GMR
The Original

REFINING OPHTHALMIC DISEASE CLASSIFICATION USING DEEP LEARNING TECHNIQUES

Soumya Das[1], Ashok Kumar Bhoi[2], Basanta Kumar Swain [3], Dillip Ranjan Nayak[*4]

[1]Assistant Professor in the Department of Computer Science and Engineering, Government College of Engineering, Kalahandi, Bhawanipatna, Odisha 766003, Email ID: s.das@gcekbpatna.ac.in
[2]Assistant Professor, Department of Computer Science and Engineering, Government College of Engineering, Kalahandi, Bhawanipatna, Odisha, 766003, Email ID: akbhoi@gcekbpatna.ac.in
[3]Associate Professor Department of Computer Science and Engineering, Government College of Engineering, Kalahandi, Bhawanipatna, Odisha, 766003, Email ID: bkswain@gcekbpatna.ac.in
[4]Associate Professor, Department of Computer Science and Engineering, Government College of Engineering, Kalahandi, Bhawanipatna, Odisha, 766003, Email ID: drnayak@gcekbpatna.ac.in
*Corresponding Author**:** Dillip Ranjan Nayak, Email ID: drnayak@gcekbpatna.ac.in

ABSTRACT
EYE sickness has emerged as one of the famous sicknesses in the world. One 10th of the population within the world suffers from eye disorder of different shape. Major cause of this ailment is inadequate information, preservation of hygiene, eating regimen, age and lack of awareness. In this paper early detection has been attempted through use of deep studying methods for the eye detection. A number of the situations like Glaucoma, cataract, everyday eye health and diabetic retinopathy are being identified early by using deep learning methods.
Labelled retinal photos are taken from public repository as dataset. MobileNetV2 is used as a characteristic extractor because of its performance in restrained material gear which will acquire the temporal and spatial features of eye pix, a hybridized version with CNN+BiLSTM is developed in supplement with the switch mastering. The accuracy of the version is improved by using great tuning, augmentation, records pre-processing and early stopping strategies which is covered within the education pipeline. The accuracy exceeds 90% which makes it healthy for use in the actual-global health care surroundings. The results show its performance in combining CNN with LSTM in category of clinical imaging dataset and the significance of deep gaining knowledge of in ophthalmology.

KEYWORDS: Attention Mechanism, BILSTM, Deep Learning, Computer-Aided Diagnosis, Eye Disease Prediction, Fundus Images, Image Classification, MobileNetV2, Model Evaluation, Retinal Disease, Squeeze-and- Excitation Block, Transfer Learning

1. INTRODUCTION

In 21st century, deep mastering and AI methodologies have substantially shaped development. these techniques have proved their effectiveness in offering right diagnosis in exclusive forms of eye related situations. In conventional approach, eye sickness analysis was achieved manually by way of healthcare experts. however traditional technique is time eating and is rather vulnerable to human blunders .for this reason, integration of artificial Intelligence and device studying strategies allows in stepped forward prediction in analysis of eye sickness. synthetic intelligence includes three kinds: preferred, slender and great AI and every of which pursuits to offer all the capabilities of human belief and intelligence.
Synthetic intelligence is a multidisciplinary which is a mixture of numerous methodologies like natural language processing, wise sensing structures, experiential and algorithmic getting to know, traditional device gaining knowledge of techniques and robotics and these elements together makes the machines to recognize, analyse, and acknowledge to complicated environment that mimics the functions of human belief. In the healthcare domain, especially for diagnosis eye-related issues, system gaining knowledge of provides very beneficial outcomes like in programs with image interpretation. Computerized text analysis and speech recognition. The improvement of gadget getting to know is prompted by traditional system learning strategies as well as modern day deep gaining knowledge of strategies. Specifically ophthalmology acts as a frontrunner in harnessing these methodologies to examine ophthalmic pictures and offer faster disorder detection. these smart structures propose the possibility to decide situations such as glaucoma and diabetic retinopathy efficaciously, also inside the unavailability of skilled clinical expert. With the developing dependence on AI, deep getting to know and ML strategies in healthcare, their obligation of mechanizing and enhancing accuracy of analysis is turning into very vital. The human eye, being delicate and complex, is willing to a numerous kind of ailments. If these conditions are not dealt with and diagnosed then severe outcomes can result in partial or complete loss of vision. some eye sicknesses are normally benign, others are progressive and may develop from a young age, deteriorating with time if no longer handled earlier than. some of the examples of such states accommodates:

1.1 Diabetic Retinopathy (DR): World health business enterprise reviews that Hyperglycemia comprises of plasma glucose whose depth when exceeds 7.0 mmol/L may cause metabolic deregulation. [4].Sustainable growth in blood

glucose introduces pathological shifts in neural and vascular tissues, that affects positive organs like the brain, kidneys, heart and retina. Especially, retinal microvasculature is vulnerable to hyperglycemic loss, which can lead to Diabetic Retinopathy (DR)—a state of affairs that indicates by using the degradation of retinal blood vessels. DR is one of the primary motives of person outbreak of vision incapacity and blindness. parent 1 display the morphological changes noticed inside the fundus of a DR affected patient.

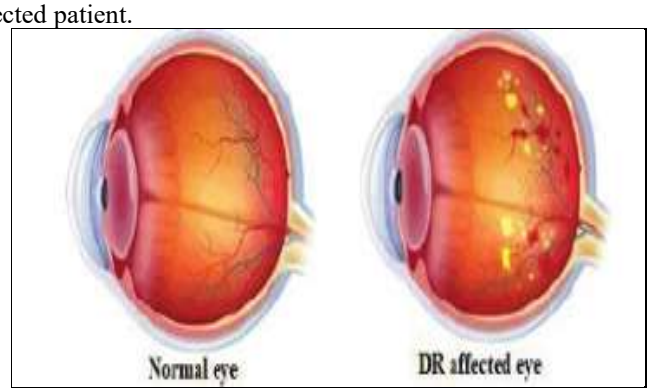

Figure 1. DR Affected Eye

1.2 Glaucoma: It is a widely emerging ophthalmic condition that relatively impacts the optic nerve, which leads to additional impairment. This neuro degeneration is mostly connected with raised intraocular pressure (IOP) which if not managed well then it may direct to complete blindness. In figure 2, the pathological outcomes of glaucoma on the retinal fundus is provided featuring the structural damage triggered by raised ocular pressure.

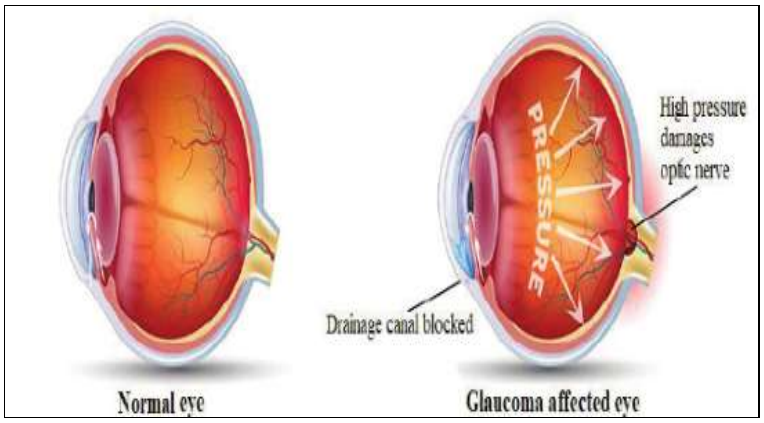

Figure 2. Glaucoma Affected Eye

1.3 Cataracts: Cataracts is a commonly emerging ocular disorder marked with the opacification of the eye's inherent lens that results in advancing visual damage. Cataracts occur commonly within the elderly people and is highly linked with uncertain aspects like tobacco use, extended disclosure to ultraviolet (UV) radiation and ageing [8]. Diagnosis mostly includes a tonometry to analyze intraocular stress, thorough ophthalmologic analysis with visual acuity trial, lens transparency and sit lamp test beneath a pupil expansion. The common treatment includes phacoemulsification surgery to erase the clouded lens and hence replaces with an intraocular lens (IOL) which is embedded to replenish vision [9]

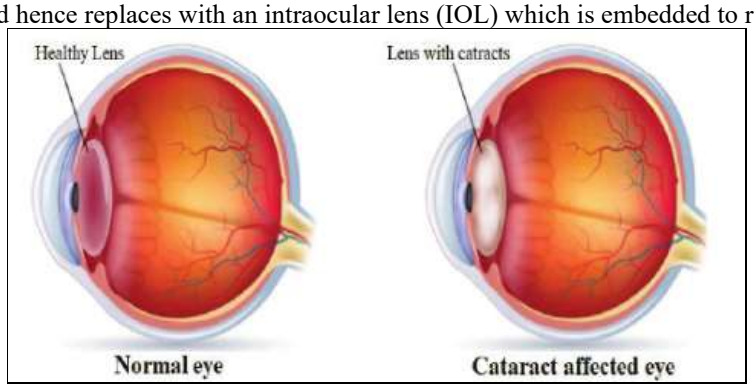

Figure 3. Cataract-Affected Eye

The paper is finally organized as below, section 2 and 3 show the related work with materials and methodology. Section 4 provides experimental results and section 5 gives the conclusion part.

2. RELATED WORK

With the improvement of various methodologies in the past years, many researches confirm the ability of AI-ML methods, that are associated with eye disease detection are mechanized. The VGG-19 architecture is used to categorize fundus

images from the ODIR dataset. Even though multiple binary tasks include high accuracy, they face extreme class disparity in the dataset. To remove this problem, the multiclass problem are transformed into multiple binary classifications, attaining 98.13% accuracy in some disease pairs. Sushma K. Sattigeri et al. (2022) had given a deep learning-based approach using CNNs that identify various eye conditions such as cataracts, conjunctivitis, and uveitis. Their approach, which combined visual features from single and dual-eye images, reported accuracy levels of 96% and 92.31%, respectively. However, the model's results were affected by inconsistencies in the dataset collected from local clinics and online sources. NoufBadha et al. (2022) experimented with both traditional machine learning classifiers and a CNN based on ResNet152. While the deep learning model outperformed ML classifiers with an accuracy of 84%, the study highlights the inadequacy of classical ML methods for high-stakes medical diagnosis. Grzegorz Meller et al. (2020) trained a CNN using the ODIR dataset to distinguish between normal and cataract fundus images. Although the model reached a peak accuracy of 93% in early experiments, overall performance was limited due to a lack of diverse class representation in the validation set. A hierarchical deep learning approach was proposed by Hao Gu et al. (2020), which addressed multi-label and multi-task classification of corneal diseases using a novel taxonomy-guided CNN. Their model achieved an AUC above 0.91 and rivalled ophthalmologist performance, but was constrained to ocular surface diseases only. Finally, Pawan Kumar Upadhyay et al. (2022) introduced a Coherent Convolution Neural Network (CCNN) for classifying OCT images into four classes, including diabetic macular oedema and drusen. Their modified VGG-16-based model achieved 97.16% accuracy, demonstrating the importance of pruning and hyper parameter tuning for improved diagnostic reliability.

3. MATERIALS AND METHODS

3.1 Retinal Fundus Images Fundus pictures is used to get high decision visualizations of internal shape of eye which encompass retina, macula, optic disc and posterior pole. those snap shots form an integral part to diagnose and determine numerous pathological exams together with diabetic retinopathy, glaucoma and cataract. in this take a look at 4217 annotated fundus photos were taken from a publicly to be had dataset in kaggle. The snap shots are grouped to 4 classes Cataract, Glaucoma, Diabetic Retinopathy and ordinary (healthy) retinas. each picture has been uniformly resized to 256×256 pixels. determine 1 represents a reference fundus photo used for the class undertaking

3.2 Pre-Processing

facts pre-processing possess an critical assignment for photo first-class enhancement which similarly results in correct segmentation and classification. In presence of noise fundus pictures frequently suffer because of illumination, contrast. here pix were resized, pixels had been normalized and filters had been applied to lessen noise.

pics are rescaled to a selection between [-1, 1] using lambda normalisation method. For noise elimination Gaussian blur and non-nearby approach de-noising are used. Denoising preserves texture and area details by way of averaging similar patches across the photo. color space adjustments had been also done to transform RGB pics to HSV format for higher highlighting chromatic functions.

3.3 Image Enhancement and Filters

image photograph filtering and enhancement play an crucial part in optimising datasets for deep getting to know models, in particular when running with raw or noisy records. with the aid of applying a sequence of pre-processing techniques, the pix can be delicate to extract the most useful functions, making them extra appropriate for the training system. below are extra explanations of the steps involved:

3.4 Gaussian Blur:

To reduce noise and to smoothen the picture Gaussian blur is used. on this approach a bell shaped curve is used to calculate the weighted average of neighbouring pixels. This filter is used to reduce excessive frequency noise. Gaussian blur keeps the structural integrity of the photograph which in flip ensures the critical functions continue to be visible. The important step increases the clarity of the picture and makes the version to examine required patterns with out being tormented by the pixel-level noise.

$$G(x,y)=1/(2\pi\sigma^2)\, e^{(-(x^2+y^2)/(2\sigma^2))} \quad (1)$$

3.5Non-nearby approach Denoising:

As in keeping with classical method this method serves more complicated results. Denoise of every pixel is finished by the use of the weighted common of equal patches. This approach provides exceptional components like textures and edges a great deal better than other techniques.

3.6Filtered photograph

a) switch gaining knowledge of the use of MobileNetV2

MobileNetV2 is a very effective convolutional neural network for the applications of embedded and cell vision. MobileNetV2 is a successor of MobileNetV1. MobileNetV2 presents large accuracy with the aid of the exploitation of specific architecture primarily based on intensively divisible convolutions and reversed residuals. depth-smart Separable Convolutions: they're utilized in MobileNetV2 which divides the unique convolution system into two parts:

- point-sensible Convolution: A 1x1 convolution matrix is a mixture of the consequences related to the depth-wise convolution.
- depth-wise Convolution: every and every channel input is one by one filtered.

This approach facilitates in reducing the cost of computation without affecting the vividness of the version and is good for the resource and cell constrained environment.

b)Inverted Residual Blocks:

The inverted residual block is the main feature of MobileNetV2. It exploits an powerful architecture inside the following mannerinput is first improved in a better dimensional area the usage of a light-weight 1x1 convolution (factor-sensible convolution).First linear bottleneck is applied then depth-clever separable convolution is implemented unique input is

mixed with the version output thru bypass connection

c)Linear Bottleneck:

This bottleneck layer allows recover effectiveness by using proscribing the variety of operations inside the network without compromising the version's potential to study. To lessen general complexity very last output has to have fewer channels for this linear bottleneck layer is used.

d)ReLU6 Activation:

In place of fashionable ReLu , ReLu6 is used as activation function in MobileNetV2. ReLu6 constricts the output to the range [0,6]. in this way big activations are averted which is proved deadly in mobile and embedded devices with restricted hardware resources.

e)international average Pooling (hole):

Upon completion of the very last convolutional layers, MobileNetV2 applies international common Pooling (gap) to compress the spatial size of the function maps. This system yields a steady, fixed-length output irrespective of the input image dimensions, which is finally forwarded to a completely connected layer to have in addition improvements.

Implementation details in the Context of MobileNetV2:

1. **Data Loading:**

The training and validation datasets are loaded from a directory. Then, the dataset is separated into training and authentication subsets, with training data consisting of 80% of the data, and the authentication and testing data are 10% respectively for each. The images which are in the dataset are altered to (224, 224) to match MobileNetV2's input size, and a batch dimension of 32 is used.

2. **Data Preprocessing:**

The pre-process input function from MobileNetV2 is used to adjust the images to the appropriate format for feeding into the model. This function normalises the pixel values and performs any required transformations to match the preprocessing steps used during the training of MobileNetV2 on ImageNet.

3. **Transfer Learning Setup:**

The base model, MobileNetV2, is loaded with weights.

The feature maps produced by the base model are first processed through a GlobalAveragePooling2D layer, which compresses their spatial dimensions to create a concise and informative feature vector. This is followed by a dense (fully connected) layer comprising 128 units with ReLU activation for non-linear transformation. Then a soft-max-initiated outer-layer is added that contains num_classes neurons equivalent to the total target classes in the dataset.

4. **Model Compilation:**

Configuration of the system is done using Adam Optimizer that is adopted vastly because of its adaptive learning and effectiveness. The loss function used here is sparse class cross entropy which are best used for multiclass pattern recognition. Moreover, the system is checks the accuracy thoroughly during the training period.

5. **Model Training:**

The base system is trained using fit method for 10 iterations. At this part, the base system of MobileNetV2 is preserved and the latest fully associated layers are only trained.

6. **Unfreezing for Fine-Tuning:**

After the initial part of training, the base system is temporarily not associated to allow optimizing. This process involves unlocking the deeper layers of the pre trained network, allowing them to be retrained for better alignment with the specific task, all while maintaining computational efficiency.

7. **Recompiling the Model:**

After unfreezing the layers, the model is again compiled with a lower learning rate(1e-5). Fine-tuning is typically done with a lower learning rate to prevent large updates to the pre-trained weights, ensuring that they don't change too much from their learned values.

8. **Callbacks:**

Early Stopping: This call back observes the validation loss during training and halts the process and is there is no improvement detected over a set number of consecutive epochs (with a patience value of 3). This mechanism helps to avoid overfitting and unnecessary computation.

Reduce LR On Plateau: Whenever the validation loss stagnates for two successive epochs, this callback automatically decreases the learning rate by a component of 0.5, enabling the framework to converge more effectively in later training stages.

9. **Fine-Tuning:**

Finally, the framework is refined for an additional 20 epochs with the callbacks applied to optimise the model further. This phase adjusts the weights of the deeper layers of the MobileNetV2 model to improve its accuracy on the new dataset.

3.7 Hybrid Model Hybrid Model Architecture: CNN + BiLSTM

To capture spatial and sequential patterns in the images, a hybrid model architecture was used. The extracted features from MobileNetV2 were reshaped into a sequential format and passed through Bidirectional Long Short-Term Memory (BiLSTM) layers. BiLSTMs learn contextual correlations in both backwardsand forward orientations, which is useful in recognising complex visual patterns in medical imaging.

Model Architecture

Here, the study gives a deep learning-based architecture for image classification, which includes the Convolutional Neural Networks (CNNs), Squeeze-and-Excitation (SE) blocks, and Bidirectional Long Short-Term Memory (BILSTM) layers. The proposed model uses MobileNetV2 fundamental backbone by transfer learning, which increases with the additional components that focus at fine tuning the attribute extraction and grabbing the time-related connections. The model's main parts and their specifications are discussed below:

a. **MobileNetV2 Backbone**

It acts as the main attribute extractor in the architecture of the proposed framework. It is a light weight CNN that is framed for supply control surroundings. MobileNetV2 provides robust performance while achieving decreased computational burden by implementing the depth-wise removable convolutions. Here, the transfer learning enables faster convergence and also leverages feature representation learned from a large-scale dataset.

In the implementation, the initial 100 layers of the MobileNetV2 backbone are frozen to retain general-purpose features, while the remaining layers are fine-tuned. This selective training approach helps minimise over-fitting and allows to have the domain-specific characteristics by the model.

b. **Squeeze-and-Excitation (SE) Block**

The Squeeze-and-Excitation (SE) block is incorporated into the model to enhance its representational power by capturing channel-wise relationships. It works by first applying Global Average Pooling (GAP) to the feature maps, after which two fully connected layers—one for reducing the dimensionality and the other for recalibrating the features. The result is a channel-specific attention map provided to the input attribute map, strengthening important channels and diminishing irrelevant ones. This attention technique model makes the model to rank the main features that can show greater performance in the complicated image classification jobs.

c. **Bidirectional LSTM (BiLSTM) Layers**

BiLSTM layers are introduced to increase the learning capability of the model in sequenced data and to have elongated range of connections. This helps the system to use the contextual knowledge from both past and future datapoints. In image classification case, the feature maps generated by CNN layers act as sequences and the BiLSTM layers enable the system attachment between the attributes throughout the image's spatial aspects.

The BiLSTM layers are situated after the convolutional and SE layers, allowing the model to learn temporal dependencies between features before making the final classification decision.

d. **Convolutional and Fully Connected Layers**

After the layers of SE and BILSTM some convolutional layers are added to consider the high-level attributes which follows with max-pooling that is responsible for down sampling. These layers of convolution use RELU activation function which can be used for complex environment and hence we can add more complicated model additional to this framework. To increase the convergence speed and the stability of training we apply normalization method.

In the end, after the LSTM layers, it is connected to fully linked dense layers having RELU which is next layered by soft-max activation for multi-category classification. There is generation of probability over all the categories by soft-max layers in which all the output specifies the likelihood that each of the input images concerns to a particular class.

3.8Model Compilation And Training

The system compilation is done using Adamn optimizer having a learning rate of 1×10−51 \times 10^{-5}1×10−5 to fine tune the weights of MobileNetV2 which are pre-trained with less modification. The employment of a loss function that is a sparse class cross-entropy which is best for the multi-pattern classification work with integer labelling. Overfitting can be prevented and convergence rate can be improved during training by using early stopping criteria and reducing the LR On Plateau callbacks. The Early Stopping callback terminates the training process if the verification loss fails to improve for five consecutive epochs, while Reduce LR On Plateau adjusts the learning rate if the verification loss plateaus

.

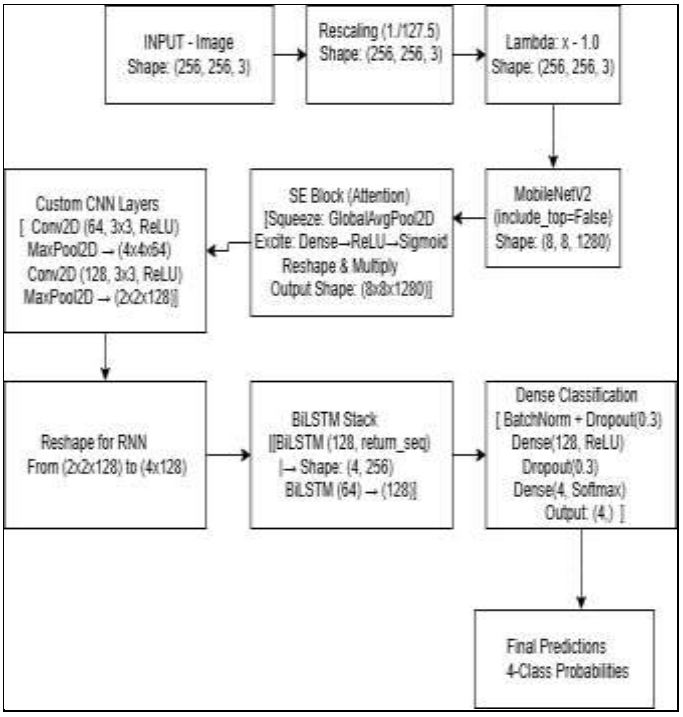

Figure 4. Proposed Architecture

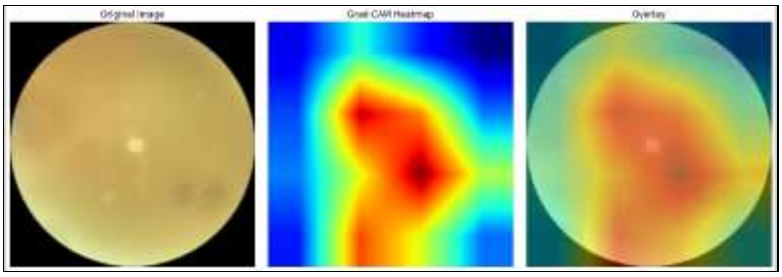

Figure 5. GRAD CAM Representation

3.9Explainable AI (XAI) Techniques

Two types of XAI techniques, which are GRAD-CAM and LIME are implemented, which basically give visually improved regions that need to focused on and can provide more clarification about the diseases as well.

a)Gradient-weighted Class Activation Mapping (Grad-CAM)

Gradient-weighted Class Activation Mapping (Grad-CAM) is an explainable AI (XAI) technique that offers visual insights into the decisions made by convolutional neural networks (CNNs). It creates class-specific heatmaps, and these heatmaps highlight the regions that strongly influence the framework's forecast, thus enhancing the interpretability of the model, an essential feature in sensitive areas such as medical imaging and fault detection.

Unlike saliency maps, which focus on the input layer, Grad-CAM operates on the spatial dimensions of intermediate convolutional layers. This is because convolutional layers maintain spatial hierarchies and semantic information, which are critical for understanding the parts of the image of the image model attends to in making its forecast.

The following implementation consists of two parts:

1. Generating the Grad-CAM heatmap
2. Overlaying the heatmap on the original input image

Generating the Grad-CAM Heatmap:

The function constructs a gradient model using the input model's final convolutional and output layers.

It computes the gradient of the class prediction (either user-specified or the top predicted class) by having the activations of the final convolutional layer.

Here, a weighted mixture of the attribute maps that produces the raw heatmap, which is then normalized to the range [0,1][0, 1][0,1].

Overlaying the heatmap on the original input image

Converts the input tensor to a NumPy array and ensures it's in RGB format.

Rescales the heatmap to the original image size and applies the JET colourmap for better visual distinction.

Superimposes the heatmap on the original image using an adjustable transparency parameter alpha.

Displays three plots side-by-side: original image, heatmap, and the overlay for comprehensive visual analysis.

b) Local Interpretable Model-Agnostic Explanations (LIME)

LIME (Local Interpretable Model-Agnostic Explanations) is an interpretability method that can be applied to explain individual predictions from complex machine learning models, regardless of their type. Unlike Grad-CAM, which is specifically designed for convolutional neural networks (CNNs), LIME can be used with any black-box model, including deep learning networks

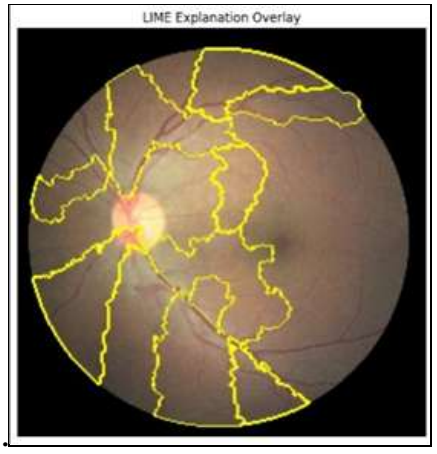

Figure 6: Lime Explanation

LIME provides an idea of using a local model which is better interpretable and simple than the whole complex model. This can be achieved by modifying the input data, gathers the system's classifications for these modified inputs and trains a substitue model which is estimated with the actual model using original datasets. LIME then identifies the input components that were further impactful in producing the framework's output, offering a valuable way to understand and validate machine learning models, particularly in sensitive areas like medical image analysis. In image classification, for instance, LIME highlights superpixels (groups of connected pixels with similar colour and texture), which gives important

information which have the most significant impact on the model's prediction for a given class.
The following implementation consists of two parts:
a) Generating the LIME explanation
b) Visualizing the explanation as an overlay on the original input image.

4. FINDING AND ANALYSIS

The dataset used for this research consists of 4,217 retinal fundus images, which were obtained from a publicly available Kaggle repository. The dataset contains 4 classes: Diabetic Retinopathy (1,098 images), Cataract (1,038 images), Normal (1,074 images) and Glaucoma (1,007 images). As discussed in table 1 the images are evenly distributed across four classes: cataract, diabetic retinopathy, glaucoma, and normal. These high-resolution images capture detailed retinal structures necessary for disease detection. The images were being resized to 256×256 pixels and normalized before being used for model training.

Table no. 1: Dataset Details

Label	Disease Type	Count
0	Glaucoma	1007
1	Cataract	1038
2	Normal	1074
3	diabetic retinopathy	1098

The proposed hybrid architecture (CNN + BiLSTM) was trained and evaluated using an 80:10:10 split for training, validation, and testing datasets. The training was conducted on Google Colab with GPU acceleration. This partitioning approach shows that the framework is trained on the majority of the data, allowing it to learn effectively, while also having separate datasets for validation and testing to assess its performance. The model initially used **MobileNetV2** as a frozen feature extractor and was later fine-tuned on deeper layers.
These results highlight that the proposed model is highly accurate and reliable, positioning it as a strong candidate for deployment in real-world medical image classification tasks, especially in the diagnosis of eye diseases.
Experiments were carried out in two phases: before fine-tuning and after fine-tuning. Then the model's performance was monitored by analysing the training and verification accuracy and loss curve during both phases.

4.1 Before Fine-Tuning

The model initially achieved final training accuracy of 0.9019 and final validation accuracy of 0.8814
As shown in the following Figure 9, the training accuracy increased steadily while the validation accuracy plateaued after a few epochs. Similarly, training and validation losses as depicted in Figure 10 both decreased consistently, indicating that the model was learning effectively with no signs of over fitting.

4.2 After Fine-Tuning

The performance was further improved through fine-tuning of deeper layers in the MobileNetV2 architecture.
After fine-tuning: Final Training Accuracy of 0.9843 and Final Validation Accuracy of 0.9205 are obtained. The training accuracy sharply improved, while the validation accuracy also showed noticeable gains, suggesting improved generalization capabilities. As seen in Figure 11, training loss significantly dropped to 0.5907, while validation loss was reduced to 0.8014, further confirming enhanced performance.

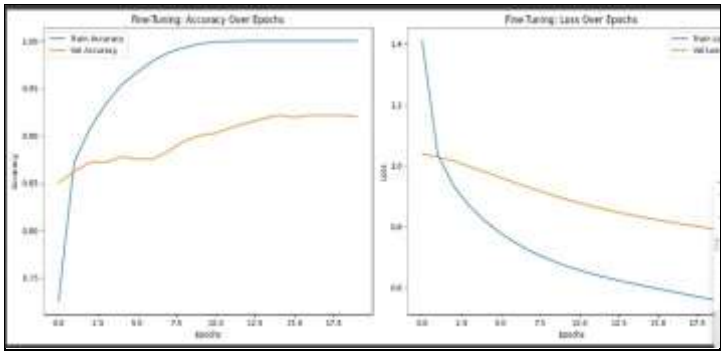

Figure 7. MobileNetV2 before Fine-Tuning

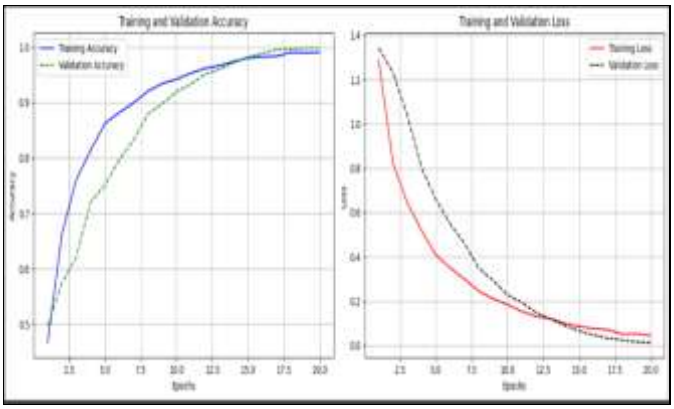

Figure 8.MobileNetV2 after Fine-Tuning

4.3 Base Model Results

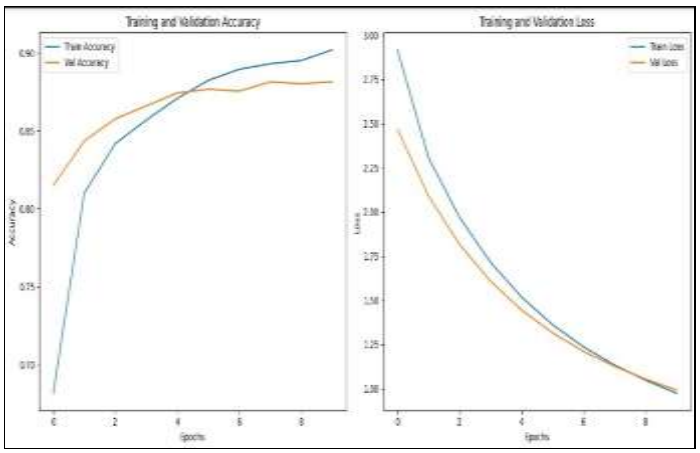

Figure9 . Base Model Performance

In this section a model is proposed that is based on deep learning for eye disease classification by combining MobileNetV2, a Squeeze-and-Excitation (SE) block, and Bidirectional Long Short-Term Memory (BILSTM) layers for feature extraction and classification. The model was fine-tuned on the eye images consisted by the dataset, which consists of four different classes: glaucoma, cataract, normal, and diabetic retinopathy.

The MobileNetV2 architecture was initially pre-trained on the Image Net and for further advance we had used the fine-tuning on our dataset. To improve the model's performance, we integrated an SE block. This block helps the model prioritise the most informative features, leading to improved classification accuracy. Furthermore, a Bidirectional LSTM layer is used for sequential context learning. The model as depicted in figure 11 demonstrated outstanding performances, achieving a final training accuracy of 99.05% and a validation accuracy of 99.83%. The test accuracy was also 99.83%, showcasing the model's excellent generalisation ability. The integration of transfer learning with MobileNetV2, SE blocks, and Bi LSTM layers proved to be highly effective for the eye disease classification task, delivering both high accuracy and robustness.

Table No. 2

Epoch	Training Accuracy	Validation Accuracy
1	0.3806	0.4970
2	0.6284	0.5748
3	0.7378	0.6215
4	0.7985	0.7204
5	0.8522	0.7505
6	0.8771	0.7965
7	0.8959	0.8321
8	0.9142	0.8791
9	0.9321	0.8978
10	0.9327	0.9191
11	0.9528	0.9329
12	0.9591	0.9509
13	0.9610	0.9599

14	0.9717	0.9720
15	0.9784	0.9817
16	0.9840	0.9881
17	0.9817	0.9955
18	0.9847	0.9962
19	0.9862	0.9981
20	0.9900	0.9983

The specifically, the results suggest that the architecture that we have proposed is proper for the real-world projects and applications, as well as in medical image analysis, particularly in the detection of eye diseases. The model demonstrated a steady improvement in both training and validation accuracy over 20 epochs as stated in table 2. Initially, the training accuracy was relatively low at 38.06%, with a validation accuracy of 49.70% in the first epoch. However, as training progressed, the model quickly learned meaningful patterns from the data. By epoch 10, training accuracy had surpassed 93%, and validation accuracy reached over 91%, indicating significant learning and good generalisation. From epoch 11 onwards, the model consistently achieved high accuracy, with both training and validation metrics improving in parallel. The final training accuracy reached 99.00%, while the validation accuracy peaked at 99.83% in the 20th epoch. This consistent increase without over fitting reflects a well-optimised learning process.The results confirm that the model not only learned efficiently but also generalised well to unseen data. These outcomes validate the effectiveness of the chosen architecture and training strategies, such as transfer learning and the integration of SE blocks and Bi LSTM layers.

4.4 Performance Measures

Table no. 3

Class	Precision	Recall	F1-score
Cataract	0.9961	1.00	0.9981
Diabetic retinopathy	1.00	1.00	1.00
Glaucoma	1.00	1.00	1.00
Normal	0.9972	1.00	0.9986

The following performance measures have been used as summarized in table 3 and 4.

1. Precision

Precision is a metric that is used to evaluate the performance of a classification model. It is calculated as:

$$\textbf{Precision} = TP / (TP + FP) \quad (2)$$

2. Recall (Sensitivity or True Positive Rate)

Recall is a metric used to measure how well a classification model identifies all actual positive cases. It is defined as:

$$\textbf{Recall} = TP / (TP + FN) \quad (3)$$

3. F1-Score

It is computed as:

$$\textbf{F1-Score} = 2 \times (\textbf{Precision} \times \textbf{Recall}) / (\textbf{Precision} + \textbf{Recall}) \quad (4)$$

4. Accuracy

Accuracy measures the overall correctness of the model; it is computed as:

$$\textbf{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

Table no.4

Metric	Precision	Recall	F! - Score	Support
Accuracy	-	-	1.0000	4217
Macro Average	1.0000	1.0000	1.0000	4217
Weighted Average	1.0000	1.0000	1.0000	4217

From Table No. 3 and 4, we get to know that the model demonstrates an outstanding classification performance across all four classifications, such as glaucoma, cataract, diabetic retinopathy and normal. For the diabetic retinopathy and glaucoma classes, the model provides ideal results in precision, recall, and F1-score (all equal to 1.00), which indicates the flawless detection and classification of these conditions. The cataract has a precision of 0.9961, a recall of 1.00 and a F1-score of 0.9981, similarly, the normal class unveils precision of 0.9972, a recall of 1.00 and a F1-score of 0.9986. These results adhere to the model's sturdiness and high consistency for distinction between healthy and pathological eye conditions.

Table no. 5

Label	Disease Type	Count
0	Glaucoma	1007
1	Cataract	1038
2	Normal	1074
3	diabetic retinopathy	1098

In Table No. 5, it is significant that the model delivers an overall accuracy of close to 1.00 on the test set. Meanwhile both macro average and weighted average values for precision, recall, and f1-score are also 1.000, which indicates the model's remarkable performance across all classes.

To approach the performance of the proposed framework in classifying eye diseases, training was conducted using the MobileNetV2 architecture with transfer learning. The confusion matrix is provided in figure 10.

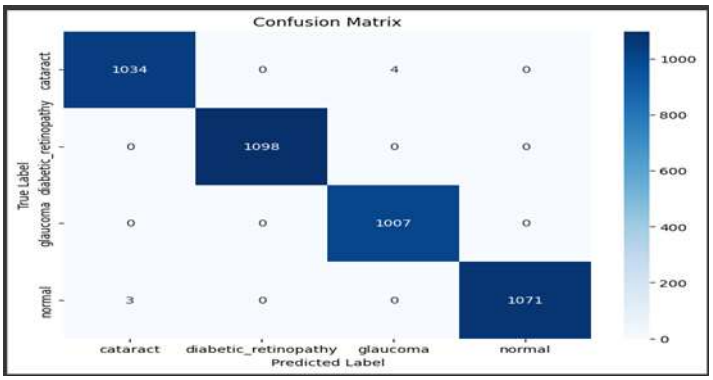

Figure 10. Confusion Matrix

5. CONCLUSION

To come across commonplace eye diseases the proposed model makes use of a aggregate of MobileNetV2, Squeeze and Exclusion (SE) blocks and BiLSTM layers. This hybrid model has been tested on a properly-dependent dataset from Kaggle after pre-processing and augmentation

The model did test accuracy of ninety nine.83%, with F1-scores of one.00 throughout all training. those effects exceed several triumphing techniques, which include conventional CNN architectures and ML-based fashions, signifying the usefulness of mixing spatial function extraction with temporal series modelling. furthermore, Explainable AI (XAI) strategies consisting of Grad-CAM and LIME were implemented to provide transparency in predictions, that's especially critical in scientific settings.

This version has enormous potential for real-world disposition in ophthalmology, in particular in supporting early identification and broadcast in far off or aid-restrained areas. Its high achievement and reliability make it a highly-priced device for ophthalmologists and healthcare professionals.

In future the model may be extended to OCT image datasets incorporating real time classification and better detection of retinal diseases

REFERENCES

[1] S Das ,S Mishra., & M Senapati, ‖ Improving time series forecasting using elephant herd optimization with feature selectionmethods‖,2021, Journal of Management Analytics, 8(1), 113–133. https://doi.org/10.1080/23270012.2020.1818321.

[2] M. Nayak, S. Das, M.R. Senapati, ―Improving flood prediction with deep learning methods‖, Journal of The Institution ofEngineers (India): Series B, vol.103, pp. 1189–1205, 2022.

[3] M. Nayak, S. Das, U. Bhanja, M. R. Senapati, ―Predictive analysis for cancer and diabetes using simplex method based socialspider optimization algorithm‖, IETE Journal of Research, vo. 69, no.10, pp.7342–7356, 2023.

[4] S. Das, M. Nayak, M. R. Senapati and J. Satapathy, "Medical Data Classification Using Velocity Enhanced Whale OptimizationAlgorithm," ICACFCT, 2021, pp. 18-22

[5] S Das, A Patra, S Mishra and M R Senapati,‖A self-adaptive fuzzy-based optimised functional link artificial neural network model for financialtime series prediction‖2016,pp 55-77https://doi.org/10.1504/IJBFMI.2015.075358

[6] S Das, A Patra, S Mishra, MR Senapati ,‖A self-adaptive fuzzy-based optimised functional link artificial neural network model forfinancial time series prediction ―Neural Computing and Applications,Springer

[7] P Mohapatra, S Das,‖ Stock market prediction using bio-inspired computing: A survey‖International Journal of Engineering Science and Technology 5 (4), 739.

[8] S Das, M Nayak, MR Senapati, J SatapathyMedical data classification using velocity enhanced whale optimization algorithm 2021First International Conference on Advances in Computing and Future Communication Technologies

(ICACFCT) ,IEEE.
[9] Maria Inês Almeida; Alvaro Rosa; Helena Castelão Figueira Carlos Pestana ―Emerging trends of artificial intelligence inhealthcare: a bibliometric outlook‖ International Journal of Healthcare Technology and Management,Inderscience,2024 Vol.21No.1,DOI: 10.1504/IJHTM.2024.136548.
[10] Abbas Sorkhi;,FaramarzPourasghar ―Artificial intelligence application in healthcare management: a review of challenges‖,International Journal of Healthcare Technology and Management,Inderscience,2024 Vol.21 No.1DOI: 10.1504/IJHTM.2024.136547
[11] S. K. Sattigeri, N. Harshith, G. N. Dhanush, K. A. Ullas, and M. S. Aditya, ―Eye disease identification using deep learning‖, IRJET, vol. 9, no.7, pp. 974-978, 2022.
[12] N. Badah, A. Algefes, A. AlArjani, and R. Mokni, ―Automatic eye disease detection using machine learning and deep learningmodels‖, Pervasive Computing and Social Networking, pp. 773–787, 2023
[13] G. Meller, ―Ocular disease recognition using convolutional neural networks,‖ Towards Data Science, 2020..
[14] H. Gu et al., ―Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs,‖ Sci. Rep.., vol. 10, no.1, p. 17851, 2020.
[15] P. K. Upadhyay, S. Rastogi, and K. V. Kumar, ―Coherent convolution neural network based retinal disease detection usingoptical coherence tomographic images,‖ J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 10, Part B, pp. 9688–9695, 2022