International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Detection of Diabetic Retinopathy Using Convolutional Neural Networks

Vedant Gulhane[1], Manthan Tayade[2], Amruta Thakare[3], Samruddhi Shirbhate[4], Vedant Bhagat[5], Prof. Aniket Thakur[6]

[1,2,3,4,5]Department of Information Technology, Sipna College Of Engineering and Technology, Amravati,
[6]Assistant professor Department of Information Technology, Sipna College Of Engineering and Technology

Abstract:
Diagnosing diabetic retinopathy (DR) from colour fundus images is a challenging and time-consuming task, requiring experienced clinicians to detect numerous small features and interpret a complex grading system. In our paper, we suggest using a Convolutional Neural Network (CNN) approach to automate DR diagnosis and accurately classify its severity from digital fundus images. Our developed CNN architecture, combined with data augmentation techniques, is capable of identifying intricate features crucial for classification, such as micro-aneurysms, exudates, and haemorrhages on the retina, enabling automatic diagnosis without manual intervention. We trained this network using a high-performance graphics processing unit (GPU) on the publicly available Kaggle dataset and achieved impressive results, especially for high-level classification tasks.
In our experiments, utilizing a dataset comprising 3600 images, our proposed CNN attained an accuracy of 87% when validated against 500 additional images. These results demonstrate the effectiveness of our CNN approach in automating DR diagnosis with high accuracy.

Keywords: Convolutional Neural Network, Deep learning, Diabetic retinopathy, exudates, Fundus, Micro-aneurysms, Retina.

1. Introduction

Individuals afflicted with diabetes are susceptible to a condition termed "Diabetic retinopathy," a grave ocular ailment capable of inducing vision loss and blindness in those with diabetes. The markedly elevated levels of blood sugar precipitate substantial harm to the retinal blood vessels. Leakage of fluid from these vessels initiates swelling or thickening of the macula, impeding blood circulation. Occasionally, abnormal blood vessel proliferation on the retina emerges. Each of these circumstances may culminate in irreversible vision impairment[5]. Initially, asymptomatic, diabetic retinopathy progressively exacerbates, imperiling vision. Early detection plays a pivotal role in safeguarding visual health. Manifestations may be absent in the initial phases, but as the ailment advances, individuals may encounter challenges such as reading difficulties, distant vision impairment, visual floaters, clouded vision, nocturnal vision impairment, vision instability, compromised colour perception, shadow-induced dark areas in vision, and eventual complete vision loss.

A. Proliferative Diabetic Retinopathy:

When left untreated, diabetic retinopathy advances to a more severe stage known as proliferative diabetic retinopathy (PDR). In this stage, abnormal growth of new blood vessels occurs in the retina, leading to increased pressure within the eyeball as these vessels disrupt fluid flow, causing detachment of the retina from the back of the eye. Additionally, blood may leak into the gel-like substance (vitreous) at the centre of the eye. These factors contribute to damage to the optic nerve, responsible for transmitting inverted images from the eye to the brain via the blind spot, ultimately resulting in vision loss. To expedite the diagnostic process and provide accurate predictions, a Convolutional Neural Network (CNN) approach has been developed. CNN technology, already proven effective in various domains such as healthcare [1,2] and intelligent automation[3], is leveraged in this study. By harnessing its capabilities, the CNN model is applied to diagnose diabetic retinopathy from eye images and classify them based on severity accurately. This system is designed to automatically diagnose diabetic retinopathy without requiring user intervention.

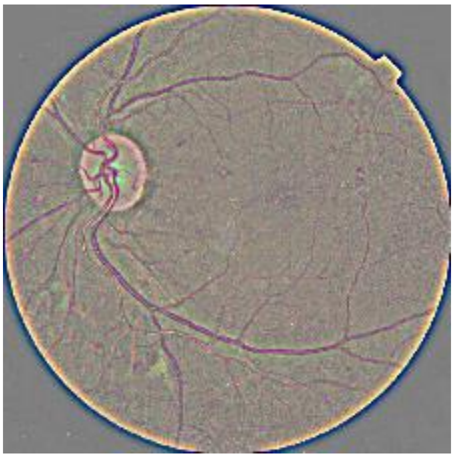


Figure 1: Proliferative Diabetic Retinopathy

B. Non-Proliferative Diabetic Retinopathy:

A person diagnosed with non-proliferative diabetic retinopathy (NPDR) experiences leakage from tiny blood vessels, resulting in swelling of the retina. Macular oedema, characterized by swelling of the macula, is a primary cause of vision loss in diabetic patients. Another condition affecting vision is macular ischemia, where the closure of retinal blood vessels prevents blood from reaching the macula, leading to the formation of small particles known as exudates. NPDR can be further categorized into three types based on symptom severity:

- Mild NPDR: Characterized by a few microaneurysms.

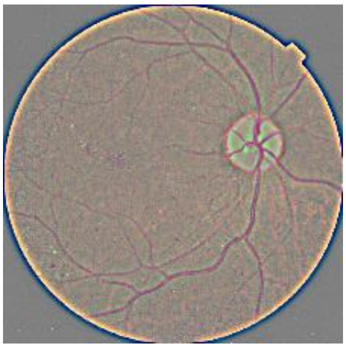


Figure 2: Mild Non-Proliferative Diabetic Retinopathy

- Moderate NPDR: Identified by the presence of cotton-wool spots and haemorrhages.

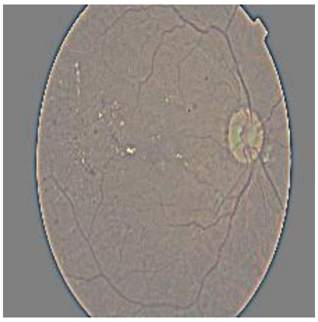


Figure 3: Moderate Non-Proliferative Diabetic Retinopathy

- Severe NPDR: Marked by intra-retinal hemorrhaging in four quadrants of the eye—two with venous beading or one with intra-retinal microvascular abnormality.

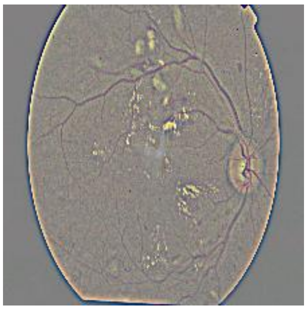


Figure 4: Severe Non-Proliferative Diabetic Retinopathy

C. Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs), a subset of deep learning techniques, have demonstrated remarkable efficacy in image analysis and understanding, particularly in domains like medical imaging [6]. Architectures tailored for processing image data emerged as early as the 1970s [9], offering valuable solutions and outperforming alternative methods in tasks such as handwritten character recognition[10]. However, it was not until significant advancements in neural network techniques, and the availability of enhanced computing power facilitated by graphical processing units (GPUs), that CNNs became viable for more intricate image recognition challenges [11]. Presently, large-scale CNN models are adept at handling complex image recognition tasks involving numerous object categories with exceptional accuracy [6].

Automated grading, especially with CNNs, faces two primary challenges. First, achieving a balanced sensitivity (accurately identifying patients with diabetic retinopathy) and specificity (correctly identifying patients without diabetic retinopathy) is particularly demanding, especially with national criteria, which entail a five-class problem encompassing normal, mild diabetic retinopathy (DR), moderate DR, severe DR, and proliferative DR [8]. Additionally, overfitting poses a significant concern in neural networks, exacerbated by skewed datasets that lead the network to excessively focus on the most prevalent class. In our dataset, less than three per cent of the images originated from the fourth and fifth classes, necessitating adjustments to ensure the network could effectively learn the features of these images.

This paper introduces a deep learning-based CNN approach for classifying diabetic retinopathy in fundus imagery, a medically relevant imaging task discussed earlier and one that has been extensively studied in the past[8]. To the best of our knowledge, this is the paper addressing the five-class classification of diabetic retinopathy using a CNN methodology. We propose several novel techniques to tailor the CNN to our extensive dataset, followed by an analysis of its performance and a detailed examination of its capabilities.

2. Architecture of Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) stand as the cornerstone in modern deep learning frameworks, particularly renowned for their efficacy in image analysis tasks, prominently including but not limited to object detection and image classification. At the heart of CNNs lies their ability to process images, which are essentially represented as multi-dimensional arrays or matrices of pixel values. This unique architecture leverages the convolution operation as its fundamental building block, revolutionizing the way computers interpret and analyze visual data.

Convolutional Layer: Serving as the bedrock upon which every CNN model is constructed, the convolutional layer orchestrates the extraction of intricate patterns and features from input images. These layers are composed of a multitude of filters, each characterized by its own set of weights. These filters systematically traverse the input image, performing convolution operations by computing the dot product between their weights and the corresponding segments of the input image. This process generates what is commonly referred to as a "feature map," which encapsulates salient information regarding various visual features present in the input image, such as edges, corners, and textures.

Pooling Layer: Immediately succeeding the convolutional layers [6], the pooling layer serves a dual purpose: it aids in dimensionality reduction and computational efficiency. Through operations like max pooling and average pooling, the pooling layer systematically down samples the feature maps generated by the preceding convolutional layers, effectively reducing their spatial dimensions. This dimensionality reduction not only alleviates the computational burden imposed by subsequent layers but also promotes translational invariance, thereby enhancing the network's ability to generalize to unseen data. An example of a max pooling layer with sample values is illustrated in Figure 1:

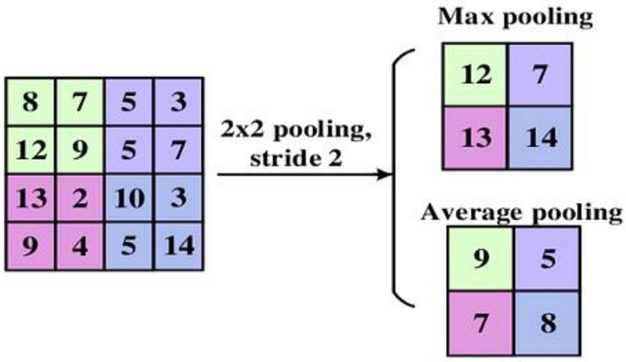


Figure 5. Max pooling layer of the CNN.

The primary aim of this layer is to decrease the size of the convolved feature map to reduce computational costs. This is performed by decreasing the connections between layers and independently operating on each feature map. Depending upon the method used, there are several types of Pooling operations. We have Max pooling and average pooling.

$$\frac{W - F}{S} + 1$$

Fully Connected (FC) Layer: As the CNN progresses deeper into its architecture, the feature maps produced by the preceding layers are eventually flattened into one-dimensional arrays. These flattened representations serve as the input to the fully connected layers, also known as dense layers. Comprising a dense arrangement of neurons interconnected with every neuron in the preceding layer, these fully connected layers play a pivotal role in synthesizing high-level abstractions from the low-level features extracted by the earlier layers. This hierarchical feature extraction culminates in the network's ability to discern intricate patterns and make accurate predictions**.** An example demonstrating the flattening process, which converts features into a one-dimensional form, is depicted in Figure 2.

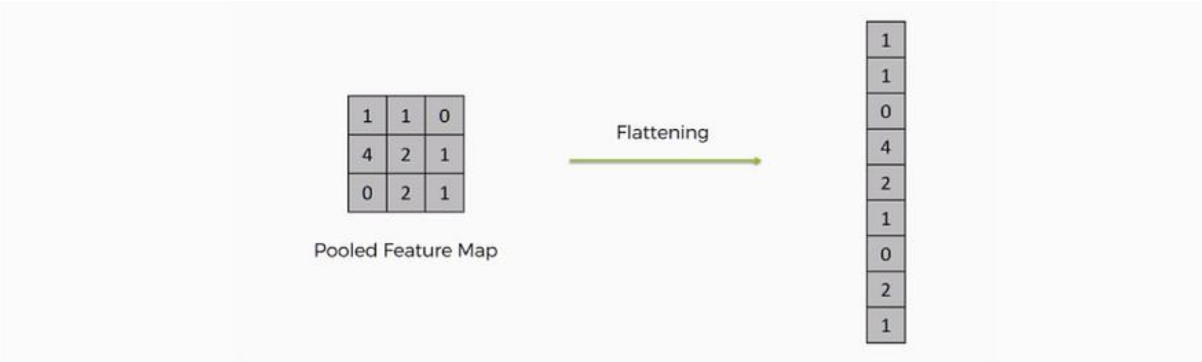
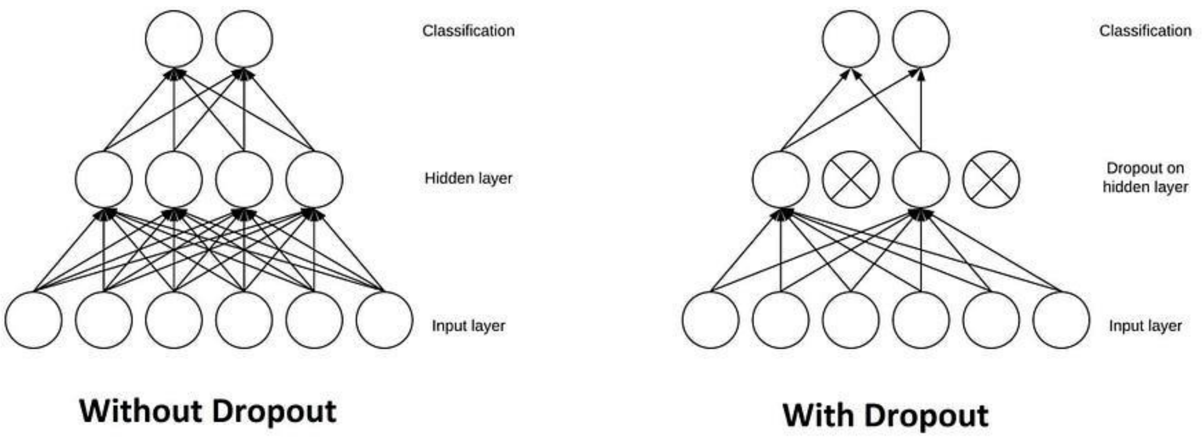


Figure 6. Flattening Layer

Dropout: Amidst the marvels of deep learning lies the lurking menace of overfitting, wherein the model inadvertently memorizes the training data instead of learning generalized patterns. Dropout, a regularization technique, emerges as a formidable weapon in the battle against overfitting. During the training phase, dropout randomly deactivates a fraction of neurons within the network with a predetermined probability. This stochastic dropout of neurons not only encourages the network to learn robust representations but also imbues it with a degree of resilience against overfitting. In the proposed model, a dropout rate of 0.2 is implemented to enhance accuracy, indicating that 20% of neurons are dropped during training [7].

Figure 7:



A. Activation function: The activation function serves as the gateway through which information flows within the neural network, imbuing it with the capacity to model complex, non-linear relationships inherent in real-world data. Among the plethora of activation functions at our disposal, several stand out

for their distinct characteristics and applications. Sigmoid and SoftMax functions are commonly used for binary classification tasks, while SoftMax is preferred for datasets with more than two classes [7].
Sigmoid: Characterized by its smooth, S-shaped curve, the sigmoid activation function finds widespread utility in binary classification tasks, where the output is constrained to the range [0, 1]. Despite its ubiquity, sigmoid's propensity for vanishing gradients renders it less suitable for deep neural networks.
tanH (Hyperbolic Tangent): Bearing resemblance to the sigmoid function, tanH is symmetric around the origin and maps input values to the range [-1, 1]. This centeredness around zero facilitates faster convergence during training, thereby mitigating the issue of vanishing gradients encountered with sigmoid activation.
Softmax: As the quintessential activation function for multi-class classification tasks, softmax normalizes the output of a neural network into a probability distribution over multiple classes. By assigning probabilities to each class, softmax furnishes the network with the capability to make informed, probabilistic predictions.
ReLU (Rectified Linear Unit): The reigning champion among activation functions in modern deep learning, ReLU introduces a thresholding mechanism that efficiently mitigates the vanishing gradient problem. By virtue of its simplicity and computational efficiency, ReLU has become the de facto choice for many neural network architectures, offering unparalleled training speeds and convergence rates.

3. Structure and Method:

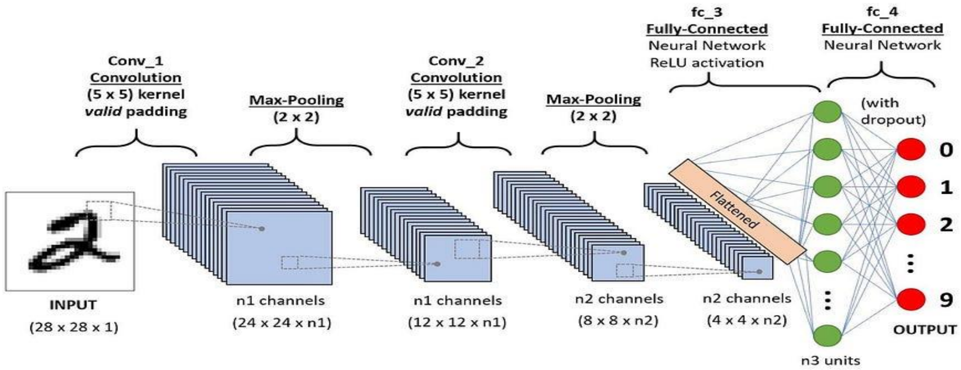
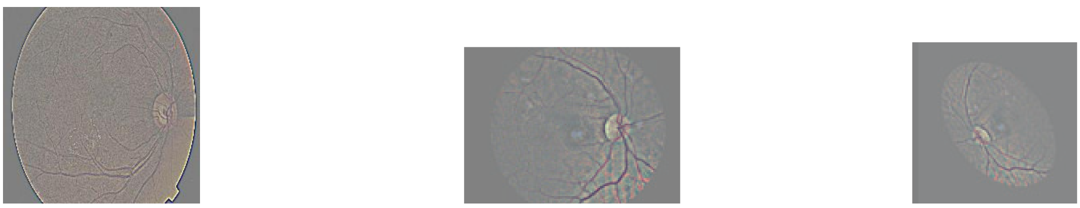


Figure 8. CNN Architecture

The architecture of our neural network [7], depicted in Fig 8, was determined following an extensive review of the literature on various image recognition tasks. The incorporation of additional convolutional layers is believed to enable the network to grasp deeper features. For instance, while the initial layers focus on learning basic features like edges, the ultimate convolutional layer, situated at the end of the network, is intended to capture more intricate features relevant to the classification of diabetic retinopathy (DR), such as hard exudate.



(a) Original image (b) Pre-processed image (c) Augmented image

Fig 9: Illustration of the preprocessing and augmentation processes

The network commences with convolutional blocks accompanied by activation functions, followed by batch normalization after each convolutional layer. As the number of feature maps expands, we transition to employing one batch normalization operation per block. Max pooling operations are consistently conducted using kernel sizes of 3x3 and strides of 2x2. Following the final convolutional block, the network undergoes flattening to condense into a one-dimensional representation.
To mitigate overfitting, we apply weighted class weights proportional to the image count in each class. Additionally, dropout is implemented on dense layers to alleviate overfitting until reaching the dense five-node classification layer, which employs a softmax activation function for classification prediction. The activation function employed is the leaky rectified linear unit (ReLU) with a leakage parameter set to 0.01 to prevent excessive reliance on specific nodes within the network. Similarly, L2 regularization is applied to weights and biases in the convolutional layers. Gaussian initialization is utilized to expedite the initial training phase. The optimization process is driven by the widely adopted categorical cross-entropy loss function.

A. Dataset, Hardware and Software:

The testing dataset utilized in our study was obtained from the Kaggle coding website (https://www.kaggle.com). This dataset comprises over 3600 images, each with approximately 6 million pixels, depicting various scales of retinopathy. To facilitate training on this sizable dataset, the images were resized, and our Convolutional Neural Network (CNN) model was executed on a high-performance GPU, specifically the NVIDIA GTX 1650 ti. The NVIDIA GTX 1650ti is equipped with 896 CUDA cores and leverages the NVIDIA CUDA Deep Neural Network library (cuDNN) for efficient GPU-based learning. Utilizing this setup, approximately 500 images were uploaded to the GPU memory at any given time. The machine learning backend was implemented using the deep learning package Keras (http://keras.io/).

B. Preprocessing:

The dataset included images from patients of different ethnicities and age groups, with varying levels of lighting in the fundus photography. This caused differences in pixel intensity values, which didn't relate to the classification. To fix this, we used color normalization with the OpenCV package (http://opencv.org/), shown in Fig 9 (b). The images were also high resolution, taking up a lot of memory. So, we resized them to 512x512 pixels to keep the important details while making sure they fit within the memory capacity of the NVIDIA 1650ti.

C. Training:

CNN was initially pre-trained on a subset of 6000 images until reaching a significant proficiency level. This preliminary training was essential to expedite the classification process without squandering considerable training time. Subsequently, after 120 epochs of training on the initial images, the network underwent further training on the complete set of 3600 training images for an additional 20 epochs.
Neural networks are prone to severe overfitting, particularly in datasets like ours, where the majority of images belong to a single class, indicating no signs of retinopathy. To address this challenge, we integrated real-time class weights into the network. With each batch loaded for back-propagation, the class weights were dynamically updated based on the ratio of images classified as having no signs of DR in the training batch. This adaptive adjustment significantly mitigated the risk of overfitting to any specific class.

4. Result:

From the dataset, 1500 images were set aside for validation purposes. Running these validation images through the network required 188 seconds. For the five-class problem at hand, specificity refers to the proportion of patients correctly identified as not having diabetic retinopathy (DR) out of the total number of patients truly not having DR, while sensitivity denotes the proportion of patients correctly identified as having DR out of the total number of patients with DR. Accuracy represents the proportion of patients correctly classified.

The final trained network yielded the following performance metrics: 93% specificity, 87% accuracy, and 30% sensitivity. In terms of numerical classification within the network, the classes were designated as follows: 0 for No DR, 1 for Mild DR, 2 for Moderate DR, 3 for Severe DR, and 4 for Proliferative DR.

5. Conclusion:

Our study demonstrates the feasibility of addressing the five-class problem for national diabetic retinopathy (DR) screening using a CNN approach. Our network exhibits promising capabilities in learning the requisite features for accurately classifying fundus images, particularly in accurately identifying proliferative cases and cases with no DR. As observed in prior studies with large datasets, achieving high specificity often comes at the expense of lower sensitivity [12]. Our method yields comparable results to previous approaches without relying on feature-specific detection and utilizing a more generalized dataset.

The potential advantage of employing our trained CNN lies in its ability to swiftly classify thousands of images per minute, enabling real-time diagnosis upon image acquisition. In practical scenarios, images are typically forwarded to clinicians for grading, delaying accurate assessment during screening appointments. The trained CNN facilitates rapid diagnosis and immediate response to patients. Notably, the network achieves these results with only one image per eye.

The network demonstrates proficiency in detecting healthy eyes, likely due to the abundance of healthy eye images in the dataset. However, during training, discerning mild, moderate, and severe DR cases posed challenges, leading to lower sensitivity in these categories. Advancements in CNNs allow for deeper networks capable of better capturing intricate features. While our network's results are promising within a conventional network topology, future iterations aim to tailor the network to specific fundus image features, such as vessels and exudates.

In conclusion, CNNs hold promise in learning to identify features of diabetic retinopathy in fundus images. As networks and datasets continue to evolve, CNNs offer potential utility to DR clinicians by providing real-time classifications.

References

1. Chazhoor, A.; Sarobin, V.R. Intelligent automation of invoice parsing using computer vision techniques. Multimedia. Tools Appl.**2022**, 81, 29383–29403.
2. Sanket, S.; Vergin Raja Sarobin, M.; Jani Anbarasi, L.; Thakor, J.; Singh, U. Narayanan, S. Detection of novel coronavirus from chest X-rays using deep convolutional neural networks. Multimedia. Tools Appl. **2022**, 81, 22263–22288.
3. Kumar, S.L. Predictive Analytics of COVID-19 Pandemic: Statistical Modelling Perspective. Walailak J. Sci. Technol. (WJST) **2021**,18, 15583.
4. Available online: {Kaggle link} (accessed on 13 June 2022).

5. Sector, S.P., et al. State of the Nation 2012. *Diabetes UK* 2013.
6. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.
7. medium.com: convolutional-neural-networks-cnn-architectures-explained.
8. Harry Pratta, Frans Coenenb, Deborah M Broadbent, Simon P Hardinga, Yalin Zhenga. Convolutional Neural Networks for Diabetic Retinopathy International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016,6-8 July 2016, Loughborough, UK
9. Fukushima, K.. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36(4):193–202.
10. Cun, Y.L., Boser, B., Denker, J.S., Howard, R.E., Hubbard, W., Jackel, L.D., et al. Advances in neural information processing systems 2. Citeseer. ISBN 1-55860-100-7; 1990, p. 396–404.
11. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–1958.
12. Fleming, A.D., Philip, S., Goatman, K.A., Prescott, G.J., Sharp, P.F., Olson, J.A.. The evidence for automated grading in diabetic retinopathy screening. Current Diabetes Reviews 2011;7:246 – 252.