Gowri Dinesh Nair

Registration Number: 200968001

B.Tech Data Science and Engineering

Section B, Semester V

Deep Learning Project:
Cataract Identification
Phase 2: Literature Review

Regardless of whether the process is automated, cataract identification is primarily done based on color fundus photographs or slit-lamp images. The former are preferred for training AI models due to their simplicity, as training using slit-lamp images tends to be computationally expensive. And while automated cataract identification is still a growing field, it is dominated by two popular approaches: transfer learning and combinations of machine learning and deep learning. Deep learning models prove time and time again their unparalleled ability to extract features from image data regardless of the classifiers used, and a variety of such models are explored below.

One of the simplest yet most successful models was developed by Zhang et al.[1], who used a deep convolutional neural network to both detect and grade cataracts from color fundus photographs. CNNs are known for their highly accurate image classification, but come with the drawback of high training time due to the use of max pooling and large number of layers typically used. This particular model was trained on 4004 fundus images from the Beijing Tongren Eye Center's clinical database, and is similar to AlexNet architecture in the sense that it consists of 8 layers; 5 convolutional layers and 3 fully connected layers. Zhang et al. also studied the effect of using G-filters (green channel filters) to eliminate uneven illumination and the reflection of eyes in the images. The model was shown to perform better when the training images were preprocessed using the G-filter, and had a 94% accuracy for cataract detection and 87% accuracy in cataract grading.

As mentioned previously, research in this domain also consists of combinations of machine learning and deep learning approaches. This is discussed in depth by two noteworthy teams, the first of which Pratap and Kokil[2], who utilized a pre-trained CNN model (AlexNet) to extract features that were then fed into a binary SVM classifier. AlexNet revolutionized image classification due to having a greater speed than the traditional CNN, but leaves room for improvement in terms of learning features. And used alone, SVMs face difficulties with performance on data that is large in size or having high noise. But used together, these models have proven to exceed expectations. Trained on 400 images obtained from over ten datasets (including HRF image database, STARE, and DIARETDB0), the model grades images into four degrees of severity ("non cataracts", "mild", "moderate", and "severe") established through consultation with ophthalmologists. In preprocessing, here too the G-channels are extracted from the color fundus images. This better identifies minute details in the color photographs as well as reduces preprocessing time to 1/3rd of what it would be using the RGB images as is. Pratap and Kokil's model is one of the best-performing among all those examined in this paper, with a 100%

accuracy for cataract detection and 92% accuracy for cataract severity grading when tested on 400 novel images.

The second group to use this combined approach is Dong et al.[3], who experimented with the results of feature extraction using a wavelet feature extractor and a CNN model built using Caffe framework. These features were used to train and test both an SVM classifier and Softmax classifier, whose performances were compared. The results were analyzed, and the best performances were obtained from using a softmax classifier that was trained using features extracted by the CNN. This model was best for both cataract identification and grading (categorizing the images into "normal", "slight", "medium", or "severe") with accuracies of 94% and 90%, respectively.

Imran, A., Li, J., Pei, Y. *et al.*[4] examined and compared the results of several deep learning models for cataract identification, including GoogleNet, AlexNet, ResNet, VGGNet, and ensemble models consisting of all four. Before training and testing the models with the Tongren Eye Center's database of 8030 images, the images were preprocessed using baseline normalization, a separate function to remove the black space around the images, and non-local means denoising. Fundi were classified into "normal", "mild", "moderate", or "severe" using all the models. The ensemble consisting of all four baseline models and a bidirectional LSTM has the highest accuracy of around 97%, while the VGGNet model had the highest specificity of all the models at around 98%, despite being the slowest to train (due to the high number of parameters). These results are explained by ensembles' lower variance and bias, which provides better generalizability than the four simple baseline models, but comes with the cost of low interpretability. On the other hand, having greater depth than the standard CNN, VGGNet models are known for having higher accuracies, but are limited by their susceptibility to the vanishing gradient problem.

In pursuit of efficiency in addition to accuracy, Li et al.[5] developed both ResNet-18 and ResNet-50 models trained on data of 8030 fundus images (also from the Beijing Tongren Eye Center's database). The ResNet models were chosen for their relatively shorter training time as compared to deep convolutional neural networks. They were modified by the replacement of the final fully connected layer with global average pooling, and the addition of a convolutional layer with 1*1 filters. The ResNet-18 model, which was used to detect the presence of cataract due to its simplicity, used binary cross entropy as the loss function. The ResNet-50 model was used to categorize cataracts into either "non cataracts", "mild", "moderate", or "severe" (established by professional graders) and used categorical cross entropy. The team found that the ResNet-18 model yielded its greatest accuracy of 94% when trained using 90 epochs, while the ResNet-50 model performed with an accuracy of 87% after training for 80 epochs.

Comparatively, few research papers exist on the use of Inception models (excluding GoogleNet) in cataract identification from fundus photographs or slit-lamp images. They have, however, been used in the extraction of surgical phases, real-time cataract grading, and further understanding from video data [6][7]. Despite requiring more training data for good performance, Inception models are generally known for their relatively lower computational complexity and less training time (as compared to standard CNNs). This can be seen from one paper by A. Raza et al. [8] using an InceptionV4 model on 601 color fundus photographs from Kaggle's "cataract dataset" [9]. They attained an accuracy of 96% without image augmentation and 86% with image augmentation. Preprocessing here involved cropping, resizing, and contrast enhancement of the images followed by noise removal using an averaging filter.

It is evident from the above papers that the success and speed of a cataract identification model is heavily reliant on appropriate preprocessing and the use of transfer learning models. Although a variety of methods and models exist and have yet to be experimented on, deep CNNs, ResNet models, VGGNet models, and the softmax function are consistently used among the best-performing classifiers. Further in this project, I will be experimenting with at least three deep or transfer learning models to observe their classification abilities on the Kaggle "Ocular Disease Recognition" dataset[10]. Keeping a feed-forward CNN as the baseline model, the goal is to attain more insight into the architectures of these models by attempting to match or improve these results through my own implementations.

Gowri Dinesh Nair

Registration Number: 200968001

B.Tech Data Science and Engineering

Section B, Semester V

Citations:

1. Linglin Zhang *et al.*, "Automatic cataract detection and grading using Deep Convolutional Neural Network," *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, 2017, pp. 60-65, doi: 10.1109/ICNSC.2017.8000068. (https://ieeexplore.ieee.org/abstract/document/8000068?casa_token=mcfvW6XUftoAAAAA:ngRRgqqmW-gc8oWnuqkdQELp3CRFE852UtN76frLzoDJX-EG-wriG32GUiLiMtDcNWcE9sy2hw)

2. Turimerla Pratap, Priyanka Kokil, Computer-aided diagnosis of cataract using deep transfer learning, Biomedical Signal Processing and Control, Volume 53, 2019, 101533, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2019.04.010 . (https://www.sciencedirect.com/science/article/pii/S1746809419301077)

3. Y. Dong, Q. Zhang, Z. Qiao and J. -J. Yang, "Classification of cataract fundus image based on deep learning," *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, pp. 1-5, doi: 10.1109/IST.2017.8261463.(https://ieeexplore.ieee.org/abstract/document/8261463/authors#authors)

4. Imran, A., Li, J., Pei, Y. *et al.* Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. *Vis Comput* 37, 2407–2417 (2021). https://doi.org/10.1007/s00371-020-01994-3 . (https://link.springer.com/article/10.1007/s00371-020-01994-3#citeas)

5. J. Li et al., "Automatic Cataract Diagnosis by Image-Based Interpretability," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, pp. 3964-3969, doi: 10.1109/SMC.2018.00672. (https://ieeexplore.ieee.org/abstract/document/8616669?casa_token=PLQf6yPds-sAAAAA:Y9sUIGRViBgoijlATbYNzrt29rfXA7Muks6zikkJI2nZDr4oaU6reb5EymXsfsBfk9Lfw0vOqg)

6. Morita, S., Tabuchi, H., Masumoto, H. et al. Real-Time Extraction of Important Surgical Phases in Cataract Surgery Videos. Sci Rep 9, 16590 (2019). https://doi.org/10.1038/s41598-019-53091-8 (https://www.nature.com/articles/s41598-019-53091-8#citeas)

Gowri Dinesh Nair

Registration Number: 200968001

B.Tech Data Science and Engineering

Section B, Semester V

7. Hu, S., Luan, X., Wu, H. et al. ACCV: automatic classification algorithm of cataract video based on deep learning. BioMed Eng OnLine 20, 78 (2021). https://doi.org/10.1186/s12938-021-00906-3 (https://link.springer.com/article/10.1186/s12938-021-00906-3#citeas)

8. A. Raza, M. U. Khan, Z. Saeed, S. Samer, A. Mobeen and A. Samer, "Classification of Eye Diseases and Detection of Cataract using Digital Fundus Imaging (DFI) and Inception-V4 Deep Learning Model," 2021 International Conference on Frontiers of Information Technology (FIT), 2021, pp. 137-142, doi: 10.1109/FIT53504.2021.00034. (https://ieeexplore.ieee.org/abstract/document/9701378)

9. “cataract dataset”, 2019. (https://www.kaggle.com/datasets/jr2ngb/cataractdataset)

10. Peking University, Shanggong Medical Technology Co., Ltd., “Ocular Disease Recognition”, 2019. (https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k/metadata?datasetId=611716&sortBy=voteCount&searchQuery=inception)