**Mini Project Report on**



**Detection of Diabetic Retinopathy using Convolution Neural Network (CNN) and Explainable Artificial Intelligence (XAI)**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Ashutosh Saurabh Pandey 2021145**

***Under the Mentorship of***

**Jyotir Moy Chatterjee**

**Assistant Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January-2025**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Detection of diabetic retinopathy using Convolution Neural Network (CNN) and Explainable Artificial Intelligence(XAI) ”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Jyotir Moy Chatterjee, Assistant Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name :-Ashutosh Saurabh Pandey University Roll no :-2021145

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 1 |
| Chapter 2 | Literature Survey | 4 |
| Chapter 3 | Methodology | 6 |
| Chapter 4 | Result and Discussion | 10 |
| Chapter 5 | Conclusion and Future Work | 12 |
|  | References |  |

**Chapter 1**

**Introduction**

Diabetic retinopathy is the main cause of blindness and low vision all over the world, and mostly in people who have endured the disease for many years. With its rapid growth around the world, diabetic retinopathy has been one of the pressing public health concerns and a signal for immediate diagnosis and treatment[2]. This work discusses some issues concerning early detection and classification of diabetic retinopathy with the use of sophisticated techniques of image processing and machine learning[3].

A diagram of a diabetic retinopathy

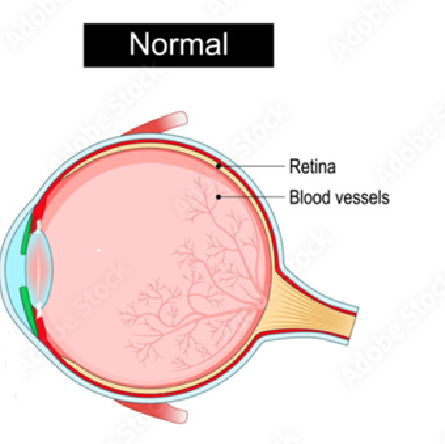
Description automatically generated 



Figure 1.1:- Image of the eye during diabetic retinopathy[6]

This study can contribute to the development of more precise, efficient, and accessible diagnostic tools. Diabetic retinopathy diagnosis by a simple manual evaluation of the retinal images by an ophthalmologist is time-consuming and inaccessible in some low-resource settings. The automated approach is an opportunity in these areas to ensure timely intervention and enhanced outcomes for the patients[2].

The primary impetus for this research arises from an increasingly pressing call for scalable solutions that can assist professionals in the medical field to identify stages of retinopathy, ranging from mild to more severe degrees, and from preventing further deterioration of vision while alleviating further burdens on healthcare services.

This chapter defines the problem statement and describes the existing approaches toward the detection of diabetic retinopathy and their limitations. It also formulates the objectives of the study, which are to develop a robust image classification model capable of distinguishing between normal and affected retinal images, to integrate explainability techniques for enhanced clinician trust in the model, and to evaluate the solution's effectiveness in real-world scenarios[2].

Addressing such goals, the contributions of this study to automated analysis of medical images would thus provide improved diagnostics, specifically diabetic retinopathy diagnosis.

**1.1 Introduction**

Diabetic retinopathy is a critical condition resulting from damage of the blood vessels of the retina from prolonged high sugar levels in the blood. If left untreated and unnoticed, DR can cause permanent loss of vision[3].

Current Challenges: - Manual diagnosis: Time-consuming with expert professionals and not easily accessible in resource-scarce areas.

Automated detection methods are plagued by non-interpretability and poor generalization to diversified datasets.

**1.2 Problem Statement**

Diabetic retinopathy is the most common cause of blindness in the world. For the diabetic patient, vision loss is irreversible unless DR is identified early and treated on time. AI has the potential to automatically diagnose DR from retinal images. Its clinical applicability, however, remains very limited owing to the following reasons:

* Lack of Interpretability [5]
* Poor Generalization[2]
* Overfitting and Model Bias

Proposed Solution: - This is a research proposal in the development of a strong, interpretable, and accurate AI model for automated detection of DR. The solution will comprise the following elements:

* Utilization of CNNs for classification
* Integration of SHAP (Shapley Additive Explanations) for the provision of visual explanations for aiding in clinician trust [5]
* Data augmentation and regularization for the purpose of providing generalization and reducing model bias.

**1.3 Objectives**

* Development a CNN to classify retinal images as normal or indicative of diabetic retinopathy.
* To enhance the model’s explainability by SHAP (Shapley Additive Explanations Technique).
* To ensure the model achieves a good accuracy rate in detecting DR.
* To provide interpretability visual feedback to build trust on model decision making.

.

The challenges in the above descriptions to address the field of automated medical image analysis. Precisely, this is an attempt at further improving DR diagnosis with XAI techniques in mind, also taking into consideration issues of interpretability, robustness, and scalability. This developed system will be able to significantly change the patterns of DR diagnosis, particularly among underserved populations, and become a key component for early diagnosis and treatment[2][3].

**Chapter 2**

**Literature Survey**

Detection of diabetic retinopathy has been the core focus in the area of medical imaging because, as a prevention technique, early detection helps prevent further loss of vision. Several researchers have explored automation as a way of solving this problem.

**Gulshan, V., Peng, L., Coram, M., et al. (2016)**

Gulshan and fellow researchers developed CNN-based models with a high performance for automated diabetic retinopathy detection, though sensitive and highly specific on very particular datasets used. The application has the prospect of early diabetic retinopathy screening and has reduced clinician workload.

However, it lacked interpretability, which was one of the biggest limitations because it did not reveal the features behind its predictions. This made validation of the outputs by clinicians quite difficult and left questions about the trustworthiness of the model in real-world clinical settings. [2].

**Abramoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018).**

Abramoff et al. introduced an AI system for referable DR detection, which became the first to receive FDA approval. Thus, their model was trained with a comprehensive dataset and performed well on specific benchmarks, which always relates this work to significant milestones about AI in the healthcare sector.

The system has difficulty generalizing to diverse datasets from different populations and imaging conditions, which calls for more robust training pipelines and diversity in datasets toward consistent performance on various real-world scenarios [3].

**Raj, A., Kumar, S., & Singh, K. (2020).**

Abramoff et al. introduced the AI system, detecting referable DR, the first to achieve approval from FDA. Their model achieved excellent results when trained over the comprehensive dataset as well as against specific benchmarks; it indeed has been an AI-driven significant health milestone.

The system did not generalize to other diverse datasets coming from other populations and different imaging conditions. This limitation underscores the need for stronger training pipelines and diverse datasets in order to consistently perform well in all scenarios of the real world. [4].

**Lundberg, S. M., & Lee, S.-I. (2017)**

Lundberg and Lee created SHAP, or SHapley Additive exPlanations, a framework for generating consistent, visually interpretable explanations of AI predictions. SHAP has been widely applied across finance, healthcare, and other domains in efforts to improve model transparency.

While SHAP is promising for making AI models interpretable in medical imaging, integrating it into the workflow for DR detection remains largely unexplored. Applications so far have been centered on general AI tasks, and few examples have used SHAP to provide insights relevant to a clinical setting, such as localization of lesions or grading of DR severity [5].

**Chapter 3**

**Methodology**

The steps taken to complete the projects are :-

* Data and image preprocessing
* Model development
* Training
* Evaluation
* Shap

**3.1 Dataset and Image Preprocessing**

**Dataset:-** Asia Pacific Tele-Ophthalmology Society(APTOS) diabetic retinopathy dataset[1] was used for training and evaluation." containing 3362 images 1805 of no diabetic retinopathy and 1857 of diabetic retinopathy at different stages.

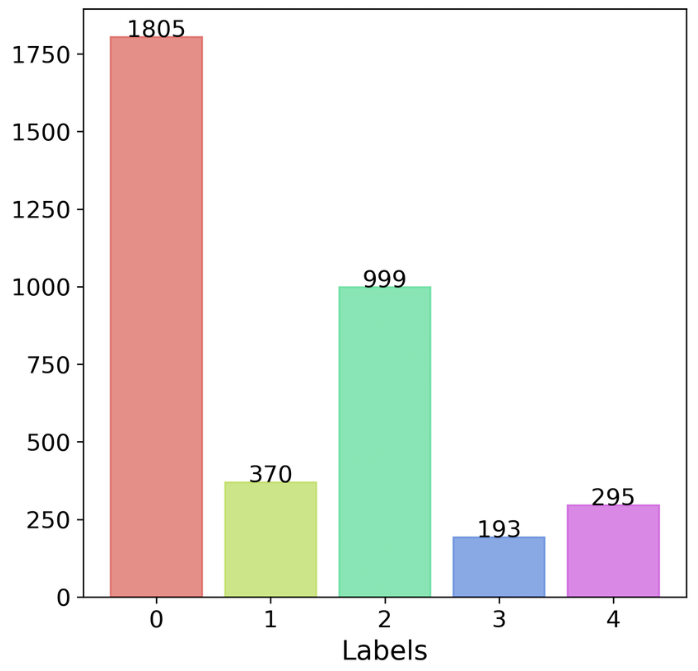


Figure 3.1:-Data distribution

Converted original multi-class labels (severity levels 0 to 4) to binary labels, where any severity greater than 0 is labeled as 1 meaning Retinopathy present

A graph of a number of retinopathy cases

Description automatically generated.

Figure3.2:-Data changed diff labels to 0,1 for binary classifiation

**Image Preprocessing**:

* Images are resized to a fixed size (128x128 pixels).
* Gaussian blur is applied to each image to reduce noise.
* The images are normalized to have pixel values in the range of 0 to 1.
* Split the dataset into training, validation, and testing sets using an 80/20 split. The test set is further divided into 50% validation and 50% test.

A close up of a ball

Description automatically generated A blue circle with a black background

Description automatically generated

Figure 3.3:- Image before processing (left) and after processing (right)[1].

**Model Development**:

Created a CNN model with multiple convolutional layers, max-pooling, batch normalization, dropout, and a final sigmoid activation for binary classification.

**Architecture**:

A Convolutional Neural Network (CNN) is designed with:

* Two convolutional layers followed by max-pooling layers to extract features from the images.
* Batch normalization to stabilize the learning process.
* A dense fully connected layer with ReLU activation to learn complex patterns.
* A final dense layer with a sigmoid activation function to output a binary classification (0 or 1).

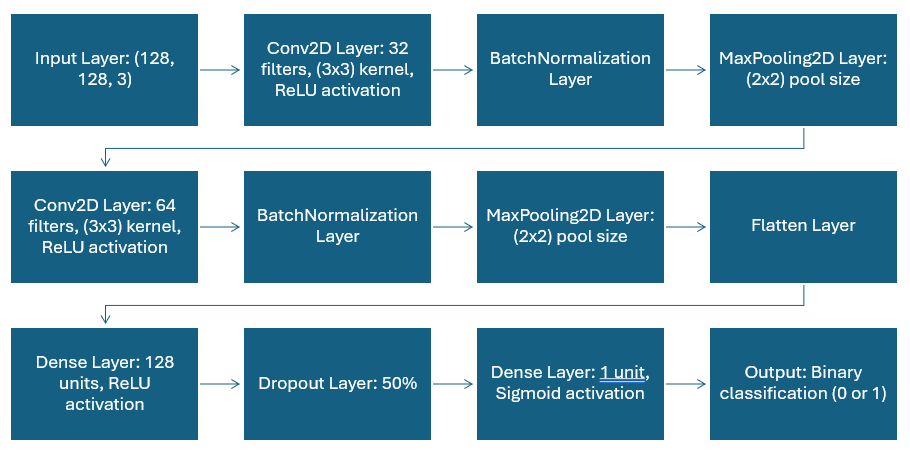


Figure 3.4: -Model Architecture

The model is compiled with:

* Adam optimizer for efficient training.
* Binary cross-entropy loss function, as this is a binary classification problem.
* Accuracy as the evaluation metric.

**3.3 Training**

* **Data Augmentation**: During training, images are augmented using random transformations such as rotation, translation, zoom, and horizontal flip to increase the variety and size of the training set, helping improve generalization.
* **Callbacks**:
  + Early Stopping to prevent overfitting by stopping training if the validation loss doesn’t improve
  + ReduceLROnPlateau to decrease the learning rate when the validation loss plateaus.
* The model is trained for up to 20 epochs, and the best model weights are restored based on validation loss.
  1. **Model Evaluation: -**
* After training, the model is evaluated on a separate test set (20% of the data).
* The final test accuracy is reported.

**3.5 SHAP: -**

* A background dataset is selected (a subset of the training set) to provide context for the SHAP explanation.
* The SHAP explainer is set up using the Gradient Explainer, which is suitable for CNN models.
* SHAP values are computed for a given test image to identify the parts of the image that contributed most to the model’s prediction.
* SHAP values are put on the input image, highlighting important regions.
* The values are normalized and visualized using a color map (e.g., coolwarm) to show which areas of the image were most important for the classification.

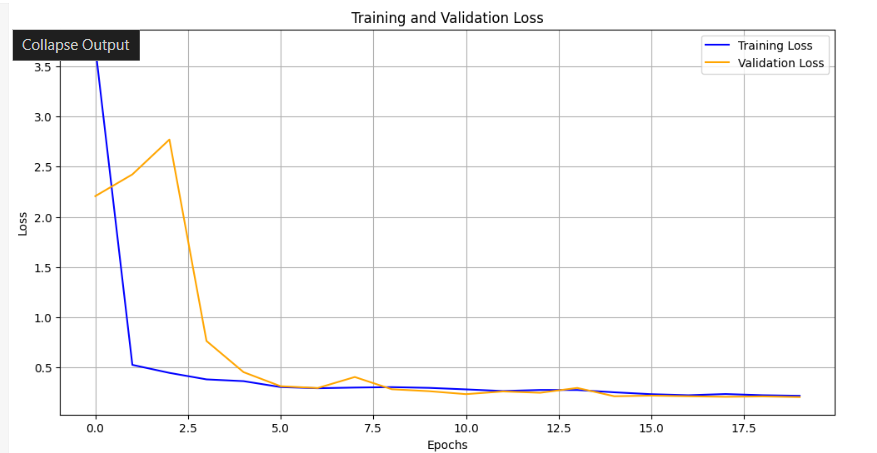
**Chapter 4**

**Result and Discussion**

The model has:-

accuracy of 94.00%

Val accuracy –93.44%

A graph with blue lines and a line

Description automatically generated

Figure 4.1:-Training and validation Loss and accuracy

* **Prediction**: For a given image, the model predicts whether the image shows retinopathy or not, with a confidence score (between 0 and 1).
* **SHAP Interpretation**: The SHAP values are mapped on the image showing which parts of the retina were most important for the prediction ( model interpretability).

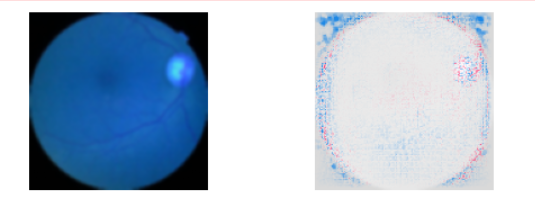


Figure 4.2:- showing calculated shap values[1]

* **Potential Use**: The ability to identify and explain model decisions can help clinicians trust and understand the AI model, potentially aiding in more accurate diagnoses and faster identification of diabetic retinopathy.

**Chapter 5**

**Conclusion and Future Work**

**Conclusion:-**

The study successfully developed a CNN-based model for binary DR classification, achieving high accuracy (94%) and integrating SHAP for enhanced interpretability. The visual explanations provide a valuable tool for clinicians to verify model predictions, fostering trust in AI-based systems.

**Future Work: -**

* Extend the model to multi-class classification to detect varying DR severity levels.
* Incorporate larger and more diverse datasets to improve generalization.
* Exploring different data preprocessing techniques.
* Explore advanced explainability techniques which will help in more depth explanation of the images.

**References**:

1. Asia Pacific Tele-Ophthalmology Society (APTOS), "APTOS 2019 Blindness Detection Dataset," 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection.
2. V. Gulshan, L. Peng, M. Coram *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.
3. M. D. Abramoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, "Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices," *NPJ Digital Medicine*, vol. 1, no. 1, p. 39, 2018, doi: 10.1038/s41746-018-0040-6.
4. A. Raj, S. Kumar, and K. Singh, "Transfer Learning for DR Detection Using Pre-Trained CNN Models," *Computer Methods and Programs in Biomedicine*, vol. 184, p. 105271, 2020, doi: 10.1016/j.cmpb.2020.105271.
5. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017, doi: 10.48550/arXiv.1705.07874.
6. Shutterstock, "Diabetic Retinopathy Images," Shutterstock. Available: <https://www.shutterstock.com/search/diabetic-retinopathy>. [Accessed: Jan. 8, 2025].