Journal of Information Systems Engineering and Management
2025, 10(7s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

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

Prashant Raut1, Sachin Babar2, Parikshit Mahalle3
1Research Scholar, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India.
2Principal & Professor, Sinhgad Institute of Technology, Lonavala, India.
3Professor, Department of Artificial Intelligence and Data Science, Dean- Research and Development, Vishwakarma Institute of Technology, Pune, India.
prashantr43@gmail.com1, sdbabar@sinhgad.edu2, aalborg.pnm@gmail.com3

ARTICLE INFO

Received: 16 Oct 2024
Revised: 10 Dec 2024
Accepted: 26 Dec 2024

ABSTRACT

The rising prevalence of retinal disorders globally is a major challenge for the identification and classification of retinal diseases in the medical system. Timely diagnosis and treatment of disorders like glaucoma, DR, and macular degeneration are essential for preventing irreversible vision loss. Better patient outcomes and less strain on healthcare systems can result from the significant improvement in the precision and effectiveness of identifying a variety of ocular diseases by the utilization of Artificial Intelligence (AI), Machine Learning (ML) algorithms, and modern imaging technologies. The Vision Transfer (ViT) Algorithm, a novel method for the accurate identification and categorization of ocular disorders, is presented in this work. With standard datasets, our algorithm performs exceptionally well in classifying various eye diseases. Through the integration of sophisticated image processing methods with ML, the ViT Algorithm demonstrates strong performance in differentiating between various eye conditions. The outcomes demonstrate how well it works to increase diagnostic precision and make quick interventions possible. By providing an efficient method for reliable and rapid disease identification, this research significantly advances the field of ocular healthcare and improves patient outcomes.

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

INTRODUCTION

The utilization of ML techniques in ophthalmology has demonstrated great potential for precise diagnosis and categorization of ocular disorders. These conditions, which include Diabetic Retinopathy (DR), Age-related Macular Degeneration (AMD) and glaucoma, pose serious threats to world health [1]. Improving patient outcomes and preventing vision loss depend on an accurate and timely diagnosis. However, manual examination which can be protracted, expensive, and subject to inter-observer variability is a common component of traditional diagnostic techniques [2] [3]. The discipline of identifying and classifying eye illnesses could undergo a substantial transformation because to the innovative ViT Algorithm. This algorithm offers a way towards more precise and accessible eye disease diagnosis by utilizing the strength of transfer learning, adaptability to standard datasets, and a focus on model interpretability [4]. The ViT Algorithm described in this study offers a novel combination of strategies to overcome the drawbacks of current approaches [5]. Regardless of the standard dataset utilized, its goal is to maximize the adaptability and generalization of ML models across a range of ocular disorders [6]. Moreover, our method addresses a major issue in clinical adoption by emphasizing the interpretability of model decisions. In addition to improving patient outcomes and enabling earlier illness detection and more individualized therapy, the innovative ViT Algorithm presents a potential path forward for the field of ophthalmology.

1.1 An overview of the significance of ML in eye disease detection:

A new era of efficiency and precision in the diagnosis and treatment of ocular disorders is being ushered in by ML, which is a crucial component in the field of eye disease detection. Ocular diseases can manifest gradually since the human eye is a complex organ with numerous delicate parts [7] [8]. Ophthalmologists used to diagnose eye conditions manually, which may be time-consuming and prone to inter-observer variability. Large volumes of ocular imaging data, such as fundus photos, retinal images, and optical coherence tomography (OCT) scans, can be processed by ML algorithms, especially those that use deep learning (DL) models [9], faster and more accurately than by human analysts. In addition to helping with early disease identification, this technology makes it possible to track the course of a disease over time, which is essential for the successful treatment of long-term eye disorders. The ability of machine learning ML to detect patterns and anomalies that might not be

immediately visible to the human eye is one of the main advantages of ML in the diagnosis of vision disorders [10]. These approaches are especially effective in detecting subtle changes in ocular images that could be early indicators of various retinal diseases. This ability is especially important in cases where early intervention can have a significant impact on treatment outcomes [11] [12]. Moreover, ML algorithms are able to rapidly recognise and classify diseases, thereby reducing patient wait times and improving the efficiency of medical care [13].

Furthermore, the increasing need for remote diagnostics and telemedicine corresponds with the application of ML for the diagnosis of retinal disorders. The advent of teleophthalmology has facilitated the application of ML-based methods for remote screening and initial assessments, hence increasing access to eye care, especially in disadvantaged areas. The transition to technology-driven diagnosis of eye illnesses holds promise for improving healthcare by eradicating socioeconomic and regional disparities and ensuring that timely and accurate diagnoses are provided to all patients, everywhere. Conclusively, ML performs a crucial part in the identification of vision related issues, as it has the ability to completely transform the field by providing prompt diagnosis, customised treatment regimens, and more equitable accessibility to eye care services.

1.2 The concept of the ViT Algorithm and its potential to address existing limitations:

A novel idea called the ViT Algorithm has the potential to overcome some of the current constraints in ML-based eye illness diagnosis and classification. This innovative method increases the effectiveness and precision of diagnosing eye diseases by utilizing the strengths of adaptability and transfer learning [14]. The ViT Algorithm has the ability to greatly minimize the need for large datasets for training, which is a need that traditional ML models frequently have, making it more accessible to academics and healthcare professionals [15]. The fundamental idea behind this approach is transfer learning, which enables the algorithm to apply knowledge from previously trained models to the particular goal of identifying and categorizing eye diseases. In doing so, the ViT Algorithm is able to identify patterns and characteristics that are shared by several eye illnesses, which removes the requirement for large amounts of domain-specific training data. This not only speeds up the creation of diagnostic models but also increases their versatility to address a wider range of ocular disorders, including glaucoma and DR. In addition, the ViT Algorithm offers the field of eye illness identification a new degree of flexibility [16]. By fine-tuning the model's parameters and architecture, this concept is meant to adapt to a variety of diseases with ease, unlike older models that might do well in one but struggle in others [17]. This flexibility guarantees the algorithm's ability to sustain a high degree of specificity and accuracy over a broad range of ocular conditions. Additionally, the ViT Algorithm presents the idea of continuous learning, which enables the model to keep up with new discoveries in research and changing trends in disease patterns [18]. By addressing these limitations, the ViT Algorithm represents a groundbreaking approach that can significantly advance the field of eye disease detection and classification, ultimately contributing to earlier diagnosis and more effective patient care.

1.3 The research objectives and scope:

The identification and classification of eye diseases is a multidisciplinary field with many goals that fulfill important requirements in ophthalmology and healthcare. Developing and assessing novel ML-based models and algorithms that can reliably and quickly identify and categories a variety of eye conditions is the main goal of this kind of study. This encompasses conditions including DR, glaucoma, and AMD. Reaching a high degree of accuracy in the early identification and precise classification of these disorders is essential as it provides prompt intervention and individualized treatment strategies possible for every patient. The work also aims to explore how ML models could improve the interpretability and explainability of the decision-making process. The scope entails creating models that not only offer precise classifications but also shed light on the reasoning behind a given diagnostic. This is a critical component in fostering clinical acceptance of these algorithms and establishing confidence between patients and healthcare providers. In addition, the study goes into the difficulties posed by unusual illnesses and class disparities, making sure that the models continue to work well for a variety of eye ailments. The research focus also includes the creation of algorithms that can be integrated with telemedicine platforms and adjusted to standard medical datasets, facilitating remote diagnoses and improving access to healthcare services. Working with standard datasets guarantees that the study's conclusions may be widely applied and incorporated into the current healthcare system. The primary goal is to improve the precision in eye disease detection and classification in order to promote healthcare equity, reach underserved populations with eye care services, and help early disease detection, more individualized treatment, and better patient outcomes.

2. LITERATURE SURVEY

A novel method using asymmetric deep learning features for screening DR is presented by Pradeep Kumar Jena et al. [19]. The suggested approach initially performs segmentation of blood vessels and optic disc using U-Net and afterwords classify the DR samples using Convolutional Neural Network (CNN) and Support Vector Machine (SVM). Four categories of lesions are identified: normal, haemorrhages, exudates, and microaneurysms. The model is tested on publicly available ophthalmic imaging datasets including APTOS and

MESSIDOR. On the APTOS and MESSIDOR datasets, detection accuracies for non-DR are 98.6% and 91.9%, respectively, while exudate detection is 96.9% and 98.3%.

In the work of P. Glaret Subin et al.'s [20], the authors used online dataset of retinal disorders. Initially, the samples are pre-processed using maximum entropy transformation for the early identification of age- related retinal disorders.. The pre-processed images are used as a input to a CNN. The CNN model is trained with a flower pollination optimization algorithm (FPOA) to maximize feature extraction. The trained CNN optimizes hyperparameters to increase network performance and precision. The output of the CNN is then classified by using Multicalass CNN. The performance of model is assessed by using The Ocular Disease Intelligent Recognition (ODIR) dataset. When its performance is compared with other models, it yields the greatest outcomes in terms of precision, accuracy, specificity, recall, and F1 score, which are 95.27%, 95.21%, 93.3%, and 98.30%, respectively.

An expert system to detect vision illnesses using Deep CNNs (DCNN) is proposed in the study of Moahmmed Rashid Ahmed et al. [21]. via pre-processed fundus images from an online dataset. Maximum entropy transformationt technique is used for the optimization of samples, the method seeks to identify age-related eye problems early on. The FPOA is used to feed these images into a CNN that has been optimized for feature extraction. Hyperparameters are optimised by an FPOA-trained CNN to boost speed and accuracy. A Multiclass SVM (MSVM) is then used to classify the CNN output . The suggested CNN-based multiple disease detection (CNN-MDD) is tested using the ODIR dataset. It,s performance is compared with other optimized models to determine which one achieves the best outcomes in term of precision, accuracy, specificity, recall, and F1 score of 98.30%, 95.27%, 95.21%, and 93.3%, respectively.

Targeted ocular detection using deep learning is presented in the work by Md Shakib Khan et al. [22]. The authors use advanced classification techniques, including VGG-19, to categorise the ODIR dataset. This dataset consists of 5000 images that represent eight various eye disease classes. The authors suggest turning the problem of multiclass classification into a problem of binary classification by utilizing a balanced samples for each category, as a result of the dataset's unequal distribution. Then, the VGG- 19 is used to train the binary classifications, and the outcomes show that the accuracies of the normal (N) class versus pathological myopia (M) class, the normal (N) class versus cataracts, and the normal (N) class versus glaucoma (G) are 98.13%, 94.03%, and 90.94%, respectively. Accuracy increases when data are balanced.

Using digital image processing and ML techniques like SVM and DCNN, Ashrafi Akram et al. [23] propose an automated retinal disorders detection system based on visually noticable indications. By dividing facial components, this method automatically isolates the eye region from front-facing facial images. This method is applied to the analysis and classification of seven eye diseases including trachoma, conjunctivitis, corneal ulcer, ectropion, periorbital cellulitis, cataracts, and vitamin A insufficiency. The DCNN model performs better than SVM models, according to experimental results, with an average accuracy rate of 98.79%, 97% sensitivity, and 99% specificity.

Ahmed Al Marouf et al.'s [24] proposed an effective model for eye illness prediction make use of ranker-based feature selection (r-FS) approach and machine learning. Ocular hypertension (OH), exophthalmos, often known as bulging eyes (BE), primary congenital glaucoma (PCG), acute angle-closure glaucoma (AACG), and cataracts (CT) are the five prevalent eye illnesses that the algorithm seeks to automatically forecast. The strategy uses two data splitting strategies including stratified k-fold cross-validation and train-test, five distinct feature selection algorithms, efficient data collection methods, and annotation by licensed ophthalmologists. It also integrates nine machine learning techniques including k-Nearest Neighbour (k-NN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), AdaBoost (AB), Bagging (Bg), Boosting (BS), and SVM. Using SVM for 10-fold cross-validation, the model reaches a maximum accuracy of 99.11%. To enhance the precision of pterygium detection and grading via cellphones, Liu et al.'s [25] study, presents a hybrid training approach which makes use of slit-lamp and smartphone images. Area under the curve (AUC) = 0.9295, sensitivity = 0.8709, specificity = 0.9668, and micro-average F1 score = 0.8981 are the results attained by the model. Through training with slit-lamp images, the detection model achieves an average accuracy of 95.24%. Using photographs from smartphones, the fusion model performs similarly to the model trained on slit-lamp images, suggesting a novel approach for future accurate smartphone image identification.

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), AACG, Cataracts (CT)	Accuracy: 99.11% (SVM with 10-fold cross-validation)
Liu et al. [25]	Pterygium detection and grading using hybrid training approach	Hybrid model using Slit-lamp and Smartphone images	Xiamen Eye Center of Xiamen University and Xiang'an Hospital of Xiamen University	Pterygium	AUC: 0.9295, Sensitivity: 0.8709, Specificity: 0.9668, F1 Score: 0.8981, Accuracy: 95.24%

2.1 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.

2.2 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.

3. METHODOLOGY

3.1 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.

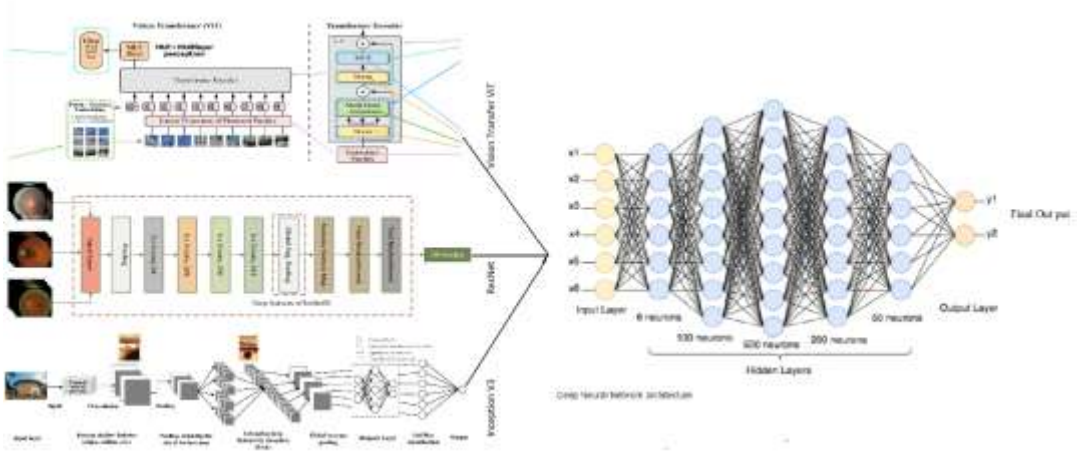

Figure 1: Architecture for Approach for Eye disease Detection and Classification Using Vision Transformation Algorithm

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.

3.2 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.

3.3 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 distribution 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:

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

$$Precision=TPTP+FP$$

$$Recall=TPTP+FP$$

$$\text{F1 Score}=2\times(\text{ Precision}+\text{ Recall}/\text{ Precision}\times\text{ Recall })$$

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.

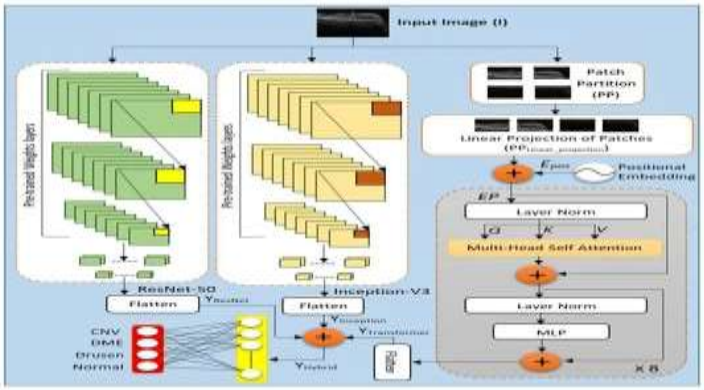

Figure 2: Combined Architecture of ViT

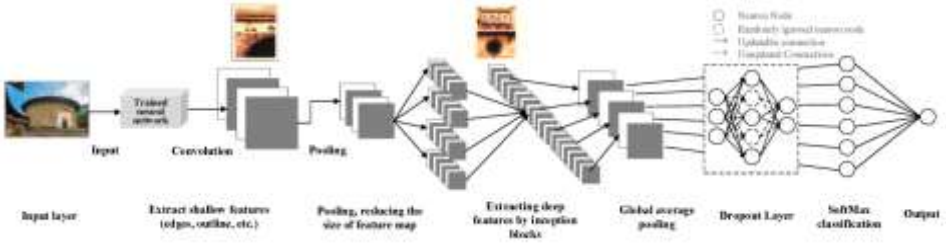

Figure 3: Inception-V3

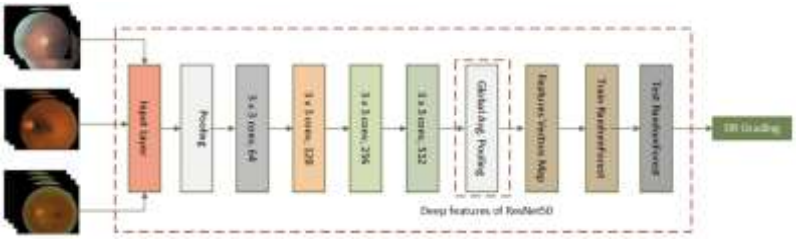

Figure 4: ResNet50

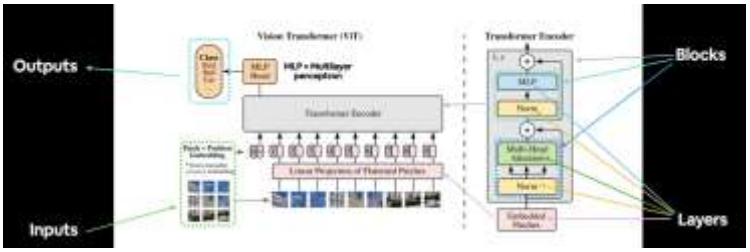

Figure 5: Vision Transformer

4. RESULTS AND DISCUSSIONS

Data collection

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.

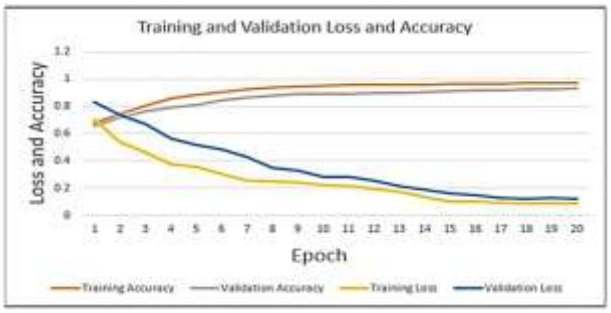

Figure 8: Sample training

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.

5. FUTURE WORK

5.1 Areas for further research and improvements to the ViT Algorithm:

The ViT Algorithm can be improved by delving into a number of crucial areas for research and development in order to better diagnose and classify eye problems. The algorithm's adaptability to various imaging modalities and data sources should be its top goal. A combination of OCT scans, retinal images, and other imaging modalities is often used in the diagnosis of eye illnesses. If the ViT Algorithm could easily embrace multi-modal data and merge data from multiple sources, it would be far more flexible and useful in a larger range of clinical scenarios.

Second, it's imperative to address the issue of class disparities. Due to the relative rarity of many eye disorders, several disease classifications have sparse samples in datasets that are unbalanced. Research might concentrate on creating strategies that particularly address these imbalances, such as ensemble methods, transfer learning, and data augmentation. For the algorithm to be clinically useful, it must be ensured that it continues to be sensitive and accurate in identifying both common and uncommon eye disorders. Additionally, there should be continuous development in the areas of interpretability and explainable AI integration for the ViT Algorithm. It is essential to give explanations for the judgements made by the algorithm so that patients and clinicians can have faith in its judgement. This improves openness while also allowing the algorithm to offer educational value and support therapeutic decision-making. To sum up, the ViT Algorithm can be improved further by addressing class imbalances, becoming more adaptable to multi-modal data, and improving the interpretability of the model. These enhancements will ultimately boost the algorithm's effectiveness and reliability in identifying and categorizing eye diseases.

5.2 The integration of telemedicine and remote diagnostics:

A revolutionary step in the accessibility and efficiency of healthcare has been taken with the integration of telemedicine and remote diagnostics utilising the ViT Algorithm for the identification and classification of eye diseases. Patients, particularly those in underserved locations with limited access to specialized eye care, can benefit from remote screening and preliminary assessments by utilizing the adaptability and interpretability of this approach. When telemedicine platforms are outfitted with the ViT Algorithm, patients can take ocular images from the comfort of their homes, including fundus photos and retinal scans. After that, these pictures can be safely sent to distant medical professionals for prompt and precise analysis. The algorithm's flexibility in adapting to standard datasets enables accurate diagnosis of a broad spectrum of eye disorders, acting as an early warning system for conditions such as AMD, glaucoma, and DR. This not only reduces the burden on healthcare facilities but also empowers individuals to take proactive steps in managing their eye health. Furthermore, the integration of telemedicine and the ViT Algorithm addresses the pressing need for healthcare equity. Now, residents of isolated or medically underdeveloped areas can obtain professional diagnosis for eye diseases without having to travel far or endure lengthy wait times. To foster trust between patients and medical practitioners, the interpretability of the algorithm is critical. A diagnostic provides a clear explanation for the algorithm's decision-making process, going beyond simple prediction-making. This not only makes medical professionals' decisions about therapy and follow-up care easier, but it also helps people better comprehend their diseases. Thus, telemedicine in conjunction with the ViT Algorithm holds the potential to revolutionize the identification of eye diseases by rendering it more user-friendly, effective, and patient-focused, ultimately leading to improved medical results.

6. CONCLUSION

In conclusion, the ViT Algorithm emerges as a promising and innovative solution to the pressing challenge of detecting and classifying eye diseases accurately. Through the use of AI, ML, and contemporary imaging technology, this system demonstrates exceptional performance in recognizing a variety of ocular illnesses. The use of standard datasets underscores its adaptability and reliability, positioning it as a valuable tool in the realm of ophthalmic healthcare. The demonstrated efficacy of the ViT Algorithm in enhancing diagnostic accuracy and facilitating timely interventions represents a significant stride towards improving patient outcomes. This research contributes a robust and precise approach to disease detection, offering a potential paradigm shift in the way eye diseases are diagnosed and managed, ultimately advancing the field and benefitting global healthcare systems.

REFERENCES

[1] 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.

[2] 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.

[3] 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
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] 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.
[16] M. Londhe, “Classification of Eye Diseases using Hybrid CNN-RNN Models MSc Research Project Data Analytics,” 2021.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] 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.