Weighted Clustering for Deep Learning Approach in Heart Disease Diagnosis

BhandareTrupti Vasantrao[1]

Department of Computer Science and Engineering
Alliance College of Engineering and Design
Alliance University, Bangalore, India

Dr. Selvarani Rangasamy[2]

Alliance College of Engineering and Design
Alliance University, Bangalore
India

Abstract—**An approach for heart diagnosis based on weighted clustering is presented in this paper. The existing heart diagnosis approach develops a decision based on correlation of feature vector of a querying sample with available knowledge to the system. With increase in the learning data to the system the search overhead increases. This tends to delay in decision making. The linear mapping is improved by the clustering process of large database information. However, the issue of data clustering is observed to be limited with increase in training information and characteristic of learning feature. To overcome the issue of accurate clustering, a weighted clustering approach based on gain factor is proposed. This approach updates the cluster information based on dual factor monitoring of distance and gain parameter. The presented approach illustrates an improvement in the mining performance in terms of accuracy, sensitivity and recall rate.**

Keywords—*Learning approach; weighted clustering; heart disease diagnosis; gain factor*

I. INTRODUCTION

Heart diseases are rapidly increasing in recent past due to uneven living style and a highly variant environment conditions. The automaton of heart diagnosis for an early alarming is hence a primal need in today's leaving. The advancement in recent technologies in data mining has brought out new possibilities of early diagnosis and alarming of heart disorderness based on vital parameter analysis. The existing system uses the past learning parameters from a large database in making decision to monitoring values. However the rapid increases in data set and the availability of new data exchange and storage facilities constraint the mining performance due to large search overhead. Hence, an enchantment to the existing approach of data mining is required in making decisions to Heart diagnosis. In developing approach for automation in heart diagnosis and early alarming various methods were presented in past. In context to the development of learning approach for heart diseases in [1] a new classification model based on feature fusion model is presented. The features of vital parameters in heart diagnosis are fused with medical monitored parameters for improvising the accuracy. In [2] a semantic co-ranking process for heart disease diagnosis is presented. This approach is however limited with the diversity in data base entry. In [3] a fast convergence learning approach based on distance parameter is presented. In [4] a decision system based of Bergman divergence condition in training and testing for heart disease diagnosis is presented. The distance parameters in feature vectors are used in developing a decision.

In [5,6] a matrix learning approach for co-ranking of data base features is presented. The approach outlines a random walk method in deriving the best possible decision for given test parameters. A modified divergence approach is proposed in [7]. This approach develops a hyper graph model for classification using learning of feature vector matrix. The diversity of feature vector in database however limits the proposed approach of classification. In [8] a relevance coding for heart disease diagnosis is presented. This approach develops a classification model based in the distribution of feature vectors in a wider domain. In [9, 10] a decision system based on personalized information in cardiovascular disorderness is presented. This approach addressed the variation parameter of learning feature in making decision. In [11] an approach of diagnosis for heart disease using semantic features is proposed. This approach defines a new relevance coding based on feature correlation and re-ranking operation in large dataset for heart diseases diagnosis. A fusion approach for process features based on the variation of feature vector is outlined in [12, 13]. The presented approach defines a correlative function model based on the magnitude parameter of vitals given. The issue of scaling and varying diversity is not been addressed in the presented approach. In [14] an approach of random updation and mapping for feature vector in heart disease diagnosis is presented. The presented approach limits the misclassification performance by a robust online diagnosis of feature vector. In [15] a naive Bayes method for heart disease diagnosis is presented. This proposed approach develops a classification approach based on hidden knowledge based on continuous data in the database. In [16] content based represented of vital parameters for heart disease diagnosis is presented. This approach monitors the variation among the feature parameters in developing a rank value for decision making. The increase in data base feature is observed to be effective in learning performance, however, the constraint of search overhead and misclassification performance under semantic feature limits the current mining approaches in early diagnosis for heart disease. To improve the performance based on search overhead, delay and accuracy metric a new weighted clustering approach based on cluster gain parameter is proposed. In [19] new MAE approach is developed to estimate the heart disease estimation using machine learning algorithms. For prediction of health system and risk different machine learning approaches are outlined in [20]. The existing approach of machine learning based on feature values used for training. The training overhead of the classification system is based on the number of learning features used. The more details of

feature value results to higher accuracy however with a cost of processing overhead. Towards minimizing feature overhead, in a recent approach [18] a feature fusion approach is developed based on data and feature values. The fusion model is developed based on the magnitude of each parameter in observation. The fusion model minimizes the processing overhead by reducing the number of training features; however the proposed approach is developed based on discrete monitoring value of the vitals. Wherein fusion model minimizes the processing overhead, the fusion of feature based on magnitude is limited with external distortion. A deviation in magnitude value due to a spike or jitter results to a misclassification. The effect of distortion is hence needed to be minimized to improve estimation accuracy. With the objective of improving the fusion accuracy and noise suppression a new time variant feature value with medical significance characteristic is proposed. This proposed approach presented a new mean of feature fusion using a time line monitoring of feature variation and observing the classification improved due to the fusion. The rest of this paper is outlined in six sections. The existing feature clustering approach and heart diagnosis approach is presented in Section II. Section III outlines the approach of proposed heart diagnosis using weighted clustering approach. Section IV presents the results obtained for the developed system. A discussion to the observations obtained is presented in Section V with conclusion briefed in Section VI.

II. HEART DISEASE DIAGNOSIS

In an early detection of heart disease the learning features of data base has a vital role. In this presented work Cleveland dataset [17] is used for learning and analysis of the developed approaches.

A. Dataset Description

As observed, there are multiple observing parameters for diagnosis for heart diseases. In [17] multiple physiological parameters is presented for early diagnosis of heart disease. For the implementation of the proposed work Cleveland Dataset is used. Most of the researchers used this Cleveland Dataset as the benchmark dataset. The dataset consists of 76 attributes out of which majority of Computational Techniques have chosen only 14 attributes. The 14 attributes that we have considered along with their details are as follows:

1) Age: A slow varying parameter specifying patient age.

2) Sex: A non-varying parameter specifying male or female patient.

3) Chest pain (CP): this is of four types, Typical angina (angina), Atypical angina (abnang), Non-anginal pain (notang), or Asymptomatic (asympt) pain. The conditions of this feature are defined by the past monitoring of patient history and pain in chest with illness conditions.

4) Trestbps: the BP condition of a patient under rest condition.

5) Chol: Defines cholesterol level in mg/dl in a patient.

6) Fbs: Fasting blood sugar level.

7) Restecg: This gives the patient ECG under rest condition.

8) Thalach: This defines the maximum heart rate observed.

9) Exang: A Boolean parameter defining the induced angina in exercise.

10) Oldpeak: Depression value obtained during exercise.

11) Slope: The slop parameter defines the ST region during exercise.

12) Ca: Indicate numbers of major vessels defined by fluoroscopy.

13) Thal: Indicates the heart status condition of a patient.

14) Num: (class attribute) values are 0 for healthy and 1, 2, 3, 4 for unhealthy.

The Cleveland heart disease dataset has five class attributes indicating either healthy or one of four sick types.

In the existing approach the parameters are developed for feature fusion to reduce the processing overhead, where the features are fused based on the data level and feature level of the observation. Wherein the fusion is based on the numerical and normalized values of the parameters, the fusion are magnitude based. This fusion limits the observation to a discrete time observation which has the probability of error in real time diagnosis.

In this work a new feature fusion approach is proposed based on the time line characteristic of measuring parameter in consideration with medical significance of the feature. This approach fuse the features based on the medical significance and time line variation of the monitoring parameter which results in more accurate decision at a lesser processing time compared to existing approach.

A single feature taken in isolation cannot figure out all individuals' risk of heart disease. Hence many features are required to diagnose it. Among the observed parameters features such as cholesterol, heart rate, hypertension (blood pressure), resting ECG, diabetes, blood sugar, stress, exercise induced angina and old age are significant in predicting heart disease. Out of these factors, eight of them are medically significant from the list of Cleveland heart disease dataset [18]. They are age, chest pain type, resting blood pressure, fasting blood sugar, cholesterol, maximum heart rate, resting heart rate and exercise induced angina. If medically significant features were neglected, then it has every chance to run into the risk of incorrect diagnosis. Cleveland heart disease data has got both continuous and discrete types of data. In the classification stage a deep learning approach is made to cluster the observing parameters and perform classification based on the modified feature set. For this cause, the testing accuracy of medical feature fusion with a weighted clustering is compared with the existing approach. The accuracy, sensitivity, specificity, recall, precision and F-measure parameters were used in evaluation of the classification performance.

B. Diagnosis Process

The process of heart disease diagnosis is developed based on the training feature and the classification method applied. In the presented diagnosis system a set of feature from Cleveland dataset is used for diagnosis. The process of decision making is developed based on the normalized distance metric as outlined in [1]. For a set of feature vector (Fv) given as,

$$Fv = \{f\,v_{1,1}, f\,v_{1,2}, \ldots, fv_{nl,nl}\} \qquad (1)$$

The classification is obtained using distance metric. The decision is derived using the minimization function of distance vector given as,

$$Selected\ distance\ (D) = \min(dist_i) \tag{2}$$

Where,

$dist$ defines the eclidian distance faotrgivne as,

$$Euclidian\ Distance,\ dist_i = \sqrt{\sum_{i=1}^{n} Qf - Tf_i} \tag{3}$$

Where, Qf is query sample feature, and

Tf is the trained database feature.

A normalized distance factor for the decision using Max-Min criterion is used given as,

$$Decision(f_{ti}, f_{li}) = \frac{dist(f_{ti}, f_{li}) - dist_{min}(f_{ti}, F)}{dist_{max}(f_{ti}, F) - dist_{min}(f_{ti}, F)} \tag{4}$$

Here, 'dist' represents Euclidian distance of feature of test sample and database trained features. The distance on testing feature (f_{ti}), selected feature set (f_{li}) and database feature (F) is computed. d_{max}, d_{min} indicate maximum and minimum Euclidian distance of testing signal with dataset feature. The decision is develop as a normalized ratio of maximum and minimum distance parameter of testing and training feature vectors.

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

III. Weighted Clustering with Time Line Monitoring

The distance based classification has a limitation of arbitrary variation in the magnitude of feature vector and a large search overhead due to linear search over the database. To overcome the addressed issue, a weighted clustering approach based on gain parameter is presented. In the representation of database information in a classification process, the raw database is mapped to a normalized feature set which has a normalized effect of magnitude variation. The process of feature monitoring is presented in "Fig. 1".

The updation normalized feature is derived as a mean difference of feature vector.

For a set of randomly distributed feature vector fv_i, the normalization is given by,

$$F(fv_1, fv_2, \ldots\ldots fv_n) = \sum_{i=1}^{n} \sum_{fv \in F} \left\| fv - \overline{fv} \right\|^2 \tag{5}$$

Here, fv is the feature vector in database and $\overline{fv}$ defines the mean of dataset given by,

$$\overline{fv} = \frac{1}{n} \sum_{fv \in F} fv \tag{6}$$

In the retrieval process the updated features are correlated over a set of test sample feature in making distance following minimum distance criterion. The illustration of search process is shown in "Fig. 2".

The distance vector is computed as a correlation metric C, given by,

$$C(fv, cl) = \sqrt{\sum_{i=1}^{n} (fv_i - cl_i)^2} \tag{7}$$

Where, the classification is developed based on the correlation of the feature vector (fv_i) and selected feature cluster (cl_i) for i^{th} test feature vector. However, the search overhead is considerably high for a large database. To minimize the effect a clustering approach is proposed. This presented approach cluster the available dataset into k-clusters based on the distance metric and the gain parameter. The cluster gain (CGn) attained due to updation in cluster is given by,

$$CGn = \frac{R(fv_i) - R(fv_i|ft_i)}{R(ft_i)} \tag{8}$$

Here, R defines the redundancy of a feature. The redundancy of a feature in a class is defined by the redundancy factor (R) given by,

$$R(f) = -\sum_{i=1}^{n} Pr(C_i) \log_2(Pr(C_i)) \tag{9}$$

where $Pr(.)$ is the probability function, and $Pr(C_i)$ defines the probability of ith information in a cluster C_i.

The magnitude variation has a direct impact on the clustering performance hence a weighted updated of the feature vector is used given as,

$$Upt(C, fv) = \sqrt{\sum_{i=1}^{n} \omega_i (C_i - fv_i)^2} \tag{10}$$

The weight parameter defines the tuning of the computed gain for an information fv under magnitude variation given by,

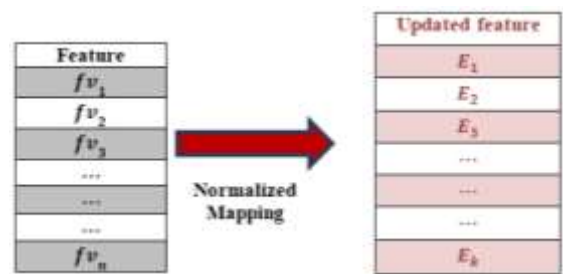$$\omega_{i+1} = \omega_i + dist(fv_i, C_i) \tag{11}$$
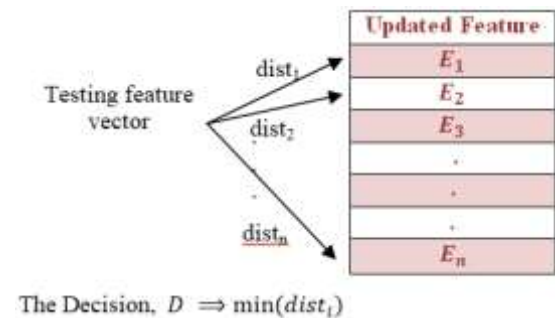

Fig. 1. Mapping Process for Feature Vector.

Fig. 2. Search Correlation over the Updated Dataset.

The weight parameter is updated by the distance parameter $dist(fv_i, C_i)$.

The convergence of the updation is given as a maximization function of cluster gain given as,

$$F_{convg} = \max\left(CGn_{\omega_i, fv_i, C)}\right) \tag{12}$$

The cluster values are updated based on the convergence of the feature value monitored for a period of time. For a time period the aggregated gain is computed and the cluster is updated based on the maximum cluster gain of a class. The classification of the developed system is performed using a multi class support vector machine (SVM). The decision is made for a set of training feature passed to the SVM architecture.

IV. RESULT OBSERVATION

The testing of developed approach is made over a Cleveland dataset where $1/3^{rd}$ part of the database is used for training and remaining is used for testing. The approach presented is outlined in Matlab tool and validated for the retrieval Accuracy (Acc), sensitivity, specificity, Recall Rate, F-measure and computation time parameter. The parameters are computed as,

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

The sensitivity is given,

$$sensitivity = \frac{TP}{TP + FN} \tag{14}$$

The specificity is given as,

$$specificity = \frac{FP}{TN + FP} \tag{15}$$

Recall rate is given by,

$$Recall = TP/(TP + FN) \tag{16}$$

The precision is given as,

$$Precision = TP/(TP + FP) \tag{17}$$

And F_Measure is computed as,

$$F\_measure = (2 * Recall * Precision)/(Recall * Precision) \tag{18}$$

The observing parameters are developed with the observing factors listed in Table I.

The confusion matrix for the heart diagnosis is illustrated in Table II.

TABLE I.   OBSERVATION FACTOR

(True Positive) TP	True match of test query feature and training feature
(True Negative) TN	False match of test query feature with true training feature
(False Negative) FN	False match of test query feature and training feature
(false positive) FP	True match of negative test query feature and training feature

TABLE II.   CONFUSION MATRIX FOR HEART DIAGNOSIS

observation	True match	False match
True match	TP	FP
False match	TN	FN

The observation for the developed system is computed for different test cases measuring the evaluating parameters. The observation for the developed approach under different test cases of healthy, type-1, type-2, type-3 and type-4 attributes is listed in Table III.

The developed approach of weighted cluster based classification is compared with the existing approach of random clustering and linear searching method. Observation plots for the healthy case samples are shown in "Fig. 3-10".

"Fig. 3" shows the accuracy comparison obtained for the developed approaches in comparisons with the existing classification approach of linear search and random clustering. The linear search method correlates the test feature vector to the database feature in a linear manner and develops decision based on the maximum correlation value. The average distance of the mapping is considered in making the decision. Random clustering is developed based on the manual selection of features for classification. The set of features are manually picked from the test sample to pass to the classifier model for decision. The observation of accuracy illustrated an improvement of 8% as compared to random clustering and 18% as compared from linear search method.

The Sensitivity of the developed method is shown in "Fig. 4". The proposed method shows a Sensitivity of 91% wherein the existing approach of linear search and random clustering shows a Sensitivity of 87% and 77% respectively.

TABLE III.   ANALYSIS OF DEVELOPED CLASSIFICATIONS FOR CASE HEALTHY TYPE

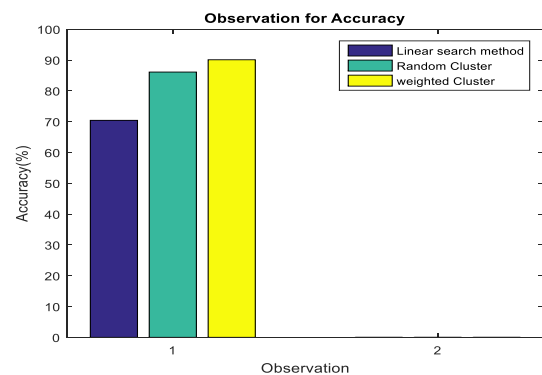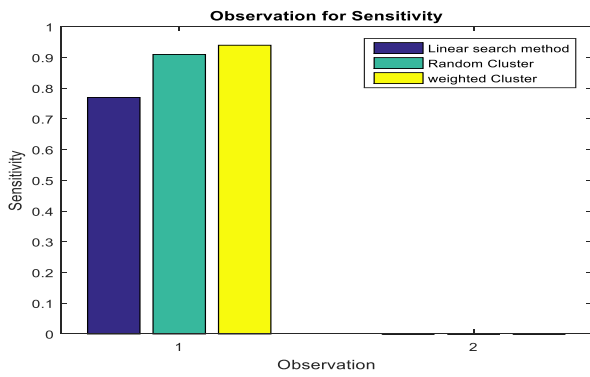| Method	Accuracy (%)	Sensitivity	Specificity	Recall	Precision	Time (sec)
Linear search method	70.42	0.77	0.45	0.77	0.79	0.64
Random cluster	86.09	0.91	0.48	0.91	0.92	0.43
Weighted Cluster	90.08	0.94	0.5	0.94	0.95	0.27

Fig. 3.   Observation of Accuracy Plot for Healthy Case.

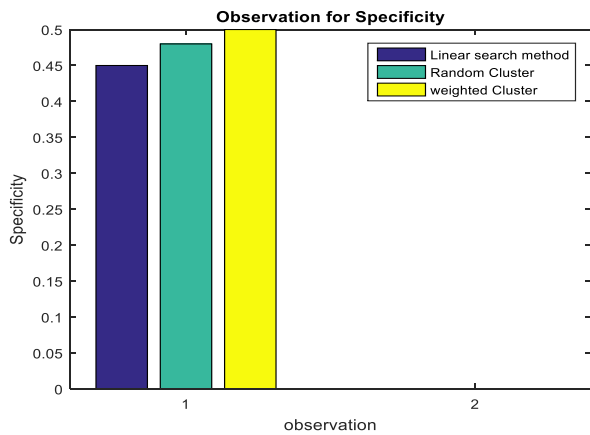Fig. 4.   Observation of Sensitivity Plot for Healthy Case.

The Precision of the developed method is shown in "Fig. 7". The proposed method shows a Precision of 93% wherein the existing approach of linear search and random clustering shows a Precision of 1% and 79% respectively. An analysis of the developed approach for different test types is summarized in Table IV.

Processing time defined as the computation time for classification is shown in "Fig. 8". The proposed method takes a computation time of 0.26 sec in classification, wherein the existing approach of linear search and random clustering observe a computation time of 0.43 and 0.65 sec, respectively.
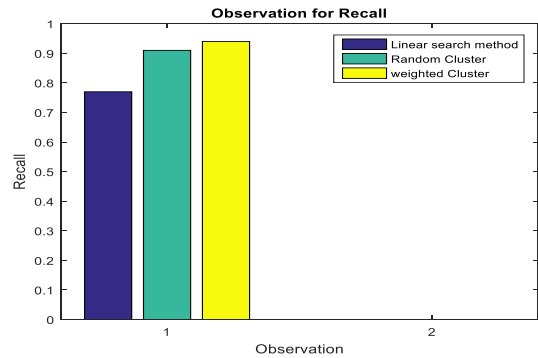

Fig. 5.   Observation of Specificity Plot for Healthy Case

Fig. 6.   Observation of Recall Plot for Healthy Case.

The specificity of the developed method is shown in "Fig. 5". The proposed method shows a specificity of 51%. The existing approach of linear search and random clustering shows a specificity of 46% and 44%, respectively.

The recall rate plot is shown in "Fig. 6". The proposed method obtain a recall rate of 91% and the existing approach of linear search and random clustering shows a recall rate of 89% and 77%, respectively.
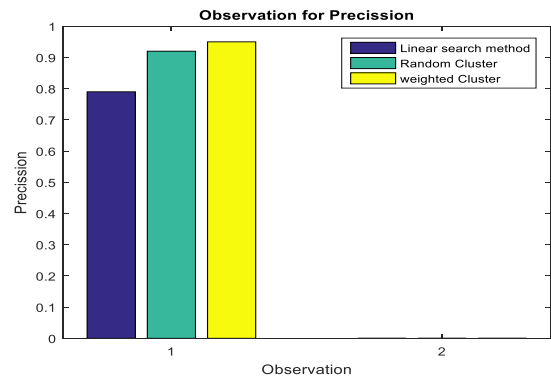

Fig. 7.   Observation of Precision Plot for Healthy Case.

TABLE IV.   ANALYSIS OF DEVELOPED CLASSIFICATIONS FOR CASES TYPE-1-4

| Case Type	Method	Accuracy	Sensitivity	Specificity	Recall	Precision	Time (sec)
Type - 1	Linear search method	77.35	0.82	0. 81	0.87	0.52	0.64
	Random cluster	85.22	0.91	0.84	0.87	0.421	0.43
	Weighted Cluster	89.08	0.93	0.95	0.98	0.745	0.27
Type - 2	Linear search method	83.35	0.89	0.78	0.77	0.65	0.58
	Random cluster	86.22	0.86	0.84	0.87	0.72	0.41
	Weighted Cluster	90.08	0.94	0.95	0.91	0.85	0.20
Type - 3	Linear search method	79.35	0.86	0.82	0.83	0.77	0.78
	Random cluster	86.22	0.91	0.84	0.89	0.84	0.56
	Weighted Cluster	91.08	0.95	0.91	0.93	0.89	0.31
Type - 4	Linear search method	63.35	0.67	0.78	0.79	0.78	0.81
	Random cluster	88.22	0.88	0.84	0.88	0.90	0.53
	Weighted Cluster	92.08	0.95	0.92	0.91	0.94	0.33
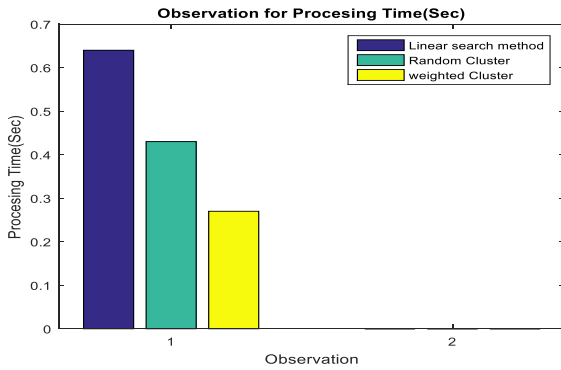
Fig. 8.   Observation of Processing Time (Sec) Plot for Healthy Case

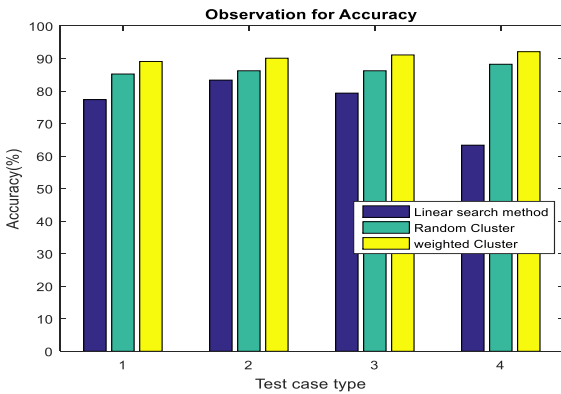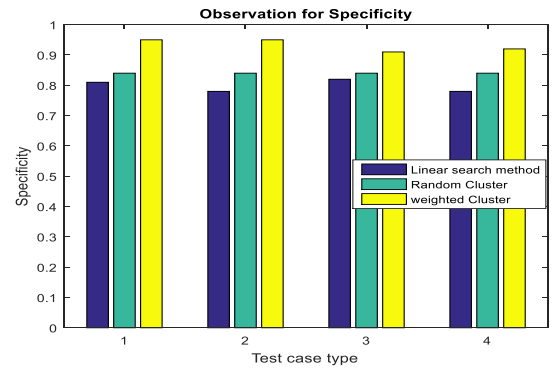Fig. 9.   Observation of Accuracy Plot for Type 1-4 Case.

Observation for four types of effects refereed in Cleveland data set is compared for accuracy. The observation plot for accuracy for the three developed classification approach shown in "Fig. 9". The observation shows an accuracy of 89% for type-190% for type-2, 91% for type-3 and 92% for type-4 cases.

Observation plot for Sensitivity for the three developed classification approach shown in "Fig. 10". The observations show a Sensitivity of 93% for type-1, 94% for type-2, 95% for type-3 and 95% for type-4 cases.

For Specificity for the three developed classification approach is shown in "Fig. 11". The observations show a Specificity of 95% for type-1, 95% for type-2, 91% for type-3 and 92% for type-4 cases.

Recall for the three developed classification approach is shown in "Fig. 12". The observations show a Recall of 98% for type-1, 91% for type-2, 93% for type-3 and 91% for type-4 cases.

Precision for the three developed classification approach is shown in "Fig. 13". The observations show a precision of 74% for type-1, 85% for type-2, 89% for type-3 and 94% for type-4 cases.

Processing time for the three developed classification approach is shown in "Fig. 14". The observations show a Processing time of 0.27 sec for type-1, 0.20sec for type-2, 0.31sec for type-3 and 0.33 sec for type-4 cases.

Fig. 10.   Observation of Sensitivity Plot for type 1-4 Case.

Fig. 11.   Observation of Specificity Plot for type 1-4 Case.
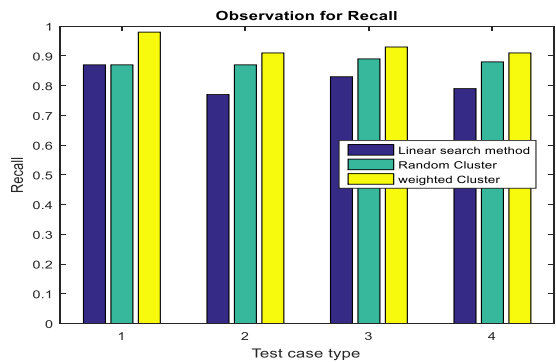

Fig. 12.   Observation of Recall Plot for Type 1-4 Case.
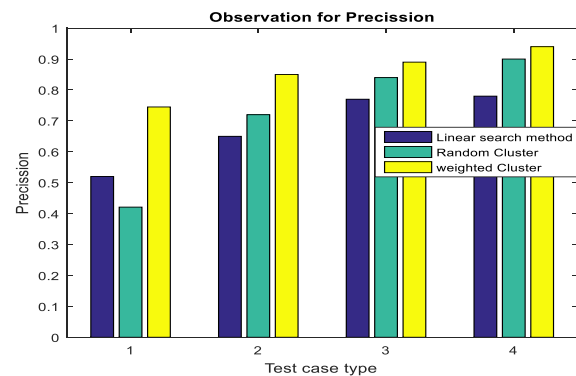

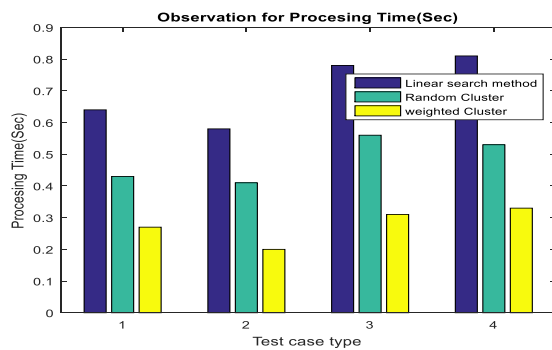Fig. 13.   Observation of Precision Plot for Type 1-4 Case.

Fig. 14. Observation of Processing Time Plot for Type 1-4 Case.

V. Discussion

Diagnosis of heart disease based on the vitals measured is a primary need for automated diagnosis and classification system. In the presented work a weighted clustering approach for feature fusion is proposed. The analysis of the proposed approach for classification accuracy, sensitivity, specificity, recall, precision and computation time resulted in more efficient observations compared to the existing approach of linear search and random clustering method. The process of time line monitoring and denoising results in faster processing resulting g in a minimization of 0.4sec in processing time compared to linear search method. The accuracy of classification is observed to improve by 20% compared to linear search method. For varying test type parameters from 1-4 it is observed that higher accuracy for type 4 case is observed. The test time is lower for type-1 case. Wherein the classification due to linear search has high search time, clustering based approach is suitable in minimizing the processing time with higher accuracy. The random clustering selects the features from the list of observation based on the magnitude values. The distortion impacts the retrieval accuracy. However, the proposed approach illustrates higher classification accuracy with lower time of processing. The analysis is focus on testing the proposed method onto real time vital parameters as future work.

VI. Conclusion

The diagnosis of heart disease using weighted clustering is presented in this paper. The approach of weight updation and clustering of large database feature based on cluster gain factor is proposed. This approach developed a new clustering approach of a distributed feature set in Cleveland data set into specified cluster to improve the classification performance. The search overhead in mining of features for classification is minimized by the formations of sub clusters. The decision approach of feature update into cluster is made more accurate with the dual factor of class attribute and cluster gain factor. The observations of the developed system illustrated an improvement in accuracy for the proposed system compared to existing approach with a reduced time of computation due to cluster search approach.

References

[1] Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam , Daehan Kwak, Amjad Ali f , Muhammad Imran g , Kyung-Sup Kwak h , "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," Information Fusion, No.63, pp-208–222, 2020.

[2] Rohé M M, Datar M, Heimann T, et al. "SVF-Net: Learning deformable image registration using shape matching, International Conference on Medical Image Computing and Computer-Assisted Intervention," Springer, Cham, 2017: 266-274.

[3] Gao Z, Xiong H, Liu X, et al., " Robust estimation of carotid artery wall motion using the elasticity-based state-space approach," Medical image analysis, 2017, 37: 1-21.

[4] Zhen X, Zhang H, Islam A, et al., "Direct and simultaneous estimation of cardiac four chamber volumes by multioutput sparse regression," Medical image analysis, 2017, 36: 184-196.

[5] Gao Z Zhao S, Gao Z, Zhang H, et al., "Robust Segmentation of Intima–Media Borders With Different Morphologies and Dynamics During the Cardiac Cycle," IEEE journal of biomedical and health informatics, 2017, 22(5): 1571-1582.

[6] E. J. Benjamin, S. S. Virani, C. W. Callaway, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, and F. N. Delling, et. al., "Heart disease and stroke statistics-2018 update: a report from the American Heart Association," Circulation, Vol. 137, No. 12, pp. 67- 492, 2018.

[7] Xu L, Huang X, Ma J, et al., "Value of three-dimensional strain parameters for predicting left ventricular remodeling after ST-elevation myocardial infarction," The International Journal of Cardiovascular Imaging, 2017, 33(5):663-673.

[8] Thiago M. Nunes, Victor Hugo C. de Albuquerque,et. al., "Automatic microstructural characterization and classification using artificial intelligence techniques on ultrasound signals," Expert Systems with Applications, 2013, 40(8):3096-3105.

[9] Khamparia A, Saini G, Gupta D, et al.,"Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network," Circuits, Systems, and Signal Processing, 2019(May).

[10] Albuquerque V H C D, Nunes T M, Pereira D R, et al.,"Robust automated cardiac arrhythmia detection in ECG beat signals," Neural Computing & Applications, 2018, 29(3):1-15.

[11] W. H. Organization and others, "World health statistics 2017: monitoring health for the SDGs, sustainable development goals," 2017.

[12] R. I. Kementrian Kesehatan, "Riset kesehatan dasar," Jakarta Kementerian. Kesehatan. RI, 2013.

[13] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Time series classification using multi-channels deep convolutional neural networks," in International Conference on Web-Age Information Management, 2014, pp. 298–310.

[14] D. R. Chowdhury, M. Chatterjee, and R. K. Samanta, "An artificial neural network model for neonatal disease diagnosis," Int. J. Artif. Intell. Expert Syst., vol. 2, no. 3, pp. 96–106, 2011.

[15] G. Subbalakshmi, K. Ramesh, and M. C. Rao, "Decision support in heart disease prediction system using naive bayes," Indian J. Comput. Sci. Eng., vol. 2, no. 2, pp. 170–176, 2011.

[16] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," Expert Syst. Appl., vol. 40, no. 1, pp. 96–104, 2013.

[17] https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[18] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," Expert Systems with Applications, No.40, pp- 96–104, 2013.

[19] Santosh A. Shinde, Dr. P. Raja Rajeswari, "A Novel Hybrid Framework for Cuff-Less Blood Pressure Estimation based On Vital Bio Signals processing using Machine Learning," International Journal of Advanced Trends in Computer Science and Engineering, Vol.9, Issue 2, March-April 2020, pp. 1556-1561.

[20] Mr. Santosh A. Shinde , Dr. P. Raja Rajeswari, "Intelligent health risk prediction systems using machine learning: a review," International Journal of Engineering & Technology, Vol.7, Issue 3, 2018, pp. 1019-1023.