International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol. 9, Issue.3, March 2021 E-ISSN: 2347-2693

MSVM Based Technique Used To Detect Diabetic Retinopathy at Early Stage

Kaveri Devi[1*], Arshdeep Kaur[2]

[1,2]SSGI Dinanagar, India

Corresponding Author: kaverid1313@gmail.com

DOI: https://doi.org/10.26438/ijcse/v9i3.3440 | Available online at: www.ijcseonline.org

Received: 05/Mar/2020, Accepted: 10/Mar/2021, Published: 31/Mar/2021

Abstract- Diabetic retinopathy causes the life of eye decay considerably. There are stages associated with the DR. Early detection of DR could lead to the adverse affect of DR to be minimised. Techniques have been devised to tackle and identify the problems of DR at early stage. This paper presents the comprehensive review of techniques such as machine learning and deep learning, used for the purpose of detection of DR and also performs the comparative analysis of parameters used for the same. The proposed algorithm uses MSVM algorithm that discovers more patterns to detect disease accurately. The results will help in predicting quicker and more accurate disease so that it lead timely treatment of the patients.

Keywords-Diabetic retinopathy, machine learning, deep learning, datasets.

I. INTRODUCTION

Diabetic retinopathy (DR) is a chronic disease related with the eye retina which presently comprises of one of the most common causes of blindness and loss of vision. [1]The incidental statistics indicate that DR is the primary cause of blindness in people of working age of the present era. [2]DR is an outcome of diabetes-mellitus, illness which elevates the concentration of glucose in blood. This unusually high glucose levels damage the eye vessel endothelium infuriating set of damages related to the illness. Although having diabetes does not necessarily entail vision mutilation, about 2% of the patients affected by this disease are blind and 10% undergo vision deprivation after 15 years of diabetes as a result of DR complications. [3]Vision-threatening retinopathy is rare in type 1 diabetic patients in the first 3–5 years of diabetes or before puberty. During the next two decades, nearly all type 1 diabetic patients develop retinopathy. Up to 21% of patients with type 2 diabetes have retinopathy at the time of first diagnosis of diabetes, and most develop some degree of retinopathy over time. The estimated prevalence of diabetes for all age groups worldwide was 2.8% in 2000 and will be 4.4% in 2030.

[4]Despite DR being an incurable disease, if the illness is detected and treated in its early stages visual impairment can be avoided in 98% of cases. In this respect, though laser photocoagulation has established to be a successful treatment for preventing major loss of vision produced by DR yet the early detection of the illness is still a difficult task since people affected by it do not recognize symptoms until visual loss develops which usually happens in the later disease stages, when treatment loses its effectiveness. This is why the prone diabetic population has to be examined periodically by public or private health systems in search of early DR signs.[5] However, this preventive action involves a daring confrontation by the health systems due to the high number of ophthalmologists and material resources needed to attend so many patients requiring ophthalmologic revision.

General step in the analysis of image artifacts are listed as under

a. Pre-processing

[6]This phase is used in order to filter the noisy part of the image and enhance the image set presented for examination. Noise handling mechanism that could be incorporated includes median filtering, Gaussian smoothening, adaptive median filter etc. after the pre-processing phase, feature extraction mechanism is performed.

b. Feature extraction

[7]This mechanism is utilized as a part of request to extricate the basic features out of the picture. These features are utilized to separate valuable data about the disease introduce inside the image. Feature extracted could include Mean, Median, Mode, Kurtosis, Std Deviation, mean deviation etc.

c. Segmentation

[8]Features are extracted from enhancement image are examined. Basic segment of the image is extricated and pointless parts are dispensed with from the image. The basic parts are represented with white area and superfluous parts are represented with dark segment. After the image is sectioned the features are removed once more. These features are compared against the training set features in order to identify the lesion if any.

d. Classification

[9]The features(Mean, Median, Mode, Kurtosis, Std Deviation,mean deviation etc.) values so extracted are compared against the disease characteristics. These characteristics are examined for fitting in classes of disease. If disease falls into the category of any disease then disease is predicted. For classification, algorithms like K-means, Decision Tree, SVM etc. can be used.

In order to tackle the large datasets special branch of machine learning known as Deep learning can be used. The critical characteristic of deep learning is the handling of large dataset. Problem with deep learning is the handling of small dataset. However size of dataset presented for evaluation could be large enough to be tackled through machine learning. So this field of machine learning provides useful mechanism for analyzing Diabetic Retinopathy.

Next section presents the literature survey of techniques used to detect the diabetic retinopathy and extract useful mechanism for the detection process.

II. LITERATURE SURVEY

Diabetic retinopathy is a serious and broadly spread eye disease. It is the commonest reason of legal blindness in the working-age population of created nations[10]. Diabetic retinopathy happens when diabetes harms the blood vessels inside the retina, leaking blood and fluids into the tissue. This liquid fluid produces microaneurysms, hemorrhages, hard exudates, and cotton wool spots (a.k.a., soft exudates)[11]. Diabetic retinopathy is a noiseless disease and may just be perceived by patients when changes in the retina have advanced to a level where treatment becomes difficult and even impossible.

The expanding number of diabetic retinopathy cases overall requires to strengthen the creation of instruments to determine diabetic retinopathy. Programmed recognition of diabetic retinopathy will save time and efforts. In this manner, S.rubhini et al[12]. proposed a technique for programmed discovery of microaneurysms in retinal fundus pictures. Maher et al [2] already assessed a choice based emotionally supportive network for programmed screening of non-proliferative diabetic retinopathy. Truth be told, support vector machines were utilized by Maher et al. [4] in the computerized analysis of non-proliferative diabetic retinopathy. A few picture pre-preparing strategies have additionally been proposed keeping in mind the end goal to distinguish diabetic retinopathy in [7], [13], [14]. However, regardless of all these past works, mechanized discovery of diabetic retinopathy still remains a field for development [1].

There are different Machine Learning algorithms- such as Decision Tree algorithm, Naive Bayes Classifier, Neural Network show, k-Nearest Neighbor algorithm, Support Vector Machine, K-implies, Bisecting K-implies calculation, Association Rule mining calculations and so on. Among them some are utilized for classification, some to cluster purposes and some others for finding effectively interpretable guidelines for taking appropriate choice. Choice Tree calculation is an extremely well known calculation for classifications. It for the most part uses Information Gain (IG) as the model for part on a quality[15]. The quality with the most astounding IG is picked as the part characteristic at each level. Be that as it may, Information Gain has a few weaknesses like it favors the property which has substantial number of unmistakable esteems. So if there is a quality in the dataset like product ID, at that point Information Gain approach will like to part on item ID as in light of the fact that this trait can remarkably recognize each tuple in the set and this would bring about a substantial number of parcels (the same number of as there are qualities), everyone will have only one tuple. Since there will be no records with various class marks in each segment, the obliged data to arrange data set D in view of this apportioning in light of Information Gain rule would be Info product ID(D)=0. In this way, the data picked up by apportioning on this quality is most elevated. Thusly, such a dividing is futile for classification. So it won't sum up the model[16]. Along these lines, IG based approach is not successful for a wide range of datasets. For example, if a dataset has diverse characteristics with various quantities of particular esteems, it inclines toward the traits with more quantities of unmistakable esteems as part qualities however some different properties with less number of unmistakable esteems might be more significant for classification. Because of this purpose behind some dataset IG based approach does not give satisfactory exactness. To defeat this we have proposed an approach which utilizes the idea of Correlation Ratio or CR as the part foundation. This strategy has no such biasness. It considers that characteristic for classification which is significant enough to distinguish no less than one result class[17]. The general CR technique is reasonable for quantative data. Be that as it may, our proposed CR based approach is appropriate to ostensible or absolute characteristics. In this paper we have dealt with 1) Proposing one element choice strategy in light of Correlation Ratio approach which is reasonable for ostensible traits. 2) Building Classification demonstrate utilizing Decision Tree which will utilize our proposed include choice approach. 3) Comparing our proposed approach with Information Gain based Decision Tree and Gain Ratio based Decision Tree. We have seen from the outcome sets that our strategy performed better for a few sorts of infections than Information Gain based and Gain Ratio based models.There are many advantages to evaluating and measuring results identified with Health care, including enhanced care conveyance and cost decrease. This has prompted a gigantic move towards using a data-driven approach in the medicinal services field. This is happening at different sizes of determination running from the person to the national level[18]. At the individual level, there is an expansion of individual observing gadgets for action, rest, and different factors that effect Health care. Such checking is required to prompt better patient engagement and enhanced ailment observing. At the national level, there is an expanding pattern towards Open Data and straightforwardness from government foundations around the globe. The US Federal Government keeps on discharging data from the Center for Medicare and Medicaid Services (CMS). Pharmaceutical organizations are likewise discharging data to general society, for example through the Yale University YODA extend. Itaudit a significant number of the advantages given by this open data development for medicinal services . A noteworthy test in separating an incentive from the accessible data is that there are numerous datasets. In spite of the fact that early adopters might have the capacity to utilize individual Health care observing gadgets, it is difficult for them to comprehend or get to the data assembled at a more extensive level. For example, they might need to know which nearby healing facilities are the best to treat their condition. In spite of the fact that this data could be accessible in national registries, it may not be available and adjustable at the individual level. Consequently, there is a hole between the person's data and interests and worldwide scale data. This requires scaffolds be worked to enable simple access to data at various levels of reflection. A major hindrance is that the data made accessible at the national level, e.g. through the data.cms.gov site in the USA, does not have an easy to use interface[19].

III. DATASET

The image dataset used is DIARETDB0 consisting of 3 categories of 200 eye fundus images. Resizing operation manually as well as automated mechanism is posted upon to fit into the input layer of the network. The images were captured and resized to 77x100 with 3 color channels.

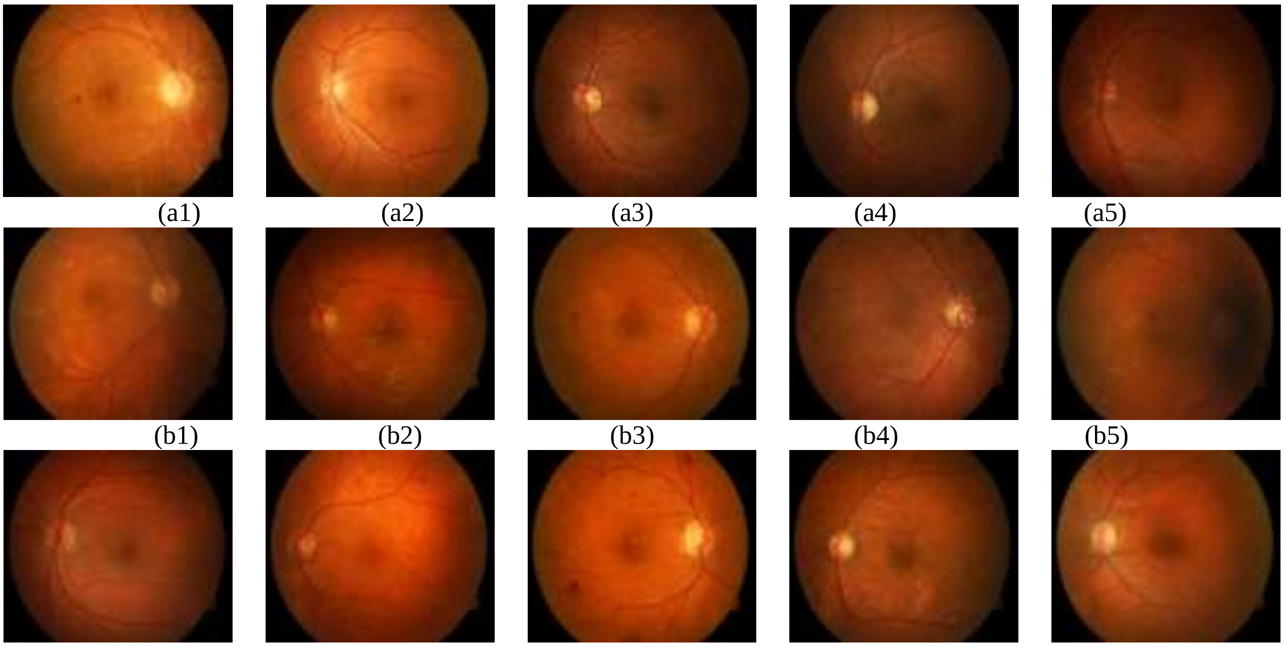


Figure 1: Fig. 1.a) mild non-retinopathy images: (b) moderate non-retinopathy images, (c) Severe non-proliferative retinopathy images.

The 200 pictures are bundled in 3 sets, one for each ophthalmologic division, utilizing the PNG format. In addition, an Excel record with therapeutic conclusions for each picture is given. In this work, we utilize the pictures of only one ophthalmologic division containing 48 pictures with mild, 48 with moderate and 48 with severe DR cases.

IV. PROPOSED SYSTEM

A. Pre-Processing

Preprocessing instrument utilized in this writing contains commotion taking care of alongside resizing task. Commotion dealing with is finished utilizing Gaussian sifting system. This channel is fit for taking care of motivation clamor alongside smoothening activity. In the wake of taking care of commotion, re-estimating activity is finished. Re-measuring is done to display the uniform information to the information layer. Resizing is finished utilizing condition 1.

$$Resized_G = Resize(G_{Smoothene\ d}, [70\ 100])$$

Equation 1: Resized image

This resized image set obtained is passed to the network for further processing.

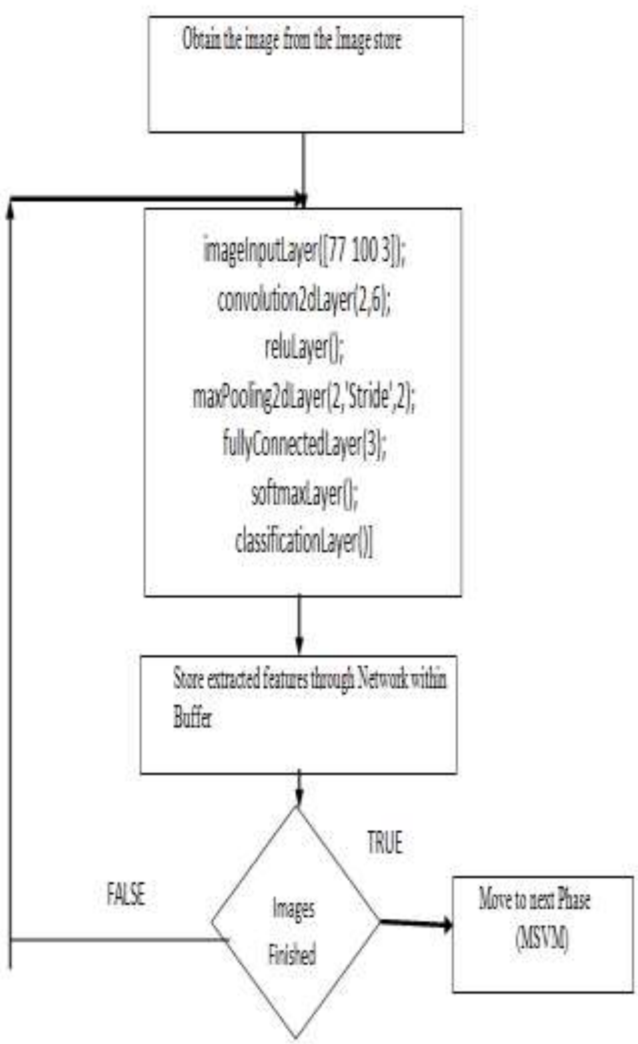


Figure 3.2: Flow of Training Process

To perform grouping MSVM is utilized as the last stage. Bolster vector machine uses guidelines based condition to accurately reach to the arrangement of the given issue including exception discovery or superfluous locales. [12] proposed this framework to determine unclassified locale. MSVM are utilized to understand the characterization results. Ideal hyper planes are characterized to decide if the acquired estimations of enrollment capacities fulfill the hyper plane (D(x)) or not.

Satisfaction Criteria D(X)>1

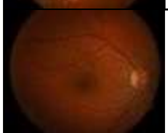
One dimensional membership function $m_{i_j}(x, y)$ is defined for determining optimal separating hyper planes $(x) = 0$ as follows
If values of diagonal are equal (i==j)
The procedure of classification is listed as follows

3. If the pixel value x is such as Di(x)>0 and is satisfied only for that class then it is fed into that class.
4. If $D_i(x) > 0$ and x lies between various classes then classify the data into the class with maximum $D_i(x)$
5. If $D_i(x) <= 0$ and x lies between various classes then classify the data into the class with minimum $D_i(x)$

V. RESULTS AND DISCUSSION

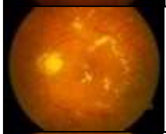
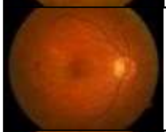
The presentation of the framework is investigated by the utilization of parameters, for example, precision, explicitness and affectability.
The illness location and expectation is given however precise arrangement, result as far as plots is given as under
For level 1 DR picture set exactness is given as under

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

Image set	**Accuracy with Deep Learning and decision tree classifiers (%)**	**Accuracy with MSVM (%)**
	76	82
	78	85
	79	84
	78	83

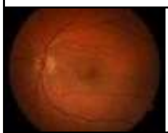
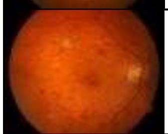
For level 2 retinopathy image set accuracy is given as under

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

Image set	**Accuracy with Deep Learning and decision tree classifiers(%)**	**Accuracy with MSVM(%)**
	78	82
	79	84
	78	82
	77	81
	76	82

For level 3 image set accuracy is given as under

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

Image set	**Accuracy with Deep Learning and decision tree classifiers(%)**	**Accuracy with MSVM(%)**
	78	82
	77	81
	75	80
	76	81
	78	82

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

Table 4: Result comparison in terms of accuracy, sensitivity and specificity

Image setname	**Parameters**	**Existing (%)**	**Proposed(%)**
Level 1 DR(Mild)	Accuracy Specificity Sensitivity	75 73 78	81 80 82

Level 2 DR(Moderate)	Accuracy Specificity Sensitivity	77 79 78	82 82 84
Level 3 DR(Severe)	Accuracy Specificity Sensitivity	78 79 78	83 84 82

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.

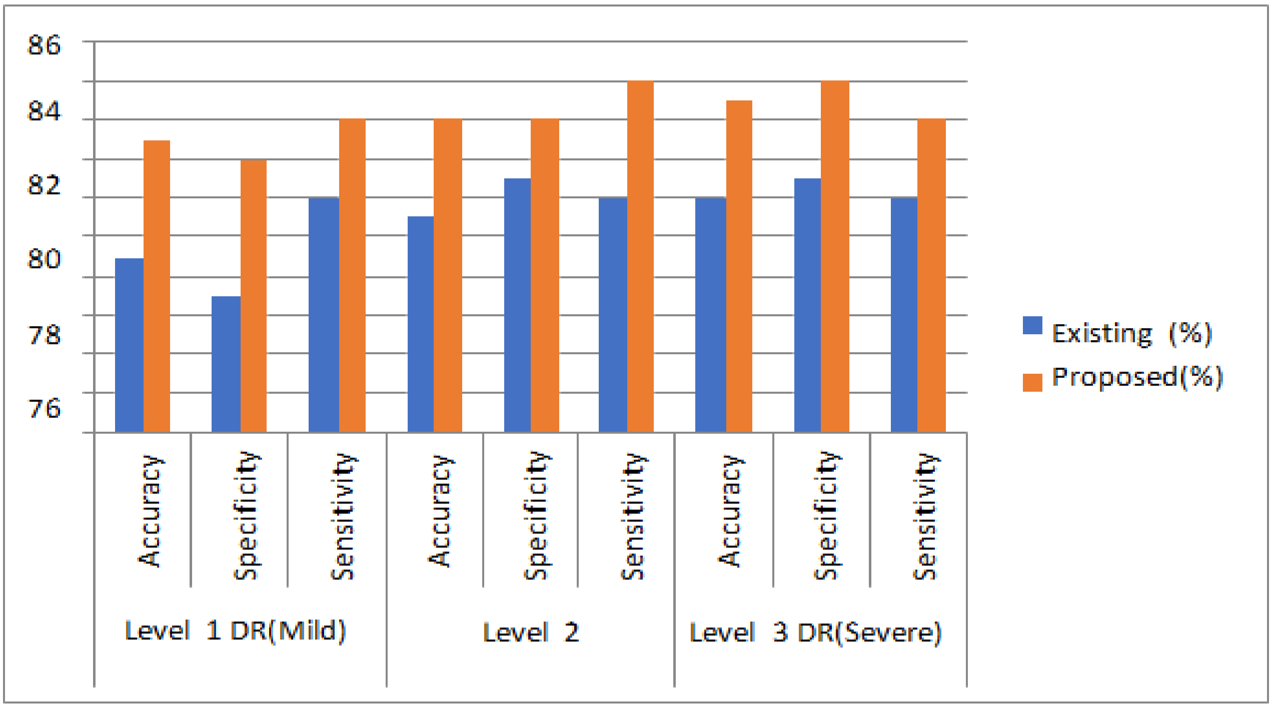


Figure 5.1: Confusion matrix:

Results and execution examination as demonstrated through the plot demonstrates that profound learning joined with multi bolster vector machine yield better outcome.

VI. CONCLUSION

The present review has addressed different deep learning, machine learning techniques that are deployed for DR diagnosis. The techniques and algorithms that were extensively used are convolutional neural networks, artificial neural networks, support vector machines, fuzzy logic, bag of words algorithms and so on. A comparison pertaining to the supremacy of deep learning techniques over existing machine learning techniques is also stated in terms of data requirements, learning, etc which is the need of the hour as the medical database is ever increasing phenomena demanding faster results. Though these studies are vast, this spectrum of research requires more rigorous investigation in terms of classification with minimal errors.

REFERENCES

[1] P. Adarsh and D. Jeyakumari, “Multiclass svm-based automated diagnosis of diabetic retinopathy,” *Int. Conf. Commun. Signal Process.*, **pp. 206–210, 2013.**

[2] R. Maher and M. Dhopeshwarkar, “Automatic detection Non-proliferative Diabetic Retinopathy using image processing techniques,” *J. Eng. Res. Appl.*, **vol. 6, no. 1, pp. 122–127, 2016.**

[3] R. Gargeya and T. Leng, “Automated Identi fi cation of Diabetic Retinopathy Using Deep Learning,” *Ophthalmology*, **pp. 1–8, 2017.**

[4] S. M. Student, C. Landran, and G. Kaur, “Review on: Detection of Diabetic Retinopathy using SVM and MDA,” *Int. J. Comput. Appl.*, **vol. 117, no. 19, pp. 975–8887, 2015.**

[5] I. Ntroduction, “Diabetic Retinopathy Classification using SVM Classifier,” *ACM Comput. Surv.*, **vol. 6, no. 7, pp. 7–11, 2017.**

[6] S. Shetty, K. B. Kari, and J. A. Rathod, “Detection of Diabetic Retinopathy Using Support Vector Machine (SVM),” *IEEE 17th Int. Conf. Parallel Distrib. Syst.*, **vol. 23, no. 6, pp. 207–211, 2016.**

[7] M. J. Paranjpe and P. M. N. Kakatkar, “Automated Diabetic Retinopathy Severity Classification using Support Vector Machine,” *Int. J. Res. Sci. Technol.*, **no. 3, pp. 86–91, 2013.**

[8] V. V. Kumari, N. Suriyaharayananm, and C. T. Saranya, “Feature Extraction for Early Detection of Diabetic Retinopathy,” in *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*, **pp. 359–361, 2010.**

[9] V. Ramya, “SVM Based Detection for Diabetic Retinopathy,” *Iccad*, **vol. V, no. I, pp. 11–13, 2018.**

[10] R. Gargeya and T. Leng, “Automated Identi fi cation of Diabetic Retinopathy Using Deep Learning,” *Ophthalmology*, **pp. 1–8, 2017.**

[11] M. S. Haleem, L. Han, J. Van Hemert, B. Li, and A. Fleming, “Retinal Area Detector from Scanning Laser Ophthalmoscope (SLO) Images for Diagnosing Retinal Diseases,” **vol. 2194, no. MARCH, 2014.**

[12] 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.**

[13] S. M. Student, C. Landran, and G. Kaur, “Review on: Detection of Diabetic Retinopathy using SVM and MDA,” *Int. J. Comput. Appl.*, **vol. 117, no. 19, pp. 975–8887, 2015.**

[14] V. Ramya, “SVM Based Detection for Diabetic Retinopathy,” *IEEE*, **vol. V, no. I, pp. 11–13, 2018.**

[15] A. Mukhlas and A. Ahmad, “Data Mining Technique : Towards Supporting Local Co-operative Society in Customer Profiling , Market Analysis and Prototype Construction,” no. May, **pp. 109–114, 2016.**

[16] R. D. Q. D. Lqglu, D. Lv, D. Frqvlghuhg, I. R. U. Wkh, D. Dqdo, S. Lq, W. K. H. Olwhudwxuh, V. D. U. Pdq, and G. Lq, “A Review of Data Mining and Solar Power Prediction,” **vol. 0, pp. 3–7, 2016.**

[17] M. Bakon, I. Oliveira, D. Perissin, J. J. Sousa, and J. Papco, “A Data Mining Approach for Multivariate Outlier Detection in Postprocessing of Multitemporal InSAR Results,” **vol. 1, pp. 1–8, 2017.**

[18] E. E. Brown, “Improving Privacy Preserving Methods to Enhance Data Mining for Correlation Research,” **pp. 3–6, 2017.**

[19] K. Ahuja, “Data Elimination Based Technique for Mining Frequent Closed Item Set,” **pp. 1–4.**

