Evaluation of Neural Networks in retinal disease classification from OCT images on mobile platforms

University of California Riverside

Professional Project Design

Faddy Sunna

MSOL Data Science
December 15, 2019

Summary

- Biomedical data (sensory, medical imaging) is being created faster than humans can interpret.
- Transforming global healthcare with Artificial Intelligence is dependent on being able to extract meaningful knowledge and insights from unstructured, unlabeled, and high dimensional biomedical data. Here I train a Convolutional Neural Network and evaluate its ability to classify medical images on a mobile platform.

Introduction – *CNN*

- Medical data requires an experts time and judgment to properly interpret.
- Artificial intelligence is the solution for the data surplus. The proven and preferred deep learning frameworks for image classification are based on Convolutional Neural Networks (CNN).
- CNN’s are a form of supervised deep learning, modeled after the visual cortex of a cat.
- Images using a layered architecture of neurons. The groups of neurons each detect a specific shape, and communicate with the other groups in order to form a full image & understanding.

Introduction - *CNN*

- CNN's where initially developed in the 1980's, and later popularized in the late 1990's.

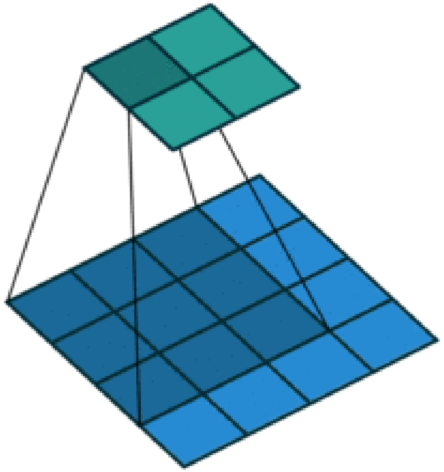
- LeCun co-authored a paper which introduces CNN's and how he applied them to recognize digits on bank checks (LeCun et al., 1998).

- This was the most successful implementation of CNN's at the time.

Convolution Layer (1st layer)

- *Convolution layer* - In this layer, features are extracted from the input image. The input image is scanned by a sliding window that reads a portion of the image at time as RGB vectors for each pixel.

- Calculations are performed on these values to identify features such as borders or specific shapes. The depth of features identified depends on the number of filters used in the CNN. With more filters,more features can be identified, allowing for a more fully connected network.



Non-linearity layer (2nd layer)

- This layer takes values and reduces them to a number between 0 and 1.
- Large negative numbers become 0, and large positive numbers become 1, creating a piecewise activation function.
- The most commonly used activation function is Rectified Linear unit(ReLU).

Pooling Layer (3rd layer)

- Used to reduce the dimensionality of the feature map. Reducing noise and computation requirements.

- CNN’s use max pooling, which creates a spatial neighborhood in the feature map, and selects the largest value from that group. Separating the most prominent features.

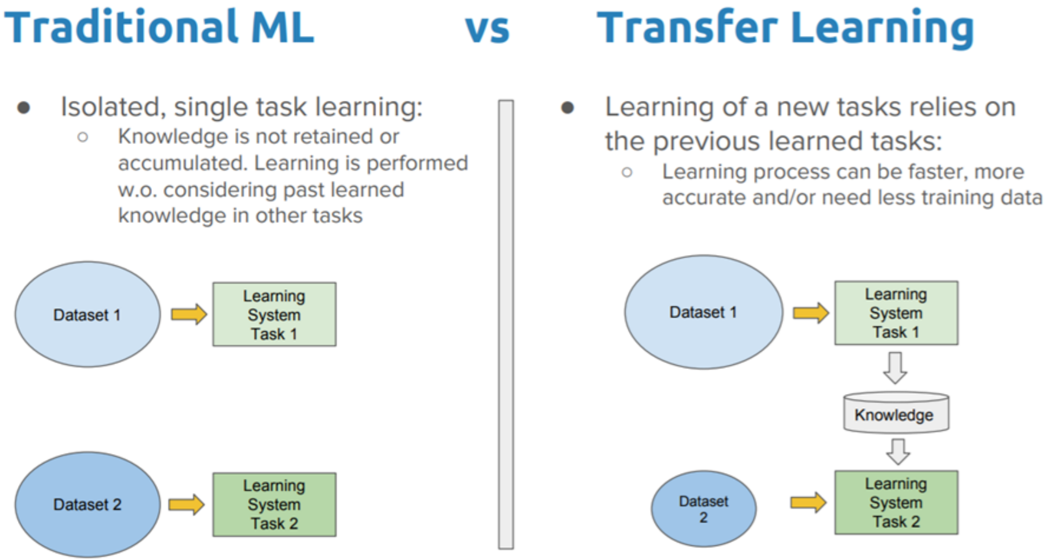
Fully Connect Layer (4th layer)

- The fully connected layer completes the CNN framework. The outputs from the Pooling and Convolutional layers represent the features of the input image.

- These outputs are passed into the fully connected layer, where it learns non-linear combinations of the features and performs classification of the image.

Transfer Learning

- Utilize pre-trained weights and networks
- Reduce computational cost
- Reduce dataset size requirements



Project Plan

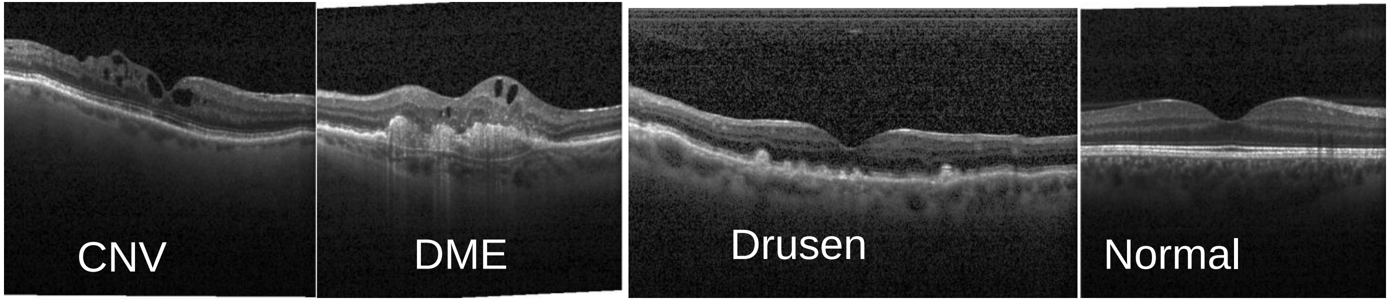
Dataset type: OCT Images - 3 retinal diseases

1. (CNV) -Choroidal neovascularization
2. (DME) -Diabetic Macular Edema.
3. (Drusen)

1. Train traditional CNN → Evaluate
2. Train CNN with Transfer learning → Evaluate
3. Convert to mobile formats (Tensorflow lite and Tensorflow lite quantized)
4. Evaluate / Benchmark on mobile device

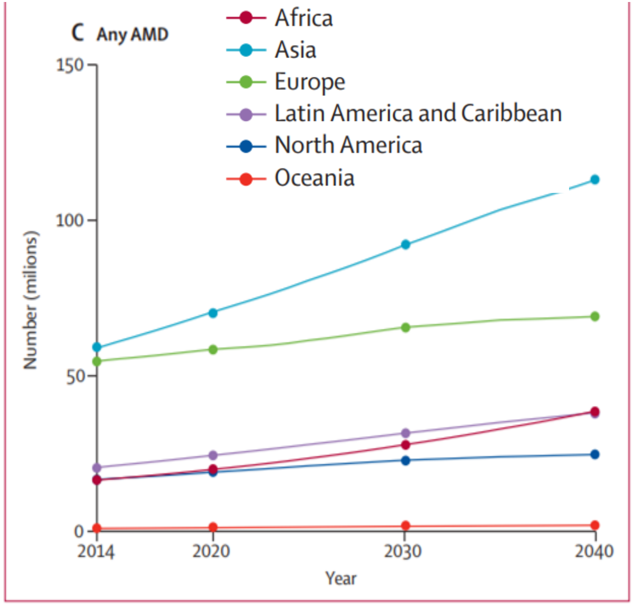
Project significance – OCT images

- One study conducted on mice/rats shows “High-resolution spectral-domain OCT provides unprecedented high-quality 2D and 3D in vivo visualization of retinal structures “(Ruggeri et al,. 2007).
- Opthamologists rely heavily on OCT for the diagnosis of many eye related diseases. It is one of the most common ophthalmic diagnosis techniques with approximately 30 million scans performed each year globally (Swanson & Fujimoto, 2017).



Project significance – Retinal Disease

- The leading cause of blindness in the United States is Age Related Macular Degeneration (AMD), and with the growing aging population, the amount of people affected by AMD is rapidly growing (Bressler et al, 2004).
- Approximately 170 million people affected by some form of AMD globally, and that number is expected to increase to 258 million by 2040, with Asia leading by a significant margin (Wong et al, 2014).
- *Treatment:* Anti–vascular endothelial growth factor therapy (Anti-VEGF)

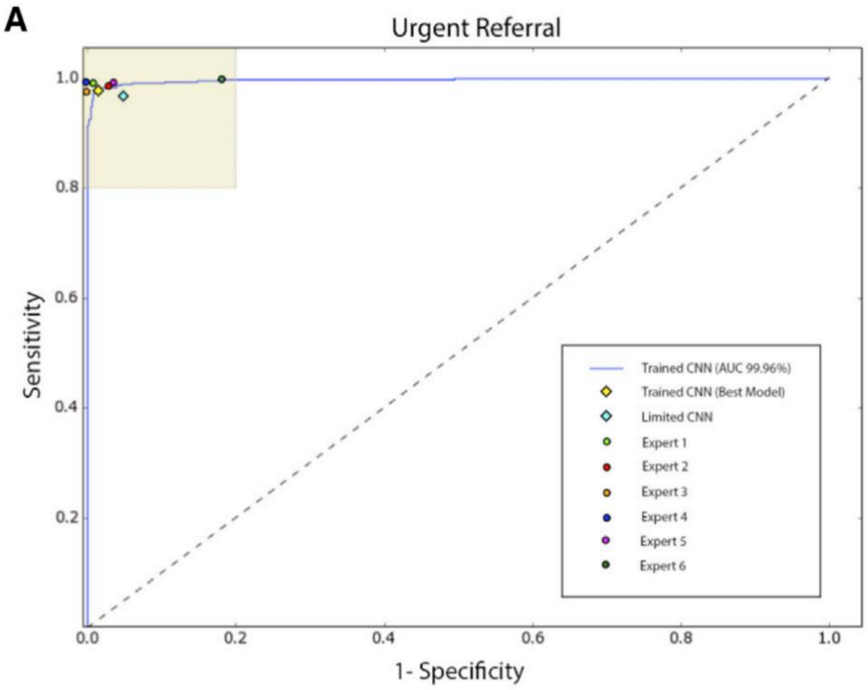


Literary Review (Kermany et al., 2018)

- In 2018 a combined research group implemented a CNN to classify CNV, DME, and Drusen in OCT images.They used transfer learning, fine tuning the InceptionV3 model to fit their dataset of 103,802 training images.

- “Six experts with significant clinical experience in an academic ophthalmology centers were instructed to make a referral decision on each test patient using only the patient’s OCT images” (Kermany et al., 2018).

Literary Review (Kermany et al., 2018)



Literary Review (Fauw et al., 2018)

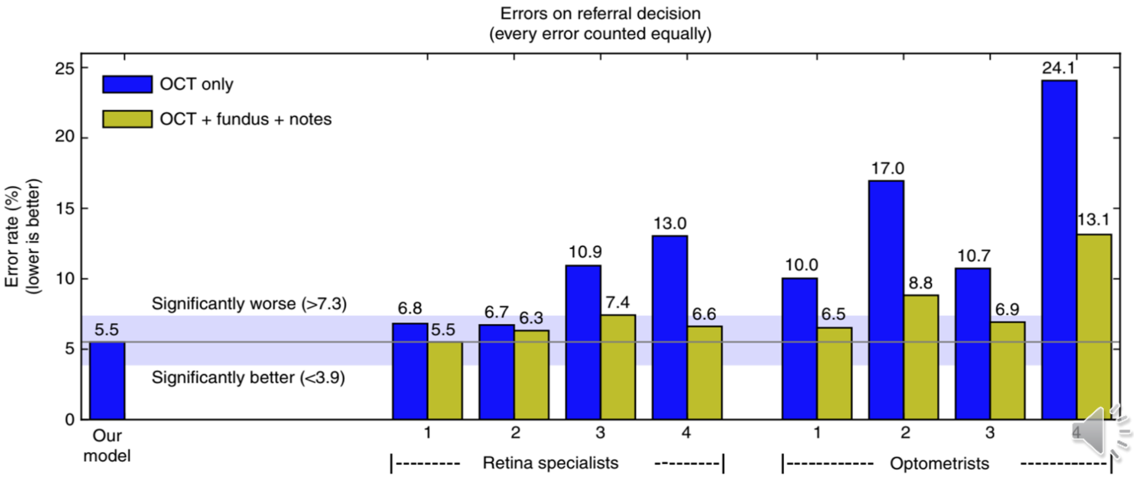
- A study completed in 2018 at the University College London in the UK used the U-Net image segmentation architecture to segment the raw OCT images into 3-D tissue maps of 15 different classes such as anatomy, physiology, and image artifacts.

- Classification was then performed on these 3-D tissue maps.

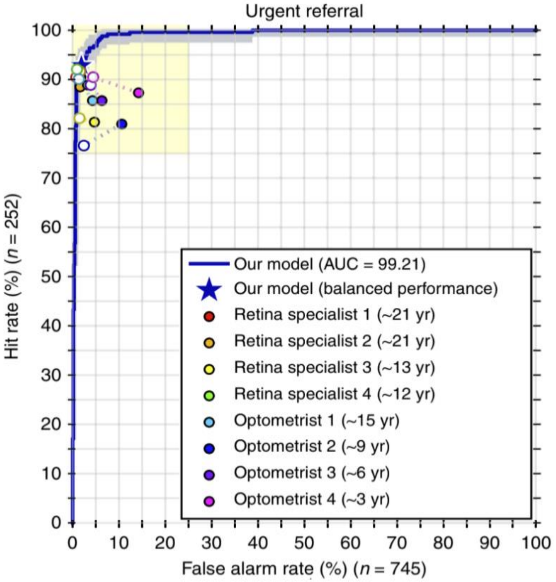
- Their results were compared with a group of retina specialists, and ophthalmologists.

Literary Review (Fauw et al., 2018)

c



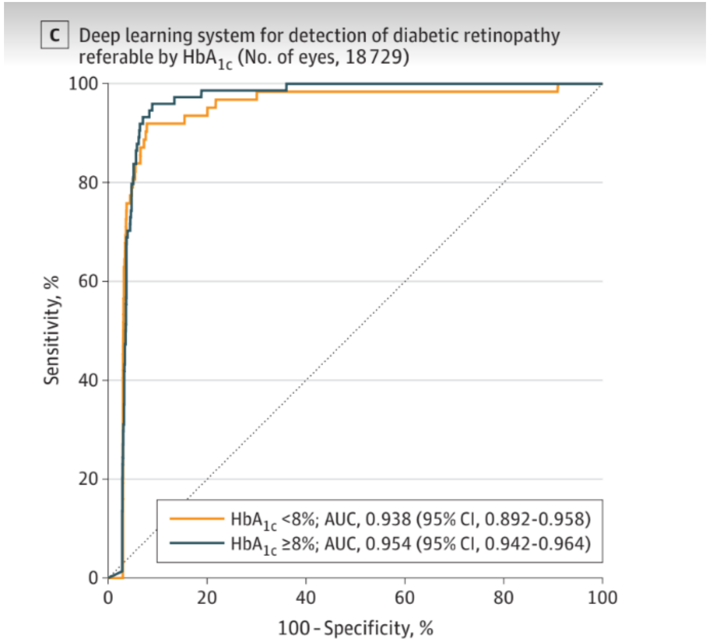
a



Literary Review (Shu Wei Ting et al., 2017)

- A collective research group in 2017, was also able to build a deep learning system for classifying AMD from OCT images.

- They achieved scores comparable to professional graders.



Literary Review (Kim et al., 2018)

- “The performance of the system has been characterized by measuring optical parameters such as the power throughput, and lateral resolution. These compared favorably to currently available commercial OCT systems” (Kim et al., 2018)

- 7 lb vs 30-50lb
- $7200 vs 30-50k

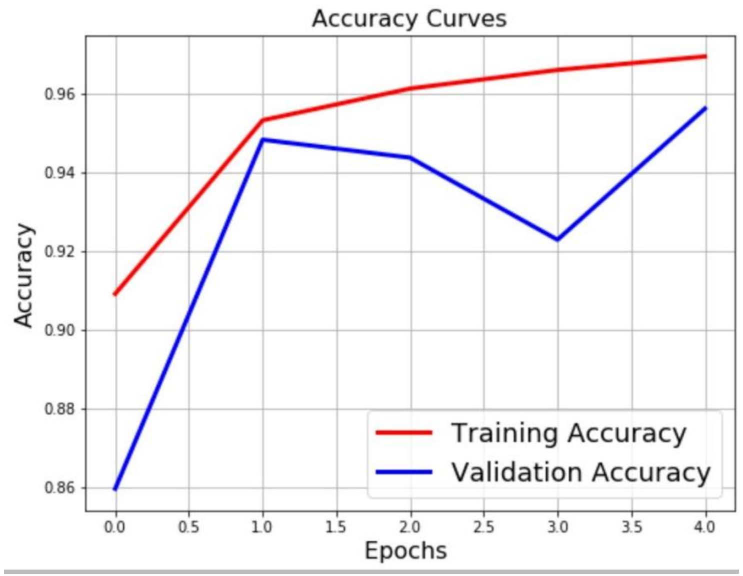
Dataset

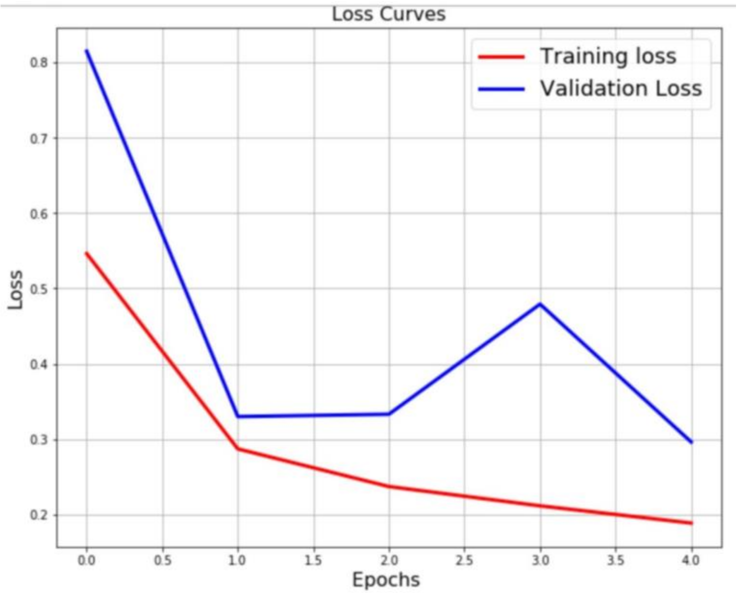
- Dataset size: (Original images are 512x496 JPEG)
- Training images: 75,534
- Validation images: 3,136
- Testing images: 5,519

Methodology

- Model A: Transfer learning of MobilenetV3
  - Convert to Tensorflow Lite
  - Convert to Tensorflow lite Quantized
- Model B: Custom 16 layer CNN
  - Convert to Tensorflow Lite
  - Convert to Tensorflow lite Quantized

Results – Custom CNN training





Results – Model Evaluation

Model	Accuracy%	Precision%	Recall%	AUC%
Custom CNN	96.82	93.96	93.28	98.59
MobileNet	96.40	93.29	92.24	98.85

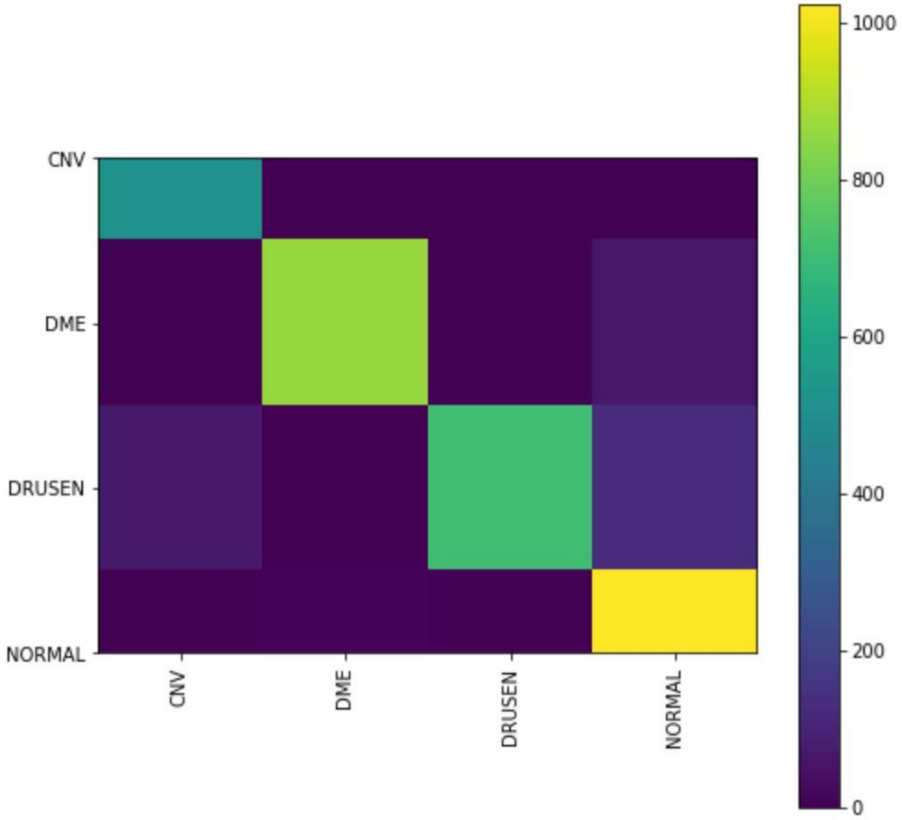
```
              precision    recall  f1-score   support

         CNV       0.83      1.00      0.91      1248
         DME       0.99      0.91      0.95      1648
      DRUSEN       0.97      0.81      0.88       939
      NORMAL       0.95      0.97      0.96      1684

    accuracy                           0.93      5519
   macro avg       0.94      0.92      0.93      5519
weighted avg       0.94      0.93      0.93      5519
```

```
Confusion Matrix
[[1243    2    3    0]
 [  74 1505    8   61]
 [ 153    2  762   22]
 [  25    9   11 1639]]
Classification Report
```

Results- Custom CNN Test images



Results

	CNV			DME			DRUSEN			NORMAL		
	TF Model	TFLite	TFLite quantized	TF Model	TFLite	TFLite quantized	TF Model	TFLite	TFLite quantized	TF Model	TFLite	TFLite quantized
0	0.95829248	0.95829248	0.95960754	2.5699352e-05	2.5699426e-05	2.0680596e-05	0.041437149	0.04143713	0.040186416	0.00024467986	0.0002446805	0.00018533628
1	0.00019736317	0.00019736358	0.00018975955	0.00018287783	0.00018287837	0.0001478895	0.0011347258	0.0011347259	0.0011862472	0.99848503	0.99848503	0.99847609
2	0.0014009299	0.0014009286	0.0016324027	0.99855763	0.99855775	0.99832326	2.7641916e-05	2.7641945e-05	3.0202847e-05	1.3747087e-05	1.3747142e-05	1.4070794e-05
3	0.99920851	0.99920851	0.99905914	1.6094945e-06	1.6095008e-06	9.4361684e-07	0.00078116223	0.00078116113	0.00093453005	8.6886885e-06	8.6887067e-06	5.3236795e-06
4	0.022990908	0.022990948	0.016506733	0.00021810408	0.00021810553	0.0001886825	0.97676796	0.97676796	0.98328352	2.3094484e-05	2.3094795e-05	2.1071799e-05
5	0.0033075109	0.0033075202	0.0027783259	0.00019818443	0.00019818501	0.00016738854	0.99585479	0.99585479	0.99649423	0.00063956575	0.00063956727	0.00056015415
6	0.0033777673	0.0033777754	0.0034889809	0.99559659	0.99559659	0.99551433	0.00053718389	0.00053718465	0.00054585823	0.00048847543	0.00048847729	0.00045069522
7	0.001673167	0.0016731726	0.0017431362	0.99810141	0.99810141	0.99798977	7.0556416e-05	7.0556482e-05	8.5697655e-05	0.00015495456	0.00015495412	0.00018148043
8	0.01233921	0.012339187	0.014930082	0.98659432	0.98659432	0.98342574	0.00069185154	0.00069185079	0.0010823043	0.00037462518	0.00037462698	0.00056193455
9	0.011657038	0.011657049	0.011138487	0.00054709241	0.00054709445	0.00047843347	0.0032421879	0.0032421912	0.0030108478	0.98455369	0.98455369	0.98537225
10	0.98546356	0.9854635	0.98576152	9.5356576e-05	[illegible]	6.3001789e-05	0.01435851	0.014358523	0.014121768	8.2532184e-05	8.2532257e-05	5.3723157e-05
11	0.0027921302	0.0027921409	0.0023168847	0.0046558655	[illegible]	0.0032814941	0.014595116	0.014595116	0.012751363	0.97795689	0.97795689	0.98165029
12	0.99941111	0.99941111	0.99953079	5.6526517e-07	5.6526306e-07	4.1211652e-07	0.000582685	0.00058268529	0.00046486018	5.6441813e-06	5.6441768e-06	4.0339132e-06
13	0.1725966	0.1725966	0.17971835	0.5836243	0.58362514	0.5957821	0.053985845	0.053985521	0.054170053	0.18979321	0.18979274	0.17032948
14	0.0031488559	0.0031488661	0.0025683267	5.210692e-05	5.210716e-05	4.1775733e-05	0.99677545	0.99677533	0.99737108	2.3614921e-05	2.3615052e-05	1.8827517e-05
15	0.00063893909	0.00063894247	0.00064096635	0.00083200872	0.0008320167	0.00077276304	0.00054069463	0.00054069667	0.0005357819	0.99798834	0.99798834	0.99805045
16	0.081420861	0.081420995	0.066421144	9.0702837e-05	9.0703521e-05	8.5200278e-05	0.91847587	0.91847575	0.93348259	1.2545355e-05	1.2545474e-05	1.09899e-05
17	0.83797365	0.83797365	0.86012709	0.00039845932	0.00039846206	0.00027914744	0.16082785	0.16082783	0.13890609	0.00080006482	0.00080007024	0.00068770145
18	0.035589881	0.035589889	0.040328681	0.00107837	0.0010783722	0.0010506456	0.0068004848	0.0068004848	0.007192459	0.95653129	0.95653129	0.95142817
19	0.15441976	0.15441987	0.17031507	0.00046638885	0.00046638885	0.00039314304	0.0039977245	0.0039977301	0.0041169068	0.84111607	0.84111601	0.82517487
20	0.014396656	0.014396684	0.013302542	0.0002057205	0.00020572128	0.00019691931	0.98527193	0.98527193	0.98635745	0.00012566427	0.00012566475	0.00014304191
21	0.0027355866	0.0027355962	0.0025495342	0.04292801	0.04292826	0.030062923	0.002817913	0.0028179202	0.0027966606	0.95151848	0.95151818	0.96459091
22	0.99338853	0.99338853	0.99538028	0.0019272936	0.0019273056	0.0010964894	0.0027597875	0.002759794	0.0022945614	0.0019243357	0.0019243412	0.0012286283
23	0.33719698	0.33719695	0.4058485	0.63943833	0.63943833	0.56897491	0.0028870448	0.0028870392	0.0029146378	0.020477701	0.020477684	0.022261931
24	0.0017471978	0.0017472011	0.0015120971	0.0024387464	0.0024387564	0.0016792814	0.0023520729	0.0023520726	0.002096463	0.99346209	0.99346197	0.99471223
25	0.078944601	0.078944623	0.078812785	0.9200232	0.9200232	0.92025435	0.00043528687	0.00043528582	0.00040398	0.0005968351	0.00059683569	0.00052897818
26	0.062707044	0.062706873	0.057463381	0.11483739	0.11483753	0.094144739	0.003842952	0.0038429562	0.0039418205	0.81861258	0.81861264	0.84445006

Results – mobile benchmark

Model format	Average inference time (ms)
Tensorflow lite	186
Tensorflow lite quantized	110

Conclusion

- I was able to train a CNN classifier that performs comparably with the studies conducted by (Kermany et al., 2018) and (Fauw et al., 2018) that were expert validated.
- The Lite un-quantized model performed similarly to the original Tensorflow model. While the quantized model had a shown variations in the confidence of its predictions.
- The Tensorflow lite model performed very well and showed little degradation from the original model. This promising for the future of AI, because it opens the door for devices such a light and mobile OCT imaging system than can also diagnose a patient on the spot.

references

1. Ergun, Erdem, et al. “Incidence of Patients Presenting with Exudative Maculopathy and Neovascular Retinal Disease in an Urban Population.” *Wiener Klinische Wochenschrift*, vol. 116, no. 21-22, 2004, pp. 737–743., doi:10.1007/s00508-004-0262-2.
2. Fauw, Jeffrey De, et al. “Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease.” *Nature Medicine*, vol. 24, no. 9, 2018, pp. 1342–1350., doi:10.1038/s41591-018-0107-6.
3. Ignatov, Andrey, et al. “AI Benchmark: Running Deep Neural Networks on Android Smartphones.” *Lecture Notes in Computer Science Computer Vision – ECCV 2018 Workshops*, 2019, pp. 288–314., doi:10.1007/978-3-030-11021-5_19.
4. Jaffe, Glenn J., and Joseph Caprioli. “Optical Coherence Tomography to Detect and Manage Retinal Disease and Glaucoma.” *American Journal of Ophthalmology*, vol. 137, no. 1, 2004, pp. 156–169., doi:10.1016/s0002-9394(03)00792-x.
5. Kaiser, P. K., and D. V. Do. “Ranibizumab for the Treatment of Neovascular AMD.” *International Journal of Clinical Practice*, vol. 61, no. 3, 2007, pp. 501–509., doi:10.1111/j.1742-1241.2007.01299.x.

references

6. Kermany, Daniel s, et al. “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning.” *Cell*, vol. 172, no. 1122, ser. 1131, 2018. *1131*.
7. Kourtis, Lampros C., et al. “Digital Biomarkers for Alzheimer’s Disease: the Mobile/Wearable Devices Opportunity.” *Npj Digital Medicine*, vol. 2, no. 1, 2019, doi:10.1038/s41746-019-0084-2.
8. Krizhevsky, Alex, et al. “ImageNet Classification with Deep Convolutional Neural Networks.” *Communications of the ACM*, vol. 60, no. 6, 2017, pp. 84–90., doi:10.1145/3065386.
9. Lecun, Y., et al. “Gradient-Based Learning Applied to Document Recognition.” *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324., doi:10.1109/5.726791.
10. Lee, Susan S., and Michael R. Robinson. “Novel Drug Delivery Systems for Retinal Diseases.” *Ophthalmic Research*, vol. 41, no. 3, 2009, pp. 124–135., doi:10.1159/000209665.

references

10. Miotto, Riccardo, et al. “Deep Learning for Healthcare: Review, Opportunities and Challenges.” *Briefings in Bioinformatics*, vol. 19, no. 6, 2017, pp. 1236–1246., doi:10.1093/bib/bbx044.Piau, Antoine, et al.
11. “Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review (Preprint).” 2018, doi:10.2196/preprints.12785.
12. Song, Ge, et al. “Design and Implementation of a Low-Cost, Portable OCT System.” *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, 2018, doi:10.1364/ots.2018.of3d.7.
13. Ting, Daniel Shu, et al. “Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images from Multiethnic Populations with Diabetes.” *American Medical Association* , vol. 318, no. 22, ser. 2211-2223, 2017, pp. 1–13. *2211-2223*, doi:10.1001/jama.2017.18152.
14. Wong, Wan Ling, et al. “Global Prevalence of Age-Related Macular Degeneration and Disease Burden Projection for 2020 and 2040: a Systematic Review and Meta-Analysis.” *The Lancet Global Health*, vol. 2, no. 2, 2014, doi:10.1016/s2214-