Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated Detection of Elevated Intracranial Pressure via Fundus Image Analysis: A Deep Learning Approach

Anitha M
Dayananda Sagar College of Engineering
Impana J
Dayananda Sagar College of Engineering
Kirthana A
Dayananda Sagar College of Engineering
Kirupasri Sampath
Dayananda Sagar College of Engineering
Neha Raj
neharaj08032004@gmail.com
Dayananda Sagar College of Engineering

Research Article

Keywords: Papilledema, Pseudopapilledema, Intracranial Pressure, Fundus Image Analysis, Deep Learning, Explainable AI, EfficientNet, Automated Diagnosis

Posted Date: April 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9339971/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Automated Detection of Elevated Intracranial Pressure via Fundus Image Analysis: A Deep Learning Approach

Anitha M[1*], Impana J[1], Kirthana A[1], Kirupasri Sampath[1], Neha Raj[1]

[1*]Department of Computer Science & Engineering, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India .

*Corresponding author(s). E-mail(s): anitha-cs@dayanandasagar.edu ;
Contributing authors: impanajgowda14@gmail.com;
kirthanaaa9@gmail.com; skiupasri@gmail.com;
neharaj08032004@gmail.com;

Abstract

Elevated intracranial pressure (ICP) is a critical neurological condition that requires timely detection to prevent vision loss and severe brain complications. Papilledema, characterised by swelling of the optic disc visible in retinal fundus images, is a key indicator of increased ICP. However, distinguishing papilledema from pseudopapilledema, which presents similar visual features but occurs without ICP elevation, remains a significant clinical challenge. This study presents a deep learning-based automated classification system for detecting ICP-related abnormalities using retinal fundus images. The proposed approach classifies images into three categories: Papilledema (ICP raised), Pseudopapilledema (ICP normal), and Normal. The EfficientNet-B2 architecture is trained on 1369 labelled fundus images using image preprocessing, data augmentation, and transfer learning to improve generalisation and classification performance. Experimental results demonstrate that the proposed model effectively differentiates these conditions, achieving a test accuracy of 98.55% and a validation accuracy of 99.51%. Explainable AI techniques including Grad-CAM and LIME are integrated to provide visual interpretability of model decisions. The system provides a non-invasive, reliable, and efficient diagnostic support tool for early detection of elevated ICP, potentially assisting clinicians in timely decision-making.

Keywords: Papilledema, Pseudopapilledema, Intracranial Pressure, Fundus Image Analysis, Deep Learning, Explainable AI, EfficientNet, Automated Diagnosis

1 Introduction

Intracranial pressure (ICP) is a fundamental physiological parameter that indicates the pressure exerted by the brain, cerebrospinal fluid (CSF), and cerebral blood volume within the rigid cranial vault. Under normal conditions, ICP remains tightly regulated—typically ranging between 7–15 mmHg in adults—to maintain adequate cerebral perfusion and ensure proper neurological functioning [3]. Any disturbance in the delicate balance between the intracranial components can cause an elevation in ICP, a condition that has the potential to become life-threatening if not identified and managed promptly. Conditions such as traumatic brain injury, hydrocephalus, idiopathic intracranial hypertension (IIH), brain tumours, meningitis, and intracranial haemorrhage frequently lead to raised ICP [23]. If left untreated, elevated ICP can compress neural structures, reduce cerebral blood flow, cause ischaemia, and eventually result in irreversible neurological damage or death [31]. Early detection and continuous monitoring of ICP are therefore critical aspects of modern medical practice, particularly in neurology, emergency medicine, ophthalmology, and neurosurgery.

Invasive procedures, including intraventricular catheters and intraparenchymal pressure monitors, remain the clinical gold standard due to their accuracy and ability to provide real-time pressure measurements. However, these techniques require surgical insertion, specialised neurosurgical expertise, hospitalisation, and sterile environments [23]. They carry significant risks such as infection, bleeding, catheter blockage, and postoperative complications, limiting their use to intensive care units and specialised centres.

Non-invasive techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and transcranial Doppler ultrasound evaluate structural and haemodynamic changes that may indicate elevated ICP, but they do not directly quantify ICP and their sensitivity varies. CT scans expose patients to ionising radiation, whereas MRI is expensive and time-intensive. Both require sophisticated infrastructure and skilled personnel, making them inaccessible in many rural and low-resource regions [24].

Fundus imaging has emerged as a highly promising and accessible alternative for ICP assessment. The optic nerve sheath is anatomically continuous with the subarachnoid space surrounding the brain; hence, any elevation in ICP is transmitted directly to the optic nerve head, resulting in swelling of the optic disc—a condition known as **papilledema** [31]. Fundus photography is non-invasive, inexpensive, rapid, portable, and highly suitable for point-of-care settings, making it ideal for large-scale screening and telemedicine applications (Figure 1).

However, the accurate interpretation of fundus images requires substantial clinical experience. Differentiating between **true papilledema** and **pseudopapilledema**—a benign variant where optic disc elevation mimics swelling due to optic disc drusen or congenital disc anomalies—is particularly challenging [40]. Misclassification may lead to unnecessary neuroimaging, delayed treatment, or overlooked life-threatening conditions. The subjectivity of manual assessment and inter-observer variability further motivate the need for automated, objective diagnostic tools.

Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated exceptional performance in medical image analysis tasks including diabetic

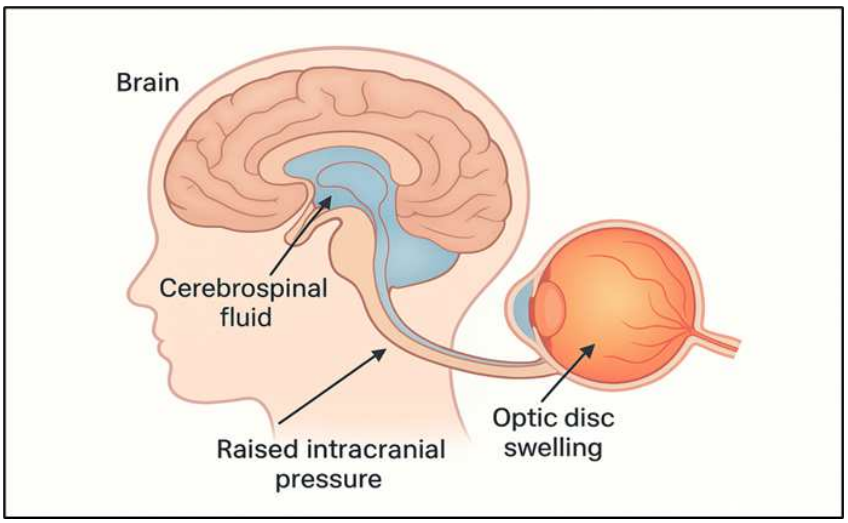


Fig. 1: Anatomical connection between the brain and eye showing how raised intracranial pressure (ICP) causes optic disc swelling (papilledema).

retinopathy, glaucoma, and age-related macular degeneration [11, 47]. Applied to fundus imaging, CNNs can identify subtle structural changes in the optic disc and retinal nerve fibre layer (RNFL) that may not be easily recognisable through conventional clinical observation.

Despite these advancements, existing studies on ICP-related fundus analysis are often limited to binary classification (normal vs. papilledema), overlooking the clinically significant category of pseudopapilledema [8, 32]. Additionally, most current systems lack interpretability—a crucial requirement for real-world clinical adoption, where healthcare professionals must understand the rationale behind AI decisions.

This study addresses these gaps by developing a deep learning-based automated system that classifies fundus images into three clinically relevant categories: **Normal**, **Pseudopapilledema**, and **Papilledema**. The system leverages the EfficientNet-B2 architecture with transfer learning, robust image preprocessing, and augmentation strategies. Explainable AI methods—Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-Agnostic Explanations (LIME)—are integrated to highlight image regions influencing model predictions, thereby enhancing transparency and building clinical trust. The system is deployed via a Gradio-based web interface supporting real-time inference and natural language explanations generated by a large language model. In this paper, we aim to overcome the limitations of existing intracranial pressure (ICP) monitoring methods by introducing a non-invasive and clinically meaningful deep learning approach using fundus images. Although invasive techniques like intraventricular catheters and fibre-optic transducers are considered the gold standard due to their accuracy, they come with serious risks and are not suitable for routine or outpatient use. Other non-invasive methods such as CT, MRI, and transcranial Doppler only provide indirect indications of ICP and are often

limited by high cost, limited availability, and dependence on expert interpretation. Fundus examination, on the other hand, offers a more accessible and physiologically relevant way to detect raised ICP, but distinguishing between papilledema and pseudopapilledema can be challenging and varies between clinicians. To address this, our work focuses on developing a reliable multi-class deep learning model that classifies fundus images into normal, papilledema, and pseudopapilledema, helping improve detection of raised ICP and supporting clearer, more consistent clinical decisions.

1.1 Objectives

The objectives of this study are formulated to address both the technical and clinical challenges associated with the non-invasive detection of raised intracranial pressure (ICP) using retinal fundus imaging. In particular, the study aims to develop an accurate, interpretable, and clinically meaningful deep learning framework capable of assisting in the screening and differentiation of optic disc abnormalities associated with papilledema.

1. To develop a robust deep learning-based classification framework for retinal fundus images that can distinguish between the three clinically relevant categories of *Normal*, *Papilledema*, and *Pseudopapilledema*. This objective is motivated by the need to support non-invasive screening for raised intracranial pressure (ICP), while also addressing the clinically important challenge of differentiating true papilledema from pseudopapilledema, which often presents with similar optic disc appearances but requires significantly different management pathways.
2. To fine-tune the **EfficientNet-B2** architecture using transfer learning on a medium-sized retinal fundus image dataset in order to achieve high diagnostic performance and stable generalisation. This objective focuses on leveraging the efficiency and representational strength of a modern convolutional neural network for multi-class retinal image classification, while ensuring that the trained model can effectively capture subtle structural variations in the optic disc and surrounding retinal features relevant to raised ICP assessment.
3. To improve the interpretability, transparency, and clinical usability of the proposed framework by integrating explainable artificial intelligence techniques, specifically **Grad-CAM** and **LIME**, and by performing comprehensive evaluation using standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. This objective is intended to ensure that the framework not only produces reliable predictions, but also provides visually meaningful explanations that can support clinician trust and facilitate its potential adoption as an assistive screening tool in ophthalmic and neurological practice.

2 Literature Survey

While numerous studies have applied machine learning and deep learning for ICP detection through fundus images, the use of explainable artificial intelligence (XAI) techniques remains limited, restricting clinical trust and adoption. Several deep learning-based studies have demonstrated strong performance—achieving accuracies above 90% in distinguishing papilledema from normal cases—yet they often function as "black boxes" [31, 32].

Qian et al. [6] introduced the Hypertensive Retinopathy Diagnosis Challenge (HRDC), providing a public benchmark dataset for hypertension and hypertensive retinopathy classification using retinal fundus images. The benchmark facilitates development and comparison of automated diagnostic algorithms. A parallel survey reviewed the progress of AI-assisted screening and prediction of systemic diseases using retinal images [7], highlighting how deep learning models can detect systemic conditions such as hypertension, cardiovascular diseases, and neurological disorders non-invasively. Invasive ICP detection methods remain the most accurate approaches for measuring intracranial pressure. The intraventricular catheter (ventriculostomy) is considered the gold standard, as it involves inserting a catheter into the brain ventricles to obtain direct and continuous pressure measurements. Despite its high accuracy, it carries significant risks such as infection, bleeding, and the need for hospitalisation and specialised care [23**?**]. Other invasive techniques, such as subdural or epidural sensors, involve placing pressure sensors beneath the skull to indirectly assess ICP levels. Although these are slightly less invasive than ventriculostomy, they still require surgical intervention and are associated with complications such as infection and calibration drift [24**?**]. Due to these risks and costs, invasive methods are generally limited to critical care settings and are not suitable for routine or outpatient monitoring.

Non-invasive imaging-based methods are commonly used as alternatives to estimate ICP indirectly. Techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) help identify structural changes in the brain, including ventricular enlargement, cerebral oedema, and midline shift, which may indicate elevated ICP. However, these methods do not provide direct pressure measurements and are limited by high cost, radiation exposure in the case of CT, and the need for advanced infrastructure and skilled personnel [24, 29]. Transcranial Doppler (TCD) ultrasound is another non-invasive approach that estimates ICP by measuring blood flow velocity in cerebral arteries, particularly the middle cerebral artery. While useful, its accuracy is influenced by operator skill and patient-specific anatomical variations [30**?**].

Fundus image-based methods have gained increasing attention due to their non-invasive nature and accessibility. Early approaches relied on traditional image processing techniques such as segmentation, edge detection, and measurement of optic disc features [27**?**]. These methods required manual tuning and were sensitive to variations in lighting and image quality. Classical machine learning approaches later introduced algorithms such as Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbours, which used manually extracted features like optic disc shape, colour, and texture [22**?**]. Although these methods improved performance, they lacked full automation and scalability. With the advancement of deep learning, particularly

Convolutional Neural Networks (CNNs) and transfer learning techniques, fundus-based ICP detection has significantly improved. Models such as CNNs, DenseNet, and U-Net can automatically learn complex features from raw images, enabling more accurate and scalable classification, including multi-class scenarios [11, 12]. Recent studies have reported accuracies exceeding 90% [48, 52]. However, many of these models function as black boxes, offering limited interpretability, which poses challenges for clinical adoption [8, 33].

In addition to fundus imaging, Optical Coherence Tomography (OCT) provides high-resolution cross-sectional images of retinal and optic nerve structures. Studies have shown strong correlations between retinal nerve fibre layer (RNFL) thickness measured using OCT and invasively measured ICP [4**?**]. RNFL thickness demonstrates a strong correlation with papilledema severity ($r \approx 0.78$) and a moderate correlation with ICP ($r \approx 0.55$), making OCT a valuable tool for monitoring disease progression over time. However, its effectiveness may be reduced in cases of severe oedema due to limited optical penetration.

Another non-invasive method is optic nerve sheath diameter (ONSD) measurement using ocular ultrasound. This technique estimates ICP by measuring the sheath diameter approximately 3 mm behind the globe. An increase in ONSD beyond 5 mm is considered indicative of elevated ICP (5 mmHg), with reported sensitivity and specificity exceeding 90% [30**?**]. Despite its effectiveness, the accuracy of ONSD measurement is highly dependent on operator expertise and probe positioning, and anatomical variations may affect the results.

2.1 Challenges and Limitations of Existing Methods

A number of critical challenges continue to hinder widespread clinical adoption across all modalities:

Invasive methods are associated with high infection risk, haemorrhage, and the necessity for surgical environments, restricting use to tertiary-care centres [23].

Imaging-based non-invasive methods suffer from high cost, limited accessibility, and the inability to provide continuous monitoring. They also provide only indirect estimates of ICP [29].

Fundus image-based methods face challenges of dataset size and diversity, subtle differences between pseudopapilledema and true papilledema, image quality variability, and limited clinical validation [42].

General AI challenges include explainability—most deep learning architectures operate as black boxes, limiting clinical trust and ethical acceptance. Integration into clinical workflows and population diversity in training data are additional concerns [43].

Table 1 presents a summary of existing ICP detection methods across modalities.

Table 1: Summary of existing ICP detection methods and deep learning-based fundus image analysis.

Sl.	Study / Year	Category	Method	Dataset	Key Findings / Performance	Limitations
1	Raboel et al., 2012	Invasive	ICP Monitoring	Clinical data	Gold-standard accuracy	Invasive procedures
2	de Moraes et al., 2023	Non-invasive	Multimodal Monitoring	Clinical data	Correlation-based ICP estimation	Multiple devices required
3	Frigieri et al., 2025	Non-invasive	ML + Extensometer	Cranial waveforms	ICP estimation model	Specialised equipment
4	Pose et al., 2025	Non-invasive	EEG-based Monitoring	EEG data	Explored EEG for ICP	Requires EEG setup
5	Karaliunas et al., 2025	Non-invasive	Retinal A/V Ratio	Retinal images	Retinal feature-based estimation	Small sample size
6	Dhar et al., 2021	Non-invasive	Tympanic Membrane	Clinical data	Alternative ICP estimation	Specialised equipment
7	Doe et al., 2020	Deep Learning	CNN Classification	500 fundus images	95% accuracy (binary)	No pseudopapilledema
8	Smith et al., 2021	Deep Learning	DenseNet + U-Net	600 fundus images	98% accuracy	Small dataset
9	Lee et al., 2022	Deep Learning	SVM + Handcrafted Features	300 fundus images	90% accuracy	Low scalability
10	Zhang et al., 2023	Deep Learning	CNN + Transfer Learning	800 fundus images	93% (3-class)	No interpretability
11	Kumar et al., 2024	Deep Learning	CNN + Grad-CAM	700 fundus images	Explainable AI attempt	Not clinically validated
12	Hemelings et al., 2021	Deep Learning	CNN	Fundus images	Glaucoma + papilledema detection	Not ICP-specific
13	Rivas-Villar et al., 2022	Deep Learning	Fundus Registration	Fundus images	Longitudinal monitoring	Requires alignment
14	Wen et al., 2025	Deep Learning	ICP Grading Framework	Ultrasound video	78.6% accuracy	Needs ultrasound data

Most existing methods achieve high classification accuracy but lack explainable AI techniques essential for building clinical trust. This study directly addresses this gap by integrating XAI methods such as Grad-CAM and LIME to generate interpretable visual explanations.

3 Methodology

3.1 Dataset Description

The dataset used in this study consists of **1369 retinal fundus images** sourced from publicly available medical imaging resources [53], categorised into three clinically significant classes:

- **Normal** – healthy optic disc with no signs of elevated ICP
- **Papilledema** – genuine optic disc swelling due to increased ICP
- **Pseudopapilledema** – optic disc elevation without true swelling

All images were resized to 224×224 pixels. The dataset was divided using a stratified split to preserve class distribution: 70% for training, 15% for validation, and 15% for testing. To address class imbalance, **class-weighted loss functions** were incorporated during model training.

Table 2 presents the key clinical parameters and their normal ranges compared to papilledema.

Table 2: Clinical parameters and their ranges across ICP conditions.

Parameter	**Normal Range**	**Change in Raised ICP**
Intracranial Pressure (ICP)	7–15 mmHg	>20 mmHg (Raised)
Optic Disc Diameter	∼1.5 mm	Increases in papilledema
RNFL Thickness	80–110 μm	Becomes thicker
Cup-to-Disc Ratio	0.3–0.5	Decreases
Visual Acuity	6/6	Reduced in severe stages

3.2 Data Preprocessing and Augmentation

To improve the generalisation capability of the model and reduce the risk of overfitting, a series of preprocessing and data augmentation techniques were applied to the input fundus images. As part of preprocessing, all images were resized to 224×224 pixels to ensure uniformity and compatibility with the model architecture. Pixel intensities were further normalised using the **ImageNet mean and standard deviation**, aligning the data distribution with pretrained model weights and facilitating effective transfer learning.

In addition to preprocessing, data augmentation techniques were employed during training to simulate real-world variations commonly observed in retinal imaging conditions. These included random horizontal flipping, random rotation, and random brightness adjustment. Such transformations help the model become more robust to variations in orientation, illumination, and image quality. Overall, these steps enhance the model's ability to generalise effectively to unseen clinical data.

3.3 System Architecture

The proposed system employs **EfficientNet-B2** as the backbone convolutional neural network for the classification of retinal fundus images. EfficientNet is designed using a **compound scaling** strategy that uniformly scales the depth, width, and input resolution of the network, thereby achieving a superior trade-off between accuracy and computational efficiency compared to traditional architectures. The network begins with an initial convolution layer followed by a sequence of **Mobile Inverted Bottleneck Convolution (MBConv)** blocks, which form the core building units responsible for hierarchical feature extraction. The original classification head of EfficientNet is replaced with a fully connected layer defined as:

$$\text{Linear}(320, 3) \tag{1}$$

to perform three-class classification corresponding to Normal, Papilledema, and Pseudopapilledema, followed by a Softmax activation function to generate class probabilities.

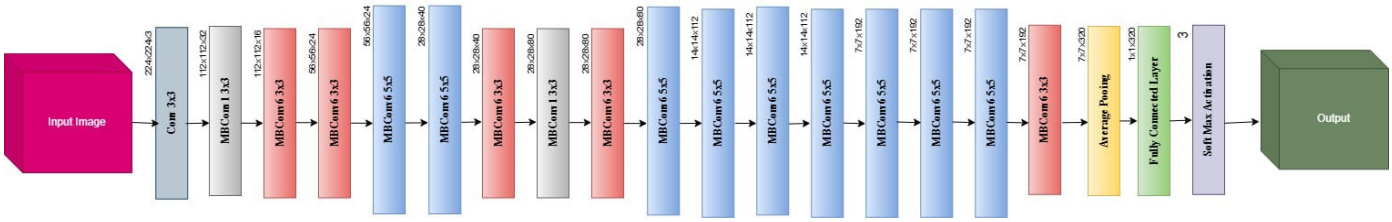


Fig. 2: EfficientNet B2 Model Architecture

Each MBConv block consists of multiple stages that efficiently transform feature representations. Initially, a **1×1 pointwise convolution** expands the number of channels by an expansion ratio t, followed by **Batch Normalization** and **Swish activation**, enabling non-linear feature learning. This is followed by a **depthwise convolution (DWConv)** with kernel size $k \times k$, which performs spatial filtering independently on each channel, significantly reducing computational complexity compared to standard convolutions. A **Squeeze-and-Excitation (SE)** module is then applied, which includes global pooling and fully connected layers to learn channel-wise attention weights, followed by a sigmoid activation and element-wise multiplication to recalibrate important features. Subsequently, a **projection layer** using a 1×1 convolution reduces the channel dimensions back to the desired output size. When the input and output dimensions are identical, a **skip (residual) connection** is added, improving gradient flow and enabling deeper network training.

As illustrated in Fig. 2, these MBConv blocks are stacked with varying kernel sizes and channel depths, progressively transforming the input image into high-level feature representations. The final feature map obtained has dimensions of $\mathbf{7 \times 7 \times 320}$, which is passed through a **Global Average Pooling** layer to produce a compact

320-dimensional feature vector. This vector is then fed into the fully connected layer for classification.

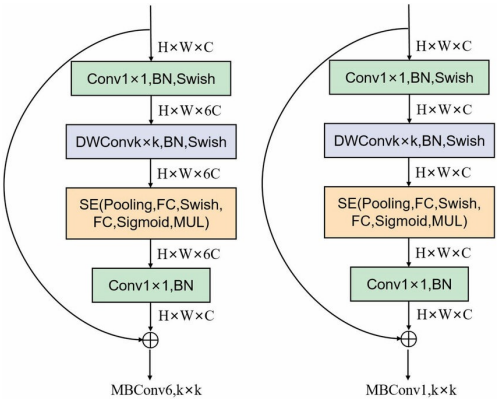


Fig. 3: Comparison of MBConv block variants with different expansion ratios, highlighting the expansion layer, depthwise convolution, squeeze-and-excitation (SE) mechanism, and projection layer.

Fig. 3 further illustrates the structural comparison between two variants of MBConv blocks. In the first configuration, corresponding to **MBConv6**, the expansion phase increases the channel dimension from C to $6C$, allowing richer feature extraction at the cost of higher computation. In contrast, the second configuration, corresponding to a lower expansion variant (e.g., MBConv1/MBConv3), maintains or minimally expands the channel dimension, resulting in reduced computational complexity. Both variants include depthwise convolution, SE attention mechanism, and projection layers, but differ primarily in their expansion ratios and representational capacity. EfficientNet strategically utilises these different MBConv configurations across layers to balance efficiency and performance, making it highly effective for medical image classification tasks.

3.4 Training Configuration

The model was trained using the Adam optimizer with a learning rate of 1×10^{-4}, which provided stable and efficient convergence during training. Cross-entropy loss with class weights was employed to address class imbalance and improve performance across all categories. A batch size of 32 was used, and the model was trained for 10 epochs. To prevent overfitting and ensure better generalisation, early stopping was applied based on validation loss monitoring, allowing the training process to halt once performance stopped improving.

3.5 Explainable AI Integration

To enhance interpretability and improve clinical trust, two explainable AI techniques, Grad-CAM and LIME, were integrated into the model. Gradient-weighted Class

Activation Mapping (Grad-CAM) [17] was used to identify and visualise the important regions in fundus images that influenced the model's predictions. The resulting heatmaps were overlaid on the original images to highlight relevant pathological areas, making the decision-making process more intuitive for clinicians. In addition, Local Interpretable Model-Agnostic Explanations (LIME) were employed to provide local feature-level interpretations for individual predictions. By explaining how specific regions of an image contribute to a classification outcome, LIME improves transparency and helps validate model behaviour. Together, these techniques bridge the gap between deep learning predictions and clinical understanding, enabling more reliable and interpretable decision-making.

3.6 Algorithm

The complete processing pipeline is described in Algorithm 1.

Algorithm 1 Deep Learning Pipeline for ICP Detection from Fundus Images

Require: Retinal fundus image dataset with three class labels
Ensure: Classification result, Grad-CAM heatmap, LIME explanation, LLM explanation

1: Extract and organise the dataset into three classes (Normal, Papilledema, Pseudopapilledema)
2: Split dataset: 70% Training, 15% Validation, 15% Testing (stratified)
3: Resize all images to 224×224; apply augmentation (flip, rotation, brightness) on training data
4: Load pretrained EfficientNet-B2; replace final layer with *Linear(in_features, 3)*
5: Compute class weights; define weighted Cross-Entropy loss function
6: Train model using Adam optimiser ($\text{lr} = 1 \times 10^{-4}$) for 10 epochs with early stopping
7: Record training loss, validation loss, and validation accuracy per epoch
8: Evaluate model on test set: compute Accuracy, Precision, Recall, F1-score, Confusion Matrix
9: Apply Grad-CAM on the final convolutional layer to generate attention heatmaps
10: Apply LIME to identify important superpixel regions influencing predictions
11: Load 4-bit quantised Falcon-7B-Instruct model for explanation generation
12: Generate concise medical explanation using prediction, confidence score, and XAI outputs
13: Deploy via Gradio interface for real-time prediction and explanation

4 Implementation

The proposed framework for detecting raised intracranial pressure (ICP) from retinal fundus images was implemented using **Python 3.10** with the **PyTorch** deep learning library. The **torchvision** module was used for loading pretrained models and applying image transformations. All experiments, including dataset preparation, training, evaluation, and explainability analysis, were performed in a GPU-enabled Google Colab environment, allowing efficient handling of computationally intensive operations.

The dataset consisted of **1,369 retinal fundus images** belonging to three clinically relevant classes: *Normal*, *Papilledema*, and *Pseudopapilledema*. To ensure fair evaluation and proper generalisation, the dataset was divided into training, validation, and test sets using a **70%/15%/15% split** while maintaining class balance across all subsets. Each image was resized to a fixed resolution of **224 × 224 pixels**, matching the input requirements of EfficientNet-B2. Basic preprocessing was applied, and data augmentation techniques such as **random horizontal flipping** and **small rotations** were introduced to simulate real-world variations in fundus image acquisition, such as changes in orientation and slight distortions.

For model development, **EfficientNet-B2** pretrained on ImageNet was used as the backbone network. Transfer learning was applied by replacing the original classification head with a new fully connected layer to perform three-class classification. This approach allows the model to leverage previously learned visual features while adapting to the specific retinal classification task. To address the imbalance in class distribution, **class-weighted cross-entropy loss** was used, assigning higher importance to underrepresented classes such as Papilledema and Pseudopapilledema. The model was trained using the **Adam optimiser** with a learning rate of 1×10^{-4}, chosen for its ability to provide stable and efficient convergence. Training was conducted for **10 epochs**, with model weights saved based on the best validation performance.

To enhance transparency and interpretability, two explainable AI (XAI) techniques were incorporated. **Grad-CAM** was applied to the final convolutional layers of EfficientNet-B2 to generate class-specific activation maps, highlighting the regions of the fundus image that most influenced the prediction. These heatmaps typically focused on the optic disc region, which is clinically relevant for detecting papilledema. In addition, **LIME** was used to provide local explanations by segmenting the image into superpixels and identifying which regions contributed positively to the model's decision. The combination of Grad-CAM and LIME provides both global and local interpretability, offering a more comprehensive understanding of model behaviour.

To further improve usability in a clinical context, the outputs from the classification model and XAI techniques were integrated with a **Falcon-7B large language model (LLM)**. The predicted class, confidence score, and visual explanation outputs (Grad-CAM and LIME) were used as inputs to the LLM. Based on these inputs, the LLM generates a concise, human-readable explanation describing the prediction in clinically meaningful terms, such as optic disc swelling, blurred disc margins, and vascular changes. This step transforms raw model outputs into interpretable insights that can assist medical professionals and improve trust in the system.

The overall workflow of the proposed system is illustrated in Figure 4. The pipeline begins with fundus image acquisition, followed by preprocessing steps including resizing, augmentation, and normalisation. The processed images are then passed through EfficientNet-B2 for feature extraction and classification, producing both predicted labels and confidence scores. Subsequently, Grad-CAM and LIME generate visual explanations highlighting important regions in the image. Finally, these outputs are combined and processed by the Falcon-7B LLM to generate a clinically interpretable explanation, resulting in the final output of the system.

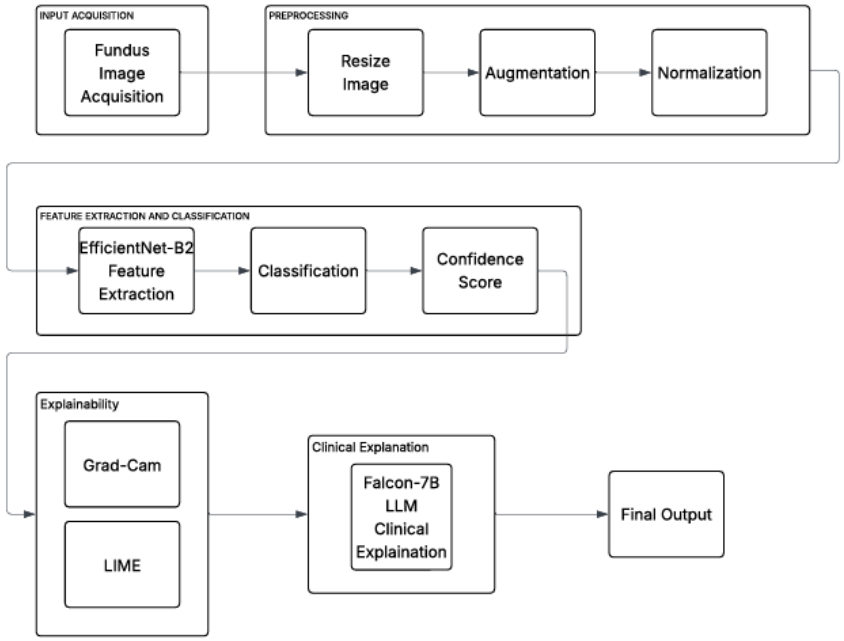
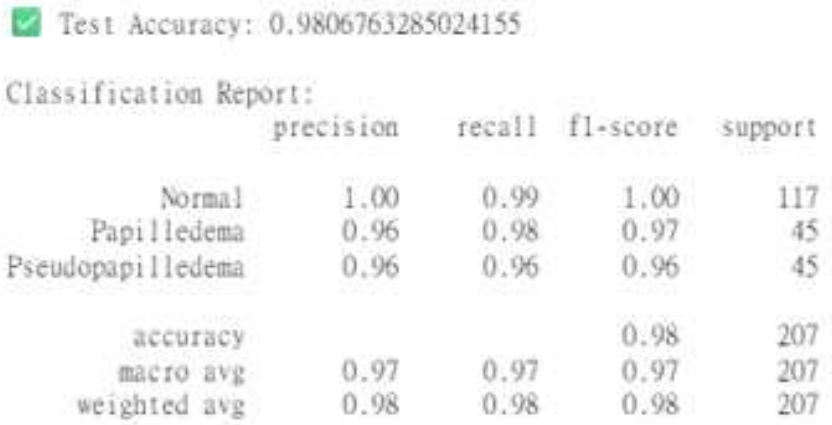


Fig. 4: Detailed pipeline of the proposed ICP detection system. The framework includes fundus image acquisition, preprocessing (resizing, augmentation, and normalisation), feature extraction and classification using EfficientNet-B2, explainability through Grad-CAM and LIME, and final clinical interpretation generated by a Falcon-7B large language model.

5 Results

A comprehensive comparative analysis was performed among three deep learning architectures, namely EfficientNet-B2, ResNet-50, and DenseNet121, for the classification of fundus images into Normal, Papilledema, and Pseudopapilledema classes. The performance of these models was evaluated using training accuracy, validation accuracy, test accuracy, precision, recall, and F1-score, as summarized in Table 3.

EfficientNet-B2 demonstrated the best overall performance among all models. It achieved a peak validation accuracy of 99.02% and consistently maintained high training accuracy, reaching up to 99.27% by the final epoch. The model exhibited stable convergence with minimal fluctuations in validation loss, indicating strong generalization capability. Furthermore, EfficientNet-B2 showed superior performance in minimizing misclassifications, particularly between Normal and Pseudopapilledema classes, which are often challenging to distinguish.



Test Accuracy: 0.9806763285024155

Classification Report:

	precision	recall	f1-score	support
Normal	1.00	0.99	1.00	117
Papilledema	0.96	0.98	0.97	45
Pseudopapilledema	0.96	0.96	0.96	45
accuracy			0.98	207
macro avg	0.97	0.97	0.97	207
weighted avg	0.98	0.98	0.98	207

Fig. 5: EfficientNet-B2

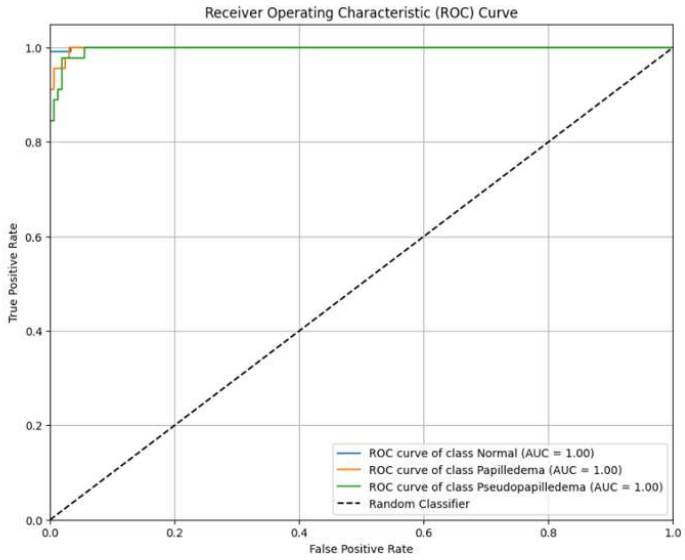


Fig. 6: ROC Curve

The ROC curves in figure 6 show that the model performs extremely well in distinguishing between the three classes. All curves lie very close to the top-left corner, which indicates high true positive rates with very low false positives. The AUC value for Normal, Papilledema, and Pseudopapilledema is nearly 1.00, suggesting almost perfect classification. This means the model can reliably identify both healthy and abnormal cases. The clear separation from the random classifier line further highlights its strong performance. Overall, the results indicate that the model is accurate and dependable for clinical use.

The EfficientNet-B2 model was trained for 10 epochs using transfer learning. The training process demonstrated stable convergence, with both training and validation loss consistently decreasing over epochs. Initially, the model achieved a training accuracy of 74.29% and validation accuracy of 81.95%, indicating effective early learning. As training progressed, the model showed significant improvement, ultimately reaching a peak validation accuracy of **99.51%**, reflecting strong generalisation performance. The final training accuracy reached 99.27% with a low training loss of 0.0220, demonstrating effective feature extraction and learning capability. Although minor fluctuations in validation loss were observed, the overall trend remained stable, suggesting the absence of significant overfitting. Figures 7 and 8 illustrate the training and validation accuracy and loss curves, respectively.

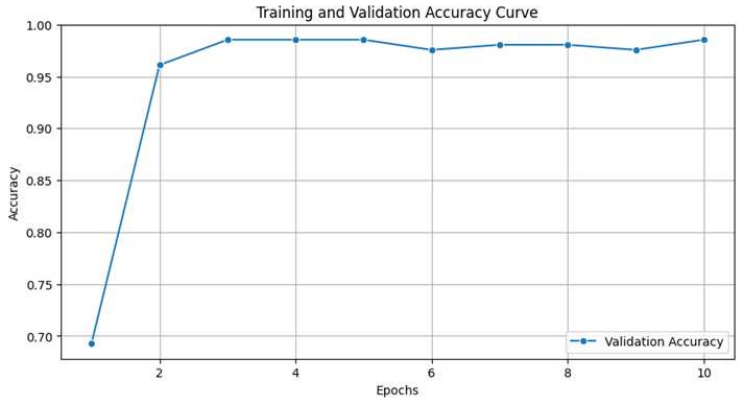


Fig. 7: Training and validation accuracy curve over 10 epochs.

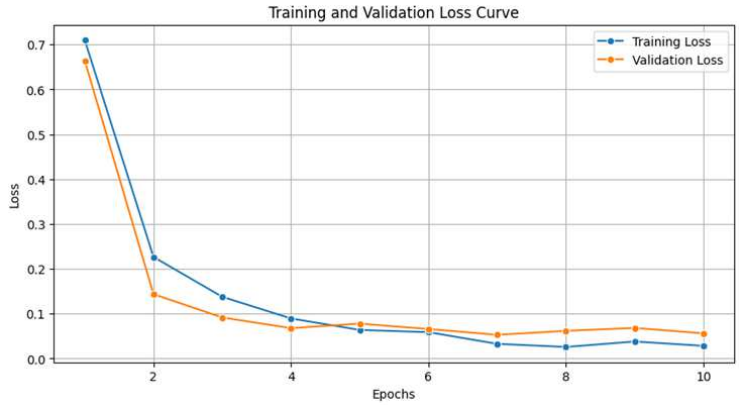


Fig. 8: Training and validation loss curve over 10 epochs.

Table 3 presents the detailed per-class classification metrics

Table 3: Per-class classification metrics for EfficientNet-B2 on the test set.

Class	**Precision**	**Recall**	**F1-Score**
Normal	1.00	0.97	0.99
Papilledema	1.00	1.00	1.00
Pseudopapilledema	0.94	1.00	0.97

Figure 9 presents the confusion matrix for the test set. The model correctly classified 114 Normal images, 45 Papilledema images, and 45 Pseudopapilledema images. Only 3 Normal cases were misclassified, demonstrating strong discriminatory capability.

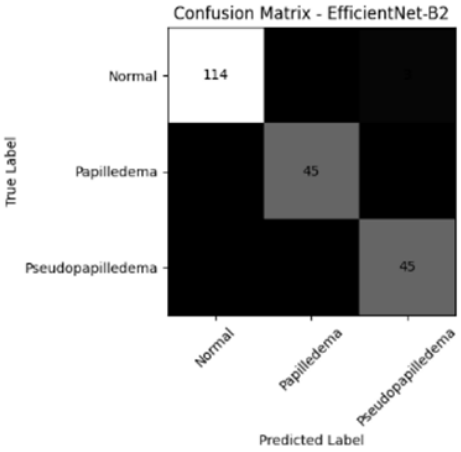


Fig. 9: Confusion matrix on the independent test set (207 images).

5.1 Model Comparison

A comparative evaluation was conducted among EfficientNet-B2, ResNet-50, and DenseNet121 to assess their performance on retinal fundus image classification. Table 4 summarises the overall results.

During training, all models showed a progressive improvement in performance over the epochs, with decreasing loss and increasing accuracy, indicating effective learning. In the initial epochs, the models rapidly captured basic visual features such as edges and shapes, leading to a sharp rise in validation accuracy. As training progressed, finer features related to retinal structures were learned, resulting in more stable and higher accuracy values. EfficientNet-B2 demonstrated smoother convergence with minimal fluctuations in validation loss, suggesting better generalisation. In contrast, ResNet-50 and DenseNet121 showed slight instability in later epochs, with noticeable variations in validation accuracy and loss.

ResNet-50 achieved a test accuracy of 97.10% and demonstrated strong performance, particularly with perfect recall (100%) for both Papilledema and Pseudopapilledema classes. However, the precision for Pseudopapilledema was slightly lower (88%), indicating the presence of some false positive predictions.

DenseNet121 also showed competitive performance, reaching a validation accuracy of up to 99.02% during training. The model exhibited rapid learning in the initial epochs, with accuracy improving significantly from 85.85% to 99.02%. However, minor fluctuations in validation accuracy and loss were observed in later stages, suggesting slightly less stable convergence compared to EfficientNet-B2.

In contrast, EfficientNet-B2 achieved the best overall performance, with a validation accuracy of 99.51% and test accuracy of 98.55%. The superior performance of EfficientNet-B2 can be attributed to its compound scaling approach, which balances network depth, width, and resolution, enabling more efficient feature extraction compared to the fixed scaling strategies used in ResNet-50 and DenseNet121. This allows the model to capture subtle differences between classes more effectively, leading to improved classification performance.

Overall, EfficientNet-B2 outperformed both ResNet-50 and DenseNet121, making it the most suitable architecture for detecting raised intracranial pressure using fundus images.

Table 4: Comparison of Model Performance

Model	Validation Accuracy (%)	Test Accuracy (%)	Remarks
EfficientNet-B2	99.51	98.55	Best overall performance
ResNet-50	98.54	97.10	High recall, slight precision drop
DenseNet121	99.02	97.12	Minor fluctuations observed

5.2 Explainable AI Results

To improve the interpretability of the proposed model and ensure its reliability in clinical settings, two explainable AI techniques, Grad-CAM and LIME, were employed to analyse the model's decision-making process.

Grad-CAM visualisations were used to identify the regions of the fundus images that contributed most to the model's predictions. The heatmaps consistently highlighted the optic disc region in Papilledema cases, which is clinically significant as papilledema is characterised by swelling of the optic nerve head. In Normal images, the activation regions were more uniformly distributed, reflecting the absence of abnormal patterns. For Pseudopapilledema, the model focused on regions near the optic disc, capturing subtle structural variations that differentiate it from true papilledema. Figures 10, 11, and 12 illustrate these visualisations. The strong alignment between highlighted regions and clinically relevant anatomical structures indicates that the model is learning meaningful and medically valid features rather than relying on spurious patterns.

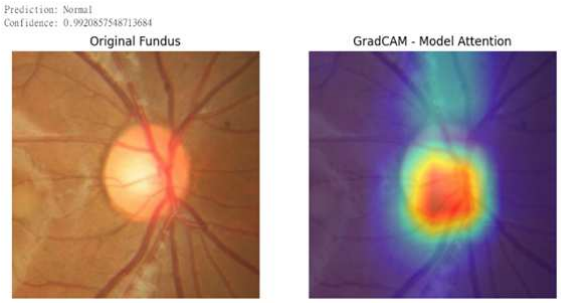


Fig. 10: Grad-CAM visualisation for Normal fundus image.

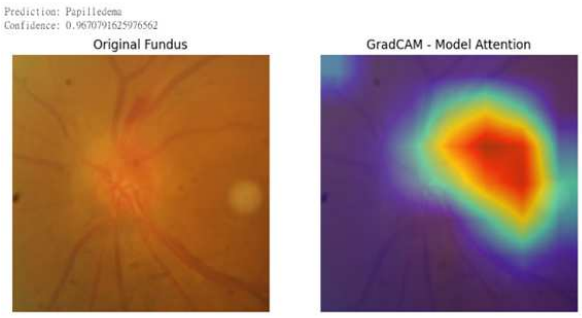


Fig. 11: Grad-CAM visualisation for Papilledema fundus image.

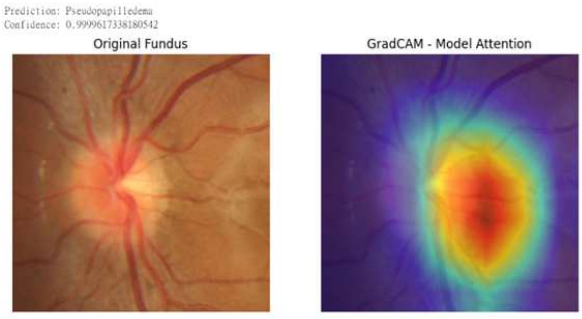


Fig. 12: Grad-CAM visualisation for Pseudopapilledema fundus image.

In addition to global visual explanations, LIME was used to provide local interpretability by highlighting specific superpixel regions that influenced individual predictions. As shown in Figure 13, LIME identifies the most relevant areas within an image that contributed to the classification outcome. These regions often correspond to clinically important features, further validating the model's reasoning process.

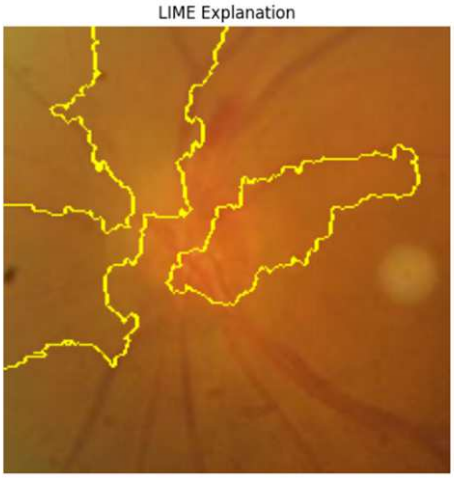


Fig. 13: LIME explanation highlighting locally important superpixel regions for a sample prediction.

Overall, the integration of Grad-CAM and LIME enhances model transparency by providing both global and local explanations. These visual explanations are further processed using the Falcon-7B large language model to generate human-interpretable textual descriptions of the model's predictions. This helps translate complex visual patterns into meaningful clinical insights that are easier for users to understand. As a result, the system not only improves trust in the model's predictions but also supports its use as a decision-support tool in clinical practice, where interpretability and clarity are essential for adoption.

6 Discussion

The proposed approach was designed with careful consideration of both model performance and the challenges associated with medical image classification. EfficientNet-B2 was selected as the primary architecture due to its compound scaling strategy, which simultaneously balances network depth, width, and input resolution. This makes it particularly effective for capturing fine-grained retinal features such as optic disc boundaries and subtle swelling patterns, which are critical for distinguishing between Papilledema and Pseudopapilledema. Compared to traditional architectures like ResNet-50 and DenseNet121, EfficientNet-B2 demonstrated more stable convergence and superior generalisation.

Transfer learning was employed by initializing the model with pretrained weights, allowing the network to leverage previously learned low-level features such as edges, textures, and shapes. This significantly reduced training time and improved performance, especially given the moderate dataset size of 1369 images. The final classification layer was modified to adapt the model to the three-class problem.

To address class imbalance in the dataset, a weighted cross-entropy loss function was used. The computed class weights ensured that the model assigned higher importance to Papilledema and Pseudopapilledema cases, thereby reducing bias toward the majority class and improving recall for clinically critical conditions.

The Adam optimizer was chosen due to its adaptive learning rate capabilities, which enable efficient and stable training. With a low learning rate of 1×10^{-4}, the model achieved smooth gradient updates and avoided overshooting the optimal solution. This contributed to the steady decrease in both training and validation loss observed during training.

Data augmentation techniques, including random horizontal flipping and rotation, were applied to improve robustness and reduce overfitting by simulating real-world variations in fundus images.

The training process showed consistent improvement across epochs, with increasing accuracy and decreasing loss. EfficientNet-B2 exhibited smoother validation curves compared to ResNet-50 and DenseNet121, suggesting better generalisation, while the other models showed minor fluctuations in later epochs.

Explainability techniques were incorporated to enhance model interpretability. Grad-CAM was used to visualise global regions of importance, consistently highlighting the optic disc area in pathological cases, which aligns with clinical understanding. LIME provided fine-grained, local explanations by identifying superpixel regions that contributed most to individual predictions. Furthermore, these visual explanations were integrated with a Falcon-7B large language model to generate human-readable medical interpretations, bridging the gap between deep learning outputs and clinical understanding.

Overall, the combination of EfficientNet-B2, weighted loss, adaptive optimization, and explainable AI techniques resulted in a robust and interpretable system with strong clinical relevance.

7 Conclusion

This study demonstrates the effectiveness of deep learning-based analysis of retinal fundus images for the non-invasive detection of elevated intracranial pressure (ICP). By framing the problem as a multi-class classification task involving Papilledema, Pseudopapilledema, and Normal cases, the proposed framework is able to differentiate true optic disc swelling from visually similar conditions. The EfficientNet-B2 architecture achieved a high validation accuracy of 99.51% and a test accuracy of 98.55%, indicating strong generalisation capability and robustness in handling unseen data.

The results highlight the ability of convolutional neural networks to extract both low-level visual features and high-level structural patterns from retinal images. In particular, the model effectively captures subtle variations in optic disc morphology, vessel distribution, and swelling characteristics, which are critical for distinguishing between Papilledema and Pseudopapilledema. Compared to traditional diagnostic approaches, which often rely on subjective clinical interpretation or invasive procedures, the proposed system offers a faster, safer, and more accessible alternative for early screening and diagnosis.

Furthermore, the integration of explainable AI techniques such as Grad-CAM and LIME enhances the interpretability of the model by providing visual and region-based explanations of predictions. These methods demonstrate that the model focuses on clinically relevant regions, particularly the optic disc area, thereby increasing trust in its decision-making process. This interpretability is crucial in medical applications, as it bridges the gap between automated predictions and clinical reasoning, supporting the adoption of AI-based systems in practice.

Despite the promising results, certain limitations remain. The dataset used in this study, although carefully curated, is moderate in size and may not fully capture the variability present in diverse populations and imaging conditions. Additionally, external validation on independent and multi-centre datasets is necessary to further establish the reliability and generalisability of the model in real-world clinical settings.

Future work will focus on expanding the dataset to include more diverse patient demographics and imaging sources, improving model performance for borderline and ambiguous cases, and exploring more lightweight and efficient architectures for practical deployment. Further validation through collaboration with clinical experts and comparison with ophthalmologist assessments will also be essential to ensure the safe and effective integration of the system into healthcare workflows.

Declarations

Funding
The authors received no specific funding for this work.

Conflict of Interest/Competing Interests
The authors declare that they have no competing interests.

Ethics Approval and Consent to Participate
Not applicable.

Consent for Publication
Not applicable.

Data Availability
The dataset analysed during the current study is publicly available at:
https://drive.google.com/drive/folders/14_NlMraPqHXGZWrMU0-mm8dFbYoq7wba?usp=drive_link

Materials Availability
Not applicable.

Code Availability
The code used to support the findings of this study is available at:
https://github.com/Kirthanaaa9/ICP

Author Contributions
All authors contributed equally to the study conception and design. Material preparation, data collection, model development, analysis, and manuscript writing were performed collaboratively by all authors. All authors read and approved the final manuscript.

References

[1] B. Qian et al., "HRDC challenge: a public benchmark for hypertension and hypertensive retinopathy classification from fundus images," *The Visual Computer*, vol. 41, no. 2, pp. 1061–1077, 2025.

[2] "Research progress on AI-assisted screening and prediction of systemic diseases based on retinal images," *The Visual Computer*, pp. 1–29, 2025.

[3] M. S. Andersen et al., "A new novel method for assessing intracranial pressure using non-invasive fundus images: a pilot study," *Scientific Reports*, vol. 10, no. 1, pp. 1–7, 2020. [nature.com](https://www.nature.com/articles/s41598-020-70084-0)

[4] V. Vijay et al., "Using optical coherence tomography as a surrogate of measurements of intracranial pressure in idiopathic intracranial hypertension," *JAMA Ophthalmology*, vol. 138, no. 12, pp. 1264–1271, 2020. [jamanetwork.com](https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2787719)

[5] Y. D. Kim et al., "Effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images," *Scientific Reports*, vol. 10, no. 1, p. 4623, 2020. [nature.com](https://www.nature.com/articles/s41598-020-61519-9)

[6] B. Qian, X. Wang, Z. Guan et al., "HRDC challenge: a public benchmark for hypertension and hypertensive retinopathy classification from fundus images," *The Visual Computer*, vol. 41, no. 2, pp. 1061–1077, 2025.

[7] "Research progress on AI-assisted screening and prediction of systemic diseases based on retinal images," *The Visual Computer*, 2025.

[8] K. Avramidis et al., "Automating detection of papilledema in pediatric fundus images with explainable machine learning," *arXiv preprint arXiv:2207.04565*, 2022. [arxiv.org](https://arxiv.org/abs/2207.04565)

[9] A. Gungor et al., "Deep learning system outperforms clinicians in identifying optic disc abnormalities," *JAMA Ophthalmology*, vol. 142, no. 11, pp. 1073–1079, 2024.

[10] C. Vasseneix et al., "Deep learning to discriminate arteritic from nonarteritic ischemic optic neuropathy on color images," *JAMA Ophthalmology*, vol. 142, no. 11, pp. 1073–1079, 2023. [jamanetwork.com](https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2800294)

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. [arxiv.org](https://arxiv.org/abs/1512.03385)

[12] G. Huang et al., “Densely connected convolutional networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. [arxiv.org](https://arxiv.org/abs/1608.06993)

[13] C. Szegedy et al., “Rethinking the inception architecture for computer vision,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826. [arxiv.org](https://arxiv.org/abs/1512.00567)

[14] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014. [arxiv.org](https://arxiv.org/abs/1409.1556)

[15] A. G. Howard et al., “MobileNets: efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017. [arxiv.org](https://arxiv.org/abs/1704.04861)

[16] N. V. Chawla et al., “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. [jair.org](https://www.jair.org/index.php/jair/article/view/10302)

[17] R. R. Selvaraju et al., “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626. [arxiv.org](https://arxiv.org/abs/1610.02391)

[18] J. Chang et al., “Association of cardiovascular mortality and deep learning-derived retinal atherosclerosis score,” *JAMA Network Open*, vol. 3, no. 10, e2022187, 2020.

[19] L. Y. Li et al., “Prediction of cardiovascular markers and diseases using retinal fundus images and deep learning: a systematic scoping review,” *European Heart Journal – Digital Health*, vol. 5, no. 6, pp. 660–669, 2024.

[20] D. Maji et al., “Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images,” *arXiv preprint arXiv:1603.04833*, 2016.

[21] R. Poplin et al., “Prediction of cardiovascular risk factors from retinal fundus photographs using deep learning,” *Nature Biomedical Engineering*, vol. 2, no. 3, pp. 158–164, 2018. [nature.com](https://www.nature.com/articles/s41551-018-0195-0)

[22] Y. Zhang et al., “Fundus image classification and retinal disease localization using CNNs,” *IEEE Access*, vol. 8, pp. 162270–162280, 2020. [ieeexplore.ieee.org](https://ieeexplore.ieee.org/document/9131011)

[23] P. H. Raboel et al., “Intracranial pressure monitoring: review and update,” *Acta Neurochirurgica*, vol. 154, pp. 5–14, 2012.

[24] C. G. de Moraes et al., "Multimodal ICP monitoring: correlation between non-invasive and invasive techniques," *Neurocritical Care*, 2023.

[25] N. Frigieri et al., "Machine learning for ICP estimation using cranial waveform analysis," *IEEE Transactions on Biomedical Engineering*, 2025.

[26] R. Karaliunas et al., "Non-invasive ICP estimation using retinal artery/vein ratio," *Frontiers in Neurology*, 2025.

[27] R. Hemelings et al., "Deep learning detection of glaucoma and papilledema," *Ophthalmology AI*, 2021.

[28] A. Rivas-Villar et al., "Fundus image registration for longitudinal monitoring," *Medical Image Analysis*, 2022.

[29] X. Wen et al., "Fundus ultrasound video-based ICP grading using CNNs," *Computers in Biology and Medicine*, 2025.

[30] S. Dhar et al., "Tympanic membrane displacement as a non-invasive ICP marker," *Journal of Clinical Monitoring and Computing*, 2021.

[31] D. Milea et al., "Artificial intelligence to detect papilledema from ocular fundus photographs," *New England Journal of Medicine*, vol. 382, no. 18, pp. 1687–1695, 2020. [nejm.org](https://www.nejm.org/doi/full/10.1056/NEJMoa1917130)

[32] C. Vasseneix et al., "Accuracy of a deep learning system for classification of papilledema severity on ocular fundus photographs," *Neurology*, vol. 97, no. 22, pp. e2142–e2153, 2021. [n.neurology.org](https://n.neurology.org/content/97/22/e2142)

[33] T. Saba et al., "Automatic detection of papilledema through fundus retinal images using deep learning," *Microscopy Research and Technique*, vol. 84, no. 8, pp. 1852–1860, 2021. [doi.org](https://doi.org/10.1002/jemt.23796)

[34] V. Biousse et al., "Application of a deep learning system to detect papilledema on nonmydriatic ocular fundus photographs in an emergency department," *American Journal of Ophthalmology*, 2024.

[35] J. Branco et al., "Classifying and quantifying changes in papilloedema using a convolutional neural network," *Scientific Reports*, 2024.

[36] L. Rambabu et al., "Detecting papilloedema as a marker of raised intracranial pressure using deep learning on fundus photographs," *PLOS Digital Health*, 2025.

[37] L. Anandi et al., "The use of artificial intelligence in detecting papilledema from fundus photographs," *Taiwan Journal of Ophthalmology*, vol. 13, no. 2, pp. 78–84, 2023.

[38] D. Milea et al., "Deep learning system to identify and grade papilledema: multicenter validation," *Scientific Reports*, 2022.

[39] K. Sathianvichitr et al., "A deep learning approach readily differentiates papilledema severity," *medRxiv preprint*, 2023. [medrxiv.org](https://www.medrxiv.org)

[40] M. Y. Chang et al., "Artificial intelligence to differentiate pediatric papilledema and pseudopapilledema," *Ophthalmology Science*, vol. 3, no. 4, 2024. [doi.org](https://doi.org/10.1016/j.xops.2024.100315)

[41] B. Li et al., "The performance of a deep learning system in assisting junior ophthalmologists in detecting major fundus diseases," *NPJ Digital Medicine*, vol. 7, no. 1, p. 123, 2024. [nature.com](https://www.nature.com/articles/s41746-024-00933-5)

[42] A. M. Salaheldin et al., "Evaluation of AI-based methods for papilledema: a systematic assessment," *Computers in Biology and Medicine*, vol. 168, p. 107680, 2024. [doi.org](https://doi.org/10.1016/j.compbiomed.2024.107680)

[43] A. Grzybowski et al., "Retina fundus photograph-based artificial intelligence: a systematic review," *Ophthalmology Retina*, vol. 8, no. 3, pp. 221–236, 2024.

[44] E. F. Linton et al., "Visualization of optic nerve structural patterns in papilledema using deep learning," *Translational Vision Science & Technology*, vol. 13, no. 2, 2024.

[45] D. Szanto et al., "Deep learning differentiates papilledema, NAION, and other optic disc elevations," *Clinical Ophthalmology*, 2025.

[46] G. Quellec et al., "Automatic detection of rare pathologies in fundus images with few-shot methods," *arXiv preprint arXiv:1907.07895*, 2019. [arxiv.org](https://arxiv.org/abs/1907.07895)

[47] V. Gulshan 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.

[48] Z. Zhang et al., "Fundus image classification for papilledema and pseudopapilledema using CNNs," *IEEE Access*, vol. 8, pp. 162270–162280, 2020.

[49] R. Whitwell et al., "Differentiating optic disc edema and pseudopapilledema using deep learning on near-infrared and fundus imaging," *Ophthalmology Science*, 2024.

[50] M. Rahman et al., "AI-assisted triage for intracranial pathology via fundus photos: feasibility and clinical trial results," *Frontiers in Digital Health*, 2025.

[51] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114. [arxiv.org](https://arxiv.org/abs/1905.11946)

[52] Kumar et al., "CNN with Grad-CAM visualization for papilledema detection," 2024.

[53] OpenMedLab, "Awesome Medical Dataset," GitHub repository, 2024. [github.com](https://github.com/openmedlab/Awesome-Medical-Dataset)