2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-6654-5401-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/PDGC56933.2022.10053137

Study on IOT based Smart Disease Detection Model on Machine Learning Techniques for Healthcare Applications

[1]Priyanka Dhaka

Research Scholar Guru Gobind Singh Indraprastha University and Assistant Professor,

Maharaja Surajmal Institute, GGSIPU Janakpuri Delhi 110058

priyankadhaka744@gmail.com

[2]Ruchi Sehrawat

Assistant Professor

University School of Information and Communication Technology, GGSIPU Delhi

ruchi.sehrawat@ipu.ac.in

Abstract- The wearable medical devices are integrated with the internet of things (IoT) for smart healthcare applications to improve the quality of service. The information regarding the patient's details updated through the IoT devices helps to provide the medication based on the present condition and hence the severity can be reduced instantly. Machine learning (ML) is utilized for making better decisions through clustering and classification methods that efficiently deal with the quality of information. Hence, this research introduces an analysis of smart disease prediction techniques for healthcare applications. The analysis of diseases such as heart disease, Alzheimer's disease, and diabetic retinopathy disease based on machine learning is employed by reviewing the conventional detection methods along with the achievements and research gaps. In addition, the analysis based on the performance metrics, dataset, and tools utilized is also devised for the development of the novel smart disease detection technique with more efficient accuracy of detection.

Keywords: detection, machine learning, Alzheimer's disease, diabetic retinopathy, heart disease

I. INTRODUCTION

The development of information technology and embedded systems leads to the development of the Internet of Things (IoT) that provides the interaction among the devices and humans through the virtual environment and physical scenario. In many of the IoT-based applications, the information is gathered through smart cities, healthcare, homes, transportation, and other smart environments using digital devices. In healthcare and medical care applications, the development of sensors and devices based on IoT is an emerging research area. Here, the person-centric environment transformation from the hospital-centric system is essential due to the enhancement of the expensive in the healthcare sector. Thus, using the person-centric healthcare scenario, the healthcare history can be accessed in real-time distantly from anywhere and anytime. The information concerning the patient supports the healthcare team or doctors to diagnose easily based on the parameters included. The parameters utilized for the healthcare monitoring of heart disease are ECG, glucose level, weight, blood pressure, temperature, and heart rate. The main reason behind the occurrence of heart disease is obesity, diabetes, high blood pressure, smoking, and several other factors that affect a healthy life and makes life risk. The neurodegenerative disorder and the chronic disease that affects the function of the brain is termed Alzheimer's disease (AD), which is a chronic disease. The vision loss of the diabetic patient is termed Diabetic retinopathy (DR) [1], which is painful and expensive. Thus, for the disease diagnosis, there is a requirement to an automatic device with minimal cost and a non-invasive manner.

The disease detection based on the machine learning technique is cost-efficient and easy to implement busing the decision support device. The decision supportive method helps the doctors to make a better decision regarding the health condition of the patients by diagnosing the disease in the early stage. Here, the generalization is the detecting of unobserved things from the past training experience and the representation is nothing but the deriving and evaluating the functions using the various occurrences of data.

The major contribution of the study is to analyze the traditional disease detection and prediction techniques based on machine learning approaches. Here, for the smart healthcare applications, the diseases such as diabetic retinopathy (DR) detection, heart disease prediction, and detection of Alzheimer's disease (AD) are categorized and analyzed based on the ML. Then, the analysis based on the tools utilized for implementation of the developed method, dataset utilized for the evaluation, and the performance metrics are analyzed and depicted. Finally, the research gaps are detailed for the researchers to develop a novel machine learning-based disease detection and prediction technique for the proper medication in the early stage to reduce the severity of the disease.

The remaining sections of the study are: Section 2 details the analysis of conventional machine learning techniques for disease detection and prediction. Section 3 elaborates the analysis of the existing methods in terms of dataset utilized, performance metrics, and tools utilized. The research gaps are explained in section 4 and section 5 concludes the work.

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on May 11,2023 at 17:08:24 UTC from IEEE Xplore. Restrictions apply.

II. Motivation

The conventional disease detection methods using the machine learning technique face several challenges such as inaccurate detection, computational complexity, failure to implement in real time applications motivate to analyze the machine learning methods of disease detection along with the research gaps to develop a novel technique by fulfilling the challenges.

2.1 Related Works

The conventional disease detection techniques based on machine learning is elaborated in this section along with the achievements and the research gaps.

2.2 Categorization

The traditional method of disease prediction based on machine learning is categorized into three different categories such as diabetic retinopathy (DR) detection, heart disease prediction, and detection of Alzheimer's disease (AD). The illustration of the disease detection techniques is presented in Figure 1.

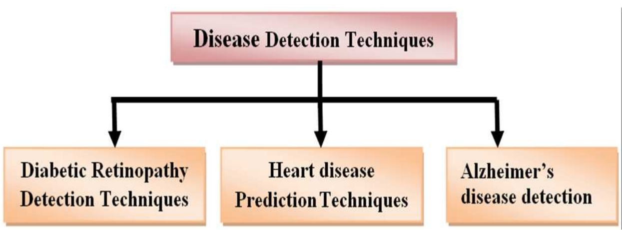

Figure 1: Categorization of different disease detection techniques

2.2.1 Diabetic Retinopathy Detection Techniques

The related works associated with the detection of diabetic retinopathy (DR) are detailed in this section. The machine learning-based DR detection was employed by [2] [3] [4] [5]. The angiography images-based detection of DR was utilized by [2], in which the feature extraction was devised for the attainment of texture information. Then, using the machine learning technique the classification of the disease was employed. The introduced logistic regression regularized with the elastic net penalty (LR-EN) obtained better performance enhancement in terms of specificity and sensitivity for the early diagnosis. KNN based DR detection was used by [3] from the retinal fundus image. The developed method classifies the DR from the extracted informative features obtained from the segmentation. Here, initially, the noise removal and the contrast of the input image were enhanced in the pre-processed stage and the post-processing was also devised to enhance the picture purity. Finally, the DR classification was devised from the extracted features. DR detection through the highlighting of blood vessels and the removing of redundant features for the effective classification was utilized by [4]. Here, the segmentation of the image was employed for the detection of exudates and optical disk elimination. Then, the feature extraction was devised for the collection of the most informative features and finally, the classification was performed using the SVM and RF. The DR detection using an automatic process was utilized by [5], in which the input is pre-processed and then statistical-based features were extracted for the classification. In addition, the PCA was utilized for the reduction of dimensionality. Finally, the classification was performed using logistic regression, SVM, bagged tree, and Fine KNN, in which the maximal accuracy was evaluated using the logistic regression-based detection of GR. The machine learning-based classifier was developed by [6] using the multipath CNN, in which the multipath CNN was utilized for the extraction of the most informative features. The classification based on the decision tree (DT) with javaC4.5 obtained reduced error pruning. The developed method obtained accurate and early detection of the DR from the fundus image for early diagnosis and cure.

A technique named HIMLA, a hybrid inductive machine learning algorithm was utilized by [7] for DR detection through segmentation approach. Initially, the normalization was devised and then segmentation based on the encoding and decoding was employed for the enhancement of the image's quality. Followed by, feature extraction taking place for the classification of the image from the selected features. The developed method obtained accurate and timely detection. The DR detection using the lacunarity and multifractal geometry features was employed by [8], in which the features extraction was devised to obtain the most informative features correlated with the variations in the morphology of the retina. Finally, the SVM was utilized for the detection of the DR. The detection of the hard executed in the retinopathy image was developed by [9] for the detection of the severity of DR. Here, initially, the fundus image was pre-processed for the extraction of the most informative features and then the concatenation of these features was employed and fed as input to the classifier named CNN for the detection of the DR.

2.2.2. Heart disease Prediction Techniques

Heart disease predictions developed by conventional researchers are reviewed in this section. A technique for the prediction of heart disease with minimal false alarming was designed by [10] using the machine learning technique. In this, the feature selection technique was introduced for the effective prediction, in which the attributes were selected initially and the n-gram based feature selection was employed using the weights. Here, the developed feature optimization based strategy of prediction obtained better performance with minimal processing overhead and alarming rate along with maximal labeled prediction accuracy. The image fusion-based machine learning was designed by [11] for heart disease prediction. The precision of the developed method was enhanced through feature fusion, which results in reliable prediction. The ANN provides a better result compared to the other classifiers such as SVM, NB, and KNN. In addition, the pre-processing of the input was employed for the removal of noise from the image. An artificial intelligent base heart disease prediction was developed by [12] using the optimization based strategy. In this, the backpropagation

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on May 11,2023 at 17:08:24 UTC from IEEE Xplore. Restrictions apply.

strategy was utilized for the tuning of the weights of the neural network using the stochastic algorithm. The random forest-based prediction was developed by [13] using the data taken from the kaggle database and evaluated the performance using 80% training data and 20% testing data.

A supervised learning-based heart disease prediction was developed by [14] using the UCI Machine repository data. The developed method obtained enhanced performance through minimal learning time and over-learning. Here, the Bayesian optimization was utilized to train the classifier named SVM for tuning the kernel and the weight function. In addition, the pre-processing of the input was employed for the removal of artifacts and the noise. The heart disease prediction based on anomaly detection was designed by [15] using the clustering-based technique. In this, the anomalies were detected through the clustering approach in which the optimal value was determined by the Silhouette approach and the detected anomalies were removed. Then, five classifiers such as logistic regression, naive Bayes, support vector machine, random forest, and K-nearest neighbor were utilized for the prediction. In addition, imputation-based pre-processing was employed for the missing data handling.

The prediction of heart disease using the optimization based feature selection was developed by [16] using the decision tree-based clustering approach. In this, the entropy-based feature selection, partition, and clustering were made. The target label distribution was employed initially for the partition of the dataset. The reduction of the feature dimension was achieved through the clustering devised by considering the class-based partition. The heart disease prediction using the reinforcement approach was developed by [17] through the time series network. In this, the level of the coronary artery occlusion was predicted from the information of the patient's body comprised of 21 different attributes. The prediction was employed using the neural network and four different classes of coronary heart disease were made by the introduced approach. Here, the simultaneous prediction enhances the accuracy through the multi-task parameter sharing strategy.

A hybrid decision support approach was introduced by [18] for heart disease prediction. In this, the most informative features were extracted, and from which the most significant features were selected using the genetic algorithm. Finally, using the machining learning technique the prediction was devised and obtained best performance using the RF technique. The heart disease prediction using the combined deep and machine learning was developed by [19]. Here, the Lasso algorithm was utilized for the extraction of the most informative features, and the duplicated data was removed from the selected attributes. In this, the deep learning classifier utilizes the flattening layer for the avoidance of overfitting issues. Finally, the outputs obtained from the machine learning and deep learning techniques were compared and the introduced K-Neighbour technique obtained better performance with the application of the pre-processing of the input data.

2.2.3. Alzheimer's disease detection

The traditional methods developed for the detection of Alzheimer's disease are detailed in this section. The early detection of Alzheimer's disease (AD) using machine learning was designed by [20]. Initially, the data cleansing was employed in the pre-processing stage and then the correlation among the features was obtained using the generation of the correlation matrix. Finally, the machine learning-based classification was employed, in which seven classifiers' performances were compared. Here, the logistic regression approach obtained better performance in terms of classification accuracy. AD detection through the EEG-based approach was employed by [21] using machine learning. In this, initially, the sub-bands of the EEG were filtered using the discrete wavelet transform, and then using the Burg approach the PSD was evaluated and from the calculated PSD the inter-hemispheric coherence was estimated. Then, the evaluated features, the classification of disease were performed using the Bagged tree and obtained the reproducible result with elevated training accuracy. The data fusion-based AD classification was used by [20] for the enhanced prediction accuracy using the machine learning approach. In this, three different real datasets with stroke data and vascular dementia data were fused together before classifying the AD to enhance the accuracy of detection. In addition, the methods such as wrapper technique and filter-based approach for the selection of the significant features. Finally, classification was performed using machine learning techniques such as KNN, RF, and NB, in which the NB obtained better performance.\

he classification of the AD based on mild, moderate, and severe categories was utilized by [20] through the deep learning strategy along with the optimization technique. Initially, the low pass filter was used for the removal of noise from the input signal before the feature extraction, in which a discrete Fourier transform was utilized. Finally, the AD classification was employed using the CNN and obtained lower accuracy; hence Adam optimization was utilized to train the classifier, which further enhances the classification accuracy with the reduction of computational complexity and loss. h. In addition, the overfitting issues were avoided using the regularization strategy based on the formulation of a multi-task approach. Finally, the AD was detected from the correlated data and obtained a minimal error rate compared to the baseline technique. A trajectory-based AD prediction was employed by [22] using the ensemble-based machine learning approach. The T1-weighted attributes were utilized for the classification from the pre-processed image. Here, the longitudinal features were tracked and several attributes were extracted to enhance the accuracy of the detection capability.

III. Analysis and Discussion

The analysis based on the dataset, tools utilized, and performance metrics are detailed in this section.

3.1 Analysis using Dataset

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on May 11,2023 at 17:08:24 UTC from IEEE Xplore. Restrictions apply.

The analysis using the dataset is depicted in Figure 2. Here, from the analysis, most of the references utilized the dataset from the UCI machine learning repository, followed by the Messidor [23], Kaggle [24], and ADNI [25] datasets are utilized by the same number of literature. The UCI dataset is utilized by 7 references and the Messidor, Kaggle, and ADNI datasets are used by 3 references each and the other dataset such as CHASE dataset [26], IDRiD [27], OASIS [28], Physionet [29], and real datasets are utilized by some of the references by one each.

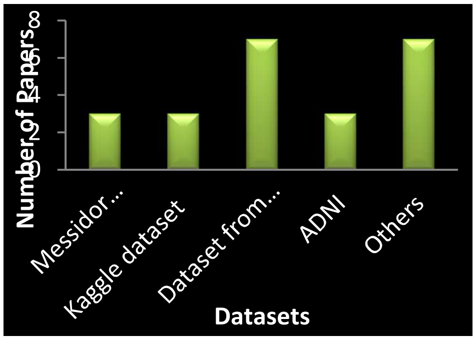

Figure 2: Analysis using the dataset

3.2 Analysis Using Performance Metrics

Accuracy: Based on the input, or training, data, accuracy is the metric used to indicate which model is best at finding relationships and patterns between variables in a dataset.

Sensitivity: Sensitivity is the probability that a test will result in a true positive outcome and is also known as true positive rate.

Specificity: Specificity is the probability that a test will result in a true negative outcome and is also known as true negative rate.

Precision: The number of positive class forecasts that really fall within the positive class is measured by precision.

Recall: Recall measures how many correct class predictions were produced using all of the successful cases in the dataset.

F1 score: The accuracy of a test is measured by the F-score or F-measure. It is derived from the test precision and recall, where precision is the proportion of true positive results to all positive results, including those incorrectly identified as positive, and recall is the proportion of true positive results to all samples that should have been identified as positive.

Figure 3 depicts the analysis based on the performance metrics such as accuracy, sensitivity, specificity, F-score, precision, recall, and other methods. Here, most of the references utilized the accuracy for the evaluation of the disease detection techniques, which is used by 22 references. Besides, specificity is used by 12, sensitivity by 11, precision is used by 10, recall is used by 8, F-measure is used by 7 references, and 6 other metrics such as ROC, False positive rate, error, and Mathew correlation coefficient.

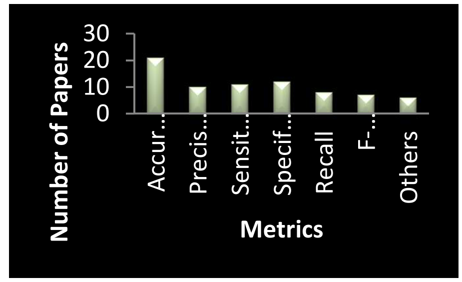

Figure 3: Analysis using performance metrics

3.3 Research Gap

The challenges faced by disease detection are detailed in this section. Several methods were traditional methods for the detection of DR are reviewed and the challenges faced by them are: the method utilized by [8] failed to detect the malignant stage of DR, some techniques failed to train the network [30], failed to utilize the optimization techniques [2] while training the network that enhances the accuracy of detection further. In addition, several other challenges such as inappropriate feature selection make the computation complex and provide inaccurate detection.

The challenges faced by the traditional heart disease prediction techniques are failure to extract the features and feature fusion that reduces the computational complexity [12], non-applicability of the real-life scenario [18] due to the overfitting issues [19], and failure to utilize the optimal feature selection and classifier training to enhance the performance of the system. Besides, failed to use evaluate the performance using other dataset formats except for labeling data [10]. In addition, failure to use optimization to enhance the accuracy failed to develop the generalized method [14] and lower accuracy are some of the major challenges.

The challenges faced by the conventional Alzheimer's Disease methods are failed to enhance the classification accuracy through the feature reduction strategy [21], and generalization was not employed through the classifier and the analysis of result whether the classifier affects the results of the detection by algorithm. The method failed to incorporate the accurate algorithm, failed to classify other diseases related to the EEG signal, and failure to use the deep learning technique, which can increase the accuracy through training and optimization using the fused dataset. In addition, the failure to optimize the feature correlation and failed to consider some of the correlation features to enhance the accuracy.

The diabetic retinopathy is able to perform the automated method, but it needs to be fully developed and its parameters fine-tuned [31]. The specificity and accuracy are slightly lower since exudates are taken into consideration to such a considerable amount [32]. Classifiers for diabetic retinopathy lesions are difficult to use, complex, and frequently inaccurate [33].

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on May 11,2023 at 17:08:24 UTC from IEEE Xplore. Restrictions apply.

IV. CONCLUSION

The disease diagnosis is significant in the early stage to avoid the disease severity and enhance the health conditions. This research proposed a study on disease detection techniques for smart healthcare applications based on machine learning techniques for the diseases such as diabetic retinopathy (DR) detection, heart disease prediction, and detection of Alzheimer's disease (AD). The smart healthcare application by utilizing the IoT helps the patient to track the previous health history and proper medication can be provided in time by accessing the record anytime and anywhere. Several disease detection and prediction techniques are analyzed based on the machine learning technologies along with the dataset used, tools utilized, and performance metrics used. Finally, the research gaps from the traditional methods are analyzed and it helps the research scholars to develop a novel system with more accurate detection and prediction in the near future.

REFERENCES

[1] Laxmi Math and Ruksar Fatima, "Adaptive machine learning classification for diabetic retinopathy", Multimedia Tools and Applications, vol. 80, No. 3, pp: 5173–5186 , 2021.

[2] Zhiping Liu, Chen Wang, Xiaodong Cai, Hong Jiang, and Jianhua Wang, "Discrimination of Diabetic Retinopathy From Optical Coherence Tomography Angiography Images Using Machine Learning Methods", IEEE Access, Vol. 9, pp. 51689 - 51694 , 2021.

[3] Avni Singhal, Vimal Kumar, Niranjan Lal, "Detection of Diabetic Retinopathy using Machine Learning", Nat. Volatiles & Essent. Oils, vol. 8No. 5, pp: 2561 - 2574, 2021.

[4] Anusha Shetty, Charithra Kini T, Deepti Viola Barboza, Ashmitha Kalkur, and Siju V Soman, "Diabetic Retinopathy Disease Detection Using Machine Learning", International Journal of Scientific Research and Engineering Development-Volume 4 Issue 3, June 2021.

[5] Mrs. Pooja Rathi and Dr. S. M. Ghosh, "Automated Diabetic Retinopathy Prediction using Machine Learning Classification Algorithms", Turkish Journal of Computer and Mathematics Education Vol.12, No. 11 , pp. 3020- 3030, 2021.

[6] S. Gayathri, Varun P. Gopi, and P. Palanisamy, "Diabetic retinopathy classification based on multipath CNN and machine learning classifiers", Physical and Engineering Sciences in Medicine, vol. 44, No.3, pp. 639–653, 2021.

[7] Mohamed H. Mahmoud, Salman Alamery, H. Fouad, Amir Altinawi, and Ahmed E. Youssef, "An automatic detection systemof diabetic retinopathy using a hybrid inductive machine learning algorithm", Personal and Ubiquitous Computing, p1-15. 2021.

[8] Mohamed M. Abdelsalam and M. A. Zahran, "A Novel Approach of Diabetic Retinopathy Early Detection Based on Multifractal Geometry Analysis for OCTA Macular Images Using Support Vector Machine", IEEE Access, Vol. 9, pp. 22844 – 22858, 2021.

[9] Akey Sungheetha, "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network", Journal of Trends in Computer Science and Smart technology, Vol.03, No. 02, Pages: 81-94, 2021.

[10] Fuad Ali Mohammed Al-Yarimi, Nabil Mohammed Ali Munassar, Mohammed Hasan Mohammed Bamashmos, and Mohammed Yousef Salem Ali, "Feature optimization by discrete weights for heart disease prediction using supervised learning", Soft Computing, vol. 25, No. 1, PP. 1821–1831 , 2021.

[11] Diwakar, Manoj & Tripathi, Amrendra & Joshi, Kapil & Memoria, Minakshi & Singh, Prabhishek & Kumar, Neeraj, "Latest trends on heart disease prediction using machine learning and image fusion", Materials today: proceedings, vol. 37, No.4, Pages 3213-3218 2020.

[12] Sudarshan Nandy, Mainak Adhikari, Venki Balasubramanian, Varun G. Menon, Xingwang Li, and Muhammad Zakarya, "An intelligent heart disease prediction system based on swarmartificial neural network", Neural Computing and Applications, 2021.

[13] Madhumita Pal, and Smita Parija, "Prediction of Heart Diseases using Random Forest", Journal of Physics: Conference Series, 2021.

[14] Nayak, Gouri & Patro, Sibo & Padhy, and Dr. Neelamadhab, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective", Informatics in Medicine Unlocked, vol. 26, 2021.

[15] Ripon, Rony & Sarker, Iqbal & Hossain, Syed & Anwar, Md & Nowrozy, Raza & Hoque, Moshiul & Furhad, Md., "A Data-Driven Heart Disease Prediction Model Through K-Means Clustering-Based Anomaly Detection", SN Computer Science, vol. 2, No. 2, 2021.

[16] G. Magesh and P. Swarnalatha, "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction", Evolutionary Intelligence, vol.14, No. 1, PP. 583–593, 2021.A

[17] Wenqi Li, Ming Zuo, Hongjin Zhao, Qi Xu, and Dehua Chen, "Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks", Methods, vol.198, Pages 96-106 , February 2022.

[18] Pooja Rani, Rajneesh Kumar, Nada M. O. Sid Ahmed, and Anurag Jain, "A decision support system for heart disease prediction based upon machine learning", Journal of Reliable Intelligent Environments, vol. 7, No. 6, PP. 263–275 2021.

[19] Nichenametla, Rajesh & Maneesha, T. & Krishna, Hari., "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Engineering and Technology (UAE). 7. Pp. 363-36, 2018

[20] Alroobaea, Roobaea & Mechti, Seifeddine & Haoues, Mariem & Rubaiee, Saeed & Ahmed, Anas & Andejany, Murad & Bragazzi, Nicola & Sharma, Dilip & Kolla, Bhanu & Sengan, Sudhakar, "Alzheimer's Disease Early Detection Using Machine Learning Techniques", Wireless Personal Communications, 2021.

[21] Mehrnoosh Sadat Safi, and Seyed Mohammad Mehdi Safi, "Early detection of Alzheimer's disease from EEG signals using Hjorth parameters", Biomedical Signal Processing and Control, vol.65, 2021.

[22] "real dataset" from https://scikit-learn.org/stable/datasets/real_world.html

[23]"Messidor dataset" from (https://paperswithcode.com/dataset/messidor-1).

[24] "Kaggle dataset" from (https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset).

[25] "ADNI datasets" from (https://adni.loni.usc.edu/data-samples/)

[26] "CHASE dataset" from https://paperswithcode.com/dataset/chase-db1

[27] "IDRiD dataset" from https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid

[28] "OASIS dataset' from https://paperswithcode.com/dataset/oasis

[29] "Physionet dataset" from https://paperswithcode.com/dataset/physionet-challenge-2020

[30] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. Narasimha Prasad, "Diabetic retinopathy detection and classification using capsule networks", Complex & Intelligent Systems, 2021.

[31] Rubini, S. Saranya, and A. Kunthavai, "Diabetic retinopathy detection based on eigenvalues of the hessian matrix," Procedia Computer Science, vol. 47 pp. 311-318, 2015.

[32] Jahiruzzaman, Md, and ABM Aowlad Hossain, "Detection and classification of diabetic retinopathy using K-means clustering and fuzzy logic," In 2015 18th International Conference on Computer and Information Technology (ICCIT), pp. 534-538. IEEE, 2015.

[33] Pires, Ramon, Sandra Avila, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha, "Beyond lesion-based diabetic retinopathy: a direct approach for referral," IEEE journal of biomedical and health informatics, vol. 21, no. 1, pp. 193-200, 2015.

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on May 11,2023 at 17:08:24 UTC from IEEE Xplore. Restrictions apply.