Detection of Cataract Using Deep Learning Models

Zorawar Jaiswal
Department of Computer Science
California State University, Northridge
Northridge, CA, USA
zorawar.jaiswal.556@my.csun.edu

Abhishek Verma
Department of Computer Science
California State University, Northridge
Northridge, CA, USA
abhishek.verma@csun.edu

***Abstract*— Cataract refers to the opacity of the eye lens that deteriorates vision, leading to either partial or even complete visual impairment, and a shallower quality of life. Early detection of Cataract may result in mitigating crucial surgeries and easier tracking of the cataract stage. This paper compares two state-of-the-art models: EfficientNet B7 and MobileNet V2, to classify images as normal or cataract. The study is the first to utilize and set a baseline performance on the novel dataset created by Azri. The images varied greatly in terms of size, aspect ratio, and average intensity, increasing the need for stricter preprocessing. Both models exceed performance standards on general deep learning problems and achieve up to 100% accuracy, opening the door to uncharted avenues in mitigating cataract-related ailments on low-resource end devices.**

Keywords*— *Cataract Detection, Retina, Image Classification, Deep Learning, CNN, EfficientNet, MobileNet

I. Introduction

Cataract refers to the opacity of the eye lens that leads to deteriorating vision [23]. In the year 1948, Chace discussed and reasoned about age-related cataract with various conjectures, such as diet, blood sugar changes, and hormonal imbalances, causing the above condition. Medical research and experimentation over the years have yielded insights into the probable cause of age-related cataracts.

Allen & Vasavada explained that with age, new fibers of the lens grow without the older ones being replaced. Over the years, a brownish-yellow tinted pigment accumulates in the lens, reducing the transmission of light. They also explain that malnutrition, diabetes, smoking, and dehydrating diseases acted as extrinsic factors to aggravate cataract formation in people throughout the developing and developed nations [24].

The causes for cataract in simple terms could be classified into (i) age-related, (ii) congenital or present since birth, (iii) due to trauma in the region, or (iv) due to diseases such as diabetes [23]. The medical field, however, prefers to classify them with the region of the lens impacted – (i) Nuclear cataract, where the contrast and color intensity are slightly reduced, leading to difficulty in day-to-day activities such as recognizing faces from afar, (ii) Cortical, where the light scatters from specific locations leading to issues like nighttime driving due to glare, and (iii) Subcapsular Cataract that disables quality vision in well-lit conditions impacting daytime driving and reading [23, 25, 26].

Since Cataract restricts the amount of light reaching the lens, leading to poor vision, and it is a common condition that occurs with age in most societies, the early detection of cataract is vital to mitigating costly surgeries and poor quality of life [24, 25]. This sets the background for our research.

The advancements of deep learning in the medical field have significantly contributed to supplementing interpretation for medical professionals, ranging from identifying nuclei in histopathology to detecting breast cancer in mammograms [27]. The advent and application of Convolutional Neural Networks (CNN) [28] to Deep Belief Networks [27] have opened avenues of application for unsupervised learning to assist medical professionals in decision-making and medical analysis.

The rest of this paper describes in section II the related works on deep learning, cataract, and medical imaging in general, followed by the Dataset Description in section III, the Methodology in section IV, Experiment Setup in section V, Results and Discussion in section VI, and Conclusion.

II. Related Works

A. Deep Learning for Eye Images

A plethora of problems are addressed using deep learning on retina and eye-based datasets. The most common application of deep learning with eye-related datasets besides medical diagnosis are pupil and blinking detection. The application, however, varies since some research tasks utilize it to avoid accidents while driving [4-6], while others perceive it as most important for Augmented and Virtual Reality systems and gaming [5]. Novel work also includes webcam-based eye tracking with multiple classes predicted, such as free viewing, fixation, or smooth pursuit [3]. Navaneethan et al. [7] highlights the importance of pupil detection for the calibration of medical equipment like tonometers and keratometers.

Ahmed et al. [1] correlated the eye movement of children with Autism Spectrum Disorders (ASD) and utilized Deep Learning techniques to decipher sequences of vision projection points in children to detect ASD. They implemented state-of-the-art Convolutional Neural Network (CNN) algorithms – GoogleNet and ResNet 18, with Support Vector Machine (SVM) up to 99% accuracy.

Sun et al. [2] utilized SVM, K Nearest Neighbors (KNN), ResNet 18, and VGG 16 networks, amongst other models, to diagnose Alzheimer's Disease (AD) in eye movement-tracked

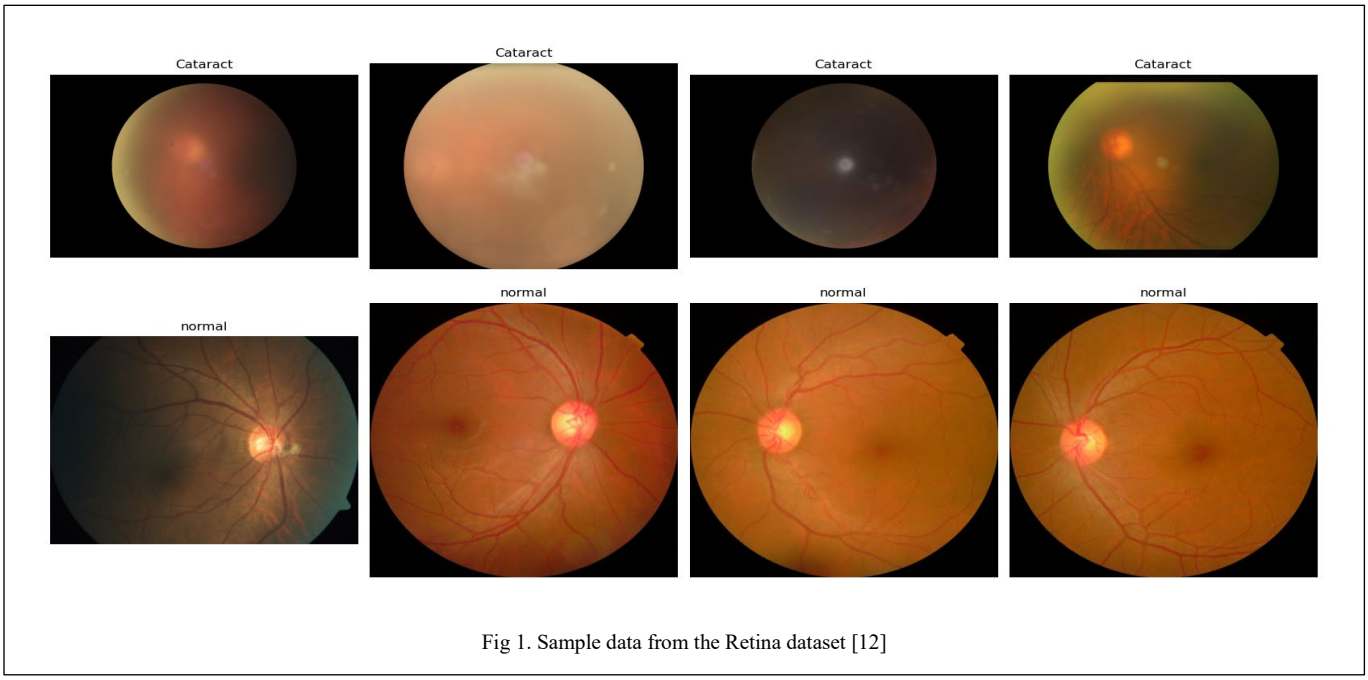

Fig 1. Sample data from the Retina dataset [12]

data and achieved an accuracy of 85%. Similarly, El Hmimdi et al. [8] predict dyslexia from images involving different tests like reading tasks, vergence and saccade.

Deep Learning on eye images for medical diagnosis is another vastly researched topic. Research problems include tumor detection such as melanoma, freckles, nevus, and melanocytoma [10]. The Deep Learning techniques generally used for experimentation are image processing (for segmentation) and CNNs [10].

Sarki et al. [9] discuss the trends in diabetic eye disease detection based on fundus images. The trend of disease detection ranges from diabetic retinopathy to glaucoma, and cataract, with Messidor being one of the most researched datasets. They emphasize the usage of the green channel in images for superior contrast (higher luminosity) [9, 11], and image segmentation techniques to minimize the impact of overdrawn borders. For cataract detection, transfer learning-based techniques were able to achieve slightly under 93% while CNN-based techniques achieved a varying range of 81-94% accuracy in general.

B. Deep Learning on Medical Images

While research solves the invention part of a solution to a problem, it is still incomplete until the solution reaches the masses. Keeping this in mind, MobileNet [16] is often utilized to deploy the solution model for a Deep Learning problem to the end-user embedded or edge devices for ease of access. Ly, Verma, and Bein, in their work on Skin Cancer recognition, deployed the end model on iOS and Android edge devices for the masses [29]. Similarly, Jiménez-Gaona et al. went the extra

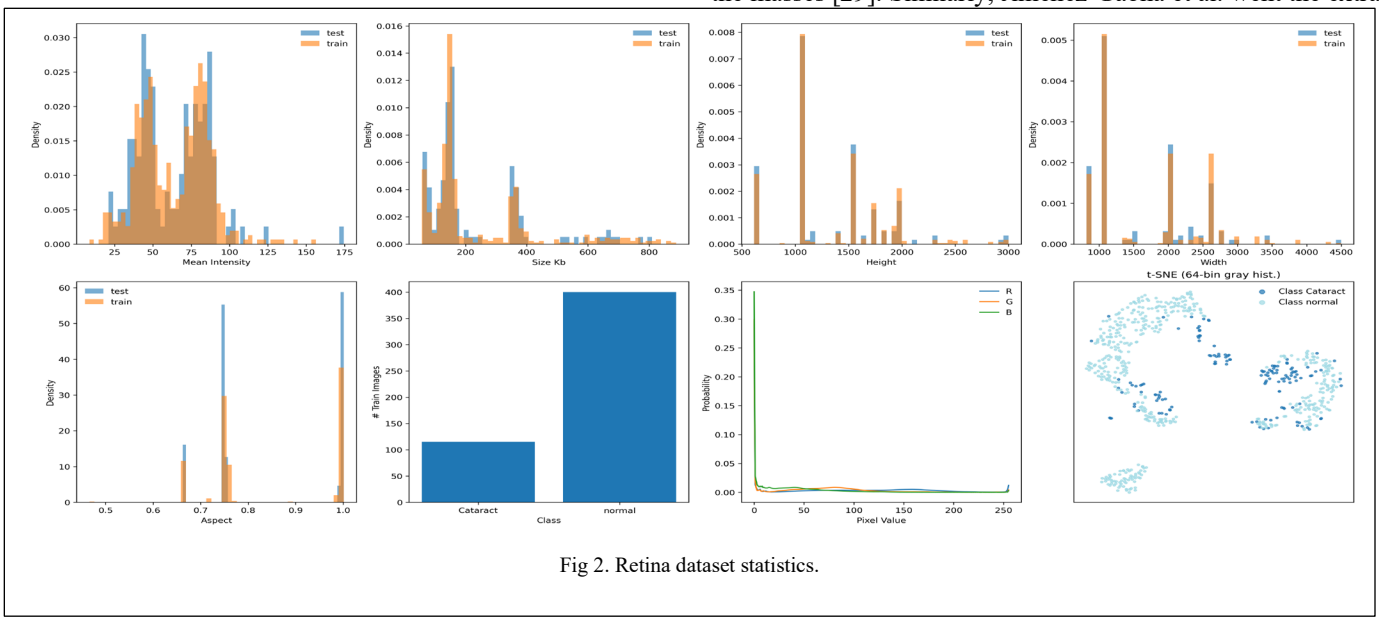

Fig 2. Retina dataset statistics.

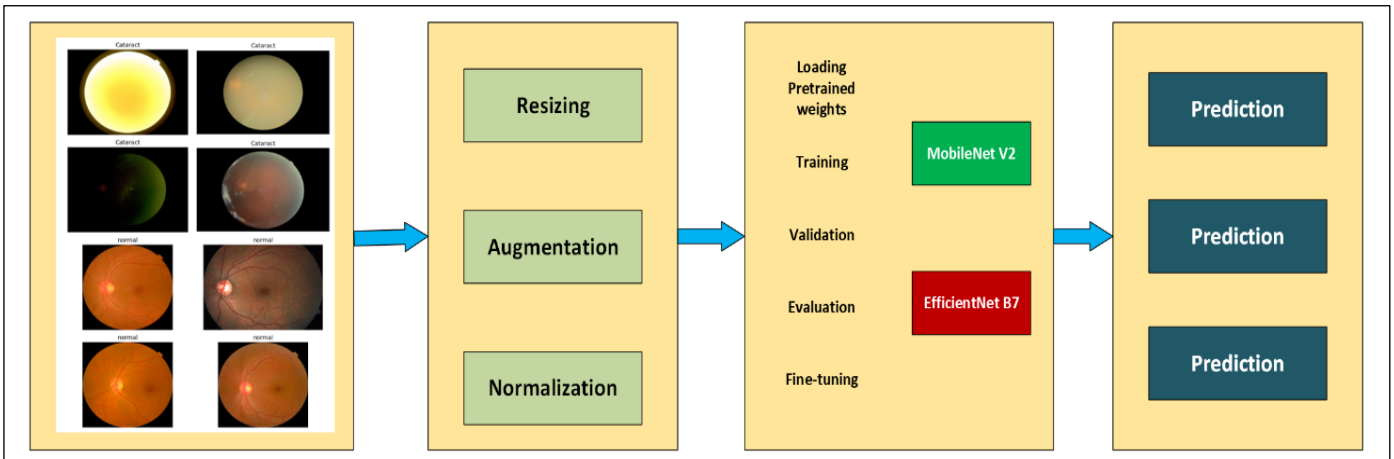

Fig 3. System pipeline for the research methodology.

mile and deployed their React application to accept mammographic images for classification into Malignant and Benignant classes, promoting public access [30].

III. Dataset Description

Azri [12] released a collection of datasets ranging from Brain MRI, Chest CT Scans, Gastrointestinal Tract diseases, Lung Radiography, and a Retina dataset. This research focuses on their Retina dataset for cataract detection. The original Retina dataset consisted of various diseases such as Glaucoma, Macular Degeneration, Diabetic Retinopathy, and Cataract.

The final subset of the dataset selected for experimentation consisted of two classes: normal and cataract. The two classes were imbalanced, with the normal class containing 500 images in total and the cataract class containing 143 images. Fig 1 shows sample images from the Retina dataset.

Fig 2 gives a brief overview of the exploratory data analysis performed on the Retina dataset. The mean grayscale intensity of the images throughout the dataset shows a right-skewed trend with the intensity ranging from 0 to 255. The Y-axis showing density refers to the probability density. The height and width plots represent the number of pixels on the X-axis and highlight the diversity of the dataset in terms of image size, strengthening the need for resizing of images and a center crop discussed in the next section.

The second last plot represents the color channel distribution in terms of the mean RGB spectrum, with blue being the most densely present throughout the data. The final t-SNE plot is derived from the mean grayscale intensity histogram by applying PCA. Each color on the t-SNE represents a class, and mixed clusters of different colors indicate the need for models to extract complex features for higher accuracy.

IV. Research Methodology

The dataset was split into train and test sets. The train/test split was set to 80/20. With the 80/20 split, training consisted of 400 normal and 115 cataract pupil images, while the test consisted of 100 normal and 28 cataract images. The next section discusses the preprocessing steps applied in detail.

The size and aspect ratio of images from the Retina dataset [12] varied in both inter and intraclass. The size ranged from 60 to 800 Kb for the normal images and from 60 to under 700 Kb for cataract images. Preprocessing was separately applied to train and test images.

The test images were (i) resized on the shorter side to 256 pixels, (ii) they were then center-cropped to 224 pixels, (iii) converted to tensors, and finally (iv) normalized on the mean and the standard deviation. The train images were subject to similar preprocessing steps with the addition of random horizontal flipping after the center cropping and before converting the image to tensors. Fig 3 shows the preprocessing steps in brief. The training images were subjected to random horizontal flipping to increase the robustness of the system against overfitting. The resizing of the shorter side to 256 pixels ensures that minimum information is lost in translating the image for appropriate feature extraction. Fig 3 gives a detailed view of the entire system pipeline.

A. Deep Learning Models Used (EfficientNet B7, MobileNet)

EfficientNet B7 was the highest scaled model from Tan and Le's published work [13]. They based their work on the idea that balancing out the network depth (count of layers), the width (count of channels), and the resolution in a Convolutional Neural Network may lead to a better performance. Scaling networks up was vital for increased accuracy with the increase in availability of data [13]. They explain that an increased resolution would require more layers (depth) and channels to capture better patterns and devise a grid search to optimize and find the adjusted coefficients based on (1).

$$Width^2 * Resolution^2 * Depth \approx 2 \qquad (1)$$

$$Such\ that\ Width \geq 1,\ Resolution \geq 1,\ \&\ Depth \geq 1$$

The disk size of EfficientNet B7 was about 245 MB and the number of parameters were approximately 64 million. The EfficientNet B7 on the ImageNet dataset [14] was approximately 8 times smaller than the Gpipe architecture [15] in terms of parameters, yielding similar performance.

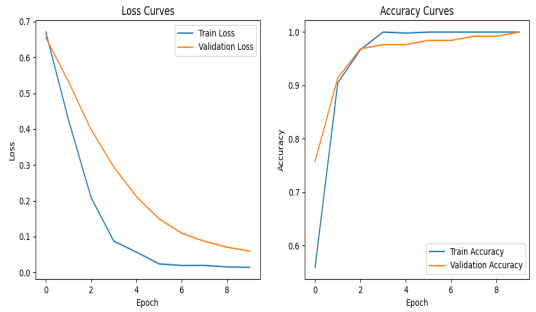

Fig 4. The loss and accuracy curve for EfficientNet B7

The original MobileNet architecture [16] was aimed at designing Neural Networks for embedded and mobile vision applications. They introduced the concept of depth-wise separable convolutions to reduce the computational load and the total parameters. Traditionally, convolutions involved applying a filter to all channels of the image simultaneously, but depth-wise convolutions changed that to applying one filter for each channel, leading to lower parameters. The idea of MobileNet was further improved by Sandler et al. by adapting inverted residual blocks to slowly increase the channels (narrow to wide), perform depth-wise convolution, and finally reduce channels (narrowing down) hence the term – inverted residual blocks [17].

Our research study utilized the MobileNet V2 architecture [17] trained on the ImageNet V2 weights.

V. Experiment Setup

A. Model training and Evaluation metrics

Model training was done on a Linux server with A6000 GPU having 48 GB VRAM. The training of both these models was done with fully unfrozen layers, 10 epochs each, with a batch size of 80. The training time for EfficientNet B7 was under 280 seconds with 64 million parameters, and MobileNet followed the trend with under 220 seconds for training with just 2.23 million parameters. Table 1 summarizes the details of both models.

Evaluation metrics for performance included accuracy and loss curves, the t-SNE plot, confusion matrix derived from the

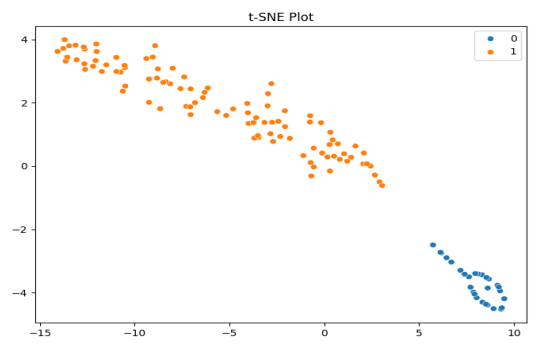

Fig 5. The t-SNE plot for EfficientNet B7

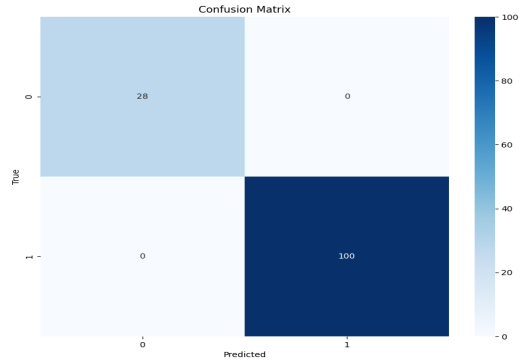

Fig 6. The the confusion matrix for EfficientNet B7

True Positive, False Negative, True Negative, and False Positive count, and the class wise accuracy report. The experiment was limited to 10 epochs due to early convergence of the models.

B. Software Setup

The programming code was written in Python 3.10. Jupyter Notebooks [18] served as the integrated development environment (IDE). Pytorch [19] was selected as the primary Deep Learning Framework for loading data, computing the tensors, and establishing and manipulating the model parameters. Matplotlib [20] was utilized to primarily visualize the output, including but not limited to the display of selectively sampled image data and the visualization of trends – loss and accuracy curves, and the class-wise accuracy. Seaborn [21] was the secondary tool utilized to visualize and contrast the t-SNE plot and the confusion matrix of various models. Numpy [22] was highly useful to manipulate arrays, compute calculations such as accuracy from model predictions, and to temporarily store the feature vectors.

VI. Results and Discussion

Figures 4-9. Show the experiment results. Both EfficientNet B7 and MobileNet V2 were able to capture the overall essence of the dataset and scored nearly perfect accuracy, with EfficientNet B7 reaching a perfect 100% and MobileNet V2

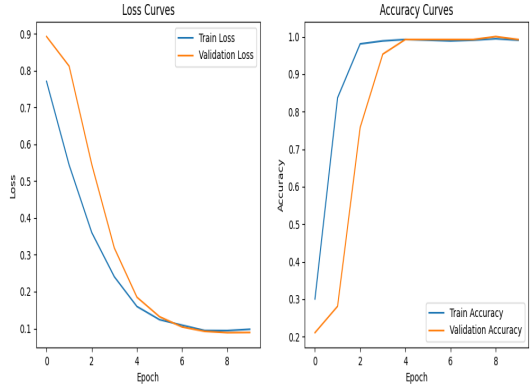

Fig 7. The loss and accuracy curve for MobileNet V2

TABLE I. TRAINING DETAILS FOR THE RETINA DATASET, 80/20 SPLIT FOR 10 EPOCHS

Model	GPU Memory Used	Batch Size	Total Training Time	Best Epoch	Disk Space	No. of Parameters	Training Accuracy (%)	Validation Accuracy (%)
EfficientNet B7	44.4/49 GB	80	278s	10	245 MB	63 M	100	100
MobileNet V2	1.8/49 GB	10	219s	5	26 MB	2.23 M	100	99

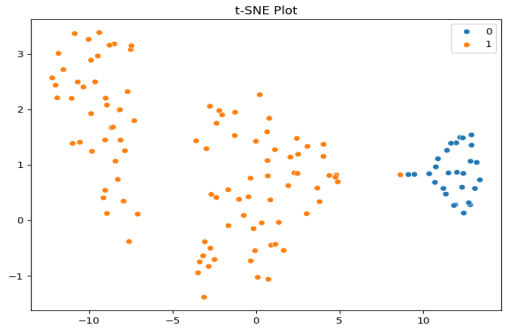

Fig 8. The t-SNE plot for MobileNet V2.

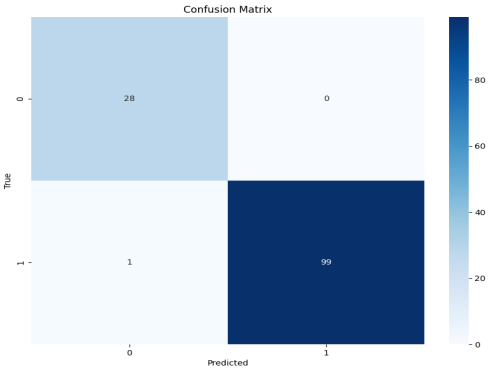

Fig 9. The the confusion matrix for MobileNet V2.

with a 99% on testing data. Both models tend to converge on the accuracy graph for their training and testing accuracy.

The training time (Table 1) was similar for both models, hence, MobileNet offers a good balance of size and cost. On comparing the training and validation accuracy curves of both EfficientNet B7 and MobileNet V2, the MobileNet V2 training and validation accuracies tend to intersect and converge faster than the former. This can be attributed to better learning, while the EfficientNet B7 was a slow learner for the early epochs with the tendency to converge by the 9th epoch. Figure 10 shows examples of correctly classified images for the Retina dataset.

VII. CONCLUSION

Both MobileNet V2 and EfficientNet B7 were able to extract features and correctly classify the images as Cataract or Normal with perfect or nearly perfect accuracy. The models underwent

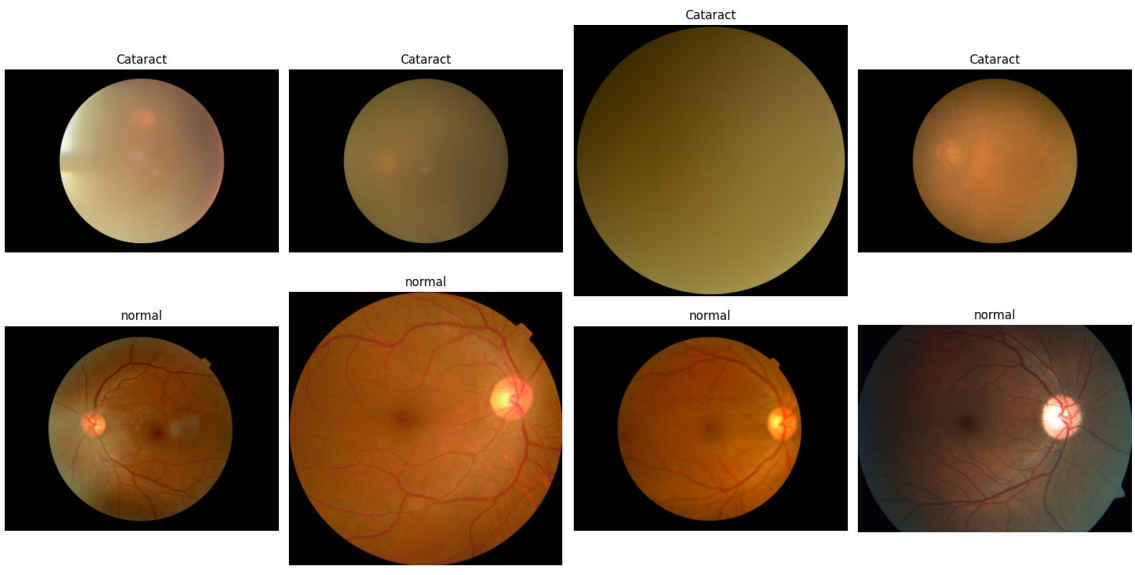

Fig 10. Examples of correctly classified images from the Retina dataset [12]

exceptional learning curves given the limited data availability and yet were able to correctly classify the images with at least 99% accuracy.

On comparing the two, MobileNet V2 gives a reasonable performance for just a fraction of cost of the total parameters of EfficientNet B7, opening the doors to integrating the model into practical handheld devices, offering agility and accuracy.

The future work in this domain could expand upon classifying the retina images in video frames in real-time and may additionally focus on classifying further eye-related disease conditions.

REFERENCES

[1] I. A. Ahmed *et al.*, "Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and Deep Learning Techniques," *Electronics*, vol. 11, no. 4, p. 530, Feb. 2022. doi:10.3390/electronics11040530

[2] J. Sun, Y. Liu, H. Wu, P. Jing, and Y. Ji, "A novel deep learning approach for diagnosing alzheimer's disease based on eye-tracking data," *Frontiers in Human Neuroscience*, vol. 16, Sep. 2022. doi:10.3389/fnhum.2022.972773

[3] S. Saxena, L. K. Fink, and E. B. Lange, "Deep learning models for webcam eye tracking in online experiments," *Behavior Research Methods*, vol. 56, no. 4, pp. 3487–3503, Aug. 2023. doi:10.3758/s13428-023-02190-6

[4] A. A. Jordan, A. Pegatoquet, A. Castagnetti, J. Raybaut, and P. Le Coz, "Deep learning for eye blink detection implemented at the edge," *IEEE Embedded Systems Letters*, vol. 13, no. 3, pp. 130–133, Sep. 2021. doi:10.1109/les.2020.3029313

[5] K. I. Lee, J. H. Jeon, and B. C. Song, "Deep learning-based pupil center detection for fast and accurate eye tracking system," *Lecture Notes in Computer Science*, pp. 36–52, 2020. doi:10.1007/978-3-030-58529-7_3

[6] H. Rahman, M. U. Ahmed, S. Barua, P. Funk, and S. Begum, "Vision-based driver's cognitive load classification considering eye movement using machine learning and Deep Learning," *Sensors*, vol. 21, no. 23, p. 8019, Nov. 2021. doi:10.3390/s21238019

[7] S. Navaneethan, P. Siva Satya Sreedhar, S. Padmakala, and C. Senthilkumar, "The human eye pupil detection system using bat optimized deep learning architecture," *Computer Systems Science and Engineering*, vol. 46, no. 1, pp. 125–135, 2023. doi:10.32604/csse.2023.034546

[8] A. E. El Hmimdi, Z. Kapoula, and V. Sainte Fare Garnot, "Deep learning-based detection of learning disorders on a large scale dataset of Eye Movement Records," *BioMedInformatics*, vol. 4, no. 1, pp. 519–541, Feb. 2024. doi:10.3390/biomedinformatics4010029

[9] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic detection of diabetic eye disease through deep learning using fundus images: A survey," *IEEE Access*, vol. 8, pp. 151133–151149, 2020. doi:10.1109/access.2020.3015258

[10] A. Sinha, A. R P, and N. N. S, "Eye tumour detection using Deep Learning," *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pp. 1–5, Mar. 2021. doi:10.1109/icbsii51839.2021.9445172

[11] H. H. Vo and A. Verma, "New Deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space," *2016 IEEE International Symposium on Multimedia (ISM)*, pp. 209–215, Dec. 2016. doi:10.1109/ism.2016.0049

[12] Nur Rusyidah Azri, "Medical Imaging Datasets for Multimodal Disease Detection and Diagnosis Research", IEEE Dataport, August 10, 2024, doi:10.21227/kebq-7h82

[13] M. Tan and Q. V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2020. doi:https://arxiv.org/abs/1905.11946

[14] O. Russakovsky et al., "Imagenet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, Apr. 2015. doi:10.1007/s11263-015-0816-y

[15] Huang, Y., Cheng, Y., Chen, D., Lee, H., Ngiam, J., Le, Q. V., and Chen, Z. Gpipe: Efficient training of giant neural networks using pipeline parallelism. arXiv preprint arXiv:1808.07233, 2018.

[16] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Jun. 2018. doi:10.1109/cvpr.2018.00474

[18] T. Kluyver *et al.*, "Jupyter Notebooks - a publishing format for reproducible computational workflows," *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 2016. doi:10.3233/978-1-61499-649-1-87

[19] Paszke *et al.*, "PyTorch: an imperative style, high-performance deep learning library," *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 8026–8037, Dec. 2019. doi:10.5555/3454287.3455008

[20] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. doi:10.1109/mcse.2007.55

[21] M. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, Apr. 2021. doi:10.21105/joss.03021

[22] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. doi:10.1038/s41586-020-2649-2

[23] R. R. Chace, "Treatment of cataract," *The American Journal of Nursing*, vol. 48, no. 3, p. 150, Mar. 1948. doi:10.2307/3457680

[24] D. Allen and A. Vasavada, "Cataract and surgery for Cataract," *BMJ*, vol. 333, no. 7559, pp. 128–132, Jul. 2006. doi:10.1136/bmj.333.7559.128

[25] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An Automated Cataract Detection System using deep learning for fundus images," *IEEE Access*, vol. 9, pp. 128799–128808, 2021. doi:10.1109/access.2021.3112938

[26] H. Morales-Lopez, I. Cruz-Vega, and J. Rangel-Magdaleno, "Cataract detection and classification systems using Computational Intelligence: A Survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1761–1774, Jun. 2020. doi:10.1007/s11831-020-09440-2

[27] G. G et al., "Deep learning with unsupervised and supervised approaches in medical image analysis," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1580–1584, Apr. 2022. doi:10.1109/icacite53722.2022.9823491

[28] R. Yamashita, M. Nishio, R. K. Do, and K. Togashi, "Convolutional Neural Networks: An overview and application in Radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018. doi:10.1007/s13244-018-0639-9

[29] D. Bein, A. Verma, and P. Ly, "Skin cancer recognition with novel deep learning methodology on mobile platform," *International Journal of Computational Vision and Robotics*, vol. 1, no. 1, 2024. doi:10.1504/ijcvr.2024.10064682

[30] Y. Jiménez-Gaona *et al.*, "BraNet: A Mobil application for breast image classification based on Deep Learning Algorithms," *Medical & Biological Engineering & Computing*, vol. 62, no. 9, pp. 2737–2756, May 2024. doi:10.1007/s11517-024-03084-1