Augmenting Diabetic Retinopathy Severity Prediction with a Dual-Level Deep Learning Approach Utilizing Customized MobileNet Feature Embeddings

Jyostna Devi Bodapati[1]*, Rajasekhar Konda[2]

[1] Department of Advanced Computer Science and Engineering, Vignan's Foundation for Science Technology and Research, 522213 Guntur, India

[2] Software Engineering Manager & IEEE Senior Member, 94536 San Francisco, California, USA

* Correspondence: Jyostna Devi Bodapati (jyostna.bodapati82@gmail.com)

Abstract: Diabetic retinopathy, a severe ocular disease correlated with elevated blood glucose levels in diabetic patients, carries a significant risk of visual impairment. The essentiality of its timely and precise severity classification is underscored for effective therapeutic intervention. Deep learning methodologies have been shown to yield encouraging results in the detection and categorisation of severity levels of diabetic retinopathy. This study proposes a dual-level approach, wherein the MobileNetV2 model is modified for a regression task, predicting retinopathy severity levels and subsequently fine-tuned on fundus images. The refined MobileNetV2 model is then utilised for learning feature embeddings, and a Support Vector Machine (SVM) classifier is trained for grading retinopathy severity. Upon implementation, this dual-level approach demonstrated remarkable performance, achieving an accuracy rate of 87% and a kappa value of 93.76% when evaluated on the APTOS19 benchmark dataset. Additionally, the efficacy of data augmentation and the handling of class imbalance issues were explored. These findings suggest that the novel dual-level approach provides an efficient and highly effective solution for the detection and classification of diabetic retinopathy severity levels.

Keywords: Color fundus images; Diabetic retinopathy (DR); Pre-trained ConvNet (PCN); Deep neural network (DNN); VGG Net, Xception; InceptionResNetV2 (Inception V4); Machine learning models (ML)

1 Introduction

Diabetic retinopathy (DR), a debilitating ocular condition, inflicts substantial damage on the retina and, if unchecked, can precipitate irreversible vision loss. This condition is a consequence of the body's impaired glucose processing and storage mechanisms, typically associated with diabetes. Elevated blood glucose levels wreak havoc on blood vessels, including those within the retina, instigating a series of pathological stages [1]. As delineated in Table 1, distinct clinical manifestations correspond to various DR phases, initiating with non-proliferative diabetic retinopathy (NPDR). This stage is characterized by blurred or foggy vision resulting from retinal tissue swelling. If unaddressed, NPDR can escalate to proliferative diabetic retinopathy (PDR), marked by the formation of fragile blood vessels that hinder retinal and vitreous circulation [2]. Progression to PDR can evoke complications such as scar tissue formation, glaucoma, and vitreous hemorrhage, each of which can progressively damage the optic nerve and lead to severe vision loss or total blindness [3]. The paramount importance of early detection and intervention in DR is underscored by Figure 1, which visually represents various lesion types observable in retinopathy patients.

Effective classification and staging of DR severity are vital for preventing vision loss, as proper management can avert over 90% of cases [4]. This is particularly pertinent given the anticipated escalation in global diabetes incidence and prevalence, which is projected to reach epidemic proportions by 2030 [5]. Therefore, the imperative for developing robust diagnostic and intervention strategies for DR is increasingly pressing.

In the context of this escalating public health challenge, this paper introduces a novel, dual-level deep learning approach to enhance the prediction of DR severity levels. Employing custom MobileNet feature embeddings, the

approach promises to provide a highly effective solution for the early detection and classification of DR, thus significantly contributing to the global efforts to combat this debilitating condition.

Table 1. Clinical signs of diabetic retinopathy at different stages

| Severity	Clinical Signs	Type
No. DR	Asymptotic	-
Mild DR	Dotted and blotted hemorrhages, microaneurysms, cotton wool spots	Non-Proliferative (NPDR)
Moderate DR	More pronounced signs of moderate DR	
Severe DR	Notable increase in the severity of hemorrhages, venous beading, and intra-retinal microvascular abnormalities	
Proliferate DR	Vitreous hemorrhage, neovascularization, and retinal detachment	Proliferative (PDR)
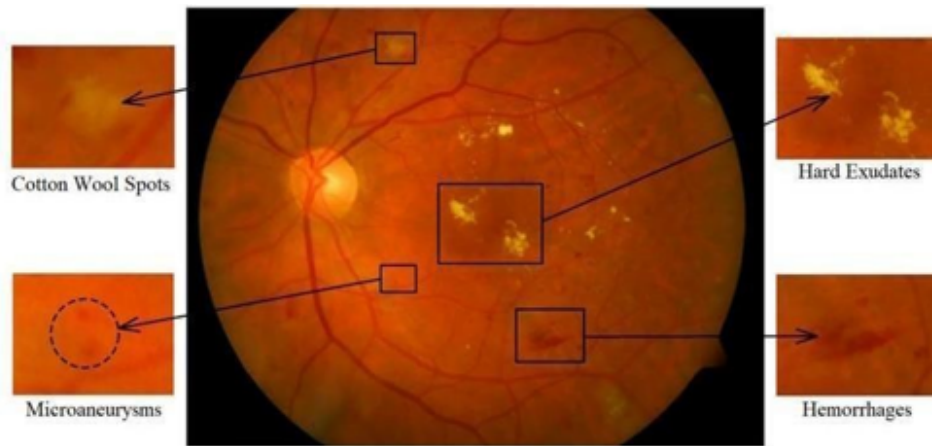
Figure 1. Color fundus imaging: Depicting various lesion types in the retinas of patients with retinopathy

Consistent dilated eye examinations are emphatically recommended for individuals with an extensive history of diabetes, primarily to detect early signs of diabetic retinopathy. Implementing preventative strategies, such as meticulously managing blood glucose levels through medication, dietary modifications, and physical activity, abstaining from alcohol and smoking, and controlling hypertension, may decelerate or even mitigate the onset of diabetic retinopathy [6]. The conventional diagnosis of DR primarily involves retinal screening, however, the incorporation of machine learning algorithms for automated diagnosis could potentially serve as a valuable adjunct for ophthalmologists. Comprehensive eye examinations encompass patient history, visual acuity measurements, and the assessment of ocular structures to determine the severity of DR. Additional diagnostic tests may include fluorescein angiography and retinal imaging or tomography. The integration of machine learning algorithms can augment the diagnostic process, serving as a supportive tool for ophthalmologists by facilitating the automated diagnosis of diabetic retinopathy. This could potentially enhance diagnostic precision and improve clinical decision-making processes. The therapeutic approach to DR is customized according to the disease's stage. During the initial stages, vigilant monitoring and efficient regulation of blood sugar levels are of paramount importance. Intraocular injections of medication can mitigate the progression of DR by inhibiting the formation of abnormal blood vessels. In more advanced stages, a series of laser treatments can be employed to obliterate abnormal blood vessels, potentially causing some peripheral vision loss to preserve the more crucial central vision [7].

The advent of automated methods for screening and diagnosing DR constitutes a promising stride in the realm of medical imaging. These methodologies, leveraging techniques such as deep learning, fuzzy logic, and deep feature extraction, have exhibited impressive proficiencies in identifying DR lesions, including hard exudates, hemorrhages, and microaneurysms [8]. The incorporation of these methodologies into clinical settings harbors the potential to significantly aid healthcare professionals in attaining accurate and prompt diagnoses, ultimately fostering improved patient outcomes. Furthermore, the Sequentially Embedded Surrogate (SeS) method, as proposed by Verma et al., presents a valuable approach for enhancing the efficiency of training deep learning models [9]. This is particularly pertinent in real-world applications where expedient training processes are paramount. By augmenting training efficiency, the SeS method bolsters the feasibility and practicality of deploying AI-equipped systems for DR diagnosis

in clinical settings. Collectively, these studies highlight the substantial promise of AI in medical image analysis, specifically within the scope of fundus image analysis [10].

Automated models, including machine learning approaches, aimed at detecting and classifying diabetic retinopathy (DR), are presently in their nascent stages of development. Additional research is mandated to enhance accuracy and generalizability, and to address inherent limitations such as the necessity for voluminous training datasets and potential data bias [11]. The cost and accessibility of technological equipment capable of capturing high-resolution retinal images may pose implementation challenges in certain regions or amongst specific populations [12]. Despite these obstacles, the incorporation of machine learning techniques in the field of DR detection and classification exhibits exceptional potential [13].

Current research endeavors are vigorously pursuing an array of techniques and algorithms specifically designed to tackle various facets of the disease. These areas of focus encompass microaneurysm detection, identification of retinal lesions, disease stage classification, and the detection of new vessels in instances of proliferative diabetic retinopathy [14].

For example, Long et al. [1] proposed a method that involves dynamic thresholding and fuzzy C-means clustering (FCM), subsequent to which a support vector machine (SVM) is deployed for recognition. One limitation of this approach, however, is its dependence on manual parameter tuning, thereby restricting its applicability across diverse datasets [15].

Similarly, a technique proposed by Haloi et al. [2] leverages mathematical morphology, Gaussian scale space, and support vector machine classification for exudate detection and categorization. Despite this method demonstrating impressive sensitivity and prediction rates, it might be susceptible to image quality and may struggle to demonstrate robust generalizability across new datasets [16].

The two-step method introduced by Eftekhari et al. [3] presents significant advancements in microaneurysm (MA) detection. However, a potential pitfall of this CNN-based method is the risk of overfitting to the training dataset. Conversely, the method developed by Bodapati [17] aims to augment CNN training efficiency via the selective sampling of misclassified negative samples. A potential trade-off with this strategy, however, is that it may inadvertently extend the overall training time.

In the context of Srivastava et al. [5], their introduction of novel filters to distinguish red lesions from blood vessels in the detection of microaneurysms and hemorrhages exhibits superior performance compared to existing methodologies. It should be noted, though, that their approach relies on patch-level processing, which may make it vulnerable to variations in image quality [18].

Haloi et al. [2] introduced an innovative deep neural network model incorporating dropout layers and a maxout activation function, achieving exceptional accuracy in spotting microaneurysms (MAs) [19]. However, the detection of MAs alone may not provide a comprehensive diagnosis of diabetic retinopathy (DR). Several models, including those proposed by Akram et al. [7], and Welikala et al. [10, 11], have demonstrated encouraging results in discerning the severity of DR [20]. Despite these advances, these models could require manual parameter tuning, may lack broad applicability to new datasets, and carry the risk of overfitting to the training dataset [21]. Addressing these limitations and probing solutions that offer a more holistic and resilient evaluation of disease progression will undoubtedly fortify the effectiveness of DR diagnostic methodologies [22].

Our proposed method presents an innovative and efficient strategy for predicting the severity levels of diabetic retinopathy (DR), harnessing the capabilities of deep learning techniques. Our approach comprises a series of strategic steps, including fine-tuning, regression model training, and learning embeddings to establish a robust solution. The initial phase involves pre-processing the raw retinal images, followed by their integration into a custom-tailored MobileNetV2 architecture for fine-tuning. A key distinguishing feature of our approach is the adaptation of the MobileNet architecture into a regression model, explicitly designed to predict DR severity levels. Furthermore, the trained model is employed for learning embeddings from the fundus images. To rigorously assess the efficacy of our proposed methodology, comprehensive experiments were conducted using the APTOS19 dataset, which contains retinal images. The results of these experiments strongly endorse the effectiveness of our approach, as it achieves a remarkable accuracy rate of 87% in predicting DR severity levels. This significant level of accuracy underscores the potential of our methodology in facilitating precise predictions and early detection of the disease. By providing such accurate forecasts, our approach has the potential to play a crucial role in the early identification and subsequent treatment of diabetic retinopathy, ultimately leading to improved patient outcomes.

2 Literature Review

Diabetic retinopathy (DR) presents a significant risk to vision, emphasizing the critical need for early detection and management. This review delves into the evolving realm of automated DR detection methodologies, with the objective to scrutinize, categorize, and critically appraise their merits, limitations, and potential implications for clinical practice. The methodologies examined in the literature can be systematically categorized into distinct subgroups as displayed in Table 2, each of which is analyzed in this section.

Table 2. Sub-categories of the diabetic retinopathy methods explored in the literature

Category	Reference
Traditional Techniques	Long et al. [1]
	Haloi et al. [2]
Deep Learning Paradigms	Eftekhari et al. [3]
	Bodapati and Balaji [4]
	Zeng et al. [18]
Hybrid Approaches	Akram et al. [7]
	Casanova et al. [8]
Microaneurysms and Hemorrhages	Srivastava et al. [5]
	Welikala et al. [11]
New Vessels and Proliferative DR (PDR)	Welikala et al. [10]
	Welikala et al. [11]
Deep Learning for DR Features	Zeng et al. [18]
Microaneurysm Detection	Mateen et al. [19]

Traditional Techniques: Some studies have employed conventional image processing methods for DR detection. Long et al. [1] introduced an algorithm that amalgamates dynamic thresholding, fuzzy C-means clustering, and support vector machine (SVM) classification to accurately pinpoint hard exudates (HEs). While these methods demonstrate proficiency under controlled conditions, their susceptibility to fluctuations in lighting and image quality warrants attention. Haloi et al. [2] proposed a method that blends a Gaussian scale space approach with mathematical morphology and SVM classification for exudate detection. This category traces the historical trajectory of traditional methods and their contextual effectiveness.

Deep Learning Paradigms: A distinct group of studies has embraced deep learning, particularly Convolutional Neural Networks (CNNs), for DR detection [23]. Eftekhari et al. [3] designed a two-step CNN architecture for microaneurysm (MA) detection, explicitly addressing data imbalance to enhance sensitivity compared to prior models. Bodapati and Balaji [4] revolutionized CNN training by implementing selective sampling of misclassified negative samples, which expedited training without sacrificing accuracy. Zeng et al. [18] delved into multiple neural network models—backpropagation, deep neural networks, and CNNs— for the identification of critical DR features. Their exploration underscores the dynamic and evolving nature of deep learning methodologies.

Hybrid Approaches: Some studies, such as those by Akram et al. [7], have ventured into hybrid models that amalgamate the strengths of diverse methodologies. These approaches adeptly fuse the Gaussian Mixture Model and m-Medoids based modeling with SVM and ensemble classifiers, achieving noteworthy accuracy in detecting MAs and hemorrhages.

Microaneurysms and Hemorrhages: Initiatives such as those by Srivastava et al. [5] and Welikala et al. [11] delve deeply into microaneurysm and hemorrhage detection. Srivastava et al. [5] introduced innovative filters and utilized Multiple Kernel Learning to enhance feature extraction and machine learning. Welikala et al. [11] strategically integrate vessel segmentation and Random Forests classification for a comprehensive DR analysis.

New Vessels and Proliferative DR (PDR): Welikala et al. [10, 11] offer dual classification systems for detecting new vessels, aligning with the clinical aim of diagnosing proliferative DR. These studies establish a critical foundation for tracking disease progression. Welikala et al. [11] presented an automated method for detecting new vessels in retinal images, leveraging two vessel segmentation approaches and a dual classification system. This method proves effective as a screening tool for the early detection of proliferative diabetic retinopathy (PDR), delivering improved sensitivity and specificity compared to traditional line operator approaches [24].

Welikala et al. [11] proposed an automated method for detecting new vessels in retinal images using a dual classification approach. Their methodology incorporates binary vessel maps and local morphological features to generate 21-D feature vectors. To enhance classification performance, they employ a genetic algorithm-based feature selection approach, achieving high sensitivity and specificity rates. This method demonstrates considerable potential for automated PDR diagnosis.

Deep Learning for DR Features: Zeng et al. [18] introduced an automated deep learning model for identifying key precursors of diabetic retinopathy (DR) using fundus images. Their exploration included backpropagation neural networks (NN), deep neural networks (DNN), and convolutional neural networks (CNN) [24]. Deep learning models exhibited superior accuracy compared to NN models, effectively quantifying DR features and severity levels [25].

Microaneurysm Detection: Rahim et al. [14] developed an automatic screening system focused on microaneurysm detection in color fundus images. Their approach blends feature extraction methods, the circular Hough transform, and fuzzy histogram equalization for preprocessing, resulting in enhanced microaneurysm detection during the fundus image preprocessing stage [26].

In a broader context, traditional techniques provide interpretability and context-specific efficacy, yet they are vulnerable to real-world variations [27]. Conversely, deep learning models excel in processing vast datasets and deciphering intricate patterns, albeit at the cost of significant computational resources [28]. The domain of hybrid models presents potential, but with the caveat of potential complexity and increased resource demands [29]. The synthesis of these diverse methodologies has the potential to transform DR screening and diagnosis, thereby facilitating timely intervention and ultimately improving patient outcomes [30].

The studies considered in this review exemplify the judicious application of diverse techniques, encompassing fuzzy C-means clustering, Gaussian scale space approaches, convolutional neural networks, and hybrid classifiers [31]. Collectively, these approaches yield high sensitivity, low false positive rates, and state-of-the-art accuracy, effectively addressing the critical need for early detection in publicly available datasets [32]. The potential of these algorithms to significantly enhance DR monitoring and early detection, while extending their applicability to other medical image analysis domains, remains an impressive testament to their innovation and impact [33].

3 Proposed Methodology

The primary aim of this study is to introduce a versatile classification model specifically designed for the identification of diabetic retinopathy severity levels. Given the inherent scarcity of available training data, our proposed methodology is strategically constructed to maximize the utility of the limited dataset at our disposal. In pursuit of our objective, we emphasized the creation of an adept feature representation mechanism for retinal images. Our methodology comprises several key steps, with the relationship between these steps depicted in Figure 2.
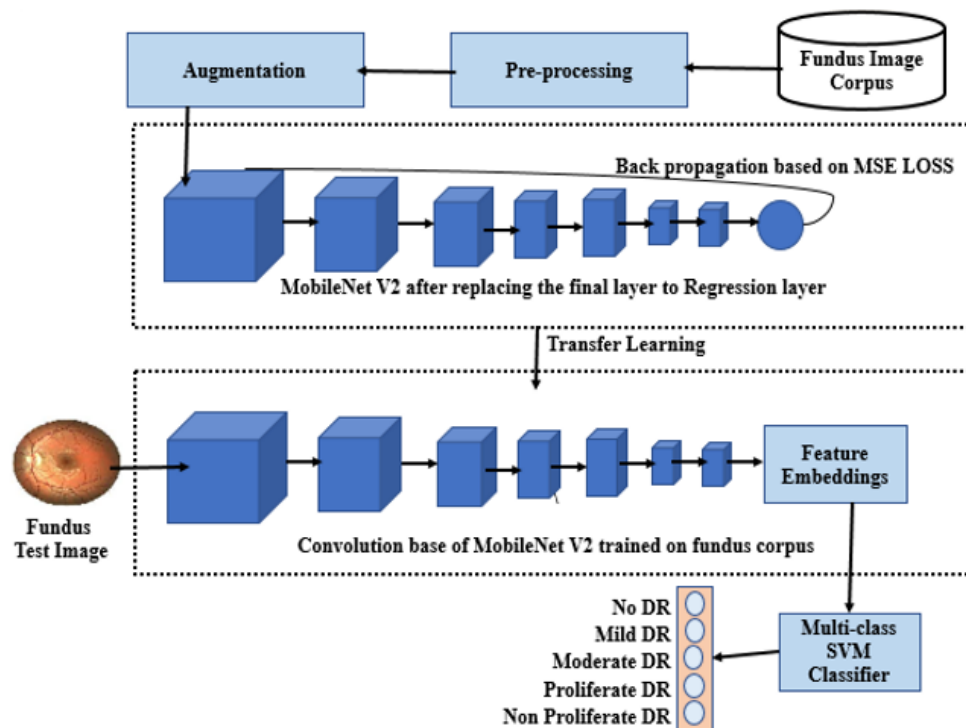

Figure 2. Workflow of the proposed dual-level deep learning approach

Our proposed approach unfolds through a series of sequential stages, each contributing to the comprehensive framework: Image pre-processing, Data augmentation, Customizing the MobileNetV2 architecture, Fine-tuning the customized model, Learning Retinopathy Image Embeddings, Training a model for classification, and ultimately, model evaluation for disease prediction tasks. The process initiates with the resizing of retinal images, standardizing them to a resolution of 224×224 pixels. To enhance dataset diversity, we judiciously employ data augmentation techniques, ensuring robust model generalization. Notably, the MobileNetV2 architecture undergoes strategic customization. Following the model training phase, the model is readied for feature extraction from fundus images, capitalizing on the knowledge it has garnered. The customized MobileNetV2 model commences its training with the adoption of the pre-trained weights from the original MobileNetV2 model, which was trained on the comprehensive ImageNet dataset. During the fine-tuning process, the convolutional layer weights are frozen, while the fully-connected layers are updated by fine-tuning the model with the DR dataset. The model employs regression to predict the severity value, which is an ordinal variable. The extracted features are then classified using classifiers such as the Support Vector Machine (SVM). Below, we delve into the details of each of the modules used in our proposed work.

Image Pre-processing: The image pre-processing phase is a critical step where retinal images undergo a series of transformations, setting the stage for subsequent analysis. The initial step in this stage is the uniform resizing of retinal images to a standardized resolution of 224×224 pixels. This resizing is crucial to synchronize the images with the dimensions that the pre-trained MobileNetV2 architecture is designed to handle. As a result, compatibility is assured between the retinal images and the architecture, paving the way for seamless integration [33].

Data Augmentation: Data augmentation is a technique that amplifies the volume and diversity of data without necessitating additional data collection. This is accomplished by generating different variations of the existing data, thereby effectively increasing the dataset size. The proposed approach employs image processing techniques, such as zooming, cropping, flipping, and rotating, to augment the original images [34]. The augmented images are amalgamated with the initial dataset and used to train the models. This strategy aids in mitigating overfitting and enhancing model generalization. By exposing the model to variations of the training data, it can recognize the same object under different conditions, thereby increasing its robustness to real-world data variations.

Customizing MobileNetV2: In our proposed approach, MobileNetV2 serves as the fundamental architecture for feature extraction. A key attribute of MobileNetV2 is its employment of depth-wise separable convolutions, a technique which significantly reduces the volume of parameters and computations required for processing [35]. Despite this reduction, the methodology maintains an impressive level of accuracy. The unique combination of competitive accuracy, compact model size, and computational efficiency that MobileNetV2 displays positions it as an ideal candidate for the crucial role of feature extraction. This design choice optimizes resource utilization and contributes to faster inference times, which are vital for real-time applications. Compared to its predecessor, MobileNetV1, MobileNetV2 has exhibited superior performance and capability, solidifying its standing as an effective solution for feature embedding across diverse applications. Its incorporation within our proposed approach for diabetic retinopathy severity classification showcases its capacity to meet the demands of the task while maintaining efficiency and accuracy.

We perform a strategic customization of the MobileNetV2 architecture to adapt it for efficient feature extraction tasks. Specifically, the final output softmax layer is replaced with a dense layer featuring a sigmoid activation function. This architectural adjustment aligns the model's suitability for regression tasks, enabling it to predict severity grades as continuous values.

Fine-tuning Modified MobileNetV2: The customized MobileNetV2 model is subjected to fine-tuning to optimize its performance in predicting diabetic retinopathy (DR) severity levels. Before fine-tuning, the DR images are augmented, and the pre-trained weights of the MobileNetV2 architecture – which were originally trained on the extensive ImageNet dataset – are utilized as initial weights for the network. These foundational weights equip the customized model with base knowledge. During the fine-tuning process, the weights of the convolutional layers remain unaltered, while only the fully-connected layers are updated using the specific DR image dataset. This approach stems from the presumption that the convolutional layers of pre-trained networks, having been exposed to extensive datasets, inherently possess effective feature extraction capabilities. The fine-tuning phase is steered by the goal of minimizing the sum of squared loss, a metric that is particularly suitable for regression tasks. The formula for the loss function is defined in Eq. (1).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{1}$$

where, Y_i represents the target value to be predicted and $\hat{Y}_i$ denotes the prediction generated by the modified MobileNetV2 architecture. The fine-tuning process aims to adjust the network's parameters to minimize the deviation between the predicted and actual target values.

Learning Retinopathy Image Embeddings: In our proposed approach, the feature extraction process takes place after the fine-tuning of the Modified MobileNetV2 network. The pre-trained MobileNetV2 model, specifically fine-tuned for the task of DR severity classification, forms the foundation for this pivotal phase of feature extraction. As part of this process, fundus images are fed into the input layer of the network, commencing the extraction of the most pertinent features from the convolutional layers inherent to the MobileNetV2 architecture. For each input image, a total of 512 features are extracted, sourced from the outputs of the first and second fully connected layers (256 features from fc1 and 256 features from fc2) within the network. These specific layers have demonstrated proficiency in capturing and encapsulating essential patterns and information learned by the network during the fine-tuning process. The dimensions of the input remain constant, ensuring a consistent and uniform feature extraction process across a diverse range of images. By extracting 512 features from each image, a compact and efficient representation of the fundus image is attained. These features serve as a condensed representation of the image, capturing the most relevant information for DR severity classification.

MobileNet Features: MobileNetV2 is chosen for feature extraction in our proposed approach due to its competitive accuracy, smaller model size, and computational efficiency. It is specifically designed for mobile and embedded

devices, which makes it suitable for efficient computation in environments with limited resources. MobileNetV2 employs depth-wise separable convolutions to reduce parameters and computations while preserving accuracy. Its architecture incorporates linear bottleneck blocks with inverted residuals and shortcut connections to enhance nonlinearity and gradient flow. With its flexibility in width and resolution parameters, MobileNetV2 provides improved performance over its predecessor, MobileNetV1, making it a favoured choice for feature embedding in various applications.

Disease Prediction Module: In the disease prediction module, the extracted features are processed and classified using classifiers such as Support Vector Machines (SVM). SVMs are widely used for disease prediction in medical imaging due to their versatility and ability to handle both linearly and non-linearly separable data. They aim to find a hyperplane that maximally separates data points in an N-dimensional space. For linearly separable data, a linear SVM can be used, while for non-linearly separable data, a non-linear SVM with a kernel function is employed to transform the data into a higher-dimensional space. Common kernel functions include linear, sigmoid, radial basis function (RBF), and polynomial kernels.

Common kernels within the context of SVM include the following in Table 3:

Table 3. Types of popular kernels suitable with support vector machines

Kernel Name	Mathematical Expression	Parameters
Linear kernel	$K(x, y) = x^T y$	None
Polynomial kernel	$K(x, y) = \left(x^T y + c\right)^d$	c, d
Radial basis function (RBF) kernel	$K(x, y) = \exp\left(-\,gamma * \|x - y\|^2\right)$	gamma
Sigmoid kernel	$K(x, y) = \tanh\left(gamma * x^T y + r\right)$	Gamma, r
Laplacian kernel	$K(x, y) = \exp(-\|x - y\|/\sigma)$	σ

Indeed, kernel selection hinges on the unique features of the specific problem and the inherent characteristics of the data. The Radial Basis Function (RBF) kernel is particularly esteemed for its proficiency in manoeuvring complex patterns and nonlinear data. However, the pivotal aspect lies in the realm of experimentation, which aids in identifying the most suitable kernel and corresponding parameters tailored to each distinct problem scenario.

4 Experimental Results

Our experimental investigations were designed to evaluate the effectiveness of our proposed method for detecting and grading Diabetic Retinopathy (DR). This section provides a synopsis of the dataset utilized, the evaluation metrics, and the comparative studies conducted.
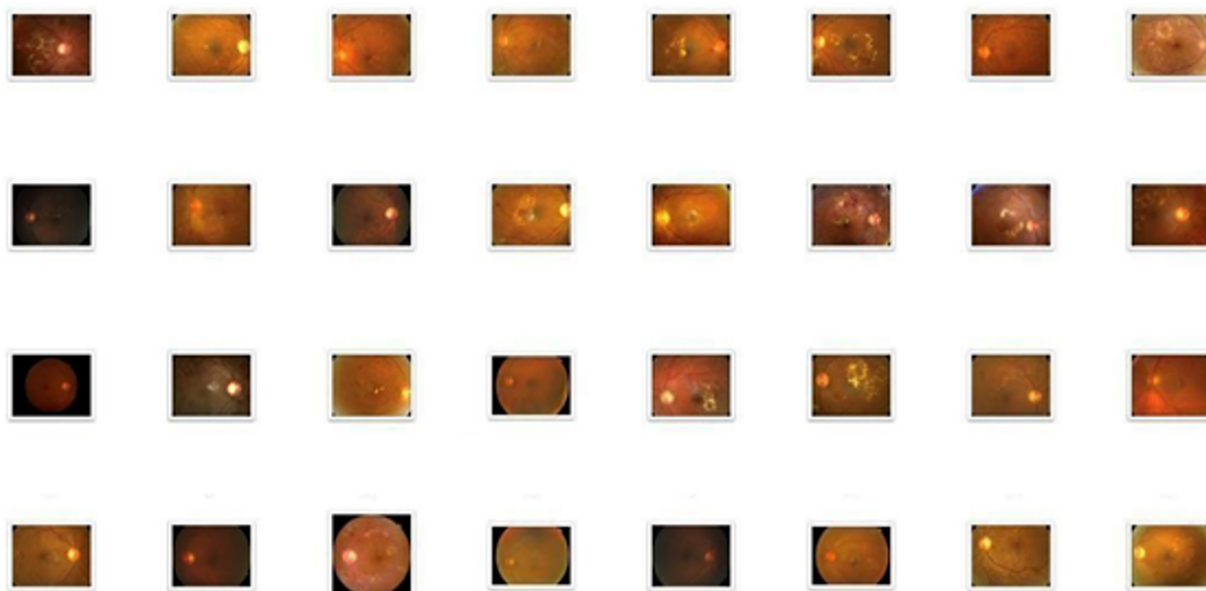

Figure 3. Sample dataset images from APTOS 2019 diabetic retinopathy dataset

Dataset Overview: The dataset is composed of retinal images captured through fundus photography, obtained from the APTOS 2019 Blindness Detection Challenge hosted on Kaggle. These images were meticulously graded on a scale of 0 to 4, each grade reflecting a different level of DR severity. The dataset encompasses images sourced from

a variety of clinics, captured under diverse imaging conditions, and utilizing a range of scanners over an extensive timeframe. The distribution of severity levels within the APTOS2019 dataset is as follows: No DR – 180 samples, Mild DR: 370 samples, Moderate DR: 999 samples, Severe DR: 193 samples, Proliferative DR: 295 samples, with a total of 3662 samples. Figure 3 shows sample images from the APTOS 2019 dataset.

Table 4. Performance of traditional ML models with different embeddings for diabetic retinopathy severity prediction

Embedding Type	Classifier	Accuracy	Precision	F1-Score	Recall
	SVM	79	62	61	78
	MLP	79	62	69	78
Raw Pixel Values	ADABOOST	74	65	68	74
	KNN	72	67	69	72
	Random Forest	79	73	70	78
	SVM	79	62	69	78
	MLP	79	69	78	62
ResNet101	ADABOOST	77	69	71	76
	KNN	72	66	68	72
	Random Forest	77	69	71	76
	SVM	79	62	69	78
	MLP	79	62	69	78
DenseNet169	ADABOOST	79	75	73	79
	KNN	73	68	70	73
	Random Forest	78	69	70	78
	SVM	79	62	69	78
	MLP	79	62	69	78
VGG-19	ADABOOST	74	63	67	74
	KNN	65	63	64	66
	Random Forest	77	61	68	76
	SVM	79	62	69	78
	MLP	79	62	69	78
VGG-16	ADABOOST	77	70	72	76
	KNN	71	66	68	70
	Random Forest	77	70	72	76
	SVM	79	62	69	78
	MLP	79	62	69	78
MobileNetV2	ADABOOST	79	73	72	78
	KNN	70	62	66	70
	Random Forest	77	67	70	76

Evaluation Metrics: Our experimental analysis employs crucial evaluation metrics including Accuracy, Precision, Recall, and F1 Score, to conduct an in-depth assessment of the performance of our proposed method for diabetic retinopathy (DR) severity prediction.

Accuracy serves as a measure of the proportion of correctly classified instances amongst the total number of samples. While it offers an overarching perspective of the model's correctness, its validity might be compromised in scenarios where class imbalance is present.

Precision measures the proportion of correctly predicted positive instances out of the total instances predicted as positive. In our context, it gauges the aptitude of the model in accurately predicting a specific DR severity level when it asserts that level.

Recall, also known as sensitivity or true positive rate, quantifies the ratio of correctly predicted positive instances to the total number of genuine positive instances. It signifies the model's capacity to identify all occurrences of a particular severity level.

The F1 Score represents the harmonic mean of Precision and Recall. It provides a balanced measure that takes into account both false positives and false negatives, making it an invaluable metric for evaluating classification models in circumstances where class imbalance or unequal misclassification costs are present.

The Kappa Statistic accounts for the potential of agreement occurring by chance. It quantifies the concurrence between predicted and expected classifications beyond what can be achieved by random chance. This statistic, with its range from -1 to 1, assists in understanding how well the model's predictions correspond with the true classes, considering the possibility of random agreement. The formula for calculating the Kappa Statistic is: Kappa = (Accuracy - Expected Accuracy) / (1 - Expected Accuracy).

Collectively, these metrics offer comprehensive insights into the efficacy of the model in accurately identifying and predicting the presence or absence of the disease.

Experimental Setup: For the purpose of our experimentation, we utilized a high-end GPU system equipped with an Intel Xenon processor, a 2 TB hard disk drive, and 32 GB of DDR4 RAM. The GPU in use is the NVIDIA GeForce GTX with a memory capacity of 16 GB. All experiments were conducted using Python 3.6.6, with TensorFlow 1.14 and Keras 2.3 serving as the backbone for building and training the neural network models.

Experimental Results: Our experimental exploration comprised two distinct phases, aimed at evaluating the performance of various machine learning models in predicting the severity of Diabetic Retinopathy (DR). Initially, the focus was on training a spectrum of traditional machine learning models using various feature embeddings. These embeddings encompassed representations from diverse pre-trained deep models as well as raw pixel values. In the subsequent phase, we employed renowned traditional machine learning models, including Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), AdaBoost, K-Nearest Neighbors (KNN), and Random Forest classifiers. These models were harnessed to predict the severity of diabetic retinopathy.

Table 4 elucidates the performance of various traditional machine learning models such as SVM, MLP, ADABOOST, KNN, and Random Forest classifiers, utilizing different types of feature embeddings for the task of diabetic retinopathy Severity Prediction. The embeddings employed include Raw Pixel Values, ResNet101, DenseNet169, VGG-19, VGG-16, and MobileNetV2.

From the derived results, it is unequivocally apparent that MobileNet embeddings outshine other embeddings, demonstrating superior performance across all classifiers. This suggests that MobileNet is particularly well-suited for the specific task of Diabetic Retinopathy Severity Prediction. Moreover, the SVM classifier transcends all other classifiers for this task, regardless of the type of embedding used, indicating SVM as a fitting choice of classifier for this task.

In examining the individual performances of embeddings, both DenseNet169 and VGG-16 exhibited commendable results across most classifiers. However, the ADABOOST and KNN classifiers didn't deliver satisfactory performances with these embeddings. This underscores the importance of classifier selection when working with different types of embeddings.

Table 5. Performance comparison of SVM when different pre-trained embeddings are used to represent diabetic retinopathy images

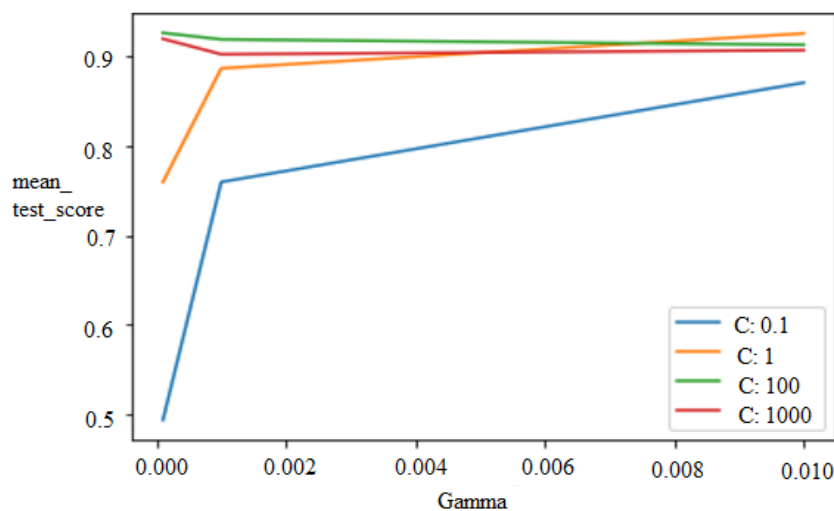| Model	Accuracy	Precision	F1-Score	Recall
ResNet+SVM	79	62	69	78
DenseNet+SVM	79	62	69	78
VGG-19+SVM	79	62	69	78
VGG-16+SVM	79	62	69	78
MobileNetV2+SVM	87	79	73	70

Figure 4. Mean test scores of the model with respect to Gamma and C while performing the grid search

Table 5 reveals that the combination of MobileNet embeddings with the SVM classifier delivers an accuracy of 79%, precision of 62%, recall of 78%, and an F1-score of 69%. This combination emerges as the most effective for

diabetic retinopathy Severity Prediction. These findings furnish crucial insights into the potential applicability of our proposed method for practical clinical implementations, particularly in the field of diabetic retinopathy severity evaluation. A Grid Search approach was employed during the experiments to ascertain optimal values for the hyperparameters, such as C, kernel type, and gamma values (Figure 4).

It is noteworthy that traditional machine learning models, specifically SVM, consistently outperform the other classifiers used in this study, which include MLP, AdaBoost, KNN, and Random Forest. This result underscores the ongoing viability and competitiveness of traditional machine learning models in predicting diabetic retinopathy severity. Broadly, our study exhibits the potential of synergizing deep learning models with traditional machine learning models for the prediction of diabetic retinopathy severity. Among these, the combination of MobileNetV2 embeddings with SVM emerges as the most potent approach, showcasing its potential to bolster the precision and effectiveness of diabetic retinopathy severity prediction. This amalgamated insight positions our research at the vanguard of advancements in both the realms of deep learning and traditional machine learning for clinical applications, particularly in diabetic retinopathy assessment.

5 Conclusions

This research endeavors to construct an automated diabetic retinopathy (DR) severity level prediction model utilizing retinal images. The proposed model harmoniously integrates an SVM classifier and deep feature extraction techniques to categorize images into varying severity levels based on the extracted features. The study offers a promising pathway towards automating DR screening, thereby alleviating the workload of ophthalmologists. Nevertheless, the model's efficacy could be further amplified by incorporating additional deep learning techniques and larger datasets. Experimental results corroborate that the amalgamation of the MobileNetV2 model and SVM classifier demonstrates superior performance relative to other models. Our proposed model, evaluated on the APTOS2019 dataset comprised of 3297 retinal images, achieves an admirable accuracy of 87%. Fine-tuning MobileNetV2 as a regression model prior to learning embeddings enhances performance, while the introduction of data augmentation and class imbalance management techniques further bolsters the accuracy of DR severity level classification. The proposed methodology holds considerable promise to expedite early DR detection and management, consequently reducing the risk of vision loss in diabetic patients.

In future research pursuits, addressing challenges associated with data imbalance, augmenting the model's robustness in real-world contexts, and emphasizing transparency in the decision-making processes inherent to automated DR detection systems should be prioritized. These efforts will significantly contribute to the evolution of accurate and ethically responsible automated DR assessment technologies.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. C. Long, X. X. Huang, Z. Q. Chen, S. Pardhan, and D. C. Zheng, "Automatic detection of hard exudates in color retinal images using dynamic threshold and SVM classification: Algorithm development and evaluation," *BioMed Res. Int.*, vol. 2019, pp. 1–10, 2019. https://doi.org/10.1155/2019/3926930

[2] M. Haloi, S. Dandapat, and R. Sinha, "A gaussian scale space approach for exudates detection, classification and severity prediction," *arXiv preprint*, 2015. https://doi.org/10.48550/arXiv.1505.00737

[3] N. Eftekhari, H. R. Pourreza, M. Masoudi, K. Ghiasi-Shirazi, and E. Saeedi, "Microaneurysm detection in fundus images using a two-step convolutional neural network," *BioMed. Eng. OnLine*, vol. 18, p. 67, 2019. https://doi.org/10.1186/s12938-019-0675-9

[4] J. D. Bodapati and B. B. Balaji, "Tumor AwareNet : Deep representation learning with attention based sparse convolutional denoising autoencoder for brain tumor recognition," *Multimed. Tools Appl.*, vol. 2023, pp. 1–19, 2023. https://doi.org/10.1007/s11042-023-15557-w

[5] R. Srivastava, L. X. Duan, D. W. K. Wong, J. Liu, and T. Y. Wong, "Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels," *Comput. Methods Programs Biomed.*, vol. 138, pp. 83–91, 2017. https://doi.org/10.1016/j.cmpb.2016.10.017

[6] L. Wu, P. Fernandez-Loaiza, J. Sauma, E. Hernandez-Bogantes, and M. Masis, "Classification of diabetic retinopathy and diabetic macular edema," *World J. Diabetes*, vol. 4, no. 6, pp. 290–294, 2013. https://doi.org/10.4239/wjd.v4.i6.290

[7] M. U. Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," *Comput. Biol. Med.*, vol. 45, pp. 161–171, 2014. https://doi.org/10.1016/j.compbiomed.2013.11.014

[8] R. Casanova, S. Saldana, E. Y. Chew, R. P. Danis, C. M. Greven, and W. T. Ambrosius, "Application of random forests methods to diabetic retinopathy classification analyses," *PLoS One*, vol. 9, no. 6, p. e98587, 2014. https://doi.org/10.1371/journal.pone.0098587

[9] K. Verma, P. Deep, and A. G. Ramakrishnan, "Detection and classification of diabetic retinopahy using retinal images," in *2011 Annual IEEE India Conference, Hyderabad, India*, 2011, pp. 1–6. https://doi.org/10.1109/INDCON.2011.6139346

[10] R. A. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, and S. A. Barman, "Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 247–261, 2014. https://doi.org/10.1016/j.cmpb.2014.02.010

[11] R. A. Welikala, M. M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, and S. A. Barman, "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy," *Comput. Med. Imaging Graph.*, vol. 43, pp. 64–77, 2015. https://doi.org/10.1016/j.compmedimag.2015.03.003

[12] S. Roy chowdhury, D. Koozekanani, and K. K. Parhi, "Dream: Diabetic retinopathy analysis using machine learning," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1717–1728, 2013. http://dx.doi.org/10.1109/JBHI.2013.2294635

[13] L. F. Porter, N. Saptarshi, Y. Fang, S. Rathi, A. I. Den Hollander, E. K. De Jong, S. J. Clark, P. N. Bishop, T. W. Olsen, T. Liloglou, V. R. M. Chavali, and L. Paraoan, "Whole-genome methylation profiling of the retinal pigment epithelium of individuals with age-related macular degeneration reveals differential methylation of the SKI, GTF2H4, and TNXB genes," *Clin. Epigenetics*, vol. 11, no. 1, p. 6, 2019. https://doi.org/10.1186/s13148-019-0608-2

[14] S. S. Rahim, C. Jayne, V. Palade, and J. Shuttle worth, "Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening," *Neural Comput. Appl.*, vol. 27, pp. 1149–1164, 2016. https://doi.org/10.1007/s00521-015-1929-5

[15] S. Dutta, B. C. S. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyenger, "Classification of diabetic retinopathy images by using deep learning models," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 1, pp. 89–106, 2018. http://dx.doi.org/10.14257/ijgdc.2018.11.1.09

[16] V. Kakani, B. Varun, J. D. Bodapati, and K. R. Sekhar, "Post-COVID chest disease monitoring using self-adaptive convolutional neural network," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India*, 2023. https://doi.org/10.1109/I2CT57861.2023.10126288

[17] J. D. Bodapati, "SAE-PD-Seq: Sequence autoencoder-based pre-training of decoder for sequence learning tasks," *Signal Image Video Process.*, vol. 15, pp. 1453–1459, 2021. https://doi.org/10.1007/s11760-021-01877-7

[18] X. L. Zeng, H. Q. Chen, Y. Luo, and W. B. Ye, "Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30 744–30 753, 2019. https://doi.org/10.1109/ACCESS.2019.2903171

[19] M. Mateen, J. H. Wen, S. Nasrullah, Song, and Z. P. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, 2019. https://doi.org/10.3390/sym11010001

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, 2014. https://doi.org/10.48550/arXiv.1409.1556

[21] F. Chollet, "Xception: Deep learning with depth wise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA*, 2017, pp. 1251–1258. https://doi.org/10.1109/CVPR.2017.195

[22] J. D. Bodapati and B. Bagepalli Balaji, "Self-adaptive stacking ensemble approach with attention based deep neural network models for diabetic retinopathy severity prediction," *Multimed. Tools Appl.*, vol. 2023, pp. 1–20, 2023. https://doi.org/10.1007/s11042-023-15120-7

[23] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA*, 2018, pp. 8697–8710. https://doi.org/10.1109/CVPR.2018.00907

[24] J. D. Bodapati, R. Sajja, and V. Naralasetti, "An efficient approach for semantic segmentation of salt domes in seismic images using improved UNET architecture," *J. Inst. Eng. India Ser. B*, vol. 104, pp. 569–578, 2023.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002. https://doi.org/10.1109/TPAMI.2002.1017623

[26] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review,"

Neural Comput., vol. 29, no. 9, pp. 2352–2449, 2017. https://doi.org/10.1162/neco_a_00990

[27] X. Y. Deng, Q. X. Luan, W. T. Chen, Y. L. Wang, M. H. Wu, H. J. Zhang, and Z. Jiao, "Nanosized zinc oxide particles induce neural stem cell apoptosis," *Nanotechnology*, vol. 20, no. 11, p. 115101, 2009. https://doi.org/10.1088/0957-4484/20/11/115101

[28] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint*, 2016. https://doi.org/10.48550/arXiv.1605.07678

[29] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, 2018. https://doi.org/10.1155/2018/7068349

[30] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada*, 2009, pp. 41–48. https://doi.org/10.1145/1553374.1553380

[31] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada*, 2014, pp. 3320–3328.

[32] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA*, 2017, pp. 4278–4284.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017. https://doi.org/10.1145/3065386

[34] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *13th European Conference on Computer Vision, ECCV 2014, Zurich, Switzerland*, 2014, pp. 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

[35] M. C. Cheng, P. M. Liao, W. W. Kuo, and T. P. Lin, "The Arabidopsis ETHYLENE RESPONSE FACTOR1 regulates abiotic stress-responsive gene expression by binding to different cis-acting elements in response to different stress signals," *Plant Physiol.*, vol. 162, no. 3, pp. 1566–1582, 2013. https://doi.org/10.1104/pp.113.221911