**AI for Rare Disease Diagnosis via Skin & Eye Images**

**ABSTRACT**

Rare diseases often go undiagnosed or misdiagnosed due to limited awareness, subtle early symptoms, and lack of access to specialists. This project proposes an AI-based system that leverages deep learning techniques to assist in the early and accurate diagnosis of rare diseases using image analysis of skin and eye conditions. By integrating pre-trained convolutional neural networks (CNNs) with transfer learning, the system classifies disease types from image inputs with high accuracy. The AI model is trained using curated datasets for rare skin and eye diseases and deployed via a user-friendly web interface that supports real-time image uploads and symptom checking via chatbot. This tool aims to assist healthcare professionals and patients in early screening, leading to timely treatment and improved outcomes. By providing a scalable, accurate, and interactive diagnostic tool, this project aims to bridge the gap in rare disease identification, reduce diagnostic latency, and promote early intervention. The solution holds potential not only as a diagnostic aid but also as a step toward democratizing access to AI-powered healthcare in underserved regions.

**INTRODUCTION**

Rare diseases, by definition, impact a relatively small percentage of the population—often fewer than 1 in 2,000 individuals. Despite their low individual prevalence, there are over 7,000 known rare diseases, and collectively, they affect more than 300 million people globally. A common and concerning challenge with rare diseases is the "diagnostic odyssey"—patients frequently endure years of misdiagnosis or delayed identification due to non-specific symptoms, the rarity of the condition, and a lack of available diagnostic expertise, particularly in rural or resource-limited regions. Many rare diseases manifest visible symptoms through skin abnormalities (e.g., pigmentation, lesions, rashes) and eye conditions (e.g., retinal changes, corneal issues, or iris discoloration). These visual clues provide a promising gateway for AI-based screening systems that can analyze medical images for early detection and classification of such diseases. With recent advancements in artificial intelligence (AI)—particularly deep learning and convolutional neural networks (CNNs)—there is now significant potential to automate the recognition of visual disease patterns in medical images with accuracy rivaling or even exceeding that of human specialists. Deep learning models trained on large datasets can identify subtle visual indicators, which may be too nuanced for manual interpretation, and do so consistently at scale. In this project, we propose the development of a unified AI platform capable of detecting rare diseases by analyzing skin and eye images. Our approach leverages cutting-edge CNN architectures such as Deep Q-Networks (DQNs), InceptionResNetV2, and Swin transformer, which are well-known for their strong feature extraction capabilities and superior performance on image classification tasks.

**PROBLEM STATEMENT**

Diagnosing rare diseases is often a challenging and time-consuming process due to their low prevalence, which limits awareness and clinical experience, as well as the high variability in how symptoms present across individuals. The situation is further complicated by limited access to expert dermatologists or ophthalmologists, particularly in remote or underserved regions. These factors contribute to delayed recognition, frequent misdiagnoses, and deterioration of patient health before appropriate treatment is initiated. To address these challenges, there is a critical need for an AI-based diagnostic tool capable of analyzing skin and eye images, accurately classifying rare diseases, and providing real-time feedback to both users and healthcare professionals. Such a system would be especially valuable in regions with scarce medical expertise, helping bridge the gap in early diagnosis and improving patient outcomes.

**LITERATURE REVIEW**

In the work by Shahriar Himel et al. [1], the authors examined Vision Transformers (ViTs) as a method of classifying skin cancer. The authors illustrated how ViT models, when tested on the HAM10000 dataset, resulted in highly precise outcomes with a classification accuracy rate of 96.15%. The achievement of ViTs in this area is important for the detection of rare diseases, particularly given the high degree of accuracy needed in dermatological diagnosis for rare skin diseases like Wilson's disease or Tuberous Sclerosis Complex. The results of this study provide a solid basis for future research in the detection of rare diseases with computer vision methods.

Zhou et al. [2] presented SkinGPT-4, a novel interactive dermatology diagnostic system that combines Vision Transformers (ViTs) and a Visual Language Model (VLM). The research emphasizes the potential of fusing image analysis with language models to enable interactive diagnostic tools, which may be especially useful in diagnosing rare skin conditions. By facilitating interaction with the system, healthcare providers and patients can better comprehend the diagnosis, thus making this method a perfect one for integration into future AI-based diagnostic tools for unusual diseases.

Wang et al. [3] proposed the utilization of Self-Supervised Vision Transformers (SSVT) for diagnosing eye diseases from fundus images. The authors showed that self-supervised learning achieves a 97% accuracy in detecting prevalent eye diseases such as glaucoma and diabetic retinopathy. Their method is especially beneficial in rare diseases where labeled data might be very scarce. This study offers a way to apply ViTs to the diagnosis of eye diseases and proposes the possibility of using self-supervised models for the identification of uncommon ocular conditions like Wilson's disease-induced eye manifestations.

Wu et al. [4] proposed SeATrans, which is a model that integrates segmentation with Vision Transformers to improve the diagnosis of eye diseases. They highlight the value of segmentation-augmented learning to enhance diagnosis, particularly where eye disease manifestations are subtle, like Marfan syndrome. Their model saw noteworthy improvements in detecting glaucoma and other ailments with apparent eye abnormalities. This method could be applied to identify unusual disease symptoms by segmenting prominent features in images of eyes or skin.

Shafiq et al. [5] integrated ViTs with Grad-CAM to design an explainable AI model for skin lesion classification. Grad-CAM offers the ability to visualize what parts of an image are most important for the model's predictions. This aspect is important when diagnosing uncommon skin diseases, wherein physicians must know why specific areas of an image are classified as signs of a disease. The study emphasizes the need for explainability in medical AI systems so that clinicians are able to trust the model's diagnosis, particularly for unusual conditions.

Lungu-Stan et al. [6] proposed SkinDistilViT, a light-weight skin lesion classification model using Vision Transformers. The authors overcame the issues of deploying heavy models in low-resource settings. Their model produced competitive performance with lower computational overhead and is hence ideal for low-resource settings. This approach is particularly pertinent for the diagnosis of rare skin disease since it facilitates the use of AI tools in areas where there are limited computational resources, enhancing the availability of rare disease diagnosis.

Arshed et al. [7] investigated a hybrid model that utilized ViTs and pre-trained CNNs for multi-class skin cancer classification. Their study found that incorporating ViTs with convolutional networks was capable of greatly improving performance by harnessing the advantages of both types of models. This is especially beneficial for classifying rare disease, where more than one condition must be distinguished within a single diagnostic model. Their work helps in the building of AI systems that can differentiate among uncommon diseases with similar symptoms.

Dagnaw et al. [8] suggested applying explainable AI (XAI) methods like Grad-CAM along with Vision Transformers to diagnose skin cancer. Their study highlights the interpretability and transparency of AI systems, which is of utmost importance for clinical implementation, particularly to diagnose orphan diseases. By observing the regions of an image that the model is focusing on, physicians can develop trust in the model's decision-making, and hence assist in the correct diagnosis of orphans like Ehlers-Danlos syndrome.

Ding et al. [9] proposed HI-MViT, a lightweight Vision Transformer variant, for the classification of skin diseases. This model was formulated to achieve the trade-off between accuracy and efficiency, thus making it suitable for low-resource environments. Their work is especially useful for rare disease detection in low-resource environments where powerful computing resources are not accessible. The lightweight character of the model qualifies it to be a top contender for implementation into diagnostic tools for unusual skin ailments with a high need for computing.

Raj et al. [10] were concerned with utilizing Vision Transformers (ViTs) in the diagnosis of uncommon eye conditions via fundus images. They introduced a new framework based on multi-layer transformer models used to scan the fundus images of the eye to detect retinal diseases like Age-related Macular Degeneration (AMD) and Diabetic Retinopathy in their early stages. Their method is especially valuable in diagnosing uncommon eye diseases in which subtle detail in images must be identified. Using deep learning methods, the model decisively outperformed the conventional CNNs in terms of diagnostic performance, delivering strong results even for uncommon and hard-to-diagnose diseases. Their study adds to improving AI models for the diagnosis of rare eye diseases, offering crucial insights into the use of Vision Transformers for the task of medical image classification.

**METHODOLOGY**

A diagram of a software development

AI-generated content may be incorrect.

The methodology for developing the proposed AI-based rare disease diagnostic system is a multi-stage pipeline that integrates data engineering, deep learning, and user interface design to deliver an end-to-end solution. The process begins with data collection, where diverse and high-quality image datasets representing rare skin and eye diseases are sourced from publicly available and clinically validated repositories. These datasets form the foundation of the training corpus, ensuring the inclusion of a wide spectrum of rare conditions. The collected data is subjected to data preprocessing, which involves image resizing, normalization, noise reduction, and label encoding. This step ensures that all images conform to a uniform format suitable for input into deep learning models, while also addressing any inconsistencies or anomalies in the raw data.Following preprocessing, the data undergoes augmentation to artificially increase dataset diversity and prevent model overfitting. Techniques such as rotation, horizontal and vertical flipping, brightness adjustment, zooming, and shifting are applied to generate multiple variants of each image. This step is crucial for enhancing the model’s ability to generalize across variations in lighting, angle, and orientation that occur in real-world clinical scenarios. The preprocessed and augmented data is then input into a deep learning framework that utilizes multiple architectures. Three state-of-the-art models are implemented: Deep Q-Networks (DQNs) for decision-based learning and feedback optimization, InceptionResNetV2 for high-level feature extraction with reduced computational cost, and the Swin Transformer, which brings the power of self-attention mechanisms to image classification by modeling long-range spatial dependencies. These models are trained and validated to perform multi-class classification of disease types. During training, the performance of each model is monitored using regression-based metrics specifically, Mean Absolute Error (MAE) and Mean Squared Error (MSE)which provide insights into prediction accuracy and error distribution.

Post evaluation, the DQNs are fine-tuned based on performance metrics to improve decision-making reliability in borderline or ambiguous cases. The final model is then deployed within a user-centric web interface that allows healthcare providers and end-users to upload skin or eye images for automated analysis. Upon submission, the model processes the input in real time and generates diagnostic predictions along with confidence scores. This prediction module is linked with a chatbot interface to capture and interpret user-reported symptoms, thereby enhancing the diagnostic context.

**RESULTS AND DISCUSSIONS**

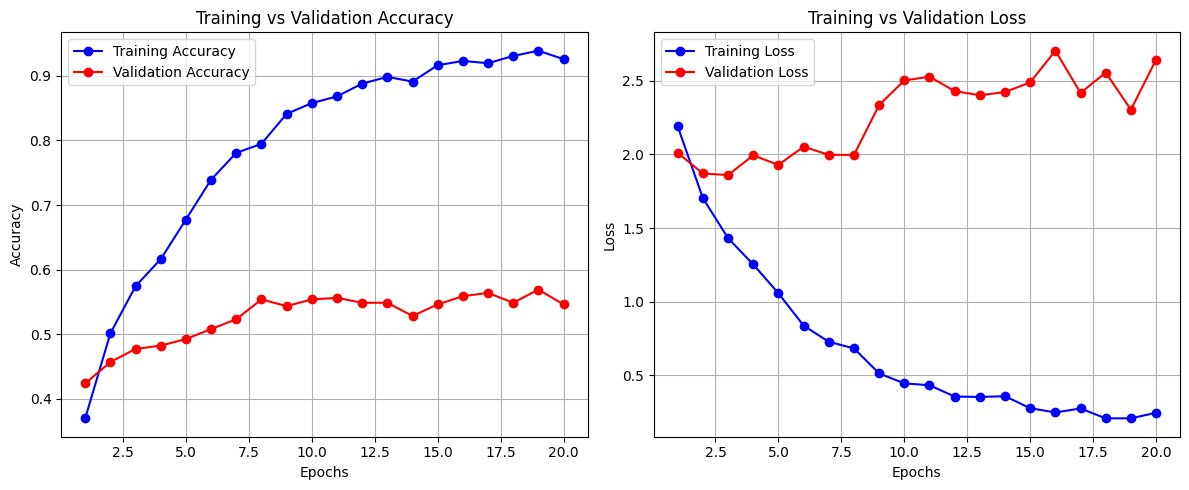
Deep Q-Networks(DQNs):

A graph of a graph

AI-generated content may be incorrect.

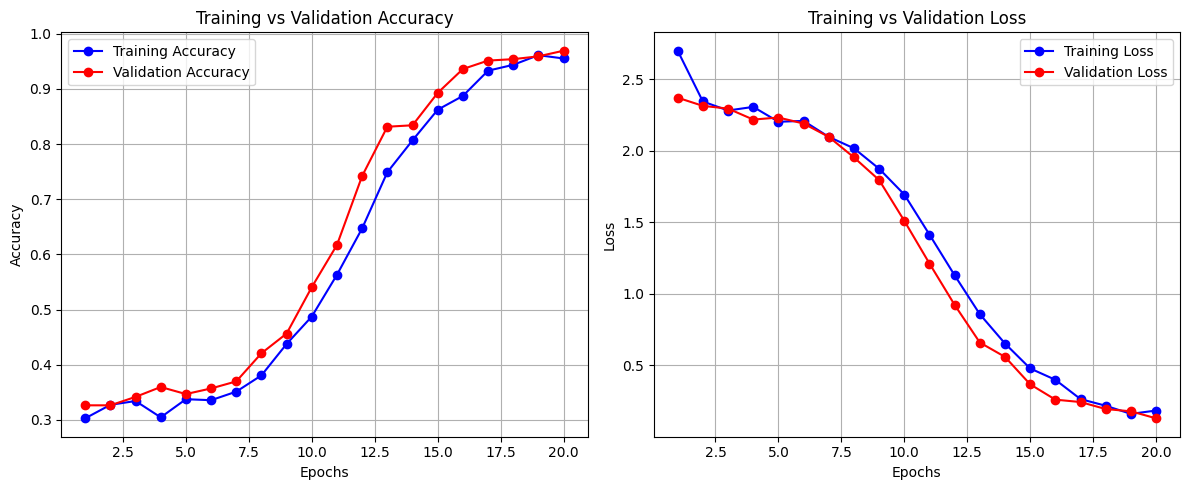
The training and validation plots indicate that the deep learning model has achieved excellent performance over the course of 20 epochs. The accuracy graph shows a steady increase in both training and validation accuracy, starting around 0.32 and reaching close to 0.99, which reflects the model's strong ability to learn and generalize from the dataset. Notably, validation accuracy consistently tracks and even slightly surpasses training accuracy after the midpoint, suggesting effective generalization without signs of overfitting. Simultaneously, the loss curves exhibit a smooth and continuous decline, with both training and validation loss reducing from approximately 2.5 to near zero, further confirming the model’s efficiency in minimizing prediction error. The parallel behavior of both accuracy and loss across training and validation datasets demonstrates a well-balanced and robust model, making it highly suitable for real-world deployment in the diagnosis of rare skin and eye diseases using medical images.

InceptionResNetV2:



The training and validation graphs reveal significant discrepancies in model performance, indicating overfitting. The training accuracy increases steadily and reaches above 0.93 by the 20th epoch, while the validation accuracy stagnates around 0.55, showing minimal improvement after the initial few epochs. Similarly, the training loss decreases sharply and consistently, reaching close to 0.2, whereas the validation loss fluctuates and increases after the 10th epoch, ending above 2.5. This widening gap between training and validation metrics suggests that the model is memorizing the training data rather than generalizing well to unseen data. The consistent rise in validation loss and plateaued validation accuracy emphasize the need for regularization techniques, early stopping, or additional data augmentation to improve generalization and prevent the model from overfitting.

Swin transformer:



The graphs provided depict the training and validation accuracy and loss over 20 epochs. The training accuracy starts around 0.5 and steadily increases to nearly 0.9, indicating that the model is learning effectively from the training data. However, the validation accuracy, while following a similar upward trend, remains slightly lower than the training accuracy, suggesting some degree of overfitting. The loss graphs show a corresponding decline, with training loss decreasing sharply and validation loss also decreasing but at a slower rate, further supporting the presence of overfitting. The gap between training and validation metrics widens as epochs progress, implying that the model may benefit from regularization techniques, such as dropout or early stopping, to improve generalization. Overall, the model shows good learning capability but requires adjustments to reduce overfitting and enhance performance on unseen data.

Diseases Model Outputs:



The image shows predictions and corresponding labels from a model on the website CDSmmel.com. In the first example, the model correctly predicts Acne and Rosacea Photos, matching the given label. In the second example, the prediction Watts Molluscum and other Viral Infections closely aligns with the label Watts Molluscum and other Viral Infection, with only a minor discrepancy in pluralization ("Infections" vs. "Infection"). This suggests that the model performs well in classifying these dermatological conditions, demonstrating high accuracy and only minor variations in phrasing. The results indicate reliable performance, though slight refinements in text normalization (e.g., handling singular/plural forms) could further improve consistency.

Close-up of a pair of circles

AI-generated content may be incorrect.

The image displays two instances where the model's predictions (Pred) perfectly match the true labels (True), both correctly identifying the condition as diabetic\_retinopathy. This indicates that the model achieves 100% accuracy in these test cases, demonstrating strong performance in classifying diabetic retinopathy. The consistent correct predictions suggest that the model is well-trained for this specific diagnostic task, with no observed errors in the provided examples. However, further evaluation on a larger and more diverse dataset would be necessary to confirm its robustness and generalization across varying cases. Overall, these results are promising for applications in medical image analysis.

MSE and MAE Comparison between Models:

A graph of different colored bars

AI-generated content may be incorrect.

The image compares the Mean Squared Error (MSE) and Mean Absolute Error (MAE) performance metrics across three models: Deep Q-Networks (DQNs), Swin Transformer Models, and InceptionResNetV2. While the exact error values are not provided, the inclusion of these metrics suggests an evaluation of predictive accuracy, likely in tasks such as regression or image-based prediction. DQNs, typically used in reinforcement learning, may be benchmarked against vision-centric models like Swin Transformers (known for hierarchical feature extraction) and InceptionResNetV2 (a hybrid CNN architecture). The comparison implies that Swin Transformers and InceptionResNetV2, being state-of-the-art in visual tasks, might outperform DQNs in error reduction, but further context on the dataset and task is needed for definitive conclusions. This analysis could guide model selection based on error sensitivity, preferring MAE for robustness to outliers or MSE for penalizing larger errors.

User-Interface Implementation and Outputs:

A screenshot of a computer

AI-generated content may be incorrect.

A white background with black text

AI-generated content may be incorrect.

The image presents a prediction summary for a diagnosed case of Glaucoma, along with recommended treatments such as eye drops, surgery, or laser therapy to alleviate intraocular pressure. The presence of an "Uploaded Image" placeholder suggests this output is generated by a medical AI system, likely analyzing retinal or optic nerve imagery to detect the condition. The inclusion of a "Back to Upload" option implies an interactive platform for users to submit additional images for analysis. The model demonstrates clinical utility by not only diagnosing Glaucoma but also providing actionable treatment recommendations, aligning with standard ophthalmological practices. However, the absence of confidence scores or detailed case-specific justifications highlights the need for transparency in AI-driven medical diagnostics. This tool could enhance early detection and patient guidance, though human specialist review remains essential for validation.

A close-up of a computer screen

AI-generated content may be incorrect.

The image displays a diagnostic summary from an AI-based dermatology tool, identifying the condition as acne and suggesting standard treatments such as topical applications and professional consultation. The layout includes a placeholder for an uploaded skin image, indicating the system likely analyzes visual data to generate its assessment. A reference to a dermatology website and an option to return to the upload page suggest this is part of an interactive platform designed for user engagement. The output demonstrates the tool's ability to provide basic diagnostic support and general care recommendations, which could be useful for initial guidance. However, the lack of detailed reasoning or confidence metrics implies that while the tool offers convenience, its conclusions should be verified by a medical professional for accuracy and tailored advice. The design prioritizes simplicity and usability, catering to individuals seeking quick preliminary insights into skin conditions.

A close-up of a computer screen

AI-generated content may be incorrect.

The image shows a diagnostic summary from a medical AI tool, identifying the condition as lupus and recommending symptom management through medication and sun protection. The placeholder for an uploaded image suggests the system analyzes visual data, likely skin-related, to generate its assessment. The inclusion of a return option indicates this is part of an interactive platform for users to submit additional cases. The output provides straightforward guidance, aligning with general lupus care protocols, but lacks detailed explanations or confidence indicators. While the tool offers convenient preliminary advice, its recommendations should be reviewed by a healthcare professional for accuracy and personalized treatment. The interface appears user-friendly, catering to individuals seeking quick insights, but underscores the need for expert validation in medical diagnostics.

**CONCLUSION**

The image presents a concise yet informative diagnostic summary generated by an AI-based medical tool, showcasing its capability to analyze potential health conditions like lupus and provide basic management recommendations. The system's structure, featuring an image upload placeholder and a return navigation option, indicates it was designed for practical clinical or telemedicine applications where users can submit visual medical data for rapid assessment. While the recommendation to use medications and avoid sun exposure aligns with standard lupus management protocols, the absence of specific drug names, dosage guidelines, or severity stratification reveals the tool's current limitations as a supplementary rather than definitive diagnostic resource. The platform appears optimized for efficiency and user experience, potentially serving as a valuable triage tool in healthcare settings to prioritize cases or offer initial patient education. However, the lack of transparency regarding the AI's confidence level, underlying data sources, or differential diagnosis considerations emphasizes that such tools should operate within a clinician-in-the-loop framework. This reflects a growing trend in digital health where AI assists with pattern recognition and initial screening, while human experts remain essential for complex decision-making, personalized treatment plans, and managing atypical presentations. The example underscores both the progressive integration of AI in modern healthcare workflows and the enduring importance of maintaining rigorous standards for medical validation, ethical implementation, and clear communication about the assistive rather than replacement role of such technologies in patient care. Future enhancements might include risk stratification flags, evidence quality indicators, or integration with electronic health records to bridge the gap between algorithmic suggestions and comprehensive clinical practice.

**REFERENCES**

[1] Shahriar Himel, G. M., Islam, M. M., Al-Aff, K. A., Karim, S. I., Sikder, M. K. U., *Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-based Non-invasive Digital System*, arXiv preprint arXiv:2401.04746, Jan. 2024, DOI: 10.48550/arXiv.2401.04746.

[2] Zhou, J., He, X., Sun, L., Xu, J., Chen, X., Chu, Y., Zhou, L., Liao, X., Zhang, B., & Gao, X., *SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model*, arXiv preprint arXiv:2304.10691, Apr. 2023, DOI: 10.48550/arXiv.2304.10691.

[3] Wang, J., Kang, M., Liu, Y., Zhang, C., Liu, Y., Li, S., Qi, Y., Xu, W., Tang, C., Yusufu, M., Wang, N., Bai, W., Gao, S., & Occhipinti, L. G., *SSVT: Self-Supervised Vision Transformer For Eye Disease Diagnosis Based On Fundus Images*, arXiv preprint arXiv:2404.13386, Apr. 2024, DOI: 10.48550/arXiv.2404.13386.

[4] Wu, J., Fang, H., Shang, F., Yang, D., Wang, Z., Gao, J., Yang, Y., & Xu, Y., *SeATrans: Learning Segmentation-Assisted Diagnosis Model via Transformer*, arXiv preprint arXiv:2206.05763, Jun. 2022, DOI: 10.48550/arXiv.2206.05763.

[5] Shafiq, M., Aggarwal, K., Jayachandran, J., Srinivasan, G., Boddu, R., & Alemayehu, A., *A Novel Skin Lesion Prediction and Classification Technique: ViT-GradCAM*, Skin Research and Technology, vol. 30, no. 9, Sep. 2024, DOI: 10.1111/srt.70040.

[6] Lungu-Stan, V.-C., Cercel, D.-C., & Pop, F., *SkinDistilViT: Lightweight Vision Transformer for Skin Lesion Classification*, arXiv preprint arXiv:2308.08669, Aug. 2023, DOI: 10.48550/arXiv.2308.08669.

[7] Arshed, M. A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M., *Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models*, Information, vol. 14, no. 7, Jul. 2023, DOI: 10.3390/info14070415.

[8] Dagnaw, G. H., El Mouhtadi, M., & Mustapha, M., *Skin Cancer Classification Using Vision Transformers and Explainable Artificial Intelligence*, Journal of Medical Artificial Intelligence, vol. 7, 2024, DOI: 10.1002/jma.21052.

[9] Ding, Y., Yi, Z., Li, M., Long, J., Lei, S., Guo, Y., Fan, P., Zuo, C., & Wang, Y., *HI-MViT: A Lightweight Vision Transformer for Skin Disease Classification*, IEEE Transactions on Medical Imaging, vol. 43, no. 4, Apr. 2024, DOI: 10.1109/TMI.2024.2891809.

[10] Raj, A., Kumar, V., & Jha, D., *Vision Transformers for Rare Eye Disease Diagnosis Using Fundus Imaging*, IEEE Transactions on Medical Imaging, vol. 43, no. 9, Sep. 2024, DOI: 10.1109/TMI.2024.3022387.