Single-modality and joint fusion deep learning for diabetic retinopathy diagnosis

Sara El-Ateif*, Ali Idri*
*ENSIAS, Mohammed V University, Rabat, Morocco

Read paper

ENSIAS

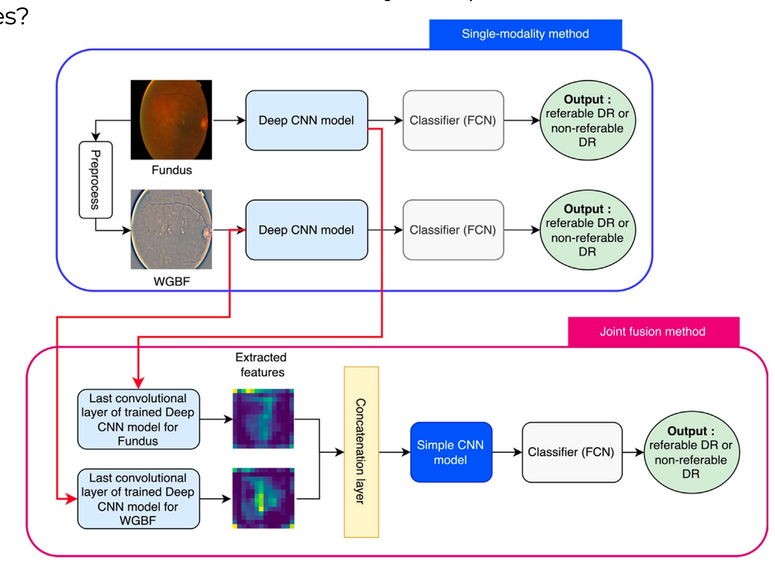
DEEP LEARNING INDABA

Introduction

We trained **VGG19 , ResNet50V2, DenseNet121, InceptionV3, InceptionResNetV2, Xception and MobileNetV2**) using two approaches : **single-modality and joint fusion** using **Fundus and Weighted Gaussian Blur Fundus (WGBF)** (i.e. preprocessed Fundus images).
To evaluate the performance of these models we used :

(1) **Accuracy, sensitivity, specificity, precision, F1-score, and area under the curve (AUC)** over two datasets: **APTOS 2019 blindness detection** and **Messidor-2.**
(2) **Scott-Knott Effect Size difference (SK ESD)** statistical test to **cluster the DL techniques** into most significant groups with non-negligible differences,
(3) **Borda count voting method to rate the best models** found in the best cluster of the SK ESD test [20].

This study addressed the following four research questions:
(RQ1): What is the overall performance of DL models using a single-modality for DR classification? Is there any single- modality DL architecture that outperforms others?
(RQ2): How does a modality impact the diagnostic performance of a DL technique?
(RQ3): What is the overall performance of the joint fusion DL models in DR classification?
(RQ4): How do joint fusion DL architectures perform in comparison with single-modality DL models?
(RQ5): How do the results of this study compare with those of state-of-the-art approaches?



Methodology

Generate the WGBF modality:

We generate the second modality WGBF following the steps described in [23] :
1) Creating a binary label: We relabeled the target variable that contained the five DR stages from 0 to 4 to **non-referable DR or 0 (from 0 to 1),** and **referable DR or 1 (from 2 to 4).** Referable DR refers to moderate or worse non-proliferative diabetic retinopathy and/or diabetic macular edema.
2) Cropping the images: We cropped the Fundus images to distinguish the retina from the background.
3) Resizing the images: We **resized the images to 224×224 pixels** to have the same radius and fit the preprocessing requirements of all the seven DL models.
4) Graham algorithm: We **subtracted the local average color** from the Fundus images and **mapped 50% gray to the local average** to make the blood vessels and lesion areas more explicit [24].

Training and testing processes:

Step 1: We train the single-modality (ImageNet pre-trained seven DL CNNs) models at first and save the weights for each modality (Fundus and WGBF).
Step 2: We then load the saved weights for each respective modality, extract the features using same DL model for both modalities and then concatenate the extracted features from the Fundus and WGBF modalities and feed them to the Simple CNN model that extracts joint features from both concatenated features and then that new features matrix is fed to the FCN (classifier) to predict if a certain patients has referable DR and non-referable DR.

Results

Evaluation and comparison of single-modality architectures (RQ1)

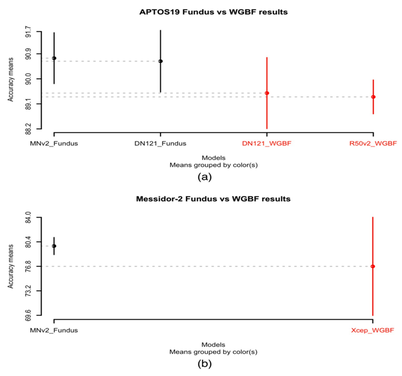
Table 3
Statistics on SK ESD ranks of DL models for the APTOS19 modalities.

Model	Scott-Knott ESD rank APTOS19 Fundus	WGBF	Sum of SK ESD ranks (SR)	Sum of SK ESD ranks differences (SDR)
Xception	2	5	7	3
MobileNetV2	1	2	3	1
DenseNet121	1	1	2	0
ResNet50V2	3	1	4	2
InceptionV3	2	4	6	2
InceptionResNetV2	4	5	9	1
VGG19	4	3	7	1

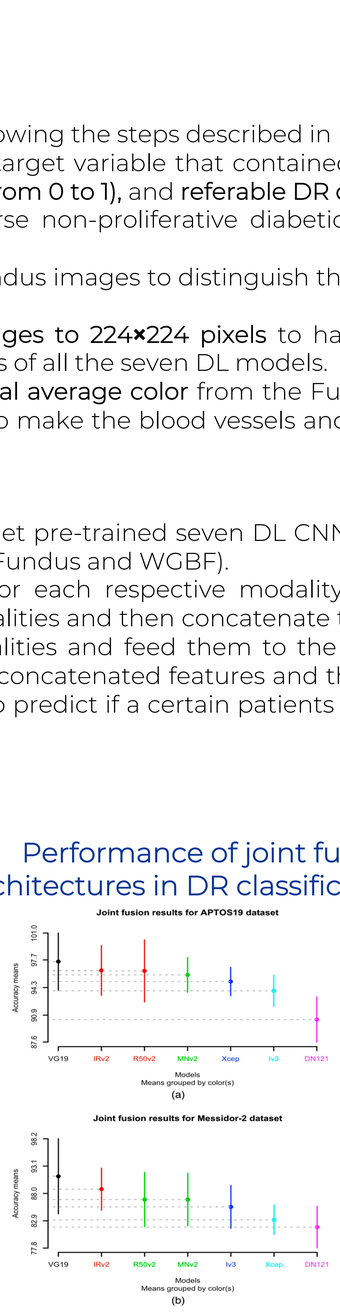
Table 4
Statistics on SK ESD ranks DL models for the Messidor-2 modalities.

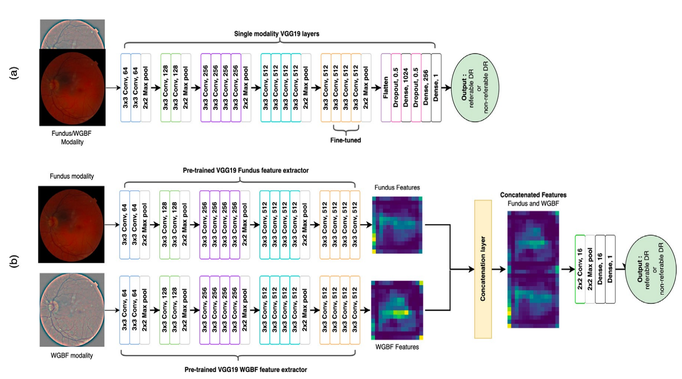
Model	Scott-Knott ESD rank Messidor-2 Fundus	WGBF	Sum of SK ESD ranks (SR)	Sum of SK ESD ranks differences (SDR)
Xception	3	1	4	2
MobileNetV2	1	3	4	2
DenseNet121	6	4	10	2
ResNet50V2	4	2	6	2
InceptionV3	2	2	4	0
InceptionResNetV2	5	3	8	2
VGG19	5	3	8	2

Modalities impact on the performance of the DL models (RQ2)

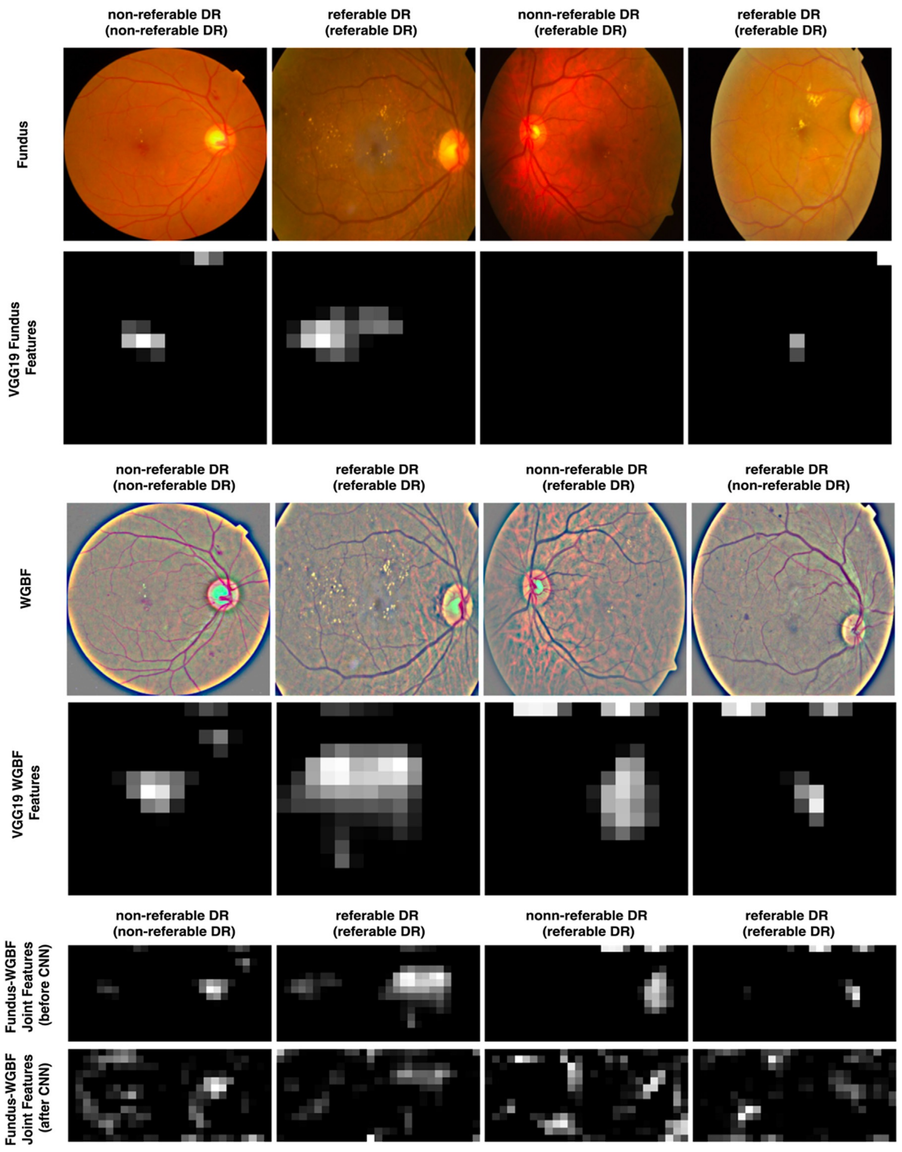


Performance of joint fusion DL architectures in DR classification (RQ3)





Comparison of the best performing model with state-of-the-art models (RQ5)



Conclusion

(RQ1): What is the overall performance of DL models using single-modality in DR classification? Is there any single- modality DL architecture that outperforms the others? The best model in terms of SR and SDR:

- **APTOS19 dataset** is the **DenseNet121 model (accuracy = 90.63%),** as it ranked first across the Fundus and WGBF modalities and obtained the best scores.
- **Messidor-2 dataset** best was **InceptionV3 (accuracy = 75.25%).**

(RQ2): How does a modality impact the diagnostic performance of a DL architecture? The **Fundus DL models were ranked top** in all of the empirical evaluations, and thus determined the Fundus modality to be the most favorably influential on the DL technique performance in comparison with the modality WGBF.

(RQ3): What is the overall performance of joint fusion DL models in DR classification? The **best joint fusion DL model was VGG19 for both APTOS19 and Messidor-2**. Additionally, the worst joint fusion DL model was DenseNet121, with accuracy values of **90.35%** and **81.71%** over the APTOS19 and Messidor-2 datasets, respectively.

(RQ4): How do joint fusion DL architectures perform in comparison with single-modality DL models? **Joint fusion DL models outperformed single-modality DL models** over both APTOS19 and Messidor-2 in DR diagnosis.

(RQ5): How do the results of this study compare with state-of-the-art approaches? Joint fusion VGG19, the best ranked model, is slightly worse than the Attention Fusion network performance with a difference of 5.6% in AUC and with an increase of 8% in AUC in comparison with the Cascaded Framework state-of-the-art models on the Messidor dataset.

References

[20] S. Elmidaoui, L. Cheikhi, A. Idri, A. Abran, Predicting software maintainability using ensemble techniques and stacked generalization, CEUR Workshop Proc. 2725 (2020) 1–16 .
[23] C. Lahmar, A. Idri, On the value of deep learning for diagnosing diabetic retinopathy, Health Technol. (2021) (Berl)., Oct., doi: 10.1007/ s12553- 021- 00606-x .
[24] B. Graham, Kaggle diabetic retinopathy detection competition, University of Warwick, 2015 .
[26] H. Zerouaoui, A. Idri, Deep hybrid architectures for binary classification of medical breast cancer images, Biomed. Signal Process. Control 71 (2022) 103226 Jan., doi: 10.1016/j.bspc.2021.103226 .

Acknowledgements

We thank the Google Ph.D. Fellowship program for the support provided to Sara El-Ateif.

Google Research