Gene based disease prediction and medicine providence through Consortium Reliant Visage Prognostication Model for IoT Health Monitoring

Narahari Narasimhaiah[1], Dr. R.PRAVEEN SAM [2]

[1] Research Scholar, Research and Development Center, BharathiarUniversity, Coimbatore, India.narasimhaiah.narahari@gmail.com

[2]Professor, Department of CSE, G.Pulla Reddy Engineering College, Kurnool, India.rpraveensam.cse@gprec.ac.in

Abstract:

Identifying genes associated with disease plays an extremely important role in the diagnosis and treatment of disease.However, prevailing research carries out only the topological structure of gene that declines the genome frequency and can disclose the inherent properties of disease-genes could increase more computational complexity.In addition, it reduces the population diversity hence those are slow down the classification which leads to overfitting of gene molecules that achieve very low accuracy during prediction.Hence, in this paper efficiently proposed a Disease-Gene Reliant Visage Prognostication (DG-RVP) Model,inorder to predict thediseasewhich contains Double Two Extrication (DTE) to extracts the features that are weighted by the homogeneity scores it strengthens the genome frequency. Once feature extraction completed Quantum Coyote Diacritic (QCD) Algorithm needs to improve feature selection through each subset of features represented the quantized individual search position in the region. To optimize a selected featureCatenation-Adore Emissary based Genetic Algorithm (CAE-GA)is implemented, which avoids the early convergence with familiarizing the genetic operators.Based on thepredicted disease Mutual Filtering Algorithmis included that provide the medicine through taking account of noise and bias from gene expression.The outcome shows the proposed model can predict gene-disease-drug association's superior to futuristic.

Keywords: Disease-Gene Reliant Visage Prognostication (DG-RVP) Model, Quantum Coyote Diacritic (QCD) Algorithm, and Catenation-Adore Emissary based Genetic Algorithm (CAE-GA), Internet of Things (IoT)

1. Introduction:

The Internet of Things (IoT) is one of the rapidly growing trends of ICT. Significant milestones have been made in various application domains, such as e-Health, business, home services,

military, and automation [1],etc. IoT is aimed at interconnecting communication devices in application domains and gathering data to provide consumers with secure and reliable services. In the field of applications for e-Health care, remote connecting devices may exchange information among patients, expert physicians and/or healthcare experts based on Artificial Intelligence (AI). Information exchange via medical [2] devices that are linked to the cloud via the Internet cloud platform for quality services is also referred to as the Internet of Medical Things (IoMT). IoMT is a newly emerging technology [3] for interconnecting medical devices and software applications to improve the effectiveness, reliability, and accuracy of medical devices in the healthcare sector. Approximately 3.7 million medical devices are in use recently, and the IoMT trade is expected to hit 136.8 billion / year worldwide by 2021 [4]. The definition of IoMT is now primarily used for activities such as remote patient monitoring, drug order tracking, and mobile wearable health devices. IoMT also offers a wide range of services to medical experts, such as providing input to medical staff, details on equipment and their configurations [5] as patients and experts require. IoMT offers fast and ultimate access to various reports [6] that assist surgeons during surgery in operation theatres. Ever since the first disease gene was found in 1949, thousands of genes have been recognized as being associated [7] with the disease.

Identifying interactions between disease-genes lets us identify disease pathways, find diagnostic signs and therapeutic goals [8], which further leads to new approaches and medications for treatment. High-throughput technologies generally predict a few hundred candidate genes [9], and validating all of these candidates requires a considerable amount of time and expense. Therefore, a widely used method is to first computationally predict [10] / prioritize candidate genes associated with the diseases under consideration, and then experimentally validate a subgroup [11] of candidates based on the results of computational prediction in order to greatly improve the yield of the experiments.

At present, different types of data have been used to predict interactions of the disease-gene, and networks of protein-protein interaction (PPI) [12] are the most commonly used proof. Previous algorithms attempted to predict interactions between the disease-gene [13] by using the topological structure of PPI networks directly. Universal PPI networks downloaded from online databases, however, generate lots of false positives [14], and the prediction accuracy cannot be

further improved only by using them. Scientists, therefore, appear to combine more data types with PPI networks to predict interactions [15] of the disease-gene.

Over the past three decades, the use of image processing [16] and computer vision-based techniques has been commonly used for screening. In this regard, techniques such as mammograms (X-rays) [17], magnetic resonance imaging (MRI) [18], ultrasound (sonographer) and thermography are generally used to detect and diagnose disease. Nevertheless, biopsy [19] provides the most accurate diagnosis with reassurance.

In the selection of features, a mathematical method may manage unknown problems while maintaining the sense of the attributes. Pawlak's proposed [20] Rough Set Theory (RS) is used as a method for evaluating data dependencies and reducing the number of dataset attributes.

For the evaluation of subset features, rough set theory is an objective for defining and solving unknown problems [21]. In order to apply rough set theory to deal with heterogeneity reduction attributes [22], some researchers have suggested methods of dealing with rough set in the positive field. Build a rough-set [23] variable precision reduction model based on upper and lower approximation. Therefore, from the aforementioned issues and considerations, the proposed technique efficiently tackle the previous issues outperform the gene- disease-drug association in the emerging field of healthcare monitoring.

2. Literature survey:

Zeng et al [24] suggested a novel method called REGENT to combine multiple gene networks with GWAS data to prioritize specific genes associated with diseases. In particular, we leveraged network representation learning, a technique recently developed to analyze social networks, in order to learn compact and stable embedding from multiple gene networks. To integrate this learned gene embedding with the GWAS results, we developed a hierarchical statistical model and derived an effective model estimation and prediction inference algorithm. Using GWAS data from six complex diseases, we showed that REGENT had outperformed existing methods for identifying known disease-associated genes.

Luo et al [25] propose a variety of learning-based approaches to predict disease genes by suggesting that the disease and its related genes should be compatible in a number of lower dimensions. The 10-fold cross-validation experiments show that the area under the receiver operating characteristic (ROC) curve (AUC) produced by our approach is 0.7452 with high-quality OMIM dataset gene-disease associations, which is higher than the PBCF (0.5700)

competing process. 9 Of the top 10, gene-disease associations predicted, current literature can be verified, which is better than the PBCF result (6 of the top 10 associations predicted).

Manogaran et al [26] The Big Data Processing System is proposed in this paper to combine climate and health data, and to find the connection between climate parameters and dengue incidence. This framework is demonstrated in a Hadoop Distributed File System (HDFS) environment, with the help of the MapReduce programming model, Hive, HBase and ArcGIS. The following weather parameters for the study are Tamil Nadu with the help of IoT weather sensor devices and NCEP, such as minimum temperature, maximum temperature, wind, precipitation, solar, and relative humidity.

Chen et al [27] To seek to refine this, propose a new feature reduction mechanism based on a fish swarm algorithm (FSA). The approach is then extended to the problem of identifying suitable subsets of features in the rough set reduction cycle of the neighborhood. To find the optimal subsets and a fitness function to evaluate the best solutions, we define three foraging fish behaviors. We build the neighborhood feature reduction algorithm based on FSA and design some experiments compared to a feature reduction method of a heuristic neighborhood. Experimental results show that the neighborhood reduction method based on the FSA is suitable for dealing with numerical data and more possibility of finding an optimal reduction.

Luan et al [28] A novel Artificial Fish Swarm Algorithm (AFSA)-based attribute reduction algorithm and the rough set are proposed. In the later phase of iterations, AFSA adopts a slow convergence rate, normal distribution function, Cauchy distribution function, multi-parent crossover operator, mutation operator and modified minimal generation gap model to boost AFSA. Based on improved AFSA and rough set, the attribute reduction algorithm takes full advantage of the improved AFSA and rough set, which are quicker, more efficient, simpler, and easier to implement. Datasets to validate the aforementioned new method are collected in the UC Irvine (UCI) Machine Learning Repository.

In [24] does not consider the attribute reduction, [25] have attained the unbalanced genome frequency level, [26] technique with achieved the very lower accuracy and disease prediction only did by one genomic information. [27] Offers very poor performance in population diversity, [28] lesser convergence speed while featuring selection. Consequently from the above mentioned issues in this paper develop a unique model to tackle those issues and predict the disease in healthcare monitoring.

3. Disease-Gene Reliant Visage Prognostication (DG-RVP) Model:

Identifying genes associated with disease plays an extremely important role in the diagnosis and treatment of disease, which requires a convenient optimal solution in real-time depending on the physiological state of the patient, the environment and other relevant data to acquire ample protection or prevention of diseases based on genomic info. In that premature prediction of diseases is one of the primary causes of mortality among humanity. Cloud-based e-Health care systems for early prediction of diseases are emerging at a rapid pace as Information and Communication Technology (ICT) developments such as 5 G technology and Internet of Medical things (IOMT), etc. As for prophesying diseases entrenched in genes massively attains parallel encourages up researchers to fumigate the in-depth exploratory analysis and speed the discovery of potential disease associated with genes.

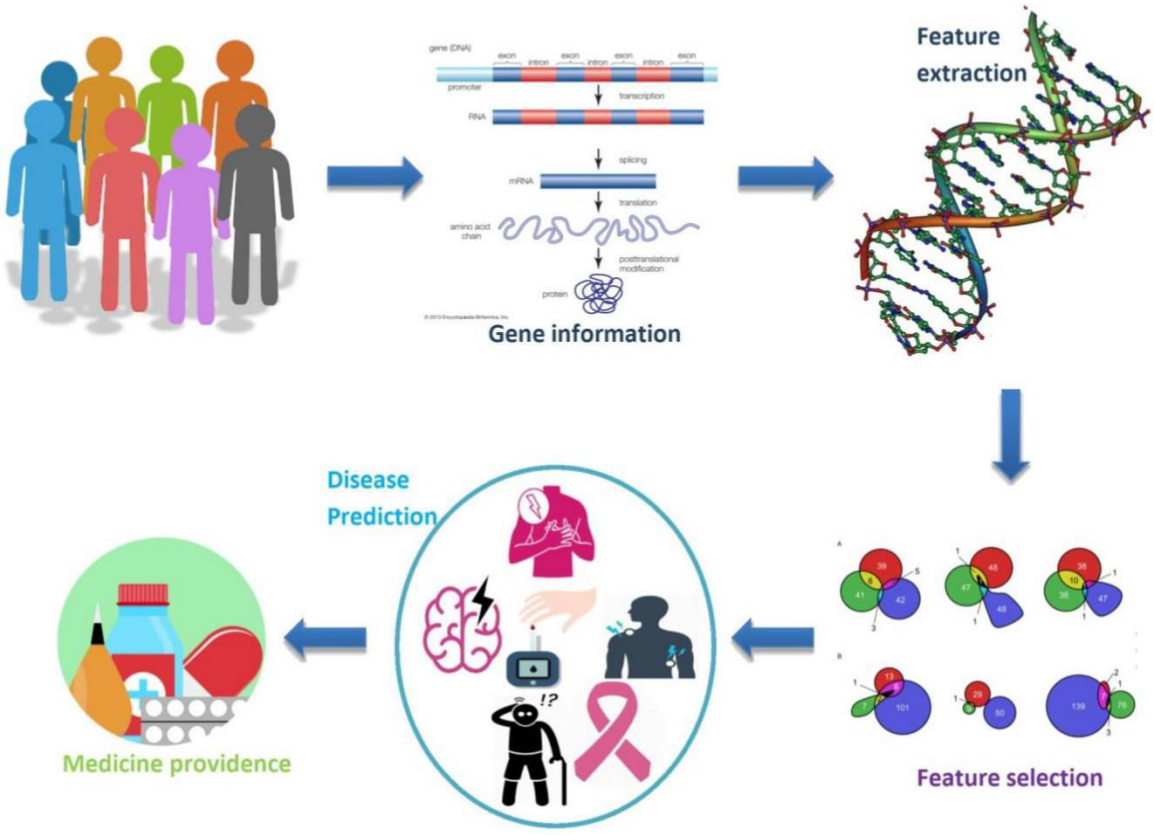


Figure: 1- Architecture of gene-disease-drug association

Since modern, different types of data have been used to predict disease depends on gene information that widely used protein-protein interaction (PPI) networks. Thus it endeavored to predict interactions with the disease-gene by explicitly using the topological structure of PPI

networks with multimodal deep learning. However,Universal PPI networks downloaded from online databases produce lots of false positives during feature extraction, and just by using them is hard to train and more attention is needed to choose appropriate hyperparameters that fail to enhance the genome frequency. Therefore, attribute reduction and prediction accuracy cannot be further enhanced. In addition, the selection of pertinent features from the regions of interest is carried out in the previous algorithm, which combines different types of clinical and non-clinical data which may reveal the intrinsic properties of diseases and genes. Therefore combine these multiple types of information could increase the computational time.

Wisdom of significant features selection is unpredictable in advance due to the complexity of the problem and for automatically generated processing. Consequently, for better representation of the gene environment, many specific feature sets are created, additional features would provide added perceptive control over the data, but in reality, more features contribute to diminishing the population diversity then slowdown the classification model. And also redundant functionality leads in the classifier is over fitting of gene molecules hindrance. Along with this identify the proper medicine from genomic information also has more challenges in prevailing researches. Therefore, this paper proposed a novel tactic for intensifying the genome frequency, highly reducing attributes, enhances population diversity and improves convergence speed, thus embellishing the greater accuracy while gene information based disease prediction with medicine providence.

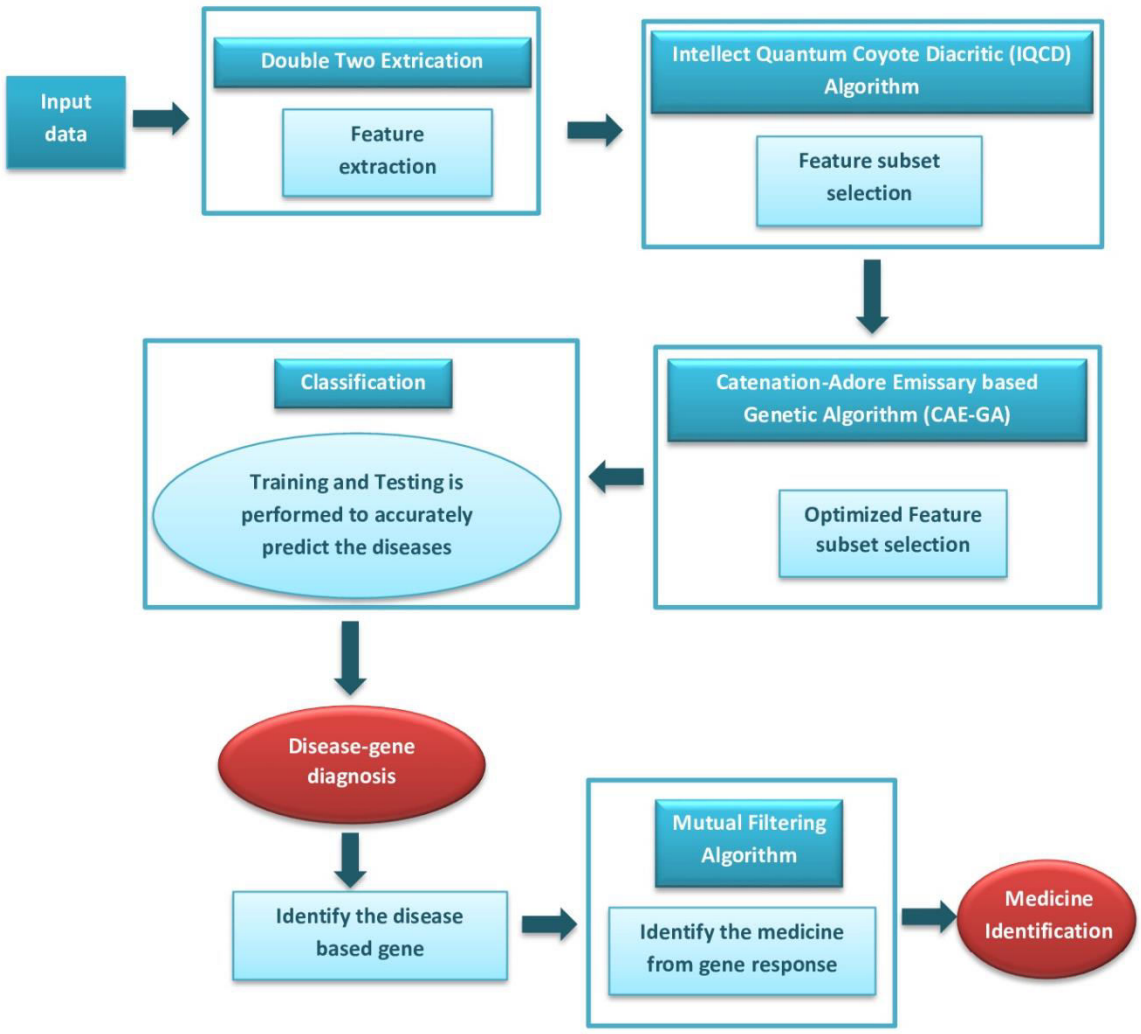


Figure-2: Block diagram of our proposed model

In this paper efficiently proposed the Disease-Gene Reliant Visage Prognostication (DG-RVP) Model for attains ridiculousintegrity to predict diseases legislated by extracting and selecting the features over embellish the genome frequency, population diversity, convergence speed and attributes reduction. Initially, the model entails Double Two Extrication (DTE) for relevant feature extraction that includes anatomy based etymological features, exponential based cohesion features, contagion analogy features, and chromatin analogy features. In DTE extracts the edges are weighted by the homogeneity scores, therefore thus it strengthens the genome frequency. After feature extraction proposed model needs to select the feature subsets, hence, an exotic model suppresses Quantum Coyote Diacritic (QCD) Algorithm, which is responsible for attribute reduction in the data classification task which improve the evaluation scheme and

search strategy of feature subsets through quantum computing and lopsided deposit contention, in that each subset of features represented the individual search position in the region.

At last from the selected featureto optimize the feature selection, the model suppresses with Catenation-Adore Emissary based Genetic Algorithm (CAE-GA). Hence it avoids the early convergence via a one-dimensional emissary structure that familiarizingthe interaction of genetic operators, which auspiciously conserves the population diversity. Finally, the novel model acquires the optimized selected features from input data then to performs the classification to identify/predict the disease based on genomic information. Then, to identify the medicine, the proposed model introduced the effectual Mutual Filtering Algorithm. That models the intrinsically low dimensionality of the data for gene expression while taking account of noise and bias. The outcome of the proposed model attains the greater more accuracy to predict the disease based gene with tackling the major challenges like lower genome frequency, poorly reducing the attributes, small population diversity, and less convergence speed. Therefore the proposed model effectively overcome those issues and predicts the diseases based on genomic information and identifies the medicine from gene responses. Consequently, the detailed explanations of proposed models are described in follows.

3.1: Double Two Extrication (DTE) for Feature extraction:

In this section, the main consideration of extraction of features from the gene by novel DTE, Once, proper segmentation of cell artifacts takes place, the extraction of specific features from the regions of interest for better predictions.

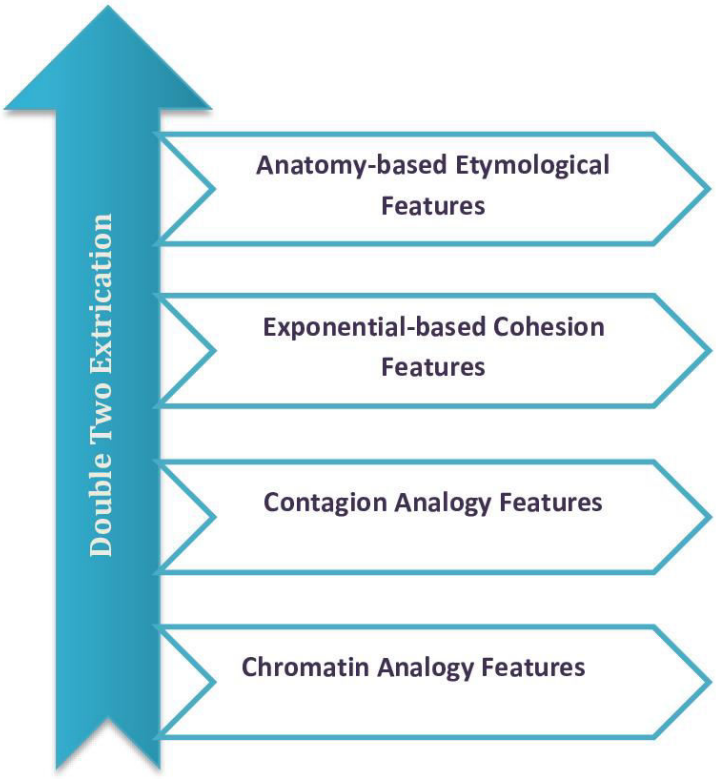


Figure: 3-Categorization of Doubles Two Extrication

The cellular level features concentrate mostly on isolated cell characteristics. For isolated objects with limited information and unknown distribution, the classification task is accomplished by extracting together with other features, such as histogram of oriented gradients and wavelet-based features, features based on shape morphology and texture. In the proposed approach consider four types of arrangements such as anatomy-based etymological, exponential-based cohesion, contagion analogy, and chromatin analog features are extracted for detection and classification.

3.1.2: Anatomy-based Etymological Feature:

The nucleus shape is described for different types of cell differentiation with the aid of morphological features. Due to the variety of nucleus morphologies, different morphological characteristics are extracted, such as area, perimeters, aspect ratio, solidity, eccentricity, shape signature, perimeter, compactness, extent, length of the major axis, and length of the minor axis. The detailed explanation here is as follows:

Area: The area of an object in a 2D image is determined by calculating the pixel region of the nucleus. This can be given, mathematically, by:

$$Ar = \sum_{i=1}^{n} \sum_{j=1}^{l} P(i, j) \tag{1}$$

Where Ar is the nucleus region and P is the area of interest (ROI) with n rows and l columns.

Perimeters: The perimeter is the distance by which each neighboring pair of pixels connects around the nucleus boundary. The simplest way to measure the nucleus perimeter is by counting the total number of edge pixels pertaining to the object. The following mathematical formula can be used to indicate:

$$Pe = \sum_{n,l \in Bop} P(n,l) \tag{2}$$

Where Pe is the perimeter, P is the ROI with n rows and l columns and Bop described the boundary of the pixel.

Aspect ratio: This feature is the best characteristic for distinguishing circular and non-circular objects or needle-like forms. Aspect Ratio value range is between 0 and 1. The value closer to 1 will be more elongated (as in our case with malignant cells), whereas the value closer to 0 will show benign cells.

Solidity: Solidity is the ROI area a ratio to their convex hull. Also, it is an essential feature, calculated as follows:

$$Solidity = \frac{Area}{ConvexHull} \tag{3}$$

Eccentricity: Eccentricity may be determined by area ratio and square perimeter. The estimation is done using the following equation:

$$\text{Eccentricity} = \frac{\text{Ar}}{\text{Pe}^2} \tag{4}$$

Where the variables $A_r$is the area and $P_e$is the perimeter.

Circularity: It is the circularity measure of the malignant cell nuclei, that is, how circular the cell nucleus is. It is represented, mathematically, as follows:

$$Circularity = \frac{Pe^2}{4\pi \times Ar} \tag{5}$$

Where pe is the Perimeter. The circularity value ranges between 0 and 1 if the circularity of the nucleus is equal to 1, then it is considered circular, if not elongated in shape.

Roundness: Malignancy tends to change nucleus chromosomes that disfigure nucleus shape. That leads to changes in the circularity and roundness of the nucleus. Roundness measures the distance between the malignant cell's borders to the area's center. Mathematically, the equation can be calculated as follows:

$$Roundness = \frac{1-\theta}{Dis\tan ce} \tag{6}$$

Where, θ displays the mean distance divergence given by the Distance variable. Mathematically, the following equations are calculated for and θ Distance:

$$\theta = \frac{\sqrt{\sum_{n,l}\left(\left\|\delta_o^p - P(n,l)\right\| - Dis\tan ce\right)^2}}{Area} \tag{7}$$

$$\text{Distance} = \frac{\sum_{\text{n,l}}\left(\left\|\delta_{\text{o}}^{\text{p}} - P(n,l)\right\|\right)}{Area} \tag{8}$$

Here, $\delta_{\text{o}}^{\text{p}}$ show the mean value of pixels in the area of a cell and $P(n,l)$ pixel value in the area of a cell.After analyzing the anatomy-based etymological feature, this work extracts the exponential based cohesion feature.

3.1.2: Exponential-based Cohesion feature:

Statistical methods using local features can be implemented in cohesion dependent applications to analyze the spatial distribution of nucleus values at every pixel in the image. These cohesion based characteristics are used to understand the microscopic image of the level pixel in the breast cytology. Statistical derivatives of the first order are used for the classification in the proposed approach. Below are some important cohesion features and their mathematical representations:

$$Mean = \frac{1}{B}\sum_{j=1}^{B} Y_j \tag{9}$$

$$SD = \sqrt{\frac{1}{B-1}\sum_{j=1}^{B}\left(Y_j - M\right)^2} \tag{10}$$

$$Skewness = \frac{1}{SD^3}.\sum_{j=1}^{B}\left(Y_j - M\right)^3 \tag{11}$$

$$Kurtosis = \frac{1}{SD^4}.\sum_{j=1}^{B}\left(Y_j - M\right)^4 \tag{12}$$

$$Entropy = -\sum_{j=1}^{B} O(Y_j) \log_2 O(Y_j) \tag{13}$$

$$Energy = \sum_{j=1}^{B} Y_j^2 \tag{14}$$

Based on the above 9 -14 equation this paper extracts the exponential based features, thus it described the parameters/ statistics of the gene. Then to extract the contagion and chromatin based features.

3.1.3: Contagion Analogy Features:

In the PPI-based model, the gene-gene interaction network mapped from the PPI network is known to be the network of gene similarities. This strategy is chosen because protein interactions may have similar functions and protein interactions that indicate the functional similarities between the respective genes. Meanwhile, the topological structure of the PPI network is also valuable when extracting features with node2vec instead of constructing another gene similarity network.

The disease analogy network Nw_{di}^{PPI} is structured according to the theory of disease module. A disease module within an interactome is a sub graph consisting of the disease-related genes. Let MD_1= (C1, W1) in the interactome (gene-gene interaction network) denote the disease module d1. C= {g_{11}, g_{12} . . .g_{1n1}} is a set of d_1-associated disease genes, and W_1 is a set of interactions between them. MD_2= (C2, W2) is another module with a similar definition for the disease. The similarity between two disease modules MD_1 and MD_2 can be calculated as follows

$$ana_{ppi}(MD_1, MD_2) = \frac{\sum 1 \le a \le n_1 G_{D2}(f1_a) \sum 1 \le m \le n_2 G_{D1}(f2_m)}{n_1 + n_2} \tag{15}$$

Where $G_D(f) = avg\left(\sum_{f_j \in MD} sim(f, f_j)\right)$ measures the relation between gene f and disease module MD, which is the sum of the transformed analogies between f and the genes in disease module MD. Given two genes f_1, $f_2$in the PPI network, their transformed analogy is calculated by

$$ana(f_1, f_2) = \begin{cases} 1, & f_1 = f_2 \\ e^{-sp(f_1, f_2)}, & otherwise \end{cases} \tag{16}$$

Where sp(f_1, f_2) is the shortest path length between $f_1$and f_2 in the PPI network. The bigger the similarity transformed, the closer the relation between f1 and f2.

Once the similarities between modules MD_1 and MD_2 are calculated, similarities between diseases di_1 and di_2 can be obtained by normalizing the similarities of the module as follows:

$$ANA_{ppi}^{di}(di_1, di_2) = \frac{2 * ana_{ppi}(MD1, MD2)}{ana_{ppi}(MD1, MD1) + ana_{ppi}(MD2, MD2)} \quad (17)$$

Specifically, edges are added to Nw_{di}^{PPI} for each disease and its top most similar diseases obtained equation 17. These edges are weighted by the analogy scores of their two connected diseases.

3.1.4: Chromatin Analogy (CA) Features:

Similar to the structure of Nw_{di}^{PPI}, the CA-based analogy networks are also built by the KNN algorithm, The GA-based analogy networks are also constructed using the KNN algorithm, except that the similarities between diseases and genes are calculated using CA instead of PPI.

CA database provides a set of vocabulary for describing the gene products based on their cellular functions. CA defines three types of ontology: the biological process, the cellular component and the molecular function.

All of the terms of CA exist as directed acyclic graphs (DAGs) where nodes represent terms whereas edges represent semantic relations. We use the approach developed by this study to measure the semantic similarities of CA terms and genes.

Let $DAG_U = (S_U, R_U)$ represents CA term U, Where Su contains all of U's successor CA terms in the DAG, and Ru contains U's semantic relationship with other terms in Su. Each term m in Su has a U-related B-value:

$$\begin{cases} B_U(m) = 1, if & m = U \\ B_U(m) = \max\{w_e * B_U(m') | m' \in childrenof, m\}, & otherwise \end{cases} \quad (17)$$

Given $DAG_U = (S_U, R_U)$and $DAG_V = (S_V, RV)$ for two CA terms U and V, the semantic analogy of these two terms is computed by:

$$BCA(U, V) = \frac{\sum_{m \in S_U \cap S_V} (B_U(m) + B_V(m))}{\sum_{m \in S_U} B_U(m) + \sum B_V(m)} \quad (18)$$

The semantic analogy of one CA term m' and set of CA terms $CA = \{m_1, m_2, \ldots, m_l\}$is denoted as:

$$ana_{ca}(m', CA) = \max_{1 \le j \le l}(BCA(m', m_j)) \quad (19)$$

Then the functionality of two genes f_1 and f_2, annotated by CA term set $CA_1 = \{m_{11}, m_{12}, \ldots, m_{1i1}\}$ and $CA_2 = \{m_{21}, m_{22}, \ldots, m_{2i2}\}$, is calculated by:

$$ANA_{CA}^{C}(f_1, f_2) = \frac{\sum_{1\le i\le n1} ana_{ca}(m_{1i}, CA_2) + \sum_{1\le j\le n2} ana_{ca}(m_{2j}, CA_1)}{n_1 + n_2} \tag{20}$$

The analogy of two diseases di_1 and di_2, associated with two sets of genes G_1= {f_{11}, f_{12},…,f_{1n1}},G_2={f_{21}, f_{22}, …,f_{2n2}} is defined as:

$$\mathrm{ANA}_{\mathrm{ca}}^{\mathrm{di}}(di_1, di_2) = \frac{\sum_{1\le i\le n1} BC(f_{1i}, DG_2) + \sum_{1\le j\le n2} BC(f_{2j}, DG_1)}{n_1 + n_2} \tag{21}$$

Where $\mathrm{BC(f', DG)} = \max_{1\le i\le 1}\left(ANA_{ca}^{c}(f', f_i)\right)$

Consequently, from the above mentioned feature extraction based on four levels of features that outperform the extraction of gene information. After feature extraction then in this paper proposed the QCD algorithm for feature selection with quantum computing, details follow.

3.2: Quantum Coyote Diacritic (QCD) Algorithm:

In order to solve the attribute reduction, the QCD algorithm was represented in three key modules by the paper. The first module included quantum in the collection and initialization of instances of a coyote. In the second module, the binary coyote solution was built from the way to (third module) update a coyote optimizer's search agent.

3.2.1: Irregular Set Assessment Function:

Using the unpredictable synchronicity, the irrelevant features are removed initially, the workload of feature selection is reduced, and the credibility of the feature subset evaluation is increased. It can be seen from the below equation 22

$$Gi\langle K|M\rangle = I(K) - I\langle K|M\rangle \tag{22}$$

Where, $Gi\langle K|M\rangle$ is information gain based on information entropy I (K) and conditional entropy $I\langle K|M\rangle$. Here if the variables K and M are not related, the information gain $Gi\langle K|M\rangle$=0, otherwise $Gi\langle K|M\rangle$>0, the larger value of information gain between K and M variables, attains stronger correlation. Therefore, information gain can be used quantitatively to access the dependency between two variables. However, the variable unit and the variable value can affect the gain of information, so further homogenization is necessary. Then unpredictable synchronicity UN(K,M) is a normalized information gain and is defines as

$$\mathrm{UN(K,M)} = \frac{Gi\langle K|M\rangle}{I(K) + I(M)} \tag{23}$$

Here the unpredictable synchronicity UN (K, M) satisfies $0 \le UN(K,M) \le 1$ and UN (K, M) =0 means that two random variables K and M are independent of each other, while (K, M) =1 means that the two random variables K and M are completely related.The appropriate features are selected in the dataset according to the above principle, and the redundant function can be removed to reduce the workload of feature selection, thus enhancing classification accuracy.

The evaluation method is as follows for evaluating the association between the conditional feature subset Q and the decision feature DF in the information system

$$EF = \varpi * \varphi q(DF) + \mu * \frac{|Q - O|}{|Q|} \tag{24}$$

$\varphi_o(DF)$ is the dependence of restrictive feature subset O with respect to decision feature DF. $|O|$,described the subset O of the selected conditional features and satisfies $O \subset Q; |Q|$ described the cardinality of the feature conditional. The effect of dependence between the conditional feature sub-set O and the decision function D is expressed by both ϕ and μ. That is, it affects the reduction rate of O.

$\varphi \in [0,1], \mu = 1 - \varphi$ This shows that the quality of classification and the length of the sub-set attributes have divided meanings for the task of attribute reduction.

The following equations in order to calculate the degree of dependency between each conditional function sub-set and decision characteristic QCD follows. According to S1, the three partitions of an unrecognized entity are a subset of the conditional function

$$OMN(S_1) = \{\{o_1, o_3, o_5, o_7\}, \{o_2, o_6\}\{o_4, o_8\}\} \tag{25}$$

The three partitions of an indistinguishable object as compared with sub-sets of conditional property S_2 are:

$$OMN(S_2) = \{\{o_1, o_5\}, \{o_2, o_3, o_6, o_7\}, \{o_4, o_8\}\} \tag{26}$$

In brief, the two partitions of indistinguishable objects are compared to the decision feature "class":

$$DF1 = \{o | class(o) = 0 = \{o_1, o_3, o_5, o_7\}\} \tag{27}$$

$$DF1 = \{o | class(o) = 0 = \{o_2, o_6\}\} \tag{28}$$

The algorithm determines the positive area that can evaluate the OMN and the indiscernible relationship between the features S1, S2 and the decision function DF, and finally measures the

dependency of the conditional subset of feature and decision feature. The dependencies between the set of decision function and S1 and S2 are as follows:

$$\varphi s_1(class) = \frac{p^{+}_{s_1}(class)}{\mathrm{O}} \tag{29}$$

$$\varphi s_2(class) = \frac{p^{+}_{s_2}(class)}{\mathrm{O}} \tag{30}$$

In equation 29 and 30 represent the dependencies of S1 and S2. After that, this paper represents the quantum individuals.

3.2.2: Quantum Representation of Individuals:

This paper used the angle of rotation to describe qubit, the angle of rotation is given in Figure 4. Quantum coyote entity for the collection of features: QCDx (the xth quantum coyote in a quantum group) corresponds to a vector $\zeta_x.\zeta_x = (\varpi_{x1,\ldots,}\varpi_{xn})$ of variables $\varpi_{xy}, with\ \varpi_{xy} \in \left[0, \frac{\pi}{2}\right] for(1 \le z \le n)$

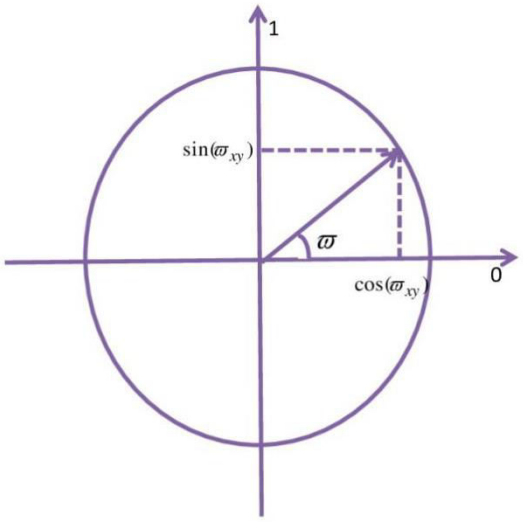


Figure: 4-Rotation Angle to describe the qubits

Every QCDx quantum coyote solution is a qubit series, measured as follows:

$$QCA_x = \begin{bmatrix} \cos(\varpi_{x1})|.....\cos(\varpi_{xn})| \\ \sin(\varpi_{x1})|.....\sin(\varpi_{xn})| \end{bmatrix} \tag{31}$$

Then represents the individuals of quantum this paper performs the feature selection based on quantum computing.

3.2.3: Feature selection based on quantum computing:

There are two states for each subset of features in the app selection: pick or not choose (1 or 0). The two-dimensional quantum superposition state of 1 and 0 can be easily described for this function. What needs to be done is to control the probability of updating the superposition when 1 or 0 is obtained, so that in the field of quantum computing the problem of selection of features can be solved.

The population of the quantum coyote is represented by qubit sequences of length n, where n is the cardinal number of conditional attributes, and each qubit determines the selection probability of the subset of function. QCDx represents the individual of the quantum coyote, and the angle of rotation is ϖ_{xy}. The formulation in question is as follows:

$$\left|QCA_{x,y}\right. = \cos(\varpi_{xy})\left|0\right\rangle + \sin(\varpi_{xy})\left|1\right\rangle and\left(\cos^2(\varpi_{xy})\right) + \sin^2(\varpi_{xy})\left.\right| = 1) \tag{32}$$

Each quantum coyote diacritic QCD_0 is initialized by:

$$QCA_0 = \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & \ldots\ldots \cos\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \ldots\ldots \sin\left(\frac{\pi}{4}\right) \end{bmatrix} \tag{33}$$

Where $\cos^2(\pi/4)$ is the probability that condition dependency feature k is designated, and $\sin^2(\pi/4)$ is the probability that condition dependency feature k is excluded, and then execute the quantum measurement for QCD.

3.2.4: Quantum Measurement for QCD:

Quantum coyote (QCDx) solutions are used in the quantity measurement process to produce a binary coyote (BCAx) solution by the projection of qubits. For a quantum bit, the interval [0, 1] is used to generate a random number z. If $z > \sin^2(\varpi_{xy})$, selecting the corresponding conditional feature is set to 1; otherwise, the value is 0 and the corresponding conditional feature is rejected. Therefore, a quantum superposition solution includes several binary solutions, due to the superposition condition of the qubits. Each qubit, however, determines the likelihood of selecting or rejecting the corresponding feature. Only certain binary solutions are extracted from quantum solutions during the quantum measurement step, and the selection is guided by the probability of quantized coding.

Algorithm: 1- Measurement of binary quantum coyote individual

Step: 1- Initialize quantum coyote individual and conditional feature set (C= $\{c_1, c_2 \ldots c_x\}$).

Step: 2- Then for each qubit of QCD_x, real value is generated between [0, 1]. If real value is greater than doubles of trigonometric function as $\sin^2(\varpi_{xy})$, then binary coyote individuals are assigned be one.

Step: 3-As well real factor is defined by combination of xth quantum real factor with conditional feature sets. i.e. $Z \leftarrow Z_x \cup c_{x,y}$.

Step: 4- Otherwise binary coyote individuals are to be zero. Finally proposed algorithm obtains binary quantum coyote individual.

A feasible feature selection solution of a binary coyote algorithm is constructed by using the quantum measurement method. The algorithm at first chooses no function. For each individual, the feature is selected to refer to a condition according to $z > \sin^2(\varpi_{xy})$, and the operation of quantum measurement is repeated until all features are searched. The following algorithm 2 constructs the feature selection feature BCAx feasible solution by observing the quantum coyote's QCDx.

3.2.5: Update binary coyote position:

The method of using the algorithm for binary coyote optimization starts with the initialization of the individual binary coyote in algorithm 2. Alpha, beta, and delta estimate the possible positions of the subset (prey) feature during the iteration processes, and each potential subset feature scheme updates its distance from the prey.

Algorithm: 2-Minimization of attribute reduction

Step: 1- Initially calculate the fitness value of coyote individuals, search coyote for alpha, beta and delta. Then initialize the coyote parameters.

Step: 2- Though time taken is less than maximum iteration, for each omega coyote calculate the fitness function value of binary individual coyote.

Step: 3- Then for each search coyote if there is search coyote (alpha, beta, and delta) position that needs to be replaced.

Step: 4- At that moment update the all parameters a, A and C as well update the present search coyote position.

Step: 5- Finally proposed algorithm obtains the minimal attribute reduction and alpha coyote values.

In this work can use $\vec{A}$ in [0, 2] to make the coyote diverge from the prey. When $\vec{A}$ >1,the solution for the candidate tends to stay away from the target, when $\vec{A}$ <1, the solution for the candidate moves to closer from the target. The $\vec{C}$ vector in [0, 2] plays an important role in circumventing local optimal stagnation. $\vec{C}$ >1 means that $\vec{C}$ it emphasizes the role of avoiding local optimum. If $\vec{C}$ <1, the part $\vec{C}$ will get weakened at random. Finally, the algorithm for coyote optimization concludes by satisfying the end criteria for obtaining an alpha coyote.

Algorithm:3-Feature subset optimization

Step: 1-Initialize the *m* quantum coyote individuals QCD_0.

Step: 2- Get the group *BCD*0 of *m* binary coyote from *QCD*0 from the algorithm 1

Step: 3- Search the minimal feature subset Z_f of each binary coyote BCD_f from the algorithm 2.

Step: 4- Then to evaluate an each Z_f corresponding of binary coyote BCD_f using fitness function.

Step: 5- while time taken is less than maximum iterations, for all binary coyote individual are equal to one.

Step: 6- After defined the step 5, then evaluate the feature subset Z_f using fitness function.

Step: 7- Finally update the best feature subset based on minimum attribute reduction. Thus it obtains the optimal feature subset.

QCD's aim was to look for the potential optimal subset of features. First, it initialized some individuals with a quantum coyote in the optimized search space of the subset of features. Then, by algorithm 2 we got the group BGW0 of n binary coyote from QGW0. Binary coyote individuals explored the set space by using algorithm 3 to search for prey and hunt prey. Coyote individuals have searched out this space to find the best three omega coyote solutions to update alpha, beta, and delta positions. Before the maximum iterations of the algorithm were reached, the fitness of individual grey coyote was calculated to determine whether to update the positions of search coyote.Finally, the dependency concept proposed by the rough set theory is used to evaluate the selection of features, and it could determine whether a subset of conditional properties is an optimal solution. Hereafter in this paper proposed the optimization of feature selection based on genetic algorithm.

3.3: Catenation-Adore Emissary based Genetic Algorithm (CAE-GA):

Awareness of significant features is uncertain in advance because of the complexity of the issue and for automated processing. Therefore, a lot of specific feature sets are set for better domain representation. In theory, additional features would allow more perceptive control over the data, but in reality, more features cause the learning process to be slowed down. Similarly, one of the disadvantages of unnecessary features causes the classifier to over-fit. The GA dependent selection technique is used in the proposed method for selecting optimal apps. The GA is one of the popular Evolutionary Algorithm (EA) styles and is widely used in a number of computer vision applications that are focused on natural feature selection. GA deals with feature population space, where a catenation of iterative processes for creating a new version is performed. To obtain satisfactory results, GA builds up a systematic population of time-based solutions provided by chromosomes. In the assessment process, a fitness function calculates the meaning of the response. Mutation and crossover are the two basic functions which have a crucial effect on the fitness value. Based on roulette wheel techniques, chromosomes with higher fitness values are selected for the next generation. Similarly, some of the genes are changed at random in the mutation process. Crossover is a genetic phenomenon merging separate characteristics from the pair of subsets into a new subset. Offspring is replaced by the previous population using the alternative approach of diversity or exclusivity to generate a new population in the future generation. The literature discusses numerous GA related feature selection

algorithms.Due to its simplicity, precision, and efficacy, Catenation-Adore Emissary based Genetic Algorithm (CAE-GA) is proposed in this paper.

Algorithm: 4-CAE-GA

Step: 1- Initialized the Emissary catenation of 0th generation D^0, and update the best solution of population $popu^0_{best}$ and the generation is initialized from 0.

Step: 2- Then to perform the dynamic neighboring competitive selection processing and update the emissary catenation in the t-th generation, obtaining mid-catenation between D^t and D^{t+1}.

Step: 3- For each emissary in mid catenation $D^{t+1/3}$, performs the cross over on it and obtain the mid catenation value.

Step: 4- For each emissary in mid catenation $D^{t+2/3}$, performs the mutation processing on it, and obtaining $D^t{}_{end}$ emissary catenation after mutation processing.

Step: 5-Then estimate the best emissary Emi^{ct}_{best} in $D^t{}_{end}$, and compare Emi^{ct}_{best} and Emi^{t-1}_{best} hence the energy value of Emi^{ct}_{best} is greater than then energy value of Emi^{t-1}_{best}, then Emi^{t}_{best} value is belongs to Emi^{ct}_{best}. Otherwise Emi^{t}_{best} belongs to Emi^{t-1}_{best} and D^{t+1} belongs to $D^t{}_{end}$.

Step: 6- whether the above steps are satisfied then obtain the output value as best emissary of t-th generation. Or t-th generation belongs to increment of t-th generation then re-performs the steps 2-6.

The CAE-GA approach is implemented as a population structure thus the arrangement of the population is simpler compared to other systems of emissaries.

Similarly, it is computationally efficient that effectively prevents early convergence. The incorporation of genetic operators such as adaptive mutation and crossovers in the CAE-GA effectively check for the global optima and maintain the diversity of the population. In addition, a competitive selection approach is used to enhance search capability through a dynamic selection of neighbors. Illustration 5 shows various features derived from photographs of the segmented breast cytology. The matrix of features is created from the extracted features, which are then passed through CAE-GA for the selection of the optimal features.

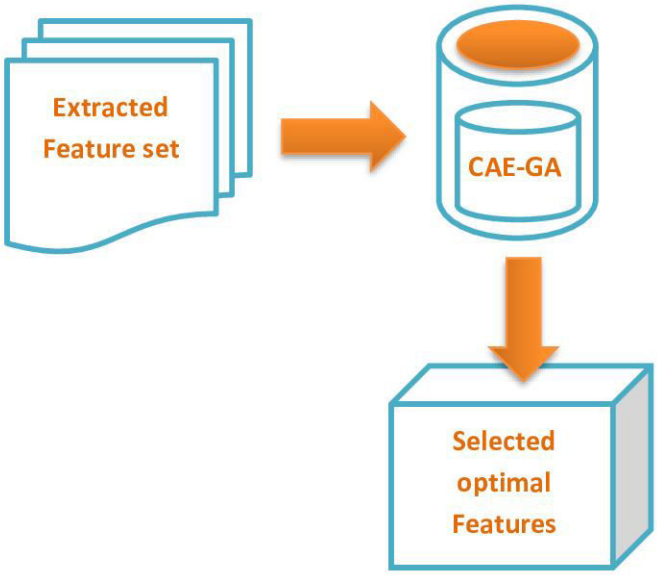


Figure: 5- CAE-GA optimal feature selection

CAE-GA is used for the optimum collection of features and the problem of global numerical optimization, resulting in a sufficient classification with higher precision. The CAE-GA incorporates both complex neighboring genetic operators and catenation-like emissary structure to achieve a higher capacity for optimization. An emissary signifies a candidate solution to the optimization problem in the catenation-like emissary structure. The evolution happens through emissary interaction. Similarly, the neighboring emissaries collaborate and work with their colleagues to improve the performance with genetic operators.

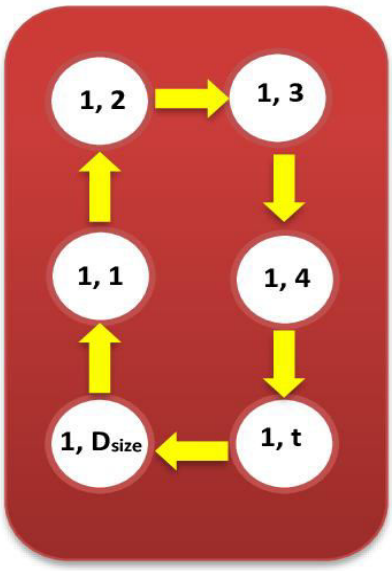


Figure:6- Emissary Catenation Structure

All emissaries are linked in one catenation in the CAE-GA system, known as the emissarycatenation, which is defined by D as shown in Fig. 6. At Fig. 6, Circles representemissaries, while the mark indicates the emissary's location in the catenation inside the circle. The emissary can communicate with neighbors both in the back and out. The catenation starts with the first node-specified emissary (1, 1) (where 1 represents a one-dimensional emissary structure) and communicates with its neighboring node (1, 2), which further communicates to the D_{size} node through the catenation.

Therefore, from the extracted and optimized selected features this research performs the classification based on classification, predicts the disease from genome information. In addition, this paper introduced the medicine providence strategy which follows.

3.4: Mutual Filtering Algorithm:

The estimation of the gene expression response to the drug is essentially undetermined. The typical way to address this problem in literature is to judiciously pick a small number of transcriptomic features by sophisticated feature selection methods such as sparse regression or other gene ranking techniques that use prior knowledge in the form of protein-protein interactions (PPI's) and genetic interactions. The existing data sets include both gene expression data from different cell lines as well as their reaction to different drugs, where each drug's response is measured only for a subset of cell lines. The data on gene expression experimentally measured is inherently noisy and not generally ' oriented. This paper, therefore, employs a novel mutual filtering approach to model the intrinsically low dimensionality of the data on gene expression while taking into account of noise and bias. The detailed procedure of the MF algorithm is illustrated in the below figure 7.

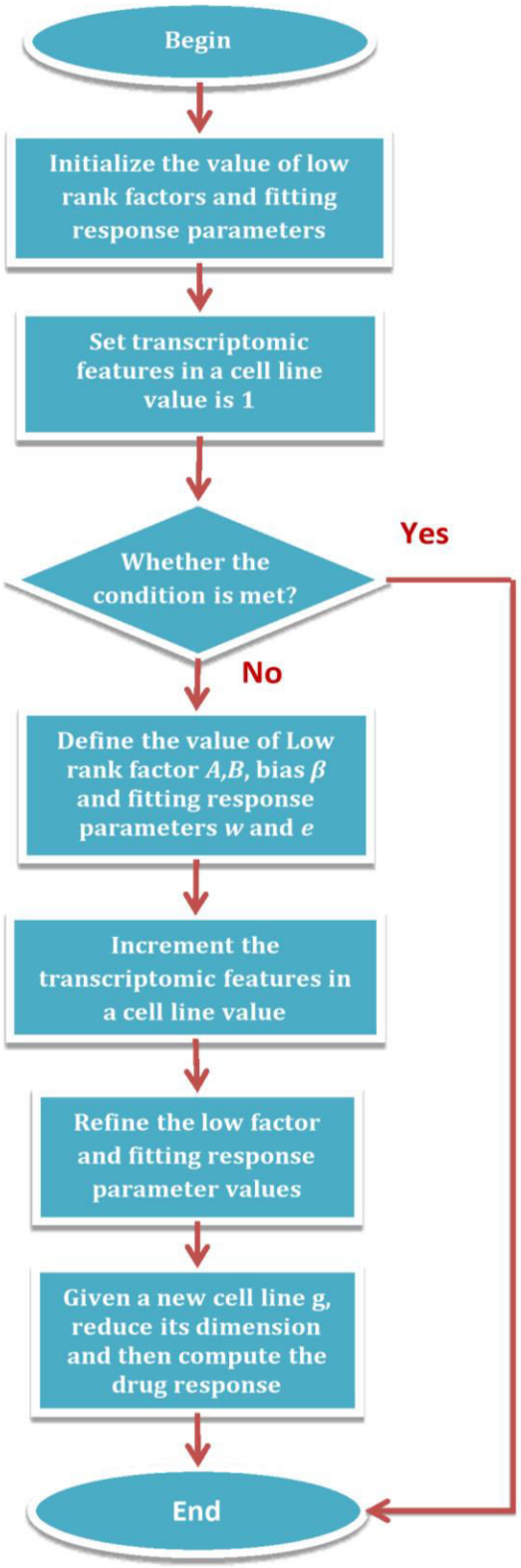


Figure:7-process flow of mutual filtering algorithm

Bias β is not just a gene expression bias vector; it actually plays an important role in seeking a specific dimensionality reduction matrix and helping estimate drug response from the latent space.

Consequently, this paper proposed a model predicts the disease and providence of medicine based on gene association in an effectual manner through representation and analysis of features with quantum and genetic algorithmic way. Hence proposed model can be efficiently used for IoT based medical healthcare monitoring applications. Further, the competence of the proposed model is proved through experimentally in the next section.

4. Result and discussion:

This section effectivelydescribes the proficiency of proposed work by extracting the features then selecting the appropriate features and optimized the selected features for predicting the disease based on gene finally to perform the filtering for identify the medicine. The competence of the proposed system is described by comparing the obtained experimental results with the conventional approaches.

4.1 System Specification:

The proposed system has been implemented in MATLAB/SIMULINK to demonstrate competent security. The system specifications are,

Platform	**MATLAB 2018b**
OS	**Windows 7**
Processor	**Intel core i5**
RAM	**8GB RAM**

4.2 Simulation Results:

Tables 1and 2 described the proposed model predict the disease and medicine based on gene id. Here that prediction performed with accurate features characteristics only. Therefore, it illustrates the proficiency of the proposed model when predicting the disease as well asthe identification of medicine.

Table: 1- Disease Prediction based on Gene

Si.No	Gene Id	Gene Symbol	Disease
1	2	A2M	Cancer
2	1	A1BG	Cancer
3	4	HLADRB1	Diabetes
4	5	MTHFD1L	Heart
5	8	SMAD3	Heart
6	7	MIA3	Heart
7	3	NAT1	Cancer
8	9	CXCL12	Heart
9	6	PSRC1	Heart

In table 1 the gene symbol as A2M, A1BG, HLADRB1, MTHFD1L, SMAD3, MIA3, NAT1, CXCL12, and PSRC1 with the corresponding disease is Cancer,Cancer, Diabetes, Heart, Heart, Heart, Cancer, Heart and Heart respectively.

Table: 2-Gene based Medicine identification

Si.No	Gene ID	Gene Symbol	Drug Name
1	1	A1BG	Camptoth
2	2	A2M	Vinblastine
3	3	NAT1	Cisplatin
4	2	A2M	Vinblastine
5	2	A2M	Vinblastine

In table 2 describe the gene Id is 1, 2, 3, 2, 2 gene symbol is A1BG, A2M, NAT1, A2M, A2M, and drug name is Camptoth, Vinblastine, Cisplatin, Vinblastine, Vinblastine respectively.

4.3:Performance evaluation:

In order to evaluate the performance of the proposed system following evaluation, metrics are taken into account, they are over all accuracy, precision, F-measure, prediction accuracy, recall, sensitivity, specificity, true positive rate, and false-positive rate.

4.3.1: Accuracy

Accuracy is the most important metric for determining the efficiency of a classification system. This is also used as a statistical measure of how well a classification test classifies a condition correctly or excludes this. The ratio of the number of correctly labeled samples to the total number of samples is taken as.

$$Accuracy = \frac{TP + TN}{P + N} \tag{34}$$

4.3.2: Specificity:

Specificity is a statistical method for measuring rating test performance. It is the responsibility of the specificity to measure the accuracy provided a specific class is set. The specificity is the probability that a negative outcome will be obtained when an attack is really negative.

$$Specificity = \frac{TN}{TN + FP} \tag{35}$$

Where, the result is a true positive (TP), where the model correctly predicts the positive type. A true negative (TN) is likewise an outcome where the model correctly predicts the negative type. And a false negative (FN) is an outcome where the negative class is incorrectly predicted by the model. A false positive (FP) is an outcome, where the model predicts the positive class incorrectly.

4.3.3: Recall:

The recall is the percentage of correctly expected positives to all actual class observations.

$$\text{Re}\,call = \frac{TP}{TP + FN} \tag{36}$$

4.3.4: F1-measure:

F1 Score could be a better measure to use if we need to aim for a balance between Precision and Recall AND there is an unequal class distribution (a large number of Real Negatives).

$$f1 = 2 \times \frac{precision * \mathrm{Re}\, call}{precision + \mathrm{Re}\, call} \tag{37}$$

4.3.5:Sensitivity:

True-Positive (TP) assessment such as the person has a disease. It is given, mathematically, by:

$$Sensitivity = \frac{TP}{TP + FN} \tag{38}$$

4.3.6: Precision:

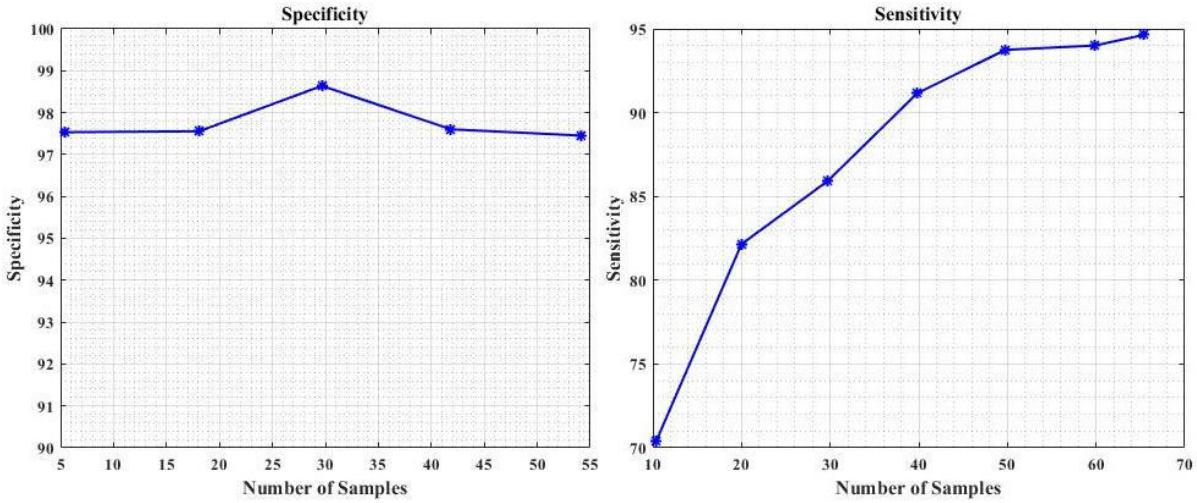
Precision (also known as the positive predictive value) is the fraction of appropriate instances among the instances retrieved. It is given, mathematically, by:

$$precision = \frac{TP}{TP + FP} \tag{39}$$

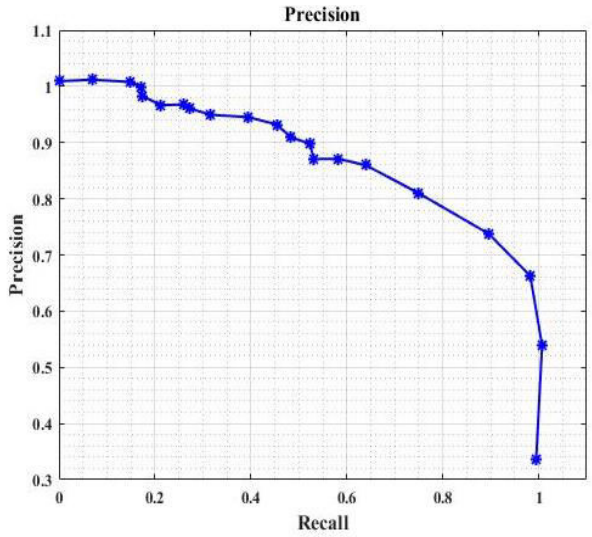
4.3.7: False Positive (FP)-Rate:

The percentage of negative cases not correctly counted as positive cases. It is given, mathematically, by:

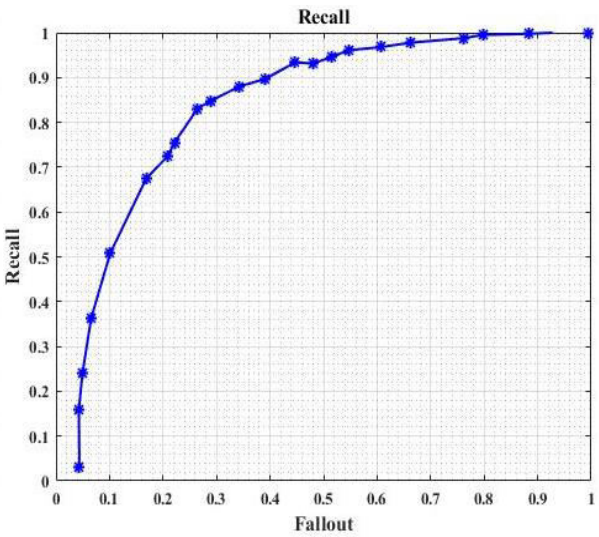
$$FalsePositiveRate = \frac{FP}{FP + TN} \tag{40}$$



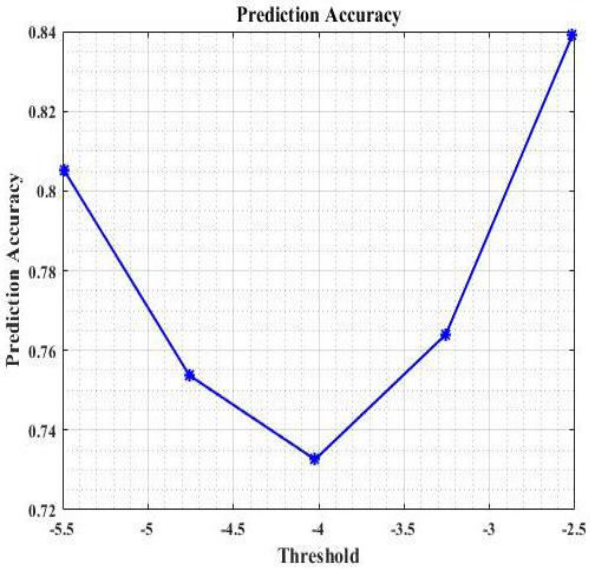
(a) **(b)**


Precision
Precision
Recall

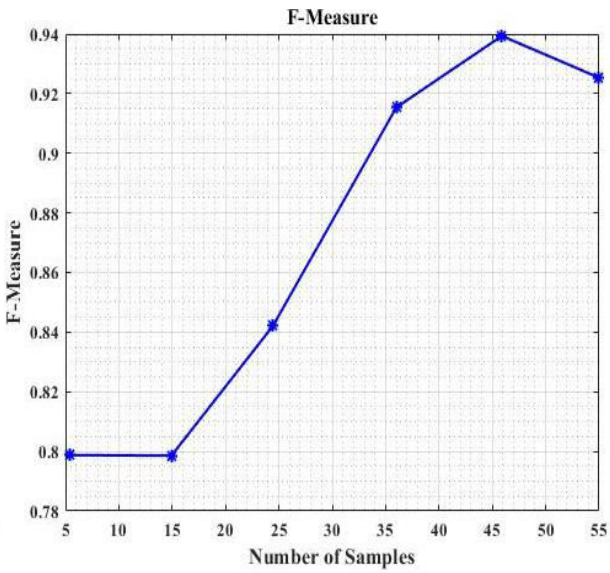
(c)


Recall
Recall
Fallout

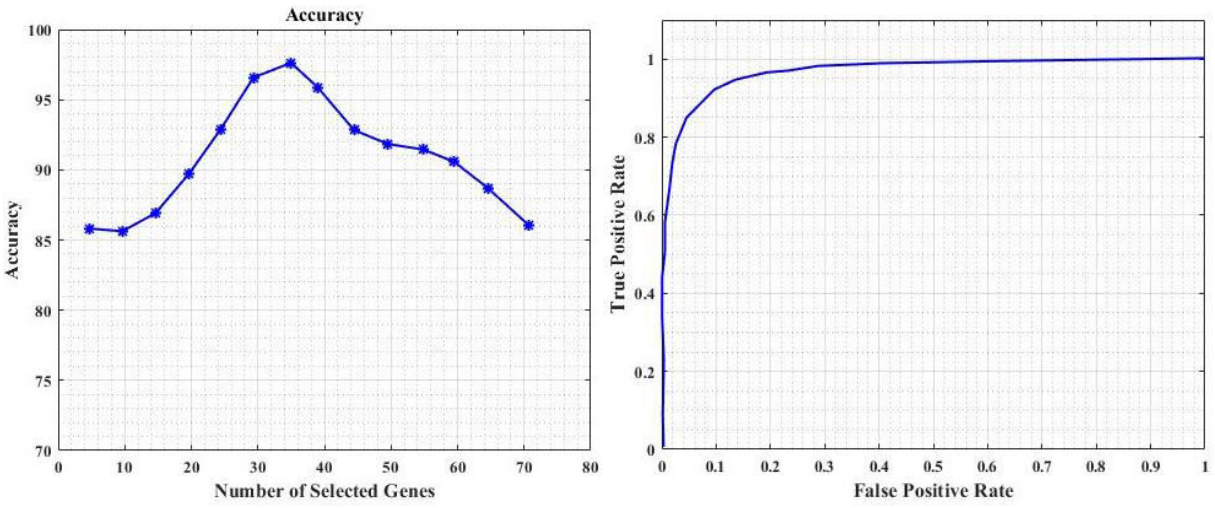
(d)


Prediction Accuracy
Prediction Accuracy
Threshold

(e)


F-Measure
F-Measure
Number of Samples

(f)

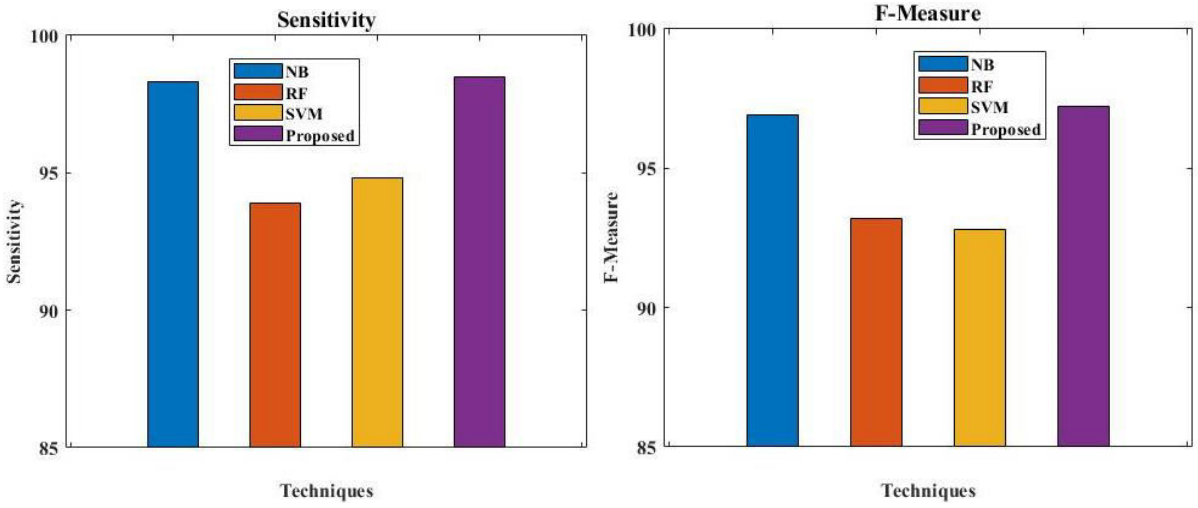


(g) **(h)**

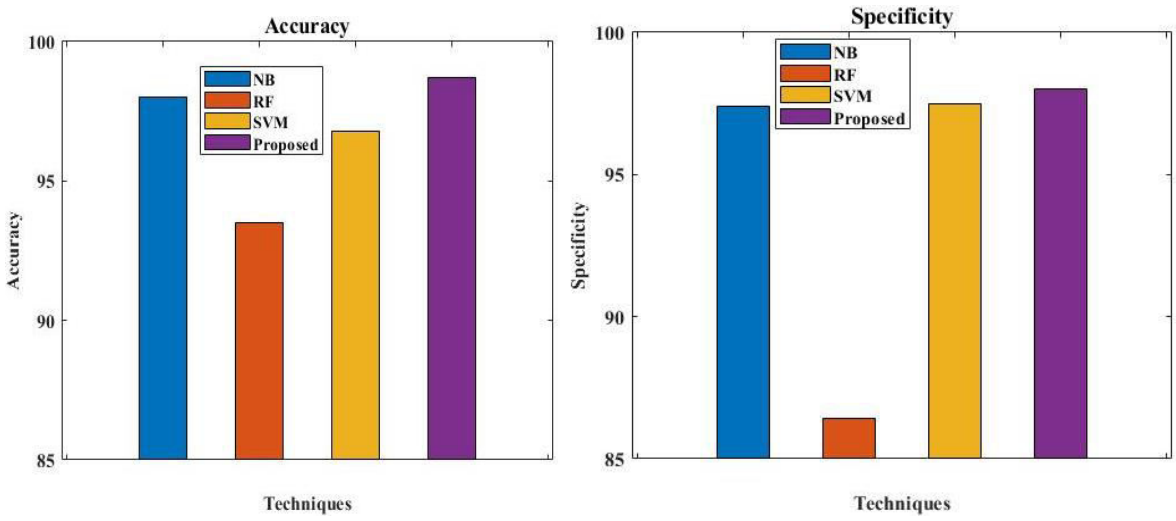
Figure: 8- overall Performance Evaluation (a) Specificity (b) Sensitivity (c) precision (d) Recall (e) prediction accuracy (f) F-Measure (g) over all accuracy (h) ROC curve

4.4: Comparison Results:

In this section, the performance of prediction accuracy, sensitivity, F-measure, and specificity is compared with existing techniques such as NB [29], RF [30], and SVM [31]. Thus it pronounced our proposed system for gene-disease-drug association attains in IoT based health monitoring in a great effectual and accurate manner.



(a) **(b)**



(c) **(d)**

Figure: 9-comparative analysis of proposed with existing (a) sensitivity (b) F-measure (c) Accuracy (d) Specificity

Consequently, from the result section, it clearly illustrates proposed models outperformed in disease, medicine prediction and achieves greater more accuracy through extracting the finite details of features with selective and optimized the value of the feature. Which attains based on genome information of patients.

5. Conclusion:

Integrating multiple types of data with the machine learning model is a challenging task, especially for predicting disease-genes-medicine correlation where there isa limited number of known associations.Hence, the model proposed in this paper for an e-Health care service to effectively and reliably predict disease and to define the medicine based on patient Genetic information. In which extract, select and optimize the features based on quantum and genetic strategy from the large volume of the database. Thus the classification attains greater more accuracy is 98.7% while gene-based disease prediction as well as medicine providence. Compare to the prevailing methods shownthe efficacy of the proposed model achieves sensitivity is 98.5%, specificity is 98% and F-measure is 97.21% in IoT based healthcare monitoring.

References:

[1] G.M. Cooper, R.E. Hausman, The cell, 85, Sinauer Associates Sunderland, 2000.

[2] W.H. Organization, W.H.O.M. of Substance Abuse Unit, Global status report on alcohol and health, 2014, World Health Organization, 2014.

[3] P. Boyle, B. Levin, et al. World cancer report 2008, IARC Press, International Agency for Research on Cancer, 2008.

[4] N.R. Council, et al., Saving women's lives: strategies for improving breast cancer detection and diagnosis, National Academies Press, 2005.

[5] F.A. Spanhol, L.S. Oliveira, C. Petitjea, n, L. Heutte, A dataset for breastcancerhistopathological image classification, IEEE Trans. Biomed. Eng. 63(7) (2016) 1455–1462.

[6] R. Rubin, D.S. Strayer, E. Rubin, et al., Rubin's pathology: clinicopathologicfoundations of medicine, Lippincott Williams & Wilkins, 2008.

[7] A.N. Pandya, N.P. Shah, Breast fine needle aspiration cytology reporting:a study of application of probabilistic approach, Indian Med. Gaz. 147 (2)(2013) 54–59.

[8] Muratli, N. Erdogan, S. Sevim, I. Unal, S. Akyuz, Diagnostic efficacy andimportance of fineneedle aspiration cytology of thyroid nodules, J. Cytol.Indian Acad. Cytol. 31 (2) (2014) 73.

[9] R.A. Castellino, Computer aided detection (CAD): an overview, CancerImaging 5 (1) (2005) 17.

[10] Y. Li, X. Zeng, L. Han, P. Wang, Two coding based adaptive parallel cogenetic algorithm with double agents structure, Eng. Appl. Artif. Intell. 23 (4) (2010) 526–542.

[11] K. Leon, D. Mery, F. Pedreschi, J. Leon, Color measurement in Lab unitsfrom RGB digital images, Food Res.Int. 39 (10) (2006) 1084–1091.

[12] K. Zuiderveld, Contrast limited adaptive histogram equalization, in:Graphics Gems IV, Academic Press Professional, Inc., 1994, pp. 474–485.

[13] D.T. Martin, S. Sandoval, C.N. Ta, M.E. Ruidiaz, M.J. Cortes-Mateos, D. Messmer,A.C. Kummel, S.L. Blair, J. Wang-Rodriguez, Quantitative automatedimage analysis system with automated debris filtering for the detection ofbreast carcinoma cells, ActaCytol. 55 (3) (2011) 271–280.

[14] N. Otsu, A threshold selection method from -level histograms, IEEETrans. Syst. Man Cybern. 9 (1) (1979) 62–66.

[15] References Davies, S.; Russell, S. NP-completeness of searches for smallest possible feature sets. In Proceedings of the AAAI Symposium on Intelligent Relevance, Berkeley, CA, USA, 4–6 February 1994; pp. 37–39.

[16] Gheyas, I.; Smith, L. Feature subset selection in large dimensionality domains. Pattern Recognit. **2010**, 43, 5–13.

[17] Vergara, J.; Estévez, P. A review of feature selection methods based on mutual information. Neural Comput. Appl. **2014**, 24, 175–186.

[18] Alba-Domínguez, M., López-Lera, A., Garrido, S., Nozal, P., González-Granado, I., Melero, J., Soler-Palacín, P., Cámara, C., and López-Trascasa, M. (2012). Complement factor i deficiency: a not so rare immune defect. characterization of newmutations and the first large gene deletion. Orphanet Journal of Rare Diseases,7(1), 42.

[19] Bademci, G., Cengiz, F., Foster II, J., Duman, D., Sennaroglu, L., Diaz-Horta, O.,Atik, T., Kirazli, T., Olgun, L., Alper, H., et al. (2016). Variations in multiplesyndromic deafness genes mimic non-syndromic hearing loss. Scientific Reports,6, 31622.

[20] Bengio, Y. et al. (2009). Learning deep architectures for ai. Foundations andTrends®in Machine Learning, 2(1), 1–127.

[21] Chen, B., Wang, J., Li, M., and Wu, F.-X. (2014). Identifying disease genes byintegrating multiple data sources. BMC Medical Genomics, 7(Suppl 2), S2.

[22] Chen, B., Li, M., Wang, J., Shang, X., and Wu, F.-X. (2015). A fast and high performance multiple data integration algorithm for identifying human disease genes. BMC Medical Genomics, 8(Suppl 3), S

[23] Li, Y., Wu, F.-X., and Ngom, A. (2016b). A review on machine learning principles for multi-view biological data integration. Briefings in Bioinformatics, 19(2), 325–340.

[24] Wu, M., Zeng, W., Liu, W., Zhang, Y., Chen, T. and Jiang, R., 2017, November. Integrating embeddings of multiple gene networks to prioritize complex disease-associated genes. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 208-215). IEEE.

[25] Luo, P., Tian, L.P., Chen, B., Xiao, Q. and Wu, F.X., 2018, June. Predicting gene-disease associations with manifold learning. In *International Symposium on Bioinformatics Research and Applications* (pp. 265-271). Springer, Cham.

[26] G. Manogaran, D. Lopez, N. Chilamkurti, In-mapper combiner based mapreduce algorithm for processing of big climate data, Future Gener. Comput. Syst. 86 (2018) 433–445.

[27] Chen, Y.; Zeng, Z.; Lu, J. Neighborhood rough set reduction with fish swarm algorithm. Soft Comput. **2017**, 21, 6907–6918.

[28] Luan, X.Y.; Li, Z.P.; Liu, T.Z. A novel attribute reduction algorithm based on rough set and improved artificial fish swarm algorithm. Neurocomputing **2016**, 174, 522–529.

[29] Chaudhary, K., Naidu, R., Rai, R. and Gawai, N., 2019. Hybrid Architecture of Heart Disease Prediction System using Genetic Neural Network.

[30] Zhang, Y., Yi, T., Ji, H., Zhao, G., Xi, Y., Dong, C., Zhang, L., Zhang, X., Zhao, J. and Liao, Q., 2019. Designing a general method for predicting the regulatory relationships between long noncoding RNAs and protein-coding genes based on multi-omics characteristics. *Bioinformatics*.

[31] Wang, X.B., Cui, N.H. and Ming, L., 2019. Identification of a blood-based 12-gene signature that predicts the severity of coronary artery stenosis: An integrative approach based on gene network construction, Support Vector Machine algorithm, and multi-cohort validation. *Atherosclerosis*, *291*, pp.34-43.

[32] Yogesh Hole et al 2019 J. Phys.: Conf. Ser. 1362 012121