International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 2 (2019), pp. 204-211

Deep Learning based Human Diseases Pattern Prediction Technique for High Dimensional Human Diseases Data Sets

N.K. Sakthivel[1,*], N.P. Gopalan[2], S. Subasree[3]

[1]*Research Scholar, Bharath University, Chennai, Tamil Nadu, India.*

[2]*Professor, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India.*

[3]*Nehru Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India.*

Abstract

For decades, more and more experimental researches have collectively indicated that microRNA (miRNA) could play a vital role in many important biological processes and thus it also the pathogenesis of human complex diseases. It is also noticed that the resource and time cost requirement for processing data in traditional biological method is more expensive and thus more and more focusing have been paid to the enhancement of effective and accurate computational mechanisms for predicting potential associations between diseases. To focus towards this, researchers identified that Gene is not responsible for many Human Diseases and instead, diseases occur due to interaction of different group of genomes that is responsible for different diseases. Hence it is very important to analyze and associate the complete genome sequences and its associations to understand or predict various possible human diseases. To identify and predict the associations between diseases, We earlier proposed the Deep Learning based Intelligent Human Diseases-Gene Association Prediction Technique (IHDGAP). It employs Convolution Neural Network (CNN) algorithm which contains multiple number of hidden layers which is helping to predict Gene Patterns and its Associations to predict Human Diseases. This Research Work revealed that our previous Model, IHDGAP unable to classify boundary region. To address this issue to enhance Classification Accuracy, this research work proposed an enhanced Classifier called Deep Learning based Human Diseases Pattern Prediction Technique (ECNN-HDPT), where an Enhanced CNN is employed. The proposed model is implemented and analyzed carefully in terms of Processing Time, Memory Usage/Utilization, Accuracy, Sensitivity, Specificity and FScore. From the experimental results, it is noticed that the proposed deep learning mechanism improves the performances of the proposed classifier in terms of Classification Accuracy, Sensitivity, Specificity, and FScore as compared with our previous model Gene Signature based Hierarchical Random Forest (G-HRF) and Intelligent Human Diseases- Gene Association Prediction Technique (IHDGAP). However, it was noticed that the proposed model consumes relatively more Memory and Processing Time as we employs SVM-ECNN to train and test Patterns. Though it takes relatively more time to train the Data Sets, it is established that the Classification Accuracy is higher as compared with our previous Classifier.

Keywords: Gene Hierarchical based Random Forest(G-HRF), Intelligent Human Disease Gene Association Prediction(IHDGAP), Deep Learning, Convolution Neural Network(CNN),Confidence Divider, SVM-ECNN, Association mining, Prediction Accuracy.

I. INTRODUCTION

DNA Microarrays designed to focus for measuring the transcriptional levels of DNA and RNA transcripts. The signature of Gene Expression in the biomedical field used to identify a few Human Disease Patterns[1,2,17,18]. Associating genes with genotypes or phenotypes is demanding research topic in bioinformatics which is called as disease-gene association research. This is also called as identification or prediction of diseased genes.

From the literature survey, it was noticed that the identification and recognition of gene diseases have been a long goal of biomedical research. It helps Researchers to understand the gene function, the interactions and pathways towards improvement and contributions of medical care. There may be more number of traditional methods of gene analysis mechanisms are available but all these methods are having its own unique disadvantages. Even though the association analysis mechanism work well to a set of selected functional set of genes, the selection of genes are not straight forward. Thus we unable to apply the specialized knowledge and hence it is considered as a limitation of this association analysis.

They are many network-based algorithmic approaches have been proposed and identified for classifying Gene-Disease Associations, but most of these methods simply focus to view the objects in gene-phenotype heterogeneous networks as the same type and it does not focus the different meaning behind

[*]*Corresponding author E-mail:nksakthivel@gmail.com*

the gene path. To address the above identified issues, this research work focuses to identify the association between the gene associations to predict various diseases.

To identify and predict the associations between diseases, this research work was proposed earlier the Deep Learning based Intelligent Human Diseases- Gene Association Prediction Technique for High Dimensional Human Diseases Data Sets (IHDGAP). This will predict the association between the diseases. It employs Convolution Neural Network (CNN)[22,23,24,26] algorithm which contains multiple number of hidden layers which is helping to predict Gene Patterns and its Associations to predict Human Diseases.

As our previous model fails to classify uncertain data/data in boundary region, the reclassification using SVM-CNN[22,25,27] is needed and it is performed in the proposed model. This helps to reclassify Diseases Patterns with high higher Classification Accuracy.

The remaining sections of this work are organized as follows: Section 2 and Section 3 describe our previous models i. G-HR Gene Signature based HRF Cluster and ii. Deep Learning approach and Association Predictions and Intelligent Human Diseases Prediction Technique through Gene Association Prediction Mechanism (IHDGAP). The identified issues were discussed in Section 4 and the proposed Model called an Human Diseases Pattern Prediction Technique (ECNN-HDPT) is discussed in Section 5. Section 6 presents the findings and performance Analysis of the proposed model ECNN-HDPT and concluded in Section 7.

II. G-HR : GENE SIGNATURE BASED HRF CLUSTER

Identifying Gene Signatures[8,9,10,11] for predicting the various Gene Patterns with highest accuracy is most essential and that could be employed to build high accuracy Gene Classifier/Predictor for clinical tests and applications[1,6,7,16,19]. Thus an efficient Gene Signature based HRF Cluster called G-HR was proposed. This is our previous proposed model. The Procedure is elaborately discussed in the following section. The architecture of the Genetic Signature based Hierarchical Random Forest is shown in the Fig. 1 to achieve better Pattern prediction and classification accuracy

II.I G-HR Procedure

The procedure of the G-HR[1,6] method is follows. This can identify gene sets that are associated with genes expression and its subset clusters. It will form Clusters based on the distances of points which can calculate with Euclidean Distance Model. This model was capable of merging clusters depends on its sizes.

It is capable of eliminating noises and outlies so that the misclassification can be reduced which will help to maximize the classification accuracy.

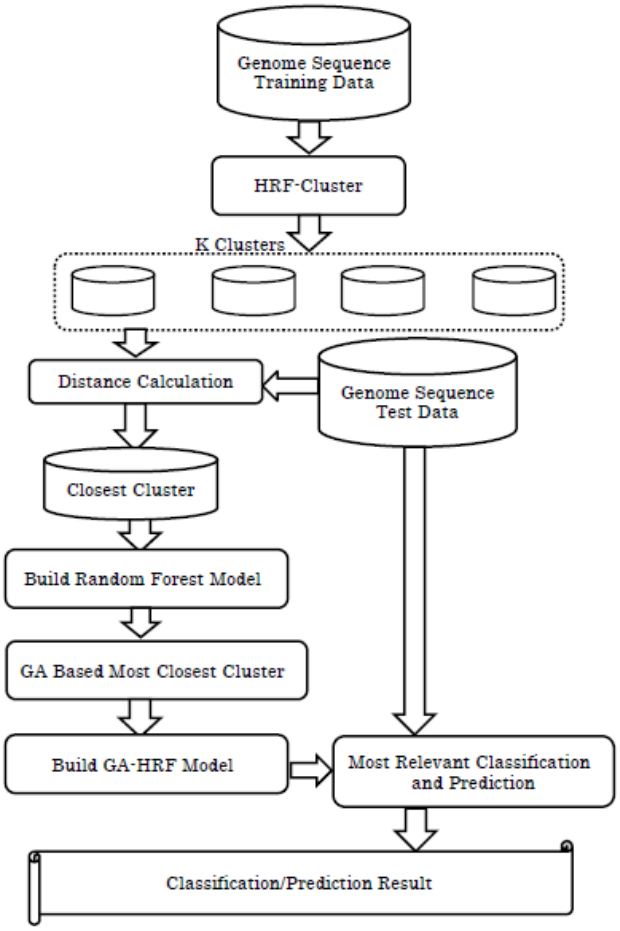

Fig. 1 Genetic Signature based Hierarchical Random Forest Cluster (G-HR Cluster)

The Closest Cluster built by Hierarchical Random Forest Model[8,9,10] was further optimized through Genetic Algorithm based Hierarchical Random Forest Model. As a whole, this proposed model achieves higher classification accuracy.

Step 1: *Collect Genome Sequence Training Data*

Step 2: *Create Multiple Clusters through Euclidean Distance*

Step 3: *Find Similar Clusters based on distance Calculated*

Step 4: *Find Clusters with less points and merge together through Hierarchical Cluster*

Step 5: *Validate through Hierarchical Random Forest*

Step 6: *Minimize Misclassification Rate through GA-HRF*

Step 7: *Maximize Area Under Curve (AUC) Measurement*

Step 8: *Select Most Closest Cluster through GA-HRF*

Step 9: *Remove Redundant Clusters through Spearman Rank Correlation Model*

$$\rho = 1 - \frac{6\sum d^2}{N(N^2 - 1)} \quad (1)$$

III. DEEP LEARNING APPROACH AND ASSOCIATION PREDICTIONS

The Machine Learning methods were introduced by researchers to improve Classification Accuracy for predicting target diseases patterns. In this section, the features of Machine Learning and its association predictions were discussed.

III.I. Deep Learning approach

Deep Learning[3,4,5] is a Machine Learning Technique in which a model learns and predict the classification patterns directly from Images, Text, or Sound. Deep Learning uses a Neural Network Architecture. The term "deep" refers to the number of layers in the network if more layers, the deeper the network. Traditional Neural Networks contain only minimum number of layers, while Deep Networks can have hundreds of layers.

Deep Learning is especially well-suited for better Prediction/Classification and a few applications are like Classification and Prediction of various Diseases[13,14,15,20,21], Face Recognition, Text Translation, Voice Recognition, and Advanced Driver Assistance Systems, including, Lane Classification and Traffic Sign Recognition.

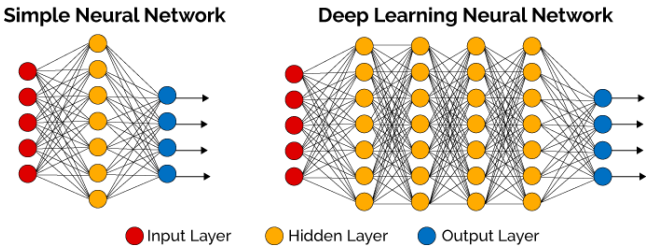

Fig. 2 Deep Learning Neural Networks

As shown in the Fig. 2, a Deep Learning Neural Network combines multiple nonlinear processing layers, using simple elements operating in parallel and inspired by biological nervous systems[12]. It consists of an Input Layer, several Hidden Layers, and an Output Layer.

The layers are interconnected via Nodes, or Neurons, with each hidden layer using the output of the previous layer as its input.

III.II. Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP)

This Research work was focused to improve the Classification/Prediction Accuracy of Classifier. It proposed and deployed Deep Learning based Convolution Neural Network(CNN) algorithm to enhance the performances of Classifier in terms of Processing Time, Memory Usage/Utilization, Accuracy, Sensitivity, Specificity and FScore.

III.II.I. Convolution Neural Network (CNN)

A Convolutional Neural Network(CNN) is one of the most popular algorithms for Deep Learning. Like other Neural Networks, a CNN is composed of an Input Layer, an Output Layer, and many Hidden Layers in between.

Feature Detection Layers are responsible for performing one of three types of operations on the data. It performs either Convolution or Pooling or Rectified Linear Unit (ReLU). Convolution puts the input data through a set of convolutional filters, each of which activates certain features. Pooling simplifies the output by performing nonlinear down sampling, reducing and identifying the number of parameters that the network needs to learn.

Rectified Linear Unit (ReLU) allows for faster and more effective training by mapping negative values to zero and maintaining positive values. These three operations are repeated perform over tens or hundreds of layers, with each layer learning to detect different features and improve the accuracy.

The following Fig. 3 shows the architecture of CNN with its Operations.

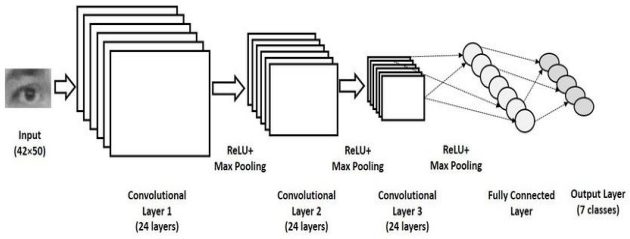

Fig. 3 Architecture of CNN

The architecture of Deep learning based Convolution Neural Network(CNN) is shown in Fig. 3. It consists of six layers, one input layer, two convolutional layers and two sub-sampling layers and an output layer. As shown in the Architecture, the convolutional layers are labeled C_i , and the sub-sampling layers are labeled S_i, where i is the layer index. The C_i layer is obtained by performing a convolution of the previous layer data and adding a bias.

The aim of the convolution operation is to enhance the original characteristics and remove the noise information. The S_i layer is obtained by calculating the average of four inputs and it will be multiplied with the average of training coefficient and these coefficient with bias and forward the result through a sigmoid function. The main focus of subsampling is to reduce the data processing without losing the useful information.

Each neuron is defined as follows n(l, m, j) where i , m and j denote the layer, map and neuron's position in the map, respectively. The value of a neuron is defined as $v^i_m(j)$,

$$v^i_m(j) = f\left(x^l_m(j)\right) \quad (2)$$

where f depends on the layer, and $x^i_m(j)$, represents the scalar product between a set of input neurons in layer and the weight

connections between these input neurons in the layer *l*-1 and the neuron number *j* in map *m* in layer *l.*

Let define $x^{i}_{m}(j)$ first convolution layer as follows.

$$x_m^l(f) = w(1,m,0) + \sum_{i=0}^{i<k} I_{i,j}\, w(1,m,i) \quad (3)$$

and define $x^{i}_{m}(j)$ for other convolution layers

$$x_m^l(j) = w(l,m,0) + \sum_{i=0}^{i<k} v_m^{i-1}\,(j*k+i) w(1,m,i) \quad (4)$$

where $I_{i,j}$ stands for original input data in "Input Layer". i is the index of each element in the kernel and the value of i is {0,1,2}. k denotes the size of the kernel, and the value of k in this work is "3". $w(l,m,i)$ denotes the weight of each connection and $w(l,m,0)$ is the weight of bias. The convolutional layer aims to find the most useful information for the classification.

III.II.II. Disease Semantic Similarity

The main aim of this method is to identify the relationship among different diseases that can be represented using Direct Acyclic Graph (DAG). Specifically an arbitrary disease D, where T(D) consisted of Node D itself and all its ancestor node, E(D) is a corresponding edge set, consists of directed edges pointing to child nodes to parent nodes. The DAG can be calculated as follows.

$$DAG(D) = \big(D, T(D), E(D)\big) \quad (5)$$

The semantic value of disease D is calculated as

$$DV(D) = \sum_{d \epsilon T(D)} D_D\,(d) \quad (6)$$

$$\begin{cases} D_D(d) = 1 & f\ d = D \\ D_D(d) = \max(\Delta, D_D(d) \,/\, d \epsilon\ children\ of\ d\,\} & if\ d \neq D \end{cases} \quad (7)$$

Where Δ was the semantic contribution factor. For a given disease D, negative correction exist between D and another disease d, and the contribution score of d of disease D. Disease locating in the same layer would contribute the same score to semantic value of disease D. The Semantic Similarity(SS) is calculated between disease d(i) to disease d(j) is calculated as follows:

$$SS\big(d(i), d(j)\big) = \frac{\sum_{t \epsilon T(d(i)) \cap T(d(j))} (D_{d(i)}\,(t) + D_{d(j)}\,(t))}{DV\big(d(i)\big) + DV(d(j))} \quad (8)$$

III.II.III. Computation of Scoring Matrix

The main aim of identifying the semantic similarity is to identify the association between the diseases. If the association is exist are set to 1 and the elements which represent all unknown associations are set to 0. For each element a new adjacency matrix is calculated. The changed values are ranked from all samples. After getting ranks for all samples, the threshold calculation is done. If the rank is less than the threshold value, the prediction is negative otherwise the prediction is positive. For each threshold True Positive Rate (TPR – Sensitivity) and False Positive Rate (FPR – Specificity) can be calculated.

Sensitivity = True Positive / (True Positive + False Negative) (9)

Specificity = True Negative /(True Negative + False Positive) (10)

IV. IDENTIFIED PROBLEM

The Gene Signatures for predicting the various Gene Patterns with highest accuracy is most essential and that could be employed to build high accuracy Gene Classifier/Predictor. This is needed for clinical tests and applications. G-HR is an efficient Gene Signature based clustering mechanism which is used to identify the multiple clusters to predict the accuracy of Gene Classification and Predictions. It is capable of eliminating noises and outliers so that the misclassification can be reduced which will help to maximize the classification accuracy.

But however it is predicted that, if we associate group of genes that were responsible for diseases, then the Diseases Prediction Accuracy will be better than that of G-HR. ie our previous model unable to classify or predict Gene data in better manner as we didn't group disease associated genes. We proposed Deep Learning based Intelligent Human Diseases- Gene Association Prediction Technique (IHDGAP) to improve Classification Accuracy. But study reveals that it fails to classify properly at the boundary region.

To address the above mentioned issue, this research work proposed an efficient Technique called, Deep Learning based Human Diseases Pattern Prediction Technique for High Dimensional Human Diseases Data Sets (ECNN-HDPT). The proposed Model was developed based on SVM-CNN and this hybridization is called as Enhanced CNN(ECNN). The proposed model will classify and predict the Gene Patterns with better accuracy as compared with our previous model G-HR and IHDGAP.

V. PROPOSED ARCHITECTURE

The proposed architecture of the Deep Learning based Human Diseases Pattern Prediction Technique called ECNN-HDPT is shown in the Fig. 4. The proposed model was designed to predict the various Human Diseases Patterns based on Enhanced CNN. That is the Enhanced CNN is a Hybrid System developed with SVM and CNN. This Model is the enhanced Model of our previous Classifier, Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP). That is the Confidence Divider and SVM-CNN is introduced in the proposed Model to enhance Classification Accuracy.

The various steps that involved in the proposed model is clearly shown in the Fig. 4.

As a whole, this proposed model achieves higher classification accuracy by using Enhanced Convolution Neural Network (ECNN) through SVM and Deep Learning Mechanisms. To achieve the proper Classification in Boundary / Uncertainty

Data, the proposed model involves SVM-CNN for higher Classification Accuracy with Confidence Divider. The following steps are used to improve the classification accuracy of proposed model Human Diseases Pattern Prediction Technique.

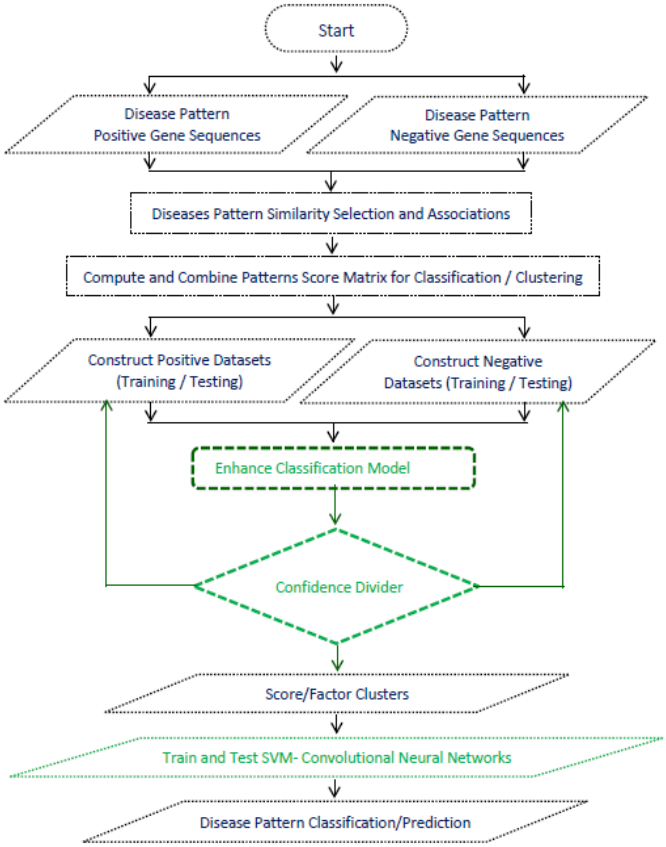

Fig. 4 Proposed Architecture of the Human Diseases Pattern Prediction Technique (ECNN-HDPT)

Step 1: Collect Genome Sequence Training Data from Database

Step 2: Consider both the Positive Gene Sequences and Negative Gene Sequences and Classify the input patterns into the mentioned two Gene Sequences Classification.

Step 3: Identify the similarity of Gene Sequences and Close Association among Gene Sequences.

Step 4: Compute the Pattern Score Matrix for the selected Gene Sequences

Step 5: From the Pattern Score Matrix, construct effective Positive and Negative Datasets for Training and Testing as well

Step 6: Enhanced Classification Model Calculate Confidence Values to correctly classify the current Inputs. Ie Correct Classification Level is towards 1 and Misclassification Level is towards 0.

Step 7: Calculate the new scores for the newly constructed Datasets.

Step 8: Input the Constructed Datasets to Train and Test SVM-Convolution Neural Network (SVM-CNN) to enhance Classification Accuracy

Repeat for Dataset Optimization for achieving higher accuracy with Association Rules.

Step 9: Compare the accuracy periodically.

Step 10:Record the Disease Pattern classifications/Prediction.

VI. PERFORMANCE ANALYSIS

The experimental set up and simulations are carried out by this research work by using the Genome Sequence Data Sets, Master.MER. This was downloaded from NCBI for thorough study.

Simulations are conducted to examine the performances in terms of classification and prediction abilities of the proposed Deep Learning based Human Diseases Pattern Prediction Technique (ECNN-HDPT). The results were compared with our previous Models, Gene Signature based HRF Cluster (G-HR) and Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP).

This work considered 10 different Genome Genes Data Sets categories for predicting possible diseases and each category has 50,000 records and in total there are 500000 records used for performance analysis of the proposed model. The experiments were repeated number of times and for classifying and predicting possible diseases were recorded.

The performances of the above discussed Genome Classifiers have been studied in terms of in terms of Processing Time, Memory Usage/Utilization, Classification Accuracy, Sensitivity, Specificity and FScore.

This Research Work has developed the Interfacing Tool with the help of VC++ Programming Language with R programming for computation to extract and validate the Gene Expressions which are downloaded from NCBI. The validated data is fed into BioWeka for analysing the proposed Genome Classifiers in terms of Processing Time, Memory Usage/Utilization, Accuracy, Sensitivity, Specificity and FScore.

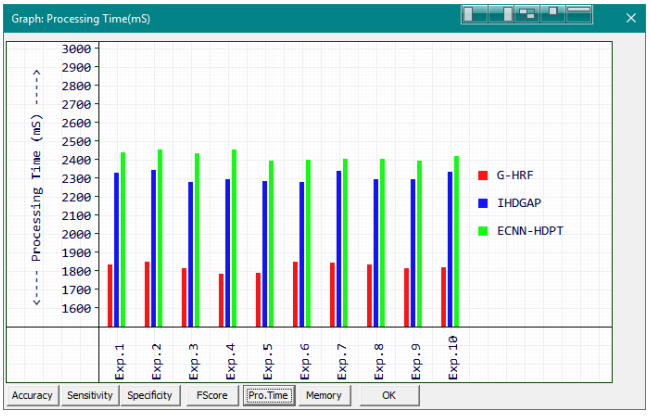

Fig. 5 Processing Time vs Classifiers (Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

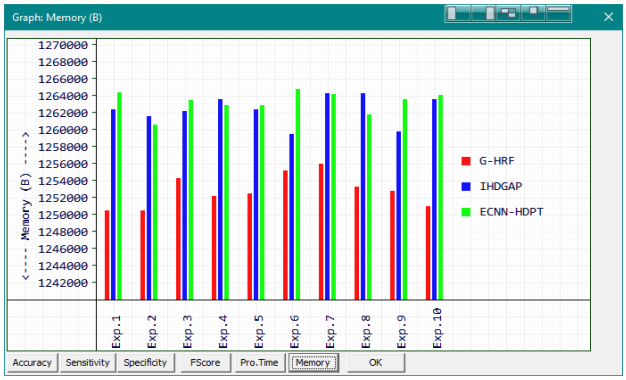

Fig. 6 Memory Usage vs Classifiers(Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

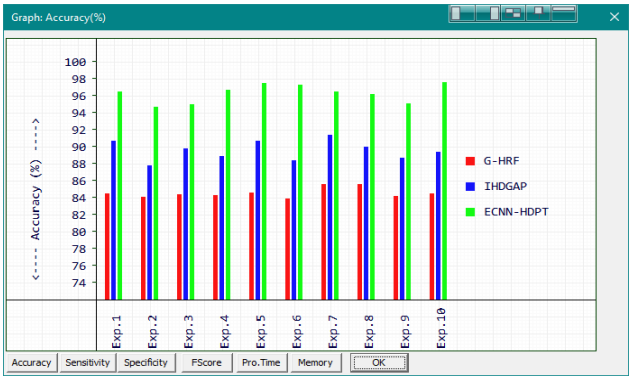

Fig. 7 Classification Accuracy vs Classifiers (Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

The experimental results of the proposed Model, Human Diseases Pattern Prediction Technique (ECNN-HDPT) is compared with our previous model called Gene Signature based HRF Cluster (G-HRF) and IHDGAP in terms of Processing Time, Memory Usage/Utilization, Accuracy, Sensitivity, Specificity and FScore and analyzed thoroughly. From the results, it was noticed that the proposed classifier is performing well which are shown in the figures Fig. 7, Fig. 8, Fig. 9, Fig. 10.

From the Fig. 5 and Fig. 6, it was clearly observed that the Processing Time and Memory Usage of our proposed Model, ECNN-HDPT is relatively high as compared with G-HRF and IHDGAP Classifiers as the proposed model employs SVM-CNN and Confidence Divider for training/testing the data to achieve better classification accuracy.

From the Fig. 7, it was clearly noticed that the Classification Accuracy of the proposed model ECNN-HDPT is better than that of our previous Classifiers, G-HRF and IHDGAP.

From the Fig. 8 and Fig. 9, it is observed that the proposed model ECNN-HDPT is performing well in terms of Sensitivity and Specificity as compared with our previous model called Gene Signature based HRF Cluster (G-HRF) and IHDGAP.

This is also noticed that our proposed model is reduced misclassification as compared with our previous model G-HRF and IHDGAP. That is, the prediction scores of True positive and True Negative high and False Positive and False Negative are very low.

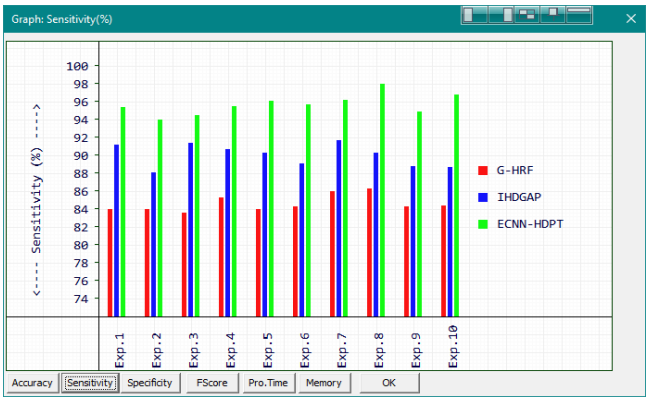

Fig. 8 Sensitivity vs Classifiers (Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

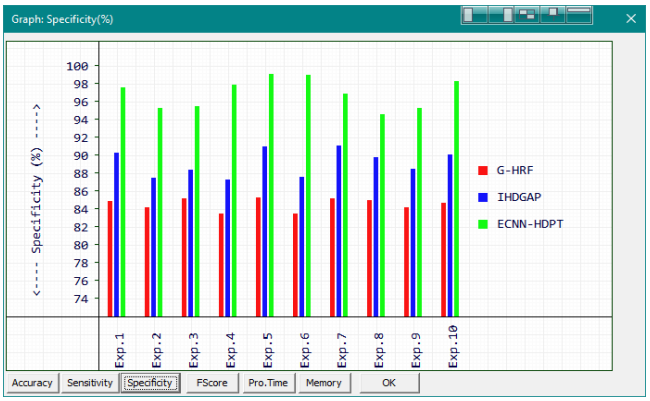

Fig. 9 Specificity vs Classifiers (Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

From the Fig. 10, it is clearly noticed that the achieved FScore of the proposed model ECNN-HDPT is better than that of G-HRF and IHDGAP. That is it is clearly established that the proposed model is classifying and predicting diseases in better manner.

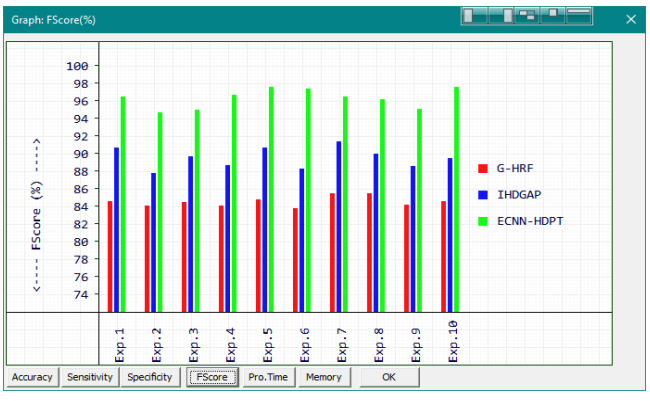

Fig.10 FScore vs Classifiers(Proposed ECNN-HDPT and our previous Models G-HRF and IHDGAP)

The performance Analysis of the proposed Intelligent Human Diseases Pattern Prediction Technique (ECNN-HDPT) in term of Classification Accuracy is shown in the Table. 1.

Table 1: Classification Accuracy for 10 Data Sets - Analysis

Parameter: Accuracy:

Data Sets	G-HRF	IHDGAP	ECNN-HDPT
1	0.85	0.91	0.97
2	0.84	0.88	0.95
3	0.85	0.90	0.95
4	0.85	0.89	0.97
5	0.85	0.91	0.98
6	0.84	0.89	0.98
7	0.86	0.92	0.97
8	0.86	0.90	0.96
9	0.84	0.89	0.95
10	0.85	0.90	0.98

VII. CONCLUSION

This Research Work is proposed an enhanced Classifier called Deep Learning based Human Diseases Pattern Prediction Technique (ECNN-HDPT). This is the Hybrid System designed with the help of SVM and CNN. The proposed model is implemented and analyzed carefully in terms of Processing Time, Memory Usage/Utilization, Accuracy, Sensitivity, Specificity and FScore. From the experimental results, it is noticed that the proposed deep learning mechanism improves the performances of the proposed classifier in terms of Accuracy, Sensitivity, Specificity, and FScore as compared with our previous model Gene Signature based Hierarchical Random Forest (G-HRF) and Intelligent Human Diseases-Gene Association Prediction Technique (IHDGAP). However, it was noticed that the proposed model consumes relatively more Memory and Processing Time as we employs SVM-ECNN to train and test Patterns. The proposed model capable of classifying and predicting the disease patterns more accurately as we use deep learning method called SVM-Convolution Neural Network algorithm and find out more Positive Gene Sequences and Negative Gene Sequences through the help of Association Mining and Confidence Divider.

REFERENCES

[1]. N. K. Sakthivel, N.P. Gopalan and S. Subasree, "G-HR : Gene Signture based HRF cluster for Predicting Human Disease", International Journal of Pure and Applied Mathematics(IJPAM), 2018.

[2]. N. K. Sakthivel N. P. Gopalan and S. Subasree, "A Comparative Study and Analysis of DNA Sequence Classifiers for Predicting Human Diseases", ACM International Conference on Informatics and Analytics, ICIA-16, August, 2016.

[3]. Shuang Cheng, Maozu Guo et.al., "MiRTDL: a deep learning approach for miRNA target prediction", IEEE/ACM Transactions on Computational Biology and Bioinformatics, November, 2015.

[4]. Lixin et.al., Chen, "DPFMDA: Distributed and privatized framework for miRNA-Disease association prediction", Pattern Recognition Letters, ScienceDirect, Elsevier, December, 2017.

[5]. Xing Chen et.al., "HAMDA: Hybrid Approach for MiRNA-Disease Association prediction", Journal of Biomedical Informatics, ScienceDirect, Elsevier, October, 2017

[6]. Thiptanawat Phongwattana, Worrawat Engchuan and Jonathan H. Chan, "Clustering-Based Multi-Class Classification of Complex Disease," 7th IEEE International Conference on Knowledge and Smart Technology (KST2015) Pp. 25-29, Chon Buri, Thailand, 2015.

[7]. Koosha Tahmasebipour and Sheridan Houghten, "Disease-Gene Association Using a Genetic Algorithm. 14th IEEE Computer Society conference on Bioinformatics and Bioengineering", Pp. 191-197, 2014.

[8]. Gregorio Alanis-Lobato, "Exploring the Genetics Underlying Autoimmune Diseases with Network Analysis and Link Prediction," Middle East Conference on Biomedical Engineering (MECBME), 2014.

[9]. Wei Hu, "High Accuracy Gene Signature for Chemosensitivity Prediction in Breast Cancer," Tsinghua Science And Technology, 530-536, Volume 20, Number 5, October 2015.

[10]. Conze, and et. al. "Random Forests on Hierarchical Multi-Scale Supervoxels for Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans," IEEE 13th International Symposium on Biomedical Imaging (ISBI), April 2016.

[11]. Desbordes Paul amd et. al., "Feature selection for outcome prediction in esophageal cancer using genetic algorithm and random forest classifier," Computerized Medical Imaging and Graphics, 2016.

[12]. M. Aishalaifa et.al., " Using context specific effect of miRNAs to identify functional associations between miRNAs and gene signatures, Bioinform, 2013.

[13]. X.Chen, CC et.al., "Wbsmda: Within and between score for mima-disease association prediction", Scientific Reports, 2016.

[14]. Li JQ et.al., "MCMDA: Matrix Completion for MiRNA Disease Association prediction, Oncotarget, 2017.

[15]. Chen X, et.al., "Long non-coding RNAs and complex diseases from experimental results to computational models", Briefings in Bioinformatics, 2016.

[16]. Xiangxiang Zeng, Member, IEEE, Yuanlu Liao, Yuansheng Liu, and Quan Zou_, Member, IEEE,

"Prediction and validation of disease genes using HeteSim Scores," IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol 14, 2017.

[17]. Z. Zahari, L. K. Teh, R. Ismail, and S. M. Razali, "Influence of DRD2 polymorphisms on the clinical outcomes of patients with schizophrenia," Psychiatric genetics, vol. 21, no. 4, pp. 183–189, 2011.

[18]. R. Kukreti, S. Tripathi, P. Bhatnagar, S. Gupta, C. Chauhan, S. Kubendran, Y. J. Reddy, S. Jain, and S. K. Brahmachari, "Association of DRD2 gene variant with schizophrenia," Neuroscience letters, vol. 392, no. 1, pp. 68–71, 2006.

[19]. H. Fan, F. Zhang, Y. Xu, X. Huang, G. Sun, Y. Song, H. Long, and P. Liu, "An association study of DRD2 gene polymorphisms with schizophrenia in a Chinese Han population," Neuroscience letters, vol. 477, no. 2, pp. 53–56, 2010.

[20]. H.-S. Hoe and G. William Rebeck, "Functional interactions of APP with the apoE receptor family," Journal of neurochemistry, vol. 106, no. 6, pp. 2263–2271, 2008.

[21]. C. Shi, X. Kong, Y. Huang, S. Y. Philip, and B. Wu, "HeteSim: A general framework for relevance measure in heterogeneous networks," IEEE Transactions on Knowledge & Data Engineering, no. 10, pp. 2479–2492, 2014.

[22]. Yuebing Zhang, Zhifei Zhang, Duoqian Miao, Jiaqi Wang, "Three-way Enhanced Convolutional Neural Networks for Sentence-level Sentiment Classification," Journal of Information Sciences, Elsevier, 2018

[23]. Z. P. Fan, Y. J. Che, and Z. Y. Chen. Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis. Journal of Business Research, 74:90–100, 2017.

[24]. M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso. Cross-domain polarity classification using a knowledge-enhanced metaclassifier. Knowledge-Based Systems, 86:46–56, 2015.

[25]. H. Fujita, T. R. Li, and Y. Y. Yao. Advances in three-way decisions and granular computing. Knowledge-Based Systems, 91:1–3, 2016.

[26]. M. S. Hajmohammadi, R. Ibrahim, A. Selamat, and H. Fujita. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. Information Sciences, 317(C):67–77, 2015.

[27]. Y. Liu, J. W. Bi, and Z. P. Fan. A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm. Information Sciences, 394-395:38–52, 2017