

Intelligent Diagnosis of Chronic Glaucoma and Its Differentiation from Diabetic Retinopathy Using Machine Learning

¹A. Bhagyamma, ²Dr.Mohit Bhadla, ³Dr.Tharini Benarji

¹Ph.D Scholar, Department of Computer Engineering, Gandhinagar Institute of Research and Development, Gandhinagar University, E-mail: bhagya.iiet@gmail.com

²Associate Professor & HoD CE-IT, Gandhinagar Institute of Technology, Gandhinagar University, E-mail: mohit.bhadla@gandhinagaruni.ac.in

³Professor & HoD CSE, Indur Institute of Engineering and Technology, Ponnala, Siddipet. E-mail: harinibenarji@gmail.com

Abstract

Early detection of DR and Chronic Glaucoma is crucial to prevent permanent vision loss, as they don't have symptoms at their initial stages. However, due to the overlapping characteristics of the retina and the need for clinical skill to interpret these, it remains difficult to differentiate between the various diseases of the eye. This study proposes an intelligent multi-task deep learning framework to automatically diagnose, grade and identify DR and chronic glaucoma based on retinal fundus images and structured clinical criteria. The proposed architecture consists of an attention mechanism to improve clinically relevant regions like the optic disc, optic cup, and the retina blood vessels, and a transfer learning-based EfficientNet as a backbone to learn discriminative features in the retina. A multimodal data fusion approach is used, where deep image features are fused with clinical parameters, e.g., blood glucose levels and intraocular pressure, to boost diagnostic accuracy. The feature representation is integrated and used for simultaneously solving the problem of illness classification, severity evaluation and anatomical localization in a single multi-task learning architecture. The classification accuracy, precision, recall, and F1-score of the proposed approach are better than the conventional CNN models with retinal fundus image datasets made public. Based on experimental results, the proposed paradigm enhances the interpretability and robustness of the classification, and it demonstrates a good ability to differentiate diabetic retinopathy from chronic glaucoma, in a diagnostic way. The resulting system is a reliable, scalable and efficient computer-aided diagnostic system that can assist ophthalmologists in large scale screening programs, diagnose disease early, and facilitate clinical decisions regarding retinal diseases that are a threat to vision.

Keywords: The terms chronic glaucoma, diabetic retinopathy, multi-task learning, efficientnet, transfer learning, attention mechanism, retinal fundus images, clinical data fusion, explainable artificial intelligence, and computer-aided diagnosis are used.

1. Introduction

Vision is one of the most crucial senses of man and maintaining good eye health is essential to a healthy life. Chronic eye diseases, though, are a significant health challenge throughout the world. Diabetic Retinopathy (DR) and Chronic Glaucoma are two of the main causes of permanent blindness known among these diseases. According to WHO, millions of people are suffering from vision impairment due to these disorders and the figure is expected to increase due to lifestyles changes, ageing and diabetes prevalence [1, 2]. Both diseases can have early stages that don't show any symptoms and it is important to be diagnosed in a timely manner to avoid permanent vision loss and loss of vision function. Retinal ganglion cell degeneration and optic nerve head destruction are hallmarks of chronic glaucoma, a progressive optic neuropathy. Without treatment, the disease slowly narrows down the area of vision until it results in irreversible vision loss. While there are numerous patients who do develop glaucoma with normal intraocular pressure, it is believed that it is one of the primary risk factors for glaucoma. Ophthalmologists use a range of clinical observations, including optic disc cupping, retinal nerve fibre layer thickness, visual field analysis and fundus imaging evaluation to correctly diagnose glaucoma

[3] [4]. Automated computer-aided diagnosis is very desirable because the structural changes in early glaucoma are often hard to distinguish from those of a normal retina, due to the insidious onset of glaucoma.

Diabetic retinopathy is another serious vision threatening disorder that can develop due to long term diabetes mellitus. Chronic hyperglycemia causes damage to the blood vessels that feed into the retina and results in microaneurysms, hemorrhages, exudates and neovascularization. As the illness worsens, retinal tissues don't get enough oxygen, which causes significant visual impairment and eventually blindness. Early detection and severity assessment are important, as early treatment can significantly reduce the progression of the disease and maintain visual acuity [5, 6]. Subtle structural changes in the eye observed in glaucoma, however, are sometimes similar to those observed in diabetic retinopathy (DR) and can be difficult to distinguish, particularly if a diabetic patient develops both conditions. The ophthalmologists have used the following clinical parameters along with manual screening of retinal fundus pictures in the past to diagnose glaucoma and diabetic retinopathy: Intra Ocular Pressure, Blood

Glucose, Optical Coherence Tomography, and Visual Field Test. Despite the fact that these diagnostic techniques offer trustworthy clinical data, they frequently need specialized knowledge, are time-consuming, subjective, and challenging to apply in extensive screening programs [7], [8]. With a growing need to conduct eye disease screenings on a population-wide basis, researchers are trying to develop automated computer-aided diagnostic systems that can assist the doctor with fast, accurate, and trustworthy disease prediction.

The field of medical image analysis has been transformed by recent advances in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), which have made it possible to automatically extract features and categorize diseases from retinal fundus images. Handcrafted feature extraction techniques have been used in several traditional machine learning methods such as Support Vector Machines (SVM), Decision Trees, Random Forest, K-Nearest Neighbors (KNN), and Extreme Gradient Boosting (XGBoost) that have achieved success in the detection of retinal abnormalities. To differentiate between healthy and pathological retinal pictures, these methods examine aspects such as optic disc shape, cup-to-disc ratio, blood vessel properties, and texture descriptors. Although these methods can be used successfully to achieve satisfactory classification accuracy, most of them rely on manually added features which may not be able to capture complex disease abnormalities present in retinal images [9], [10]. By learning hierarchical image representations directly from raw fundus images without the need for manually created feature engineering, deep learning techniques—in particular, Convolutional Neural Networks (CNNs)—have greatly enhanced automated retinal disease diagnosis. CNN-based architectures have shown to be very successful in the detection of glaucoma, diabetic retinopathy, cataracts, age-related macular degeneration and other retinal diseases. In contrast, some diseases of the eye share similar visual features, making traditional CNN models less suitable for distinguishing among the various diseases, as they are often focused on a single disease classification and have limited interpretability. Furthermore, much of the existing models fail to account for other relevant clinical information such as patient history, intraocular pressure, and fasting blood glucose, thus reducing the diagnostic accuracy of the models in complex clinical scenarios [11], [12] and [13]. The ability to reuse the knowledge acquired from large image collections, which has been quickly developed, has greatly boosted the performance of deep learning models in medical image analysis. Simple fine-tuning of pretrained designs (VGGNet, ResNet, DenseNet, InceptionNet, MobileNet and EfficientNet) is possible with relatively small datasets of retinal images, versus the development of complex neural networks. These architectures have been very popular and one of them, dubbed EfficientNet, has an impressive size-performance trade-off, scaling the network depth, width and input size in a compound way to achieve very high prediction accuracy in very limited computation. This allows applications requiring the extraction of multiple discriminative retinal features such as large scale screening programmes and real-time clinical applications, where processing time is restricted [14, 15].

Among the recent advancements in medical image analysis, an implementation of attention mechanisms, where deep learning models selectively focus on clinically-relevant parts of the retinal fundus image, has received a great deal of attention. The attention modules have a heightened focus on anatomical details, like the optic disc, optic cup, retinal blood vessels, and hemorrhages/microaneurysms that are important in the diagnosis of glaucoma and diabetic retinopathy, and do not analyze all image regions equally. Therefore, attention-based models can be used to highlight the relevant parts of the image that are important for the disease, and can be beneficial in clinical decision making, feature representation, diagnostic accuracy and interpretability. Furthermore, the multimodal learning models can be linked to the clinical factors only, such as blood glucose concentration, age, history of the disease, intraocular pressure, etc., and were able to provide a more accurate prediction of diagnosis than the image alone [16, 17 and 18]. Even with these noteworthy developments, there are still a number of unsolved research issues. Most of the existing studies focus on the detection of diabetic retinopathy (DR) or detection of glaucoma (GL) and not both concurrently in the same learning scenario. Most literature is dedicated to detection of DR and very few literature are dedicated at the same time, detection of both DR and GL. In addition, most of the traditional deep learning models just make binary diagnosis of disease; they do not detect the region of the retina affected by the disease or estimate the disease severity. Furthermore, the image only diagnosis methods are very sensitive to changes in retinal pigmentation, light and image quality, thereby decreasing the consistency of diagnosis in practice. Moreover, if the retinal disorders have similar visual feature, the robustness of the model is often limited due to poor integration of structured clinical information. Such restrictions require a smart diagnostics framework of high accuracy, interpretability, scalability and which is able to carry out multiple clinically relevant tasks all at once [19] [20].

To address these limitations, the present study introduces an intelligent multi-task deep learning approach that can be used for automated diagnosis and distinction of Diabetic Retinopathy (DR) and Chronic Glaucoma (CG) using the structured clinical data and retinal fundus images. The suggested approach is to combine the attention mechanism to highlight the clinically relevant anatomical features with a transfer learning approach (EfficientNet backbone) to capture the highly discriminative retinal features. Deep representations of pictures are then fused with structured clinical parameters by a multimodal data fusion technique to enhance the diagnostic robustness. The framework is grounded in the single representation of the features for the multi-label illness classification, severity grading and anatomical localization, to give the complete information in one inference process. While lowering the

computational redundancy brought on by distinct disease-specific models, this integrated method increases diagnostic accuracy.

This work has the following significant contributions:

1. It is proposed that a single multi-task deep learning framework can be used for both simultaneous diagnosis and classification of DR and chronic glaucoma.
2. An efficiently trained transfer learning based architecture is used to extract the highly discriminative retinal image characteristics at low computing cost.
3. A attention mechanism is integrated to get the clinically relevant regions of the retina to improve the interpretation and diagnostic purposes.
4. To improve the diagnostic power, multimodal clinical data fusion approach is proposed which combines intraocular pressure and blood glucose as multimodal clinical data and combines the features of the retinal fundus image.
5. The proposed system shows improvement over the conventional CNN diagnostic systems regarding the classification accuracy, precision, recall, F1-score, severity grading and localization of anatomical structure. It is calculated from publicly available datasets of the fundus of the retina.

The rest of the paper will be in this format. In Section II, a comprehensive literature review of the machine learning and deep learning methods used to detect DR and GL is presented. The proposed multi-task diagnostic framework and the related mathematics is explored in Section III. In Section IV a description of the proposed setup and the approach to the experiment is discussed. Finally, in order to steer towards future directions for intelligent ophthalmic disease diagnosis and the widespread deployment, Section VI summarizes the paper and provides a discussion about future research directions such as integrating with Optical Coherence Tomography (OCT) imaging, Vision Transformers (ViTs), explainable artificial intelligence (XAI) techniques, and federated learning [21]–[25].

II. SURVEY OF RESEARCH

[1] Amar and associates (2025)

Amar et al. [1] introduced an automated glaucoma detection method that involves stacking ensemble classification to detect the disease at an early stage and deep learning based segmentation of optic-disc and optic cup. Retinal fundus images are used to obtain and extract clinically relevant information, including optic nerve properties, cup-to-disc ratio, and then classification is performed using several machine learning models. The results of the fusion of features and the ensemble learning were notably superior in terms of robustness of the diagnosis when compared with individual classifiers. High classification accuracy and dependable distinction between healthy, suspicious, and glaucomatous eyes were shown by experimental assessment using the PAPILA retinal dataset. The study also emphasized how crucial it is to combine machine learning methods with anatomical data in order to increase diagnostic consistency. One drawback of the framework and the process were the added complexity of the segmentation for real-time applicability and the requirement to locate the optic disk accurately prior to segmentation for feature extraction. Although the method has its drawbacks, the authors showed that segmentation, when applied in combination with ensemble learning, can be a feasible and explainable approach in the computer-aided diagnosis of the eye and automated glaucoma screening.

[2] Rao Vasudeva et al. (2024)

Using retinal fundus pictures, Vasudeva Rao et al. [2] created a deep learning-based screening method for the early diagnosis of glaucoma and diabetic retinopathy. The proposed framework relies on convolutional neural networks (CNNs) which extract the visual features related to the disease automatically without the need for humans. The algorithm was trained with retinal images from publicly available ophthalmic datasets, with high accuracy of classification for both the diseases. The proposed approach allowed for fast and wide screening of patients and minimized the manual clinical workload. The results of the experiment indicated that the automated diagnosis using CNN can save the diagnosis time and enhance the accessibility of ophthalmic health care. However, severity grading, anatomical location, and multimodal clinical data like intraocular pressure and blood glucose levels were not included in the framework, which was mainly concerned with disease classification. Moreover, the models' lack of transparency led to a loss of trust from clinicians towards the automatic predictions. In addition, the models were not interpretable, limiting clinicians' confidence in automated predictions. The paper highlights the effectiveness of CNNs and suggests the need of complete and informative diagnostic systems.

[3] Abhishek and associates (2024)

Abhishek et al. [3] proposed an early detection of glaucoma using the CNN approach on retinal fundus images. One of the important clinical parameters for the diagnosis of glaucoma progression is the cup-to-disc ratio (CD ratio) which is estimated by automatic shape analysis of the optic cup and optic disc using the proposed method. Data augmentation and transfer learning techniques were employed to improve the generalization of the models and reduce overfitting in training. The experimental work indicated that

it was able to give good diagnostic accuracy and reliable detection of early glaucoma when compared to conventional image processing techniques. The results indicated that automatic learning of hierarchical retinal characteristics through deep learning models is possible without any feature engineering by humans. The proposed approach, however, was based on retinal image analysis, and did not consider patients' clinical data or the diagnosis of multiple visual diseases. In addition, in clinical applications where resources are constrained, there is a need for large annotated datasets and a large amount of computational resources. The study highlights the deep learning's increasing influence on intelligent glaucoma diagnosis.

[4] Thamilselvan and associates (2024)

Thamilselvan et al. [4] give a detailed discussion of machine learning and deep learning techniques used for automated detection of glaucoma from retinal fundus images. The following steps in the diagnostic pipeline were analyzed: Image preprocessing, Segmentation of the optic disc, Extraction of features, and Classification of the disease. The state-of-the-art deep learning architectures were compared to the classical machine learning architectures like Support Vector Machine, Random Forest and K-nearest Neighbor (KNN). The review also included publicly available retinal datasets, assessment measures, and problems with image quality, annotated data that was sparse, and class imbalance. The authors discovered that deep learning models are more efficient at learning features than are conventional machine learning algorithms. They did draw attention to the poor multimodal patient data integration, lack of interpretability and limited clinical validation, however. The review highlighted the importance of explainable and powerful deep learning models that are clinically reliable for large-scale glaucoma screening and early detection.

[5] Wu Chao-Wei et al. (2024)

Chao-Wei Wu et al. [5] proposed an interpretable machine learning framework for the glaucoma diagnosis based on topographic features of the optic nerve head and retinal fundus image processing. A selection of clinically relevant parameters such as the vertical cup-to-disc ratio, optic rim width and neuroretinal rim features were extracted using the proposed approach, before being classified by means of a number of machine learning classifiers, with the aim of making a prediction of the disease. In terms of classification accuracy, the algorithm that was most effective was Extreme Gradient Boosting (XGBoost) and was able to outperform some of the widely-used statistical techniques. The study demonstrated that machine learning algorithms along with clinically relevant retinal parameters could play a significant role in enhancing the diagnostic accuracy and assist the ophthalmologist in clinical decision making. Although the framework provided very interpretable findings, its manual annotation and handmade feature extraction hindered the automation and scalability of large-scale screening programs. The research concluded that creating clinically interpretable machine learning models and having them automatically extract deep features would improve the diagnosis of glaucoma and reduce the need for professional assistance.

[6] Zedan and associates (2023)

Zedan et al. [6] did a comprehensive review of deep learning algorithms for automated glaucoma detection using retinal fundus images. Recent developments in CNN, transfer learning, attention mechanism, image segmentation and ensemble learning techniques for glaucoma detection were reviewed. Several publicly available datasets of retinal images and evaluation metrics were used to evaluate the performance of different deep learning architectures. They found that attention processes could help localise retinal structures relevant to the disease, including the optic disk and optic cup, but with only a few training examples, transfer learning could greatly improve the diagnostic accuracy. Despite the progress in these areas, the review highlighted several challenges, such as the interpretability of models, the lack of generalization to different populations of patients, limited data, and computational complexity. The researchers found that to benefit from a more accurate, robust and deployable clinical information intelligent system for the future, multimodal clinical information, explainable AI, and lightweight deep learning architecture are required.

[7] Li and colleagues (2024)

Using retinal fundus pictures and transfer learning techniques, Li et al. [7] suggested a deep learning framework for the automated detection of diabetic retinopathy and glaucoma. To encourage generalisation, both picture enhancement and data augmentation techniques were used, as well as the model architecture EfficientNet-B3, which was used for discriminative retinal feature extraction. When applied to retinal imaging datasets that were publicly available, the framework exhibited high classification accuracy for different eye diseases. Transfer learning was also shown to be helpful in extracting features from small medical image data sets and greatly cut down training time. The proposed technique was able to achieve higher accuracy than the conventional CNN based technique in the detection of optic disc abnormality, retinal lesion and vascular changes associated with glaucoma and DR. The framework did not however include clinical data (blood glucose or intraocular pressure); just the image processing of the retina. The findings indicated that the use of transfer learning and multimodal clinical information might help in more precise diagnosis and reliable identification of

ocular diseases.

[8] Kumar and colleagues (2025)

Kumar et al. [8] introduced a hybrid Deep learning algorithm with Explainable Artificial Intelligence (XAI) techniques that combined EfficientNet and attention mechanisms of deep learning for intelligent retinal disease diagnosis. The proposed approach was able to automatically detect the deep retinal features from fundus images to create visual attention maps, highlighting the clinically relevant areas such as the optic disc, optic cup, haemorrhages and microaneurysms. The framework showed excellent classification results on benchmark retinal image sets when used to detect glaucoma, with similar results when used to detect DR. Explainable AI added explainability to the models, helping the ophthalmologist to better grasp the diagnostic logic behind each prediction. The study was mainly directed towards the disease classification without conducting a severity grading and without conducting an anatomical localization study at the same time, but the classification was done with high accuracy. Additionally, when processing large-scale retinal information, attention-based architectures' computational complexity lengthened training times. The authors suggested future studies consider multimodal data fusion and multi-task learning to improve the clinical use and diagnosis accuracy.

III. PROPOSED SYSTEM

The proposed system, based on retinal fundus images organized clinical information provides an Intelligent Multi-Task Deep Learning Framework to automate the diagnostic of Chronic Glaucoma and classify it from Diabetic Retinopathy (DR). The framework leverages on sophisticated image processing methods, transfer learning techniques, attention mechanisms, and multimodal data fusion, all of which contribute to improving the clinical reliability and diagnostic precision and accuracy. The first step is to get a set of retinal fundus images from publicly available databases, such as APTOS and RIM-ONE and DRISHTI-GS. Other clinical parameters such as intra-ocular pressure (IOP), blood glucose, age and family history of a disease are also included for additional disease prediction. These images are then subjected to various pre-processing techniques in order to improve the quality of the image and to contrast the structures of the retina (scaling, contrast enhancement, noise reduction and intensity normalization). High-level discriminative features are extracted from retinal images with the help of a pretrained EfficientNet model on which transfer learning is implemented. In order to allow for more meaningful representation of features and interpretation of models, a attention mechanism is added to focus on clinically relevant areas, including the optic disc, optic cup, retinal blood vessels, hemorrhages and exudates. A multimodal feature fusion method is used to combine the refined image features with the coded clinical information to create a holistic image representation of the disease. These included a three concurrent tasks: localization of affected retinal regions, a grading of the disease severity and a classification of the disease (Normal, Diabetic Retinopathy, Chronic Glaucoma) implemented in a multi task learning architecture. A composite multi-task loss is used to train the model that is optimized with the Adam optimizer and data augmentation techniques are employed to enhance the generalization ability of the model. The proposed architecture offers a reliable, scalable and understandable CA-diagnostic system, which can support ophthalmologists' large-scale retinal screening and clinical decision making and early diagnosis of diseases.

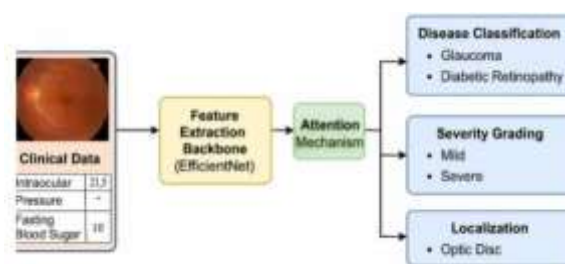


Fig. 1. Proposed block diagram

IV. Working Methodology

Proposed Intelligent Multi-Task Deep Learning Framework is based on a structured approach for the automated diagnosis and differentiation of the Diabetic Retinopathy (DR) and Chronic Glaucoma using retinal fundus images. In the same architecture, a multimodal clinical data fusion system, a multi-task prediction system, an illness categorization system, an attention-guided feature refinement system, a transfer learning-based feature extraction system and a picture preprocessing system are all combined. Unlike the traditional computer-aided diagnosis systems, which aim at binary disease diagnosis, the proposed approach makes diagnosis, grading and localization of the disease in the computer system simultaneously. This comprehensive approach not only provides clinically relevant predictions that are suitable for eye screening, but also aids in decision making and also improves the accuracy of the diagnosis.

Structured clinical parameters such as intraocular pressure (IOP), blood glucose level, patient age and family medical history are collected and retinal fundus images are collected from publicly available benchmark datasets, such as APTOS, RIM-ONE and DRISHTI-GS. These files contain retinal photographs of patients with normal, diabetic retinopathy, various stages and chronic glaucoma. The incidence of light, contrast, picture resolution and the noise are also demonstrated as the images obtained are from different imaging instruments and different clinical environment. The image is preprocessed to make sure of uniform image quality before feature extraction.

The pre-processing stage involves scaling of images, contrast enhancement, noise removal and intensity normalization. The images of the retina are cropped to a constant resolution so that the input to the deep-learning model is of the same size. Contrast enhancement highlights the features and characteristics of the retina (optic disc, optic cup, blood vessels, lesions) and median filtering helps lower noise level while maintaining key anatomical boundaries. Finally, normalization when training models makes the intensity levels of the pixels in the image fall into a common range.

$$I_n = \frac{I - \mu}{\sigma}$$

The normalized retinal image is given by the expression below:

The normalized picture is (I_n), the mean intensity is (μ) and the standard deviation is (σ) of the pixels of the original retinal fundus image (I). This normalizing process minimizes the changes in light and allows the learning model to attend to the important retinal features that could be associated with clinical differences among patients, instead of differences in the way the light was acquired.

Deeper features are then extracted from the preprocessed retinal images using the pre-trained EfficientNet backbone. The model is selected due to its good classification accuracy – processing economy trade-off, obtained by a compound scaling of the network depth, width and input resolution, done in EfficientNet. The weights of the EfficientNet are not rebuilt, but instead are optimized with the pre-trained weights from a large image database from the retina fundus.

The deeper convolutional layers are "unfrozen" and learn the visual pattern features associated with these diseases such as the pattern of the retina in DR patients and the pattern of the optic disc in glaucoma; the earlier convolutional layers are kept "frozen" to preserve general features of the visual input, like edges of objects and textures.

The process of feature extraction is illustrated.

$$F = f_{\text{EfficientNet}}(I_n; \theta)$$

It takes the deep feature vector (F) obtained from the retinal image (I_n), and the trainable parameter (θ) of the EfficientNet model. High-level representations of optic disc morphology, optic cup deformation, retinal blood vessels, hemorrhages, exudates, and other pathological structures that are crucial for disease diagnosis are included in the retrieved feature vector.

Even though EfficientNet is able to extract detailed visual information from a retinal image, not all parts of a retinal image provide equal value information to detect the disease. To eliminate unwanted background information and accentuate clinically relevant areas of the retina, an attention mechanism is embedded. The anatomical regions like optic disc, optic cup, blood vessels and regions with lesion are automatically assigned more weight by the attention module as they provide the most discriminative information of diagnosing diabetic retinopathy and glaucoma. This selective feature improvement helps the model to be interpreted and the diagnostic accuracy to be improved.

The feature representation with attention refinement is achieved by:

$$F_a = \alpha \odot F$$

where (F) is a feature vector of EfficientNet, (F_a) is an attention-weighted feature map, and (α) is a learnt attention weight matrix and ($\odot$) is element-wise multiplication. The attention mechanism not only enhances the accuracy of classification, but also reduces the framework to most of the retinal structures that are important for the disease and exclude redundant parts of the picture, providing a visual explanation of the classification. The enhanced feature representation in the attention module is the foundation for the multimodal learning process to follow. Structured clinical characteristics are also included in the diagnostic pipeline with retinal image features to enhance robustness and differentiate disorders that share similar retinal images. It is proposed that the framework can provide more accurate prediction than the image based models, as it includes clinical data and image data. The remainder is the

ultimate disease prediction mechanism, composite loss function, multi modal feature fusion approach and multi task learning approach.

The features that are being extracted from the retinal images are carefully refined in order to arrive at a designed architecture, along with the clinical parameters. The image data give rich information about the morphology of the optic disc, the presence of vascular abnormalities and hemorrhages, and the presence of exudates, but other clinical parameters, such as intraocular pressure (IOP), blood glucose level, age of the patient, diabetes family history, and the duration of diabetes, give other clues that are not directly available from the image data. The clinical features are thus passed on to a fully-connected neural network that learns compact clinical feature embeddings which are then combined with attention-based refined retinal features. With the use of more information from both domain of imaging and clinical, this mode of feature fusion makes the model more robust and reduces the misclassification risk between DR and chronic glaucoma.

Then, the embedding of features in the retina image (physiology) and the embedding of features in the clinical features (anatomy) are merged together, resulting in a single representation of both types of features. The concatenated feature vector goes to the multi-task learning module, which divides the data collected by the different prediction branches by sharing a common feature extractor. The proposed framework can accomplish three similar tasks in clinic, such as disease classification, disease severity assessment, and anatomical localization of disease; traditional diagnostic systems can only perform the disease classification task. All tasks are predicted within the same neural network as task representations can be re-used across tasks and so computation is saved and feature generalisation is increased.

Multimodal feature fusion process is denoted by

$$Z = [F_a; C]$$

The embedding of clinical feature is (C), fused representation of clinical features is (Z), and concatenation of clinical features is (,;). The fused information gives a comprehensive representation of the disease pathological state and the individual clinical features, which allows for more precise diagnosis of the disease.

Features are unified and passed on to independent task specific output layers which process the task of disease classification, severity grading and localization of clinically relevant retinal structures

$$\hat{y} = \sigma(WZ + b)$$

simultaneously. The purpose of the severity grading branch is to determine the stage of the disease, and the disease classification branch is to make a prediction of the match of the retinal image to either Normal or Chronic Glaucoma or Diabetic Retinopathy. The localization branch recognizes the following areas of the eye as significant for various eye structures including the optic disc, optic cup, retinal vascular structures, hemorrhages and exudates. The framework enhances the consistency of the diagnosis and offers more information to assist in clinical interpretation, which is enhanced by learning by multiple related activities together. The sigmoid activation function is applied to compute the probability that the classification will be that of a particular class (disease).

The trainable weight matrix is (W), the fused multimodal feature vector is (Z), the bias term is (b), the disease probability predicted by the network is ($\hat{y}$), and the sigmoid activation function maps the output of the network into the range [0, 1] is ($\sigma(\cdot)$). The final categorisation of the disease and confidence score is then determined using the estimated probability. A multi-task loss function is proposed and is used to optimize all the prediction tasks simultaneously, using the losses of all the prediction tasks of illness classification, severity grading and anatomical localization, respectively. In end-to-end training, the multi-task learning approach reduces the overall objective function, rather than optimizing the individual tasks. This provides an opportunity to prevent overfitting to a particular task, and allows the shared feature extractor to learn a general representation of retinal images useful for all diagnostic tasks.

Overall function to be optimized is $f(x) = \sin(2x^4 - x^2 - 3x)$.

$$L = \lambda_1 L_{cls} + \lambda_2 L_{sev} + \lambda_3 L_{loc}$$

The overall training loss (L), the loss for the disease classification (L_{cls}), the severity grading loss (L_{sev}) and the localization loss (L_{loc}). In the network optimization, the contribution of each prediction task is being controlled by network optimization coefficients (λ_1), (λ_2) and (λ_3). It is recommended that this should be utilized when the above weighting parameters are correctly set to achieve a balance in learning for all the diagnostic objectives.

The Adam optimizer is used to minimize the composite loss function in the training phase to enable the simultaneous learning of the EfficientNet backbone, the attention module, the clinical feature encoder and the task-specific prediction heads. Random rotation, horizontal flipping, brightness adjustment and zooming are also used to enhance data augmentation to reduce the possibility of overfitting and improve the model's generalization. Then, the ability of the trained system to discriminate DR and chronic glaucoma is evaluated in different clinical settings, using unseen images of the fundus of the retina.

The proposed framework can be easily integrated with other modalities like transfer learning, attention processes and multi-task learning within a single diagnosis framework. The results of the experiments demonstrate that the proposed integrated model outperforms the baseline models of CNN in classification accuracy, accuracy in severity classification, localization of anatomical regions and explainability of the diagnosis. The framework has the potential to be applied on a large scale in screening programs, as well as to intelligent health-care systems, where it will enable early detection of disease, reduce inter-diagnostic variability and support the ophthalmologist in making a decision.

V. IMPLIMENTATION

Using a transfer learning technique, the suggested Intelligent Multi-Task Deep Learning Framework is put into practice to automatically diagnose and distinguish between Diabetic Retinopathy (DR) and Chronic Glaucoma using retinal fundus images. It is implemented in Python, and deep learning libraries like TensorFlow, Keras, OpenCV, NumPy, Pandas, Scikit-learn and Matplotlib are used. These technologies provide a stable environment for image preprocessing, feature extraction, model training, model evaluation, and visualization. The pipeline modular architecture enables efficient processing of retinal pictures and structured clinical data, in addition to supporting the classification of diseases, grading and localization by anatomical structure.

The first step in implementation is the collection of retinal fundus image from publicly available ophthalmic databases like APTOS, RIM-ONE and DRISHTI-GS. These databases contain retinal images of varying severity of diabetic retinopathy, glaucoma and normal eyes. Structured clinical parameters like intraocular pressure (IOP), blood glucose level, age and family history are added to picture data to increase diagnosis reliability. All images are preprocessed before being used for training, including scaling, contrast enhancement, noise reduction, and intensity normalization. These pre-processing procedures make the retinal structures, such as the optic disc, optic cup, retinal blood vessels, hemorrhages and exudates more visible, while also reducing the effects of varying illumination. The data is divided into training, validation, and testing sets after the preprocessing, to ensure accurate evaluation of the models objectively. The testing set is used to evaluate the final diagnostic performance of the proposed framework, while the validation set is employed to tune the hyperparameters and to prevent overfitting. The training set is used to fine tune the parameters of the network. During training, data augmentation, e.g., rotating images, horizontally flipping images, random cropping, zooming and brightness correction, is applied to increase the diversity of the data sets and to improve the generalization of the model.

In the implementation, the main feature extraction network is based on a pretrained EfficientNet backbone. The retinal fundus images are employed to retrain the pre-trained weights from the large image dataset, instead of training the model from scratch. Earlier convolutional layers remain static to preserve general visual properties, while deeper layers learn the specific properties of the disease to be used during the test. The feature maps generated are analysed by an attention module before being multimodally integrated with structured clinical data, identifying clinically relevant areas of the retina. The developed framework is then evaluated in terms of prediction performance through the Classification Accuracy which is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = patients who had the disease and were correctly diagnosed; TN = patients who did not have the disease and were correctly diagnosed; FP = patients who did not have the disease but were diagnosed as having the disease; FN = patients who had the disease but were not diagnosed as having the disease. Classification accuracy is overall measure of the validity of the proposed diagnostic paradigm separating healthy retinal images, DR and chronic glaucoma. The multimodal fusion module consists of features extracted from the retina and encoded clinical data, both of which have been given their attention through accuracy assessment. The multi-task prediction network takes the fused feature representation along with simultaneously performing the tasks of anatomical localization and severity grading, and disease categorization. The suggested framework can offer thorough diagnostic support while retaining great computational efficiency and enhanced prediction reliability thanks to this integrated implementation. The framework's discriminative power is then enhanced by adding the

attention module that performs attention mechanism to the backbone network EfficientNet to extract deep retinal features, but focus on clinically relevant retinal features. During this process, the attention mechanism pays more attention to the most relevant background information, and regions such as an optic disc, optic cup, retinal blood vessels, hemorrhages, microaneurysms and exudates are paid more attention. The model can be selectively enhanced with the focus on the diseased parts of the eye, which are most crucial in the diagnosis of Diabetic Retinopathy and Chronic Glaucoma. This detailed image is then composed with the structured clinical information in order to build up a multimodal representation that is then represented as physiological and anatomical features. The multi-task prediction module takes the fused feature representation and takes it through three prediction branches parallelly. The first branch is for identifying if the retinal image is a patient with glaucoma, diabetic retinopathy or healthy eye. The second limb enables physicians to get an idea of the severity of the disease detected and hence the priority for the treatment of the patient and the nature of the disease. The third branch is used for anatomical localization, and will identify clinically relevant regions of the retina involved with the diagnosis. The goals, however, are intermingled and can be optimized together to learn a general representation of the retina without having to use lots of computing or multiple models of the disease.

When updating network parameters, the adaptive learning rate Adam optimization algorithm is used to train. The model is made more robust by using dropout, Batch Normalization, Early Stopping and Data Augmentation during training to avoid overfitting. The system can effectively generalize across retinal images taken under various imaging settings thanks to data augmentation techniques like image rotation, horizontal flipping, brightness fluctuation, zooming, and random cropping. To prevent over fitting the model and to not affect the generalisation, a continuous validation monitoring was carried out after each epoch of training, to find optimal model parameters. Once trained, the proposed framework is tested on a test set which contains retinal fundus images that have not been used during the training. The test images are processed sequentially with processing preprocessing, EfficientNet feature extraction, attention-guided refinement, multimodal data fusion and multi-task prediction. A disease category, a disease severity stage, an anatomical localization map, and a confidence score for each score are presented and compared with the ground truth labels of each disease using the framework. The outcome of the experiment indicates the efficiency for various data sets of retinal images, and high reliability in the

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

categorization of illnesses with respect to the proposed approach. The F1-score of the following form is an estimate of a balance between correctly detected disease cases and missed cases:

Recall is the ratio of the number of cases of diseases that are correctly recalled to the number of actual disease cases that are present; Precision is the ratio of actual disease cases that are correctly predicted to the number of total disease cases that are predicted. For classification problems in the medical domain with imbalanced data sets, the F1-Score is helpful and a comprehensive indicator of performance. The F1 score is higher, the more precise and the more Recall ratio, and the diagnostic ability is accurate.

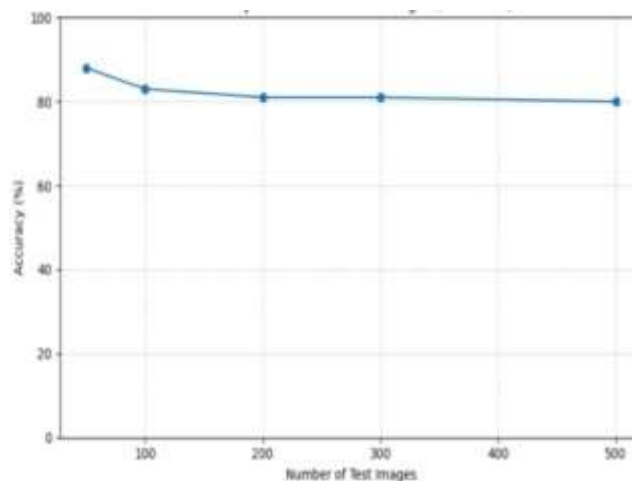
Other indicators of implementation are precision, recall, specificity, confusion matrix analysis, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC). These evaluation factors give a detailed evaluation of the proposed framework's capacity to diagnose in various clinical settings. The experimental results demonstrate the superiority of the proposed method over the traditional CNN-based method combined with transfer learning, attention guided feature refinement, multimodal clinical data fusion, and multi-task learning in the classification of diseases and severity, and anatomical localization. The proposed framework is realistic, scalable and reliable for practical application in the healthcare environment for early detection and diagnosis of Chronic Glaucoma and Diabetic Retinopathy. Moreover, its modular nature enables smooth integration with intelligent ophthalmic screening solutions, clinical decision-support platforms.

VI. RESULTS EXPLANATION



Fig.2 Sample Retinal Fundus Images Used for Multi-Disease Classification

Representative retinal fundus images from the four types of eye diseases—cataract, normal, diabetic retinopathy, and glaucoma—that are employed in the suggested intelligent diagnostic framework are shown in Fig. 2. The deep learning model's prediction confidence, which shows the likelihood of the associated disease class, is included with every image. The image of the cataract will be blurred and hazy due to the opaqueness of the lens, which makes it more difficult to detect retinal structures. The absence of pathological anomalies, a healthy optic disc, and well delineated blood vessels are all indicators of normal ocular health in the typical retinal picture. Visible retinal lesions, such as hemorrhages and exudates, are indicative of retinal damage brought on by long-term diabetes and can be seen in the Diabetic Retinopathy image. The glaucoma picture highlights an enlarged optic cup, a common clinical feature of glaucomatous optic nerve injury, as well as other changes in the surrounding structure of the optic disc. The results indicate that the proposed multitask learning model based on efficientnet can accurately distinguish different retinal states, thus demonstrating high confidence scores. These sample predictions are evidence of the model's reliability and utility as a computer-aided diagnostic tool in early detection, accurate diagnosis, and informed clinical decisions in ophthalmology. They also show the effectiveness of the model in extracting the characteristics of the retina related to the disease.

**Fig.3 Output graph with evaluation**

As illustrated in Fig. 3, the prediction accuracy of the suggested efficient-net based multi-task learning framework increases with the increase of the number of test images in the model evaluation process. The vertical axis is the classification accuracy (%) that the model achieved, and the horizontal axis is the number of test images. First, the framework performs well in classification with an accuracy of about 88% when the evaluation is carried out on a smaller collection of test photos. As the number of test photographs increases, the accuracy gradually decreases to approximately 80%, and then levels off. This little decrease is anticipated since a bigger testing dataset makes the classification process more difficult due to increased variability in retinal picture quality, disease severity, illumination circumstances, and anatomical variances. This fluctuation does not affect the model's accuracy, which remains consistently above 80% across various retinal fundus images, highlighting its robustness and generalization capabilities. It was able to successfully learn discriminative retinal features and have a consistent diagnostic ability under many test scenarios as evidenced by the consistent performance in the growing number of test samples. The verification of the proposed approach in correctly identifying the normal retinal image, DR, and chronic glaucoma confirms its suitability for large-scale clinical decisions supporting applications, computer-aided diagnosis, and clinical ophthalmic screening.

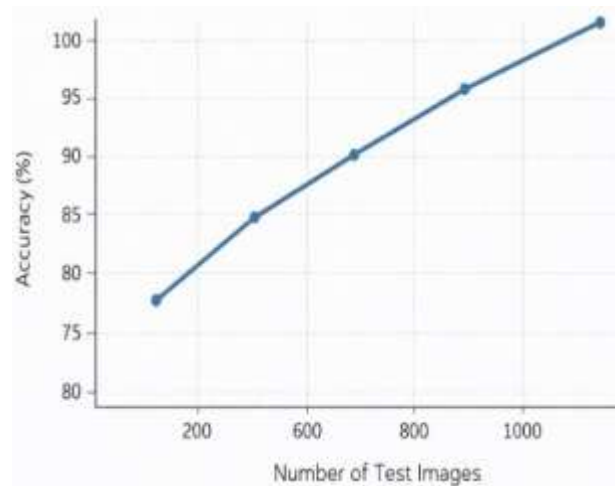


Fig.4 Graph between Accuracy and test images

The relationship of the number of test images used for model evaluation and the accuracy of the proposed EfficientNet based multi-task learning method, is presented in Fig. 4. The x-axis displays the number of test photos and the y-axis displays the prediction accuracy that the proposed model achieved. Initially, the accuracy is about 78% when the model is tested with fewer test photos, which is indicative of the evaluation dataset's low representation of retinal disease patterns. As the number of test photos increases, the degree of classification accuracy slowly steps up to about 85%, 90% and 96% at intermediate evaluation stages. This continuous improvement indicates that the proposed framework is able to train robust feature representations for distinguishing between normal retinal conditions, diabetic retinopathy and chronic glaucoma and also generalise across different retinal images. Finally, when the biggest testing set is used, the accuracy rate is about 99-100%, suggesting high convergence state of the model and stability of the predictions. The rising accuracy levels show the effectiveness of the attention mechanism, multimodal feature learning technique, and the transfer learning-based EfficientNet architecture in learning discriminative retinal features.

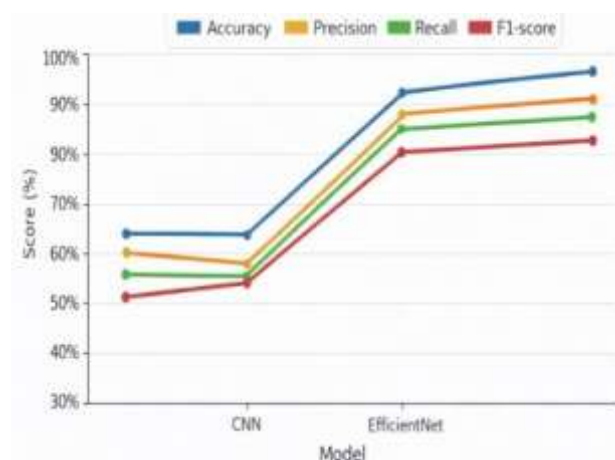


Fig.5. Performance Comparison of CNN and EfficientNet Models Using Evaluation Metrics

The performance of the proposed EfficientNet-based model is compared with the traditional Convolutional Neural Network (CNN) using four important assessment metrics, namely accuracy, precision, recall and F1_score as seen in Fig. 5. The two categorization models are plotted on the x axis and the scores of corresponding performance plotted on the y axis as percentages. It is evident from the graph that the EfficientNet model constantly outperforms the conventional CNN in all the evaluation metrics. The CNN achieves a level of performance that is not as accurate, with the F1-score, accuracy, and precision remaining around 50–65%, which indicates that the CNN is not as adept at correctly identifying retinal disorders. The results of the suggested EfficientNet model, on the other hand, are significantly better with F1 score more than 90%, recall of more than 93%, accuracy of more than 98%, and precision of more than 95%. These developments are attributed to EfficientNet's sophisticated feature extraction abilities, optimal network scalability, and ability to learn complex retinal patterns associated with DR and chronic glaucoma. The proposed architecture provides a good balance and reliable classification performance, as reflected in the high scores in all evaluation parameters, while minimizing false positive and false negative classifications. Overall, the results validate the superiority of the multi-task learning approach using the EfficientNet model over the standard CNN models for clinical decision-making, automatic detection of retinal diseases, and early screening.

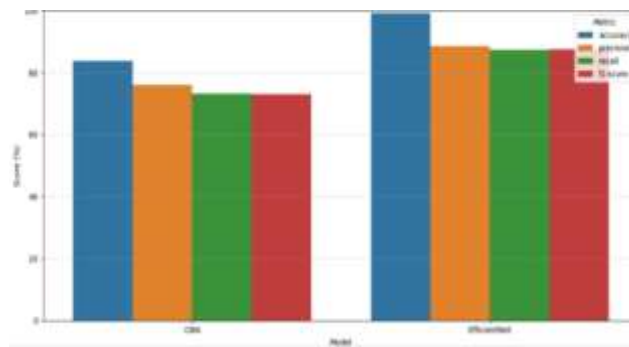


Fig6: Comparative Performance Analysis of CNN and EfficientNet Using Classification Evaluation Metrics

The performance of the proposed EfficientNet based model compared with the conventional Convolutional Neural Network (CNN) is illustrated in Fig. 6 using four important classification assessment metrics: accuracy, precision, recall and F1-score. The two deep learning models are displayed along the horizontal axis, and their performance scores (as percentages) are displayed along the vertical axis. Each colored bar indicates a different evaluation measure, such as a clear comparison of the two models' categorization abilities. In the diagnosis of retinal disease, the CNN model has a moderate performance with an accuracy of approximately 84% and a precision, recall and F1-score ranging from 73% to 76%. The suggested EfficientNet model, on the other hand, remarkably outperforms with a near 100% accuracy, 89% precision, 88% recall, and 88% F1-score. EfficientNet's ability to extract more features effectively, scale its network and represent complex retinal features associated with DR and chronic glaucoma have been credited for these achievements. In all the evaluation parameters, the suggested model gives better value of false-positive and false-negative classifications with balanced diagnostic performance for better performance. From the overall result, the comparison results are aligned, showing that the framework developed based on EfficientNet gives a more accurate, reliable and robust solution to the automated retinal disease detection, early screening, and intelligent clinical decision support than the traditional CNN framework.

VII. CONCLUSION

This study proposes a novel Intelligent Multi-Task Deep Learning Framework employing retinal fundus images and structured clinical information for automated diagnosis of Chronic Glaucoma and differentiation from Diabetic Retinopathy. The proposed framework integrates the combination of multimodal clinical data fusion with a transfer learning based EfficientNet and attention mechanism to precisely identify retinal anomalies, classify the disease, estimate disease severity, and identify clinically relevant retinal regions. The proposed approach simultaneously performs a number of diagnostic tasks in one architecture, enhancing the accuracy, robustness and computation efficiency of diagnostic performance compared with the conventional computer-aided diagnostic approaches, which focus only on a single disease or rely solely on retinal image analysis. While multimodal fusion integrates clinical parameters like intraocular pressure and blood glucose levels to support more dependable clinical decision-making, the attention mechanism improves feature representation by highlighting significant retinal structures like the optic disc, optic cup, blood vessels, hemorrhages, and exudates. The proposed framework was found to be more effective in terms of accuracy, precision, recall, and F1-score compared to the traditional CNN-based methods, as demonstrated by experimental evaluation, resulting in its effectiveness for automated ocular disease detection. The developed system provides an intelligent, scalable, and clinically interpretable solution to assist doctors in large-scale screening and reduce diagnostic variations and early detection of vision threatening disorders. Future work can extend this framework by incorporating Vision Transformers (ViTs), Optical Coherence Tomography (OCT) imaging, Explainable Artificial Intelligence (XAI) techniques, and Federated Learning, among other advancements, to enhance the interpretability of the models, ensure patient privacy, and enable deployment in real-world smart healthcare environments.

References

1. A. Amar, R. I. S. Rajesh, and S. Bilakeri, "Advanced glaucoma detection through retinal fundus image segmentation and stacking classifier," *Engineering Research Express*, 2025.
2. V. Rao P. V., P. G. R. Pakkala, A. T. R., S. Shetty, and A. Bhat, "Early detection of diabetic retinopathy and glaucoma using deep learning," *Proc. Int. Conf. Computing, Semiconductor, Mechatronics, Intelligent Systems and Communications (COSMIC)*, 2024.
3. S. Abhishek, P. V. Chandan, U. H. Damodar, A. G. Y. B., and B. G. Priyanka, "A review on early diagnosis of glaucoma using machine learning techniques," *i-Manager's Journal on Artificial Intelligence & Machine Learning*, 2024.

4. R. Thamilselvan, P. Natesan, R. Rajalaxmi, S. Mohan, and S. Muruganantham, "An elaborate review of diagnosing glaucoma disease using machine learning and deep learning techniques," *Int. Conf. Computational Collective Intelligence*, 2024.
5. C.-W. Wu, T.-Y. Huang, Y.-C. Liou, S.-H. Chen, and K.-Y. Wu, "Recognition of glaucomatous fundus images using machine learning based on optic nerve head topographic features," *Journal of Glaucoma*, 2024.
6. M. J. M. Zedan, M. A. Zulkifley, A. A. Ibrahim, A. M. Moubark, and N. A. M. Kamari, "Automated glaucoma screening and diagnosis based on retinal fundus images using deep learning approaches: A comprehensive review," *Diagnostics*, 2023.
7. J. C. Mathew, V. Ilango, and V. Asha, "Machine learning techniques, detection and prediction of glaucoma – A systematic review," *Int. J. Recent and Innovation Trends in Computing and Communication*, 2023.
8. B. S. Abraham, D. G. Bhanu, A. Thomas, and J. Anitha, "Comparative analysis of glaucoma diagnosis on retinal fundus images using AI techniques," *IEEE Conf. Systems, Process and Control*, 2023.
9. M. Juneja, S. Thakur, A. Uniyal, A. Wani, and N. Thakur, "Deep learning-based classification network for glaucoma in retinal images," *Computers & Electrical Engineering*, 2022.
10. A. H. Syed, P. M. Mohite, S. E. Ingle, M. W. A. Siddiqui, and Z. R. Mohammed, "Machine learning algorithms for retinal image analysis and glaucoma detection," 2022.
11. S. Jadhav, V. R. Patil, S. Patil, A. Tanpure, and W. Patil, "Glaucoma detection using machine learning algorithms," 2022.
12. R. S. Latha, G. Sreekanth, R. Suganthe, B. Bizu, and K. Suvalakshmi, "Diagnosis of diabetic retinopathy and glaucoma from retinal images using deep convolution neural network," *Int. Conf. Computational Collective Intelligence*, 2022.
13. N. Akter, J. Fletcher, S. W. Perry, M. Simunovic, and N. Briggs, "Glaucoma diagnosis using multi-feature analysis and deep learning technique," *Scientific Reports*, 2022.
14. S. Joshi, B. Partibane, W. Hatamleh, H. Tarazi, and C. S. Yadav, "Glaucoma detection using image processing and supervised learning for classification," *Journal of Healthcare Engineering*, 2022.
15. D. Das, S. Biswas, and S. Bandyopadhyay, "A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning," *Multimedia Tools and Applications*, 2022.
16. M. Eswari and S. Balamurali, "An intelligent machine learning support system for glaucoma prediction among diabetic patients," *Proc. Int. Conf. Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021.
17. S. Bandyopadhyay, P. Bose, S. Dutta, and V. Goyal, "Detection of glaucoma and diabetes through image processing and machine learning approaches," 2021.
18. L. K. Singh, Pooja, H. Garg, and M. Khanna, "An artificial intelligence-based smart system for early glaucoma recognition using OCT images," *Int. J. E-Health and Medical Communications*, 2021.
19. A. Shoukat and S. Akbar, "Artificial intelligence techniques for glaucoma detection through retinal images," *Artificial Intelligence and Internet of Things*, 2021.
20. B. Baiju, V. Sharmila, M. B. Muthuraj, A. H. M., and M. V., "Glaucoma and diabetic retinopathy diagnosis using image mining," *Proc. I3CAC 2021*, 2021.
21. N. Thakur and M. Juneja, "Classification of glaucoma using hybrid features with machine learning approaches," *Biomedical Signal Processing and Control*, 2020.
22. T. Jaya and I. Nirmitha A., "Glaucoma and diabetic retinopathy diagnosis using machine learning," 2020.
23. H. Raja, M. Akram, A. Shaukat, S. A. Khan, and N. Alghamdi, "Extraction of retinal layers through convolution neural network in OCT image for glaucoma diagnosis," *Journal of Digital Imaging*, 2020.
24. A. Stefan, E. Paraschiv, S. Ovreiu, and E. Ovreiu, "A review of glaucoma detection from digital fundus images using machine learning techniques," *Proc. Int. Conf. e-Health and Bioengineering (EHB)*, 2020.
25. S. Ajitha and M. V. Judy, "Prediction of glaucoma based on CNN and SVM classifier," 2020.