Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

A wavelet CNN with appropriate feed-allocation and PSO optimized activations for diabetic retinopathy grading

Chandrasekaran Raja, Santhosh Krishna B V, Balaji Loganathan, Sanjay Kumar Suman, L. Bhagyalakshmi, Mubarak Alrashoud, Jayant Giri & T. Sathish

To cite this article: Chandrasekaran Raja, Santhosh Krishna B V, Balaji Loganathan, Sanjay Kumar Suman, L. Bhagyalakshmi, Mubarak Alrashoud, Jayant Giri & T. Sathish (2024) A wavelet CNN with appropriate feed-allocation and PSO optimized activations for diabetic retinopathy grading, Automatika, 65:4, 1593-1605, DOI: 10.1080/00051144.2024.2409552

To link to this article: https://doi.org/10.1080/00051144.2024.2409552

Published online: 15 Oct 2024.

AUTOMATIKA
2024, VOL. 65, NO. 4, 1593–1605
https://doi.org/10.1080/00051144.2024.2409552

A wavelet CNN with appropriate feed-allocation and PSO optimized activations for diabetic retinopathy grading

Chandrasekaran Raja[a], Santhosh Krishna B V [b], Balaji Loganathan[a], Sanjay Kumar Suman[c], L. Bhagyalakshmi[d], Mubarak Alrashoud[e], Jayant Giri [f,g] and T. Sathish[h]

[a]Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India; [b]Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, India; [c]Department of ECE, St. Martin's Engineering College, Secunderabad, India; [d]Department of ECE, Rajalakshmi Engineering College, Chennai, India; [e]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia; [f]Department of Mechanical Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, India; [g]Division of Research and Development, Lovely Professional University, Phagwara, India; [h]Saveetha School of Engineering, SIMATS, Chennai, India

ABSTRACT

This work modifies the architecture of conventional CNN with the integration of Multi-resolution Analysis (MRA) in a CNN framework for Diabetic Retinopathy (DR) diagnosis and grading. Here, the HF sub-bands are subjected to optimized activations and are directly fed to the fully connected layers, as it encompasses edge features. Unlike FD-Relu, the proposed function preserves significant negative coefficients, compared to the S-Relu, the proposed third-order S-Relu is optimized such that it sustains the activations in the range suitable for the wavelet coefficients. The coefficients of higher-order terms of the proposed 3rd-order S-Relu are optimized with PSO, fitting the maximum energy of the wavelet sub-bands to ensure High Frequency (HF) edge preservation. The authors re-architecture 3 different CNNs published in the Retinal Image analysis field, with spatial and wavelet inputs with optimized activations. The highest accuracy of 96% is attained with the AlexNet re-architecture, with 35,126 fundus images secured from the Kaggle dataset. As we can infer the proposed re-architecture wavelet CNN outperformed the multiscale shallow CNNs, multiscale attention net, and stacked CNNs with a 6.6, 0.3, 0.7 per cent increase in accuracy. The entire implementation of the wavelet CNN is made available under source code.

ARTICLE HISTORY
Received 6 February 2023
Accepted 23 September 2024

KEYWORDS
WaveletCNN; activation function; ResNet; AlexNet; wavelet; diabetic retinopathy

1. Introduction

DR is a common chronic ailment that is diagnosed in patients with diabetes across the globe. The World Health Organization (WHO) has estimated that, in 2030, diabetes will be the world's most extreme and 7th highest death-causing disease [1]. The number of individuals diagnosed with diabetes has risen significantly over the past few decades and diabetes increases the risk of a certain type of eye disorder, of which DR is among the most severe causes of blindness in the middle-aged population [2]. In this sense, preventing human life from being affected by diabetes is quite important. Some anomalies, including lesions, are produced in the retina in the case of DR, which eventually results in non-reversible blindness and vision loss. However, the early identification and treatment of these lesions will potentially reduce blindness [3]. Early identification, which is vital to successful diagnosis, depends on professionals and it is a time-consuming process. This becomes a challenging task in areas where skilled assistants are lacking. Additionally, the manual complexity of DR screening methods creates widespread discord among the readers. Eventually, given the growing prevalence of both diabetes and related retinal disorders worldwide, manual diagnostic methods may not be able to keep pace with the demand for screening facilities [4]. Computerization is a prerequisite for improving the efficiency, accessibility and affordability of health care.

During the past two decades, several teams have researched the automated analysis of retinal colour images for DR [5–7]. DR primarily affects the blood vessels in the light-sensitive tissue (retina).

Numerous earlier pieces of research concentrated on the automatic detection of DR from fundus images. Convolutional Neural Networks (CNNs) have become highly prevalent in a variety of situations, owing to the accessibility of annotated data and the advancement of GPUs. CNNs have always shown excellent results in the exploration of complex structures in high-dimensional datasets, through their multiple-layer features [8]. The models have been effective in studying

CONTACT Santhosh Krishna B V santhoshkrishna1987@gmail.com Department of CSE, New Horizon College of Engineering, Bengaluru, Karnataka, India

the most discriminatory and often abstract elements of the image [9–11]. It is predicted that supplementing with wavelet domain inputs shall enhance the accuracy of diagnosis by diversity of features. In this work the authors attempt to re-architecture conventional CNN with multi-resolution analysis, accepting wavelet domain inputs. The new activation function is proposed for the wavelet domain inputs and is optimized. The frequency-domain Relu [12] and the spectral Reluproposed [13] are not perfect to be employed for the wavelet domain inputs.

Hence a 3rd-order spectral Relu is defined and the coefficients of higher-order terms are optimized to fit the maximum energy of the wavelet sub-bands (Horizontal (H), Vertical (V), Diagonal (D)) as in the High Frequency/Wavelet sub-bands, the edges contribute to most of the energy. A stopping criterion is imposed to restrict the energy of the activation within a certain per cent of the energy of the wavelet sub-bands before activation. Else unbounded increase of energy leads to over brightening i.e. unequalized histogram.

2. Literature review

Mookiah et al. [10] worked out discrete wavelet transformation and stationary wavelet transform functions to extract the features from the fundus images, and the optimal features were chosen to differentiate normal from DR using top-rank and AdaBoost machine learning methods. More than 90% accuracy has been obtained by classifying extreme cases of pathological signs in the eye. Optimal Hyper Analytic (OHA) Wavelet Transform has given high accuracy, specificity, and sensitivity values from retinal images [14]. Akram et al. [11] implemented a hybrid classifier that combines a Gaussian mixture model to distinguish dark and bright regions for the identification of DR-related lesions. The work [15] aptly points out the evolution of Medical Image Analysis in the Big data domain. In the literature, different forms of deep learning models are used to study applications related to images [16]. Such algorithms are Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), Deep Autoencoder (DAE), Recurring Neural Networks (RNN), Deep Belief Networks (DBN), Deep Convolutional Extreme Machine Learning (DCEML), etc., CNN received a lot of attention for work related to image processing and was addressed by the researchers. The various types of CNN architectures such as Alexnet [17], LeNet [18], Faster R-CNN [19], GoogLeNet [20], ResNet [21] etc., are found in literature. Zeng et al. [22] used a Siamese-like CNN based on binocular retinal fundus images trained for automatic DR detection with a transfer learning technique that showed high efficiency. Chandrakumar et al. [23] suggested a deep learning architecture used in retinal images to identify DR. The CNN algorithm was considered in this approach to characterize the retinal image based on convolved features and for the classification of retinal images, the authors achieved an accuracy of around 94–96%. Lam et al. [24] suggested a deep learning framework to diagnose DR. They utilized CNN architecture to discern stages of DR in the image. With this method, they obtained a classification performance of about 95%. Giancardo [25] suggested a system that depends on the characteristics of colour and wavelet decomposition from exudates candidates for training the classifiers. Utilizing a support vector machine (SVM) classifier with areas under receiver operating characteristics (AUC) around 0.88 and 0.94 the authors yielded the best performance. Theoretical characteristics of conventional methodologies to signal/image processing, such as multi-resolution analyzes using wavelets, are well studied, rendering these methods more comprehensible. Indeed, there are some previous works integrating wavelet representations into CNNs. The researchers in [26] constructed an Xception design that fuses features from various convolution layers. Through fusion, they demonstrated a significant improvement in accuracy of around 83%. The research work [27] employs powerful feature selection methods, max-relevance, and min redundancy in type 2 diabetes detection.

Oylallon et al. [28] have suggested a hybrid framework that would substitute ResNet's initial layers with a wavelet scatter network. This improved ResNet resulted in an output comparable to that of the original ResNet but has fewer trainable parameters.

In [29] AlexNet is modified to diagnose DR, where the green channel is alone processed. It consists of 8 layers (5 Convolution layers and 3 dense layers). The AlexNetin [26] is modified to accommodate an input size of 259×259. To detect DR, deep layer aggregation (combines feature maps from different CNN architectures) is done. The CNN architectures include the Inception module, residual blocks, and separable convolutions to attain an accuracy of 83 per cent in DR diagnosis. In [30] multi-scale shallow CNNs are used as base learners, each one with its receptive field. The multiple learned parameters are integrated into a common dense layer by a novel voting scheme.

In [31] a fusion of the architectures VGG, Spatial Pyramid Pooling (SPP), Network-in-network (NiN) is being used in which the VGG 16 encompasses 16 convolution layers, followed by SPP which pools the feature maps to adapt to the size of the dense layers of VGG 16. And NiN architecture uses Multi-Layer Perceptron (MLP) convolutions. In [32] multi-scale attention net is used resizing the images to size 512×512 to suit the ResNet encoder. The feature maps before and later the encoder is combined by a dedicated scaling mechanism, reaching an accuracy of 98.3%. In [33] several CNNs are stacked to diagnose DR from fundus images in the

Kaggle dataset, and the weights of the 3 custom CNN models are shared producing an accuracy of 97.92% in DR diagnosis. M. Khanna et al., in [34] shortlisted four datasets and 20 CNN-based models and tested and validated the best ones using 16 detailed experiments with fivefold cross-validation. The two best pre-trained transfer learning-based detection models can contribute clinically by offering accurate and rapid patient prediction. The authors of [35] explore the application of deep learning, specifically convolutional neural networks, for the automated prediction and grading of diabetic retinopathy using retinal fundus images. They highlight the limitations of traditional manual diagnosis, which is time-consuming, expensive and prone to human error. The study investigates the performance of various pre-trained CNN architectures, including VGG, ResNet, Inception and MobileNet, for DR grading. They employ transfer learning, fine-tuning these models on a real-world DR screening dataset. The researchers in [36] evaluated five improved deep learning architectures: U-Net, DenseU-Net, LadderNet, R2U-Net and ATTU-Net. They also introduced an enhanced customized R2-ATT U-Net model for retinopathy screening. In relevant domains with similar feature extraction and classification [37] specific DCNN architectures are proposed. The research work [38] implies ROI segmentation with a Gaussian mixture model, CNN-based DR diagnoses Alex Net and SVM.

There are shreds of evidence of combining the **wavelets with CNNin fields *other than* Retinal Image analysis** for taking the advantage of Spatial-frequency analysis.

Fujieda et al. [39] introduced wavelet CNNs (WCNNs), which were focused on the similarities between multi-resolution processing and CNN's convolution filtering and pooling operations. They suggested a Dense Net-like CNN, but the Haar wavelets (usually used in multi-resolution analysis) have been used as layers of convolution and pooling. Such layers of wavelets concatenated with the feature maps provided by the subsequent convolution blocks. This framework was most interpretable than CNNs because the input wavelet transform is generated by the wavelet layers.

All the kinds of literature discussed above extract wavelet-based features or of late use CNN to detect DR, and very few attempts are made to implement wavelet-based CNN. In the proposed design, the original image and the LF wavelet sub-band are processed in the CNN format, i.e. the convolution layers operate on them. Whereas the HF wavelet sub-bands are directly fed to the fully connected layer, as it encompasses edge features. Thus, the major contribution of the work is the integration of the Multi-Resolution Analysis in a CNN framework, with appropriate feed-allocation and optimized activation function.

In most of the works, conventional activation functions like Relu are used. It is unsaturated but not differentiable at all points. In this work, the authors carefully examine the performance of the activation function tailored for the wavelet domain and then tune the function for the range of wavelet coefficients in DIP.

The frequency-domain Relu in [12] nullifies the negative coefficient terms which correspond to significant edge feature maps in an image. Hence the authors propose an S-Relu of third-order, as the activation function weights the negative coefficients proportional to their magnitude. This is highly desired as in the detailed co-efficient sub-bands the co-efficient with high magnitude corresponds to a significant "edge" feature map and the "sign" indicates the direction of the kernel.

The spectral Relu proposed in [13] is a second-order activation function, it transforms a certain range of negative coefficients to positive but the output range of the activation is not suitable. The spectral Relu of third-order proposed in this work activates the inputs within the range that is suitable for the wavelet domain in Digital Image Processing, by constrained Particle Swarm Optimization (PSO) for an output range of [0,400]. And we propose a novel fitness function for the H, V and D sub-bands intending to increase the energy of the sub-bands after activation. The energy increase is checked with stopping criteria to restrict the energy after activation to be maintained within a certain per cent of the energy of the wavelet sub-bands to ensure histogram-equalized activations. The optimization of the co-efficient is executed offline and out of the CNN, the framework to limit the computations. In [40] a novel dual-branch transfer learning approach for the detection and grading of various stages of diabetic retinopathy using a single retinal fundus photograph. Our model utilizes two state-of-the-art CNN architectures, EfficientNetB0 and ResNet50, which have been fine-tuned through transfer learning techniques on a large multi-center dataset.

2.1. Dataset and front end

The images are extracted from the Kaggle repository [41]. In the proposed work, we have experimented with 35,126 images with 80% of the images used for training. 25,810 images are taken as Class 1 with No DR, 2443 images are taken for Class 2 – Mild DR, 5292 images are taken for Class 3 – Moderate DR, 873 images are taken for Class 4 – Severe DR, 700 images are taken for Class 5 Proliferative DR. Like some recent research works [42], the implementation is done with cooperative i-Python environment called the Google Co-laboratory. The Computer Vision Library – "Open CV" and the machine learning libraries "Keras", and "Tensor Flow" which are supported by Google Co-laboratory, are used to train CNNs with GPUs in the Google Cloud.

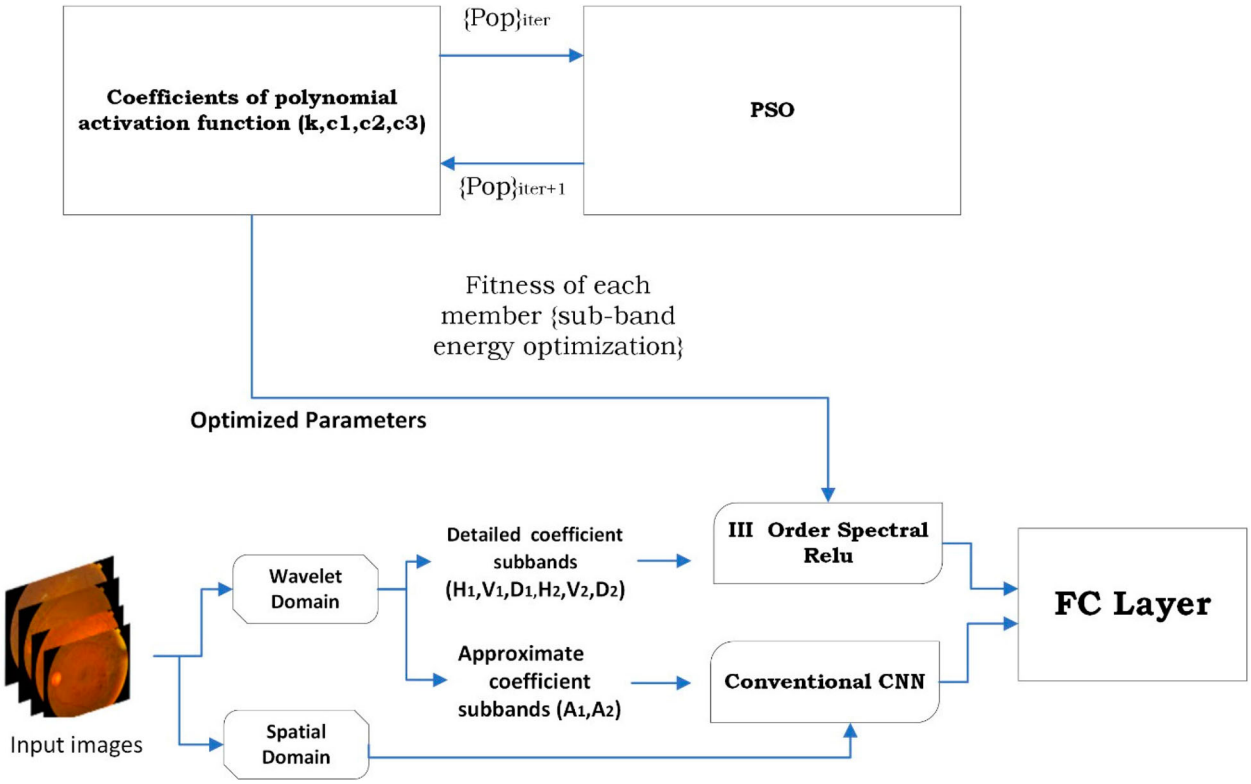

Figure 1. Overall process diagram of wavelet CNN and Optimization of activation functions.

2.2. Wavelet CNN with PSO optimized activation function

Initially the images are subjected to wavelet transform. And in addition to the image dataset, we append an approximate sub-band dataset and detailed sub-band dataset as shown in Figure 1 and in pseudo-code in Figure 2. The pseudo-code in Figure 2 shows that these 2 datasets are appended such that each image has 3 components (the image (I), approximate coefficient sub-band Φ(I), detailed sub-band ψ(I)). And as detailed in Figure 1, the approximate coefficient sub-band corresponding to level 1, 2 are A_1, $A_2$and the detailed coefficient sub-bands are H_1, V_1, D_1 and H_2, V_2, D_2 The image (I) in the spatial domain is fed to a conventional CNN. In subsequent levels of wavelet transformation, the Low Pass (Scaling) coefficients Φ(I) are also fed to the conventional CNN.

The concern for feeding approximate co-efficient sub-band (LP outputs) also to the CNN is that the inputs to the CNN are made available also in the wavelet domain, enabling multi-resolution analysis in the CNN framework.

The HF wavelet domain inputs or detailed co-efficient sub-bands ψ(I) are subjected to the activation with III order spectral Relu and then **fed directly to** Fully Connected Layer. The appropriation of the image I, approximate coefficient sub-band Φ(I), detailed sub-band ψ(I) can be inferred from the process diagram in Figure 1 and the pseudo-code in Figure 2.

The coefficients of III order activation function is optimized by PSO. PSO is initialized offline to the CNN

Algorithm: Wavelet CNN

Input d : image dataset (I images), l : labels
$d_{train}, d_{train} \in d$
for each image I in dataset do
Obtain Scaling function/LF sub-bands, $\Phi(i)$
Obtain Wavelet function/HF sub-bands, $\Psi(i)$
End for

Append Inputs Approx sub-band dataset,
$d_{approx} \leftarrow \Phi(d)$
Detailed sub-band dataset, $d_{det} \leftarrow \Psi(d)$
append d $\leftarrow d_{approx} d_{det}$
$d_{train}, d_{test} \leftarrow$ Split training and testing data

for each Image I in dataset do
I; $\Phi(I)$; $\Psi(I) \leftarrow$ Segregate feature maps/Inputs
for each sub-band $\Psi(I)$ (Offline)
Optimized parameters
$\{K, C_1, C_2, C_3\} \leftarrow$ **Optimize** III order activation function
Formulate III order activation function
End for
Fully connected Layer $\leftarrow$ Other layers of CNN $\leftarrow$ Conventional activation function $\leftarrow$ I, $\Phi(I)$
Fully connected Layer $\leftarrow$ **Optimized** III order activation function $\leftarrow \Psi(I)$
End for

Figure 2. Pseudo-code of the wavelet CNN and Optimization of activation functions.

framework as shown in Figure 1, all members with four dimensions K, C_1, C_2, C_3 where C_1, C_2, C_3 are the coefficients of the spectral terms of different order and K refers to the coefficient of the activation function. These

Table 1. Sizes of Tensors received by a Fully Connected layer for a single image.

	Size of the Input	Tensors received by a Fully Connected layer for a single image	Layers crossed by the Tensor before the Fully connected layer	Size and Channels
FC or Side-Output Layer 1	400 X 400	Original image in level-0 after Drop out	A conventional CNN – i.e. Convolution Layers, other specific layers of a CNN, Max-pool, Drop-out	98 × 98 (32 Convolution channels)
		Level 1 (Resolution 1 of wavelet transform)		
FC or Side-Output Layer 2	200 X 200	Approximate coefficient sub-band A_1 or Φ	A conventional CNN – i.e. Convolution Layers, other specific layers of a CNN, Max-pool, Drop-out	48 × 48 (32 Convolution channels)
FC or Side-Output Layer 2	200 X 200	Detailed coefficient sub-band H_1, V_1, D_1 or ψ without any drop-outs	Custom Activations – III order spectral activation function	200 × 200 (3 channels – H_1, V_1, D_1)
		Level 2 (Resolution 2 of wavelet transform)		
FC or Side-Output Layer 3	100 X 100	Approximate coefficient sub-band A_2 or Φ	A conventional CNN – i.e. Convolution Layers, other specific layers of a CNN, Max-pool, Drop-out	24 × 24 (32 Convolution channels)
FC or Side-Output Layer 3	100 X 100	Detailed coefficient sub-band H_2, V_2, D_2 or ψ without any drop-outs	Custom Activations – III order spectral activation function	100 × 100 (3 channels H_2, V_2, D_2)

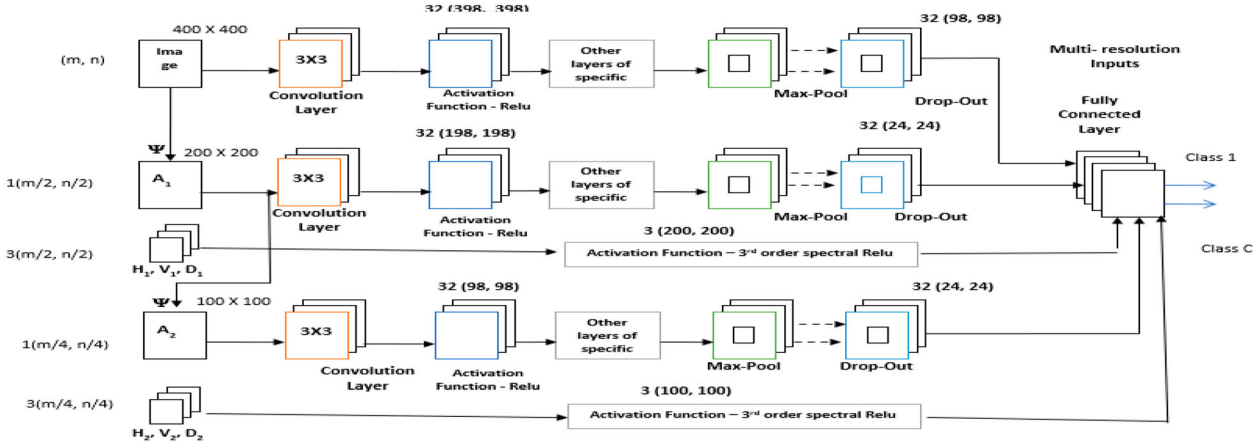

Figure 3. The proposed Wavelet CNN X (m, n) – refers to X Channels of size (m, n).

parameters are optimized fitting the higher energy of the sub-bands; edges contribute to most of the energy in detailed coefficient sub-bands. As shown in the overall process diagram in Figure 1, the optimized coefficients K, C_1, C_2, C_3 is applied to the III order spectral activation function, which is applied to the detailed coefficient sub-bands/wavelet function ψ(I). This can be inferred from the pseudo-code in Figure 2. Since there is a re-architecture the feed allocation is detailed in Table 1, with details of the tensors of the images and sub-bands.

2.2.1. The wavelet CNN

In Figure 3 the symbol Ψ represents the 2-D Wavelet transform on the Image. The "db3" wavelet is used in this work because of the proven significance of the "db3" Wavelet function in retinal image analysis [43]. Several researchers like Dua et al. have established the significance of Daubechies-3 wavelet for retinal image analysis [44].

All the above works of literature put forth that, the CNNs that involve only spatial domain filtering, the **inclusion of spectral inputs shall enhance the feature extraction process.** And as the wavelet transform decomposes an image into Low Frequency and High-Frequency sub-bands and much of the High-frequency features are extracted in the Convolution layer of the CNN, the authors do not subject the High-Frequency sub-bands to the CNN. They are fed directly to the Fully Connected Layer corresponding to the particular resolution (as shown by the straight-line connections between the HVDs and the Fully Connected Layers in Figure 3).

Also this can be inferred from Table 1 that the High-Frequency sub-bands/wavelet sub-bands, after custom activation function are directly fed to the fully connected layer.

The original image and the 2 levels of wavelet transform outputs' approximate co-efficient sub-band (Φ or A-LP outputs) are subjected to the entire CNN.

In the model shown in Figure 3, in the 1st branch, the image in the spatial domain is fed to a conventional CNN. In other words level-0 resembles a convention CNN by itself.

In subsequent levels, the Low Pass (Scaling) coefficients of each level of the transform are fed to the CNN input (as shown in the 2nd and 3rd main branches of the model). The concern for feeding approximate co-efficient sub-band (LP outputs) also to the CNN is that the inputs to the CNN are made available also in the wavelet domain, enabling multi-resolution analysis in the CNN framework. The approximate coefficient sub-bands or scaling co-efficient A_1 and A_2 correspond to levels 1 and 2 in Figure 3. The multi-resolution analysis can be inferred that the original image size (m, n) is halved at each level to (m/2, n/2) for level 1; (m/4, n/4) for level 2.

The HF wavelet domain inputs or detailed co-efficient sub-bands H_1, V_1, D_1 (in level 1) & H_2, V_2, D_2 (in level 2) in Figure 3 are **fed directly to** Fully Connected Layer after (offline) optimized activation as **additional Edge textures in the Wavelet Domain**.

The above implementation is carried out with the ability of the Keras tool to receive multiple inputs of different sizes. Table 1 shows concisely the different inputs and their sizes before the Fully Connected Layer. There are 3 Fully Connected Layers, and they are called side output layers as the Loss function is estimated for all 5 classes. For instance for a single image of size 400 × 400, from the input side, the tensors received have 5 different sizes at the Fully Connected Layers.

The mentioned sizes of different tensors are summarized in Table 1.

2.3. The CNN

Filtering is executed in the Convolution Layer over the input images. These filters extract significant edge information from the image. The CNN model in Figure 3 uses 32 filters, of size 3 × 3 with stride 1. The purpose of Rectified Linear Unit RELU is to transform the features in a linear space into a non-linear space.

In the drop-out layer, 30% of the neurones are dropped out to avoid model-over-fitting. The dimension of the Feature Maps has to be reduced for minimizing the computations. Generally, an even-sized window is moved over the image with stride 2. At each tile, the highest value is retained. The Fully Connected layer is usually the terminal layer which resembles the back-propagation stage of an ANN. The drop-out layers are the side output layers and the Loss function is estimated. In this work, to the input of the Fully connected layer, we have the tensors related to the original image, 2 levels of Approximate co-efficient sub-band (A-LP outputs, with convolution filtered and activation function applied), 2 levels of Detailed co-efficient sub-band (H, V, D-HP outputs).

The conventional AlexNet is capable of receiving input images of size 224 X 224 X 3, it then extracts features with convolution layers with filter sizes 11 X 11, 5 X 5, 3 sets of layers with filter sizes 3 X 3, each followed by Max-pooling. In the proposed wavelet CNN, as we have different sizes – 400 X 400 (original image), 200 X 200, 100 X 100 (Wavelet domain inputs) we implement the conventional Alex Net thrice (once for the image in the spatial domain and twice for the wavelet domain inputs). i.e. for instance, the image of size 400 X 400 is fed to the set of convolution layers with filter sizes 11 X 11, 5 X 5, 3 sets of layers with filter sizes 3 X 3, each followed by Max-pooling by a custom implementation. And the conventional Alex Net encompasses 3 dense layers with 4096, 4096 and 1000 neurones. We tailored the last dense layer with 2 neurones for the 2-class classification problem.

Also, the authors carefully examine the performance of the activation function tailored for the frequency domain and then tune the function for the wavelet domain. Here, the authors focus on the fact that the "edge" feature maps constitute most of the energy in the detail coefficient sub-bands as the edges mainly contribute to the discrimination and optimize the corresponding activations.

The Frequency domain Relu [12], shown in Figure 4a, fails for detailed wavelet sub-bands as it nullifies the negative coefficients,

$$Y(\omega) = \frac{1}{2}[X(\omega) + X||(\omega)||] \tag{1}$$

The Frequency domain activation function in equation (1) normalizes negative coefficients in the design of frequency domain CNN but does not work well for images in the wavelet domain. The nullification of negative coefficients as in FD Relu leads to a major loss of significant features.

2.4. Optimization of the co-efficient of the polynomial activation function

The authors in [46] and [47] propose activation functions of first order with learnable slope parameters. Polynomial functions of higher order are also researched and they exhibit curvature. In this work, the authors' trials with most of the II order activation functions (with different coefficients) lead to a wider range of outputs compared to the inputs.

1. For instance, the Spectral domain Relu (S-Relu) given by equation (2) [13] fits the frequency domain data.

$$Y(\omega) = 0.021X^2(\omega) + 0.3X(\omega) \tag{2}$$

It is observed that the output of the activation, shown in Figure 4b, is bounded in the same range for an input range [−15, 15] as claimed in the work [13]. But it is not perfectly tailored for the wavelet domain as the activation stretch-out the range for inputs in the range [−200, 200] as shown in Figure 4b.

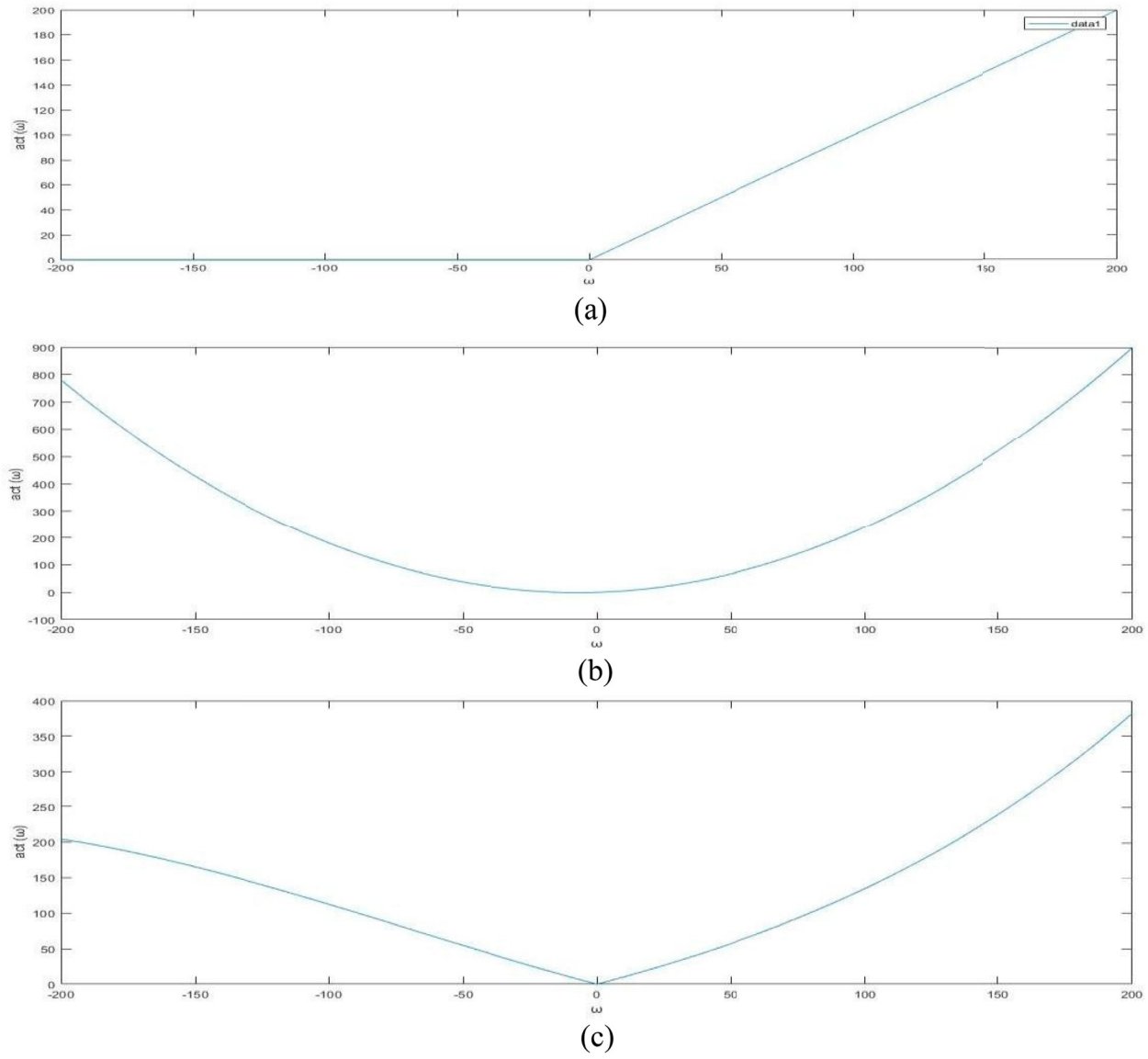

Figure 4. (a). Frequency Domain Relu (proposed in [45]). (b) Spectral Relu (Second order proposed in [12]). (c). Coefficient Optimized Spectral Relu (Third-order).

2. A parabolic activation function is also implemented to achieve linear decrease and linear increase for negative and positive coefficient inputs respectively, but the metrics are not successful.

3. The authors optimize the third-order functions as shown in Figure 4c, and the shape of the activation function. The optimization is subject to the **constraint** that the output shall be in the range [0,400] for the input range of[−200, 200] of H, V and D wavelet sub-bands.

The third-order activation function for the wavelet domain, as given by equation (3), shown in Figure 4c, fitting the conditions, transform all negative coefficients to positive and within a range that fits the wavelet domain inputs.

$$Y(\omega) = \frac{1}{K} C_1 X^3(\omega) + C_2 X^2(\omega) + C_3 X(\omega) \quad (3)$$

For the proposed activation function, shown in equation (3), the authors propose to optimize the constants K, C_1, C_2, C_3. The coefficients are optimized to fit the maximum energy of the wavelet sub-bands (Horizontal (H), Vertical (V), Diagonal (D)) as in the High Frequency/Wavelet sub-bands, the edges contribute to most of the energy. A stopping criterion is imposed to restrict the energy of the activation within a certain per cent of the energy of the input wavelet sub-bands before activation. Else unbounded increase of energy leads to over brightening i.e. unequalized histogram.

Another concern about the polynomial activations is that the higher value (near unity) for the coefficients of higher-order terms leads the weights to grow extremely [48] and cause instability. Hence the allowable range of the coefficient in equation (3) is taken as K – [5,10], C_1 – [0, 0.05], C_2 – [0, 0.05], C_3 – [0, 1]. As discussed in [48]

C_1 and C_2 are chosen less with negligible magnitude concerning unity.

In the initial step of optimization, PSO is initialized offline to the CNN framework as shown in Figure 1, with 30 members (Population size), each with four dimensions K, C_1, C_2, C_3 corresponding to equation (4).

The position and velocity are updated in PSO is executed as shown in equations (8) and (9),

$$V_{L,M}^{t} = WV_{L,M}^{t-1} + C_1 r_1 (bp(x)_L^t - p_{L,M}^t) + C_2 r_2 (bp(x)_L^t - p_{L,M}^t) \quad (4)$$

$$p_{L,M}^{t+1} = p_{L,M}^t + V_{L,M}^t \quad (5)$$

In equations (4) and (5), the terms "V" and "P" represent the velocity and position respectively; M and L represent the member and dimension respectively; "t" represents the iteration number, and "W" denotes the inertia weight that controls the impact of the previous velocity on the current velocity.

The fitness of the memth member in the iterth iteration is calculated as shown in equation (6) and the energy optimization is taken as a constraint for optimization of the parameters as shown in Figure 1.

$$Fit_{mem}^{iter} = Energy^{iter}(mem) \quad (6)$$

Where Energy represents the Energy of a sub-band.

The PSO entities and parameters are consolidated as follows:

Input – H, V, D sub-bands; Objective function – To increase the Energy of activation to the sub-bands (H, V, D); Population – Constants and co-efficient terms of the activation function – K, C_1, C_2, C_3; Dimension – 4; Allowable range – K – [5,10], C_1 – [0, 0.05], C_2 – [0, 0.05], C_3 – [0, 1]; Population size – 30; Constraint – the output of the activation function is constrained to the range [0,400]; Stopping Criteria – η% of the energy of the sub-band.

The optimized activation function exhibits a linearly decreasing relationship for the negative coefficients, as shown in Figure 4c. That is the activation function gives more weight to the negative coefficients with high magnitude. This is highly desired as in the detailed co-efficient wavelet sub-bands the **negative co-efficient with high magnitude also corresponds to a significant "edge" feature map** and the "sign" is typically the direction of the kernel. And the function is ensured to produce the best activation maps by visual examination as shown in Figure 6 in Section 4.

As shown in the overall process diagram in Figure 1, the optimized coefficients K, C_1, C_2, C_3 is applied to the III order spectral activation function, which is applied to the detailed co-efficient sub-bands/wavelet function, i.e. the III order spectral activation function is optimized for the detailed co-efficient sub-bands/wavelet sub-bands. This can be inferred from Table 1 and it is further inferred that the approximate co-efficient sub-band/scaling function is fed to a conventional CNN with conventional activation function.

Then, in the fully connected layer, each entity has its own significance and it shall be appropriate to let the network optimize the significance among the inputs fitting to the error minimization. The significance of the individual inputs to the fully connected layer is formulated to be optimized based on their Weight fitting to improve the overall accuracy, inspired by [45].

That is, a Loss value is evaluated in the forward pass but they work together for weight-up gradation in the backpropagation. The loss value is initially evaluated individually for all the layers (3 drop-outs and 2 detailed sub-bands).

The Loss function for the nth side-output node (Drop-out) is initially evaluated with the fully connected layer as per equation (7)

$$l_{side}^{n}(w^n) = -\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{C} Y_{ji} log(P_{ji}), W^n \quad (7)$$

Y_{ji} represents the probability of correspondence of label "i" to sample "j" in the ground truth. P_{ji} represents the probability of predicting sample "j" as belonging to label "i". w^n represents the Set of weights of the nth side-node.

Then there is a weight fusion step in which the Loss function for all the side-output nodes are evaluated with the fully connected layer as per equation (8) in [45], with the fusion of weights, i.e. the weights corresponding to all the side output nodes are updated.

$$l_{fuse}^{n}(w^n, w^{fuse}) = -\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{C} Y_{ji} log(P_{ji}), W^n, W^{fuse} \quad (8)$$

Then the weight updating is executed subject to the constraint in equation (13). Hence the significance of the individual inputs to the fully connected layer is formulated to be optimized as β_n in equation (9) based on their Weights fitting to improve the overall accuracy as proposed in [45].

$$Obj(W) = argmin\left(\sum_{n=1}^{C} \beta_n l_{side}^{n}(W, w^n)\right) \quad (9)$$

As there are 5 side-nodes **(n = 1–5), corresponding to** the Original Image (in the spatial domain), resolution 1 and resolution 2 inputs respectively $\boldsymbol{\beta_1}$ **to** $\boldsymbol{\beta_5}$would represent their significance in order.

3. Result and discussion

The multi-resolution analysis is experienced with "db3" mother wavelets as the wavelet has been significant in Retinal Image Analysis [44]. In Table 2, we compare

Table 2. Comparison of the metrics of the CNNs recently implemented by other researchers in the Retinal Image Analysis field.

Objective & Reference	Dataset & Number of Images	Methods	Accuracy
[29] To detect DR (2019)	Messidor – 1130 colour fundus images	Green Channel input is applied to AlexNet for DR. The AlexNet is modified to accommodate input size. It consists of 8 convolution layers and 3 dense layers.	96.6%
[30] To detect DR (2020)	Thirty Five thousand and One Hundred and Twenty Six images of 5 classes from the Kaggle dataset	Features extracted from multi-scale base CNNs are integrated into a dense layer. The base CNs are custom-defined and shallow and the integration is based on a novel voting scheme.	92%
[31] To detect DR (2021)	Thirty Five thousand and One Hundred and Twenty Six images of 5 classes from the Kaggle dataset	Multi-scale attention Nets (ResNet + 2 residual blocks fed to Atrous convolution)	98.3%
[33] To detect DR (2021)	Thirty Five thousand and One Hundred and Twenty Six images of 5 classes from the Kaggle dataset	Stacked CNNs (3 custom models) with weighed-majority-based voting for final classification. Grey-Wolf Optimization-based luminosity normalization is done.	97.9%
[49] To detect DR (2022)	2693 images – WF-OCTA	Multi-branch CNN algorithm to analyze WF-OCTA images for staging DR.	96.11%
[50] To detect DR (2022)	3622-ISBI 2018 IDRiD	InceptionV3-VGG-16 based CNN models Adaptive active contour-based blood vessel segmentation and MDNN-based grading.	95.27%
[40] To detect DR (2024)	Messidor, IDRiD, EyePACS and APTOS 2019	Optimized CNN FNU-GOA-MDNN s transfer learning, employing two state-of-the-art pre-trained models as feature extractors and fine-tuning them on a new dataset	98.50%
[51] To detect DR (2022)	KAGGLE-3662	Deep Learning (DL) and transfer learning algorithms	96.2%
Proposed Wavelet Method			98.6%

Table 3. Metrics of the recent CNN implementations in the Retinal Image Analysis field (**with regular inputs & multi-resolution inputs).**

Reference	CNN Model	CNN as explained in reference and implemented	No of images	Metrics of the existing CNN	Metrics of re-architecture Wavelet CNN
[52]	A Custom CNN	Convolution Layer with 32 filters, ReLu Activation function (2 sets), Max-pool, drop-out, dense function, followed by another set of activation, drop-out and dense	35,126 images of 5 classes from the Kaggle dataset	Accuracy = 0.82 Sensitivity = 0.86 Specificity = 0.77	Accuracy = 0.88 Sensitivity = 0.85 Specificity = 0.91
[53]	ResNet with Attention Layer	6 learned layers, 4 Convolution layers, 2 Fully Connected Layer, Softmax layer	Accuracy = 0.94	Accuracy = 0.98 Sensitivity = 0.91 Specificity = 0.95	Sensitivity = 0.93 Specificity = 0.90
[29]	AlexNet for DR	2 Convolution layers, Pool, Convolution layer, Pool, Convolution layer, Pool, Flatten, FC1, FC2, Output layer	Accuracy = 0.93	Accuracy = 0.97 Sensitivity = 0.94 Specificity = 0.91 ROC = 0.93	Sensitivity = 0.97 Specificity = 0.98 ROC = 0.96

the wavelet CNN (with appropriate feed allocation and optimized activations) with the state-of-the-art CNN models implemented for DR diagnosis with a focus on the same dataset, Kaggle. As we can infer that the proposed re-architecture wavelet CNN outperformed the Alex Net for DR [29], multiscale shallow CNNs [30], multiscale attention net [32], stacked CNNs [33] with 2, 6.6, 0.3, 0.7% increase accuracy.

In Table 3 we compare the metrics of the recent CNN implementations in the Retinal Image Analysis field with the CNNs modified for multi-resolution inputs with a 3rd-order spectral activation function.

In CNN 1 (a custom CNN [52]), the accuracy obtained is 82%, and when the wavelet input is appended by re-architecture, the accuracy is increased to 85%. In CNN 2 (ResNet with Attention Layer [53]), the accuracy obtained is 85%, and when the wavelet input is appended by re-architecture of ResNet with Attention Layer, the accuracy is increased to 89%. In CNN 3 (AlexNetfor DR [29]), the accuracy obtained is 93%, when the wavelet input is appended by the re-architecture of AlexNet, the accuracy is increased to 96%. We can infer the increase of accuracy values with re-architecture CNNs in all types of CNN.

Figure 5 shows the PSO fitness plot which is the offline (to CNN) optimization of the average energy of the 3 sub-bands (H, V, D) fitting for the maximum value. The stopping criteria are initiated for η% of the average energy of the sub-bands (H, V, D). Hence with a random choice of the population, the initial average energy of the sub-bands is 201.24 J and it increases with iterations up to 209.79 J at an early stage of 13 iterations, then the stopping criteria is imposed to ensure avoidance of histogram un-equalization of the activations as inferred subjectively from Figure 6.

Figure 6 illustrates activations offered by the S-Relu (LHS) and Optimized 3rd-order S-Relu (RHS) for the H, V and D wavelet coefficients' (Top to bottom). It can be inferred that for all the sub-bands the Optimized 3rd order S-Relu(with $C_1 = 0.0001$ and $C_2 = 0.0021$) acts as an effective activation function than the counterpart proposed for the frequency domain. The effectiveness can be inferred from the visualization of

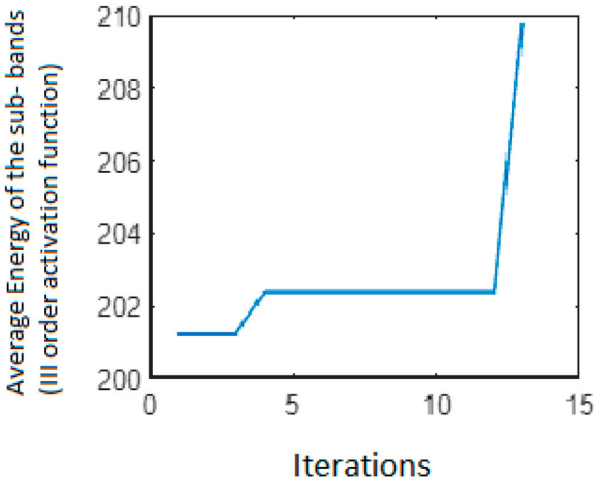

Figure 5. Plot of PSO fitness (Early stopped).

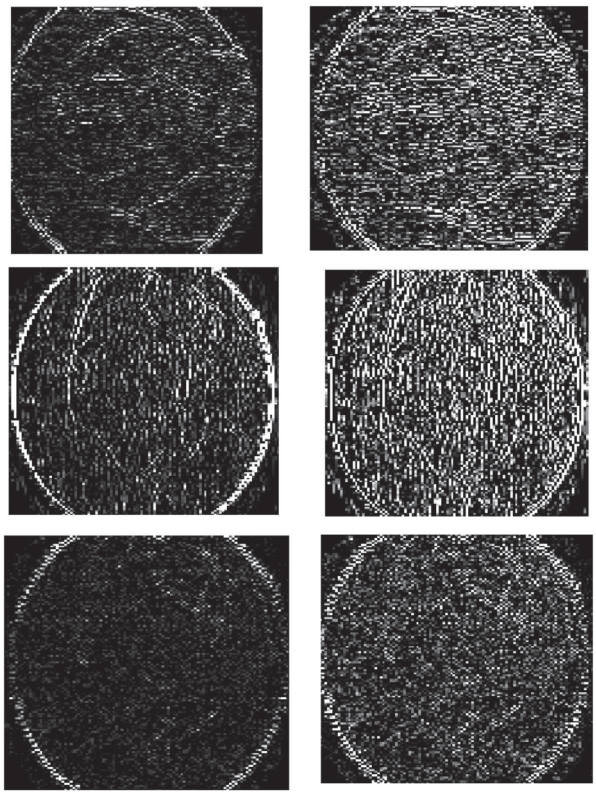

Figure 6. Illustration of Spectral Relu proposed in [12] (LHS) and 3rd order coefficient – Optimized spectral Relu (RHS) for the H, V, D wavelet coefficients (Top to bottom).

more "edge" feature maps with the tuning of the higher-order co-efficient C_1 and C_2, lower-order co-efficient C_3, and the constant K fitting the maximum energy of the corresponding sub-band.

Apart from the quantitative justification of the re-architecture Wavelet CNN in terms of the different metrics, we showcase the "learnings" of these CNN as Heat Maps by tapping their Weights. In [53] a soft attention architecture is used to forecast the heatmaps of the correspondence of the CNN training to different regions of the image. The authors [54] claim that without the heat maps of the "Weights", the CNNs are commonly perceived as black-boxes and that shortly CNN heat maps shall disclose the regions of significance for discrimination [55]. In this work, the inputs are multi-fold (1) Original Image, (2) Scaling co-efficient/Approximate sub-band (Low-Frequency), (3) Horizontal Wavelet co-efficient/Detailed sub-band (High-Frequency), (4) Vertical Wavelet co-efficient/Detailed sub-band (High-Frequency), (5) Diagonal Wavelet co-efficient/Detailed sub-band (High-Frequency). Figure 7 shows the Original Image and corresponding Heat Maps (Class 1–5 from Left to Right in all the rows). Row 1 shows the Original Images. Row 2 shows the heat Maps of the Images. Row 3 and Row 4 show the Heat Maps of Horizontal and Vertical Wavelet sub-bands. Inferring the heat maps of the images (II rows) in class 1, there is no significant Heat map outside the Disc as it is a normal image. Whereas some haemorrhages, and exudates are learned by the CNN (shown as RED patches) for class 2 and class 3(II rows, II, and III column). For classes 4 and 5, the lesions (scars) are predominantly learned in the H and V sub-band (last 2 rows, IV and V figures) compared to the heat maps of the original image (II rows, IV and V figures). The discrimination of the heat maps along each row establishes the learning of the CNN, fitting each class. In that context, the heat map discrimination between classes is dominant among Row 3 and Row 4 also (Heat Maps of Horizontal and Vertical Wavelet sub-bands). In [56] the researchers have implemented wavelet hyper-analytic activations, i.e. the wavelet domain inputs are treated with hyper-analytic activations. The authors of [56] subjected wavelet domain inputs with suitable activations but they have not implemented MRA in the framework i.e. there is only 1 level of wavelet decomposition. Whereas the proposed wavelet CNN re-architectures allocating the wavelet domain feeds of different resolution into the CNN & the detailed co-efficient wavelet sub-bands are fed directly to the fully connected layer with suitable activations (III order spectral activation function).

The proposed wavelet CNN re-architecture is unique in the way of offering MRA in CNN framework.

The wavelet CNN re-architecture implemented on the ResNet with Attention Layer produces the best classification accuracy of 98% in DR diagnosis.

Informed consent: N.A.

4. Conclusion

In this work, the authors propose a new architecture for CNN, integrating the Multi-resolution Analysis (MRA). In the proposed design, the original image and the LF (Scaling) sub-bands are processed in the CNN format. The HF (wavelet) sub-bands are directly fed to the fully connected layers as it encompasses edge features. Existing models like M-Net have successfully

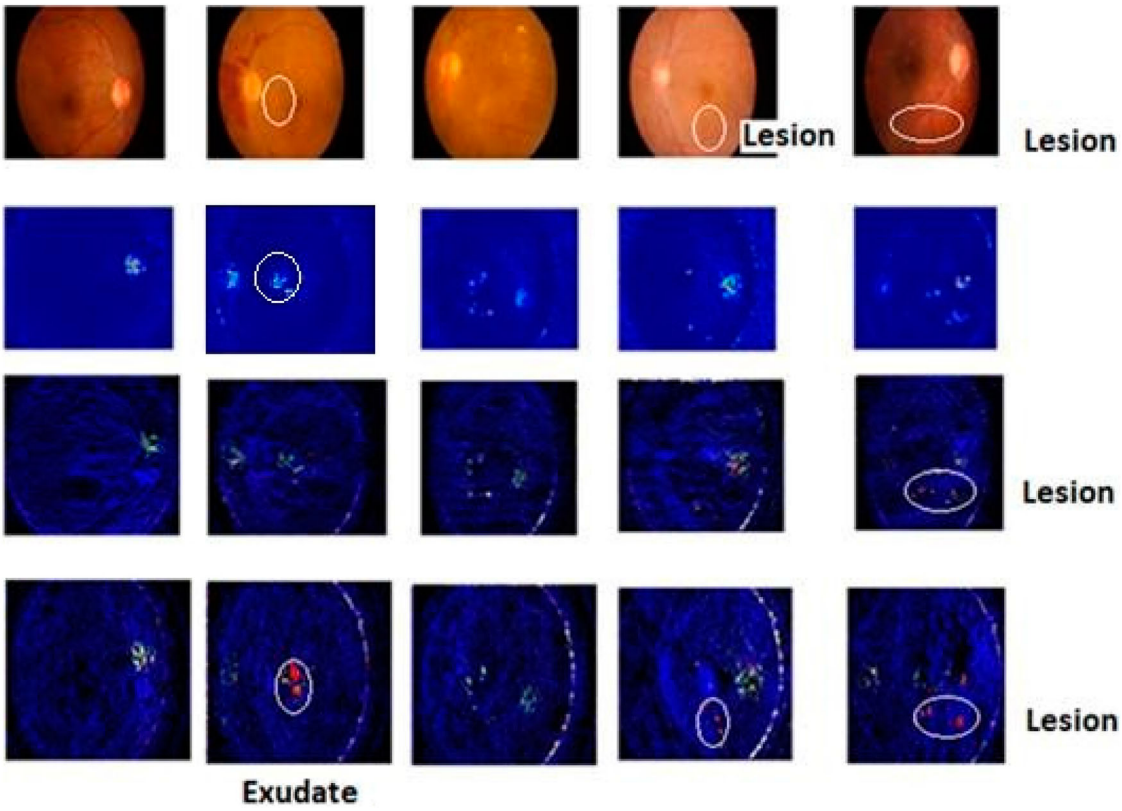

Figure 7. Original Image and corresponding Heat Maps – Class 1–5 (Left to Right) (Best Viewed in colour). Row 1 – Original ImageRow 2 – Heat Maps of Original Image. Row 3 – Heat Maps of Horizontal Wavelet Image. Row 4 – Heat Maps of Vertical Wavelet Image.

utilized wavelets in a CNN framework for Segmentation where the design is complex, and finally, the proposed 3rd-order S-Reluis optimized to preserve the maximum energy of wavelet sub-bands (and hence preserve the edge feature maps) compatible with the range of wavelet coefficients in DIP. We infer that the proposed Wavelet CNNs outperformed the conventional CNNs in terms of Accuracy, Sensitivity and Specificity. A custom CNN, ResNet and AlexNet re-architecture for wavelet inputs have shown an average 3% improvement in inaccuracy. The highest accuracy of 96% is attained with the AlexNet re-architecture. The proposed CNN outperformed the recent state of the art research with CNN re-architectures multiscale shallow CNN [30], Network in Network [31], multiscale attention net [32]. Also from the heat maps, we infer that for severe DR, the lesions (scars) are predominantly learned by the Horizontal and Vertical Wavelet sub-bands compared to the heat maps of the Image in the spatial domain, which implies the significance of multi-resolution analysis (Wavelet sub-bands) within a CNN. The novelty lies in the concern feed allocation for the spatial domain, wavelet domain inputs; optimized activations in the range suitable for wavelets in DIP.

Acknowledgment

The authors extend their thankfulness to researchers supporting project number (RSPD2024R800), King Saud University, Riyadh, Saudi Arabia.

Disclosure statement

(We) certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript "A Wavelet CNN with appropriate feed-allocation and PSO Optimized activations for Diabetic Retinopathy Grading" submitted to the Journal.

Funding

This work was supported by King Saud University [grant number RSPD2024R800].

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of data and material (data transparency)

Input Images are extracted from the Kaggle dataset [41].

Source code

For further research in this area, the Source Code is made available at https://github.com/Rajachandru/Wavelet-CNN.

ORCID

Santhosh Krishna B V http://orcid.org/0000-0002-1919-4031
Jayant Giri http://orcid.org/0000-0003-4438-2613b

References

[1] https://www.who.int/news-room/fact-sheets/detail/diabetes

[2] Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema, and related vision loss. Eye Vis. 2015;2(7).

[3] Mateen M, Wen J, Nasrullah N, et al. Exudate detection for diabetic retinopathy using pre-trained convolutional neural networks. Complexity. 2020;2020: 1–11. doi:10.1155/2020/5801870

[4] Goh JKH, Cheung CY, Sim SS, et al. Retinal imaging techniques for diabetic retinopathy screening. J Diabetes Sci Technol. 2016;10(2):282–294. doi:10.1177/1932296816629491

[5] Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Visual Sci. 2016;57(13):5200–5206. doi:10.1167/iovs.16-19964

[6] Hipwell JH, Strachan F, Olson JA, et al. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. Diabetic Med. 2000;17(8):588–594. doi:10.1046/j.1464-5491.2000.00338.x

[7] Ege BM, Hejlesen OK, Larsen OV, et al. Screening for diabetic retinopathy using computer-based image analysis and statistical classification. Comput Methods Programs Biomed. 2000;62(3):165–175. doi:10.1016/S0169-2607(00)00065-1

[8] Raja C, Balaji L. An automatic detection of blood vessel in retinal images using convolution neural network for diabetic retinopathy detection. Pattern Recognit Image Anal. 2019;29(3):533–545. doi:10.1134/S1054661819030180

[9] Pratt H, Coenen F, Broadbent DM, et al. Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci. 2016;90:200–205. doi:10.1016/j.procs.2016.07.014

[10] Mookiah MRK, Chua CK, Min LC, et al. Computer-aided diagnosis of diabetic retinopathy using multi-resolution analysis and feature ranking framework. J Med Imaging Health Inform. 2013;3(4):598–606. doi:10.1166/jmihi.2013.1210

[11] Akram MU, Khalid S, Tariq A, et al. Detection and classification of retinal lesions for grading of diabetic retinopathy. Comput Biol Med. 2014;45:161–171. doi:10.1016/j.compbiomed.2013.11.014

[12] Lin J, Ma L, Cui J. A frequency-domain convolutional neural network architecture based on the frequency-domain randomized offset rectified linear unit and frequency-domain chunk max pooling method. IEEE Access. 2020;8:98126–98155. doi:10.1109/ACCESS.2020.2996250

[13] Ayat SO, Khalil-Hani M, Ab Rahman AA-H, et al. Spectral-based convolutional neural network without multiple spatial-frequency domain switchings. Neurocomputing. 2012;364:152–167. doi:10.1016/j.neucom.2019.06.094

[14] Raja C, Gangatharan N. Optimal hyper analytic wavelet transforms for glaucoma detection in fundal retinal images. J Electr Eng Technol. 2015;10(4):1899–1909. doi:10.5370/JEET.2015.10.4.1899

[15] Mansour RF, Escorcia-Gutierrez J, Gamarra M, et al. Artificial intelligence with big data analytics-based brain intracranial hemorrhage e-diagnosis using CT images. Neural Comput Appl. 2021;35: 16037–16049.

[16] Manju K, Sabeenian RS. Robust CDR calculation for glaucoma identification. 2018.

[17] Alexnet. https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet. [accessed 2020 Feb]

[18] LeNet. http://deeplearning.net/tutorial/lenet.html. [accessed 2020 Feb]

[19] Faster R-CNN. https://github.com/rbgirshick/py-faster-rcnn. [accessed 2020 Feb]

[20] GoogLeNet. https://leonardoaraujosantos.gitbooks.io/artificialinteligence/content/googlenet.html. [accessed 2020 Feb]

[21] ResNet. https://github.com/gcr/torch-residual-networks. [accessed 2020 Feb]

[22] Zeng X, Chen H, Luo Y, et al. Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network. IEEE Access. 2019;7:30744–30753. doi:10.1109/ACCESS.2019.2903171

[23] Chandrakumar T, Kathirvel R. Classifying diabetic retinopathy using deep learning architecture. Int J Eng Res Technol. 2016;5(6):19–24.

[24] Lam C, Yi D, Guo M, et al. Automated detection of diabetic retinopathy using deep learning. AMIA Summits Transl Sci Proc. 2018;2018:147.

[25] Giancardo L, Meriaudeau F, Karnowski TP, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal. 2012;16(1):216–226. doi:10.1016/j.media.2011.07.004

[26] Kassani SH, Kassani PH, Khazaeinezhad R, et al. Diabetic retinopathy classification using a modified Xception architecture. In: 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE; 2019. p. 1–6. doi:10.1109/ISSPIT47144.2019.9001846

[27] Yu-Hang Z, Wei G, Tao Z, et al. Identification of microbiota biomarkers with orthologous gene annotation for type 2 diabetes. Front Microbiol. 2021; 12: 711244.

[28] Oyallon E, Belilovsky E, Zagoruyko S. Scaling the scattering transform: deep hybrid networks. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. p. 5618–5627.

[29] Shanthi T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. Comput Electr Eng. 2019;76:56–64. doi:10.1016/j.compeleceng.2019.03.004

[30] Chen W, Yang B, Li J, et al. An Approach to detecting diabetic retinopathy based on integrated shallow convolutional neural networks. IEEE Access. 2020;8:178552–178562. doi:10.1109/ACCESS.2020.3027794

[31] Khan Z, et al. Diabetic retinopathy detection using VGG-NIN a deep learning architecture. IEEE Access. 2021;9:61408–61416. doi:10.1109/ACCESS.2021.3074422

[32] Al-Antary MT, Arafa Y. Multi-scale attention network for diabetic retinopathy classification. IEEE Access. 2021;9:54190–54200. doi:10.1109/ACCESS.2021.3070685

[33] Kaushik H, Singh D, Kaur M, et al. Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models. IEEE Access. 2021;9:108276–108292. doi:10.1109/ACCESS.2021.3101142

[34] Khanna M, Agarwal A, Singh LK, et al. Radiologist-level two novel and robust automated computer-aided prediction models for early detection of COVID-19

Infection from chest X-ray images. Arab J Sci Eng. Aug. 2021;48(8):11051–11083. doi:10.1007/s13369-021-05880-5

[35] Khanna M, Singh LK, Thawkar S, et al. Deep learning based computer-aided automatic prediction and grading system for diabetic retinopathy. Multimed Tools Appl. Mar. 2023;82(25):39255–39302. doi:10.1007/s11042-023-14970-5

[36] Kaur J, Mittal D, Malebary S, et al. Automated detection and segmentation of exudates for the screening of background retinopathy. J Healthc Eng. 2023;2023:1–12. doi:10.1155/2023/4537253

[37] Khanna M, Singh LK, Thawkar S, et al. PlaNet: a robust deep convolutional neural network model for plant leaves disease recognition. Multimed Tools Appl. 2023;83(2):4465–4517. doi:10.1007/s11042-023-15809-9

[38] Mansour RF. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. Biomed Eng Lett. 2017;8(1):41–57. doi:10.1007/s13534-017-0047-y

[39] Fujieda S, Takayama K, Hachisuka T. Wavelet convolutional neural networks; 2018. Preprint, arXiv:1805.08620.

[40] Shakibania H, Raoufi S, Pourafkham B, et al. Dual branch deep learning network for detection and stage grading of diabetic retinopathy. Biomed Signal Process Control. Jul. 2024;93:106168. doi:10.1016/j.bspc.2024.106168

[41] https://www.kaggle.com/c/diabetic-retinopathy-detection/data

[42] Civit-Masot J, Luna-Perejón F, Vicente-Díaz S, et al. TPU cloud-based generalized U-Net for eye fundus image segmentation. IEEE Access. 2019;7:142379–142387. doi:10.1109/ACCESS.2019.2944692

[43] Raja C, Gangatharan N. Incorporating phase information for efficient glaucoma diagnoses through hyper-analytic wavelet transform. Adv Intell Syst Comput. 2015;336:325–339.

[44] Jia M, Li F, Wu J, et al. Robust QRS detection using high-resolution wavelet packet decomposition and time-attention convolutional neural network. IEEE Access. 2020;8:16979–16988. doi:10.1109/ACCESS.2020.2967775

[45] Zuo W, Zhou F, Li Z, et al. Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection. IEEE Access. 2019;7:32510–32521. doi:10.1109/ACCESS.2019.2903587

[46] Varshney M, Singh P. Optimizing nonlinear activation function for convolutional neural networks. Signal Image Video Process. 2021;15:1323–1330.doi:10.1007/s11760-021-01863-z.

[47] Qian H, Liu H, Liu C, et al. Adaptive activation functions in convolutional neural networks. Neurocomputing. 2018;272:204–212. doi:10.1016/j.neucom.2017.06.070

[48] Obla S, Gong X, Aloufi A, et al. Effective activation functions for homomorphic evaluation of deep neural networks, Vol 8; 2020.

[49] Dong B, Wang X, Qiang X, et al. A multi-branch convolutional neural network for screening and staging of diabetic retinopathy based on wide-field optical coherence tomography angiography. IRBM. 2022;43(6):614–620. doi:10.1016/j.irbm.2022.04.004.

[50] Sau PC, Bansal A. A novel diabetic retinopathy grading using modified deep neural network with segmentation of blood vessels and retinal abnormalities. Multimed Tools Appl. 2022;81:39605–39633. doi:10.1007/s11042-022-13056-y

[51] Raja Sarobin MV, Panjanathan R. Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks. Symmetry (Basel). 2022;14(9):1932. doi:10.3390/sym14091932

[52] Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images; 2019. p. 1–7.

[53] Varadarajan AV, Poplin R, Blumer K, et al. Deep learning for predicting refractive error from retinal fundus images. Invest Ophthalmol Visual Sci. 2018;59(7):2861–2868. doi:10.1167/iovs.18-23887

[54] Van Grinsven MJ, van Ginneken B, Hoyng CB, et al. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. IEEE Trans Med Imaging. 2016;35(5):1273–1284. doi:10.1109/TMI.2016.2526689

[55] Gondal WM, Köhler JM, Grzeszick R, et al. Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE; 2017. p. 2069–2073. doi:10.1109/ICIP.2017.8296646.

[56] Raja C, Balaji L. Retinopathy grading with deep learning & wavelet hyper-analytic activations. Vis Comput. 2023;39:2741–2756. doi:10.1007/s00371-022-02489-z