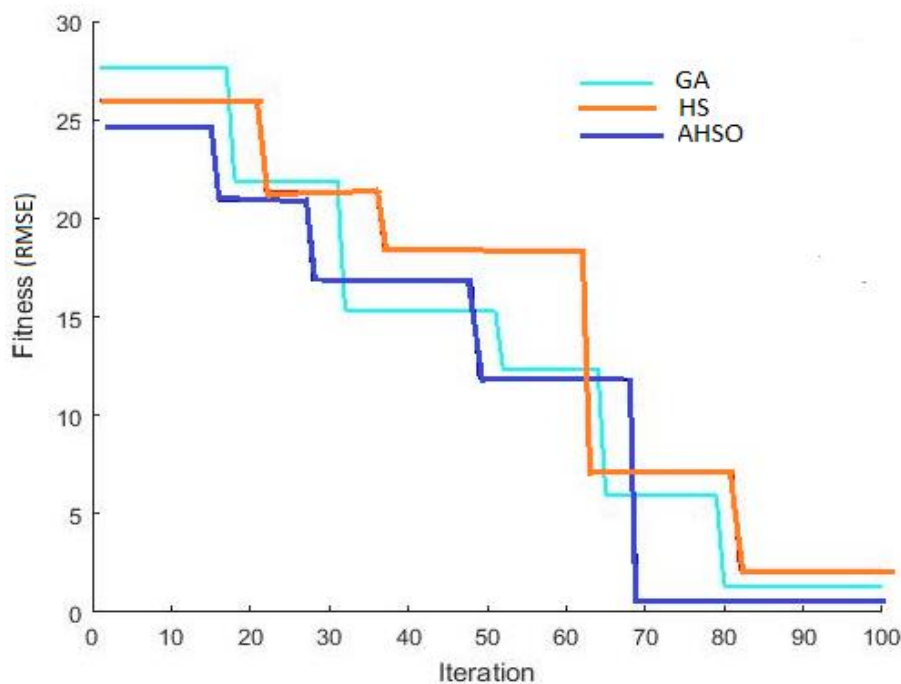


Figure 6. Convergence Graph

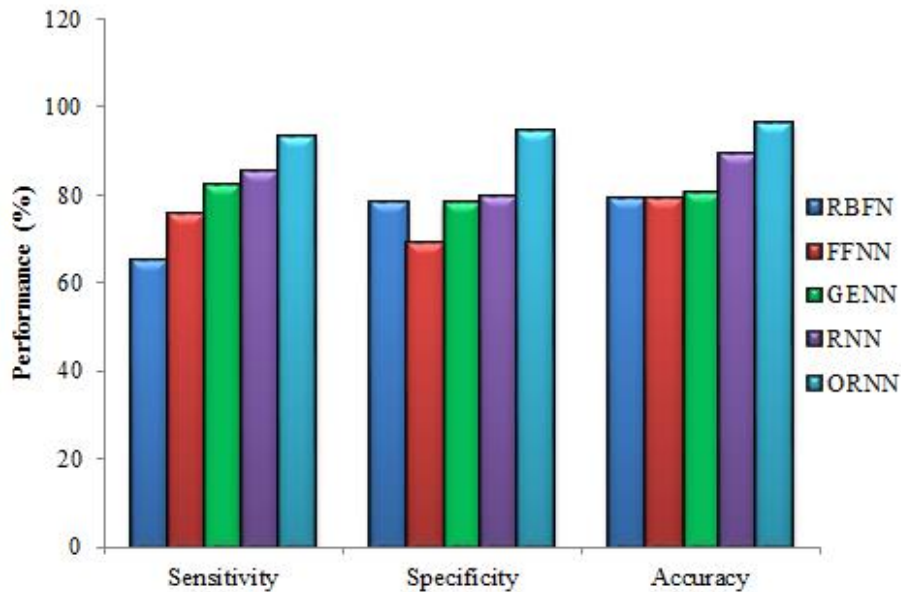


Figure 7. Comparative analysis

Figure 7 shows the comparison chart for different classifiers. The performance measures were compared with different classifiers and the optimal one was found. This chart explains that ORNN improves the performance of the proposed model. It reaches the optimal value when compared with RNN, RBFN, FFNN, and GENN. From Table 3, it was observed that the classification accuracy of the ORNN algorithm tends to be better in five datasets out of the nine. Based on SOM features, the classification was analyzed in cancer detection model. From the analysis, the general accuracy was 96.27% while the ORNN the model attained a result with general sensitivity 80.03% and similarly the specificity measure. At last, the NN system with the proposed classifier achieved highest classification accuracy.

6. Conclusion

A different approach for the detection and classification of OC is developed in this research work. From the extensive data set, the best subset features were chosen and assembled by utilizing SOM strategy in IoMT data. In this manner, the performance of the framework was assessed utilizing the performance measurements such as sensitivity, specificity, accuracy, and RMSE. The outcome exhibited that the proposed model accomplishes most extreme accuracy and it reduces the RMSE value with the correlation of existing models. The SOM and ORNN technique provided promising outcomes on classification and feature selection accuracy for ovarian cancer detection with 95% accuracy and 0.019% RMSE. The optimization (AHSO) also produced better error rate with optimal classification process. Additionally, the proposed procedure seems to be totally robotized, precise, and can be effortlessly composed as a product application to use in any PC. The proposed approach, in this manner, would enhance the general public health by expanding human lifespan, diminishing death rate in the group and guaranteeing individuals life more beneficial and free living.

ACKNOWLEDGEMENT

This research is supported in part by the Shenzhen Governmental Basic Research Grant (JCYJ20160331185848286, JCYJ20170413170301569), the National Natural Science

Foundation of China (61873349), Guangdong Province Natural Science Fund (2016A030310129), the Guangzhou Science and Technology Planning Project (201704020079, 201803010093), CAS President's International Fellowship for Visiting Scientists (2017VTA0011)

REFERENCES

1. Kozak, K.R., Su, F., Whitelegge, J.P., Faull, K., Reddy, S. and Farias - Eisner, R. Characterization of serum biomarkers for detection of early-stage ovarian cancer. *Proteomics* **2005**, 5(17), pp.4589-4596.
2. Manogaran, G., Lopez, D. and Chilamkurti, N. In-Mapper combiner based Map-Reduce algorithm for big data processing of IoT based climate data. *Future Generation Computer Systems* **2018**, 86, pp.433-445.
3. Badgwell, D. and Bast Jr, R.C. Early detection of ovarian cancer. *Disease markers* **2007**, 23(5, 6), pp.397-410.
4. Asuntha, A., Brindha, A., Indirani, S. and Srinivasan, A. Lung cancer detection using SVM algorithm and optimization techniques. *Journal of Chemical and Pharmaceutical Sciences* **2016**, 9(4), pp.3198-3203.
5. Yasodha, P. and Ananthanarayanan, N.R. Analysing big data to build knowledge-based system for early detection of ovarian cancer. *Indian Journal of Science and Technology* **2015**, 8(14), pp.1-7.
6. Alzubaidi, A., Cosma, G., Brown, D. and Pockley, A.G., A new hybrid global optimization approach for selecting clinical and biological features that are relevant to the effective diagnosis of ovarian cancer. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on* (pp. 1-8). IEEE.
7. Tan, T.Z., Quek, C., Ng, G.S. and Razvi, K. Ovarian cancer diagnosis with complementary learning fuzzy neural network. *Artificial intelligence in medicine* **2008**, 43(3), pp.207-222.
8. Ganesan, N., Venkatesh, K., Rama, M.A. and Palani, A.M.O. Application of neural networks in diagnosing cancer disease using demographic data. *International Journal of Computer Applications* **2010**, 1(26), pp. 76-85.
9. Conrads TP, Fusaro VA, Ross S, Johann D, Rajapakse V, Hiitt BA, Steinberg SM, Kohn EC, Fishman DA, Whitely G, Barrett JC. High-resolution serum proteomic features for ovarian cancer detection. *Endocrine-Related Cancer* **2004**, 11(2), pp.163-78.
10. Giftlin. and Vinolia Anandan, Relevant Studies On Different Types Of Cancer Using Neural Network: A Survey. *International Journal of Recent Trends in Engineering & Research* **2016**, 2(5), pp.89-97.
11. Woolderink, J.M., De Bock, G.H., de Hullu, J.A., Hollema, H., Zweemer, R.P., Slangen, B.F.M., Gaarenstroom, K.N., van Beurden, M., van Doorn, H.C., Sijmons, R.H. and Vasen, H.F.A. Characteristics of Lynch syndrome associated ovarian cancer. *Gynecologic Oncology* **2018**, 150(2), pp.324-330.
12. Vlahou, A., Schorge, J.O., Gregory, B.W. and Coleman, R.L. Diagnosis of ovarian cancer using decision tree classification of mass spectral data. *BioMed Research International* **2003**, 2003(5), pp.308-314.
13. Ji, L.L., Larregieu, C.A. and Benet, L.Z. Classification of natural products as sources of drugs according to the biopharmaceutics drug disposition classification system (BDDCS). *Chinese journal of natural medicines*, **2016**, 14(12), pp.888-897.
14. Wu, H.C. and Huang, S.H.S. User behavior analysis in masquerade detection using principal component analysis. In *Intelligent Systems Design and Applications. ISDA'08. Eighth International Conference* **2008**, pp.201-206.
15. Tempany, C.M, Zou, K.H, Silverman, S.G., Brown, D.L, Kurtz, A.B. and McNeil, B.J. Staging of advanced ovarian cancer: comparison of imaging modalities—report from the Radiological Diagnostic Oncology Group. *Radiology* **2000**, 215(3), pp.761-767.
16. Raad, A., Kalakech, A. and Ayache, M. Breast cancer classification using neural network approach: MLP and RBF. The 13th International Arab conference on information Technology **2012**, pp.15-19.
17. Isa, N.A.M., Hamid, N.H.A., Sakim, H.A.M., Mashor, M.Y. and Zamli, K.Z., 2004, December. Intelligent classification system for cancer data based on artificial neural network. *IEEE Conference on Cybernetics and Intelligent Systems (Singapore, Singapore)* **2004**, pp. 196-201.
18. Wanqing Wu, Sandeep Pirbhulal, Heye Zhang, Subhas Chandra Mukhopadhyay , "Quantitative Assessment for Self-Tracking of Acute Stress based on Triangulation Principle in Wearable Sensor System", *IEEE Journal of Biomedical and Health Informatics* **2018**, 2018, pp.1-1
19. Wanqing Wu, Heye Zhang, Sandeep Pirbhulal, Subhas Chandra Mukhopadhyay, Yuan-Ting Zhang, "Assessment of Biofeedback Training for Emotion Management Through Wearable Textile Physiological Monitoring System", *Sensors Journal, IEEE* **2015**, 115(12), pp. 7087-7095

20. W. Wu, S. Pirbhulal, A. K. Sangaiah, S. C. Mukhopadhyay, and G. Li, "Optimization of signal quality over comfortability of textile electrodes for ECG monitoring in fog computing based medical applications", *Future Generation Computer Systems* **2018**,86, pp.515-526.
21. Sandeep Pirbhulal, Heye Zhang, Subhas Chandra Mukhopadhyay, Wanqing Wu and Yuan-Ting Zhang. "Heart-Beats Based Biometric Random Binary Sequences Generation to Secure Wireless Body Sensor Networks", *IEEE Transactions on Biomedical Engineering* **2018**, pp.1-1.
22. Etter, J.L., Eng, K., Cannioto, R., Kaur, J., Almohanna, H., Alqassim, E., Szender, J.B., Joseph, J.M., Lele, S., Odunsi, K. and Moysich, K.B. Hereditary association between testicular cancer and familial ovarian cancer: A Familial Ovarian Cancer Registry study. *Cancer epidemiology* **2018**, 53, pp.184-186.
23. Kaur, B., Mann, K.S. and Grewal, M.K. Ovarian cancer stage based detection on convolutional neural network. *2017 2nd International Conference on Communication and Electronics Systems (Coimbatore, India), 2017*, pp. 855-859.
24. Singh, A. and Kumar, D. Novel ABC based training algorithm for ovarian cancer detection using neural network. *2017 International Conference on in Trends in Electronics and Informatics (Tirunelveli, India), 2017*, pp. 594-597.
25. Aissa, F.B, Sakkari, M., Ejbali, R. and Zaied, M. Unsupervised Features Extraction Using a Multi-view Self Organizing Map for Image Classification. *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), (Hammamet, Tunisia), 2017*, pp. 196-201.
26. Furey, T.S, Cristianini, N, Duffy, N., Bednarski, D.W, Schummer, M. and Haussler, D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **2000**, 16, pp. 906-914.
27. Liu, P., Qiu, X. and Huang, X., 2016. Recurrent neural network for text classification with multi-task learning. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16) (New York, USA), 2012*, pp.2873-2879.
28. Ibtissem, B. and Hadria, F., Unsupervised clustering of images using harmony search algorithm. *Journal of Computer Sciences and Applications* **2013**, 1(5), pp.91-99.
29. Xiong, H., Pandey, G., Steinbach, M. and Kumar, V., Enhancing data analysis with noise removal. *IEEE Transactions on Knowledge and Data Engineering* **2006**, 18(3), pp.304-319.
30. Pratiwi, D., 2012. The use of self organizing map method and feature selection in image database classification system. *International Journal of Computer Science Issues* **2012**, 9(3), pp.377-381.
31. NG, T. and Prasanna, S., Design and Development Of Medical Image Processing Techniques and to Study their Applications Using Graphical System Design in Ovarian Cancer. *International Journal of Engineering and Technology* **2016**, 2(2), pp.1252-1255.
32. Hermans, M. and Schrauwen, B. Training and analysing deep recurrent neural networks. NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems (Lake Tahoe, Nevada), **2013**, pp.190-198.
33. Mahdavi, M, Fesanghary, M. and Damangir, E. An improved harmony search algorithm for solving optimization problems. *Applied mathematics and computation* **2007**, 188(2), pp.1567-1579.
34. Antunes, M., Gomes, D. and Aguiar, R.L. Towards IoT data classification through semantic features. *Future Generation Computer Systems* **2018**, 86, pp.1-34.