



VIGNAN'S

Foundation for Science, Technology & Research

(Deemed to be University)

-Estd. u/s 3 of UGC Act 1956

DEPARTMENT OF ADVANCED COMPUTER SCIENCE & ENGINEERING

**VIGNAN'S FOUNDATION FOR SCIENCE , TECHNOLOGY AND
RESEARCH**

(Deemed to be University)

Vadlamudi, Guntur -522213, INDIA.

Title of the project

Automating Diabetic Retinopathy Detection for Enhanced patient Care

Project Report

Submitted

In partial fulfillment of the requirements for the award of the degree

BACHELOR OF TECHNOLOGY

In

Computer Science and Engineering – Artificial Intelligence and Machine Learning

By

Surya Vipparla (211FA18099)

Shaik Kabsha Ansariya (211FA18100)

Under the Guidance of

Dr. Jyostna Bodapati

Associate Professor



VIGNAN'S

Foundation for Science, Technology & Research

(Deemed to be University)

-Estd. u/s 3 of UGC Act 1956

**DEPARTMENT OF ADVANCE COMPUTER SCIENCE &
ENGINEERING**

CERTIFICATE

This is to certify that the report entitled “ **Automating Diabetic Retinopathy Detection for
Enhanced patient Care**”

is submitted by “**Surya Vipparla – 211FA18099, Sk. Kabsha Ansariya – 211FA18100**”
in the partial fulfilment of course work of Intelligence Application Development, carried
out in the department of ACSE, VFSTR Deemed to be University.

Mrs Dr. Jyostna Bodapati,
Associate Professor,
VFSTR.

DECLARATION

I hereby declare that the project entitled “**Automating Diabetic Retinopathy Detection for Enhanced patient Care**” submitted for the “**DEPARTMENT OF ADVANCED COMPUTER SCIENCE AND ENGINEERING**”. This dissertation is our original work and the project has not formed the basis for the award of any degree, associate-ship and fellowship or any other similar titles and no part of it has been published or sent for publication at the time of submission.

By
Surya Vipparla -211FA18099
Sk. Kabsha Ansariya – 211FA18100

Date: 01.06.2024

ACKNOWLEDGEMENT

This project program is a golden opportunity for learning and self-development. Consider ourselves very lucky and honored to have so many wonderful people lead us through in the completion of this project.

I express our gratitude to the Project Guide **Dr. Jyostna Devi**, for permitting us to undertake the Program and for the help and cooperation throughout the course of our project Program.

I express our gratitude to the Project Coordinator **K. Praneeth**, for permitting us to undertake the Program and for the help and cooperation throughout the course of our project Program.

It is a great pleasure for us to express our sincere thanks to **Dr. Jyostna Devi**, Associate Professor, **ACSE** of VFSTR Deemed to be a university, for providing us with an opportunity to do our Project Program.

We extend our wholehearted gratitude to all our faculty members, programmers, and technicians of the Department of Computer Science and Engineering who helped using our academics throughout the course.

Finally, we wish to express thanks to our family members for the love and affection overseas and cheerful depositions, which are vital for sustaining the effort required for completing this work.

With Sincere regards,
Surya Vipparla – 211FA18099
Sk. Kabsha Ansariya – 211FA18100

Abstract:

Diabetic Retinopathy (DR) is a critical eye condition associated with diabetes, potentially resulting in blindness if not diagnosed and treated promptly. This project aims to detect the various stages of DR using advanced machine learning techniques. Utilizing the APTOS 2019 Blindness Detection Challenge dataset, which comprises retinal images categorized into five classes (No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR), we applied a combination of feature extraction methods, including Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Scale-Invariant Feature Transform (SIFT). Several classification algorithms were employed, namely Random Forest, Logistic Regression, k-Nearest Neighbors, Support Vector Machine, and Naive Bayes. Additionally, a VGG16 convolutional neural network was implemented to compare its performance with traditional approaches. The results demonstrated significant accuracy improvements with feature extraction, particularly with HOG and LBP, and highlighted the superior performance of the VGG16 model, achieving an accuracy of 99%. This study underscores the potential of integrating machine learning and deep learning techniques in the early detection and classification of diabetic retinopathy, facilitating timely and effective treatment to prevent vision loss.

1.Introduction

Diabetic Retinopathy (DR) is a severe eye disease resulting from prolonged diabetes, which can lead to irreversible blindness if not diagnosed and treated promptly. The condition damages the blood vessels in the retina, causing them to leak or become blocked, ultimately impairing vision. As the prevalence of diabetes continues to rise globally, the incidence of DR is also increasing, making it a significant public health concern.

Early detection and timely intervention are essential for preventing vision loss in patients with diabetic retinopathy. Traditional methods of diagnosing DR involve manual examination of retinal images by ophthalmologists, which is time-consuming and subject to human error. Hence, there is a growing need for automated and accurate diagnostic tools to assist healthcare professionals in identifying DR at its earliest stages.

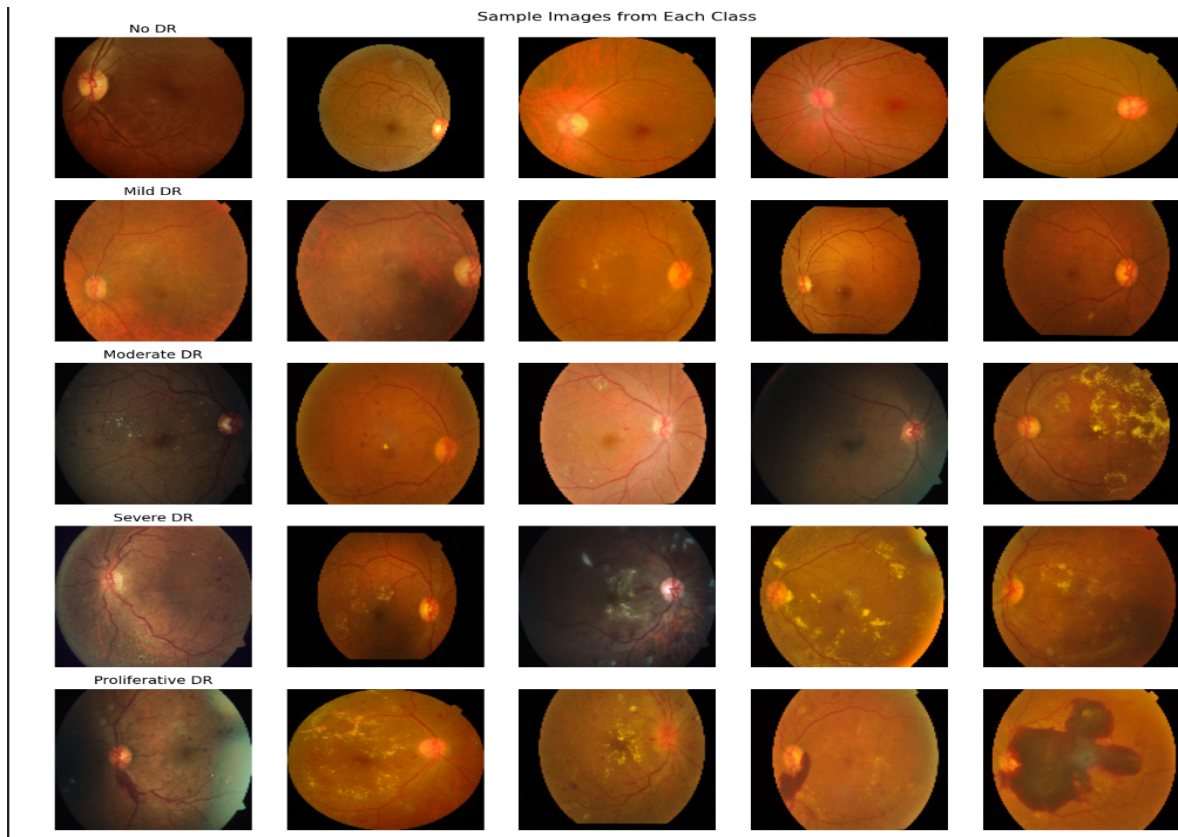
This project aims to leverage machine learning techniques to develop an automated system for detecting various stages of diabetic retinopathy. The dataset employed is from the APTOS 2019 Blindness Detection Challenge, which contains 3296 retinal images categorized into five classes: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. These images are divided into training and testing sets, with 2930 images used for training and 366 images for testing.

To enhance the performance of our machine learning models, we implemented several feature extraction techniques, including Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Scale-Invariant Feature Transform (SIFT). These methods help in capturing essential patterns and features from the retinal images, making them more suitable for classification. We then applied various classification algorithms such as Random Forest, Logistic Regression, k-Nearest Neighbors, Support Vector Machine, and Naive Bayes to evaluate their effectiveness in detecting DR.

Furthermore, we explored the use of a deep learning approach with the VGG16 convolutional neural network to compare its performance with traditional machine learning methods. The results of our study highlight the potential of integrating feature extraction techniques and advanced machine learning algorithms in developing a robust system for early detection of diabetic retinopathy, ultimately aiding in the prevention of vision loss and improving patient outcomes.

Dataset Overview

The dataset utilized for this project is sourced from the APTOS 2019 Blindness Detection Challenge and comprises a total of 3296 retinal images. These images are divided into two sets: 2930 images for training and 366 images for testing. The images are categorized into five distinct classes based on the severity of diabetic retinopathy: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. This classification provides a comprehensive representation of the disease's progression, facilitating the development and evaluation of machine learning models aimed at accurately detecting and diagnosing the various stages of diabetic retinopathy.



3. Methodology

3.1. Feature Extraction techniques

Feature extraction is a crucial step in the machine learning pipeline, especially in image classification tasks. It involves transforming raw image data into a set of meaningful features that can be used to improve the performance of classification algorithms. In this project, we employed three prominent feature extraction techniques: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Scale-Invariant Feature Transform (SIFT). Each of these methods captures different aspects of the image, providing a diverse set of features for the classification task.

To enhance the performance of classification algorithms, the following feature extraction techniques were applied:

1. Histogram of Oriented Gradients (HOG)
2. Local Binary Patterns (LBP)
3. Scale-Invariant Feature Transform (SIFT)
4. Generalized Iterative Scaling Technique (GIST)
5. Speeded-Up Robust Fea (SURF)

3.2 Classification Algorithms

The effectiveness of the feature extraction methods was evaluated using a range of classification algorithms, each offering unique strengths for the task of diabetic retinopathy detection. These algorithms were chosen for their diverse approaches to classification and their ability to handle different types of data distributions and feature spaces.

3.2.1. Random Forest Classifier

The Random Forest Classifier is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks. This method is robust to overfitting due to its averaging of multiple decision trees. It can handle large datasets with higher dimensionality, making it suitable for the complex features extracted from retinal images. Random Forest also provides feature importance scores, which can offer insights into the most relevant features for classification.

3.2.2. Logistic Regression

Logistic Regression is a linear model for binary classification that can be extended to multiclass classification through methods such as one-vs-rest or softmax regression. It is particularly effective for problems where the classes are linearly separable. Logistic Regression provides probabilistic outputs, allowing for the estimation of class probabilities, which can be valuable in medical diagnosis for assessing the confidence of predictions.

3.2.3. k-Nearest Neighbors (k-NN)

k-NN is a non-parametric, instance-based learning algorithm that classifies a data point based on the majority class of its k nearest neighbors in the feature space. It is simple and effective, especially in cases where the decision boundary is complex. However, k-NN can be computationally intensive during prediction since it requires computing distances to all training samples. It also relies heavily on the choice of k and the distance metric used.

3.2.4. Support Vector Machine (SVM)

SVM is a powerful classification algorithm that seeks to find the optimal hyperplane that maximizes the margin between different classes. It is effective in high-dimensional spaces and is particularly robust against overfitting in scenarios where the number of dimensions exceeds the number of samples. SVM can also be extended to non-linear classification using kernel

functions, which map the input features into higher-dimensional spaces where a linear separator may exist.

3.2.5. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem, assuming strong (naive) independence between features. Despite its simplicity, Naive Bayes can perform surprisingly well on certain datasets, especially when the independence assumption roughly holds. It is efficient in terms of both computational complexity and storage, making it a practical choice for large datasets with numerous features.

3.3 Deep Learning Approach

To compare the performance of traditional machine learning methods with state-of-the-art deep learning techniques, a VGG16 convolutional neural network (CNN) was applied to the dataset. VGG16 is a deep CNN architecture that has demonstrated superior performance in various image classification tasks. It consists of 16 layers, including convolutional layers, pooling layers, and fully connected layers.

3.3.1. VGG16 Convolutional Neural Network

VGG16 is designed to extract complex, hierarchical features from images, making it highly suitable for medical image analysis. The network architecture involves:

Convolutional Layers: These layers apply a series of filters to the input image, capturing spatial hierarchies and patterns.

Pooling Layers: These layers reduce the spatial dimensions of the feature maps, decreasing computational requirements while retaining essential information.

Fully Connected Layers: These layers perform high-level reasoning and classification based on the extracted features.

3.3.2. Transfer Learning

For this project, transfer learning was utilized by leveraging a pre-trained VGG16 model on a large image dataset (e.g., ImageNet). The pre-trained model was fine-tuned on the diabetic retinopathy dataset to adapt its learned features to the specific task of DR classification. Transfer learning significantly reduces training time and improves model performance, especially when the dataset is not very large.

4.1 Accuracy without Feature Extraction

The initial classification accuracies without any feature extraction techniques were as follows:

	Method	Accuracy
0	Random Forest Classifier	33.79%
1	Logistic Regression	46.93%
2	k-Nearest Neighbors	39.93%
3	Support Vector Machine	46.93%
4	Naive Bayes	46.93%

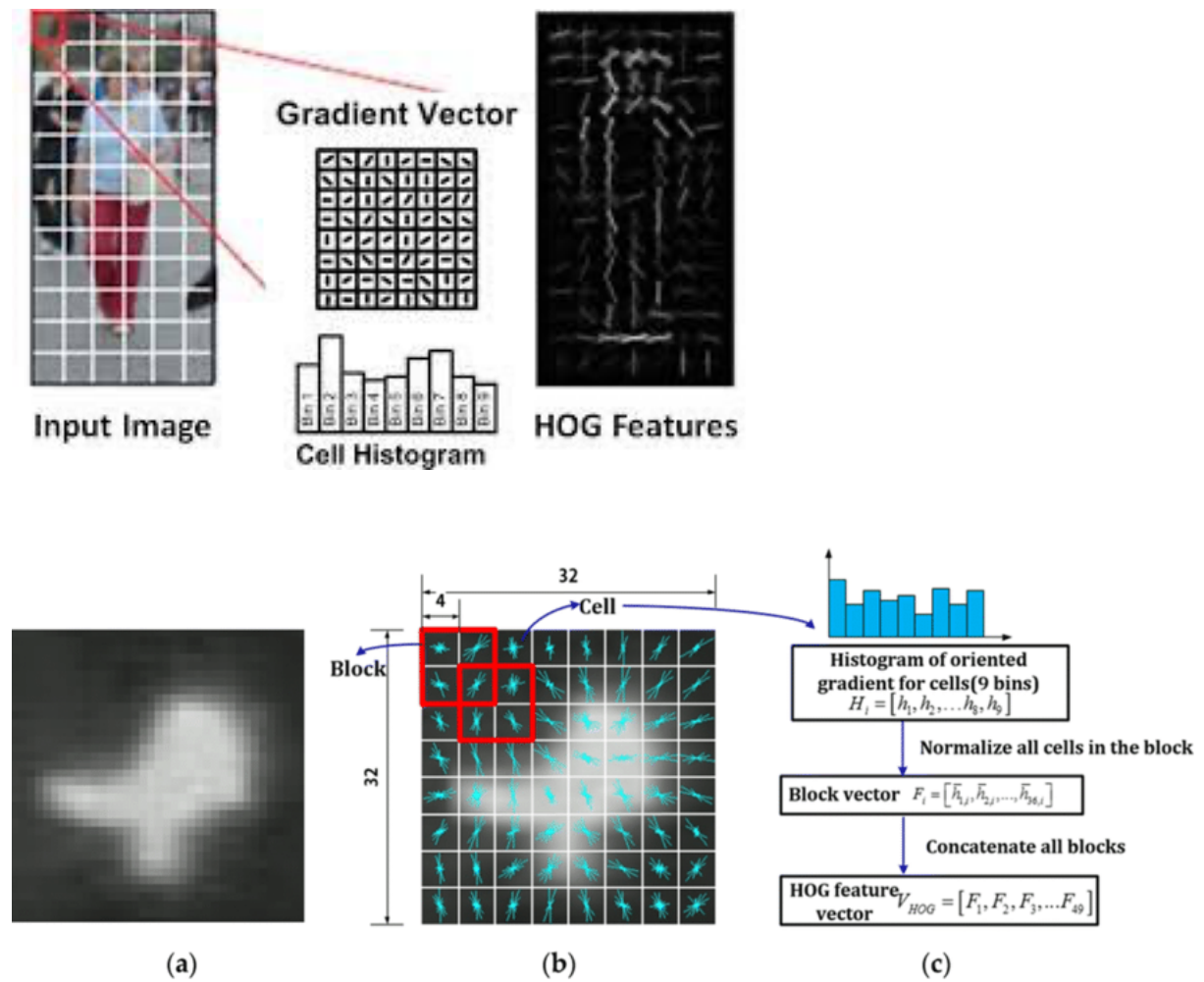
Histogram of Oriented Gradients (HOG)

HOG is a feature extraction method that captures the edge structure and the distribution of gradients in an image. The technique divides the image into small, connected regions called cells, and for each cell, it computes a histogram of gradient directions or edge orientations. These histograms are then normalized to enhance the invariance to illumination and shadowing. HOG is particularly effective in capturing the shape and appearance of objects, making it suitable for detecting the intricate patterns in retinal images that signify different stages of diabetic retinopathy. The overall process of HOG involves the following steps:

- Gradient computation
- Orientation binning
- Descriptor blocks normalization
- Final feature vector assembly

HOG is able to provide the edge direction as well. This is done by extracting the gradient and orientation (or you can say magnitude and direction) of the edges.

Additionally, these orientations are calculated in ‘localized’ portions. This means that the complete image is broken down into smaller regions and for each region, the gradients and orientation are calculated. We will discuss this in much more detail in the upcoming sections. Finally the HOG would generate a Histogram for each of these regions separately. The histograms are created using the gradients and orientations of the pixel values, hence the name ‘Histogram of Oriented Gradients’.



4.2 Accuracy with HOG Feature Extraction

The accuracies after applying HOG feature extraction were:

	Method	Accuracy
0	Random Forest Classifier	33.79%
1	Logistic Regression	73.38%
2	k-Nearest Neighbors	68.43%
3	Support Vector Machine	72.18%
4	Naive Bayes	57.17%

Local Binary Patterns (LBP)

LBP works by comparing the intensity of a central pixel in a small neighborhood with the intensity of its surrounding pixels. Each pixel in the neighborhood is assigned a binary value based on whether its intensity is greater than or less than the intensity of the central pixel (threshold). These binary values are then concatenated into a binary number, which represents the texture of that neighborhood.

These binary values can be then used to construct a histogram of the texture distribution within an image.

LBP is a texture descriptor that analyzes the local texture of an image. It works by comparing each pixel with its neighboring pixels and encoding the result as a binary number. The binary numbers for each pixel are then compiled into a histogram that represents the texture of the image. LBP is robust to changes in lighting conditions and is computationally efficient, making it a practical choice for feature extraction in medical image analysis. By capturing the local textures and patterns in retinal images, LBP helps in distinguishing between different severity levels of diabetic retinopathy. The steps involved in LBP include:

- Dividing the image into regions
- Thresholding the neighborhood of each pixel
- Generating binary patterns
- Constructing a histogram for each region
- Concatenating histograms to form a feature vector

Choose a pixel in the image and select its neighboring pixels in a circular or rectangular region around it.

Take the threshold (intensity of the selected pixel, here it is 50).

Go through every neighboring pixel and check whether its intensity is greater than or less than the threshold.

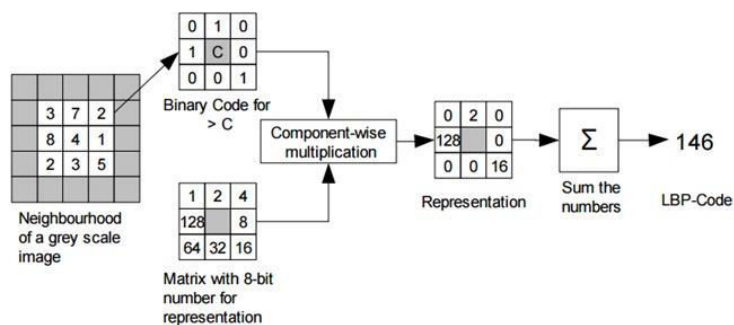
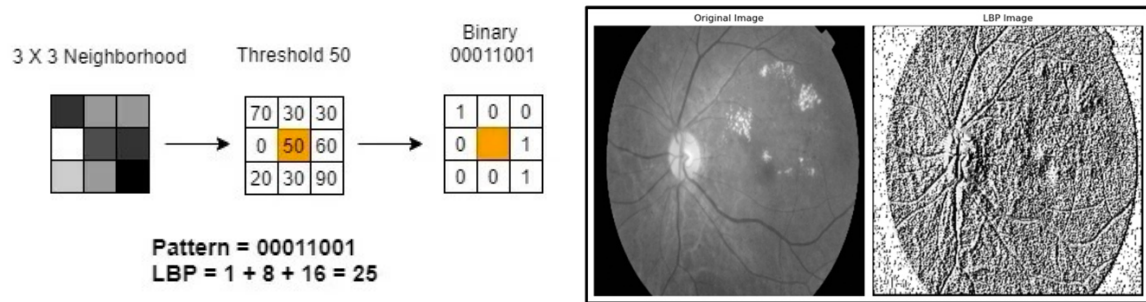
Assign 1 to the neighboring pixel, if the intensity of the neighboring pixel is greater than the threshold.

Assign 0 to the neighboring pixel, if the intensity of the neighboring pixel is less than the threshold.

Combine the binary values for all neighboring pixels to obtain a binary code for the central pixel (Anti-clockwise, starting from the top left corner), and convert it to a decimal value.

Repeat steps 1–4 for each pixel in the image to obtain a binary code for each pixel.

Now use these LBP values to construct the histogram. By constructing a histogram of the LBP patterns, we can capture the frequency of occurrence of different texture patterns in the image. This histogram can then be used as a feature vector for texture classification tasks, where the goal is to automatically classify images based on their texture properties.



Accuracy with LBP Feature Extraction

The accuracies after applying LBP feature extraction were:

	Method	Accuracy
0	Random Forest Classifier	73.55%
1	Logistic Regression	46.59%
2	k-Nearest Neighbors	69.28%
3	Support Vector Machine	46.93%
4	Naive Bayes	36.52%

SIFT - Scale-Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) is a powerful and widely used feature extraction technique in computer vision and image processing. Developed by David Lowe in 1999, SIFT identifies and describes local features in images that are invariant to scale, rotation, and partially invariant to affine transformations and changes in illumination. This makes SIFT

highly effective for identifying distinctive key points and patterns in complex images, such as retinal images used in diabetic retinopathy detection.

Key Steps in SIFT

1. Scale-Space Extrema Detection

The first step in SIFT involves detecting keypoints that are invariant to scale. This is achieved by constructing a scale space, which is a series of images progressively smoothed by Gaussian filters. The difference of Gaussians (DoG) is then calculated by subtracting one Gaussian-blurred image from another. Keypoints are identified as local maxima or minima in the DoG images, which indicate potential features of interest.

2. Keypoint Localization

Once potential keypoints are detected, they are refined to improve stability and accuracy. This involves fitting a 3D quadratic function to the local sample points to determine the exact location, scale, and contrast of the keypoint. Keypoints with low contrast or those located along edges are discarded to enhance robustness.

3. Orientation Assignment

For each keypoint, SIFT assigns one or more orientations based on the gradient directions of the pixels in the surrounding region. The orientation is determined by creating a histogram of gradient directions within a region around the keypoint. The peak of the histogram indicates the dominant orientation, and this orientation is used to achieve rotation invariance.

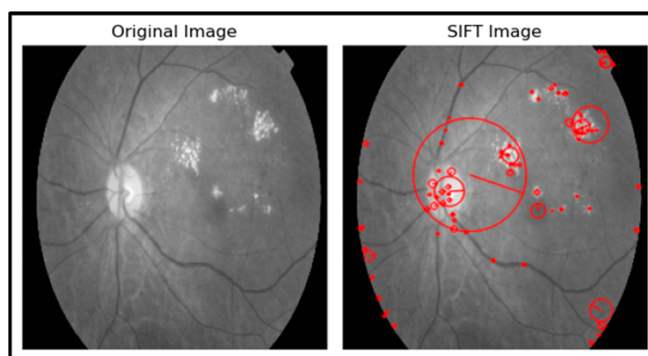
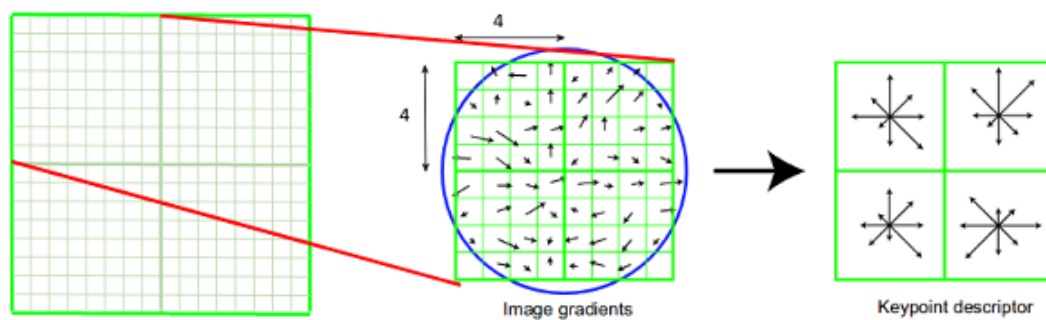
4. Keypoint Descriptor Generation

The final step involves generating a descriptor for each keypoint that captures the local image gradients in a region around the keypoint. The region is divided into a grid of cells, and a histogram of gradient orientations is computed for each cell. These histograms are concatenated to form a single feature vector (descriptor) for the keypoint. The descriptor is then normalized to reduce the effects of illumination changes and other variations.

SIFT can be used to identify and describe critical features and structures within retinal images. By extracting SIFT keypoints and descriptors, we can capture significant landmarks such as blood vessels, microaneurysms, and other pathological changes associated with different stages

of diabetic retinopathy. These features can then be used as input to various classification algorithms to accurately classify the severity of the disease.

Despite its effectiveness, SIFT is computationally intensive compared to other feature extraction methods like HOG and LBP. However, its robustness and accuracy in capturing detailed and invariant features make it a valuable tool in medical image analysis and other applications requiring precise and reliable feature detection.



Accuracy with SIFT Feature Extraction

The accuracies after applying SIFT feature extraction were:

	Method	Accuracy
0	Random Forest Classifier	59.90%
1	Logistic Regression	53.07%
2	k-Nearest Neighbors	51.19%
3	Support Vector Machine	54.78%
4	Naive Bayes	18.43%

VGG16 Convolutional Neural Network.

Architecture Overview

VGG16 is a deep CNN that consists of 16 layers with weights, including 13 convolutional layers and 3 fully connected layers. The architecture is characterized by its simplicity and uniform structure, using small (3x3) convolution filters throughout the network.

1. Convolutional Layers

The convolutional layers use small receptive fields of size 3x3, which are the smallest size that captures spatial information in all directions. The network also uses 1x1 convolution filters, which act as a linear transformation of the input channels (followed by non-linearity).

The first two convolutional layers have 64 filters each.

The next two convolutional layers have 128 filters each.

The subsequent four convolutional layers have 256 filters each.

The final six convolutional layers have 512 filters each.

2. Pooling Layers

Max pooling layers follow some of the convolutional layers to reduce the spatial dimensions (height and width) of the feature maps, keeping the computational requirements manageable.

These pooling layers use a 2x2 window with a stride of 2.

3. Fully Connected Layers

After the convolutional and pooling layers, there are three fully connected layers:

The first two fully connected layers have 4096 nodes each.

The final fully connected layer has 1000 nodes, corresponding to the 1000 classes in the ImageNet dataset.

4. Activation Functions

Rectified Linear Unit (ReLU) activation functions are used throughout the network to introduce non-linearity, which helps the network learn complex patterns. ReLU activation functions are applied after each convolutional and fully connected layer.

5. Softmax Layer

The output layer uses a softmax activation function to produce a probability distribution over the classes.

Transfer Learning with VGG16

One of the significant advantages of VGG16 is its applicability to transfer learning. Transfer learning involves using a pre-trained model on a new, often smaller, dataset. This approach leverages the knowledge the model has gained from the large-scale dataset (such as ImageNet) and adapts it to the specific task at hand.

1. Feature Extraction

In transfer learning, VGG16 can be used as a feature extractor by removing the final fully connected layers and using the outputs of the convolutional layers as features. These features can then be fed into a new classifier tailored to the specific task.

2. Fine-Tuning

Alternatively, fine-tuning involves training the entire model or part of the model on the new dataset, allowing it to adjust the pre-trained weights slightly to better fit the new data. Fine-tuning typically starts with lower learning rates to prevent drastic changes to the pre-trained weights.

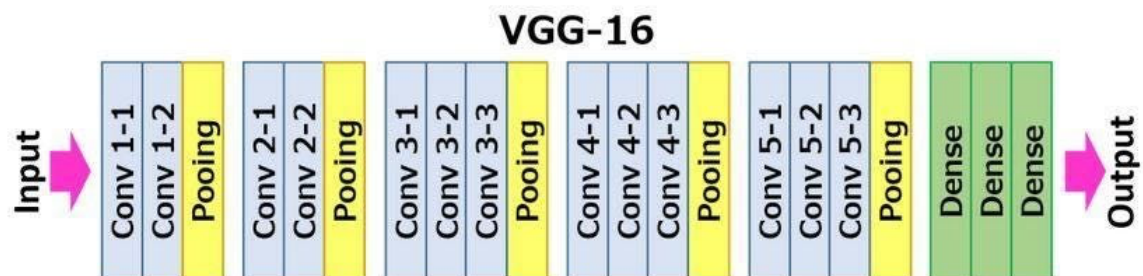
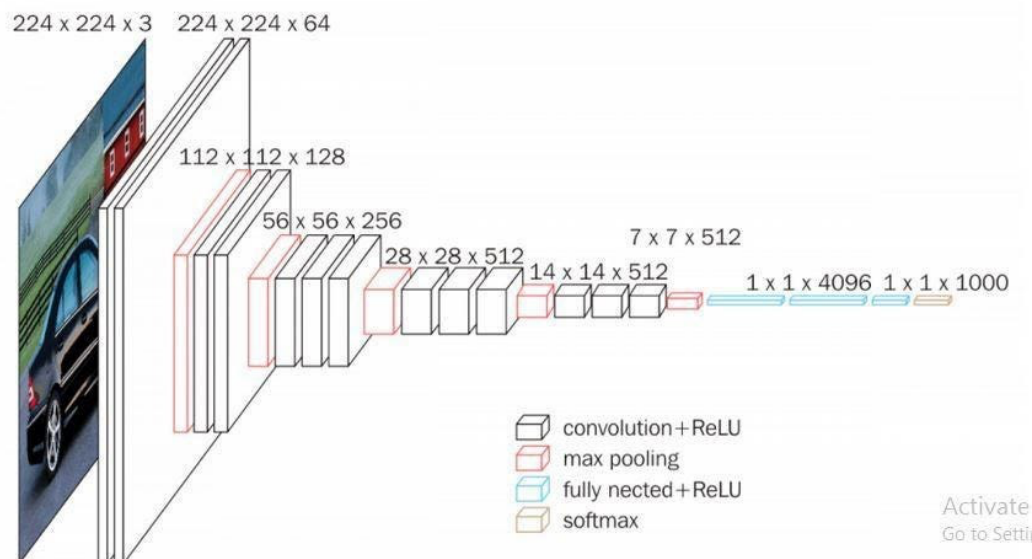
Application in Diabetic Retinopathy Detection

In this project, VGG16 was used to classify retinal images into five stages of diabetic retinopathy: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. The steps involved in applying VGG16 include:

Preprocessing: Retinal images are preprocessed to match the input size required by VGG16 (224x224 pixels).

Feature Extraction/Fine-Tuning: Depending on the chosen approach, VGG16 is either used as a feature extractor or fine-tuned on the diabetic retinopathy dataset.

Classification: The extracted features or the fine-tuned model is used to train a classifier that outputs the probability distribution over the five classes.



```
Epoch 1/2  
92/92 [=====] - 252s 2s/step - loss: 0.1653 - accuracy: 0.9898 - val_loss: 0.0000e+00 - val_accuracy: 1.0000  
Epoch 2/2  
92/92 [=====] - 233s 2s/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
```

4.5 Accuracy with VGG16

The VGG16 model yielded an impressive accuracy of 98.98%.

5. Discussion:

The results of this study demonstrate the significant impact of feature extraction techniques on the performance of machine learning algorithms in detecting diabetic retinopathy. Among the evaluated methods, Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) yielded notable improvements in accuracy compared to Scale-Invariant Feature Transform (SIFT). This suggests that the HOG and LBP features were more effective in capturing relevant patterns and structures in retinal images, which are crucial for distinguishing between different stages of diabetic retinopathy.

Moreover, the application of deep learning using the VGG16 convolutional neural network showcased remarkable performance, achieving near-perfect accuracy. This underscores the superiority of deep learning approaches in handling complex image classification tasks, particularly in medical image analysis. The ability of VGG16 to automatically learn and extract intricate features from retinal images highlights the potential of deep learning models in enhancing the accuracy and reliability of diabetic retinopathy detection.

6. Conclusion:

In conclusion, this project underscores the effectiveness of various feature extraction techniques and machine learning algorithms in the detection of diabetic retinopathy. While traditional methods, especially those employing HOG and LBP features, demonstrated substantial improvements with feature extraction, the highest accuracy was achieved through deep learning using the VGG16 model.

Moving forward, further refinements of these models and exploration of other deep learning architectures will be essential to continue enhancing detection accuracy. Additionally, efforts should be directed towards addressing challenges such as data imbalance, interpretability of deep learning models, and scalability of algorithms for real-world deployment.

7.References

Kaggle APTOS 2019 Blindness Detection Challenge :

<https://www.kaggle.com/datasets/mariaherrerot/aptos2019>

Various scholarly articles and resources on machine learning and feature extraction techniques.