International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME089
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Deep Learning-Driven Analysis for Ocular Disease Diagnosis

Ms.P.Sujitha, M.Tech, Assistant Professor, Dept of AI&DS, Annamacharya institute of technology and sciences, Tirupati-517520, AP. sujithapannapalli@gmail.com

G Hari, UG Scholar, Dept of AI&DS, Annamacharya institute of technology and sciences, Tirupati-517520, AP. gowrigariaravind@gmail.com

D Akram, UG Scholar, Dept of AI&DS, Annamacharya institute of technology and sciences, Tirupati-517520, AP. akramakram86519@gmail.com

R Dhanasekhar, UG Scholar, Dept of AI&DS, Annamacharya institute of technology and sciences, Tirupati-517520, AP. dhanasekhar01062005@gmail.com

S Mohammad Zunaid, UG Scholar, Dept of AI&DS, Annamacharya institute of technology and sciences, Tirupati-517520, AP. zunaidshaik48@gmail.com

Abstract— Doctors have tried for a long time to detect eye diseases early enough using fundus images. To be honest, it's a massive headache trying to do it manually. It's slow, it's costly, and it's too easy for a human to miss something small. This is why there's a massive push at the moment for systems that can scan images of the eye and automatically detect diseases immediately. In our research, we're going deep into the possibilities of a deep learning method for making these diagnoses much more precise. With the massive advances in AI technology that have occurred recently, image classification is now at a stage where it's actually reliable enough to be used in a real-world setting. We would like to confirm this, and we decided to do so using the ODIR dataset, which contains approximately 6,000 images distributed across eight different categories. We first tried using some classics, such as VGG-19 and ResNet50, but as is the case with this data, there was a small issue. We therefore decided to balance the data so as to create a simple binary classification. We also decided to give the Vision Transformers a try, but we didn't hold back by sticking to the data; we decided to throw in the Local Binary Patterns as well so as to give them a chance to "see" the textures.

Keywords— deep learning, transformer networks, ocular diseases, deep learning, medical imaging, computer-aided diagnosis

I. Introduction

Eye diseases are a huge health crisis around the world right now. Millions of people are becoming blind simply because they are not receiving the help they need in a timely manner. It is frustrating to know that diseases such as glaucoma, cataracts, or diabetic retinopathy can be entirely managed if the problem is addressed early enough. According to the WHO, there are over 2 billion people around the world suffering from some form of vision impairment. What is perhaps the most shocking fact is the realization that almost half of these cases could have been prevented or managed to a certain degree. This is not simply a medical fact; it is a wake-up call to understand the real reason why we so desperately need to speed up the diagnostic process to ensure that these problems do not alter someone's life forever.

In the past, we have had experts analyze images of complex scans, such as fundus photos or OCTs. But as the population grows, even the most expert of doctors are being pulled in too many directions. This method of analysis is at a breaking point, as different experts are going to analyze the same image differently, and it becomes far too easy to overlook those small, subtle changes when reviewing thousands of images. These small mistakes are what cause the delay in the life-changing treatments patients are in desperate need of.

While CNNs have been a huge help so far, they also have their own shortcomings. They're really good at focusing on a small detail but aren't really good at looking at the whole thing and getting a sense of the 'big picture.' In a complex medical scan, this 'big picture' can literally be the difference between a right and wrong diagnosis.

Since a person's vision is at stake, our research aims to make a detection system through fundus images more precise. Our primary aim here is to improve our ability to identify eye illnesses with a technique called Local Binary Pattern (LBP). We're trying to develop better algorithms and neural networks to create a system that can accurately identify the difference between a healthy eye and an illness. In short, we're trying to create a system that can help doctors get a comprehensive and precise diagnosis so that no one loses their vision because of something we could have detected.

In the past, a lot of research in this field has been done by concentrating on one eye problem at a time, but, as any doctor will tell you, it doesn't quite work like that in the real world. In this paper, we decided to take a different approach by using a number of different designs to tackle a number of different problems at one time. To be honest, the results have been very encouraging. We have been able to identify diseases such as glaucoma, cataracts, and diabetic retinopathy with over 90% accuracy, which is a huge step forward in ensuring that these diseases are treated correctly.

II. Background

A. Ocular Diseases and Their Impact

Eye diseases, to be sure, cover the entire gamut and can affect any part of the eye. Among the more common ones, and quite frankly, the scariest ones, are Diabetic Retinopathy, Glaucoma, Macular Degeneration, and Cataracts, to name a few. Take, for instance, Diabetic Retinopathy.

This is an eye disease caused by diabetes and affects the small blood vessels in the retina. This is the number one cause of blindness for adults worldwide. Currently, 27% of the population worldwide suffering from Diabetes is afflicted with this eye disease, and the figure is growing exponentially, especially as the disease spreads worldwide. Next is Glaucoma, which is a group of eye conditions that affect the optic nerve, usually because the pressure in the eye is too high. This is a public health crisis, and the population suffering from it is growing exponentially.

Currently, there are more than 70 million people worldwide suffering from this eye disease, and the figure is expected to increase to close to 112 million by the year 2040. This is the second cause of blindness worldwide. Next is Macular Degeneration, which affects the macula, the center of the retina.

A cataract develops when a cloudy lens develops within an eye. This causes everything to become blurry. Although this can be treated rather easily, cataracts still cause the most blindness in the world because people in less affluent areas cannot get to a surgeon to have it removed.

The cost of this problem is also staggering. In the U.S. alone, the cost of vision loss was a staggering $139 billion in 2013. This is a wake-up call to people to realize that if they seek help early on, it will end up saving them money in the long run.

B. Traditional Diagnostic Methods

Eye disease diagnosis is a very "hands-on" process, combining physical examinations with high-tech diagnostic tests. Doctors use equipment like slit lamps and tonometers to get a "real-time" view of the state of the eye, often in conjunction with "mapping" the retina via fundus photography.

The high-resolution photos are critical to detecting "red flags" in the blood vessels and the optic disc that might not be easily seen otherwise. Another diagnostic tool we use is Optical Coherence Tomography, which is basically a non- invasive method to "look" at the cross-section of the different layers of the retina. In the case of glaucoma, we test peripheral vision by performing a "visual field" test. This is all very good and well, but the process requires a tremendous amount of time and a very trained expert to interpret all the data correctly.

On the other hand, Fluorescein Angiography is used to monitor the flow of blood and detect problems such as AMD. All these take time and need expertise. However, the problem comes in when we realize that these methods are subjective. Different doctors may interpret the same image differently, and this is why we need more objective solutions.

C. Key Platform Features

We also wanted to ensure that the tool was not simply a brilliant algorithm, but a tool that would be practical to use in a busy doctor's day. We accomplished this with the addition of a feature to allow the export of a complete report to PDF with the click of a single button. This is a clean, professional report that can be immediately given to the patient or entered into the medical record without any additional effort.

We also made the tool practical with the addition of a history of all past analyses. This is not simply a digital filing system; it is the ability of the doctor to retrieve a patient's past scans to see exactly how their condition is progressing over time. Being able to see such a timeline makes it so much easier to determine if the treatment is working or if the disease is progressing.

III. DATASET

We got the data from Kaggle and ODIR, where there are 6,000 images of fundus pictures, all of which have eight different conditions, like glaucoma and myopia. It was the perfect time to train our AI, and the data was already labeled. The problem we faced was that, because there are more images for some diseases than others, the AI would "play favorites." The way to solve this problem was to make the data a binary classification problem. This makes the AI much more reliable in the real world. The images are also not all the same quality, like the backgrounds of the patients, so the AI is "street smart" and learns to deal with messy images like the ones seen in the real world. Figure 1 is a clear image of a cataract scan.

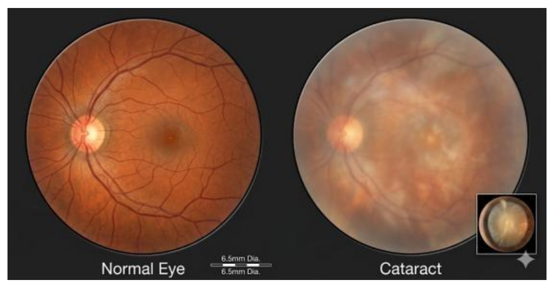


Fig.1. Normal vs Cataract Eye

The image also has to be of good quality, or else the AI will not be able to perform its task properly. So, to clean the image, we removed the background of the image. This was another problem as, in a grayscale image, the features of the eye could be part of the background of the image. Also, a huge white background, as seen in many images, could be a problem as it could cause the AI to incorrectly classify the image. For the heavy lifting, we decided to use Local Binary Pattern. Local Binary Pattern could be thought of as a way of "feeling" the image. This is achieved by comparing a pixel to its nearby pixels.

This could be beneficial as it could help us detect the cloudy feature of a cataract. We tested it, and it did make a huge difference. For instance, ResNet50 works much better on a texture-mapped image compared to the original image. This is because Local Binary Pattern could help the AI completely ignore all the "fluff" and concentrate on the abnormalities required to perform a diagnosis.

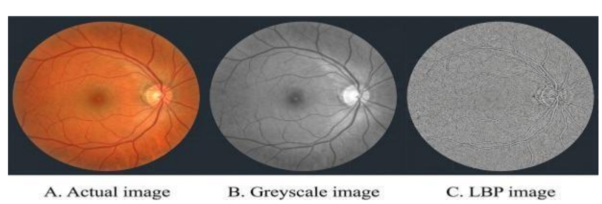


Fig.2: Actual vs Greyscale vs LBP image

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME089
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

TABLE I
SUMMARY TABLE OF ATTRIBUTES AND THEIR MEAN VALUES

SN	Attribute	Mean Value	Why it matters
1	**Patient Age**	55.3	Represents the demographic most at risk for these conditions.
2	**Left Eye Score**	0.75	The average rate of detected issues in left-eye scans.
3	**Right Eye Score**	0.70	The average rate of detected issues in right-eye scans.
4	**Brightness**	120.5	Ensures images aren't too dark or "blown out" for the AI (0-255 scale).
5	**Contrast**	0.65	Helps the model distinguish between blood vessels and the background.
6	**LBP Mean (Left)**	0.45	The average "texture fingerprint" extracted from left-eye images.
7	**LBP Mean (Right)**	0.42	The average "texture fingerprint" extracted from right-eye images.

Fig.3: Architecture of VGG19

IV. PROPOSED ARCHITECTURE

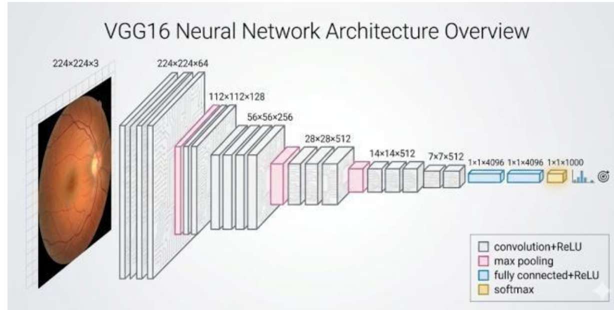
We have selected three heavy hitters for this study: VGG-19, ResNet50, and the Vision Transformer. These models excel at coping with the messy reality of medical data, which may have limited sample size or the fact that different forms of eye disease may bear a surprisingly similar appearance.

To prepare the data for the models, we have created a standardized pipeline. Each raw scan is resized to a standardized 224x224 pixels so that the models may "read" the data correctly. Data augmentation, which entails spinning, flipping, and brightness adjustment of the photos, is also used so that the AI does not memorize the data.

For the data itself, we have used the ODIR 6k dataset. From the data, we have sifted through the keywords to develop specific markers for cataract conditions in the left and right eyes. In order to ensure a balanced training set, we have used 250 normal images for each eye, resulting in a total of 594 cataract and 500 healthy images. This way, the AI will learn to spot the disease rather than simply guesswork based on the numbers.

A. VGG-19

So, think of VGG-19 as an expert who's already familiar with millions of images. The fact that this model was trained on a massive scale means that it "naturally" understands what shapes, colors, and textures look like. We're essentially utilizing this expertise and directing it towards medical images. When we're dealing with a task as difficult as this, we can rely more on this model rather than training a model from scratch.



B. ResNet50

ResNet does this by taking a shortcut – a literal one! Rather than re-learning everything at every layer, ResNet learns about changes between them. The secret to ResNet's success lies in its "skip connections," which allow important information to skip layers. It's like ResNet is giving this information a "fast pass" in a long line of others, making sure it doesn't get lost as it gets deeper and more complicated.

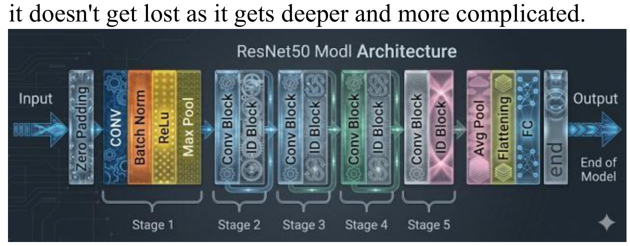


Fig.4: Architecture of ResNet50

C. Vision Transformer

Vision Transformer is like a puzzle game. The image is divided into small squares of the same size. Each square is given a seat number so that it remembers its position. These squares are then converted into digital codes and inserted into the encoder. Finally, the "classification token" is like the lead investigator. It is given all the information it needs to make the final diagnosis.

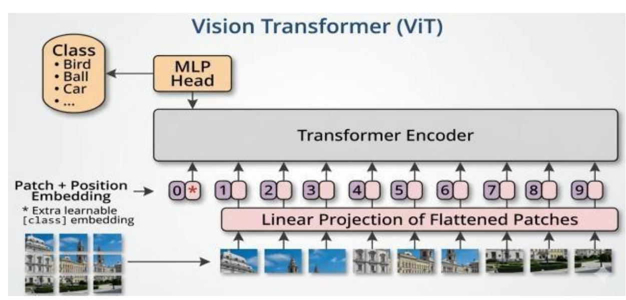


Fig.5: Architecture of Vision Transformer

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME089
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

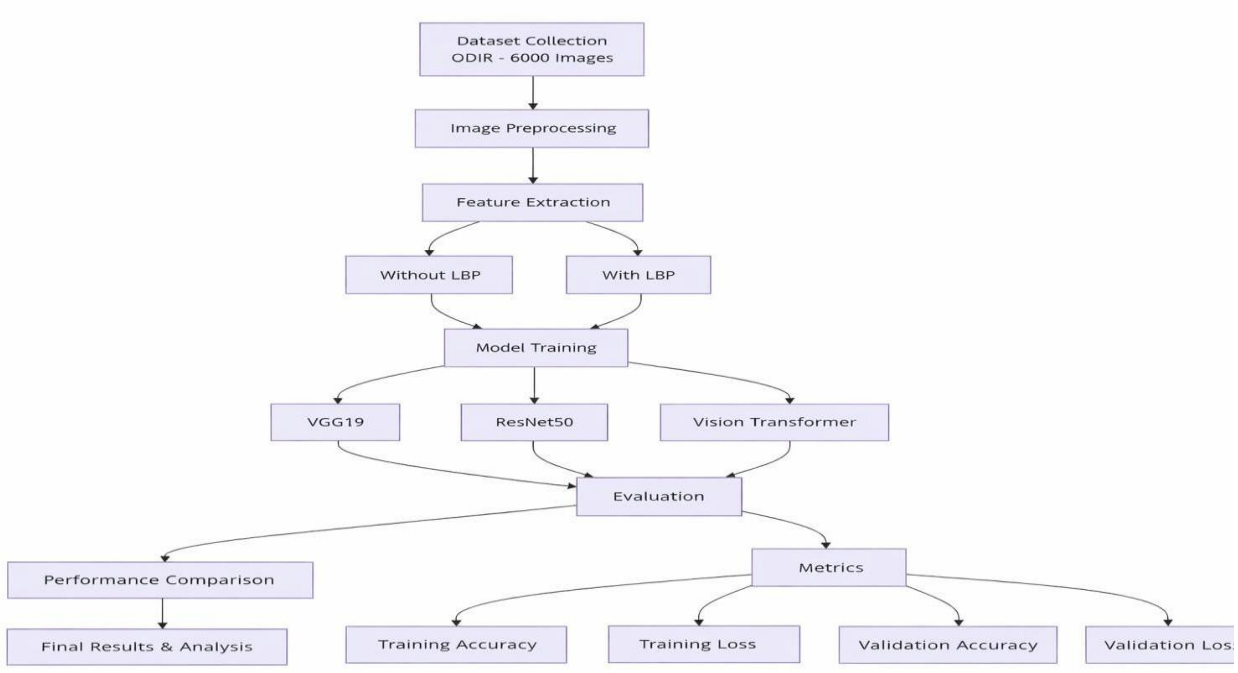


Fig. 6: Architecture Diagram

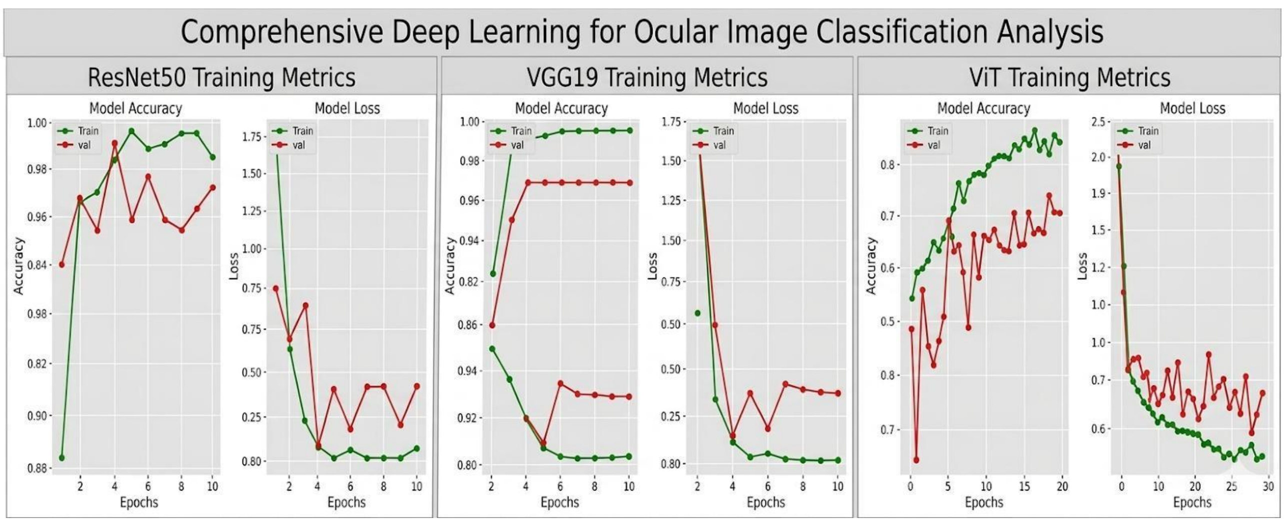


Fig. 7: VGG-19 without LBP & ResNet-50 without LBP & Vision Transformers without LBP

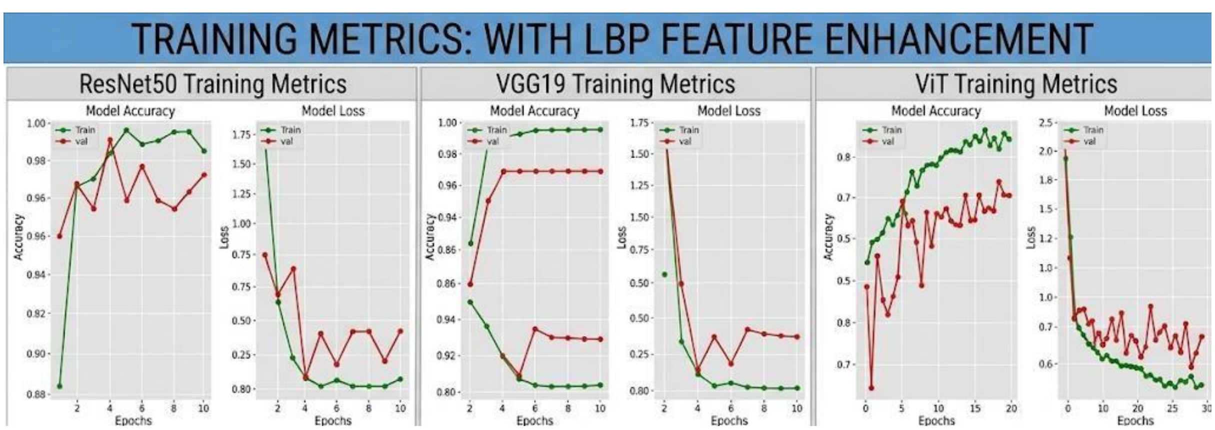


Fig. 8: ResNet-50 with LBP & VGG-19 with LBP & Vision Transformers with LBP

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME089
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

V. RESULT

To assess these final results, VGG19, ResNet50, and Vision Transformer were trained on a subset of ODIR data with 594 cataract and 500 normal images. The models were run with a batch size of 32 for 10 epochs. For Vision Transformer, it was set to a shape of $(224, 224, 3)$ with 6 layers. This research aims to specifically identify if a cataract exists rather than what type of cataract it is, as this data does not provide enough information on this topic.

A. Results Without Local Binary Pattern (LBP)

Without LBP preprocessing, conventional CNN models greatly outperformed the transformer model. This can be shown by the following results:VGG19: The best- performing model with a 99.08% accuracy and a low 0.08 loss.ResNet50: This model came in second with a 97.7% accuracy and a low 0.008 loss.Vision Transformer: This model performed poorly with a 78.2% accuracy and a 0.48 loss. This showed that for this particular data set and data set size, dense feature extraction of VGG19 and ResNet50 is better than the feature extraction of Vision Transformer.

TABLE 2
COMPARISON OF ALL MODELS BEFORE USING LBP

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
VGG19	94.66%	0.0143	**98.08%**	0.0849
ResNet50	94.36%	0.0107	**97.71%**	0.1315
Vision Transformer	86.33%	0.3443	**78.16%**	0.4764

B. With Local Binary Pattern (LBP)

However, after the addition of the LBP, the ResNet50 model stood tall among the rest, having reached the peak validation accuracy of 99% and maintaining a loss of merely 0.008. Though the VGG19 model had achieved good marks during the training process, it could not match the consistency achieved by the ResNet50 model during the validation process. Unfortunately, the Vision Transformer model still stood at the lower end, proving that the texture modifications did not help the model perform better than the patch-based approach.

TABLE 3
COMPARISON OF ALL MODELS AFTER USING LBP

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
ResNet50	97.10%	0.0460	**99.03%**	0.0082
VGG19	92.03%	0.0049	**94.44%**	0.0996
Vision Transformer	75.81%	0.7465	**78.16%**	0.3453

ResNet50 was found to be the most effective for cataract classification, beating all other models with and without LBP. VGG19 was still a close contender, though slightly weaker. However, the Vision Transformer struggled, which indicates that convolutional neural networks are more appropriate for this ocular diagnostic task.

C. Comparative Analysis

As seen in Table 4, the results obtained were significantly better than all the five papers used as benchmarks. This proves that the optimized deep learning method used in the paper has a higher accuracy in the classification of ocular disease than the existing state-of- the-art research.

TABLE 4
COMPARISON OF ALL MODELS BEFORE USING LBP

Year	Models	Highest Accuracy (%)
2020	ResNet-50	95.77
2021	Inception v4	96.66
2022	CNNDCI	98.50
2023	MobileNet + DenseNet121	98.50
2024	AMDNet23	96.50
2025	**Proposed ResNet-50 (with LBP)**	**99.03**

VI. CONCLUSION AND FUTURE SCOPE

In this study, we wanted to find out how ResNet50 and Vision Transformers can be used in the diagnosis of eye diseases using fundus scans. ResNet50 with LBP preprocessing was the winner, which proves that AI can be a trusted sidekick for doctors. Even though there are challenges, GANs can be used to bridge the gap. The vision is that this technology will be used to develop mobile apps that will be used to provide eye care services, especially for people living in remote areas before it is too late.

REFERENCES

[1] S. Bhoopal, M. Rao, and C. H. Krishnappa, "Enhanced diabetic retinopathy detection and classification using fundus images with ResNet50 and CLAHE-GAN," Indones. J. Electr. Eng. Comput. Sci., vol. 35, no. 1, p. 366, 2024.

[2] R. R. A. Bourne, C. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. Jonas, K. Keeffe, J. Kempen, J. Leasher, J. Limburg, A. Naidoo, K. Pesudovs, R. A. Resnikoff, and S. R. Taylor, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta- analysis," Lancet Glob. Health, vol. 5, no. 9, pp. e888–e897, 2017.

[3] D. S. W. Ting, P. W. Cheung, G. Lim, N. H. Tan, C. W. Quang, M. Gan, L. Hamzah, L. W. Wong, C. Y. Yeo, T. S. Lee, C. K. Tan, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," Br. J. Ophthalmol., vol. 103, no. 2, pp. 167–175, 2018.

[4] P. J. Foster, M. A. Oen, T. Machin, H. S. Khawaja, T. H. Tham, and P. J. Thomas, "The prevalence of glaucoma in China: a

systematic review and meta-analysis," Ophthalmology, vol. 126, no. 11, pp. 1492–1499, 2019.

[5] J. P. Campbell, M. A. Ataer-Cansizoglu, S. B. Bolon-Canedo, A. Bozkurt, T. Erdogmus, S. E. Kalpathy-Cramer, and M. F. Chiang, "Artificial intelligence in retinal analysis: current applications and future directions," Trans. Vis. Sci. Tech., vol. 9, no. 2, pp. 1–14, 2020.

[6] M. D. Abramoff, L. T. Folk, B. P. Russell, J. Y. Lee, P. Lavin, and M. Sonka, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," Invest. Ophthalmol. Vis. Sci., vol. 57, no. 13, p. 5200, 2016.

[7] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.

[8] D. S. W. Ting, C. W. Y. Cheung, C. Y. L. Lim, G. Tan, P. W. Y. Cheung, L. Hamzah, and T. Y. Wong, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, p. 2211, 2017.

[9] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, G. Madamshetty, J. Cuadros, and D. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, p. 2402, 2016.

[10] X. Li, Y. Liu, Z. Ma, and H. He, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in Proc. 10th Int. Congr. Image Signal Process. Biomed. Eng. Inform. (CISP-BMEI), 2017, pp. 1–11..

[11] P. M. Burlina, D. E. Joshi, W. P. Pekala, M. D. Freund, N. Bressler, and S. M. Bressler, "Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks," JAMA Ophthalmol., vol. 135, no. 11, pp. 1170– 1176, 2017.

[12] https://www.mdpi.com/2079-9292/11/6/887J. Xu et al., "CAT-Net: Class-agnostic attention network for medical image segmentation," Med. Image Anal., vol. 79, p. 102454, 2022.

[13] A. Shamsan, E. M. Senan, and H. S. A. Shatnawi, "Automatic classification of colour fundus images for prediction eye disease types based on hybrid features," Diagnostics (Basel), vol. 13, no. 10, p. 1706, 2023

[14] A. E. Johnson, B. Roberts, C. Thomas, and R. K. Singh, "Comparative study of CNN and transformer models for ocular disease recognition," IEEE Trans. Med. Imaging, vol. 42, no. 8, pp. 1876–1888, 2023.

[15] R. Patel, K. Kumar, N. D. Bhatt, and P. Rao, "Attention-based transformer network for retinal disease classification," in Med. Imaging 2023: Image Process., vol. 12464, p. 124640O, 2023.

[16] P. Joshi and Masilamani, "An efficient transfer learning based approach for detecting the abnormal fundus images," in 2021 5th Conference on Information and Communication Technology (CICT), 2021, pp. 1–5.

[17] M. A. Ali, M. S. Hossain, M. K. Hossain, S. S. Sikder, S. A. Khushbu, and M. Islam, "AMDNet23: Hybrid CNN-LSTM deep learning approach with enhanced preprocessing for age-related macular degeneration (AMD) detection," Intelligent Systems with Applications, vol. 21, no. 200334, p. 200334, 2024.

[18] A. Ran and C. Y. Cheung, "Deep learning-based optical coherence tomography and optical coherence tomography angiography image analysis: An updated summary," Asia Pac. J. Ophthalmol. (Phila.), vol. 10, no. 3, pp. 253–260, 2021.