To Improve the Result of Automatic Detection of Diabetic Retinopathy

Monika[1], Rohit Mahajan[2]
[1]*Dept. of CSE, GCET, Gurdaspur, PTU Kapurthala, Punjab, India*
[2]*Dept. of CSE, GCET, Gurdaspur, Punjab, India*

Abstract- **Diabetic retinopathy detection at early stage is critical and ignorance could lead to loss of vision also. Diabetic retinopathy detection at early stage is the prime objective of this study. To detect the diabtetic retinopathy at early stage, deep learning mechanism is collaborated with the Gaussian filtering and multi support vector machine. Multi Support vector machine is used at classification phase. Simulation is conducted within MATLAB 2017 using image processing toolbox. Operation is performed in coloured fundus images. To tackle the problems of large image set, deep learning mechanism is merged along with genetic approach. Layers are defined including input, processing and output layers along with filtering layer. Once output is generated using this network, it is fed into the CNN for segmentation. Genetic algorithm with MSVM is used at classification section to improve the accuracy of overall result. Classification results shows improvement by 2 % in terms of specificity, accuracy and error rate.**

Index Terms- Diabetic retinopathy, CNN, MSVM, Gaussian Filtering, fundus images.

INTRODUCTION

Diabetic retinopathy is the major cause of blindness within the human beings nowadays. To tackle the issues of diabetic retinopathy image processing techniques are being used. Image processing techniques used for analysis of such diseases includes machine learning and deep learning mechanism. [2]Machine learning mechanisms are categorised into supervised and unsupervised learning mechanisms. Supervised learning mechanisms are those in which guidance is required at every stage of segmentation and classification. Hence are static in nature and may not be scalable. Unsupervised learning mechanism on the other hand includes mechanisms that lead to the segmentation and classification which are scalable in nature. This means image can be used for classification which may not be a part of training set. Parallel execution of complex images required multi-heuristic approach for accurate identification of disease.[3] Early detection of critical and life threatening diabetic retinopathy can be tackled using the proposed [4]Genetic approach along with multi support vector machine. In addition deep learning mechanism is used for processing large datasets. Convolution neural network mechanism is considered for evaluation in this literature. Techniques defined for determining diabetic retinopathy is researched over and most appropriate of them are discussed in the literature survey.

Rest of the paper is organised as under: section 2 gives the literature survey of existing techniques for image mining, section 3 gives the proposed system, section 4 gives the performance analysis and results, section 5 gives the conclusion and future scope, section 6 gives the references.

LITERATURE SURVEY

Data mining approaches is the base of this literature. Analysis of existing literature provide base for proposed literature. [5] Reviewed various models and methods used within data mining. Data mining techniques development from 2005 to 2015 is reviewed and application in regards to health care is proposed. [6] Suggests the integration of medical data with data mining strategies used to form medical information system. Patient medical condition can be analyzed along with future prediction about patient's health. Hidden possibilities can be extracted using unlimited data mining techniques to make accurate health forecast. [7] Proposed multilayer perceptron in order to analyze big data corresponding to health care. As literature deals with health care of patients hence high degree of accuracy is desired. To accomplish the desired goal comparison of SVM and

multilayer perceptron on health care data set is made. Results of SVM in terms of classification are better as compared to multilayer perceptron. [8] Suggests data mining techniques used for analysis of diabetics. Support Vector Machine (SVM) is used for this purpose. Genetic approach is also analyzed for diabetic's dataset in the field of data mining. Results of SVM are obtained to be better. [9] Suggests five J.48 classifiers to predict hypertension and eight other diseases. Prediction accuracy is obtained and compared against naïve bayes approach. Results in terms of J.48 are obtained to be better. [10] Suggests hybrid approach for health care to predict diseases using Big data. Pruning based KNN is used for this purpose which used density based clustering based method integrated with KNN approach. Local outlier factor of PB-KNN is better as compared to KNN. [11] proposes SVM and neural network techniques for skin lesion detection in human body. Segmentation along with classification is performed in order to detect the diseases. [12]predict heart diseases are primary cause of death among humans in last decade. Data mining techniques are used in order to detect and predict heart diseases efficiently. [13] proposes a mechanism through which information about patient coming for checkup at hospital is stored and algorithm is applied in order to perform predictions. Data mining algorithm considered in this approach is naïve bayes. Accuracy of prediction is obtained is significant in this case.[14] suggests intelligent heart disease prediction system. Decision tree, naïve bayes and neural network technique are used for accurate analysis and prediction of disease.

Study of literature suggests that large image sets are not tackled through the existing literature. Also no optimization mechanism like genetic approach is followed. Next section gives the flow of proposed system.

PROPOSED SYSTEM

The image set is required to be operated upon by the modified support vector machine with classification. Feature extraction is performed by the use of Genetic algorithm. Rules are implemented within Matlab neural network toolbox. The detailed description of proposed system is listed as under

DATASET DESCRIPTION

The dataset which is used is derived from internet. DR dataset is used for extraction. DR dataset contains 321 images of 40 to 44 KB in size. The images are gray scale in nature. Dimension of images is 256x256.

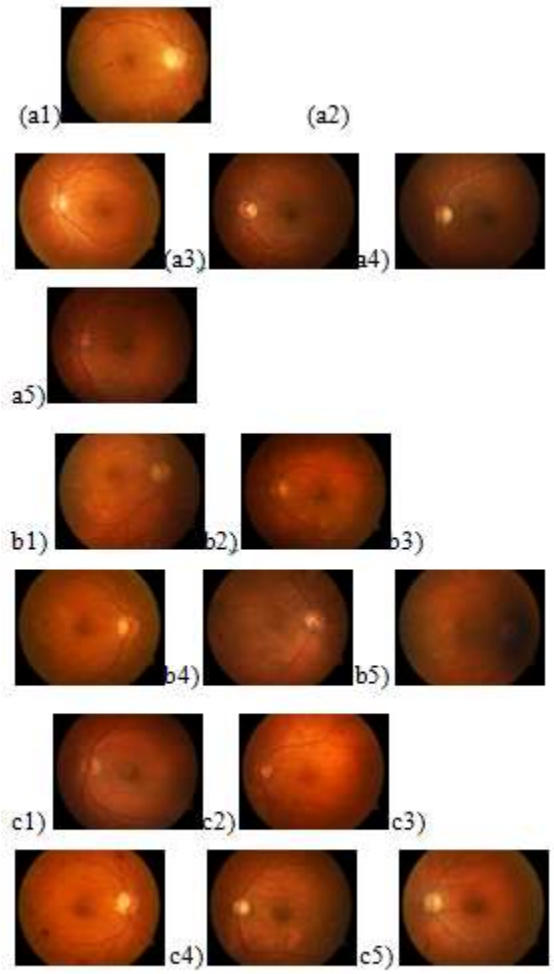


Fig. 1.a) mild non-retinopathy images: (b) moderate non-retinopathy images, (c) Severe non-proliferative retinopathy images.

DR dataset contains group of images with and without cancerous cell. Through the proposed system classification is generated to detect the disease and compared against the original image set.

APPLYING CNN ALGORITHM FOR SEGMENTATION

Convolution neural network is a class of deep learning mechanism with feed forward network. In this approach number of nodes comparable with the Diabetic retinopathy dataset tuples is taken. These

nodes construct a network. Number of nodes varies with the iterations of optimization approach used. The optimization approach collaborated with the proposed system genetic in nature. Every nodes have a weight assigned to it. These weights are adjusted with feed forward propagation mechanism. The process continues until generation expires or prescribed tolerance is achieved.

The result obtained through layers of CNN is fed to the classification stage. After performing segmentation, result is as follows

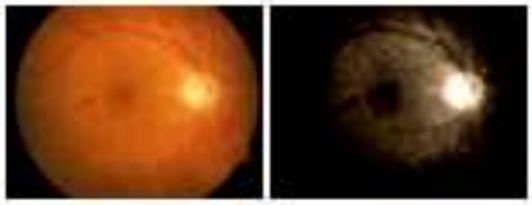


Figure 2: a) Original Image b) Segmented Image

The CNN algorithm is followed by genetic algorithm for feature selection and classification. Decision tree produce the classification results the objective function associated with each Testing and training set image is in the form of weights.

GENETIC ALGORITHM And MSVM for classification

Genetic algorithm is multiheuristic algorithm having multiple objectives associated with it. Genetic algorithm is associated with different phases. The pixels correspond to chromosomes. To perform feature extraction, selection operation takes place. The proposed system uses random selection operation. The fitness function evaluation is used to generate next population for feature extraction. The extracted features are compared against the threshold value. The threshold value is assumed to be base value above which optimality is achieved.

$$Threshold = Optimal_{base-value}$$

The fitness function evaluation takes place in order to obtain optimal results for classification.

$$F(t) = Pixel_i(Features_j > Threshold)$$

The mutation and crossover is performed only if threshold value is invalidated. Mutation and crossover is accomplished by identifying pixels having intensity values lower than desired levels. After which selection operation is performed again.

Decision tree is used for classification purpose. Classification of results required certain rules to be created. The Training rules correspond to features which are identified through genetic algorithm. The disease is detected if rules are violated. Decision tree rules are listed as under

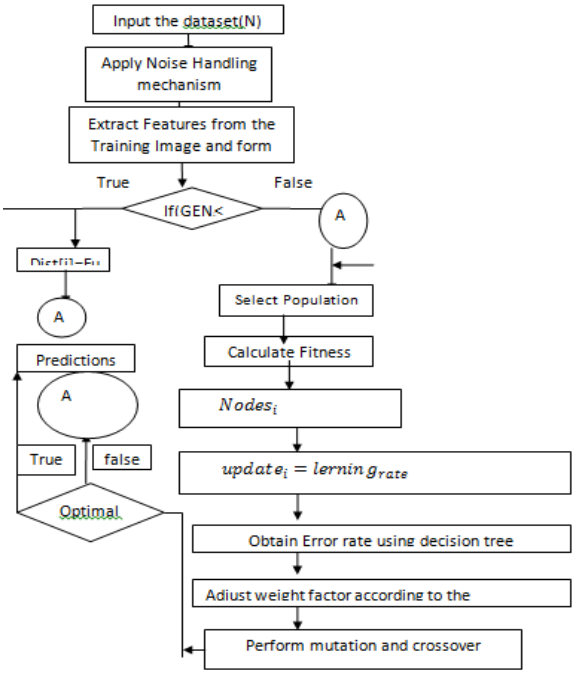
Classification rules.

Maximum value 1 and minimum value 0

If GM>0.5
If STD>0.5
If Kurtosis>0.5
If Moment>0.5
If Mean>0.5
Disease detected
Else
Disease not detected

A feature extracted from the segmented image is compared against these rules. Membership is decided only if feature lies within the range of 0 and 1. In order to detect the disease membership should be greater than 0.5.

These Memberships are obtained for specific parameters. The process is repeated until optimize values of the weights are obtained. Once obtained, correct class corresponding to the disease is predicted.



RESULT AND PERFORMANCE ANALYSIS

The performance of the system is analyzed by the use of parameters such as accuracy, specificity and sensitivity.

Accuracy is obtained by subtracting the actual result from the approximate result. In terms of predictions accuracy is obtained as

$$Accuracy = \frac{Correct_{Pre}}{Total_{Pred}}$$

Equation 5: Accuracy in terms of prediction

Sensitivity is obtained by dividing number of positive predictions to the total true positive rate.

$$\text{Sensitivity} = \frac{Correct_{Positive_{predictions}}}{Total_{Positives}}$$

Equation 6: Sensitivity evaluation formula

Specificity is another parameter used to evaluate correctness of the proposed system. It is given as under

$$\text{Specificity} = \frac{True_{Negitives}}{TP+FN}$$

Equation 7: Specificity obtaining formula

The disease detection and prediction is given though accurate classification, result in terms of plots is given as under

For level 1 DR image set accuracy is given as under

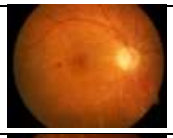
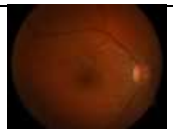
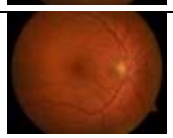
Imageset	Accuracy with Deep Learning and decision tree classifiers(%)	Accuracy with Genetic(%)
	97	99
	96	98
	97	99
	95.5	98
	96	98

Table 1: Predicted accuracy corresponding to (Mild) non-proliferative retinopathy images (level 1)

For level 2 retinopathy image set accuracy is given as under

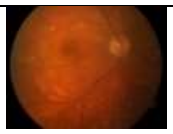
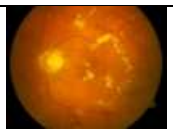
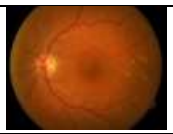
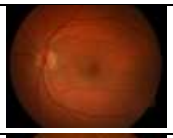
Imageset	Accuracy with Deep Learning and decision tree classifiers(%)	Accuracy with Genetic(%)
	95.5	98
	96	99
	97	98
	96	98
	96	98

Table 2: classification accuracy for (Moderate) non-proliferative diabetic retinopathy images

For level 3 image set accuracy is given as under

Imageset	Accuracy with Deep Learning and decision tree classifiers(%)	Accuracy with Genetic(%)
	96	98
	97	98
	96.5	99

	97	99
	97	99

Table 3: prediction accuracy of image set (Severe) non-proliferative diabetic retinopathy images.

Result comparison in terms of accuracy, sensitivity and specificity are given as under

Image set name	Parameters	Existing (%)	Proposed(%)
Level 1 DR(Mild)	Accuracy	96	98
	Specificity	75	79
	Sensitivity	83	86
Level 2 DR(Moderate)	Accuracy	96.5	98
	Specificity	79	80
	Sensitivity	78	87
Level 3 DR(Severe)	Accuracy	97	99
	Specificity	79	76
	Sensitivity	78	86

Classification accuracy of proposed system appears to be more as compared to existing techniques. Multiple class prediction mechanism showing higher accuracy proving the worth of study.

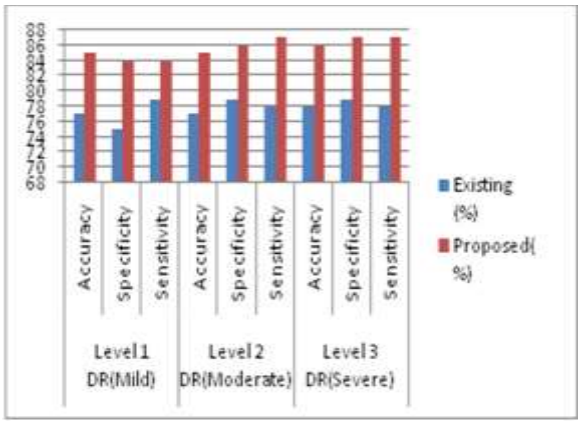


Figure 4: Confusion matrix

Results and performance analysis as indicated through the plot shows that deep learning combined with multi support vector machine yield better result.

Conclusion and Future scope

In this paper an automated system that utilizes multi support vector machine with genetic algorithm for diabetic retinopathy is proposed. Pre-processing phase is critical and is well defined using noise handling and resizing operation. Obtained images are fed into the trained network for feature extraction using genetic algorithm and classification is performed using decision tree. Hybrid approach followed gives better results.

The main objective of the proposed literature is create optimised network using strategy of genetic algorithm for better classification accuracy. Higher classification accuracy is achieved by the use of said literature. In future, proposed strategy can be examined against the reak time datasets for better evaluation of classification accuracy.

REFERENCES

[1] L. Zhou, Y. Zhao, J. Yang, Q. Yu, and X. Xu, "Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images," IET Image Process., vol. 12, no. 4, pp. 563–571, 2018.

[2] S. Shetty, K. B. Kari, and J. A. Rathod, "Detection of Diabetic Retinopathy Using Support Vector Machine (SVM)," vol. 23, no. 6, pp. 207–211, 2016.

[3] S. S. Rubini and A. Kunthavai, "Diabetic retinopathy detection based on eigenvalues of the hessian matrix," Procedia Comput. Sci., vol. 47, no. C, pp. 311–318, 2014.

[4] V. Singh, "Detection of unhealthy region of plant leaves using Image Processing and Genetic Algorithm," pp. 1028–1032, 2015.

[5] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," Procedia Comput. Sci., vol. 72, pp. 306–313, 2015.

[6] I. Ţăranu, "Data mining in healthcare: decision making and precision," Database Syst. J., vol. 5, no. 4, pp. 33–40, 2015.

[7] P. Naraei, V. Street, V. Street, and V. Street, "Application of Multilayer Perceptron Neural Networks and Support Vector Machines in Classification of Healthcare Data," no. December, pp. 848–852, 2016.

[8] "1-s2.0-S2001037016300733-main(1).pdf." .

[9] F. Huang, S. Wang, and C. Chan, "Predicting Disease By Using Data Mining Based on Healthcare Information System," 2012 IEEE Int. Conf. Granul. Comput. Predict., pp. 12–15, 2012.

[10] K. Yan, X. You, X. Ji, G. Yin, and F. Yang, "A Hybrid Outlier Detection Method for Health Care Big Data," 2016 IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Soc. Comput. Netw. (SocialCom), Sustain. Comput. Commun., pp. 157–162, 2016.

[11] M. A. Farooq, M. A. M. Azhar, and R. H. Raza, "Automatic Lesion Detection System (ALDS) for Skin Cancer Classification Using SVM and Neural Classifiers," 2016 IEEE 16th Int. Conf. Bioinforma. Bioeng., pp. 301–308, 2016.

[12] S. Sivagowry, M. Durairaj, and a. Persia, "An empirical study on applying data mining techniques for the analysis and prediction of heart disease," 2013 Int. Conf. Inf. Commun. Embed. Syst., pp. 265–270, 2013.

[13] C. Anusha, S. K. Vinay, H. J. Pooja Raj, and S. Ranganatha, "Medical data mining and analysis for heart disease dataset using classification techniques," Natl. Conf. Challenges Res. Technol. Coming Decad. (CRT 2013), pp. 1.09–1.09, 2013.

[14] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," 2008 IEEE/ACS Int. Conf. Comput. Syst. Appl., pp. 108–115, 2008.