MTCSNN: Multi-task Clinical Siamese Neural Network for Diabetic Retinopathy Severity Prediction

Chao Feng, Jui Po Hung, Aishan Li, Jieping Yang, Xinyu Zhang
{chfeng, rhung, aishanl, jiepingy, xyuzhang}@umich.edu

Abstract
Diabetic Retinopathy (DR) has become one of the leading causes of vision impairment in working-aged people and is a severe problem worldwide. However, most of the works ignored the ordinal information of labels. In this project, we propose a novel design MTCSNN, a Multi-task Clinical Siamese Neural Network for Diabetic Retinopathy severity prediction task. The novelty of this project is to utilize the ordinal information among labels and add a new regression task, which can help the model learn more discriminative feature embedding for fine-grained classification tasks. We perform comprehensive experiments over the RetinaMNIST, comparing MTCSNN with other models like ResNet-18, 34, 50. Our results indicate that MTCSNN outperforms the benchmark models in terms of AUC and accuracy on the test dataset. The code is available at https://github.com/chlorophyllccc/MTCSNN_for_Diabetic_Retinopathy_Severity_Prediction.

1 Introduction
With the global increase of diabetes, many consequences are also experiencing a growing trend in the world. Diabetic Retinopathy (DR) is a leading cause of vision loss in working-age adults. It was estimated that around 16 million Americans would be affected by Diabetic Retinopathy by 2050. To prevent people from blindness and provide early and proactive assistance, the detection of Diabetic Retinopathy should be of great importance.

In terms of medical diagnosing and clinical decision-making, grading systems based on medical images are usually applied to assess disease severity. Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the main two phases and there are in total of 5 severity levels with order information. The higher level of the label, the more severe the disease. However, though there has existed relevant work on the classification of Diabetic Retinopathy, most of the research just processed and classified different levels as independent variables regardless of the ordinal information among the labels.

In this project, a siamese neural network-based model is proposed to classify and predict Diabetic Retinopathy severity levels from standardized MNIST-like retina fundus images. Instead of treating labels as independent variables, we incorporate the ordinal information of labels for better prediction performance. To model such dependency, we propose a siamese neural network to simultaneously conduct image classification and difference regression among predicted labels. The total loss is composed of the cross-entropy loss from classification and the mean squared error from difference regression, where the additional loss can be viewed as a regularization term.
We perform comprehensive experiments over an open dataset RetinaMNIST and compare the performance of our model MTCSNN with benchmark models like ResNet-18, ResNet-34, ResNet-50 and Google AutoML Vision. Experimental results indicate that our proposed method can achieve a greater prediction performance for Diabetic Retinopathy severity. Meanwhile, the improvement also shows the capability to apply our model to solve other tasks with ordinal information like age estimation in the future.

2 Related Work

Diabetic Retinopathy has long been a main reason for severe vision loss in working-aged people. There has been numerous research done to overcome the problems. In order to better detect and diagnose Diabetic Retinopathy, conventional machine learning technique, such as K-nearest neighbor and random forest classifiers, have been widely applied. In 2005, M. Niemeijer firstly implemented a K-nearest neighbor classifier with a red lesion candidate detection system based on pixel classification.[2] The final automatic detection performance was proved to be close to the classification performance of a human expert. In [3], a random forest classifier was performed to classify the fundus photography data. As a result, random forest clearly outperformed other machine learning methods like logistic regression.

In recent years, with the development of computing resources, more advanced methods and classifiers using neural networks have been researched and implemented in the classification of Diabetic Retinopathy. In [4], Valliappan Raman and his team focused on image preprocessing and feature extraction. They firstly designed an image enhancement module to preprocess noise and then identified the blood vessels and optic disc. Then the feature extraction module extracted different features and they were fed into an Artificial Neural Network (ANN) classification algorithm to detect abnormal features and classify different stages of Diabetic Retinopathy. To further improve the existing accuracies, Suvajit Dutta and his team applied Fuzzy C-means clustering (FCM) to find out cluster levels of the training data. Then, they compared the classification performance of Feed forward Neural Network (FNN), Deep Neural Network (DNN) and a Convolutional Neural Network (CNN), VGG-16 model. The result showed that classification using DNN had the highest accuracy among the three models.

3 Dataset

3.1 Dataset Information

One of the most common causes of vision damage in working-age people is Diabetic Retinopathy, which is caused by the long-term effects of diabetes.[1] Therefore, a timely and proactive diagnosis of Diabetic Retinopathy is of great importance to preventing eye damage and blindness. In the realm of ophthalmology, ordinal information is usually collected by the different grading levels of retinal fundus photographs. To evaluate the effectiveness of our proposed model, we used standardized retina fundus images from an open dataset RetinaMNIST. RetinaMNIST comes from a large-scale biomedical MNIST-like dataset MedMNIST v2.[5] The dataset preprocessed 1600 retinal fundus images from DeepDR Diabetic Retinopathy Image Dataset (DeepDRiD).[6] There are in total 5 disease severity levels of Diabetic Retinopathy. The five levels and according images are shown in Table[1]
Table 1: The five levels of international clinical Diabetic Retinopathy (DR) severity. NPDR and PDR are two main phases: non-proliferative diabetic retinopathy and proliferative diabetic retinopathy. Adopted from [6].

| Disease Severity Level | Description          |
|------------------------|----------------------|
| Grade 0                | Healthy - No apparent retinopathy |
| Grade 1                | Mild – NPDR          |
| Grade 2                | Moderate – NPDR      |
| Grade 3                | Severe – NPDR        |
| Grade 4                | PDR                  |

Figure 1: The sample images from the source dataset (DeepDRiD), where each image comes from one level.

3.2 Data Setting

All the images in DeepDRiD were cropped and preprocessed into $3 \times 28 \times 28$ in RetinaMNIST. The sample images from RetinaMNIST are shown below, where each row and column contains 10 standardized images. There are 1600 retina fundus samples in total for this experiment. The data samples were split into three sets, where the source validation set from DeepDRiD was utilized as the test set. 90% of the source training set was used as the training set and the remaining 10% was used as the validation set.

Figure 2: Visualization of the sample images in RetinaMNIST. Each row and column contains 10 images.
4 Proposed Method

4.1 Model Selection

In this project, we choose the siamese neural network as our prediction model since it can fully utilize the ordinal information among labels, given that the dataset RetinaMNIST has the ordinal information on the degree of retinal disease. Siamese neural network consists of two identical artificial neural networks each capable of learning the hidden representation of an input vector. The two neural networks are both feed-forward perceptrons, and employ error back-propagation during training; they work parallelly in tandem and their outputs are compared at the end. The output generated by a Siamese Neural Network execution can be considered the semantic similarity between the projected representation of the two input vectors [7].

![Figure 3: An example: the architecture of SNN](image)

The reason for choosing this model is that compared with other methods of classification, the Siamese Neural Network can learn from semantic similarity from the ordinal information among the different levels, which is very applicable to the dataset RetinaMNIST.

4.2 Specific Method

We propose a Multi-task Clinical Siamese Neural Network that contains two subnetworks. The neural networks would have the same architecture and parameters, which have some convolution layers and two fully-connected branches. One is for the normal classification task and the other branch is for the regression task. During the course of training, two randomly sampled images with various labels would be fed into the model, and the model can simultaneously learn for the classification and difference regression task.

To be specific, the regression task is to make the model predict the distance between two labels of the two images fed into the net, where the distance is defined as $|y_2 - y_1|$, $y_2$ and $y_1$ are the ground-truth labels of images. We can view this distance as a kind of similarity evaluation metric, and this task can enforce the model to tell the difference in grading levels between images by itself, which is beneficial for this fine-grained problem. This is also somewhat similar to contrastive learning. In contrastive learning, the dot product between feature embedding is leveraged to compute the
similarity score, and we would like the distance of embeddings within the same class to be small, and the distance between images from different classes to be large. Here we use the distance to replace the dot product as a way of "supervised contrastive learning", and we desire the model to learn more discriminative embeddings for the task.

More importantly, using the siamese neural network architecture and this difference regression task can fully utilize the ordinal information among labels, which is not leveraged by previous work. The ordinal information is conducive to supervising a more robust model for better performance.

4.3 Training Objective

Our loss function consists of two parts, the first part is general cross-entropy loss $L_1$, and the second one is mean square error loss $L_2$, which is used for the difference regression task.

\[
    L_1 = \frac{1}{N} \sum_{i=1}^{N} -\log \frac{e^{S_{y_i}}}{\sum_{j=1}^{C} e^{S_{y_j}}}
\]

(1)

\[
    L_2 = \frac{2}{N} \sum_{i=1}^{N} (|G_{y_2} - G_{y_1}| - |y_2 - y_1|)^2
\]

(2)

\[
    \text{Loss} = L_1 + \lambda L_2
\]

(3)

By using these two loss functions, we force the model to have the ability to differentiate the grading levels of different images, while can correctly classify the level of the images. The additional loss $L_2$ can also be viewed as a way of regularization, which can help the model avoid overfitting.

Here, $N$ is the number of training data samples, $C$ is the number of classes, and $\lambda$ is the hyperparameter needed to be tuned during the training process. $S$ represents the raw logits from
the classification branch and $G$ represents the raw logits from the difference regression branch. $|\cdot|$ is the absolute value operator.

5 Experiments

5.1 Implementation Details

We use three benchmarking models ResNet-18, ResNet-34, ResNet-50 as the backbone. The source training set from DeepDRiD is divided into a 90% training set (1080) and a 10% validation set (120). The source validation set is used as the test set, which contains 400 images. We use Adam Optimizer with learning rate $1e^{-3}$, $\beta_1$ is 0.9 and $\beta_2$ is 0.999. There is no weight decay here. The batch size is 36 and the maximum epoch for training is 20. We set the hyperparameter $\lambda$ to be 0.5 during the training across different architectures. All experiments are done on NVIDIA GeForce RTX 3080 Laptop GPU.

5.2 Experimental Results

| Methods                  | AUC   | Test Accuracy |
|--------------------------|-------|---------------|
| ResNet-18                | 0.7170| 0.5225        |
| ResNet-34                | 0.7024| 0.5250        |
| ResNet-50                | 0.7076| 0.4975        |
| AutoKeras                | 0.7190| 0.5030        |
| auto-sklearn             | 0.6900| 0.5150        |
| Google AutoML Vision     | 0.7500| 0.5310        |
| MTCSNN with ResNet-18    | 0.7373| 0.5525        |
| MTCSNN with ResNet-34    | 0.7419| 0.5600        |
| MTCSNN with ResNet-50    | 0.7122| 0.5375        |

As the baseline, we trained a deep learning model using the ResNet architecture to classify the severity of retina images. We used a batch size of 128 and tuned the learning rate to $1e^{-3}$. Due to the small size of the dataset, the results were sub-optimal. Even the highly well-trained Google AutoML Vision achieved only accuracy of 0.531.

Using the carefully-tuned MTCSNN, we achieved better results than accomplished methods such as auto-sklearn, AutoKeras, and Google AutoML Vision. By fully utilizing ordinal information among labels, our approach did not have to involve massive, unwieldy models and barely added any parameters in the backbone and the classifier.
5.3 Result Analysis

As shown in Fig. 5 during the course of training, the loss on the training and validation set gradually converge, which means that our method and training strategy works. The number of training epochs is 20 to avoid overfitting. In addition, the confusion matrices shown in Fig. 6 indicate that our method can enforce the model to make predictions more accurately. Although the model still fails in some cases, it can make the predicted label approach the ground-truth label further, which means that our method can enforce the model to learn more discriminative feature embeddings for better classification performance.

Figure 5: Loss history of MTCSNN with ResNet-18, ResNet-34 and ResNet-50 (from left to right)

Figure 6: Confusion Matrix of ResNet-18, ResNet-34 and ResNet-50

Figure 7: Confusion Matrix of MTCSNN with ResNet-18, ResNet-34 and ResNet-50
6 Conclusion

In this study, we propose MTCSNN, a novel design of multi-task Diabetic Retinopathy image classification inspired by the Siamese Neural Network. Compared with ResNet-18,34,50, MTCSNN has achieved better results in challenging standard diabetic retinopathy severity classification tasks without adding parameters in the backbone and the classifier. Our MTCSNN also outperforms the model generated by Google AutoML Vision.

From the results, we confirmed the intuition that Siamese Neural Network’s property of semantic similarity solves ordinal-regression-based problems well. The linear combination of the two losses in the difference regression task implicitly includes an additional regularization term, which avoids the risk of over-fitting. Besides, since our difference regression task provides additional evidence for the relevance of the features [9], the data augmentation is implicitly performed.

As a result, our proposed multi-task learning framework is able to show robustness and universality, and we believe it will benefit other ordinal classification tasks like age estimation.

These preliminary results are promising, but there are still many challenges. One is to find the fittest backbone for certain medical image datasets, and another is to leverage ordinal information better. We expect future work to improve MTCSNN.

References

[1] Koushik Tripathy Unnati V. Shukla. Diabetic Retinopathy. StatPearls, StatPearls [Internet], 2021 August 29. https://www-ncbi-nlm-nih-gov.proxy.lib.umich.edu/books/NBK560805/.

[2] M. Niemeijer, B. van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, and M.D. Abramoff. Automatic detection of red lesions in digital color fundus photographs. IEEE Transactions on Medical Imaging, 24(5):584–592, 2005.

[3] Ramon Casanova, Santiago Saldana, Emily Y Chew, Ronald P Danis, Craig M Greven, and Walter T Ambrosius. Application of random forests methods to diabetic retinopathy classification analyses. PLOS one, 9(6):e98587, 2014.

[4] Valliappan Raman, Patrick Then, and Putra Sumari. Proposed retinal abnormality detection and classification approach: Computer aided detection for diabetic retinopathy by machine learning approaches. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), pages 636–641, 2016.

[5] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification, 2021.

[6] Deepdr diabetic retinopathy image dataset (deepdrid), 2020. https://isbi.deepdr.org/data.html.

[7] Davide Chicco. Siamese neural networks: An overview. Artificial Neural Networks, pages 73–94, 2021.

[8] Soumak Dey, Anjan Dutta, J Ignacio Toledo, Suman K Ghosh, Josep Lladós, and Umapada Pal. Signet: Convolutional siamese network for writer independent offline signature verification. arXiv preprint arXiv:1707.02131, 2017.
[9] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.