UDC 575.
https://doi.org/10.2298/GENSR2303983S
Original scientific article

ENHANCED RETINAL IMAGE CAPTIONING SYSTEM FOR IDENTIFYING AND DIFFERENTIATING SEVERITY LEVELS OF DIABETIC RETINOPATHY

V. SIVAMURUGAN[1], K.R. UTHAYAN[2], V. THANIKACHALAM[3]

[1]Department of Information Technology, Sri SivaSubramaniya Nadar College of Engineering, Chennai, India
[2]Faculty of School of Computer Science & Engineering, Vellore Institute of Technology, Chennai, India

Sivamurugan, V., K.R. Uthayan, V. Thanikachalam (2023): *Enhanced retinal image captioning system for identifying and differentiating severity levels of diabetic retinopathy*. - Genetika, Vol 55, No.3, 983-996.
The objective of this research was to formulate a clinical decision support framework leveraging AI towards utilizing retinal fundus images for the identification and categorization of the four distinct stages of diabetic retinopathy, namely proliferative, severe, moderate, and mild. The devised system architecture integrated Long Short-Term Networks (LSTM), Generative Adversarial Networks (GAN), and pre-trained convolutional neural network (CNN) models. Following an exhaustive performance analysis, the most optimal image captioning model was identified and recommended to ophthalmologists for the purpose of identifying and categorizing diabetic retinopathy. Notably, the results revealed that employing ResNet50 with LSTM, in conjunction with enhanced retinal images, yielded superior accuracy of 0.975. The proposed methodology holds transformative potential for the realm of diabetic retinopathy diagnosis and classification, facilitating early detection and intervention to mitigate vision loss in individuals affected by diabetes.
Keywords: Convolutional Neural Network (CNN), Deep Learning, Diabetic Retinopathy (DR), Generative Adversarial Network (GAN), Image Captioning System, Long Short-Term Networks (LSTM)

INTRODUCTION

Ophthalmology relies extensively on retinal fundus images to discern irregularities in ocular structures and to assess the severity of diabetic retinopathy. Fundus photography, akin to a camera's film, captures images of the retina, allowing imaging of highly scattering tissues in non-invasive real-time. This imaging technique is invaluable for ophthalmologists in diagnosing

Corresponding author: K.R. Uthayan, Department of Information Technology, Sri SivaSubramaniya Nadar College of Engineering, Chennai, India, email: uthayankr@ssn.edu.in

abnormalities associated with the retina. This study has the primary aim to develop an automated system capable of detecting and classifying the varying degrees in severity levels in diabetic retinopathy evident in fundus retinal images. An image caption generator was crafted to achieve this, leveraging pre-trained LSTM and CNN models. Diabetic retinopathy, a condition impacting small blood vessels in retina, induces swelling in surrounding tissue, including macula—a central light-sensitive tissue in retina. Progressive damage to retina's microvascular and neurovascular system can lead to irreversible vision loss in diabetic patients.

Figure 1 illustrates the four levels of diabetic retinopathy severity: mild, moderate, severe, and proliferative. Severity determination involves assessing the presence of specific retinal abnormalities, such as exudates, hemorrhages and microaneurysms. The automated system developed herein facilitates the detection and classification of these abnormalities, offering a quantitative evaluation of diabetic retinopathy severity. This system contributes to early detection and intervention, potentially averting vision loss in diabetic patients.

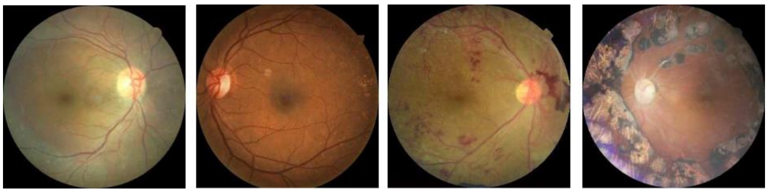

Figure 1. Four different levels of DR from Mild to Proliferative DR

Microaneurysms with hemorrhage or exudates are common in mild diabetic retinopathy. Moderate DR exhibits findings like mild DR, along with cotton wool spots or intraretinal microvascular abnormalities. The "4-2-1 rule" aids in diagnosing DR severity, with severe DR characterized by specific indicators in four quadrants. Proliferative diabetic retinopathy involves neovascularization, extending across the retina or into the vitreous cavity. This type showcases changes seen in non-proliferative DR, with new vessels crossing normal arteries and veins, indicating unregulated growth.

MATERIALS AND METHODS

Literature Survey

The design of a retinal image captioning system necessitates collaboration between CV and NLP communities. In AKARA *et al.* (2008) work, morphological operators optimized for detecting exudates in non-dilated pupils and low-contrast images achieved 80% sensitivity and 99.5% specificity, validated against expert ophthalmologists' ground-truths. DILIP *et al.* (2017) employed preprocessing techniques, extracting features for quantitative analysis, with exudate area proving effective for diabetic retinopathy prediction. BEHDAD *et al*. (2018) have proposed a double stage approach for microaneurysm identification, achieving 0.471 average sensitivity on the ROC dataset.

VELLAKANI (2020) proposed an OCT image caption system, while SUDESHNA *et al*. (2018) removed blood vessels and optic discs for classification. USMAN *et al.* (2014) classified

DR severity levels, requiring manual annotation. KEDIR *et al.* (2018) designed an automated system to detect hemorrhage, using partial least square classifier. RAMASUBRAMANIAN *et al*. (2018) proposed efficient hemorrhage detection with the same classifier used by them.

SHUNGYU *et al.* (2018) detected neovascularization in the optic disc using SVM, indicative of proliferative diabetic retinopathy. LEI *et al*. (2018) utilized a deep CNN to classify fundus images into six classes. LING *et al*. (2018) implemented a Multi-Sieving Deep Learning CNN for microaneurysm detection. LIPPI *et al*. (2019) conducted a comparative study on LSTM-generated language structure. CESC *et al*. (2018) introduced a CNN-based language model for image captioning. MIN *et al*. (2019) employed cross-domain learning for image captioning. UTHAYAN *et al*. (2021) used Genetic Algorithms and CNN for cancer detection. XIAODONG HE *et al*. (2008) proposed a CNN and RNN-based model for image captioning. YANG *et al*. (2018) developed a model for video content captioning. SIVAMURUGAN *et al*. (2020) created OCT image caption models, recommending them for clinical decision support.

Dataset

Table 1 show cases the allocation of images during training, validation, and testing, obtained from the Kaggle dataset.

Table 1. Kaggle Diabetic Retinopathy Dataset

S.No.	# Training	#Validation	# Testing	Severity of DR
1	1750.0	250.0	50	Mild
2	1750.0	250.0	50	Moderate
3	1750.0	250.0	50	Severe
4	1750.0	250.0	50	Proliferative

Recommended Approach

Figure 2 outlines the proposed methodology for generating captions for retinal images, consisting of three operational components: the training, validation, and the testing model for caption generator model. This model is employed for retinal image classification, utilizing highly trained CNN models to extract features of low-level, mid-level, and high-level. CNN models trained through Deep Learning are examined and utilized for extracting features.

Training the Pretrained model

The training process for the CNN-LSTM model involves utilizing fractional or captions that are only partially complete to accurately forecast the following word. The image attribute extractor, sequence generator, and decoder are three main parts of this training model. The pre-trained model generates attributes as input to the training model, represented as one-dimensional vectors of 32-bit floating-point numbers, stored in the features.pkl file. Multiple textual descriptions for each retinal image, based on present lesions, are created for training and validation, stored in the description.txt file.

A word embedding layer or sequence processor processes the text input, anticipating a sequence of inputs with a predetermined length. Subsequently, we have utilized a Long Short-Term Memory (LSTM) layer. Word embedding layer ignores padded values, and both the extractor of features and processor of sequences models generate a 256-element vector feature. To prevent overfitting, a 50% dropout regularization technique is employed. The decoder model combines the features from the attribute extractor and sequence predictor, with softmax prediction for dense layer to guess the succeeding word in text sequence. The training dataset is used to fit the training model, and model performance is evaluated using validation dataset after each epoch to address overfitting and quick learning issues. The trained model is saved after each epoch, and the model with the best performance is selected for further validation and testing.

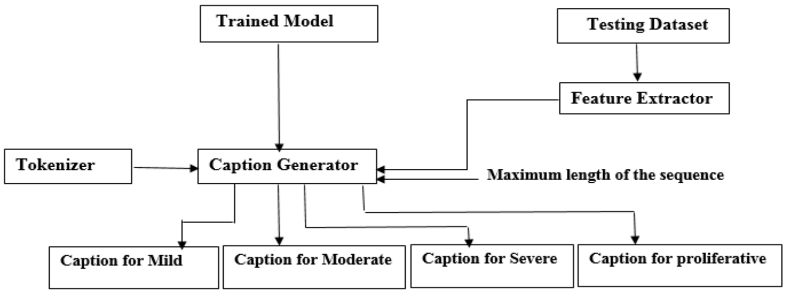

Figure 2. Proposed methodology for Image Caption Generator

Validation Model

After model fitting, its performance can be assessed on the validation dataset or holdout test data. The evaluation process includes creating descriptions for retinal fundus images in the validation dataset. The quality of the generated descriptions is measured using a standard cost function known as BLEU score. This metric gauges and checks whether the generated and true text descriptions are how much similar.

Image Caption Generation Model

The retinal image captioning model outlined in this study generates captions describing the various lesions present in the image. An innovative aspect of this research involves the creation of super-resolution retinal images. This enhancement is achieved through a combination of a Generative Adversarial Network and ResNet50, aiming to improve the accuracy of image captioning. Figure 3. delineates the different stages involved in the retinal image captioning model.

To generate captions for novel retinal images, it is imperative to utilize the model that has undergone training and the Tokenizer used in text encoding. In addition, the model requires the application of pre-trained model and the redefinition of LSTM-based training model, using the image to be captioned and the desired attribute for extraction. The training model will get input from the attributes extracted by pre-trained model. The resulting sequence of descriptions, excluding the start and end sequence tokens, forms captions illustrating various lesions and their corresponding class labels for the new retinal image. These captions play a crucial role in aiding ophthalmologists in detecting and differentiating various levels of diabetic retinopathy (DR) severity. The testing outcome from the image caption generator module with retinal images are showcased in Figures 4 to 7.

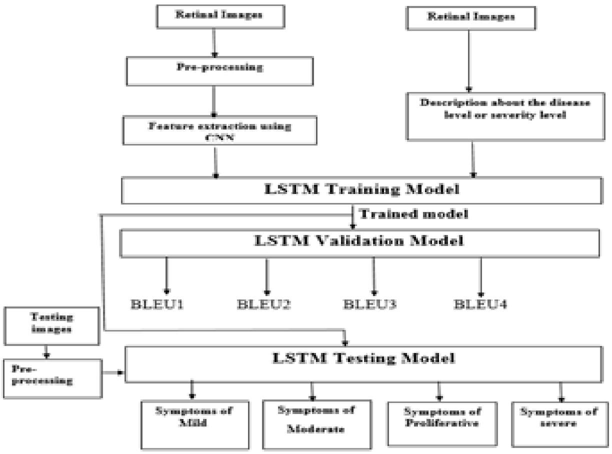

Figure 3. Various stages in the process of generating image captions

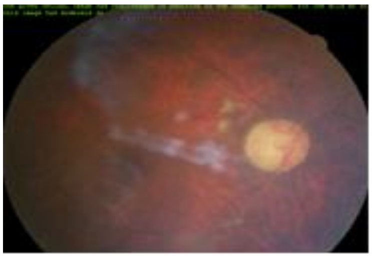

Figure 4. Generated Captions for Mild DR class

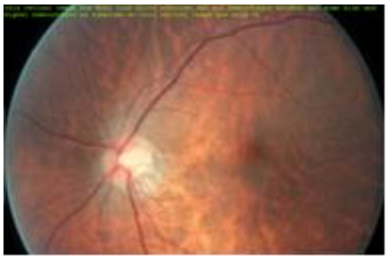

Figure 5. Predicted Captions for the Moderate CR class

Figure 6. Captions created for the Severe DR class

Figure 7. Captions made for the Proliferative DR class

RESULTS WITH DISCUSSION

Loss for Performance Analysis

To evaluate the effectiveness of image caption generators utilizing LSTM with 9 distinct pre-trained CNN models, the validation and training losses were computed. The training loss gauges the error during the model fitting process, while the validation loss measures the predictions error made by the trained model using validation images. The caption generator utilizing the Xception pre-trained model exhibited the lowest training and validation losses, indicating robust performance on the training dataset. When validated with the validation images, this model achieved a minimal validation loss of 0.119. Table 2 presents a comparative analysis of the minimum validation and training losses for 9 image captioning models, each developed with distinct pre-trained CNN model is employed in the caption generator with LSTM.

Table 2. Comparative evaluation of various pre-trained CNN models utilizing loss

Pre-Trained Model	Loss during Training	Loss during Validation
VGG16	0.128	0.124
VGG19	0.126	0.123
DenseNet121	0.130	0.121
DenseNet169	0.123	0.119
DenseNet201	0.123	0.119
ResNet50	0.122	0.120
InceptionV3	0.123	0.120
Xception	0.120	0.119
InceptionResNetV2	0.123	0.120

Confusion matrix for different pretrained CNN models

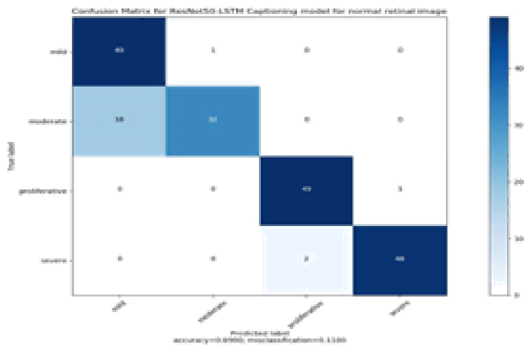

Figure 8. Confusion Matrix for ResNet50-LSTM model for the original retinal images

The confusion matrix visually represents the classification predictions of various image captioning models, organizing these predictions into columns and rows. Rows are used for the actual class, while columns are used for the predicted class. Correctly classified observations are shown in diagonal cells, while incorrectly classified ones are in off-diagonal cells. In Figure 8, the confusion matrix is presented for retinal images when utilizing the ResNet50-LSTM model for generating captions.

Figure 9 illustrates the confusion matrix depicting the classification predictions for super-resolution retinal images when employing the ResNet50-LSTM model for caption generation.

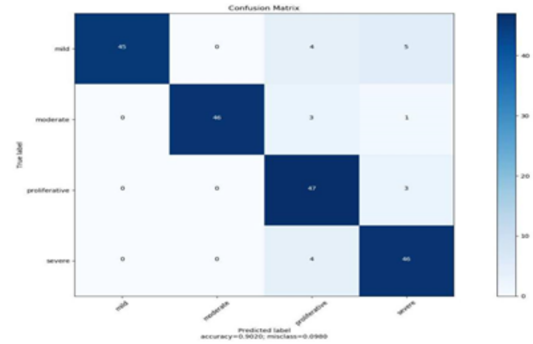

Figure 9. Confusion Matrix for ResNet50-LSTM model for the super resolution retinal images

Statistical analysis of Performance for Diabetic Retinal Image Caption Models.

For the purpose of performance analysis and comparison, various performance metrics were computed and identified. These metrics were utilized to evaluate the optimal performance of well-learned models for both images created using Super Resolution and gaussian noise superimposed images. The aim of this performance analysis is to furnish conclusive proof for recognizing best learned captioning model that will be suggested to ophthalmologists for retinal disease assessment. Table 3 offers a comprehensive overview of the statistical analysis of the performance of the retinal image captioning system.

Table 3. Performance metrics for image captioning models in Diabetic Retinal cases

CNN Models	Image Type	Kappa	Overall Accuracy	PPV_ Micro	PPV_ Micro	TPR_ Macro	TPR_Micro
VGG16	normal	.60667	.705	None	.705	.705	.705
VGG19	normal	.54	.655	None	.655	.655	.655
DenseNet121	normal	.6	.7	.70609	.7	.7	.7
DenseNet169	normal	.58	.685	.72229	.685	.685	.685
DenseNet201	normal	.6467	.735	.77465	.735	.735	.735
InceptionV3	normal	.38667	.54	.48224	.54	.54	.54
Xception	normal	.44667	.585	.63444	.585	.585	.585
InceptionResNetV2	normal	.49333	.62	.66491	.62	.62	.62
ResNet50	normal	.8533	.89	.91035	.89	.89	.89
ResNet50	Super resolution	.96667	.975	.97658	.975	.975	.975

Performance Analysis using Kappa

The assessment of the classifier's image captioning performance incorporates the use of kappa statistics. Kappa is computed by comparing observed accuracy to expected accuracy, providing a measure of how well the caption generator's classification aligns with ground truth data labeled by experts. The kappa values determine the level of agreement, categorized as excellent, good, moderate, mediocre, or inadequate. The caption generator employing ResNet50 as the feature extractor achieves a maximum kappa value of 0.8533, while testing on super-resolution images elevates the kappa value to 0.96667.

Table 4. Kappa value in deciding the agreement strength

Kappa Value	Strength of Agreement
< .20	Poor
.21 –.40	Fair
.41 –.60	Moderate
.61 – .80	Good
.8 – 1	Very Good

Overall Accuracy for Performance Analysis

The model's Overall Accuracy is determined by performing division operation between the total number of correct predictions across all severity levels and the total number of images in the process of generating captions. The caption generator utilizing ResNet50 as a feature extractor achieves an overall accuracy of 0.89. However, when tested with super-resolution images, the overall accuracy rises to 0.975. This strengthens the recommendation for ophthalmologists to utilize this image caption generator for precise classification of various severity levels in diabetic retinopathy.

Performance analysis using PPV_MACRO and PPV_MICRO

In multiclass classification, PPV_Macro is chosen to address dataset imbalances. The caption generator employing ResNet50 for feature extraction achieves the maximum PPV_Macro value at 0.91035, which increases to 0.97658 when tested on super-resolution images. This metric further supports the endorsement of this caption generator for ophthalmologists in assessing different severity levels in diabetic retinopathy.

TPR_MICRO and TPR_MACRO for analyzing performance

For the image captioning system utilizing ResNet50, the True Positive Rate is 0.89. Upon testing the model with super-resolution images, the True Positive Rate increases to 0.975. This reinforces the recommendation of this caption generator for accurate clinical decision-making.

Individual performance statistical analysis for each severity levels

To determine the best image caption generator, metrics such as Accuracy, F1Score, F2Score, Sensitivity, Specificity, and Precision are computed for each of the four severity classes: Mild, Moderate, Proliferative, and Severe DR. Based on benchmark results and minimum validation loss, the caption generator designed with ResNet50 consistently performs well across all severity levels. Tables 5 to 8 present individual performance analyses for each severity level of DR.

Performance analysis using accuracy

The accuracy assessment of image caption generators, utilizing ResNet50 as the feature extractor, reveals values of 0.905 for Mild, 0.905 for Moderate, 0.985 for Severe, and 0.98 for Proliferative Diabetic Retinopathy (DR), respectively. Additionally, when employing super-resolution images for caption generation, the accuracy surges to 0.98 for Mild, 0.98 for Moderate, 0.995 for Severe, and 0.995 for Proliferative DR, respectively. These findings indicate that the caption generation systems developed with LSTM and ResNet50 for feature extractor, have effectively produced captions for all classes. Detailed individual performance analyses for each severity level of DR—Mild, Moderate, Proliferative, and Severe—are presented in Tables 5 to 8.

Table 5. Analysis of individual performance statistics for retinal image captioning system designed using different pre-trained CNN models with LSTM for mild class

CNN Models	Image Type	Accuracy	F1 Score	F2 Score	Precision	Specificity	Sensitivity
VGG16	normal	.95	.91089	.91633	.90196	.96667	.92
VGG19	normal	.975	.95238	.98039	.90909	.96667	1
DenseNet121	normal	.885	.76768	.76305	.77551	.92667	.76
DenseNet169	normal	.895	.82353	.91078	.71014	.86667	.98
DenseNet201	normal	.98	.96078	.97222	.94231	.98	.98
InceptionV3	normal	.88	.78571	.83969	.70968	.88	.88
Xception	normal	.845	.752	.8545	.62667	.81333	.94
InceptionResNetV2	normal	.87	.75	.76772	.7222	.9	.78
ResNet50	normal	.905	.8361	.9176	.73134	.88	.98
ResNet50	Super resolution	.98	.96154	.98425	.92593	.97333	1

Table 6. Analysis of individual performance statistics retinal image captioning system designed using different pre-trained CNN models with LSTM for moderate class

CNN Models	Image Type	Accuracy	F1 Score	F2 Score	Precision	Specificity	Sensitivity
VGG16	normal	.855	.77519	.89606	.63291	.80667	1
VGG19	normal	.73	.625	.7653	.47872	.67333	.9
DenseNet121	normal	.765	.56075	.5836	.52632	.82	.6
DenseNet169	normal	.745	.53211	.5598	.49153	.8	.58
DenseNet201	normal	.89	.78431	.7936	.76923	.92	.80
InceptionV3	normal	.675	.15584	.1321	.2222	.86	.12
Xception	normal	.66	.38182	.4038	.35	.74	.42
InceptionResNetV2	normal	.72	.54839	.62044	.45946	.73333	.68
ResNet50	normal	.905	.77108	.6867	.9697	.993	.64
ResNet50	Super resolution	.98	.95833	.93496	1	1	.92

Table 7. Analysis of statistics for individual performance retinal image captioning system designed using different pre-trained CNN models with LSTM for Severe class

CNN Models	Image Type	Accuracy	F1 Score	F2 Score	Precision	Specificity	Sensitivity
VGG16	normal	.75	0.0	0	None	1	0
VGG19	normal	.75	0.0	0	0	1	0
DenseNet121	normal	.835	.61176	.55319	.74286	.94	.52
DenseNet169	normal	.88	.7551	.74559	.77083	.92667	.74
DenseNet201	normal	.755	.59504	.6642	.50704	.76667	.72
InceptionV3	normal	.705	.40404	.4016	.40816	.80667	.5
Xception	normal	.82	.48571	.3863	.85	.98	.34
InceptionResNetV2	normal	.785	.53763	.5144	.5814	.88	.5
ResNet50	normal	.985	.96997	.96386	.97595	.99333	.96
ResNet50	Super resolution	.995	.9899	.98394	1.0	1.0	.98

Table 8. Individual performance statistics analysis retinal image captioning system designed using different pre-trained CNN models with LSTM for Proliferative loss

CNN Models	Image Type	Accuracy	F1 Score	F2 Score	Precision	Specificity	Sensitivity
VGG16	normal	.85	.75	.8333	.64286	.83333	.9
VGG19	normal	.855	.71287	.71713	.70586	.9	.72
DenseNet121	normal	.915	.844	.85603	.77966	.91333	.92
DenseNet169	normal	.85	.59459	.49107	.91667	.98667	.44
DenseNet201	normal	.845	.58667	.48889	.88	.98	.44
InceptionV3	normal	.82	.67857	.72519	.6129	.84	.76
Xception	normal	.845	.67368	.65306	.25	.91333	.64
InceptionResNetV2	normal	.865	.65823	.56769	.89655	.98	.52
ResNet50	normal	.98	.9703	.9761	.96078	.98667	.98
ResNet50	Super resolution	.995	.9901	.99602	.98039	.99333	1

Performance analysis using F1 Score

The harmonic mean of recall and precision defines F1 Score, proves valuable in considering both false negatives and false positives, especially in datasets with uneven class distribution. For the caption generation system employing ResNet50 as the feature extractor, the F1 scores for Mild, Moderate, Severe, and Proliferative Diabetic Retinopathy (DR) are 0.8361, 0.77108, 0.9697, and 0.9703, respectively. When utilizing super-resolution images, these scores increase to 0.96154 for Mild, 0.95833 for Moderate, 0.9899 for Severe, and 0.9901 for Proliferative DR. These outcomes affirm the commendable performance of the ResNet50-based LSTM caption generation system, even in the presence of class distribution imbalances.

F2 Score for performance analysis

The weighted average between recall and precision defines F2 Score, proves beneficial when the cost of false negatives outweighs false positives, making it suitable for imbalanced class distribution datasets. The caption generation system employing ResNet50 as the feature extractor achieves F2 score values of 0.9176 for Mild, 0.6867 for Moderate, 0.96386 for Severe, and 0.9761 for Proliferative DR. Upon utilizing super-resolution images for caption generation, these F2 score values improve to 0.98425 for Mild, 0.93496 for Moderate, 0.98394 for Severe, and 0.99602 for Proliferative DR. These findings favor the efficacy of the ResNet50-with LSTM caption generation system for classification tasks with imbalanced class distributions.

Precision for performance analysis

Precision quantifies correct positive predictions relative to total positive predictions. The caption generation system utilizing ResNet50 as the feature extractor exhibits precision values of 0.73134 for Mild, 0.9696 for Moderate, 0.97595 for Severe, and 0.96078 for Proliferative DR, respectively. When incorporating super-resolution images for caption generation, precision values improve to 0.92593 for Mild, 1.0 for Moderate, 1.0 for Severe, and 0.98039 for Proliferative DR, respectively.

Performance analysis using Specificity (TNR)

Specificity, also known as True Negative Rate (TNR), represents correct negative predictions relative to total negatives. For the caption generation system designed with ResNet50 as the feature extractor, specificity is 0.88 for Mild, 0.993 for Moderate, 0.99333 for Severe, and 0.98667 for Proliferative DR, respectively. Furthermore, when utilizing super-resolution images for caption generation, specificity improves to 0.97333 for Mild, 1.0 for Moderate, 1.0 for Severe, and 0.99333 for Proliferative DR, respectively.

Performance analysis using Sensitivity (TPR)

Sensitivity quantifies correct positive predictions relative to total positives. The ResNet50-based caption generation system demonstrates sensitivity values of 0.98 for Mild, 0.64 for Moderate, 0.96 for Severe, and 0.98 for Proliferative DR, respectively. When utilizing super-resolution images for caption generation, sensitivity increases to 1.0 for Mild, 0.92 for Moderate, 0.98 for Severe, and 1.0 for Proliferative DR, further supporting the ResNet50-LSTM captioning model's effectiveness. These results recommend the model for eye specialists in predicting various levels of DR severity.

CONCLUSION AND FUTURE WORK

The study determined that the inclusion of super-resolution images generated through generative adversarial networks significantly improved the model's performance. The ResNet50 with LSTM demonstrated the highest accuracy at 97.5% when generating captions for various levels of diabetic retinopathy (DR) severity, as assessed by performance metrics. Additionally, the model exhibited notable performance in Positive Predictive Value Macro (97.658%) and True Positive Rate Macro (97.5%). Consequently, our research successfully devised an AI-based based support system capable of clinical decision for categorizing distinct stages of diabetic retinopathy. Further advancements in the research could involve integrating handcrafted features with CNN features, utilizing this amalgamated feature set to formulate a novel training model for caption generation.

ACKNOWLEDGEMENT

This work focused on research endeavor received funding from Sri SivaSubramaniya Nadar College of Engineering. We extend wholehearted gratitude to our subordinates at Sri SivaSubramaniya Nadar College of Engineering for their valuable insights and expertise, which significantly contributed to the success of the research.

Received, April 18th, 2023
Accepted December 12th, 2023

REFERENCES

AKARA SOPHARAK, A., A., BUNYARIT UYYANONVARA, B., SARAH BARMAN, C., WILLIAMSON (2008): Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. Computerized Medical Imaging and Graphics, *32*(8): 720–727.

BEHDAD, D., J., ZHANG, F., HUANG, M., HAAR (2018): Retinal Microaneurysms Detection Using Local Convergence Index Features. IEEE Transactions on Image Processing, *27*(7): 3300-3315.

CESC CHUNSEONG P., Y., KIM, G., KIM (2018): Retrieval of Sentence Sequences for an Image Stream via Coherence Recurrent Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, *40*: 945-957.

DILIP S. S., N., SHRUTHI, P., KHOBRAGADE (2017): Diabetic Retinal Fundus Images: preprocessing and Feature Extraction for Early Detection of Diabetic Retinopathy. Biomedical and Pharmacology Journal, *10*(2): 615-626.

KEDIR, M. A., P.G., VAN ETTEN, J.P., MARTINEZ, K.W., ROUWEN, A. K., VERMEER, L. J.VAN VLIET (2018): An Automated System for the Detection and Classification of Retinal Changes Due to Red Lesions in Longitudinal Fundus Image. IEEE Transactions on Biomedical Engineering, *65*(6): 1382-1390.

LEI, Z., Y., ZHAO, J., YANG, Q., YU, X., XU (2018): Deep multiple instances learning for automatic detection of diabetic retinopathy in retinal images. IET Image Processing, *12*(4): 563-571.

LING, D., R., FANG, H., LI (2018): Clinical Report Guided Retinal Microaneurysm Detection with Multi-Sieving Deep Learning. IEEE Transactions on Medical Imaging, *37*(5): 1149-1161.

LIPPI, M.M., A., MONTEMURRO, M. D., ESPOSTI, G. CRISTADORO (2019): Natural Language Statistical Features of LSTM-Generated Texts. IEEE Transactions on Neural Networks and Learning Systems, *30*: 3326-3337.

MIN, Y., W., ZHAO, W., XU, Y. Z., ZHAO, X., CHEN, K., LEI (2019): Multitask Learning for Cross-domain Image Captioning. IEEE Transactions on Multimedia, *21*: 1047-1061.

RAMASUBRAMANIAN, B. and S., SUNDARAMOORTHY (2018): Efficient approach for the automatic detection of hemorrhages in color retinal images. IET Journals & Magazines, *12*(9): 1540 – 1544.

SHUANG, Y., D., XIAO, Y., KANAGASINGAM (2018): Machine Learning Based Automatic Neovascularization Detection on Optic Disc Region. IEEE Journal of Biomedical and Health Informatics, *22*(3): 886- 894.

SIVAMURUGAN, V. and P., INDUMATHI (2020): An enhanced OCT image captioning system to assist Ophthalmologists in detecting and classifying eye diseases. Journal of X-ray Science & Technology, *28* (5): 975 – 988.

SUDESHNA, S. and S.P., MAITY (2018): Automatic Detection of Retinal Lesions for Screening of Diabetic Retinopathy, IEEE Transactions on Biomedical Engineering, *65*(3): 608-618.

USMAN, A.M., S., KHALID, A., TARIQ, S., KHAN, F., AZAM (2014): Detection and Classification of retinal lesions for grading of diabetic retinopathy. Computers in Biology and Medicine, *45*(1): 161-171.

UTHAYAN, K.R., S., MOHANAVALLI, B., NIVETHA, S., DHIVYA (2022): Optimized feature selection for early cancer detection, Genetika – Belgrade, *53*(3): 1297-1309.

XIAODONG, H.E. and L., DENG (2017): Deep Learning for Image-to-Text Generation. A Technical Review, IEEE Signal Processing Magazine, *34*: 109-116.

YANG, J. Z., A.I., JIANGBO, Y., BIN, A., HANJALIC, H.T., SHEN, J., YANLI (2018): Video Captioning by Adversarial LSTM. IEEE Transactions on Image Processing, *27*(11): 5600-5611.

UNAPREĐENI SISTEM ZA PREVOĐENJE RETINALNE SLIKE ZA IDENTIFIKACIJU I DIFERENCIJU NIVOA DIJABETIČNE RETINOPATIJE

V. SIVAMURUGAN[1], K.R. UTHAYAN[1], V. THANIKACHALAM[2]

[1]Departman za informacione tehnologije, Sri SivaSubramaniia Nadar, Inženjerski fakultet, Čenaj, Indija
[2]Fakultet računarskih nauka i inženjerstva, Tehnološki institut Velore, Čenaj, Indija

Izvod

Cilj ovog istraživanja je bio da se formuliše okvir za podršku kliničkom donošenju odluka koji koristi veštačku inteligenciju ka korišćenju slika fundusa mrežnjače za identifikaciju i kategorizaciju četiri različita stadijuma dijabetičke retinopatije, naime proliferativne, teške, umerene i blage. Osmišljena arhitektura sistema je integrisala dugoročne mreže (LSTM), generativne adversarijske mreže (GAN) i unapred obučene modele konvolucionih neuronskih mreža (CNN). Posle iscrpne analize performansi, identifikovan je najoptimalniji model opisa slika i preporučen oftalmolozima u svrhu identifikacije i kategorizacije dijabetičke retinopatije. Značajno je da su rezultati otkrili da korišćenje ResNet50 sa LSTM-om, u kombinaciji sa poboljšanim slikama mrežnjače, daje superiornu preciznost od 0,975. Predložena metodologija ima transformativni potencijal za oblast dijagnoze i klasifikacije dijabetičke retinopatije, olakšavajući rano otkrivanje i intervenciju za ublažavanje gubitka vida kod osoba pogođenih dijabetesom.

Primljeno 18.IV.2023.
Odobreno 12. XII 2023.