International Journal of Engineering & Management Informatics
Volume 2, Issue 2 (2026)

Fusion of Vision Transformer, Inception-V3 and ResNet50 for Efficient Eye Disease Detection

Richarv Comle[1]
Electrical and Electronic Engineering Dept. Universiti Teknologi PETRONAS Bandar Seri Iskandar 31750 Tronoh, Perak, Malaysia

Article Info

Received: 10-01-2026 **Revised:08-03-2026** **Accepted:20-4-2026** **Published:06-05-2026**

ABSTRACT
The diagnosis and categorization of retinal illnesses in the medical system is severely hampered by the increasing incidence of retinal disorders worldwide. Preventing irreversible vision loss requires prompt detection and treatment of conditions including glaucoma, DR, and macular degeneration. The use of Artificial Intelligence (AI), Machine Learning (ML) algorithms, and contemporary imaging technologies can significantly improve the accuracy and efficacy of identifying a range of ocular diseases, leading to better patient outcomes and less burden on healthcare systems. This paper presents the Vision Transfer (ViT) Algorithm, a revolutionary approach for the precise diagnosis and classification of ocular abnormalities. Our system is quite good at identifying different eye disorders using standard datasets. The ViT Algorithm performs well in distinguishing between different eye diseases by combining advanced image processing techniques with machine learning. The results show how well it improves diagnostic accuracy and enables prompt actions. This discovery greatly advances the area of eye healthcare and improves patient outcomes by offering an effective way for accurate and quick disease detection.

Keywords: ViT Algorithm, Eye Diseases, Detection, Classification, Image Processing, ML.

INTRODUCTION

The use of machine learning methods to ophthalmology has shown significant promise for accurate diagnosis and classification of ocular problems. These diseases, which include glaucoma, age-related macular degeneration (AMD), and diabetic retinopathy (DR), pose major risks to global health [1]. A prompt and correct diagnosis is essential to improving patient outcomes and reducing visual loss. However, a common element of conventional diagnostic methods is manual inspection, which may be time-consuming, costly, and prone to inter-observer variability [2] [3]. The novel ViT Algorithm has the potential to significantly change the field of diagnosing and categorizing eye diseases. By leveraging the power of transfer learning, flexibility to standard datasets, and an emphasis on model interpretability, this approach provides a path toward more accurate and accessible detection of eye diseases [4]. The study's ViT Algorithm provides a unique set of techniques to address the shortcomings of existing methods [5]. Its objective is to increase the generalization and flexibility of ML models across a variety of ocular illnesses, regardless of the standard dataset used [6]. Furthermore, by highlighting the interpretability of model choices, our approach tackles a significant problem in clinical adoption. The cutting-edge ViT Algorithm offers a promising future for the discipline of ophthalmology in addition to enhancing patient outcomes and facilitating earlier sickness identification and more tailored treatment.

1.1 A summary of ML's importance in identifying eye diseases:

ML, a key element in the area of eye illness detection, is ushering in a new age of efficiency and accuracy in the diagnosis and treatment of ocular ailments. Because the human eye is a complicated organ with many

sensitive elements, ocular disorders may appear gradually [7] [8]. In the past, ophthalmologists would manually diagnose eye disorders, which might be laborious and subject to inter-observer variability. ML algorithms, particularly those that use deep learning (DL) models [9], can handle large amounts of ocular imaging data, including fundus pictures, retinal images, and optical coherence tomography (OCT) scans, more quickly and correctly than human analysts. This technology not only aids in the early detection of diseases but also allows for the tracking of a disease's progression over time, which is crucial for the effective treatment of chronic eye conditions. Machine learning's (ML) capacity to identify trends and abnormalities that may not be

One of the primary benefits of ML in the diagnosis of visual impairments is that it is instantly apparent to the human eye [10]. These methods work particularly well at identifying minute alterations in ocular pictures that may be early signs of a number of retinal disorders. This capacity is particularly crucial when treatment results might be significantly impacted by early intervention [11] [12]. Additionally, ML algorithms can quickly identify and categorize illnesses, which lowers patient wait times and boosts the effectiveness of medical therapy [13].

Additionally, the use of ML to diagnose retinal problems is in line with the growing demand for telemedicine and remote diagnostics. Access to eye care has increased, particularly in underprivileged regions, because to the development of teleophthalmology, which has made it easier to apply ML-based techniques for remote screening and preliminary evaluations. By eliminating socioeconomic and geographical inequities and guaranteeing that all patients, wherever, get prompt and accurate diagnoses, the shift to technology-driven diagnosis of eye conditions offers promise for improving healthcare. In conclusion, machine learning plays a critical role in identifying vision-related problems since it has the potential to revolutionize the industry by offering timely diagnosis, personalized treatment plans, and more equal access to eye care services.

1.2 The ViT Algorithm idea and its ability to overcome current constraints:

Some of the present limitations in ML-based eye ailment detection and classification may be solved by a revolutionary concept known as the ViT Algorithm. By leveraging the advantages of flexibility and transfer learning, this novel approach improves the efficacy and accuracy of detecting eye conditions [14]. The ViT Algorithm is more accessible to academics and healthcare practitioners since it may significantly reduce the demand for big datasets for training, which is a need that classic ML models often have [15]. Transfer learning, which allows the algorithm to apply information from previously trained models to the specific aim of finding and classifying eye illnesses, is the basic principle behind this technique. By doing this, the ViT Algorithm eliminates the need for substantial volumes of domain-specific training data by identifying patterns and traits that are common to a number of eye conditions. This accelerates the development of diagnostic models and makes them more adaptable to treat a greater variety of ocular conditions, such as glaucoma and DR. Furthermore, the ViT Algorithm provides a new level of flexibility in the detection of eye diseases [16]. Unlike previous models that could perform well in one condition but struggle in another, this idea aims to easily adapt to a range of diseases by fine-tuning the model's parameters and design [17]. This adaptability ensures that the algorithm can maintain a high level of accuracy and specificity over a wide variety of ocular conditions. Furthermore, the ViT Algorithm introduces the concept of continuous learning, which allows the model to adapt to evolving illness patterns and new research findings [18]. By resolving these issues, the ViT Algorithm offers a novel strategy that has the potential to greatly improve the identification and categorization of eye diseases, eventually leading to an earlier diagnosis and better patient treatment.

1.3 The goals and extent of the research:

The interdisciplinary discipline of identifying and classifying eye illnesses has several objectives that meet crucial needs in ophthalmology and healthcare. The primary objective of this kind of research is to develop and evaluate new ML-based models and algorithms that can accurately and rapidly diagnose and classify a range of eye diseases. This includes diseases including AMD, glaucoma, and DR. Achieving a high level of accuracy in the early detection and accurate categorization of these illnesses is crucial because it enables timely intervention and customized treatment plans for each patient. Additionally, the initiative seeks to investigate how ML models could enhance the decision-making process's interpretability and explainability. The scope involves developing models that provide accurate classifications while also illuminating the logic of a particular diagnosis. This is essential to building trust between patients and healthcare professionals and promoting clinical adoption of these algorithms. In order to ensure that the models continue to function successfully for a range of eye conditions, the research also examines the challenges presented by uncommon diseases and class differences. In order to

improve access to healthcare services and enable remote diagnostics, the research emphasis also involves developing algorithms that can be modified for common medical datasets and integrated with telemedicine systems. Utilizing standard datasets ensures that the study's findings may be used broadly and integrated into the existing healthcare system. The main objective is to increase the accuracy of eye illness identification and categorization in order to support healthcare equality, provide eye care services to disadvantaged communities, and aid in early disease diagnosis, more tailored treatment, and improved patient outcomes.

LITERATURE SURVEY

Pradeep Kumar Jena et al. [19] describe a unique approach for screening DR using asymmetric deep learning characteristics. The proposed method first uses U-Net to segment the optic disc and blood vessels, and then uses Convolutional Neural Network (CNN) and Support Vector Machine (SVM) to categorize the DR data. There are four types of lesions: microaneurysms, hemorrhages, exudates, and normal. APTOS and other publicly accessible ocular imaging datasets are used to evaluate the model.

MESSIDOR. Exudate detection is 96.9% and 98.3% on the APTOS and MESSIDOR datasets, whereas non-DR detection is 98.6% and 91.9%, respectively.

The online dataset of retinal illnesses was utilized by P. Glaret Subin et al. [20]. In order to detect age-related retinal problems early, the data are first pre-processed using maximum entropy transformation. A CNN receives the pre-processed pictures as input. To optimize feature extraction, the CNN model is trained using the flower pollination optimization algorithm (FPOA). To improve network performance and accuracy, the trained CNN improves hyperparameters. Multicalass CNN is then used to classify the CNN's output. The Ocular Disease Intelligent Recognition (ODIR) dataset is used to evaluate the model's performance. It produces the best results when compared to other models in terms of precision, accuracy, specificity, recall, and F1 score (95.27%, 95.21%, 93.3%, and 98.30%, respectively).

Moahmmed Rashid Ahmed et al.'s paper [21] proposes an expert system that uses pre-processed fundus photos from an online dataset to identify visual diseases using Deep CNNs (DCNN). The optimization of samples is done using the maximum entropy transformation methodology, which aims to detect age-related eye issues at an early stage. These photos are fed into a CNN that has been tuned for feature extraction using the FPOA. An FPOA-trained CNN optimizes hyperparameters to increase accuracy and speed. The CNN output is then classified using a Multiclass SVM (MSVM). The ODIR dataset is used to evaluate the proposed CNN-based multiple disease detection (CNN-MDD). The model that performs the best in terms of precision, accuracy, specificity, recall, and F1 score—98.30%, 95.27%, 95.21%, and 93.3%, respectively—is compared to other optimized models.

The study of Md. Shakib Khan et al. [22] presents targeted ocular detection using deep learning. The authors classify the ODIR dataset using sophisticated classification methods, such as VGG-19. Five thousand photos from eight different kinds of eye diseases make up this collection. Because of the uneven distribution of the dataset, the authors propose using balanced samples for each category to reduce the multiclass classification issue into a binary classification problem. The binary classifications are then trained using the VGG-19. The results indicate that the normal (N) class's accuracy against the pathological myopia (M) class, cataracts, and glaucoma (G) is 98.13%, 94.03%, and 90.94%, respectively. When data are balanced, accuracy rises.

Ashrafi Akram et al. [23] offer an automated retinal problem identification method based on visually noticeable indicators using digital image processing and machine learning techniques like SVM and DCNN. This technique automatically separates the eye area from front-facing face photos by segmenting facial components. Seven eye conditions—trachoma, conjunctivitis, corneal ulcer, ectropion, periorbital cellulitis, cataracts, and vitamin A deficiency—are analyzed and categorized using this approach. Experimental findings show that the DCNN model outperforms SVM models with an average accuracy rate of 98.79%, 97% sensitivity, and 99% specificity.

An efficient model for predicting eye diseases was suggested by Ahmed Al Marouf et al. [24] using machine learning and the ranker-based feature selection (r-FS) technique. The system aims to automatically predict five common eye conditions: ocular hypertension (OH), exophthalmos, often called bulging eyes (BE), primary congenital glaucoma (PCG), acute angle-closure glaucoma (AACG), and cataracts (CT). The approach makes use of five different feature selection algorithms, two data splitting methodologies (train-test and stratified k-fold cross-validation), effective data gathering techniques, and annotation by certified ophthalmologists. Additionally, k-Nearest Neighbor (k-NN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF),

Naive Bayes (NB), AdaBoost (AB), Bagging (Bg), Boosting (BS), and SVM are among the nine machine learning methods that are integrated. The model achieves a maximum accuracy of 99.11% using SVM for 10-fold cross-validation. Liu et al.'s [25] work proposes a hybrid training technique that uses slit-lamp and smartphone photos to improve the accuracy of pterygium diagnosis and grading using smartphones. The model obtained the following results: area under the curve (AUC) = 0.9295, sensitivity = 0.8709, specificity = 0.9668, and micro-average F1 score = 0.8981. The detection model attains an average accuracy of 95.24% by using slit-lamp pictures for training. The fusion model performs comparably to the model trained on slit-lamp pictures when using smartphone photos, indicating a new method for precise smartphone image recognition in the future.

Table 1: Summary of Eye Disease Detection and Classification Using Standard Datasets

Study	Methodology	Techniques/Models Used	Datasets	Targeted Eye Diseases	Performance
Pradeep Kumar Jena et al. [19]	Segmentation of blood vessels and optic disc,	U-Net, CNN, SVM	APTOS, MESSIDOR	Normal, Haemorrhages, Exudates, Microaneurysms	APTOS: 98.6% (non-DR), 96.9% (exudates); MESSIDOR:
	classification of DR samples				91.9% (non-DR), 98.3% (exudates)
P. Glaret Subin et al. [20]	Pre-processing, feature extraction, classification	Maximum Entropy Transformation, CNN with Flower Pollination Optimization Algorithm (FPOA), Multiclass CNN	ODIR	Age-related retinal disorders	Precision: 95.27%, Accuracy: 95.21%, Specificity: 93.3%, Recall: 98.30%
Moahmmed Rashid Ahmed et al. [21]	Pre-processing, feature extraction, classification	Maximum Entropy Transformation, CNN with FPOA, Multiclass SVM (MSVM)	ODIR	Age-related eye diseases	Precision: 98.30%, Accuracy: 95.27%, Specificity: 95.21%, Recall: 93.3%
Md Shakib Khan et al. [22]	Classification of ocular diseases	VGG-19	ODIR	Eight eye disease classes	N vs M: 98.13%, N vs Cataracts: 94.03%, N vs Glaucoma: 90.94%
Ashrafi Akram et al. [23]	Automated detection and classification of eye diseases	SVM, DCNN	Local	Seven eye diseases including Trachoma, Conjunctivitis, Cataracts, etc.	Accuracy: 98.79%, Sensitivity: 97%, Specificity: 99%
Ahmed Al Marouf et al. [24]	Eye illness prediction using ML and feature selection	Ranker-based Feature Selection (r-FS), SVM, k-NN, LR, DT, RF, NB, AB, Bg, BS	Proprietary	Ocular Hypertension (OH), Exophthalmos (BE), Primary Congenital Glaucoma (PCG),	Accuracy: 99.11% (SVM with 10-fold cross-validation)

				AACG, Cataracts (CT)	

Challenges and limitations in current methods:
While classification and detection of eye have advanced significantly, ML-based methods still have several drawbacks and restrictions. Understanding these challenges is essential to understanding the state of the discipline today and identifying areas in need of additional study and advancement. The following are some of the main challenges and restrictions in this field: Limited Access to Various and Annotated Datasets: For training, a lot of ML algorithms need big and varied datasets. Nevertheless, it can be difficult to gather such datasets in the case of eye illnesses. Furthermore,
acquiring comprehensively annotated datasets encompassing a wide range of ocular disorders is frequently a barrier due to the labor-intensive and specialized nature of producing precise ground truth labels.
Inter-Observer Variability: One major problem in manual illness diagnosis is inter-observer variability. Differences in the diagnosis of eye diseases among clinicians can result in disparities in ground truth labels. ML model evaluation and training may be impacted by this heterogeneity.
Class Imbalances: Class disparities in eye diseases are common, with certain diseases being more common than others. Models may become biased in favor of the majority class as a result, which would reduce their sensitivity and accuracy when identifying rarer circumstances.
Model Interpretability: Various ML models especially those pertaining to deep learning are commonly thought of as "black boxes." Fostering clinical acceptance and confidence in a model requires an understanding of the reasoning and process involved in the diagnosis. Ensuring the interpretability of these models remains a challenging task.
Generalization to Real-World Data: It is possible that models created and tested in controlled research settings will not always translate effectively to actual clinical situations. The performance of the model might be impacted by variables like patient demographics, lighting circumstances, and differences in imaging instruments.
Privacy and Ethical Concerns: Similar to other medical data, the use of patient data for ML may give rise to issues related to privacy and ethics. It is crucial to make sure that laws like GDPR and HIPAA are followed. Finding a balance between enhancement of research research and safeguarding patient privacy is still challenging.
Clinical Validation: Comprehensive clinical validation is necessary before implementing ML models in clinical settings. This procedure can be costly and time-consuming, which prevents these technologies from being widely used.
Model Robustness: Models may not function effectively in the presence of noise or artefacts and may be sensitive to changes in image quality. Robustness in the face of such circumstances is a continuous problem.
Cost and Accessibility: The expenses associated with implementing ML-based solutions in the healthcare industry might be prohibitive for providers, particularly in environments where resources are scarce.
Regulatory Approvals: Meeting regulatory standards for medical devices and diagnostic equipment can be a time-consuming and difficult procedure, which can further delay the adoption of ML-based solutions in clinical practice.
Overcoming these challenges is essential to the further development of ML-based solutions for the diagnosis and categorization of retinal disorders and the integration of these approaches into regular clinical practice. Working together, academics, doctors, and tech developers can overcome these challenges and ensure that these technologies live up to their potential of improving patient care.
The role of standard datasets in research:
The advancement of ML-based research and development techniques for the detection and classification of ocular disorders depends on the utilization of standard datasets. These standardized datasets advance the discipline and offer a number of advantages:

Benchmarking and Reproducibility: Researchers can benchmark their algorithms using standard datasets, which offer a common foundation. Researchers are able to objectively compare the performance of various models and methodologies when they have access to a widely recognized dataset. This promotes the emergence of creative solutions and healthy competition. Additionally, it makes study findings reproducible, guaranteeing that other researchers can independently validate the findings.

Consistency and Fair Evaluation: Consistency in the evaluation of algorithms is ensured by the use of standard datasets. Since every researcher is using the same set of photos, differences that can arise from using separate data sources are eliminated. Because of this uniformity, fair and equitable evaluations are possible, which facilitates the identification of the advantages and disadvantages of various approaches.

Accessibility: Standard datasets are frequently made available to the general public, allowing researchers of all backgrounds and resources to use them. By enabling researchers from many institutions and backgrounds to work together on eye disease detection studies, this democratizes research. Researchers working in resource-constrained environments who do not have the capacity to gather their own datasets will find this accessibility to be very helpful.

Efficiency: It might take a lot of time and resources to gather and annotate datasets for study. Researchers no longer have to spend a lot of time gathering, organizing, and labelling data thanks to standard datasets. This makes it possible for academics to concentrate on creating cutting-edge algorithms and doing outcome analysis.

Comparative Analysis: Researchers can directly compare the effectiveness of their approaches with previously published results on the same dataset by using standard datasets. By comparing academics' algorithms to current state-of-the-art methods, this comparative analysis encourages innovation and advancements in the field.

Community Collaboration: Standard datasets are a valuable resource for promoting collaboration among researchers. Together, researchers can solve shared problems, exchange ideas, and improve the latest advancements in eye disease identification. Together, these efforts may result in the creation of algorithms that are more reliable and precise.

Validation of Clinical Relevance: Standard datasets frequently capture the difficulties that arise in clinical practice. Researchers can more persuasively argue for the clinical applicability of their techniques if they can show how well ML models perform on these datasets. The use of these technologies in healthcare settings may benefit from this.

Table 2 lists the widely used dataset for classifying and detecting eyes using ML techniques:

Table 2: Summary of widely used datasets for retinal disorders detection and classification

Dataset Name	Description	Number of Images	Eye Diseases Covered	Image Resolution	Source
APTOS 2019	Contains fundus images for predicting diabetic retinopathy (DR).	~3,662	Diabetic Retinopathy (DR)	Various resolutions	Kaggle
MESSIDOR	A large dataset used for the grading of diabetic retinopathy.	1,200	Diabetic Retinopathy (DR)	1440 × 960 pixels	Public domain
ODIR	A comprehensive dataset for multi-disease ocular recognition.	5,000	Multiple diseases including DR, glaucoma, AMD, etc.	224 × 224 pixels	AIROGS
DRIVE	Focuses on the segmentation of blood vessels in retinal images.	40	N/A (Blood vessel segmentation)	565 × 584 pixels	Public domain

STARE	Used for the detection of retinal vessel structures and lesions.	400	Various retinal abnormalities	605 × 700 pixels	Public domain
EyePACS	A large dataset of fundus images, primarily for diabetic retinopathy detection.	~88,000	Diabetic Retinopathy (DR)	Various resolutions	EyePACS organization
RIGA	Used for glaucoma assessment via optic nerve head segmentation.	750	Glaucoma	2144 × 1424 pixels	Public domain
IDRiD	Annotations for lesions in fundus images for DR and diabetic macular edema (DME).	516	Diabetic Retinopathy (DR), Diabetic Macular Edema (DME)	4288 × 2848 pixels	Indian Diabetic Retinopathy Image Dataset (IDRiD)
DIARETDB1	Dataset for diabetic retinopathy, featuring manually annotated lesions.	89	Diabetic Retinopathy (DR)	1500 × 1152 pixels	Public domain

Standard datasets provide many benefits, but it is vital to recognize that they might not fully capture all potential variances in clinical practice or cover the entire range of eye illnesses. Therefore, to make sure that their techniques are reliable and adaptable to clinical settings, researchers should also focus on modifying their models to handle a variety of real-world data. However, standard datasets are a useful place to start when conducting research on the identification and categorization of eye diseases. They also offer a strong basis for the creation of precise and practically applicable ML algorithms.

METHODOLOGY

ViT Algorithm-

Retinal illness identification in practical applications still requires a great deal of work, mostly because feature extraction with DL models is still the main method used. The hybrid feature extraction method presented in this work combines a Vision Transformer model with conventional pre-trained CNN models, such as Inception-V3 and ResNet-50. The features that are extracted by each model separately are eventually merged together for improved detection.

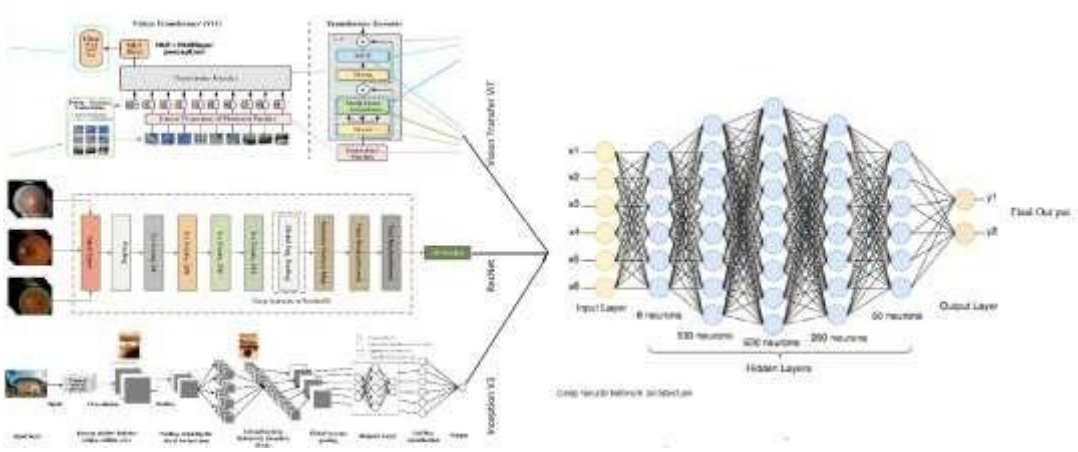


Figure 1: Architecture for Approach for Eye disease Detection and Classification Using Vision Transformation Algorithm

We employ a hybrid model in this method called Triple-Stream Conv-ViT, which combines three different models: Vision Transformer, ResNet-50, and Inception-V3. CNN-based networks Inception-V3 and ResNet-50 use convolution to extract features. They use kernels to examine correlations between neighbouring pixels in order to obtain texture information. By associating consecutive and neighbouring pixels, Inception-V3, a massive neural network, uses a multitude of filters—277 in total—to discover deep texture properties. Despite being computationally efficient, Inception-V3's complexity makes vanishing gradient issues more likely. ResNet-50 is included to improve feature extraction and reduce vanishing gradients by utilising residual connections that optimise parameters and create stronger correlations between pixels. Concurrently, the Vision Transformer concentrates on obtaining spatial pixel correlations, so streamlining the process of extracting features based on shapes.

How transfer learning, adaptability, continuous learning, and interpretability are integrated-

Inception-V3 and ResNet-50 are combined with the ViT to produce shape-based texture characteristics that boost the deep neural network's processing power and enhance classification accuracy. The principal goal of this work is the creation of a triple-stream network configuration that reliably obtains hybrid features for deep neural network classifier use in final classification. Using an attention mechanism, the Vision Transformer ViT) establishes links between local and distant pixels.

The Vision Transformer works by initially using kernels to divide the input image into small patches, much like a convolution layer would. The process produces a batch-indexed matrix with remaining dimensions represent rows, columns, and depth. Patches are integrated with positional information to group smaller patches together and scale up to bigger image sizes in order to manage the computational cost of transformers. Positional embedding uses cosine functions for odd-positioned patches and sine functions for even-positioned patches to integrate sine and cosine functions of different frequencies. Subsequently, positional embeddings are concatenated with the linearly projected patches to produce embedded patches.

Following positional embedding and linear projection, an encoder block with eight identical layers processes the patches. Normalization, multi-head self-attention (MSA), and a multi-layer perceptron (MLP) with dropout are included in each layer. Normalization and MLP layers appear after the encoder block, which concatenates the input embeddings with the MSA output. By adding a skip link from the input to the attention output, the original embedded patch is preserved for later layers, boosting the positional impact.

The utilization of standard datasets in the algorithm's development-

The Mendeley dataset that was used for preprocessing, training, testing, and evaluation is accessible to the general public. Foveal slices of the original images are used, and about 8,000 example images from routine examinations are included. To make the distribution of the test and validation sets identical to the distribut ion of the training set, adjustments are made. As such, the sample ratios in the training, testing, and validation sets are maintained constant for each class. This method guarantees that throughout processing, the model's outputs substantially resemble actual situations.

Evaluation Metrics:

Metrics such as F1 Score, Precision, Accuracy, and Recall are employed for evaluation. The representation below highlight how the dataset's imbalance makes Precision, Recall, and F1 Score more critical than Accuracy in this context:

$$Accuracy=TP+TNTP+TN+FP+FN$$

$$Precision=TPTP+FP$$

$$Recall=TPTP+FP$$

F1 Score=2×(Precision+ Recall/ Precision× Recall)

Where true positive, true negative, false positive, and false negative are represented, respectively, by the symbols TP, TN, FP, and FN. How well a model can identify a class as positive in every instance counts as its accuracy. Conversely, Precision is the ratio of true positives to the sum of false positives and true positives, and it indicates how well the model identifies a class as positive among all samples that are categorized as positive. On the other side, recall measures the model's ability to identify a class among all real instances of that class by dividing the number of true positives by the total number of false negatives. The weighted average of Precision and Recall is termed as the F1 Score.

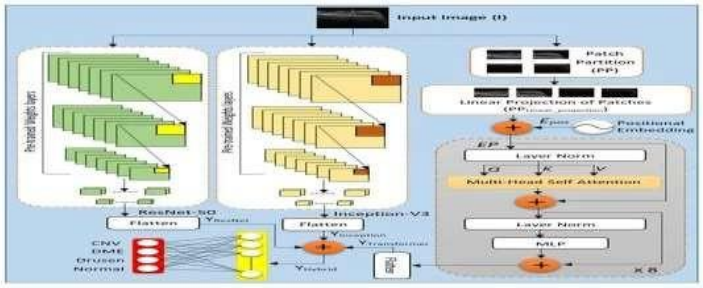


Figure 2: Combined Architecture of ViT

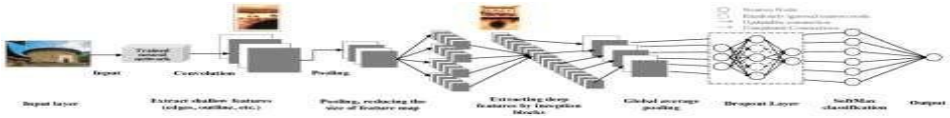


Figure 3: Inception-V3



Figure 4: ResNet50

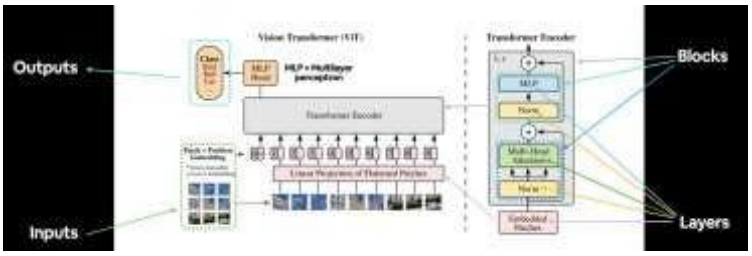


Figure 5: Vision Transformer

RESULTS AND DISCUSSIONS

Preprocessing, training, testing, and evaluation were all done using a publically accessible Mendeley dataset. It has over eight thousand sample images taken from regular checks; every image has a foveal segment of the original sample.

Model Training and Validation

Hyperparameters are changed during model training to enhance prediction accuracy. To increase the accuracy of the model, the weights are modified using the Adam optimizer using the categorical cross-entropy loss function.

Training process of ViT Algorithm

This combination of the three models, called the Triple-Stream Conv-ViT, includes Vision Transformer, ResNet-50, and Inception-V3. CNN-based networks Inception-V3 and ResNet-50 are used for convolution-based feature extraction. These CNN models are merged with the Vision Transformer to generate texture features based on shape. The deep neural network then processes these features, improving the classifier's performance in final identification. Discussion on the validation and evaluation process

The assessment employs metrics such as F1 Score, Accuracy, Precision, and Recall. Due to the dataset's imbalance, Precision, Recall, and F1 Score are given more importance over Accuracy. Accuracy reflects the model's ability to correctly identify a class as positive across all samples. On the other hand, Precision measures the proportion of true positives out of all instances classified as positive, indicating how effectively the model identifies a class as positive among all labeled positive samples.

Key finding and result from ViT algorithm

The test set is utilised for evaluation once the model has gone through 20 epochs of training. Each training session lasts roughly 20 minutes, for a total of nearly 6 hours. Every epoch, the loss and accuracy metrics for the training and validation sets are tracked. The training and validation sets' loss curves show comparable changes, however the accuracy curves for both sets demonstrate an upward trend.

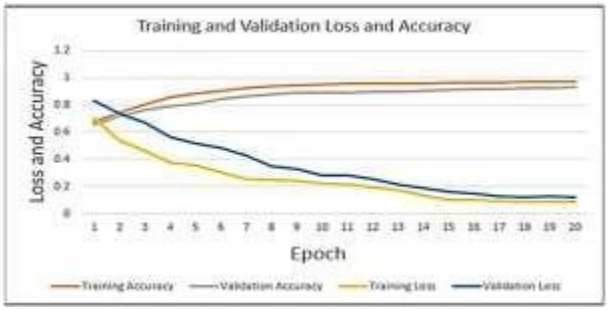


Figure 8: Sample training

The Conv-ViT model's performance is examined and assessed using sample images, with the error determined by the model's right or incorrect prediction. Hence, a variety of image from each category are selected for the review procedure. After that, the image is given to the models, and then the error is analyzed. Seven distinct models are used for the qualitative analysis, and each model was trained using the same model parameter and tuning method. Feature extractors employed including ViT, ResNet-50, and Inception-V3. Two of the classes are correctly predicted and classified by all models, while the suggested model accurately predicts every class.

The curves for training and validation accuracy indicate an increasing trend up to 16 epochs, after which they nearly plateau. Training accuracy ranges from 78.51% to 97.13% throughout 16 epochs, whereas validation accuracy ranges from 85.82% to 98.64%. Following training, the accuracy of the training and validation is 95.45% and 97.89%, respectively.

FUTURE WORK

5.1 Topics for more study and ViT Algorithm enhancements:

To better diagnose and categorize eye issues, the ViT Algorithm may be enhanced by exploring many important areas for study and development. The primary objective of the algorithm should be its flexibility to adapt to different imaging modalities and data sources. The diagnosis of eye conditions often involves a mix of OCT scans, retinal pictures, and other imaging modalities. The ViT Algorithm would be far more adaptable and helpful in a wider variety of clinical situations if it could readily accept multi-modal input and combine data from many sources.

Second, the problem of class inequality must be addressed. Many disease classes have scant samples in imbalanced datasets because many eye illnesses are very uncommon. Research may focus on developing techniques like data augmentation, transfer learning, and ensemble approaches that specifically address these imbalances. The algorithm must maintain its sensitivity and accuracy in recognizing both common and unusual eye illnesses in order to be therapeutically relevant. Furthermore, the ViT Algorithm's interpretability and explainable AI integration should be continuously improved. In order for patients and physicians to trust the algorithm's judgment, it is imperative to provide reasons for its decisions. This increases transparency and enables the algorithm to enhance therapeutic decision-making and provide educational value. In conclusion, the ViT Algorithm may be further enhanced by resolving class disparities, increasing its adaptability to multi-modal data, and enhancing the model's interpretability. In the end, these improvements will increase the algorithm's efficacy and dependability in recognizing and classifying eye conditions.

5.2 The combination of remote diagnostics and telemedicine:

The combination of telemedicine and remote diagnostics using the ViT Algorithm for the diagnosis and categorization of eye illnesses represents a breakthrough advance in the efficiency and accessibility of healthcare. By exploiting this approach's flexibility and interpretability, patients—especially those in underprivileged areas with limited access to professional eye care—can benefit from remote screening and preliminary evaluations. Patients may capture ocular images, such as fundus pictures and retinal scans, from the comfort of their homes when telemedicine systems are equipped with the ViT Algorithm. After that, these images may be securely sent to far-off medical specialists for accurate and timely analysis. A wide range of eye problems may be accurately diagnosed because to the algorithm's adaptability to standard datasets, which also serves as an early warning system for diseases including AMD, glaucoma, and DR. In addition to lessening the strain on medical facilities, this encourages people to take charge of their eye health. Additionally, the urgent demand for healthcare fairness is addressed by the combination of telemedicine and the ViT Algorithm. Residents of remote or undeveloped medical locations may now get a professional diagnostic for eye conditions without having to drive far or put up with long wait periods. The algorithm's interpretability is essential to building confidence between patients and healthcare professionals. Beyond only producing predictions, a diagnostic offers a detailed explanation of the algorithm's decision-making process. This helps patients better understand their illnesses and facilitates the choices made by medical experts on treatment and follow-up care. Therefore, telemedicine combined with the ViT Algorithm has the potential to transform the detection of eye illnesses by making it more patient-centered, efficient, and user-friendly, which will eventually result in better medical outcomes.

CONCLUSION

In summary, the ViT Algorithm shows promise as a novel approach to the urgent problem of precisely identifying and categorizing eye disorders. This system performs very well in identifying a range of eye diseases by using AI, ML, and modern imaging technologies. Its versatility and dependability are highlighted by the usage of standard datasets, making it a useful tool in the field of ophthalmic healthcare. The ViT Algorithm's proven ability to improve diagnostic precision and enable prompt treatments is a major step toward bettering patient outcomes. In the end, this study advances the field and benefits international healthcare systems by

providing a solid and accurate method of illness detection, perhaps leading to a paradigm change in the diagnosis and treatment of eye disorders.

REFERENCES

A. G. Priya Henry and A. Jude, "Convolutional neural-network-based classification of retinal images with different combinations of filtering techniques," Open Comput. Sci., vol. 11, no. 1, pp. 480–490, 2021, doi: 10.1515/comp-2020-0177.

N. Tsiknakis et al., "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," Comput. Biol. Med., vol. 135, p. 104599, 2021, doi: 10.1016/j.compbiomed.2021.104599.

P. Chandre, V. Vanarote, M. Kuri, A. Uttarkar, A. Dhore and S. Pathan, "Developing an Explainable AI Model for Predicting Patient Readmissions in Hospitals," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 587-592, doi: 10.1109/ICECAA58104.2023.1021215

R. Buettner, D. Beil, S. Scholtz, and A. Djemai, "Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings," Proc. Annu. Hawaii Int. Conf. Syst. Sci., vol. 2020-January, pp. 3216–3225, 2020, doi: 10.24251/hicss.2020.393.

S. Malik, N. Kanwal, M. N. Asghar, M. A. A. Sadiq, I. Karamat, and M. Fleury, "Data driven approach for eye disease classification with machine learning," Appl. Sci., vol. 9, no. 14, 2019, doi: 10.3390/app9142789.

J. W. Asare, P. Appiahene, E. T. Donkoh, and G. Dimauro, "Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images," Eng. Reports, vol. 5, no. 11, pp. 1–21, 2023, doi: 10.1002/eng2.12667.

V. K. Velpula and L. D. Sharma, "Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion," Front. Physiol., vol. 14, no. June, pp. 1–17, 2023, doi: 10.3389/fphys.2023.1175881.

P. R. Chandre, P. N. Mahalle, and G. R. Shinde, "Machine learning based novel approach for intrusion detection and prevention system: a tool based verification," in 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Nov. 2018, pp. 135–140, doi: 10.1109/GCWCN.2018.8668618.

P. Chandre, P. Mahalle, and G. Shinde, "Intrusion prevention system using convolutional neural network for wireless sensor network," IAES Int. J. Artif. Intell., vol. 11, no. 2, pp. 504–515, 2022, doi: 10.11591/ijai.v11.i2.pp504-515.

A. Sinha and R. S. Shekhawat, "Review of image processing approaches for detecting plant diseases ISSN 1751-9659," IET Image Process., vol. 14, no. 8, pp. 1427–1439, 2020, doi: 10.1049/iet-ipr.2018.6210.

S. Chandrappa, L. Dharmanna, and B. Anami, "A Novel Approach for Early Detection of Neovascular Glaucoma Using Fractal Geometry," Int. J. Image, Graph. Signal Process., vol. 14, no. 1, pp. 26–39, 2022, doi: 10.5815/ijigsp.2022.01.03.

L. Dai et al., "A deep learning system for detecting diabetic retinopathy across the disease spectrum," Nat. Commun., vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-23458-5.

Makubhai, Shahin, Ganesh R. Pathak, and Pankaj R. Chandre. "Prevention in healthcare: an explainable AI approach." International Journal on Recent and Innovation Trends in Computing and Communication 11.5 (2023): 92-100.

S. Prajapati, S. Qureshi, Y. Rao, S. Nadkarni, M. Retharekar, and A. Avhad, "Plant Disease Identification Using Deep Learning," 2023 4th Int. Conf. Emerg. Technol. INCET 2023, no. July, pp. 974–978, 2023, doi: 10.1109/INCET57972.2023.10170463.

Y. Tong, W. Lu, Y. Yu, and Y. Shen, "Application of machine learning in ophthalmic imaging modalities," Eye Vis. 2020 71, vol. 7, no. 1, pp. 1–15, Apr. 2020, doi: 10.1186/S40662-020-00183-6.

M. Londhe, "Classification of Eye Diseases using Hybrid CNN-RNN Models MSc Research Project Data Analytics," 2021.

Z. Li et al., "Artificial intelligence in ophthalmology: The path to the real-world clinic," Cell Reports Med., vol. 4, no. 7, p. 101095, 2023, doi: 10.1016/j.xcrm.2023.101095.

V. Mayya, S. K. S, U. Kulkarni, D. K. Surya, and U. R. Acharya, "An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images," Appl. Intell., vol. 53, no. 2, pp. 1548–1566, 2023, doi: 10.1007/s10489-022-03490-8.

P. K. Jena, B. Khuntia, C. Palai, M. Nayak, T. K. Mishra, and S. N. Mohanty, "A Novel Approach for Diabetic Retinopathy Screening Using Asymmetric Deep Learning Features," Big Data Cogn. Comput., vol. 7, no. 1, 2023, doi: 10.3390/bdcc7010025.
P. Glaret subin and P. Muthukannan, "Optimized convolution neural network based multiple eye disease detection," Comput. Biol. Med., vol. 146, no. May, p. 105648, 2022, doi: 10.1016/j.compbiomed.2022.105648.
M. R. AHMED, S. R. AHMED, A. D. DURU, O. N. UÇAN, and O. BAYAT, "An Expert System to Predict Eye Disorder Using Deep Convolutional Neural Network," Acad. Platf. J. Eng. Sci., vol. 9, no. 1, pp. 47–52, 2021, doi: 10.21541/apjes.741194.
M. S. Khan et al., "Deep Learning for Ocular Disease Recognition: An Inner-Class Balance," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/5007111.
A. Akram and R. Debnath, "An automated eye disease recognition system from visual content of facial images using machine learning techniques," Turkish J. Electr. Eng. Comput. Sci., vol. 28, no. 2, pp. 917–932, 2020, doi: 10.3906/elk-1905-42.
A. Al Marouf, M. M. Mottalib, R. Alhajj, J. Rokne, and O. Jafarullah, "An Efficient Approach to Predict Eye Diseases from Symptoms Using Machine Learning and Ranker-Based Feature Selection Methods," Bioengineering, vol. 10, no. 1, 2023, doi: 10.3390/bioengineering10010025.
Y. Liu et al., "Accurate detection and grading of pterygium through smartphone by a fusion training model," Br. J. Ophthalmol., pp. 1–7, 2023, doi: 10.1136/bjo-2022-322552.