A Comprehensive Survey on AlexNet improvements and fusion techniques

Bahaa S. Rabi[1,*], Ayman S. Selmy[1], Wael A. Mohamed[1]

[1]Dept. of Electrical Engineering, Benha Faculty of Engineering, Benha University, Benha, Egypt.

Emails: b.mohammed50868@beng.bu.edu.eg; ayman.mohamed01@bhit.bu.edu.eg; wael.ahmed@bhit.bu.edu.eg

Abstract

Machine- and deep-learning techniques have been used in numerous real-world applications. One of the famous deep-learning methodologies is the Deep Convolutional Neural Network. AlexNet is a well-known global deep convolutional neural network architecture. AlexNet significantly contributes to solving different classification problems in different applications based on deep learning. Therefore, it is necessary to continuously improve the model to enhance its performance. This survey study formally defined the AlexNet architecture, presented information on current improvement solutions, and reviewed applications based on AlexNet improvements. This work also presents a simple survey based on a fusion of AlexNet with different machine-learning techniques for recent research in biomedical applications. In the survey results for about 11 research papers for both improvement and fusion techniques of AlexNet, it was clear that the fusion was the superior one with 99.72, and the improved one was 99.7%. In the conclusion and discussion section, there was a comparison between the improved techniques and fusion techniques of AlexNet and a proposal for future work on AlexNet development.

Keywords: Artificial Intelligence; Deep Learning; AlexNet improvements; Machine learning; Convolutional Neural Networks; Fusion of AlexNet

1. Introduction

Machine Learning (ML), as a subset of artificial intelligence (AI), has been the state of the art. There are two distinct approaches to problem-solving using machines: ML and traditional programming [1], Classic programming entails clearly outlining a set of guidelines and directives that the computer must adhere to, taking into account the problem-solvers understanding and the intended result. Although it can successfully resolve well-defined issues, it can become cumbersome and ineffective when tackling complex and ambiguous problems. However, ML can learn and adapt on its own through experience, enabling it to deal with a wider range of problems more efficiently. Machine learning [2] teaches machines to change their behavior based on experience and learn from data. When it comes to solving problems, this method is more adaptable and widespread than traditional programming. A comparison of ML and classical programming is presented in Figure 1.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

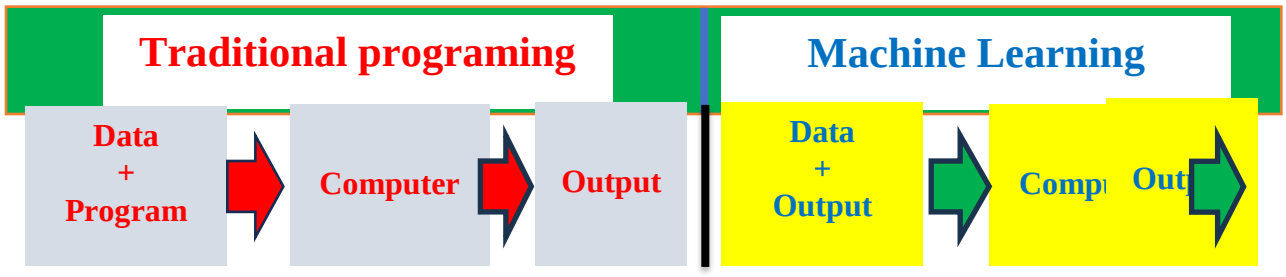


Figure 1. Comparison between Traditional Programming and ML

Machines may use ML to analyze vast volumes of data, find patterns and correlations, and forecast or take action. For complicated decision-making tasks like image identification, autonomous driving, and natural language processing (NLP), ML is therefore a good fit. ML-enhanced AI applications contain a subset known as Deep Learning (DL). Deeper variations, either linear or nonlinear, of artificial neural networks (ANNs) with numerous layers are known as DL models [3]. Due to their ability to acquire hierarchical features from a variety of input sources, including text, audio, pictures, and numbers, deep learning models are useful for solving problems related to recognition, regression, semi-supervised learning, and unsupervised learning [1]. CNN, which refers to a system or method that is capable of independently extracting distinctive CNNs with local connections and weight communication as their two noteworthy qualities, has been more and more popular in the recent few years due to advancements in computer power and data storage [4]. A local connection can be used exclusively for local feature learning that associates the convolutional layer's receptive field with the corresponding area of the original feature map.

On the other hand, weight sharing results in a large reduction in the number of parameters by indicating that convolution kernels that match have the same weight. CNNs have been employed for medical image analysis, social network image classification, and many more computer vision domains because of the combined action of these two features. One example of a historic CNN model that has many convolution layers is the Deep Convolutional Neural Network (DCNN) AlexNet, which was first introduced by Krizhevsky, Sutskever, and Hinton.

The fusion of CNNs was suggested in [5], and it was suggested to classify medical picture modalities using a variety of distinct CNN architectures. Deep image features are extracted using a fine-tuned CNN. After training multiple multiclass classifiers with the collected features, the posterior probabilities of these classifiers are pooled to perform the classification task. The testing findings showed that the suggested technique could achieve a higher classification accuracy of 96.59% by utilizing only fine-tuned CNN.

The proposed work will focus on the significance of Alex Net's bettered-style perpetration since 2012 and how it revolutionized DL applications. It will also claw into the innovative armature and use of convolutional layers that set AlexNet apart from the standard AlexNet model. This investigation will explore the impact of enhancing AlexNet on enabling researchers to attack complex problems with extraordinary delicacy and effectiveness. The purpose of the survey is to provide readers with a thorough comprehension of Alex Net's enhanced methods from the perspectives of data and models.

Also, the proposed work will focus on the significance of Alexnet fusion with different ML models in recent biomedical applications. In summary, because of the huge research volume based on Alexnet fusion with ML models proposed in multiple applications, in this version, we present only recent research, only biomedical applications based on Alexnet fusion with other ML models.

The mechanisms and strategies of an improved AlexNet approach are introduced to allow readers to grasp how these approaches work. Several existing AlexNet-improved studies have been connected and systematized. Specifically, about 10 representative AlexNet-improved approaches are introduced. In this survey, we focused more on Alex Net's improvement in the medical field, as it is a very important approach in our future research. This article is broken up into four sections. Section 2 introduces a summary of recent related work. Section 3 introduces the CCN overview. Section 4 presents the methods including, AlexNet Overview, AlexNet improvements and fusion According to biomedical applications which apply different techniques. Section 5 concludes this survey and introduces future work. A summary of the survey's determinations is as follows:

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

- This paper introduces and summarizes 6 modern approaches that have improved the representative AlexNet model.
- This paper introduces and summarizes 5 modern approaches which utilize AlexNet fusion with ML techniques for biomedical applications.

2 Related Work

In this section, a proposed review of previous works that were interested in biomedical application development based on DCNN is presented. Some previous works were based on DCNN improvement techniques, especially AlexNet, but others were based on DCNN fusion techniques. This will be summarized in the next two subsections.

2.1 Related work based on DCNN improvement techniques

An upgraded AlexNet-based algorithm for blood group interpretation was proposed by Ranxin Shen, et al [6], achieving an accuracy of 96.6%.&Arun Kumar, et al [7] suggested an enhanced AlexNet classifier based on FFT for the classification of ECG signals, achieving an accuracy of 99.7% %.&Yi Hu, et al [8] improved the AlexNet model and measured skeletal maturity automatically using images from hand X-rays, achieving an accuracy of 96.23%.&An intelligent approach to identify diabetic retinopathy using an improved AlexNet model was proposed by Sadhana, S., and R. Mallika, et al [9], achieving an accuracy of 97.81%.&Yue Duan, et al [10] To automatically segment prostate cancer magnetic resonance images, the researchers introduced the 3D AlexNet technique. Achieving an accuracy of 92.1%. &A technique for MRI abnormal brain identification using augmented AlexNet and ELM, optimized by the chaotic bat algorithm, was proposed by Lu, Siyuan, et al [11], achieving an accuracy of 92.1%. &a computer-aided diagnosis (CAD) system is presented by Shakarami, Ashkan, Hadis Tarrah, and Ali Mahdavi-Hormat [12] for the diagnosis of Alzheimer's disease, based on AlexNet-SVM technique is proposed for feature extraction and classification, achieving an accuracy of 96.39%. & in a study proposed by Alaskar, Haya, et al. [13], the effectiveness of AlexNet in identifying abnormalities in PCG signals, achieved 87% recognition accuracy.

2.2 Related work based on DCNN fusion techniques

With an accuracy of 98.3%, Sathishkumar, R., et al. [14] presented an enhanced Google Net and AlexNet Fusion Model for Deep Learning Breast Cancer Prediction. & Lung Cancer Detection Using a Fusion Approach Based on Modified AlexNet Architecture and Support Vector Machine was proposed by Naseer, Iftikhar, et al. [15], achieving an accuracy of 97.64%. &a novel method for classifying lung cancer histopathology images was presented by Sethy, Prabira Kumar, et al. [16]. It was based on the fusion of wavelets with AlexNet and achieved 97.64% accuracy. &A unique brain tumor detection model using deep feature fusion (AlexNet, Google net, etc.) and well-known machine learning classifiers (SVM, KNN) was proposed by Kibriya, et al. [17], attaining a 99.7% accuracy rate. &In order to assist clinics in assessing the health state of fetuses, Muhammad Hussain, et al. [18] developed a fusion DL system (AlexNet-SVM) for use on cardiotocographic data. Reaching a 99.72% accuracy rate& for the purpose of detecting skin cancer, Ahmed Abdelaziz, et al [19] created an intelligent system made up of an integrated deep feature fusion model. Reaching a 91% accuracy rate& with a 78% accuracy rate, Taneja, Harsh, et al. [20] used ML and DL approaches to detect the presence of COVID-1. & In order to improve the healthcare sector, Atassi, Reem, et al. [21] suggested a novel data fusion model in an IoT environment. The goal of the SSOECC-MIC model, which is a suggested model, is to build an efficient encryption scheme with an ideal key generation procedure for the IoT. The experimental results showed that the SSOECC-MIC model outperformed more modern models. & A novel strategy was put up by Kaur, Surinder, et al. [22]; they provided an algorithm for the detection of breast cancer that makes use of asymmetry analysis as the primary option. And unity at the decision level. The image feature representation is the combination of local nuclei features that were retrieved using CNN models that had been previously trained on the database. Attaining 96.91% accuracy& a totally DCNN-based automated cardiovascular disease diagnosis method was presented by Z., Abdallah, et al. [23]. In this instance, DL convolution filters and full CNNs are used to evaluate the risk of cardiovascular disease (CVD). Attaining an 88% accuracy rate. & Sheng Long, Lee, and Zare, Mohammad Reza [24]. This method combines the individual outputs of the DL models using score-based fusion. The best test accuracy achieved by the suggested combined model was 90.03 and 96.93%. & in their conclusion, Sinha, Divyanshu, et al. [25] point out that progressive fusion may be a topic for further study in the future, contingent on prior biological information or search techniques. Similarly, multimodal datasets may be able to overcome sample size constraints through the use of transfer learning. When these datasets become even more publicly available, multimodal deep learning approaches offer the chance to build complete models that are able to recognize the complex administrative cycles that underlie health and illness. In the next section of this survey, we present a simple review of CNN architecture.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

3 CCN overview

Understanding the CNN structure and operations performed by its components is essential to facilitating the AlexNet model's understanding because it is considered a DCNN model application. In this section, CNN will be explained in as much detail as possible to help researchers understand the main terms of CNN and consequently do not need to repeat this explanation when we define AlexNet and the improvements that are applied in the next section.

3.1 CNN Main Components

A straightforward convolutional network is a stack of layers, and every layer converts one volume of activation function to another with the assistance of a differentiable function.

The CNN [26] is equipped with 3 layers: a convolutional, a pooling, and a fully connected layer. These layers comprise a convolutional network architecture.

Figure 2 represents the basic architecture of the CNN, an overview of the frequently utilized layers, their definitions, and the roles they play in the network is given in Table 1.

Researchers and practitioners can create CNN designs that are more effective for a variety of applications by understanding the roles of each layer [27]. This knowledge can be valuable in a range of fields, from computer vision to healthcare.

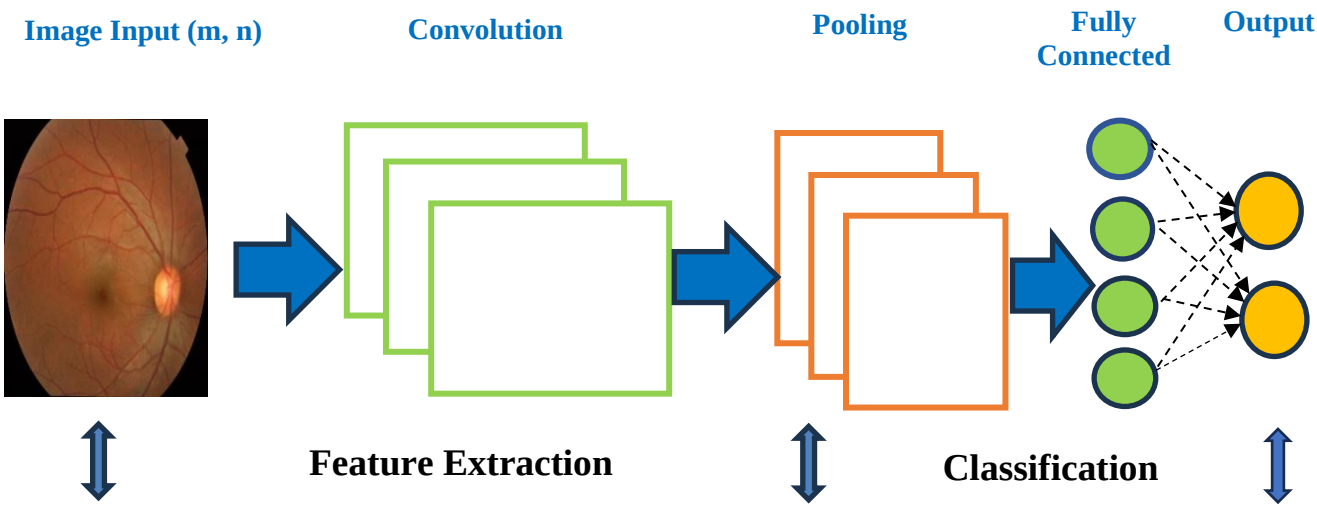


Figure 2. Basic DCNN architecture.

Table 1: CNN's layers and their functions.

Layer	Properties
Convolution Layer	The heart-building block of a CNN. Creates a collection of feature maps by acting as a convolution process on the input image using a number of learnable filters.
Pooling layer	Layer is employed to reduce the convolutional layer's feature maps' spatial dimensions. It operates independently on every feature map, down sampling each one by calculating the average or maximum value of the non-overlapping regions. Pooling offers a decrease in the computational cost of the network and makes it more resistant to tiny translations in the input picture.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

Activation layer	The output of the preceding layer is subjected to a non-linear activation function by the activation layer. By doing this, the network gains non-linearity and becomes capable of learning more complex information.
Batch Normalization layer	By subtracting the mean and dividing by the batch standard deviation, the batch-normalization layer adjusts the output of the preceding layer. This improves the network's connectivity and aids in reducing the internal covariate shifting.
Dropout layer	During training, the dropout layer randomly eliminates a portion of the neurons from the preceding layer. This forces the network to learn better features, preventing over-fitting.
Fully connected layer	A typical neural network layer called the fully connected layer integrates every neuron from the preceding layer with every neuron from the current layer.

3.2 Convolution Layer Operation

In signal processing, image processing, and computer vision, the convolution operation is a frequently used mathematical procedure. They frequently create a third signal by combining two signals or functions. This third signal, weighted by the other signal's shape, shows how one signal has affected the other. The following equation (1-3) defines the convolution operation's mathematical formula.[28]:

$$(u * g)[n]=\sum_{-\infty}^{\infty} u[s]g[n-s] \quad (1)$$

u and g are two functions (discrete or continuous), n is the position or time index of the output signal, symbol $*$ represent (Conv_Op). For discrete input signals, Eq. (1) can be written as follows:

$$(u * g)[n]=\sum_{-\infty}^{\infty} u[s]g[n-s] \triangle s \quad (2)$$

where $\triangle s$ is the sampling interval. For continuous input signals, Eq. (1) can be written as follows:

$$(u * g)(t)=\int_{-\infty}^{\infty} u(\tau)\, g(t-\tau)d\tau \quad (3)$$

where t denotes the time index of the output signal.

The Conv_Op is basic in CNN structure. To create an output feature map, it slides a filter or kernel over an input image and calculates the dot product at each location.

3.3 Feature Maps

An integral component that defines a convolutional layer's output is a feature map (Feat_Map). A two-dimensional array called Feat_Map indicates how well the input image's local areas match the filters used on them. Every Feat_Map feature captures particular elements of the input image and is compared to the activation of a neuron in the layer.

The input image is subjected to a series of filters to calculate the Feat_Map. Every filter creates a Feat_Map that draws attention to a specific pattern or feature in the input image. Combining the Feat_Maps produced by distinct filters in the same layer is possible to produce a multichannel Feat_Map that captures diverse facets of the input image.

Equation (4) defines the Feat_Map mathematical formula. [28]:

$$\mathbb{y}_{i,j,k}=\sum_{l=1}^{F}\sum_{m=1}^{F}\sum_{n=1}^{C_{in}} \mathcal{W}_{l,,m,n,k}\, \mathcal{X}_{i+l-1,j+m-1,n}+B_k \quad (4)$$

where $\mathbb{y}_{i,j,k}$ presents the output tensor's k-th Feat_Map value at location (i, j), $\mathcal{W}_{l,,m,n,k}$ presents the weight of the k-th filter at position (l,m,n), $\mathcal{X}_{i+l-1,j+m-1,n}$ presents the value of the n-th input Feat_Map at position $(i+l-1, j+m-1, n)$, and B_k presents the bias term for the k-th Feat_Map.

The filter's height and width are represented by F, and the number of channels in the input tensor is indicated by C_{in}, and each element of the k-th Feat_Map receives a scalar bias term called B_k .

The weights were learned through back-propagation and gradient descent during the training process. They ascertained the responses of each filter to the specific areas inside the input image. A sliding filter is applied to the input image during the feature map computing process, and the dot product between the filter and the appropriate local region of the

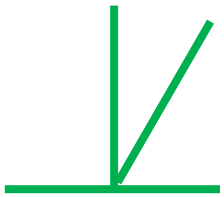
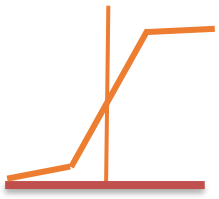
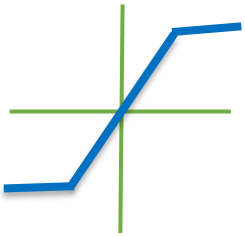
DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

input image at each position is computed. Each filter in the layer receives a duplicate set of weights that are then added together and run through an activation function to determine the final values of the Feat_Map process. This produces a set of Feat_Map that together capture various features of the input image.

3.4 Activation Function

The activation function (Act_Func) is an important element in CNNs that presents nonlinearity in the model.
This nonlinearity is vital because numerous real-world phenomena are complex. Act_Func enhances CNNs' expressiveness and ability of CNNs to model a wide range of functions.
Table 2 summarizes the commonly used Act_Funcs, their mathematical equations (5-7), illustrations, and other properties [29].

Table 2: Activation Function Types and Features.

Activation Function	Illustration	Mathematical Equation	Properties
Rectified Linear Unit (Relu)		F (i) = Maximum (0, i) (5) -i present the input to the neuron	- Multi-layer Perceptron (MLPs) and CNNs Apps. - Halting negative values while maintaining positive values. - Enables more realistic learning of challenging features in CNNs and helps control the saturation of neurons during training.
Sigmoid /Logistic		$F(i)=1/(1+e^{i})$ (6) -i present the input to the neuron	-S-shaped function - Can be applied to any real number to convert it to a value between 0 and 1. - Good option for issues involving binary classification. - Because it is saturating, it struggles with the vanishing gradient problem when employed in deep neural networks.
Hyperbolic Tangent (Tanh)		$F(i)=\frac{(e^{i}-e^{-i})}{(e^{i}+e^{-i})}$ (7) -i present the input to the neuron	- Maps every real number to a value between -1 and 1, much to the sigmoid function. - It may experience the vanishing gradient issue in DCNN, just like the sigmoid function.

3.5 Padding, Stride, and Filters

The size and resolution of the output Feat Map are determined by the parameters that CNNs need to have: padding, stride, and filters [30].

These parameters are essential for managing the quantity of data in Feat_Map and have a big impact on how the network is implemented.

- **Padding** entails surrounding the input image with more rows and columns of zeros before applying the filters. Because the convolutional layers supply the spatial dimensions of the image, this can aid in maintaining them. To ensure that the output Feat_Map has the same spatial dimension as the input image, padding is usually employed.
- **Stride** is a crucial CNN parameter that defines how far apart contiguous filter locations should be as the filter passes over an input picture. Handling the output resolution is simple when the stride is adjusted. Feat_Map. A wider stride

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

affects the output with less resolution. Feat_Map, but a lower stride corresponds to a greater resolution. Feat_Map results.

- **Filters** are used in the input image as tiny matrices to create Feat_Map. Backpropagation is used to correct the filter values after they are learned during the training phase. One of the most important factors in determining the scope and intricacy of the elements the filter captures is the filter size.

Four parameters determine the size of the output Feat_Map: stride, padding, filter size, and input picture size. Equation (8) can be used to estimate the size of the output Feat_Map [22].

$$\text{Feat\_Map}_{\text{size}}=\frac{(W-F+2P)}{S}+1 \quad (8)$$

where W is the width (or height) of the input picture, F is the width (or height) of the filter, S is the stride, and P is the amount of padding.

To build efficient neural network architectures that can precisely capture and identify complicated visual patterns in images and other data, CNNs must have padding, stride, and filter functions.

3.6 Training CNNs

To make sure the CNN delivered accurate predictions of the training data, the model's weights and biases were adjusted throughout training. An optimization process that reduces a loss function (Loss_Func) is used to do this. The distinction between the true output and the anticipated output of the model is estimated by the loss function. Stochastic gradient descent (SGD) is the optimisation algorithm most frequently used to train CNNs. Based on the gradient of the Loss_Func concerning these parameters, this algorithm modifies the model's weights and biases incrementally [31].

Through CNN training, back-propagation occurs, this is the process of employing a gradient to update a neural network's weights and biases. To adjust the weights and biases of each layer, it propagates the error from the output layer (Out_Lay) throughout the network. This is accomplished by applying the calculus chain rule to determine the derivative of the Loss_Func concerning each network parameter.

Another term is called the learning rate, this is a hyper-parameter that controls how big the weight and bias changes are for each optimization algorithm iteration. It is imperative to choose the right learning rate since an incorrect value could lead the optimisation algorithm to fail to converge, overshooting the ideal weights and biases. Conversely, an excessive value could cause the process to converge slowly.

3.7 Back Propagation

A procedure called backpropagation is used to determine the gradients of the Loss_Func concerning the weights and biases of a neural network [32]. It functions by transferring the mistake from the Out_Lay throughout the network and estimating the derivative of the Loss_Func for each network parameter using the calculus chain rule. Figure 3 illustrates how backpropagation works. The subsequent stages can be used to express the back-propagation algorithm as a mathematical procedure [28]:

1. The network output for a given input is calculated.

2. Determine the difference between the actual and expected results.

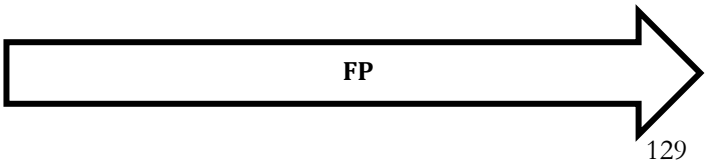
3. Determine the gradient of the Loss_Func relative to the network's output.

4. The gradient of the Loss_Func concerning the weights and biases of each layer in the network, which are represented by the following equations (9,10), can be calculated using the chain rule;

5- An optimization algorithm is employed to update the network's weights and biases.

6- Steps 1–5 are repeated for every training case.

7- Reiterating phases 1-6 until convergence or for a specific number of epochs.



DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

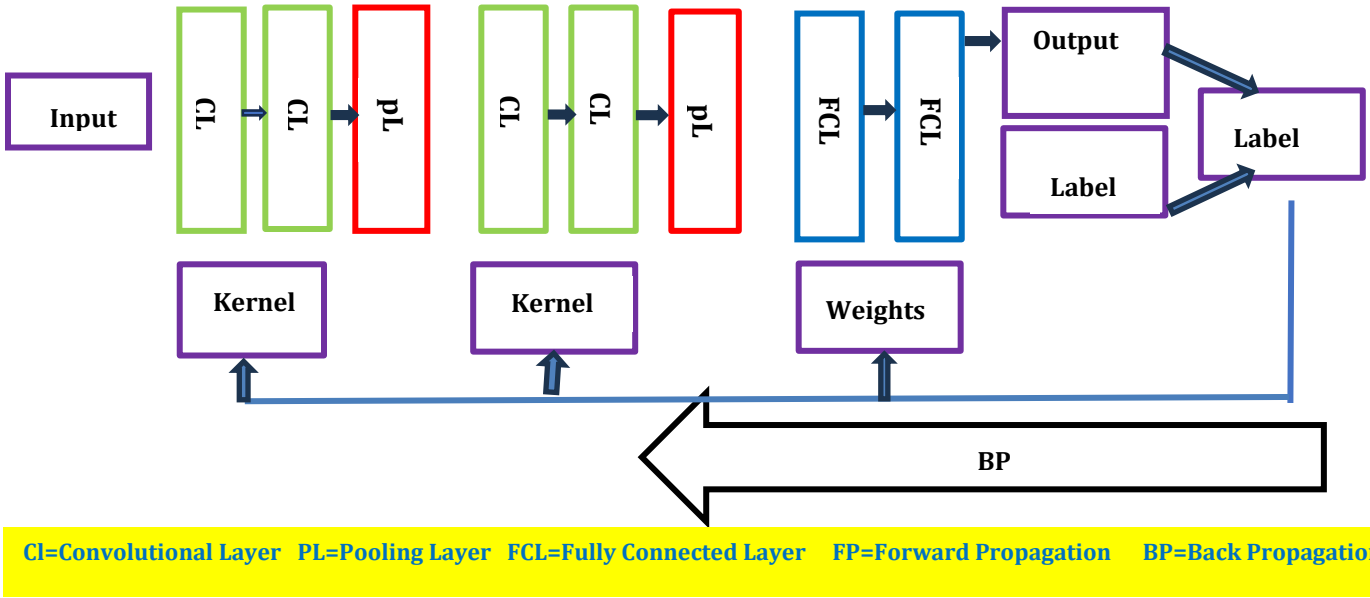


Figure 3: CNN backpropagation procedure.

3.8 Optimization Algorithms

Optimisation techniques are used to change the weights and biases of a neural network during training in order to minimize the loss function [33]. For neural network training, stochastic gradient descent (SGD) is a fairly popular optimisation technique. Based on the gradient, the SGD gradually modifies the network's weights and biases.

- **Momentum:** The gradient update can be made smoother and the concourse can run faster by including a momentum term. The following equations (9,10) represent the momentum mathematically;

$$\Lambda_t=\beta\Lambda_{t-1}+(1-\beta)\nabla_W L(W_{t-1}) \quad (9)$$

$$W_t=(W_{t-1})-\alpha\upsilon_t \quad (10)$$

Λ_t is the momentum at time t, β is the momentum coefficient, $\nabla_W L(W_{t-1})$is the gradient of the Loss_Func concerning the weights at the time $t-1$, α is the learning rate, and W_t is the updated weights.

- **AdaGrad:** The learning rate of each weight is modified by this method based on the size of the gradients that have been seen up to that point. Because of this adaptation, convergence might happen more quickly in level regions and more slowly in steep ones.

The equations for updating using AdaGrad are as follows (11,12) [12];

$$AG_t=AG_{t-1}+(\nabla_W L(W_{t-1}))^2 \quad (11)$$

$$W_t=(W_{t-1})-\frac{\alpha}{\sqrt{G_t+\epsilon}}\nabla_W L(W_{t-1}) \quad (12)$$

The diagonal matrix of past gradient sums squared up to time t is called AG_t, A tiny constant called ϵ prevents division by zero., and all other variables are defined in the previous section.

- **RMSProp:** This algorithm adjusts the learning rate for each weight by calculating the moving average of squared gradients. By doing this, the learning rate is kept from growing too quickly. The update rule for RMSProp can be defined as the next equations (13, 14) [28].

$$MG_t=\beta MG_{t-1}+(1-\beta)(\nabla_W L(W_{t-1}))^2 \quad (13)$$

$$W_t=(W_{t-1})-\frac{\alpha}{\sqrt{MG_t+\epsilon}}\nabla_W L(W_{t-1}) \quad (14)$$

where $\mathbf{MG_t}$ is the moving average of the squared gradients up to time t, other terms are as defined before.

- **Adam**: This algorithm creates an adaptable learning rate that is effective for a variety of neural networks by

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

combining ideas from AdaGrad and momentum. The mathematical equations represent Adam's algorithm (15–19) [28];

$$m_t=\beta_1 m_{t-1}+(1-\beta_1)\nabla_W L(W_{t-1}) \quad (15)$$

$$\Lambda_t = \beta_2 \Lambda_{t-1} + (1-\beta_2)(\nabla_W L(W_{t-1}))^2 \quad (16)$$

$$\widehat{m}_t = \frac{m_t}{1-\beta_1{}^t} \quad (17)$$

$$\widehat{\Lambda}_t = \frac{\Lambda_t}{1-\beta_1{}^t} \quad (18)$$

$$W_t = (W_{t-1}) - \frac{\alpha}{\sqrt{\hat{v}_t}+\epsilon}\widehat{m}_t \quad (19)$$

where are the estimates of the gradients' first and second moments, the decay rates for those moments, and the estimates of the moments' first and second moments after bias correction; other terms are defined before.

3.9 Regularization Techniques

To prevent neural networks from over-fitting, regularization techniques are employed. When a model performs well on training data but poorly on fresh, untested data, this is known as over-fitting [34]. Regularization techniques add a penalty term to the Loss_Func to prevent this. This motivates the model to have smaller weights or weights that are simpler. The most utilized regularization techniques are defined as next:

- **L1 Regularization:** It is a method that adds a value proportionate to the absolute value of the weights to reduce the Loss_Func. This can lead to feature selection and promote the model to have sparse weights. Equation (20) provides an expression for the regularized Loss_Func for the L1 regularization Function.

$$\tilde{L}(W)=L(W)+\Psi \sum_{i=1}^{n} |W_i| \quad (20)$$

L (W) is the original Loss_Func, W_i is the ith weight in the network, n is the total number of weights, and Ψ is the regularization strength.

- **L2 regularization:** It's a method for keeping a model from overfitting. It functions by increasing the Loss_Func by a penalty that is based on the square of the weights. This discourages over-fitting by encouraging the model to have smaller weights. The following equation (21), for L2 regularization, can be used to represent the regularized Loss_Func.

$$\tilde{L}(W)=L(W)+\frac{\lambda}{2}\sum_{i=1}^{n} W_i{}^2 \quad (21)$$

All the above equation terms are defined before.

- **Dropout:** In the process of training a neural network, the Dropout approach eliminates some neurons at random. This can result in better generalization performance by preventing an over-reliance on any one feature. The output of each neuron can be randomly set to zero during training with a specified probability p to apply this regularization strategy. The dropout regularization can be represented mathematically by the following equations (22–24):

$$\hat{y}=r \odot y \quad (22)$$

$$D_{(r_i=1)}=1-D \quad (23)$$

$$D_{(r_i=0)}=D \quad (24)$$

A layer's output is represented by y, and its regularized output by $\hat{y}$, a random binary mask created for every training example by r, and element-wise multiplication by ⊙.

To account for the impact of dropout, the output is scaled by 1-D and the dropout is disabled during testing.

- **Early stopping:** This method stops a model from fitting the training set too closely. During training, it keeps an eye on a validation set performance indicator like validation loss or accuracy. The training process is terminated early if the metric ceases to improve.

3.10 Evaluation Metrics

A confusion matrix, a commonly used assessment approach in image categorization, can be utilized to assess the learning performance of the CNN model. The confusion matrix allows us to distinguish between precise and imprecise predictions and helps quantify the degree to which the expected results agree with the actual values. It also sheds light on the

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

particular kinds of mistakes the model makes. A collection of test and validation data with the values of the acquired results is needed to calculate the confusion matrix.
To determine the key parameters, such as accuracy and sensitivity, etc., utilize the confusion matrix. The assessment metric is displayed in Figure 4[28].

- **True Positive (TP):** It indicates the total number of positive samples that were accurately identified.
- **True Negative (TN):** It indicates the total number of negative samples that were accurately identified.
- **False Positive (FP):** It indicates the total number of positive samples that were mistakenly detected.
- **False Negative (FN):** It specifies the total quantity of negative samples that were mistakenly detected [2].

In the next section Methods, an overview of AlexNet Model.

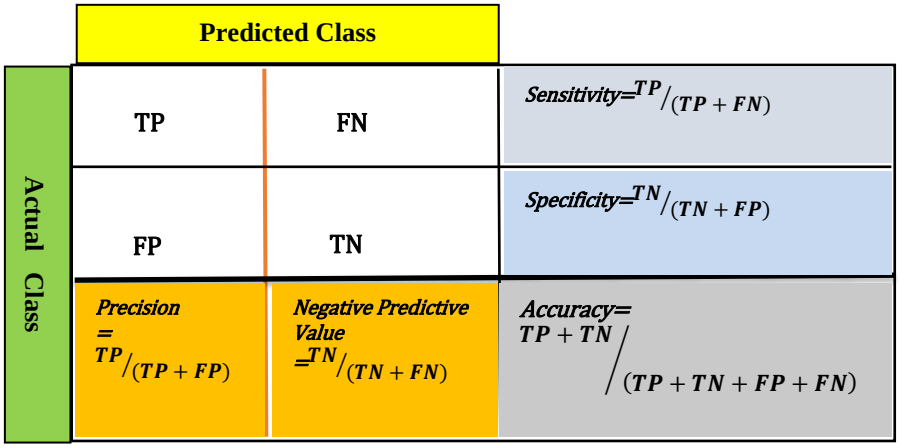


Figure 4: Confusion matrix.

4. Methods

4.1 AlexNet Overview

AlexNet is a DCNN architecture that revolutionized the field of computer vision. In 2012, Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton [35] presented it, and it easily emerged victorious in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). This breakthrough architecture demonstrated the power of deep learning in image classification tasks and paved the way for the development of more advanced neural networks in computer vision. The model boasts an 8-layer structure and unique advantages in image classification.
Figure 5 depicts the precise structure of AlexNet. The input format for the data source is 227 ×227 ×3 pixels, where 227 pixels represent the width and height of the input image and 3 indicates that the data source is operating in three-channel RGB mode. This means that it supports color images in widely used formats, meaning that additional formats are not required for the original data source. The computational processes involved in the top two layers are convolution (Conv), maximum pooling (max-pooling), and activation (Relu).
After the second layer, the output result is convolutional using 256 feature maps with a kernel size of 5, a stride of 1, and the same remaining parameters as the first layer. Only convolution and Relu procedures were carried out by the third and fourth layers. Except for not having been normalized, the fifth layer is comparable to the first layer Convert the fifth layer's output into a long vector employing a three-layer fully connected structure, then input it into a traditional neural network. The first two completely connected layers have kernels of 4,096 each. The SoftMax regression algorithm can be utilized to obtain the label value of the 1,0 0 0 nodes generated by the final layer. The AlexNet architecture can be expressed mathematically as follows:

- Conv_Lay with 96 11×11 filters, together with a Relu Act_Func and a 2×2 max Pool_Lay.
- Conv_Lay with 256 5x5 filter sizes, a Relu Act_Func, and a 2x2 max Pool_Lay in sequence.
- Conv_Lay with a Relu Act_Func and 384 3x3 filters.
- Conv_Lay with 384 filters of size 3×3 and a Relu Act_Func.
- Conv_Lay with 256 3x3 filters, a Relu Act_Func, and a 2x2 max Pool Lay as a sequence.
- 4096 units in a fully connected layer with a Relu Act_Func.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

- 4096 units in a fully connected layer with a Relu Act_Func.
- Out_Lay with a SoftMax Act_Func and 1000 unit.

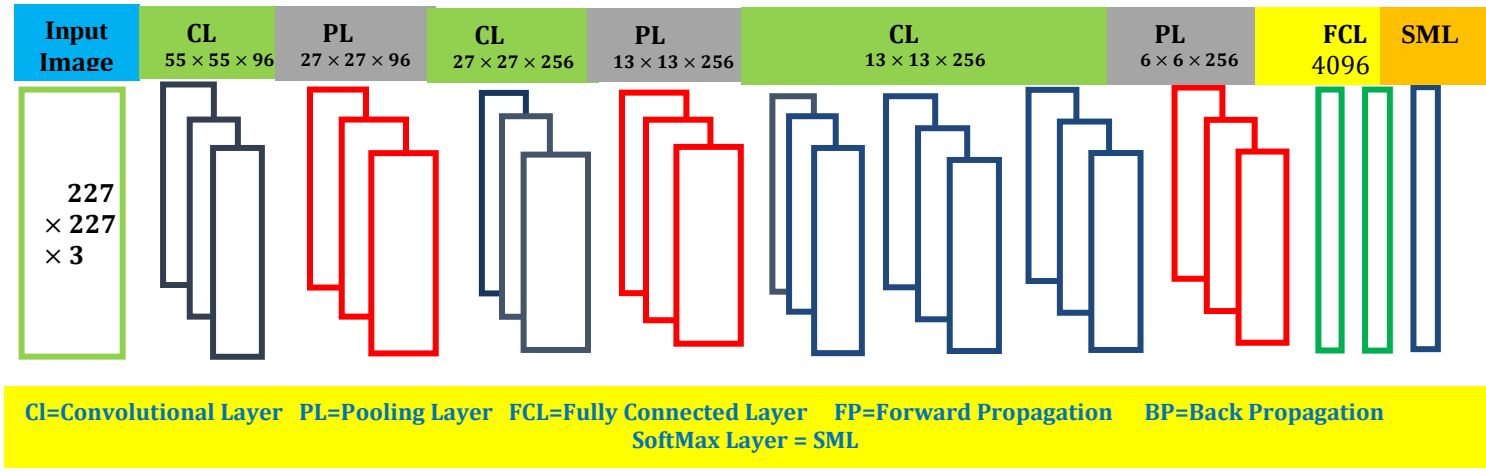


Figure 5. The AlexNet network structure

4.2 AlexNet improvements and fusion techniques overview

The advancements made to AlexNet have brought about a revolution in the DL field, by significantly improving the accuracy of image classification tasks. Its implementation in 2012 was a breakthrough moment that sparked a new era in computer vision and paved the way for many subsequent advancements. The network's innovative architecture and use of convolutional layers have allowed for the development of more sophisticated deep-learning models. This has enabled researchers to address complex visual recognition problems with unprecedented accuracy and efficiency.

The proposed work will focus on the significance of Alex Net's bettered-style perpetration since 2012 and how it revolutionized deep learning applications. It will also claw into the innovative armature and use of convolutional layers that set AlexNet apart from the standard AlexNet model. The exploration will deeply explore the impact of enhancing AlexNet by enabling experimenters to attack complex visual recognition problems with remarkable delicacy and effectiveness. Eventually, it'll conclude by pressing the ongoing advancements and unborn eventuality of deep literacy models in image bracket tasks.

This survey aims to provide readers with a comprehensive understanding of Alex Net's improved approaches from the perspective of data and models. To help readers understand how the updated AlexNet technique functions, the mechanisms and strategies behind them are presented. A number of the existing AlexNet net-improved research projects are connected and systematized. Specifically, about 6 representative AlexNet-improved approaches are introduced. Besides, in the conclusion and discussion section, there is a detailed comparison of AlexNet improvements that can be reused in future applications.

Besides, this survey presents a review of recent approaches which applied fusion techniques based on AlexNet in the biomedical applications field.

In this survey, we focused more on Alex Net's improvement and fusion techniques in the medical field, as it is a very important approach in our future research.

4.3 Research Methodology

Currently, there are a growing number of improvements being made to the AlexNet model. This has resulted in extensive research being conducted on various applications. Based on a survey, we propose a classification system for these AlexNet improvements according to the applied application. Each research paper includes a problem definition, methods. And the results obtained by the improved AlexNet approach in every approach. In this first version of our approach, the three main categories of AlexNet improvement techniques have been defined below.

- Biomedical Applications based on improved AlexNet.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

- Biomedical Applications based on the fusion of AlexNet with ML techniques.

In this proposed review for AlexNet, these 3 categories are chosen because of the great contribution of researchers to AlexNet and their importance to human progress.

4.3.1Biomedical Applications based on Improved AlexNet

4.3.1.1 First approach

Ranxin Shen [6] suggested an algorithm for interpreting blood groups based on enhanced AlexNet.
The fully automatic blood group analysis equipment has gained popularity as a topic of discussion in the field of blood type detection.
The ongoing advancement of clinical science and ongoing blood type research studies have led to the emergence of numerous techniques and formats for blood type diagnosis. The automatic blood group analysis devices have covered the microplate method and the microcolumn gel method as two of the primary international detection techniques.
The blood group analysis device is capable of identifying blood types A, B, O, and Rh as well as abnormal antibody screening and cross-blood matching.
The fully automatic blood group interpretation system, which uses the blood group analysis instrument's microcolumn gel detection method—also referred to as Card-type—is the focus of the research project. A microcolumn gel card serves as the experimental carrier in the automatic blood group analysis device, which makes use of Microcolumn Gel Detection Technology (MGDT). Another name for it is a blood group analysis tool similar to a card.
The blood group interpretation system uses a DCNN model to classify red blood cell agglutination phenomena.
To increase the reading accuracy of the current blood group interpretation system in the area of blood type detection utilizing the microcolumn gel method, the project [6] implements an upgraded AlexNet network model.
With the addition of a channel attention mechanism (SEBlock) between the first convolutional layer and the first max-pooling layer, the structure is an enhanced variant of the original AlexNet.
In DL models, attention processes are essential. Comprehending the two primary categories of attention mechanisms—spatial attention and channel attention—is crucial.
As mentioned before, CNN extracts image features as the shape of the Feat_Map is [C, H, W]. The Feat Map's shape by training is [B, C, H, W]. The addition of a spatial attention mechanism or a channel attention mechanism between the convolutional layers can improve the CNN's Feat_Map feature extraction capability, where B denotes the batch size, C the number of channels, H the height of the feature map, and W the weight of the feature map.
A channel attenuation mechanism called the SEBlock is suggested in the enhanced version of AlexNet. By explicitly establishing the connections between channels, it adaptively recalibrates the responses of channel features. These modules can be combined to form an MLP network structure, which can then be successfully generalized over a variety of datasets.
In Figure 6 [6], the SEBlock flow chart is displayed. The module comprises four main steps, as illustrated: First, the image features are extracted starting with a single image, and the feature layer U's current feature map dimension is [C, H, W]. The second step is global pooling for the feature map's [H, W] dimensions. The pooled Feat_Map results are sized between [C, H, W] and [C, 1, 1]. [C, 1, 1] can be understood to signify that there is a number that corresponds to each channel C one by one. Third, the weights of each channel are extracted directly. The weights show how each channel affects the extraction of features. The vector indicates that it has obtained each channel's weights after this stage. Fourth, the feature map U [C, H, W] is applied with the weights acquired for every channel C [C, 1, 1]. Put differently, every channel multiplies the weight that is allocated to it.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

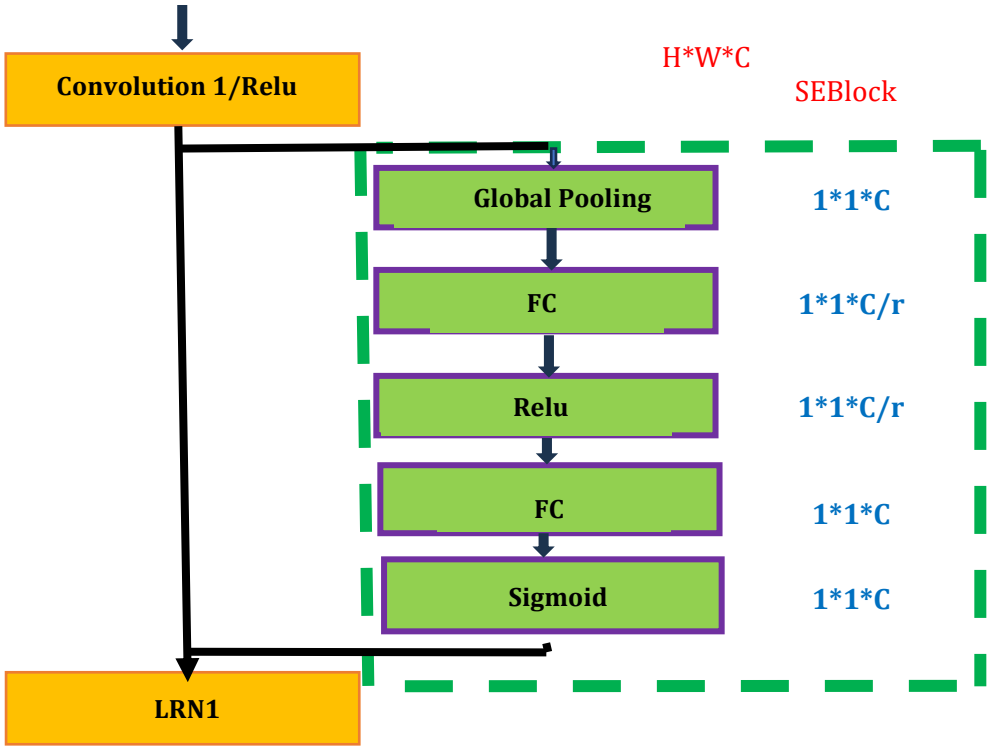


Figure 6. The SEBlock module step flow chart.

With an exceptional classification accuracy of 96.6%, the enhanced AlexNet network model demonstrates strong feature description capabilities for MGDT blood group images. Feature-distinct regions become more evident because the mixture of the channel attention method SEBlock considers the weight of each channel of the images.

4.3.1.2 Second approach

Arun Kumar, and Arvind Chakrapani [7] proposed a Classification of ECG signals using FFT based improved AlexNet classifier.

Cardiovascular disease is becoming more common and poses a major risk to one's life and health. As a result, it's critical to concentrate on this disease's diagnosis and prevention. The most popular non-invasive screening test for determining the presence of any particular heart condition is an electrocardiogram. It assists in the early detection of any anomaly by monitoring the heart's performance throughout time. Due to non-communicable and non-infectious disorders, coronary heart disease is one of the main causes of death globally. The three main contributing elements are diet, lifestyle, and occupation.

Early diagnosis and preventive interventions are essential for effective clinical treatment. The primary cause of cardiac insufficiency is abnormal functioning of the sinus node, commonly referred to as the SA node. To regulate the heart's contractions and relaxations, electrical impulses are sent and received by the SA node.

Electrical activity in the heart generates Electrocardiogram (ECG) signals. By monitoring these signals, the heart's bio-electrical activities can be precisely represented. Abnormal heart rhythms and psychological stress can be detected using this technology. ECG machines are safe and affordable; however, the signal may experience peaks due to noise and other distortions.

Artefacts include things like power line disruptions, electrode movements, and bodily motions. To ensure accurate readings, multiple adjustments are made to ECG signals to minimize noise and artefacts [7].

An enhanced FFT-based AlexNet model was put forth in this paper [7]. The model processes the incoming ECG data to produce a reduced set of functions.

A DL-based detection approach is used to extract and implement functions after the signal has been preprocessed to eliminate noise (Improved AlexNet Proposed Method).

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

As mentioned before in the AlexNet network structure section, when a hidden layer is activated before the output layer, AlexNet can generate a 4096-dimensional feature vector for every input image. With approximately 60 million parameters and 650,000 neurons, AlexNet is a large structure. The model maintains data expansion and dropout to prevent overfitting. This helps to provide more precise results and successfully reduces the problem.

This paper's suggested enhanced version of AlexNet [7] sets it apart from the conventional AlexNet classification technique.

To improve the accuracy of picture feature information, more convolution layers are added to the original AlexNet architecture and the local receptive fields are retained by the use of the max-avg pooling strategy.

The Global Average Pooling (GAP) layer replaces the original fully connected (FC) layer, reducing over-fitting without harming features. The final result remains the same even without calculating numerous network parameters, which improves network speed.

The convolution layer is augmented with an LRN layer to prevent numerical issues and avoid neuron saturation. The proposed method utilizes a BN layer after each layer's convolution.

Two smaller kernels emerge from the massive convolution kernel in a structural cascade with fewer stages. In order to improve the low-level function or the spatial information integration process, a second folding layer is added after the first. The asymmetric folding core covers the last three folding layers. The enhanced AlexNet model outperforms the original model in terms of rating accuracy, according to experiments conducted with both datasets. With this method, 99.7% accuracy, 98.3% sensitivity, 99.2% specificity, and 96.1% precision were attained.

4.3.1.3 Third approach

Study conducted by Yi Hu, Yu Lu, Xudong Zhao, and Ming He [8] enhanced the AlexNet model and used hand X-ray pictures to measure skeletal maturity automatically.

Because bones grow and develop in predictable ways, clinicians can use bone growth measurements to establish an individual's biological age. One reliable measure of a person's biological development and growth is their bone age. It is ascertained by looking at each bone's developmental stage, taking into account its size, form, location, and level of closure. In clinical care, athletic events, and legal rulings, bone age is extensively utilized as the most reliable assessment metric of biological age.

The age of the bones is significant in several professions. It aids in the analysis and diagnosis of illnesses such as metabolic, endocrine, nutritional, and genetic abnormalities in clinical medicine. Athletes' age can be reliably ascertained by their bone age, which enables them to participate in age-restricted sports. Furthermore, the growth and development rate of young athletes' bones has an impact on their function, physical fitness, and athletic skill. As a result, a key consideration in the scientific selection of athletes is bone age. In the legal system, a juvenile criminal suspect's bone age is a major factor in their conviction and sentencing. Finding out a victim's true age also aids in the investigation and gathering of evidence.

The most commonly used technique for determining the age of bones is to use left-hand X-rays, which show the fingers, palm, and wrist. The Greulich and Pyle (GP) mapping method and the Tanner and Whitehouse (TW) method [8], also referred to as the artificial average weighted maturity geometric mean method, are the two main approaches used worldwide to measure bone age. The GP mapping approach estimates bone age by comparing X-ray pictures with photos of children in atlases at different ages. This approach is straightforward, but it can also be arbitrary and untrustworthy.

This research [8] presents a new approach based on the Standards of Skeletal Development of Hand and Wrist for Chinese (CHN) technique for determining skeletal maturity evaluation. Just 14 typical hand bones are observed by this approach, which employs DCNN. The suggested flowchart illustrates the automated skeletal maturity assessment procedure utilizing the enhanced AlexNet model, as seen in Figure 7[8].

The conventional approach, which employs whole-hand evaluation, was compared with the new approach. For both approaches, the same test dataset was utilized. By employing data augmentation, the dataset was enlarged and the CNN's capacity for generalization was strengthened. The suggested approach lowers the upper bound of the absolute value of the single skeletal age mistake and increases the final skeletal age assessment's accuracy. The results of the studies demonstrated the efficacy of this approach in offering doctors more reliable, efficient, and practical diagnostic help and decision support.

In this paper [61], Good accuracy rates have been attained by four different network models. The bone age evaluation based on the modified AlexNet model presented in this study was significantly more accurate than the evaluation based on the entire hand bone. The CHN approach increases the accuracy rate by around 10% points when compared to the whole-hand classification method and whole-hand regression method. Similarly, the TOP2-CHN method increases the

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

accuracy rate by about 10% points when compared to the CHN classification method.
In particular, the TOP2-CHN classification model achieved an accuracy rate of 96.23% inside the [−1.0, 1.0] age error interval. The results of this approach demonstrate that the accuracy of the CHN and TOP2-CHN methods is higher than that of the all-hand classification and regression methods [8]. The TOP2-CHN approach is the most accurate method.
In comparison to bone age assessment based on the entire hand bone, the technique utilized in this work increases the model training duration because it requires training 14 classification models.

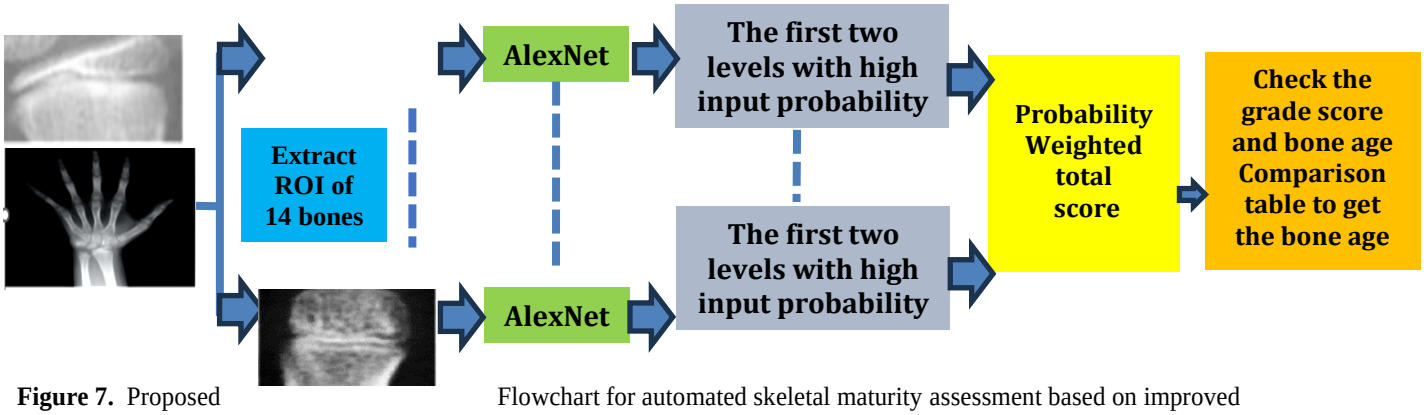
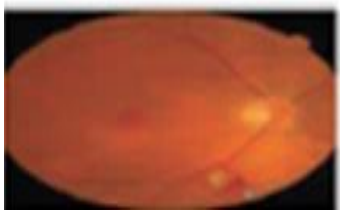


Figure 7. Proposed Flowchart for automated skeletal maturity assessment based on improved AlexNet model.

4.3.1.4 Fourth approach

Research by Sadhana, S., and R. Mallika [9] suggested an intelligent method that makes use of an enhanced AlexNet model to detect retinopathy due to diabetes.
Blindness is a serious issue caused by diabetic retinopathy (DR), which is a complication of diabetes that affects the blood vessels in the retina. It would be really beneficial if DR could be identified early in the clinical therapy process. On the other hand, early detection of DR remains a difficult task and requires a significant amount of time when done manually.
The goal of this research is to develop a quick and efficient automated method for identifying DR symptoms from retinal pictures. Using a unique histogram equalization technique for contrast enhancement and equalization at the first pre-processing stage is the first step in this approach. After that, these previously processed photos are routinely used to extract image patches.
It is common for diabetic patients to experience blindness as a result of diabetes. The World Health Organization (WHO) estimates that 422 million people globally suffered from diabetes in 2014 [9]. Because of the accumulation of damage to the small blood vessels in the retina, about 35 per cent of these patients get a specific kind of retinopathy. There is a significantly increased prevalence of DR in some patient categories.
In rural areas, around 40% of type II diabetic patients and 86% of type I diabetic patients have diabetic retinopathy (DR), with an estimated DR rate of 43%. DR is a disease that develops gradually and can lead to loss of sight.
It occurs when diabetes damages the retina of the eye, which is why it is referred to as DR. DR can cause permanent blindness, and patients are classified based on their level of risk. Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the two stages of DR.
The threat level of NPDR is modest to moderate and it is still in its early stages. PDR, on the other hand, is an advanced stage and has a severe threat level that can lead to complete blindness.
The early signs of DR are microaneurysms (MAs), which are small blood vessel swellings in the retina. Other signs of NPDR include soft exudates (EXs), hard exudates (EXs), and hemorrhages (HMs).
For the classification of diabetic retinopathy. Early diagnosis of DR can save around 90% of diabetic patients, according to various researchers. Early detection and timely treatment can prevent visual loss.
Diabetes frequently results in DR, a condition that damages the retina. It is a common sickness that is a primary cause of blindness in middle-aged and older age groups.
DR affects over 80% of persons with diabetes who have had the disease for more than ten years. A person's risk of having diabetic retinopathy increases with the length of their diabetes. A growing number of imaging platforms are available globally, and fundus photography is now often used for DR screening.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024



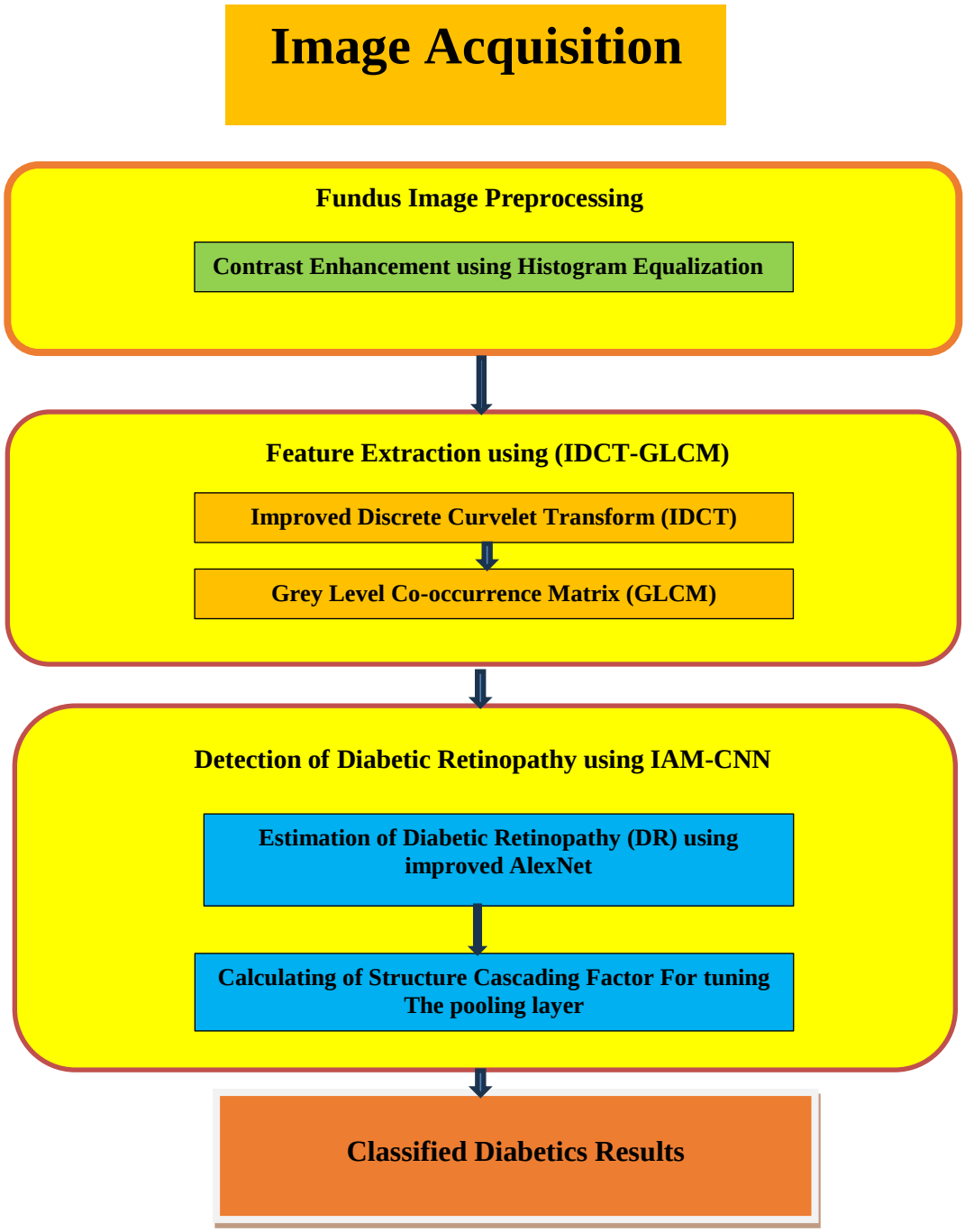


Figure 8. Overall Process of the Proposed Technique for Detection of Diabetic Retinopathy

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

This paper [9] suggested this article proposing a novel automated technique for the identification of exudates, retinal hemorrhage, and micro-aneurysms associated with diabetic retinopathy.
The detection is performed on the retinal fundus image dataset. The initial pre-processing stage involves using novel histogram equalization for contrast enhancement and equalization. Then, image patches are extracted regularly from these preprocessed images. In the second stage, features are extracted using the Improved Discrete Curvelet Transform and Grey Level Co-occurrence Matrix (IDCT-GLCM), which enhances classifier performance.
The last classifier suggested for the efficient categorization of DR from digital fundus images is named Improved AlexNet model-based.
An architectural overview of the suggested method for detecting diabetic retinopathy disease is provided in Figure 8[9]. The specific stages of the suggested method's implementation are covered in detail in this section.
With wavelets' success, many multi-resolution analytic tools have been presented recently to better depict edges and other singularities along curves.
Wavelets do not, however, contain many directional properties. Curvelets use matrix groups to effectively characterize signals at the multi-scale and multi-direction levels.
Curvelets use a higher scale value to achieve a high count of fine directions.
To achieve accurate results in this technique, manual direction is required for the fingerprint image core. A complex image enhancement process was performed before Curvelet feature extraction, which comprises image filtering, local ridge orientation, and local ridge frequency estimation across fingerprints, and binarization (turning a grayscale image to a binary image).
These are the main, intricate steps required in image enhancement.
In this proposed work, first, the authors used an improved discrete Curvelet transform (IDCT) to break down the original image into five-scale Curvelet coefficients.
The fundus image's ridge discontinuities are then smoothed by building a curvelet filter utilizing the curvelet coefficient connection at neighbouring scales.
At the coarsest scale, in the second stage, the authors produced GLCM-based texture features and grey-level co-occurrence matrices (GLCM) using curvelet coefficients.
Finally, a feature set with these combined IDCT-GLCM-based features to configure an enhanced CNN (IAM-CNN) classifier based on the AlexNet model. The best CNN architecture to handle picture categorization issues.
The following describes the procedures that were taken to put the suggested AlexNet architecture into practice.
Structured Cascading Factor-Based AlexNet Model Large spatial filter convolutions are computationally expensive. A 5x5 convolution can be split into two cascaded 3x3 convolutions to reduce the computational cost by around 28%.
To reduce the computational cost, a structure consisting of a 3x1 convolution followed by a 1x3 convolution can be used instead of a 3x3 convolution. This approach reduces computation cost by approximately 33% [(3+3)/(3x3)-1]. Additionally, the increase in computational cost due to an increase in parameter count can be minimized through convolution decomposition.
The experimental results show a considerable improvement in the performance of the AlexNet model, with a classification accuracy of about 97.81%, which is higher than the accuracy of the current model in detecting diabetic retinopathy.

4.3.1.5 Fifth approach

Research by Chen, Jun, Zhechao Wan, and Yue Duan. [10] To automatically segment magnetic resonance images of prostate cancer, the scientists presented the 3D AlexNet approach. Prostate cancer is a disease that frequently affects men with tumors. Because of the cancer's lengthy growth phase and mild symptoms, early identification is difficult. Imaging diagnosis is particularly difficult.
Medical professionals currently mostly use manual segmentation, which is a labour- and time-intensive procedure. It also greatly depends on the medical expert's experience and skills. Prostate region segmentation that is quick, precise, and repeatable remains a difficult task. Examining the 3D AlexNet network's automatic prostate image segmentation is essential.
In the suggested approach, a deep-learning convolutional neural network that has been trained on three-dimensional data is fed to the medical imaging of prostate cancer.
A technique for using the medical picture of prostate cancer as the entry point was presented in this study [10]. The DCNN receives the three-dimensional data. through the development of an enhanced 3D AlexNet technique for the automatic segmentation of magnetic resonance images related to prostate cancer.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

The traditional AlexNet uses Relu as its activation function. Relu can speed up convolutional neural network training and stop gradient dispersion during backpropagation.
The improved AlexNet introduced the PRelu activation function as a better version of the Relu activation function. Improved model training outcomes were obtained by changing the Relu function's negative axis slope from 0 to a variable parameter 'a. The equation (25) represents the PRelu activation. The Relu activation equation was explained before in Section 3.4.

$$\text{PRelu}(\chi)=\begin{cases}\chi, if\ \ \chi > 0\\ a\chi\ if\ \chi\ \leq 0\end{cases} \quad (25)$$

The impact of varying convolution kernel sizes on the precision of model identification can be observed via the PRelu activation function. Adaptive improvement based on classic AlexNet was used to create a 3D model from training sets of 500 patients with prostate cancer who had magnetic resonance imaging.
With an accuracy rate of 92.1%, specificity of 89.6%, sensitivity of 90.2%, and an Area Under Curve (AUC) of 96.4%, the model performs exceptionally well despite its straightforward construction. Between the segmentation result and the gold standard of the medical expert, the mean absolute distance (MAD) is 0.356 mm, the Hausdorff distance (HD) is 1.024 mm, and the dice similarity coefficient is 97.68%.

4.3.1.6 Sixth approach

Lu, Siyuan, Shui-Hua Wang, and Yu-Dong Zhang [11] suggested a method for MRI abnormal brain detection using enhanced AlexNet and ELM that is optimized by the chaotic bat algorithm.
The neurological system that controls our behaviour is centralized in the brain, making it the most complicated organ in the body. Regretfully, brain disorders rank among the deadliest illnesses, and a patient's survival depends on receiving a diagnosis as soon as possible.
These days, medical imaging is utilized to diagnose brain illnesses; the most often employed imaging modalities in clinical diagnosis are X-rays, CT, and magnetic resonance imaging (MRI). As MRI is non-invasive and radiation-free, it is the modality of choice for diagnosing brain diseases. In addition, it yields more lucid soft tissue imaging data than CT and X-rays, which makes it the preferred method for diagnosing brain diseases.
Medical professionals can use computer-aided diagnostic (CAD) tools to assist in decision-making based on imaging data. Using AI, the task of identifying aberrant brain patterns can be seen as an image identification and classification challenge. Classifier training and feature extraction are common components of a system used to address such image classification issues.
This paper [11] developed a new algorithm for detecting abnormal brains using ML and DL techniques, the algorithm first modifies and fine-tunes a pre-trained AlexNet on a brain MRI dataset.
Then, the authors replaced the last few layers of the modified AlexNet with an extreme learning machine. Finally, we use a novel chaotic bat algorithm to optimize the extreme learning machine for better generalization ability.
Improved AlexNet Proposal The method is defined as the BN-AlexNet-ELM-CBA model. In this research [11], the authors suggested the use of BN to improve Alex Net's dependability in identifying abnormal brains.
The wide variation in human brains is the reason for the complex distribution of brain MRIs. Because of this, the input distributions in each layer of AlexNet vary, which can make parameter training challenging and time-consuming and necessitate careful initialization. BN was created as a solution to this internal covariate shifting issue. The idea behind BN is straightforward. Because CNNs are trained in minibatch mode, BN applies a transformation that maintains the means and variances fixed, normalizing the activations of layers.
For a random variable x and its values in a mini-batch, □, the next equations define this method:

$$\square=[x_1\ x_2\ \ x_3\ \ldots\ldots\ldots] \quad (26)$$

The mean μ_s and variance σ_s^2 of x can be obtained by:

$$\mu_s=\frac{1}{n}\sum_{i=1}^{n} x_i \quad (27)$$

$$\sigma_s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_s) \quad (28)$$

So, the normalized values $\hat{x}_i$ can be obtained by:

$$\hat{x}_i=\frac{(x_i-\mu_s)}{\sqrt{\sigma_s^2+\varepsilon}} \quad (29)$$

'ε' is added as a constant to improve numerical stability. However, the learning target of the layers may not be the

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

normalized activations in some cases. Therefore, a transformation is added to the result.

$$y_i = \gamma \hat{x}_i + \alpha \quad (30)$$

where γ and α are two minibatch □ learnable parameters.
With BN, gradients become less dependent on initial parameters and training DCNNs can be accelerated. To enhance generalization, it also acts as regularization;

AlexNet has been upgraded, and it can do well in categorization. However, the last few layers, which are primarily made up of fully connected layers, have a significant impact on how well it is classified.
The authors proposed replacing these layers with an extreme learning machine (ELM), a more effective classifier model, to overcome this constraint.
Guang-Bin and Qin-Yu [36] presented the ELM training algorithm for a single-layer feedforward network (SLFN). It is anticipated that this layer substitution will increase detection accuracy even more.
SLFN (Single Layer Feedforward Network) is made up of three layers: the input layer, the hidden layer, and the output layer. 'W' and 'β' stand for the input and output weights, respectively, whereas 'b' indicates the bias in hidden nodes. 'x' and 'o' stand for the input and output, respectively. The Chaotic Bat Algorithm (CBA), a swarm intelligence optimisation technique that developed from the Bat Algorithm. The CBA employs a group of bats with viable solutions to search the solution space using certain tactics.
It is inspired by the echolocation behaviour of bats.
The bats' position, velocity, and frequency are modified in each iteration by the best solution as far discovered. The Bat Algorithm outperforms the conventional PSO in terms of optimisation and improving searching performance.
The bat placements in the CBA were updated by the authors using a chaotic map, they employed four chaotic maps for optimisation: the logistic, cosine, Gaussian, and sine maps. The experiment's findings showed that, with a sensitivity of 97.14%, specificity of 95.71%, and total accuracy of 96.43%, BN-AlexNet-ELM-CBA performed the best out of the four.

Next Table 3 summarizes the approaches mentioned in this research-based AlexNet improvements

Table 3: Summary of research based on AlexNet Improvements.

Research's Authors	**Field of Research**	**Dataset**	**Problem Definition**	**Proposed System**	**Results**
R. Shen, J. Wen and P. Zhu [6].	**Biomedical Application**	837 blood group images into 9 categories (479 training,197 validation,161 test).	The blood group interpretation system uses a DCNN model to classify red blood cell agglutination phenomena.	improved version of the original AlexNet with a channel attention mechanism (SEBlock)	Accuracy of 96.6%.
A. Kumar, A. Chakrapani [7].		MIT-BIH, 48 records.	ECG signal classification using an enhanced AlexNet classifier based on FFT.	-FFT-based improved AlexNet classifier. - FFT-based improved AlexNet classifier structure contains (Original AlexNet+BN+LRN).	99.7% precision, 96.1% specificity, 98.3% sensitivity, and 99.7% accuracy.
M. a, X., Y. Hu [8].		hengjing Hospital of China Medical University, comprising 372	An enhanced version of the AlexNet model that uses hand X-ray images to	assessment of skeletal maturity automatically using enhanced AlexNet.	Accuracy rate of 96.23%.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

		teenagers, 850 school-age children, 60 toddlers, and 384 preschoolers.	automatically evaluate skeletal maturity.		
Sadhana, S., and R. Mallika [9].		STARE images, DRIVE database	Diabetic retinopathy detection with an enhanced AlexNet model.	IDCT+ GLCM+ improved Alex Net	classification accuracy of 97.81%.
Chen, Jun, and Y. Duan. [10].		Hospital samples for 25 patients, the total samples include 500 images.	A three-dimensional AlexNet technique is suggested for automatically segmenting magnetic resonance images of prostate cancer.	AlexNet+PRELU	Accuracy rate of 92.1%, Specificity of 89.6%, Sensitivity of 90.2%.
Lu, Siyuan, Shui-Hua Wang, and Yu-Dong Zhang [11].		Brain Atlas-Harvard Medical School, including (177 abnormal samples,28 healthy controls)	A method for using the chaotic bat algorithm to optimize AlexNet and ELM for the detection of aberrant brain activity in MRIs.	BN-AlexNet+ CBA optimization	97.14% sensitivity, 95.71% specificity, and 96.43% accuracy.

4.3.2 Biomedical Applications based on the fusion of AlexNet with ML techniques.

DCNNs, specialized neural networks with grid topology, are utilized in computer vision research for object and image recognition and diagnosis problems. They utilize filtering to create output maps, demonstrating how brain research can be applied to ML applications.

Data fusion at the feature level involves combining feature map sets of feature vectors, providing rich information about unprocessed biometric data. This process enhances recognition performance and yields the best-discriminating data from diverse feature sets [37-38-39]. AlexNet has been applied in many fusion systems to improve the performance of the diagnosis and detection of diseases. In Table 4, some of the modern fusion systems that were used for biomedical applications have problems.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

Table 4: Summary of Biomedical Applications based on fusion of AlexNet with ML techniques.

RESEARCH'S AUTHORS	DISEASE TYPE, YEAR OF PUBLISHING	Dataset	Problem Definition	Proposed System	Results
Sathishkumar, R., et al [14].	Breast Cancer, 2023.	----------	The likelihood of a full recovery from breast cancer might be considerably increased by an early and precise diagnosis.	An improved Google Net and AlexNet Fusion Model for Deep Learning Breast Cancer Prediction.	98.3% Specificity, 98.3% Accuracy, 97.5% Precision, 97.3% Recall, and 98.1% F1-Score.
Naseer, Iftikhar, et al [15].	Lung cancer, 2023.	LUNA16 (1018 CT scan Images).	Lung nodule recognition in computed tomography (CT) scan pictures is essential for lung cancer identification and screening. Early diagnosis has a major impact on the prognosis and treatment plan for people with lung cancer.	a technique for automatically detecting nodules in CT images that uses the SVMs algorithm and a modified version of the AlexNet architecture, called LungNet-SVM.	99.08% specificity, 96.37% sensitivity, and 97.64% accuracy.
Sethy, Prabira Kumar, et al [16].	Lung histology, 2023.	The 5,000 digitized histopathology photos in the LC25000 Lung and Colon Histopathology Image collection	This article's goal was to create a hybrid network that included wavelet, SVM, and AlexNet to classify images related to lung histology.	Integrating wavelet, SVM, and AlexNet.	99.3% accuracy and 0.99 area under the curve (AUC).
Kibriya, Hareem, et al [17].	Brain tumors, 2022.	Glioma, meningioma, and pituitary tumor T1-weighted MRI scans were published online in 2017 by Jun Cheng.	A novel deep feature fusion-based multiclass brain tumor classification method is presented in this study.	A single feature vector is created by fusing the deep CNN features from transfer-learned architectures such as AlexNet, Google Net, and ResNet18. This feature vector is	Accuracy of 99.7%;

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

				then put into SVM and KNN for prediction.	
Muhammad Hussain, Nadia, et al... [18].	Cardiotocographic (CTG), 2022.	UCI (Machine Learning Repository) dataset.	The updated deep neural algorithm proposed in this paper can classify suspect and untapped pathogenic CTG recordings with the appropriate temporal complexity.	To lower temporal complexity, SVMs and AlexNet architecture are combined at the fully linked layers.	99.72% accuracy, 96.67% sensitivity, and 99.6% specificity, in that order.

5. Conclusions, Discussion, and Future Work

AlexNet is one of the most important models used in DL applications. In this study, we focused on two important aspects regarding the development and use of AlexNet in the field of biomedical applications.

First, improving AlexNet has a great advantage in diagnosing many diseases and extracting the most important data that helps the therapist know the best methods appropriate for the medical condition. In this study, the main two approaches [7] and [9] achieved distinguished results in accuracy of 99.7% and 97.81%, respectively, compared to other approaches based on other techniques.

Second, the approaches are based on utilizing the fusion of AlexNet into intelligent fusion systems. In this study, a deep revision of approaches focusing on the fusion of AlexNet to improve biomedical application efficacy has also given very distinguished results. The main two approaches [17] and [18] achieved distinguished results in accuracy of 99.7% and 99.72%, respectively.

From the above, we can conclude that it is possible to improve the performance of intelligent systems used in biomedical applications, whether by developing a DL model or using the fusion feature of the model without making any changes to the standard model. And another conclusion is that, according to the study and results recorded in Tables 4 and 5, the AlexNet fusion approaches had better results than improved AlexNet approaches, mostly.

Also, this research contributes to feeding a clear image to many researchers around the world through this study, which contains the latest methods used to develop the AlexNet model or use it fusioned with other models for biomedical applications, which encourages the continuous development and innovation of modern methods to reuse AlexNet in other different applications.

The future work will be an extension of this study by updating the content of studies in the field of development or fusion of the AlexNet model, by adding a new approach based on fusion of AlexNet with other DL models for developing agricultural applications, and also by expanding the study into other fields such as agricultural applications or the field of public services.

Declaration

Conflict of interest the authors declare that they have no conflict of interest.

References

[1] Ertel, W. Introduction to Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2018.

[2] Rabi, Bahaa, et al. "Automatic Classification of Gastrointestinal Diseases Based on Machine Learning Techniques." 2019 29th International Conference on Computer Theory and Applications (ICCTA). IEEE, 2019

[3] Wang, Xizhao, Yanxia Zhao, and Farhad Pourpanah. "Recent advances in deep learning." International Journal of Machine Learning and Cybernetics, springer.11 (2020): 747-750.

[4] Han, X., Y. Zhong, L. Cao, and L. Zhang. 2017. “Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification.” Remote Sensing 9: 848.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

[5] Kumar, Ashnil, et al. "An ensemble of fine-tuned convolutional neural networks for medical image classification." IEEE journal of biomedical and health informatics 21.1 (2016): 31-40.

[6] Shen, Ranxin, Jiayi Wen, and Peiyi Zhu. "Blood Group Interpretation Algorithm Based on Improved AlexNet." Electronics 12.12 (2023): 2608.

[7] Kumar M, Arun, and Arvind Chakrapani. "Classification of ECG signal using FFT based improved AlexNet classifier." PLOS one 17.9 (2022): e0274225.

[8] He, Ming, et al. "An improved AlexNet model for automated skeletal maturity assessment using hand X-ray images." Future Generation Computer Systems 121 (2021): 106-113.

[9] Sadhana, S., and R. Mallika. "An intelligent technique for detection of diabetic retinopathy using improved AlexNet model based convolutional neural network." Journal of Intelligent & Fuzzy Systems 40.4 (2021): 7623-7634.

[10] Chen, Jun, Zhechao Wan, Jiacheng Zhang, Wenhua Li, Yanbing Chen, Yuebing Li, and Yue Duan. "Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet." Computer methods and programs in biomedicine 200 (2021): 105878.

[11] Lu, Siyuan, Shui-Hua Wang, and Yu-Dong Zhang. "Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm." Neural Computing and Applications 33 (2021): 10799-10811.

[12] Shakarami, Ashkan, Hadis Tarrah, and Ali Mahdavi-Hormat. "A CAD system for diagnosing Alzheimer's disease using 2D slices and an improved AlexNet-SVM method." Optik 212 (2020): 164237.

[13] Alaskar, Haya, et al. "The implementation of pretrained AlexNet on PCG classification." Intelligent Computing Methodologies: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part III 15. Springer International Publishing, 2019.

[14] Sathishkumar, R., et al. "An Improved Fusion Model from Google Net and AlexNet to Predict Breast Cancer using Deep Learning." 2023 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2023.

[15] Naseer, Iftikhar, et al. "Lung Cancer Detection Using Modified AlexNet Architecture and Support Vector Machine." Computers, Materials & Continua 74.1 (2023).

[16] Sethy, Prabira Kumar, et al. "Lung cancer histopathological image classification using wavelets and AlexNet." Journal of X-Ray Science and Technology 31.1 (2023): 211-221.

[17] Kibriya, Hareem, et al. "A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers." Computational Intelligence and Neuroscience 2022.1 (2022): 7897669.

[18] Muhammad Hussain, Nadia, et al. "Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data." Sensors 22.14 (2022): 5103.

[19] Abdelaziz, Ahmed., N., Alia. Skin Cancer Detection Using Deep Learning and Artificial Intelligence: Incorporated model of deep features fusion. Fusion: Practice and Applications, vol. 8, no. 2, 2022, pp. 08-15. DOI: https://doi.org/10.54216/FPA.080201

[20] Taneja, Harsh, Abhinav., Apoorv. Mangal, Himanshu. , Agarwal, Naman. Detection of Covid-19 using Cough Sounds. Fusion: Practice and Applications, vol. 7, no. 2, 2022, pp. 79-90. DOI: https://doi.org/10.54216/FPA.070202

[21] Atassi, Reem., Alhassan, Fuad., Djordjevic, Milan. A New Data Fusion Model for Medical Image Encryption in IoT Environment. Fusion: Practice and Applications, vol. 8, no. 1, 2022, pp. 16-26. DOI: https://doi.org/10.54216/FPA.080102

[22] Kaur, Surinder., Dinesh, Javalkar. Chaudhary, Gopal. , Khari, Manju. Breast Cancer Detection Using Deep Learning and Feature Decision Level Fusion. Fusion: Practice and Applications, vol. 8, no. 1, 2022, pp. 50-59. DOI: https://doi.org/10.54216/FPA.080105

[23] Z., Abedallah, Almajed, Rasha. , A., Saleh. Early Detection of Cardiovascular Diseases using Deep Learning Feature Fusion and MRI Image Analysis. Fusion: Practice and Applications, vol. 8, no. 2, 2022, pp. 16-24. DOI: https://doi.org/10.54216/FPA.080202

[24] Lee, Sheng Long, and Mohammad Reza Zare. "Biomedical compound figure detection using deep learning and fusion techniques." IET Image Processing 12.6 (2018): 1031-1037.

[25] Sinha, Divyanshu, et al. "Multimodal Deep Learning Analysis for Biomedical Data Fusion." Human Cancer Diagnosis and Detection Using Exascale Computing (2024): 53-69.

[26] Oza, Parita, et al. "Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey." Neural Computing and Applications, springer.34.3 (2022): 1815-1836.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024

[27] DeLiu, H.; Zhang, C.; Deng, Y.; Xie, B.; Liu, T.; Zhang, Z.; Li, Y.F. TransIFC: Invariant cues-aware feature concentration learning for efficient fine-grained bird image classification. IEEE Trans. Multimed. 2023, 1–14.

[28] Krichen, Moez. "Convolutional neural networks: A survey." Computers 12.8 (2023): 151.

[29] Sandia, Tomasz. "Review and comparison of commonly used activation functions for deep neural networks." Springer. Bio-inspired neurocomputing (2021): 203-224.

[30] Naseri, H.; Mehrdad, V. Novel CNN with investigation on accuracy by modifying stride, padding, kernel size and filter numbers. Multimed. Tools Appl. 2023, 82, 23673–23691.

[31] Tian, C.; Xu, Y.; Zuo, W.; Du, B.; Lin, C.W.; Zhang, D. Designing and training of a dual CNN for image denoising. Knowl. -Based Syst. 2021, 226, 106949.

[32] Jaramillo-Munera, Y.; Sepulveda-Cano, L.M.; Castro-Ospina, A.E.; Duque-Muñoz, L.; Martinez-Vargas, J.D. Classification of Epileptic Seizures Based on CNN and Guided Back-Propagation for Interpretation Analysis. In Proceedings of the International Conference on Smart Technologies, Systems and Applications, Cuenca, Ecuador, 16–18 November 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 212–226.

[33] Ndong, P.S.B.; Adoni, W.Y.H.; Nahhal, T.; Kimpolo, C.; Krichen, M.; Byed, A.E.; Assayad, I.; Mutombo, F.K. A face-mask detection system based on deep learning convolutional neural networks. In Advances on Smart and Soft Computing: Proceedings of ICACIn 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 273–283.

[34] Hui, T.W.; Tang, X.; Loy, C.C. A lightweight optical flow CNN—Revisiting data fidelity and regularization. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 43, 2555–2569.

[35] Y. Keizhevsk, I Stuskever, E Hinton G, “ImageNet classification with deep convolutional neural networks[C].”, in Proceedings of the Advances in Neural Information Processing Systems. South Lake Tahoe, US, 2012, pp. 1097–1105.

[36] Guang-Bin H et al (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501.

[37] L. Yang, X. Ban, Y. Li, and G. Yang, “Multiple features fusion for facial expression recognition based on ELM,” Int. J. Embed. Syst., vol. 10, no. 3, pp. 181–187, 2018.

[38] El-Douh A, Lu S, Abdelhafeez A et al (2023) A neutrosophic multi-criteria model for evaluating sustainable soil enhancement methods and their cost 2 implications in construction. SMIJ 5(2):11.

[39] A. Abdelhafeez, H. K. Mohamed, A. Maher and N. A. Khalil, "A novel approach toward skin cancer classification through fused deep features and neutrosophic environment", *Frontiers Public Health*, vol. 11, pp. 1-15, Apr. 2023.

DOI: https://doi.org/10.54216/FPA.170210
Received: January 27, 2024 Revised: April 21, 2024 Accepted: September 22, 2024