Extracting the Trustworthy Glaucoma Features using WGMO based EvoTransform: Advanced Vision Transformer from Fundus Images

Archana E.[1], Geetha S.[1], Victo Sudha George G.[1,*]

[1]Department of Computer Science and Engineering , Dr.M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India

Emails:archana.athi@gmail.com; geethasoman@drmgrdu.ac.in; victosudhageorge@drmgrdu.ac.in

Abstract

Glaucoma is a dangerous eye illness that greatly reduces the sharpness of a person's vision. If not caught early enough, this retinal disorder can damage the optic nerve head (ONH) and cause permanent blindness. Automated glaucoma diagnosis now has tool support thanks to recent advances in deep learning besides the convenience of computing resources. The low reliability of generic convolutional neural networks has prevented their widespread usage in medical procedures, even if deep learning has improved illness diagnosis using medical pictures. While there has been a rise in the use of deep learning for glaucoma classification, very few studies have tested whether or not the models are easy to understand and interpret, which bodes well for their future use. Medical picture feature extraction using Vision Transformers is showcased in this study utilising an EvoTransform: Advanced Evolutionary Algorithm Integration in Transformer Networks named as (EvoTAEA). Combining the powers of Convolutional Neural Networks with Vision Transformers, the suggested EAT Former architecture takes advantage of their data pattern recognition in addition adaptability capabilities. The classification accuracy is enhanced by using the Wild Geese Migration Optimizer (WGMO) to fine-tune the parameters of the proposed feature extraction. The design makes use of new parts, such as the Multi-Scale Region Aggregation, Global and Local Interaction, and Enhanced EA based Transformer blocks with Feed-Forward networks. For dynamically simulating non-standard places, it also presents the Modulated Deformable MSA module. Important components of the Vision Transformer (ViT) model are covered in the study, including patch-based processing, Multi-Head Attention mechanism, and positional context inclusion. In order to give an inductive bias, it presents the Multi-Scale Region Aggregation module, which combines data from several receptive fields. The MSA-based global module is improved by the Global and Local Interaction module, which adds a local path for extracting discriminative local info. An approach to glaucoma diagnosis that integrates ResNet-50, DenseNet-201, and Xception is suggested in the study. Two publicly available datasets, ORIGA and ACRIMA, are used to evaluate the trials. This model can help with the automated diagnosis of glaucoma using fundus pictures.

Keywords: EvoTransform: Advanced Evolutionary Algorithm; Wild Geese Migration Optimizer; Multi-Scale Region Aggregation; Fundus Images; DenseNet-201; Glaucoma detection

1. Introduction

1.1. Background of Eye and Eye Detection

As a part of the visual system, the eye is able to perceive visual inputs by reacting to light. A separate area of the brain called the optic nerve is responsible for the effective transmission of visual information from the cortex [1]. Elevated intraocular pressure has contributed to the alarming rise in the incidence of glaucoma, a common eye disease. Damage to the optic nerve can lead to glaucoma and other eye diseases by obstructing blood flow to the eye. In glaucoma, an eye condition, elevated intraocular pressure (IOP) leads to a gradual degeneration of the optic nerve. Worldwide, 111.8 million cases are expected by 2040, with 47% of patients being Asian and 87% of Angle Closure Glaucoma patients being of Asian heritage [2]. People over the age of 60 are at a higher risk of developing the illness [3]. Early detection is crucial in preventing permanent vision loss and structural damage caused by glaucoma, the second leading cause of visual impairment globally.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Both open-angle besides angle-closure glaucoma are subtypes of the eye disease glaucoma. Open-angle glaucoma, which frequently manifests with no visible symptoms, accounts for nearly 90% of all glaucoma cases. A serious eye disease, angle-closure glaucoma demands rapid medical treatment. Visual impairment, reddening of the eyes, inflammation of the eye, headaches, and increased intraocular pressure are all possible outcomes. Pharmaceutical interventions and surgical procedures are among the therapeutic possibilities [4]. Tonometry, ophthalmoscopy, perimetry, gonioscopy, and pachymetry are the five conventional glaucoma tests that are typically performed during a normal examination. Tonometry analyses the pressure within the eye, ophthalmoscopy looks at the shape and colour of the optic nerve, perimetry measures the area of vision, gonioscopy checks the angle of the eye, and pachymetry measures cornea.

1.2. Background of Artificial Intelligence Models

When analysing fundus photographs, computer-aided applications can help diagnose glaucoma. Medical imaging has benefited greatly from the advancements in AI and DL, which have allowed for more efficient and effective solutions [5]. Having said that, the current literature is inadequate and faces a number of obstacles. For example, the currently available annotated detection exhibits insufficient coverage of the diverse populations and severity levels encountered in clinical practice.It might be challenging to construct reliable and applicable models with little datasets. Problems with training models can arise from datasets that have an imbalance between normal and glaucomatous cases [6]. Many people also have trust concerns when it comes to these kinds of apps because of the opaque nature of the models they utilise. As a result, there is a growing need to boost trust in computational models by making sure that physicians can understand and use them, and that decision-making is transparent [7]. It is also possible that current models will not make full use of clinical domain expertise. In order to incorporate expert views into the creation and interpretation of models, it is necessary for machine learning researchers to collaborate with ophthalmologists. Problems with computing efficiency, model size, and interaction with current clinical workflows must also be addressed in order to deploy glaucoma detection settings.

1.3. Introduction of Vision Transformer

The convolution processes of Convolutional Neural Networks (CNNs) can be tuned to match the unique properties of the data, and CNNs already have a natural tendency to spot patterns in the data. Combining inductive bias with adaptability allows CNNs to perform exceptionally well in computer vision applications [9]. In a similar vein, Vision Transformers have been a game-changer for computer vision visual jobs. Many practical uses for these networks have recently emerged, including in the fields of autonomous driving, object identification, healthcare, and defence [10]. In this investigation, we zero in on techniques for medical image categorization. Data-Efficient Image Transformer (DeiT), resilient vision transformer (MedViT), localvit, and Swin Vision Transformer are the four specific examples of Vision Transformers used. Xception, ResNet, DenseNet, MobileNet, EfficientNet, and Inception are the six convolutional neural networks that serve as the foundational models.

1.4. Problem Statement

For feature extraction and classification, previous glaucoma classification efforts have mixed handcrafted features with more conventional learning models. The usefulness of these methods in accurately identifying glaucoma detection is, however, diminished due to their limitations [11]. When using handcrafted features in conjunction with older machine learning methods, there is a potential drawback. Because they are handcrafted using domain expertise, these traits could fall short when it comes to detecting glaucoma, which can have a wide range of intricate variations. The finer points and nuanced variations among leaf species may be too much for them to depict correctly [12]. Moreover, conventional ML algorithms could have trouble understanding hierarchical representations and intricate patterns from these handcrafted data, leading to subpar classification performance. However, deep learning models have shown promising results in a number of picture categorization tasks [13]. Over fitting is a problem that deep learning models trained from starting with little datasets are susceptible to. They could miss some of the finer nuances and intricate spatial correlations included in glaucoma detection pictures.

1.5. Contribution of the Research Work

By comparing Vision Transformer to the widely used Evolutionary Algorithm (EA) and emphasising their shared mathematical formulation, this research aims to overcome these constraints and demonstrate Vision Transformer's usefulness. The paper presents a new pyramid EvoTAEA backbone that uses the proposed EA-based Transformer (EAT) block entirely, building on the success of previous EA variations. Modules for make up the EAT block's three upgraded sections. The purpose of these parts is to gather data on an individual, multi-scale, interactive level on their own. Moreover, the research creates a Task-Related Head (TRH) for the transformer backbone, which allows for flexible data fusion in the final stage. In addition, we improve the method by including a Modulated Deformable MSA (MD-MSA) module that allows for the dynamic modelling of non-standard sites. Inspired by the cooperative behaviour of swarms of wild geese migrating great distances, WGMO fine-tunes the parameters

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

of the suggested feature extraction model. The algorithm's search mechanism, which strikes a good balance between exploration and exploitation in the search space, is constructed using randomly established migration groups, synchronous migration, and free foraging.

Xception, ResNet-50, and DenseNet-201 are the three components of the fused network that do the categorization. To improve the accuracy of plant leaf classification, each model has its own set of advantages. ResNet-50's deep architecture allows for the extraction of multi-scale characteristics from glaucoma images, which are rich and detailed. Learning complicated hierarchies and patterns is made easier using ResNet-50's deep layers and residual connections. Through thick connections between layers, DenseNet-201 improves feature propagation. Because of these linkages, gradient flow and representation learning are both enhanced. Images of glaucoma can be captured with complicated fluctuations and fine-grained details by DenseNet-201. By making use of depth-wise separable convolutions, Xception models are able to reduce computational complexity without sacrificing the ability to capture fine-grained features. Because of this, Xception is a good fit for glaucoma classification tasks, which need the capture of complex structures and features.

1.6. Organization of the Work

Section 2 mentions the related works; the brief explanation of the projected model with its mathematical expression is given in Section 3; the result analysis besides its discussion is given in Section 4 and lastly, the conclusion of the work is shown in Section 5.

2. Related works

Using visualisation approaches to make the results easier to understand, Shyamalee et al. [14] demonstrates state-of-the-art deep learning methods to segment besides categorise fundus images in order to forecast glaucoma symptoms. Our segmentation and classification predictions are grounded in U-Net with attention mechanisms with ResNet50 and a modified Inception V3 architecture, respectively. On the RIM-ONE dataset, Attention U-Net achieved 99.58% accuracy for optic disc segmentation and 98.05% accuracy for optic cup segmentation using a modified ResNet50 backbone, respectively. In addition, we use Gradient-weighted Class Activation Mapping (Grad-CAM) and Grad- that show which areas affected the glaucoma diagnosis. Working with the RIM-ONE dataset, our model for segmented image classification achieves compassion of 95.59%, specificity of 99.42%, and accuracy of 98.97%. This model can aid automated glaucoma diagnosis utilising fundus pictures.

For accurate glaucoma diagnosis from retinal fundus pictures, Das et al. [15] presented CDAM-Net, an architecture. In order to extract class-specific features from fundus images, we also provide an attention module named channel shuffle dual attention (CSDA). This module consists of a channel attention block, a spatial attention block, and a channel shuffle unit. To facilitate the extraction of multi-scale characteristics from fundus pictures, the CDAM-Net is primarily composed of MFR blocks. A CSDA module follows each MFR block, which helps to further enrich the feature representation. When tested on a retinal fundus image (RFI) dataset with 1426 fundus images (837 with glaucoma and 589 without), CDAM-Net outperforms state-of-the-art methods in terms of classification accuracy. In addition, the efficacy of the CDAM-Net's individual components is evaluated using ablation experiments.

Using three primary steps—image acquisition, pre-processing, and classification—Nugraha et al. [16] performs automatic feature extraction of the retinal rim using a machine learning approach. For the normal eyes dataset, we utilised RIM-ONE, and for the glaucoma images, we utilised DRISTHI-GS. There were 154 photos that were classified; 80 of those were for glaucoma and 74 were for normal eyes. Automatic extraction and classification were tested for sensitivity, specificity, and accuracy with respect to TPM, FNP, FNP, and TNL. The top three results are 96.20%, 98.75%, and 93.24%. The results of this study demonstrated the feasibility, accuracy, and significance of using automated texture characteristics and detection.

Govindan et al., [17] that analyses fundus pictures for early-stage glaucoma using unique architectural designs, have introduced a convolutional neural network (CNN) perfect. The REFUGE, Structured Analysis of the Retina (STARE), and Online Retinal Fundus Image Database for Glaucoma Analysis and Research (ORIGA) are some of the publically available datasets used in this investigation. Models such as InceptionV3, ResNet50, AlexNet, and VGG16 are trained on retinal fundus pictures in order to identify cases of glaucoma. A hybrid model was created by merging the ResNet50 and InceptionV3 models, which both showed better performance. While the STARE dataset attained a greater level of accuracy with a score of 99.1%, the ORIGA dataset attained a score of 97.4%. An F1 Score of 99.2% was also demonstrated by the REFUGE dataset, indicating outstanding performance. Ophthalmologists and other medical professionals can now benefit from a dependable glaucoma diagnostic system according to the suggested technique, which allows for more precise mass screenings and diagnoses of the disease.

Using cutting-edge soft-computing algorithms—the Whale Optimisation Algorithm (WOA)—Singh et al., [18] has introduced a fresh and efficient approach. One of our unique contributions to the scientific community is the hybrid version (hGWWO) of these two methods. Prior work in feature selection across several domains has made

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

use of the aforementioned baseline techniques. Our best guess is that the ORIGA public dataset is where these three algorithms are making their debut in the field of glaucoma detection. This proposed methodology is centred on the treatment of glaucoma, which is becoming more common around the world. When it comes to the causes of blindness, this disease is second only to cataracts. Medical imaging specialists can diagnose glaucoma by reviewing retinal images. To confirm this illness, trained ophthalmologists must perform manual eye screenings and use retinal fundus imaging. Because of its intricacy, reliance on trained personnel and potential for bias in results, screening analysis is not a quick process. This current effort to confirm this condition from retinal fundus images makes use of an artificial computer-aided clinical decision support system (CA-CDSS) to help the medical community deal with these problems. To classify the fundus retinal pictures under inquiry, ML models are used for classification and computational methodologies inspired by nature are used for feature selection. 65 characteristics were extracted from the ORIGA dataset. Using three FS approaches based on soft computing, we extract a subset of the most important characteristics dataset. Using a 70:30 evaluation strategy, ML classifiers are trained on this subset of the data. With a specificity of 0.981, sensitivity of 0.992, precision of 0.969, and an F1-score of 0.982, the proposed approach achieved 96.8% accuracy. New initiatives that benefit the public, researchers, and ophthalmologists are highlighted in this paper.

Gao et al. [19] applied deep learning (DL) techniques, namely the YOLOv7 architecture, to automatically distinguish the optic disc and optic cup in fundus pictures and compute VCDR.. The system was resilient and accurate. In addition, we look at the seldom-discussed problem of training a DL model on one population (say, Europeans) and then using it to VCDR estimate in another. Following our model's training on 10 publicly accessible datasets, we fine-tuned it using the REFUGE dataset, which comprises images of Chinese patients. In compared to human expert judgements, the DL-derived VCDR displayed excellent precision, with a Pearson correlation value of 0.91 and a mean absolute error (MAE) of 0.0347. By showing lower MAEs and better Dice similarity coefficients, our models outperformed previous methods on the REFUGE dataset. In addition, we came up with an optimisation method that can adjust DL outcomes for different populations. Clinicians now have a potential tool at their disposal thanks to our innovative methods for detecting optic discs and optic cups and calculating VCDR. This technology increases speed and accuracy while significantly lowering the human burden involved with image assessment. This automated technique is an excellent tool for glaucoma detection since it can effectively discriminate between glaucoma and Sheraz et al., [20], have proposed non-glaucoma patients.A two-stage network for automatic glaucoma diagnosis using fundus pictures. Step 1 involves using Yolo-v4 to find and extract the optic disc from a retinal fundus picture; step two involves using ResNet-101 to identify glaucomatous or healthy discs. Disappointingly, disc localization requires bounding box ground truth, which is not included in any of the freely available retinal fundus image datasets. To that end, we have presented a method for creating ground truth that is semi-automatic; this will provide the necessary annotations for training the Yolo-v4 based model that will allow autonomous disc localization. The ORIGA dataset, which is accessible to the public, is used to test the suggested method. With an accuracy of 87.4%, precision of 89.79%, and recall of 88.7%, the recommended automated OD localization performed admirably. Alternatively, the suggested glaucoma diagnostic module achieved respectable outcomes, achieving an AUC of.920 and an accuracy of 88.5%.

3. Proposed Method

The main goal of this research study is to extract the relevant information from the fundus images and classifies it using deep learning techniques. Figure 1 shows the process of the research work. Where each block is explained in this section in detailed.

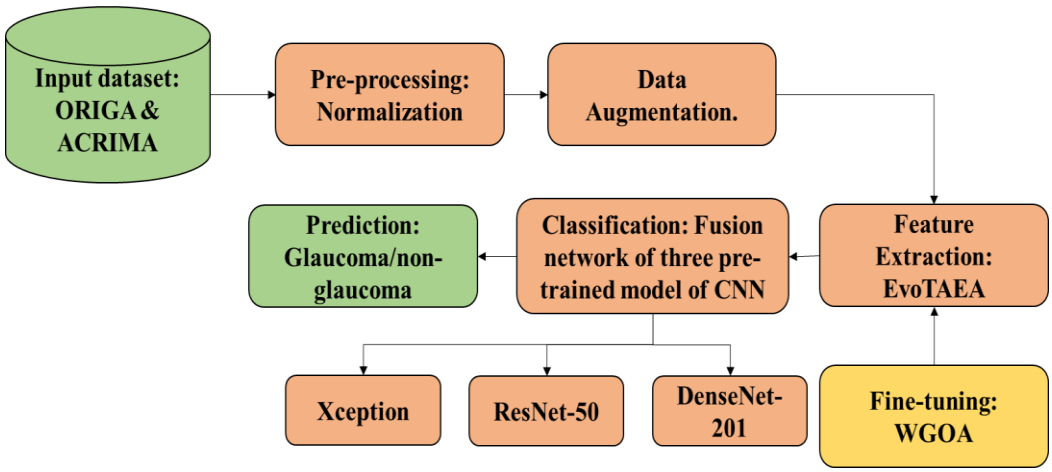

Figure 1. Workflow of the Research work

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

3.1 Dataset Details

This work takes into account two distinct fundus image datasets, namely the ACRIMA [21] and ORIGA [22] datasets, in an effort to offer a generalised solution. The images can be downloaded from https://www.kaggle.com/datasets/sshikamaru/glaucoma-detection?select=glaucoma.csv. As you can see from Table 1, the datasets are summarised. Among the relevant studies that have made use of these datasets, ORIGA stands out as the most popular.

Table 1: Summary of fundus image datasets

Dataset	Images
ACRIMA	705
ORIGA	650
Total	1355

705 fundus images were compiled from 396 glaucomatous photos besides 309 normal images in the ACRIMA dataset, which is another publicly available dataset [21]. With a combined eight years of experience, two glaucoma experts have annotated every image in the ACRIMA database. The ACRIMA database also does not have optic cup masks or segmented optic disc masks. Hence, we made use of www.apeer.com to make masks.

The ORIGA database contains 650 fundus photos, 482 of which are non-glaucomatous and 168 of which are glaucomatous. These images were retrieved from the Singapore Malay Eye Study (SiMES) [22]. All of the optic cup and disc mask pictures make up the ORIGA dataset.

At first, we alienated the dataset into three parts: training, testing, besides validation, with each part receiving 70% of the total. To maintain the class balancing property, we used distinct augmentation methods for the glaucoma and healthy classes, taking into account the quantity of photos in each. Table 2 shows how many pictures were utilised for training, testing, and validation for each dataset.

Table 2: Summary of training, validation, besides test imageries of used datasets.

Dataset	Training Image	Validation Images	Testing Images
ACRIMA	3193	718	724
ORIGA	775	228	231

3.2 Data Preprocessing

Instead of storing the whole dataset in memory all at once, the researchers in this study used Tensor Flow's dataset API, which provides an easy way to construct data pipelines. The researchers greatly improved their data handling and processing capabilities by utilising this API. Researchers made use of the Keras library's utility functions, known as Keras helpers, to load the data into the pipeline. Thanks to these tools, they were able to import the data from their specified folder and arrange it in a dataset. There was also a logical separation of the dataset into positive (glaucoma) and negative (non-glaucoma) categories.

All of the pictures in the dataset were randomly shuffled to make sure there was no bias when training. The data is more robustly trained and the model is prevented from learning biassed connections based on the input sequence if any inherent order or patterns are broken up in this step. The researchers split the dataset into 32 image batches to make training more efficient. The model's needs and the available computing resources will determine the optimal batch size.

The image pixels' values were also changed from "0 to 255" to "0 to 1." This process of scaling is used for two main reasons:

Numerical Stability: If you want your calculations or algorithm results to be more stable numerically, try working with pixel values between 0 and 1. As a result, problems like overflow and underflow are less likely to occur while working with higher pixel values.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Normalization: One method of normalisation is to scale the pixel values to a range of 0 to 1. To make comparisons and analyses easier, normalising the pixel values makes sure the data is consistently scaled across all images or channels. The model's ability to learn from the data and produce accurate predictions can be enhanced by doing this normalisation phase.

3.3 Data Augmentation

Therefore, in order to make sure the model worked with the original image collection, the pixel dimensions were modified to 224x224 pixels for the validation and training images during pre-processing and testing. After that, the photos were cropped and turned into grayscale. In addition, a zooming factor of 0.035 and a rotation range of 0.025 were employed for data augmentation; post-cropping, contrast-limited adaptive histogram equalisation (CLAHE) was also applied in order to minimise overfitting caused by the small data sizes in publically available datasets. So that they would still resemble fundus images, they were not turned upside down.

3.4 Feature Extraction using EvoTransform: Advanced Evolutionary Algorithm Integration in Transformer Networks named as (EvoTAEA)

The EvoTAEA architecture consists of four stages with varying resolutions, following the hierarchical structure [26]. It incorporates EAT blocks, which include three mixed paradigm y = f(x) + x residuals: (a) Feed-Forward Network (FFN), (b) Global and Local Interaction (GLI), and (c) Multi-Scale Region Aggregation (MSRA) modules. The down-sampling procedure between two stages is achieved using MSRA with a stride greater than 1. Additionally, we introduce a novel Modulated Deformable MSA (MDMSA) for enhanced global modelling and a TaskRelated Head (TRH) that provides a more elegant and flexible approach to the classification task.

3.4.1. Overview of Vision Transformer

The ViT model [24] processes input photos successfully using a systematic approach. At its core, it uses a method that involves segmenting each input image into a number of fixed-length, nonoverlapping patches. The next step is to apply a trainable linear projection layer to these patches; this layer will progressively increase and decrease the embedding dimension. The goal of this procedure is to get useful information out of each patch and make sure it works with everything that comes after it. One unique aspect of ViT's architecture is the incorporation of the positional context of each patch. To do this, distorted patches have positional embeddings applied to them. In order to provide background information on the absolute and relative locations of patches in a picture, position embeddings are crucial. The ViT model to improve its comprehension of the spatial relationships in the input data uses the positional context. Additionally, CLS is an extra classification token that the ViT model incorporates into the embedded patches. Important for subsequent operations, especially picture categorization, this token represents the full picture.

The CLS token, along with the patches, are thereafter inputted into an encoder to undergo transformation. Multi-Head Attention (MSA) and multi-layer perceptron (MLP) layers make up the Transformer encoder. The embedded patches are partitioned into several heads using the Multi-Head Attention method, allowing each head to learn its own unique state of self-attention. This makes it easier to capture different parts of the input data and their relationships. After that, an MLP layer receives the combined outputs from all the heads and processes them further. After each MLP operation, normalisation layers are performed to maintain stable information flow. Residual blocks are each operation. These layers enhance the robustness and efficiency of the model as a whole. The ViT model takes a holistic view by utilising a mix of MSA and MLP operations inside the Transformer encoder, patch-based processing, and positional context inclusion. By using this method, the model is able to analyse images efficiently, which in turn allows for the accurate execution of tasks like image categorization. After entering a series of inputs u into the Transformer encoder, the following outputs v are computed:

$$\hat{u} = u + MSA(Norm(u)) \quad (1)$$

$$v = \hat{u} + MLP(Norm(\hat{u})) \quad (2)$$

where Multi-Head Self-Attention (MSA), Norm denotes layer normalisation, MLP stands for multilayer perceptron. In its last prediction step, ViT employs the CLS following the transformer encoder.

A.) Multi-Scale Region Aggregation

We bring the idea of using numerous populations and diverse searching regions to enhance model performance to 2D image analysis, which is influenced by some evolutionary algorithm (EA) methods. We present a new module named Multi-Scale Region Aggregation (MSRA) as part of our research on a vision transformer. The model uses N local convolution procedures. $(Conv\ Sn, 1 \leq n \leq N)$ Walking at different speeds. These techniques successfully give an inductive bias without further position embedding procedures by combining data from several receptive fields. If we want to describe it mathematically, this is the n-th dilation operation on that changes the map x:

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

$o_n(x) = ConvS_n(Norm(x))$ (3)

s.t. $n = 1, 2, \ldots, N$,

We present the Weighted Operation Mixing (WOM) mechanism, which can be learned to mix all operations in a weighted way using a set of learnable weights and a softmax function. $\alpha_1, \ldots, \alpha_N$. Because of applying function F, the middle illustration xo can be obtained as shadows:

$x_0 = \sum_{n=1}^{N} \frac{\exp(a_n)}{\sum_{n'=1}^{N} \exp(a'_n)} o_n(x)$ (4)

F denotes the addition function in the provided formula. Concatenation is one of the fusion functions available; it produces better results but requires more parameters. The paper chooses to utilise the addition function by default for simplicity's sake. Later on, a convolutional layer, ($Conv_{\mathrm{So}}$), is employed to map representation, x_{o}, to the same number of channels as the input, x. To get the module's output, relative connections are utilised. The MSRA module, which is both the stem and the embedding patch, adds to the EvoTAEA's sleekness and consistency. This work forgoes position embedding in favour of a CNN-based MSRA, since the GLI module supplies an intrinsic inductive bias.

B.) Global and Local Interaction

The study suggests a new module called the Global and Local Interaction (GLI) module, which would improve the MSA-based global module. This is in response to the need for faster and more effective convergence of high-quality solutions, which was inspired by the implementation of local search procedures in EA variants. In addition to the global route, the GLI module also has a local path. Following the earlier-mentioned idea of a local population, the local path seeks to discriminative info pertinent to locales, whereas the global path continues to centre on modelling global information. Local features and global features are the two types of input features that are classified at the channel level. In order to make feature interactions possible, the global and local routes process these features independently. The study uses the suggested Weighted Operation Mixing method, which modifies the weights given to the local branch (α_l) and the global branch (α_g), to make sure that both paths contribute equally. We mix the outputs from these two methods to get data with the original dimensions. As an enhanced module for the existing transformer structure, this combination can be seen as a versatile plug- operation, designated as H. Keep in mind that there are a few different ways the local operation can be implemented, such as using a local MSA or a conventional convolutional layer. In contrast, MSA, D-MSA, Performer, and comparable methods can be employed by the worldwide operation.

The GLI module is based on naive convolution with MSA modules, which are chosen for this research work. By going with this option, GLI keeps its global modelling capabilities intact while improving their local ones. It is worth mentioning that, just like the vanilla ViT, our suggested structure maintains efficiency and parallelism by keeping the maximum path length between any two places at O(1). The model's efficacy and efficiency are greatly affected by the feature separation ratio (p) that is used. Model accuracy, floating-point operations per parameter (FLOPs), and ratio are all affected by the chosen ratio.

To give you a rundown, the local path is a collection of feature maps in the dimensions that are calculated point-wise. of $R^{C\times H\times W} = R^{C\times L}$, but depth-wise convolutions of size k × k are involved in both pathways. Given that there are a certain amount of global channels, $C_g = p \times C$ and the channels is $C_l = C - C_g$. The following is a study of the computational characteristics and parameter count of the upgraded GLI module:

$Params = 5C_g^2 + (2 - 2C - K^2)C_g + (k^2 + 2 + C)C$ (5)

C.) Modulated Deformable MSA

Modulated Deformable MSA (MD-MSA) is a new module that we have improved based on the distribution seen in real people. This module considers positional fine-tuning and re-weighting of each spatial patch. The QKV features are retrieved from the input feature map X using the function fqkv() in the naïve MSA method. Expressed as $QKV = f_{qkv}(X)$, where $f_{qkv} = f_q \oplus f_k \oplus f_v \oplus$ is a merging the functions f_q, f_k, and f_v in the concatenation stage. In contrast, changes are implemented by the MD-MSA method. In order to extract K and V features, the query-aware access to the map $\hat{X}$ is what distinguishes the proposed MDMSA from the original MSA. If we take an example feature map X with L places as input, we can see that the function f_{q} is practical to find the query matrix Q, denoted as $Q = f_q(X)$. This query predict deformable offsets Δl and scalars Δm for all sites in order to admission the feature map Δm:

$\Delta l, \Delta m = f_{md}(Q)$ (6)

Using the l-th position, we compute feature $\hat{X}_l$ as shadows:

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

$\hat{X}_l = S(X_l, \Delta l).\Delta m$ (7)

where Δl stands for the relative coordinate which has an unbounded range for the l-th position, whereas Δm is limited to the interval (0, 1). The bilinear interpolation function is represented by the symbol S. As a result, the new feature map map is used to calculate KV. $\hat{X}$, denoted as $KV = f_{kv}(\hat{X})$. By considering a number of positional factors, the MD-MSA method maximises performance. Algorithm 1 explains the description of proposed model.

Algorithm 1: EvoTAEA Feature Extraction

Inputs:

Input image I

Number of evolutionary algorithm iterations T

Number of wild geese N

Number of head geese M

Outputs:

Extracted features from EvoTAEA

Step 1: Image Tokenization

Divide Input Image into Patches:

Divide I into non-overlapping patches.

Each patch has a fixed length.

Linear Projection:

Apply a trainable linear projection to each patch.

Expand and contract the embedding dimension to extract meaningful information.

Step 2: Incorporate Positional Embeddings

Attach Positional Embeddings:

Attach positional embeddings to the patches.

Incorporate contextual information about the relative and absolute positions.

Add Classification Token (CLS):

Integrate a classification token (CLS) to the embedded patches.

Step 3: Transformer Encoder

Input Sequence:

Feed the sequence of patches and CLS token into the transformer encoder.

Multi-Head Self-Attention (MSA):

Compute attention weights and context vectors using multi-head self-attention.

Multi-Layer Perceptron (MLP):

Process the outputs of MSA through MLP layers.

Apply normalization layers and residual connections.

Layer Operations:

$\hat{u} = u + MSA(Norm(u))$ (1)

$v = \hat{u} + MLP(Norm(\hat{u}))$ (2)

Step 4: Multi-Scale Region Aggregation (MSRA)

Apply Local Convolution Operations:

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Use NNN local convolution operations with varying strides.

Weighted Operation Mixing (WOM):

Mix the operations using a weighted softmax function:

$x_0 = \sum_{n=1}^{N} \frac{\exp(a_n)}{\sum_{n'=1}^{N} \exp(a'_n)} o_n(x)$ (3)

Final Convolution Layer:

Apply a convolution layer to map the intermediate representation to the input's channel size.

Step 5: Global and Local Interaction (GLI)

Feature Separation:

Divide input features into local and global features.

Local and Global Paths:

Process local features using convolutions.

Process global features using MSA.

Weighted Operation Mixing:

Combine outputs from local and global paths using a softmax-weighted mixing mechanism.

Step 6: Modulated Deformable MSA (MDMSA)

Query Matrix Q:

Compute Q from the input feature map X:

$Q = f_q$ (4)

Predict Deformable Offsets and Modulation Scalars:

$$\Delta l, \Delta m = f_{md}(Q) \quad (5)$$

Resample and Reweight:

Calculate the resampled and reweighted feature:

$\hat{X}_l = S(X_l, \Delta l).\Delta m$ (6)

Compute KV:

Obtain KV from the new feature map $\hat{X}$:

$KV = f_{kv}(\hat{X})$ (7)

Step 7: Fine-Tuning using Wild Geese Migration Optimizer Algorithm (WGMOA)

Initialization:

Randomly generate the initial population of wild geese positions.

Formation of Migration Groups:

Establish migration groups with head geese at the center.

Synchronized Flight:

Update positions based on environmental information and head goose positions:

$x_i^{t+1} = x_i^t + c_1\left(x_{best}^t - x_j^t\right) + c_2(x_k^t - x_j^t)$ (8)

Free Foraging:

Allow random exploration while maintaining optimal position information:

$x_i^{t+1} = x_i^t + c_3\left(x_j^t - x_i^t + L\right) + c_4(x_{best}^t - x_j^t)$ (9)

Selection of Head Geese:

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Replace head geese with optimal individuals from each migration group.

Reduce the migration group radius L after each iteration:

$L = L, \left(1 - 0.1.\frac{t}{T}\right)$ (10)

Step 8: Final Feature Extraction

D). Fine-tuning using Wild Geese Migration Optimizer Algorithm (WGMOA)

The starting population of the WGMO algorithm is chosen at random from the solution space, with a specific sum of wild geese serving as the geese. The head geese serve as leaders for the flock of wild geese as they migrate. In the GMO algorithm, N is the population size of the wild geese and M is the head geese. The initial radius size for the migration group is set to L. ($L = ud\ -\ ld/N$).

D.1). Formation of Migration Groups.

With each cycle, the migrating groups are reformed based on where the head geese are located. Each group's members are spread at random within a radius L that centres on the head goose. It will help make the formation change and the replacement of the head geese a reality. This is the mathematical model:

$$\begin{cases} x_i^t = x_j^t & if\ i = b * (j-1) + 1, \\ x_i^t = x_j^t - L + 2L * rand(1, \dim), & else \end{cases} \quad (8)$$

Where x_i^t characterizes the site of the i-th separate at the t-th repetition ($i = 1, 2, \dots, N$). T is the determined sum of repetitions ($t = 1, 2, \dots, T$). x_j^t characterizes the site of the j-th head goose individual at the t-th iteration ($j = 1, 2, \dots, M$). b characterizes the sum of migration groups ($b = N/M$).

D.2). Synchronized Flight.

During their migration, wild geese rely on a combination of environmental cues, memories of past migrations, and their own flight experiences. As the head goose flies with the migration group, each member of the group stays in a relatively stationary position. The WGMO algorithm mimics the flying patterns of wild geese by using the synchronous flight strategy and setting all members of the migrating group to take the same sum of steps during their flight. The head goose relies on the optimal site and the site info of other head goose to update the individuals' positions in the migrating group. This is the mathematical model:

$x_i^{t+1} = x_i^t + c_1\left(x_{best}^t - x_j^t\right) + c_2\left(x_k^t - x_j^t\right)$ (9)

Where x_{best}^t characterizes the global optimal specific and x_k^t is the haphazardly selected head goose separate. x_i^t and x_j^t characterize group, correspondingly. The flight step size $c_1 \in [0, 1]$, and c_2 is calculated by

$$\begin{cases} c_2 = exp\frac{fit(j) - fit_{ave}}{fit_{worse} - fit_{best}} & fit(j) \le fit_{ave} \\ c_2 = exp\frac{fit(j) + fit_{ave} - 2fit_{best}}{fit_{worse} - fit_{best}} & fit(j) > fit_{ave} \end{cases} \quad (10)$$

where $fit(j)$ is the fitness charge of the head goose, fit_{worse}, fit_{ave}, and fit_{best} characterize the worst, average, and best fitness charge of the head geese, correspondingly, and c_2 is experience info. If $fit(j) \le fit_{ave}$, it indicates that the value of $fit(j)$ is small besides means that x_j^t is an exceptional head goose and learn supplementary info from additional goose. The exact reverse is true when $fit(j) > fit_{ave}$.

D.3.) Free Foraging.

During long-distance flights, migratory groups will inevitably rest and forage. Foraging areas for wild geese generally include lakes or other bigger bodies of water found in nature. The members of the migratory group will keep a specific link in a short region while randomly exploring according to the information provided by the head goose during the free foraging process. Simultaneously, group uses ideal location information to maintain the moving trend. The wild geese will reassemble and begin their migration once they have finished foraging. We can express this mathematically as follows:

$x_i^{t+1} = x_i^t + c_3\left(x_j^t - x_i^t + L\right) + c_4\left(x_{best}^t - x_j^t\right)$ (11)

where c_3 and c_4 are arbitrary integers between 0 and 1, used to regulate the magnitude of individuals' movement steps when foraging. To regulate how far the migratory group is from the leader geese, one uses the range, abbreviated as L.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

D.4.) Selection of the Head Geese.

The head geese, as the leaders of the whole swarm, are the most important individuals throughout the geese. To achieve high flight durability, it is necessary to replace the head geese often. Following each location update, the WGMO algorithm will select the best individuals from each migrating group to serve as the head geese of the next generation. The algorithm's capacity to strike a good balance among exploitation and exploration is greatly enhanced by this selection technique, which enables the head geese to convey excellent location info while simultaneously ensuring the dispersion of their places. Equation (12) is used to minimise the radius (L) after all the head geese have been replaced. Raising the group's membership density will boost the algorithm's exploration accuracy.

$$L = L * \left(1 - 0.1\left(\frac{t}{T}\right)\right) \quad (12)$$

Where T is the maximum sum of repetitions and t is the existing sum of repetitions.

D.5). Implementation of WGMO:

An emerging method for stochastic optimisation is the WGMO algorithm. The best solution is found by iteratively optimising all solutions from a set of randomly selected initial places inside the solution space.

D.6). Time Complexity.

An algorithm's computational efficiency and performance are of equal importance in real-world engineering applications. One of the most imperative ways to measure the algorithm's recital is with the time complexity analysis method. If the parameters N and T stay the same, this method can assess the algorithm's difficulty while precisely verifying the algorithm's computational efficiency. There are primarily three steps to the WGMO algorithm's calculating process: initialising the population (O(N)), establishing migration groups (O(N * T)), and then deciding whether to use synchronised flight or free foraging (O(N * T)). Hence, the GMO algorithm's temporal complexity is $O(WGMO) = O(N) + O(N * T) + O(N * T)$. The complexity formulation has no action and is mostly pretentious by the basic limit $N * T$. From the above study, it GMO procedure has a complexity.

E) Overall Obtained Features from the proposed model

This architecture combines Vision Transformers with evolutionary algorithm-inspired modules to extract meaningful and diverse image features. Local Texture and Shape Features are extracted by Patch-Based Embedding. The first stage of the EvoTAEA extracts local features by dividing an image into non-overlapping patches. Each patch undergoes a linear transformation, creating a feature vector for each patch. This process is similar to convolutional operations and is capable of capturing fine-grained details such as local textures, edges, and color gradients within the patches. Multi-Scale Structural Features, Edge Features, Texture Features are extracted by Multi-Scale Region Aggregation (MSRA). The MSRA module performs convolution with multiple kernel sizes (3x3, 5x5, and 7x7) to capture multi-scale information from the image. By aggregating information across different receptive fields, this module extracts. Structural features at different scales, which helps to detect both small details and larger patterns in the image. By using different scales of convolutions, MSRA captures edges and contours more effectively, which are critical for recognizing the shape of objects like the optic disc and cup in fundus images. It also captures fine-grained variations in texture, which is important in fundus images where subtle differences in texture can indicate abnormality.

The Global and Local Interaction (GLI) module is responsible for integrating both global and local information. The multi-head self-attention (MSA) mechanism models long-range dependencies across different patches of the image, capturing high-level global context such as the overall shape, structure, and spatial relationships between different regions of the image. The convolutional operation in the local path captures fine-grained discriminative details within smaller regions of the image, such as small anomalies or irregularities that might be missed by global processing alone. This combination allows the extraction of complex spatial relationships between different parts of the image, essential for detecting subtle signs of medical conditions like glaucoma. The Modulated Deformable Multi-Head Self-Attention (MD-MSA) module fine-tunes the locations where attention is applied by using modulated deformable convolutions. It dynamically adjusts the positions at which features are extracted, which makes it capable of capturing. In real-world medical images, abnormalities like lesions or optic nerve damage may not follow regular patterns. MD-MSA is designed to model these irregular locations, which are crucial for tasks such as detecting early signs of glaucoma. By fine-tuning the attention mechanism, this module ensures that spatial position is taken into account, improving the model's ability to localize features relative to the anatomy of the eye (e.g., position of the optic nerve, optic cup). The Feed-Forward Network (FFN) component applies fully connected layers to each patch token, helping to refine the extracted features. The FFN enhances the abstract features captured by the multi-scale, global, and local modules. These features are highly refined and represent complex patterns

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

such as composite shapes and textures. After passing through multiple layers, the model is able to combine low-level edge and texture features into more abstract representations that correspond to medical conditions. By transforming patch tokens individually, the FFN captures how different parts of the image relate to each other in a broader context. After feature extraction from multiple layers, mean pooling is applied over all the patch tokens. This summarizes the feature vectors across the image, producing a global feature vector that represents the overall content of the image, including high-level structures (e.g., optic disc, optic cup) in medical images.

Overall Features Extracted by EvoTAEA:

Low-Level Features:

Edges: Sharp transitions in pixel intensity, important for detecting boundaries of anatomical structures.

Texture: Variations in pixel intensity, useful for detecting tissue texture or irregularities in fundus images.

Local Shapes: Small structures like blood vessels, optic disc boundaries, and optic cup regions.

Mid-Level Features:

Contours and Geometric Patterns: Helps in recognizing larger anatomical structures.

Multi-Scale Structures: Combination of small and large-scale structures allows detection of abnormalities at different resolutions.

Positional and Spatial Relationships: Important for localizing features relative to specific regions of the eye (e.g., optic nerve location).

High-Level Abstract Features:

Complex Patterns: Representation of complex medical conditions such as the shape and ratio of the optic cup-to-disc (important for glaucoma detection).

Contextual Information: How different areas of the retina relate to each other in terms of disease progression or abnormalities.

The model is designed to detect glaucoma by extracting features such as the optic nerve head (ONH), optic cup, and cup-to-disc ratio (CDR). It captures both global structural information (size and shape of ONH) and local variations (thinning of retinal nerve fibers) which are critical for glaucoma diagnosis.

3.5 Classification using Fusion Model of pre-trained network

3.5.1. ResNet-50

ResNet-50 introduced the concept of residual connections to address the degradation problem of deep neural networks.A total of fifty layers make up the ResNet-50 architecture. These layers include pooling layers, convolutional connections. Instead of fitting the required underlying mapping directly, ResNet-50 incorporates residual blocks, which enable the functions. This innovation is the key of ResNet-50.

Two or three layers using batch normalisation and ReLU activation algorithms make up each outstanding block in ResNet-50. After those levels, a skip link goes straight from the residual block's input to its output. The issue of vanishing gradients is reduced by facilitating the gradient's flow during training through this skip connection.

Here is a mathematical representation of the skip connection:

$$y = \mathcal{F}(x, w_i) + x \quad (13)$$

where x characterizes the input to the remaining block, $\mathcal{F}(x, w_i)$ stood for the function applied by layers using weights w_i, while y stood for the residual block's output.

To address the issue of information deterioration in deep networks, ResNet-50 introduces these residual connections, which make the model deeper without sacrificing speed. Learning of more intricate and deep features is made possible by the skip connections, which improve the gradient's propagation across the network.

3.5.2. DenseNet-201

The highly linked layers and effective information flow characterize DenseNet-201. A number of dense blocks, each with numerous convolutional layers that are tightly coupled, make up the DenseNet-201 architecture. New to DenseNet-201 are dense connections, which enable direct connections between layers of varying depths, as opposed to the sequential flow of information in conventional CNN layers. The effective flow of info throughout the network is made possible by this dense connectivity, which also encourages feature reuse.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

The building blocks of DenseNet-201 are densely linked layers, which might include convolutional layers, batch normalisation, activation functions, and other similar components. In a single dense block, all of layers are joined with each layer's output. The mathematical expression for this operation of concatenation is

$x_l = [h_0, h_1, \dots, h_{l-1}]$ (14)

where x_l characterizes the output within block, and h_i represents the feature i-th layer.

To promote feature reuse and give the network access to a rich set of layer, DenseNet-201 densely connects the layers inside each dense block. As a result, the model becomes smaller and the vanishing gradient problem is reduced, making deep network training more efficient. In addition, DenseNet-201 uses transition layers to regulate the sum of feature are incorporated between dense blocks. In order to reduce data size and increase computing performance, these transition layers usually include a convolutional layer and a pooling layer.

3.5.3. Xception

Aiming to collect more spatial info and better feature illustration, Xception [25] is renowned for its novel tactic to the creation of convolutional layers. Separable convolutions that are depth-wise form the basis of the Xception architecture. These convolutions separate the spatial and channel-wise aspects of the operation. Xception improves efficiency and efficacy by introducing independent operations for each dimension, unlike standard layers that conduct convolutions over both dimensions concurrently.

Applying spatial convolutions separately for each channel besides then combining the results finished convolutions is the main notion underlying separable convolutions. It is possible to express the depth-wise separable convolution procedure mathematically as

$\Upsilon = DW(x) * \mathcal{P}W(Y')$ (15)

where X signifies the input maps, DW signifies the convolution operation, $\mathcal{P}W$ characterizes convolution process, Y' symbolizes the middle feature maps, besides Υ characterizes the output feature maps.

Xception drastically cuts down on computation and parameter requirements as compared to conventional convolutional layers by splitting the spatial and channel-wise processes. Better learning and representation capability are made possible by deeper network topologies with fewer parameters. Following in the footsteps of the ResNet architecture, Xception includes depth-wise separable convolutions as well as residual connections. In order to train extremely deep networks effectively, these connections let the network functions and deal with the problem. In addition, Xception has a modular design that includes pooling layers after each series of convolutional blocks. The network is able to pick up both simple and complex visual patterns thanks to this design's support for hierarchical feature extraction at various abstraction levels.

3.5.4. Fusion of Model Output

The categorization procedure continues with training the models and then merges their outputs. A concatenation layer can accomplish this by combining the results of several models into one larger tensor. A combined activation map from all the models is produced when the concatenation layer joins the separate along a given axis. Afterwards, the rectified linear function is used to this united representation. To increase the model's expressive power and enable the network to learn complicated patterns, the ReLU activation function adds unit. Lastly, glaucoma detection's classification layer uses the softmax activation function to get normalised likelihoods for every class.

4. Results and Discussion

The next part briefly discusses the experimental setup, assessment measures, and results analysis of the proposed model compared to existing methodologies.

4.1. Experimental Setup

A PC with an i7 processor, 16 GB of RAM, and a 6 GB Nvidia GeForce GTX 1060 GPU is required for the implementation. The Google Colab platform was used for all of the work. The proposed model was implemented using TensorFlow, which was in turn used in conjunction with Python 3.10 and TensorFlow version 2.12.

4.2. Evaluation Metrics

We assess the efficacy of our glaucoma forecast model using four metrics: specificity, sensitivity, accuracy, and precision.

Our focus now shifts to the confusion matrix, which offers a more comprehensive analysis of the model's forecasts. By examining the confusion matrix, we may ascertain the number of correct, incorrect, and misclassified results produced by our classification process. The sum of positive cases that were effectively foretold is called true

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

positives, while the sum of negative cases that were accurately predicted is called true negatives. Conversely, false positives and false negatives are measures of the proportion of positive and negative cases, respectively, that were incorrectly identified. An in-depth knowledge of the representation's prediction power and presentation in class distinction can be obtained by investigative the confusion matrix, where Figure 2 presents the confusion matrix of the proposed model and Figure 3 presents the accuracy and loss graph of the proposed model for training and testing data.

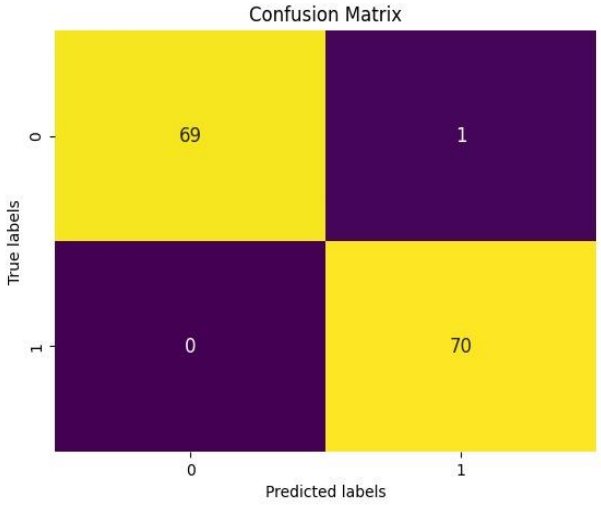

Figure 2. Confusion matrix

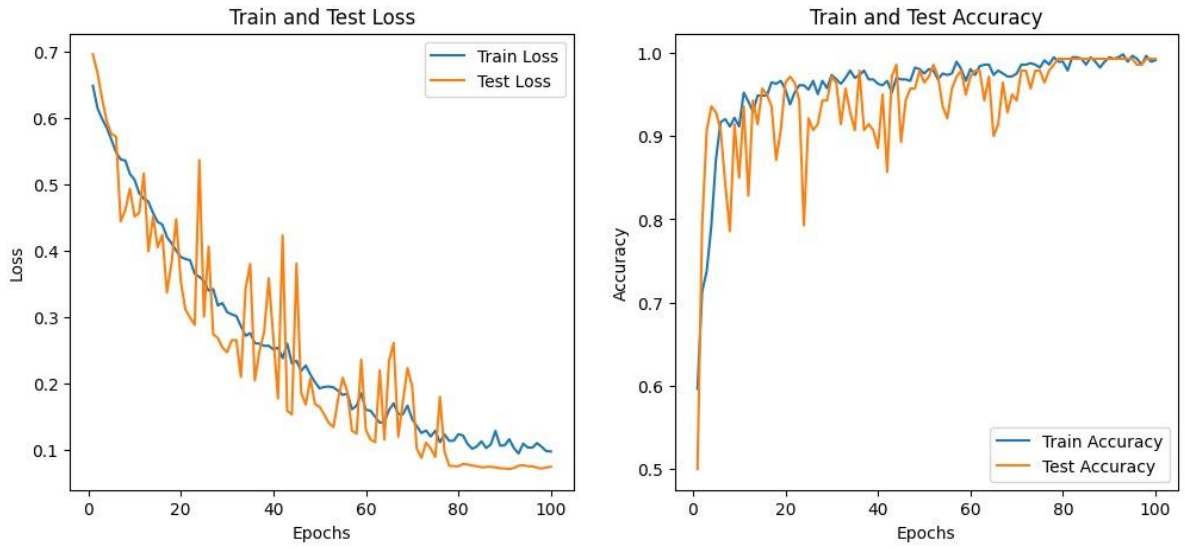

Figure 3. Accuracy and Loss of the proposed model

Next, we may build the assessment metrics mentioned earlier using the confusion matrix's results, which true positives, true negatives, false positives, besides false negatives. Some of the metrics that depend on the confusion matrix. The accuracy of our glaucoma prediction algorithm can be better understood by calculating the subsequent metrics: true positives, true negatives, false positives, besides false negatives. To fully grasp how well our algorithm detects and classifies glaucoma instances, let us move on to calculating and analysing these metrics.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (16)$$

$$Accuracy = \frac{TrueP\ ositives + TrueNegatives}{TotalPredictions} \quad (17)$$

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (18)$$

$$Specificity = \frac{TrueNegatives}{TrueP\ ositives + FalseNegatives} \quad (19)$$

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

4.3. Validation analysis of Proposed Feature Extraction

Table 3 and Figure 4 presents the different features affect the presentation of the projected feature extraction model by validating using the minimum and maximum number of features from the whole features list that is explained in the proposed section.

Table 3: Analysis of features on Proposed Model

Features percentage	Dataset	ACRIMA	ORIGA
15% (minimum)	Accuracy %	92.3	93.48
	Precision %	90.63	92.97
	Sensitivity %	91.18	93.45
	Specificity %	93.87	91.07
30% (maximum)	Accuracy %	96.24	98.91
	Precision %	91.69	97.09
	Sensitivity %	96.26	98.68
	Specificity %	97.21	97.14

Analysis of features on proposed Model as 15% features in the accuracy 92.3 also ORIGA technique of 93.48 correspondingly. Then the Precision of ACRIMA dataset achieved 90.63 also ORIGA dataset of 92.97 correspondingly. Then the Sensitivity of ACRIMA dataset as 91.18 also ORIGA dataset of 93.45 correspondingly. Then the Specificity of ACRIMA dataset as 93.87 also ORIGA dataset of 91.07 correspondingly. Then the 30% features of maximum from the whole features in the Accuracy of ACRIMA dataset as 96.24 also ORIGA dataset of 98.91 correspondingly. Then the Precision of ACRIMA dataset as 91.69 also ORIGA dataset of 97.09 correspondingly. Then the Sensitivity of ACRIMA dataset as 96.26 also ORIGA dataset of 98.68 correspondingly. Then the Specificity of ACRIMA dataset as 97.21 also ORIGA dataset of 97.14 correspondingly.

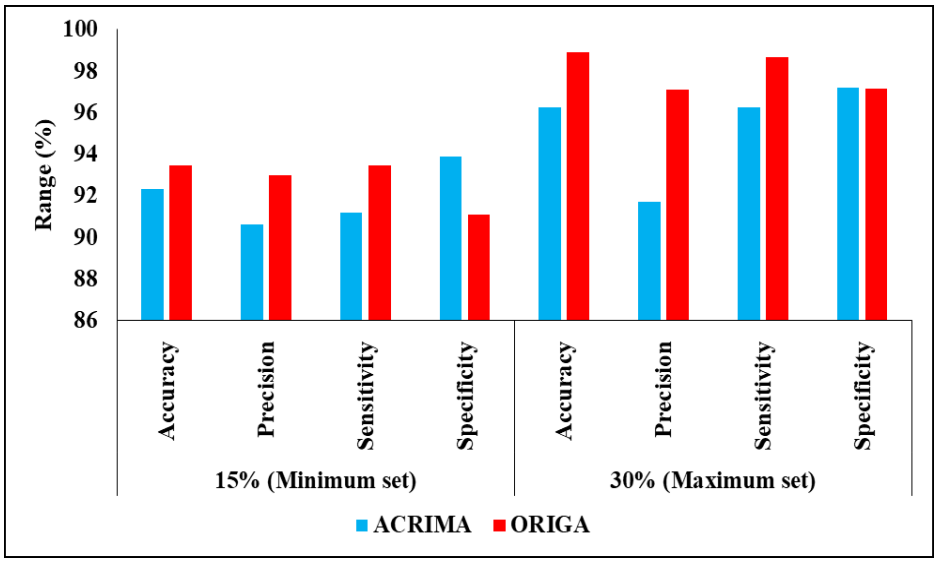

Figure 4. Visual Representation of Proposed Feature Extraction Model

Table 4 and Figure 5 represents the experimental analysis of Proposed WGMOA model for different learning rate on two datasets.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Table 4: Learning Rate Analysis of WGMOA

Learning Rate	**Dataset**	**ACRIMA**	**ORIGA**
0.001	Accuracy	90	95.8
	Precision	90.43	91.33
	Sensitivity	92.37	92.64
	Specificity	94.84	93.21
0.0001	Accuracy	95.3	97.7
	Precision	94.42	96.66
	Sensitivity	95.43	95.66
	Specificity	97.71	96.17

In the WGMOA Learning Rate Analysis, the ORIGA dataset's 95.8 corresponding accuracy and the ACRIMA dataset's <90 accuracy is both 0.001 under the learning rate condition. Next, the ORIGA dataset's precision of 91.33 corresponds to the ACRIMA dataset's 90.43 precision. Next, the ORIGA dataset's 92.64 corresponding sensitivity and the ACRIMA dataset's 92.37 sensitivity. Next, the ORIGA dataset's 93.21 Specificity and the ACRIMA dataset's 94.84 Specificity are correspondingly high. Next, under the learning rate condition, the ORIGA dataset's 97.7 and the ACRIMA dataset's 95.3 accuracy are both 0.0001. Next, the ORIGA dataset's precision was 96.66, while the ACRIMA dataset's precision was 94.42. Subsequently, the ORIGA technique's sensitivity was 95.66 and the ACRIMA dataset's sensitivity was 95.43. Next, the ORIGA dataset is 96.17 and the ACRIMA dataset's 97.71 specificity respectively.

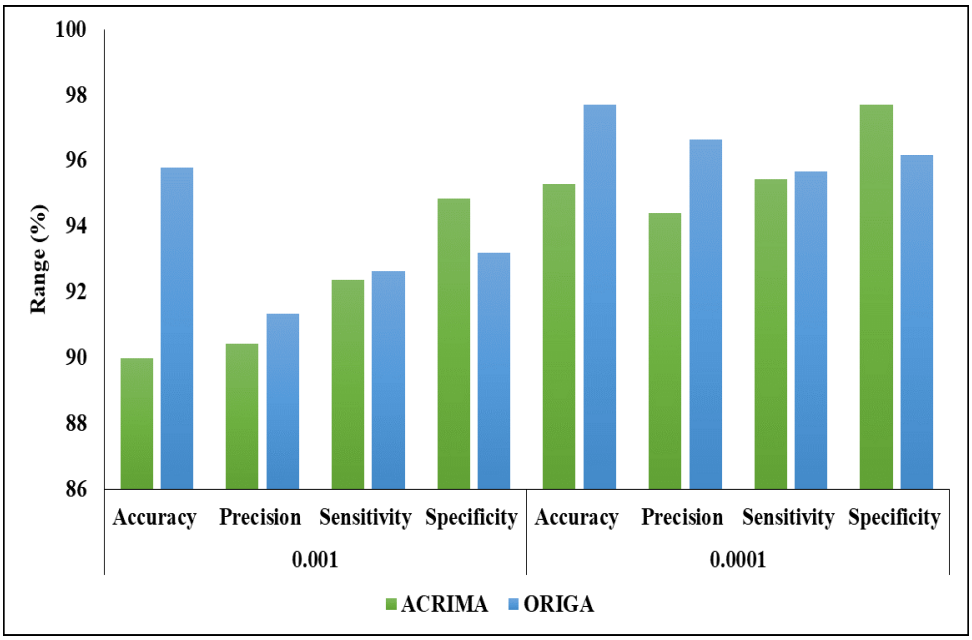

Figure 5. Graphical Representation of proposed optimizer on learning rate analysis

Table 5 besides Figure 6 shows the comparative analysis of projected FE and classifier with existing models on combined datasets. The existing models uses different datasets and therefore, it is implemented on our considered datasets and results are averaged.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

Table 5: Comparative Analysis of proposed model over existing techniques

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Inception V3 [14]	90.36	89.49	88.04	89.25
CSDA [15]	91.13	91.34	91.56	91.95
CNN [17]	91.69	92.98	90.65	90.81
ResNet-101 [20]	90.29	91.77	92.87	92.32
EvoTAEA	94.20	93.12	93.08	94.60
Fused Network	**94.25**	**93.95**	**93.59**	**94.27**

In the Comparative Analysis of proposed model over existing techniques study as Inception V3 [14] technique accuracy as 90.36 also precision of 89.49 also recall as 88.04 also f1-score as 89.25 correspondingly. Then the CSDA [15] technique accuracy as 91.13 also precision of 91.34 also recall as 91.56 also f1-score as 91.95 correspondingly. Then the CNN [17] technique accuracy as 91.69 also precision of 92.98 also recall as 90.65 also f1-score as 90.81 correspondingly. Then the ResNet-101 [20] technique accuracy as 90.29 also precision of 91.77 also recall as 92.87 also f1-score as 92.32 correspondingly. Then the EvoTAEA technique accuracy as 94.20 also precision of 93.12 also recall as 93.08 also f1-score as 94.60 correspondingly. Then the Fused Network technique accuracy as 94.25 also precision of 93.95 also recall as 93.59 also f1-score as 94.27 correspondingly.

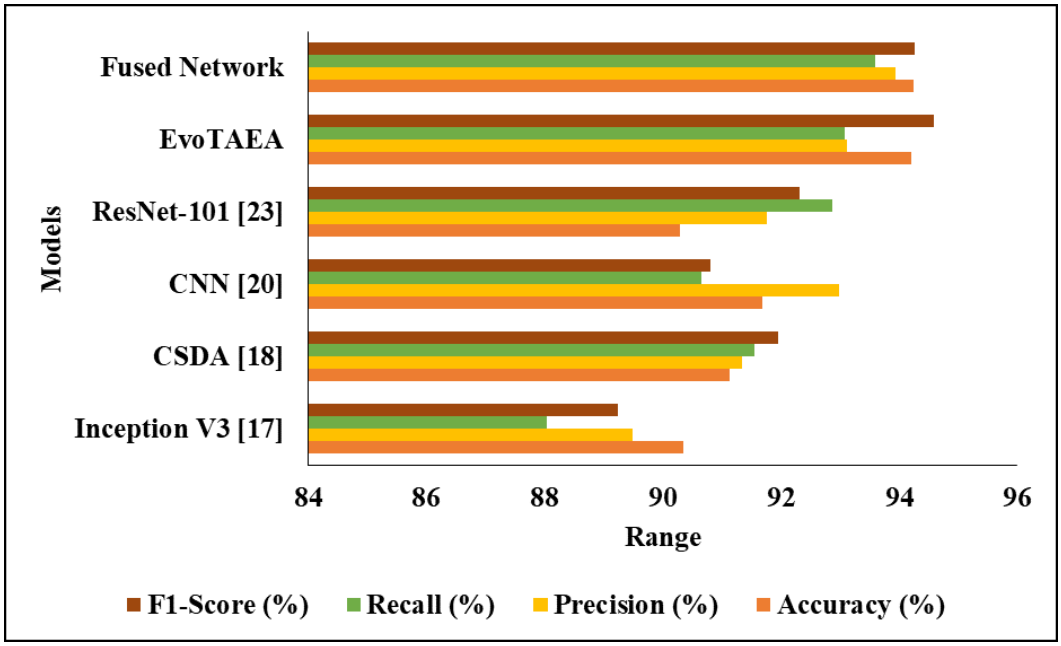

Figure 6. Visual Representation of proposed model with existing techniques

5. Conclusion

The overarching goalmouth of this research is to expand the accuracy besides efficiency of medical diagnoses by creating a novel glaucoma classification model; this will help find this eye disease earlier and save time and money. Using the Vision Transformer architecture, this research introduced a new method for extracting features from medical images. In order to improve the representation's performance, the suggested EvoTAEA architecture included modules inspired by evolutionary algorithms. These modules included Multi-Scale Region Aggregation (MSRA), Modulated Deformable MSA (MD-MSA). The WGMOA algorithm enhances the classification accuracy by fine-tuning the parameters. A new method for glaucoma classification is presented in this work. It integrates

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

the best features of ResNet-50, DenseNet-201, besides Xception representations. Using a specialised concatenation layer and a final classification layer, the suggested approach takes advantage of the maps obtained from all of these models and merges them. By reducing the impact of the vanishing gradient issue, ResNet-50—popular for its remaining connections—makes feature extraction more successful. Improved representation learning is the result of DenseNet-201's ability to reuse features and gradient flow via dense connections. The depth-wise separable convolutions that define Xception allow it to strike a nice balance between computational performance and model complexity. Using measures including specificity, accuracy, and precision, the research compared two fundus imaging datasets, ORIGA and ACRIMA. With this in mind, a computational model that makes use of deep learning techniques can be a useful tool in the fight against glaucoma. In order to help ophthalmologists and other doctors perform precise glaucoma diagnoses, this methodology has developed a dependable system. The proposed method may have some limitations, such as the fact that the classification pipeline becomes more computationally difficult due to the integration of different models. This constraint is reasonable, nevertheless, when weighed against the achieved accuracy of glaucoma categorization.

References

[1] R. Fan et al., "Detecting glaucoma from fundus photographs using deep learning without convolutions: Transformer for improved generalization," *Ophthalmology Science*, vol. 3, no. 1, p. 100233, 2023.

[2] R. Hemelings et al., "A generalizable deep learning regression model for automated glaucoma screening from fundus images," *NPJ Digital Medicine*, vol. 6, no. 1, p. 112, 2023.

[3] A. Shoukat, S. Akbar, S. A. Hassan, S. Iqbal, A. Mehmood, and Q. M. Ilyas, "Automatic diagnosis of glaucoma from retinal images using deep learning approach," *Diagnostics*, vol. 13, no. 10, p. 1738, 2023.

[4] R. Thanki, "A deep neural network and machine learning approach for retinal fundus image classification," *Healthcare Analytics*, vol. 3, p. 100140, 2023.

[5] L. J. Coan et al., "Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review," *Survey of Ophthalmology*, vol. 68, no. 1, pp. 17–41, 2023.

[6] A. Elmoufidi, A. Skouta, S. Jai-Andaloussi, and O. Ouchetto, "CNN with multiple inputs for automatic glaucoma assessment using fundus images," *International Journal of Image and Graphics*, vol. 23, no. 1, p. 2350012, 2023.

[7] B. Gunapriya, T. Rajesh, A. Thirumalraj, and B. Manjunatha, "LW-CNN-based extraction with optimized encoder-decoder model for detection of diabetic retinopathy," *Frontier Scientific Publishing Pte. Ltd.*, vol. 1095, 2023.

[8] S. Santhosh and D. V. Babu, "Retinal glaucoma detection from digital fundus images using deep learning approach," in *Proc. 7th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Feb. 2023, pp. 68–72.

[9] V. Kurilová et al., "Detecting glaucoma from fundus images using ensemble learning," *Journal of Electrical Engineering*, vol. 74, no. 4, pp. 328–335, 2023.

[10] V. A. Devi, A. Thirumalraj, B. P. Kavin, and G. H. Seng, "Securing the predicted disease data using transfer learning in cloud-based healthcare 5.0," in *Intelligent Systems and Industrial Internet of Things for Sustainable Development*, Chapman and Hall/CRC, pp. 101–117.

[11] R. Shanthakumari et al., "Glaucoma detection using fundus images using deep learning," in *Proc. Int. Conf. Circuit Power Comput. Technol. (ICCPCT)*, Aug. 2023, pp. 1887–1894.

[12] D. Chen et al., "Applications of artificial intelligence and deep learning in glaucoma," *Asia-Pacific Journal of Ophthalmology*, vol. 12, no. 1, pp. 80–93, 2023.

[13] R. K. Patel and M. Kashyap, "Automated screening of glaucoma stages from retinal fundus images using BPS and LBP based GLCM features," *International Journal of Imaging Systems and Technology*, vol. 33, no. 1, pp. 246–261, 2023.

[14] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated tool support for glaucoma identification with explainability using fundus images," *IEEE Access*, 2024.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025

[15] D. Das, D. R. Nayak, S. V. Bhandary, and U. R. Acharya, "CDAM-Net: Channel shuffle dual attention based multi-scale CNN for efficient glaucoma detection using fundus images," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108454, 2024.

[16] G. S. Nugraha, A. Juliansyah, and M. Tajuddin, "Glaucoma detection based on texture feature of neuro retinal rim area in retinal fundus image," *International Journal of Health and Information System*, vol. 1, no. 3, pp. 117–127, 2024.

[17] M. Govindan, V. K. Dhakshnamurthy, K. Sreerangan, M. D. Nagarajan, and S. K. Rajamanickam, "A framework for early detection of glaucoma in retinal fundus images using deep learning," *Engineering Proceedings*, vol. 62, no. 1, p. 3, 2024.

[18] L. K. Singh, M. Khanna, S. Thawkar, and R. Singh, "A novel hybridized feature selection strategy for the effective prediction of glaucoma in retinal fundus images," *Multimedia Tools and Applications*, vol. 83, no. 15, pp. 46087–46159, 2024.

[19] X. R. Gao, F. Wu, P. T. Yuhas, R. K. Rasel, and M. Chiariglione, "Automated vertical cup-to-disc ratio determination from fundus images for glaucoma detection," *Scientific Reports*, vol. 14, no. 1, p. 4494, 2024.

[20] K. Gunasekaran et al., "An efficient cardiovascular disease prediction using multi-scale weighted feature fusion-based convolutional neural network with residual gated recurrent unit," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 27, no. 9, pp. 1181–1205, 2024.

[21] A. Diaz-Pinto et al., "CNNs for automatic glaucoma assessment using fundus images: An extensive validation," *Biomed. Eng. OnLine*, vol. 18, no. 1, pp. 1–19, Dec. 2019.

[22] J. Sivaswamy et al., "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 53–56.

[23] W. Wang et al., "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 568–578.

[24] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," *arXiv preprint arXiv: 2010.11929*, 2020.

[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 21–26 July 2017, pp. 1251–1258.

DOI: https://doi.org/10.54216/FPA.190212
Received: December 12, 2024 Revised: February 02, 2025 Accepted: March 01, 2025