

Review of AlexNet for Medical Image Classification

Wenhao Tang¹, Junding Sun^{1,*}, Shuihua Wang², Yudong Zhang^{1,3,†}

¹1st School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P.R. China

²2nd Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

³3rd School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China

Abstract

In recent years, the rapid development of deep learning has led to a wide range of applications in medical image classification. The variants of neural network models with ever-increasing performance share some commonalities: to try to mitigate overfitting, improve generalization, avoid gradient vanishing and exploding, etc. AlexNet first utilizes the dropout technique to ease overfitting and the ReLU activation function to prevent vanishing gradient. Therefore, we focus on AlexNet, which initially contributed significantly to Convolutional Neural Networks (CNNs) research in 2012. After reviewing over 100 papers, including those from journals and conferences, we give a narrative on the technical details, advantages, and application areas of AlexNet.

Keywords: Medical Image Classification, ReLU, Neural Networks, Gradient Vanishing, CNNs

Received on 15 November 2023, accepted on 18 December 2023, published on 1 December 2023

Copyright © 2023 W. Tang *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transforming, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetel.4389

1. Early Development of Deep Learning

During the development of Deep Learning (DL), artificial neural networks and machine learning techniques have made great strides, leading to the emergence of powerful deep neural networks. The evolution of DL has gone through four stages:

Early Developments: The concept of artificial neural networks dates back to the 1940s and 1950s, with the development of perceptrons by Frank Rosenblatt [1]. However, early neural networks had limitations, and research stagnated. The perceptron, a single-layer neural network, could only model linearly separable functions, limiting its practical applications.

Revival with Backpropagation: In the 1980s, the field of neural networks experienced a resurgence with the development of the backpropagation (BP) algorithm, which allowed for efficient training of multi-layer neural networks [2]. Researchers such as Geoffrey Hinton and Yann LeCun contributed significantly during this period. However, due to the disappearance of gradients, deep

networks still face challenges in training and scaling, where gradients become weaker and weaker as they propagate layer by layer until they eventually disappear, no longer contributing to the learning process [3].

DL Renaissance: The DL Renaissance, occurring primarily in the 2010s, marked a transformative period in the field. Breakthroughs in neural network architectures, increased computational power, the availability of large datasets, and the introduction of deep learning frameworks fueled significant advancements. Section 2 will present detailed content related to the DL Renaissance.

Breakthroughs and Applications: The content related to breakthroughs in DL will be introduced in Section 3. The success of AlexNet demonstrated the power of deep convolutional neural networks (CNNs) for image recognition. In the following years, DL methods have achieved remarkable results in many areas such as natural language processing, speech recognition, robotics, and self-driving cars. This led to the widespread adoption of DL techniques and advanced deep neural network architectures.

Our paper is structured as follows: Section 2 focuses on two essential innovations in DL and describes the reasons

*Corresponding author. Email: sunjd@hpu.edu.cn

†Corresponding author. Email: yudongzhang@ieee.org

and conditions that have led to its revival. Then, Section 3 describes how the success of AlexNet has led to breakthroughs in DL. Moreover, in Section 4, we discussed several technologies that AlexNet first used, which improved its performance. Then, in Section 5, we discussed the main operational processes and constituent structures of convolutional layers. The role and some concepts of ReLU activation function and dropout technology in improving model performance are mentioned in Sections 6 and 7, respectively. Next, in Sections 8 and 9, we successively introduce various variants of AlexNet and summarize and discuss the application areas of AlexNet. We will specifically discuss the application of AlexNet in medical diagnosis in Section 10. Finally, Section 11 mentions the conclusions we have drawn based on the specific content of this article.

2. Renaissance of Deep Learning

There is resistance to the DL journey. Although the emergence of backpropagation algorithms has brought new opportunities for developing neural networks, the difficulty of training and tuning [4] has gradually increased with the rise of deep networks. Traditional DL methods rely on large amounts of labeled data [5], which comes at a significant cost, which reflects the training challenge. Researchers started exploring different solutions to solve the problem of vanishing gradients. Among them, the proposal of ResNet opens a new path for deep network training. One of ResNet's innovations is skip connection. With X denoting the output of l -th layer, $F(X)$ denoting the output of the weighted $(l + 1)$ -th layer with skip connection added, the output of $(l + 1)$ -th layer is:

$$F'(X) = X + F(X). \quad (1.)$$

By introducing a skip connection (Figure 1), the residual structure breaks the layer-stacking approach of the traditional network so that the gradient can be directly transferred between different layers, thus effectively alleviating the problem of gradient vanishing [6].

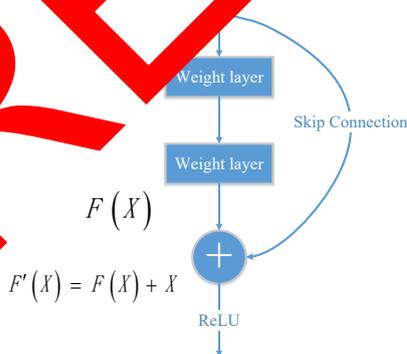


Figure 1. Diagram of skip connection.

The Batch Normalization (BN) technique appeared during this period. This technique mitigates the internal covariate bias problem by normalizing the inputs of each layer in the network [7], which speeds up the training of the network and improves the generalization ability of the model. However, it harms the model's robustness against adversarial examples [8]. Kong, *et al.* [8] further explained the vulnerability of batch normalization in terms of two aspects, such as intra- and inter-class feature distances. These neural networks affect complex model fitting and distribution approximation that traditional statistical methods cannot achieve. It has excellent potential for development compared to many machine learning algorithms [2]. It has also facilitated many innovations in the field of DL, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNN) [9]. These network architectures have achieved remarkable success in many tasks, such as computer vision and natural language processing [1]. Assume that the input of the BN layer is $B = \{x_1 \dots x_m\}$, and γ and β are learnable parameters. Then

$$y_i = \text{BN}_{\gamma, \beta}(x_i). \quad (2.)$$

The computational process of BN follows the following steps: first,

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i, \quad (3.)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \quad (4.)$$

where μ_B and σ_B^2 are the mean and variance of B , respectively. Then, normalizing x_i using mean μ_B and variance σ_B^2 , namely

$$\tilde{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}. \quad (5.)$$

The purpose of normalization is to regularize the data to a uniform interval, reduce the degree of dispersion of the data, and reduce the learning difficulty of the network. Finally, using γ and β as the reduction parameters to preserve the distribution of the original data to some extent:

$$y_i = \gamma \tilde{x}_i + \beta. \quad (6.)$$

ResNet and BN are two critical innovations for solving problems related to DL. The DL renaissance began in the 2010s when researchers devised techniques to address the challenges of training deep neural networks. Notably, in 2006, Hinton, *et al.* [10] introduced the Restricted Boltzmann Machine (RBM) and deep belief networks (DBNs), paving the way for the resurgence of DL, additionally developing the ReLU activation function, dropout regularization, and advances in hardware (e.g., GPUs) pretty accelerated training deep networks.

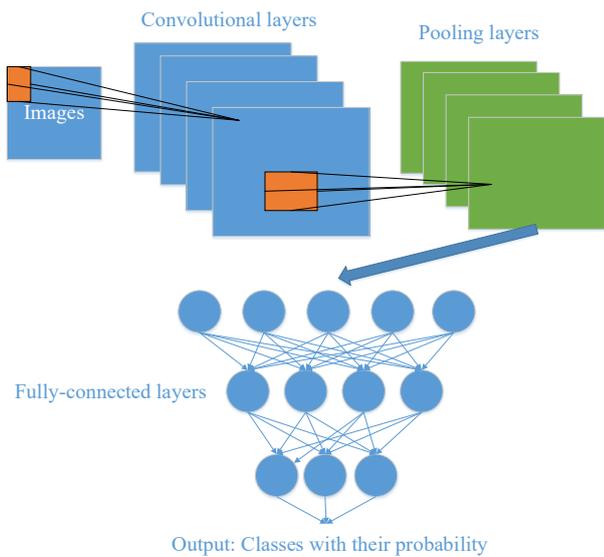


Figure 2. The simple architecture of CNN.

Today, DL is a dominant paradigm in artificial intelligence and machine learning, enabling state-of-the-art performance in tasks like image classification (achieved excellent results since 2012 [11]), language translation [5] and autonomous driving [12]. The field continues to evolve with ongoing research into advanced architectures, transfer learning (TL), reinforcement learning, and the application of DL to new domains [5], promising continued innovation and transformative breakthroughs.

3. AlexNet: A Groundbreaking Convolutional Neural Network

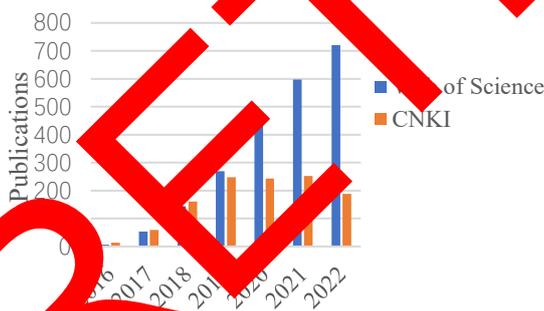


Figure 3. Publication record for topic-relevant publications related to AlexNet in Web of Science and CNKI from 2016 to 2022.

The breakthrough moment for DL came in the 2010s, driven by several key developments, including AlexNet's victory in the 2012 ImageNet competition [13], which showcased the power of Deep Convolutional Neural

Networks (DCNNs) for image recognition. As shown in Figure 3, there has been a continuous rise in the circulation of publications related to the AlexNet topic through topic-based searches on Web of Science and China National Knowledge Infrastructure (CNKI). This year, the number of topic-relevant publications related to AlexNet exceeds 500 and 200, respectively. We can also see from Table 1 that the years with the highest number of publications about AlexNet on CNKI and Web of Science are 2021 and 2022, which amounted to 252 and 716, respectively. In addition, many related works are not in these two comprehensive paper search sites. The above shows that AlexNet, in DL, is a hot research object.

Table 1. AlexNet relevant publications from 2016 to 2022.

Year	Database	
	Web of Science	CNKI
2016	7	14
2017	53	59
2018	144	161
2019	269	248
2020	467	243
2021	597	252
2022	721	188

AlexNet is an extensive network structure with 60 million parameters and 650,000 neurons [14]. It significantly outperformed traditional methods, drew attention to the potential of DL, and laid the foundation for the subsequent development of Deep Convolutional Neural Network (DCNN). AlexNet represented a significant leap forward in image classification tasks. The network's architecture was a DCNN consisting of eight layers. The eight layers are five convolutional layers and three Fully Connected (FC) layers. The role of the softmax layer is to control the output in the range of (0,1) to ensure neuron activation [14]. Eq. 7 [14] expressed the normalization process of the softmax layer.

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)}, \quad (7.)$$

where n is the dimension of the input tensor and x_i is also its i -th predicted value. One of AlexNet's groundbreaking innovations was using Rectified Linear Units (ReLU) as activation functions, greatly accelerating training convergence. AlexNet also effectively used data augmentation and dropout techniques, which helped prevent overfitting. AlexNet has had considerable significance in previous generations of CNNs and has influenced many subsequent architectures and innovative research in CNNs [15].

4. Architecture of AlexNet

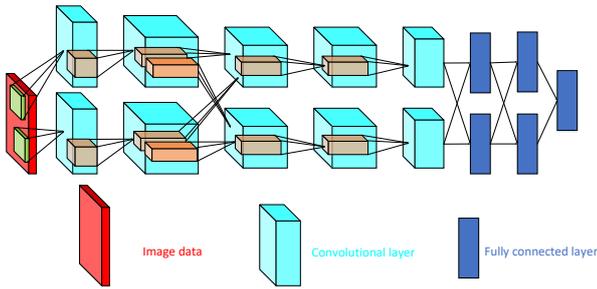


Figure 4. Architecture of AlexNet.

As shown in Figure 4, AlexNet [13] consists of five convolutional layers, with the first two being followed by max-pooling layers. Its convolutional layers extract hierarchical features from input images. They employ relatively small filter sizes (e.g., 3×3 and 5×5) and utilize many filters (e.g., 96 and 256) in the early layers. These layers are responsible for learning low-level features, such as edges and textures, and gradually building up to more complex representations.

ReLU activation functions are applied after all convolutional layers and the first two FC layers. The main advantage of ReLU is its improved gradient propagation and it produces fewer problems with gradient vanishing [16]. ReLU introduces non-linearity to the network by setting negative values to zero while all positive values to pass unchanged [17]. This non-linearity promotes learning complex and hierarchical features and accelerates training by mitigating the vanishing gradient problem. Moreover, in neural networks activation functions add nonlinear properties so that the network can learn and perform more complex tasks. ReLU, although widely used, still suffers from the computational complexity of training multi-layer fully connected ReLU neural networks [18].

AlexNet introduced Local Response Normalization (LRN) to enhance the network's generalization ability. LRN normalizes the responses of neighboring neurons, promoting competition between neurons and improving the network's robustness to variations in input data. The following is the formula for LRN:

$$b_{x,y}^i = \frac{a_{x,y}^i}{\left(k + \alpha \sum_{j=\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a_{x,y}^j)^2 \right)^\beta}, \quad (8)$$

where i denotes the i -th channel, x and y denotes the positional coordinates of this value to be normalized, k serves to prevent the occurrence of division by 0, α and β are constants; n denotes the range of the neighborhood, with the boundary cases complemented by 0.

However, BN [19] techniques emerged later to improve higher performance than LRN. The core idea of BN is to constrain the change in gradient to a transferable range,

preventing the gradient from vanishing or exploding. BN attempts to achieve a stable distribution of activation values throughout the training process by subtracting the batch mean and dividing by the batch standard deviation [20]. Jain, et al. [21] proposed a new high-performance convolutional Siamese network discussing the problem of offline signature verification. In this model, they used the BN technique instead of local response normalization to get better accuracy.

For AlexNet, three max-pooling layers follow the first three convolutional layers. These pooling layers downsample the spatial dimensions of the feature maps, reducing computational complexity while retaining essential information. Max pooling selects the maximum value within a local region to create a downsampled representation. After the convolutional and pooling layers, AlexNet incorporates three FC layers. These layers are traditional artificial neural network layers where all neurons in every layer have links with every neuron in the next layer. A softmax activation function is behind the final FC layer, which produces class probabilities, making it suitable for image classification tasks.

AlexNet introduced dropout as a regularization technique to prevent overfitting. Dropout randomly deactivates a fraction of neurons during training, reducing dependencies between neurons and improving model generalization. AlexNet was designed to take advantage of parallel processing capabilities and was one of the first DNNs to do so effectively. It utilized two GPUs during training, which allowed it to handle the large number of parameters and computations involved in DL.

5. Convolutional Layers

The primary operation performed in convolutional layers is convolution. Convolution involves sliding a set of learnable filters (kernels) over the input data. These filters are small grids that have weights associated with each element. As shown in Figure 5, the convolution operation calculates the dot product between the filter and the local region of the input data.

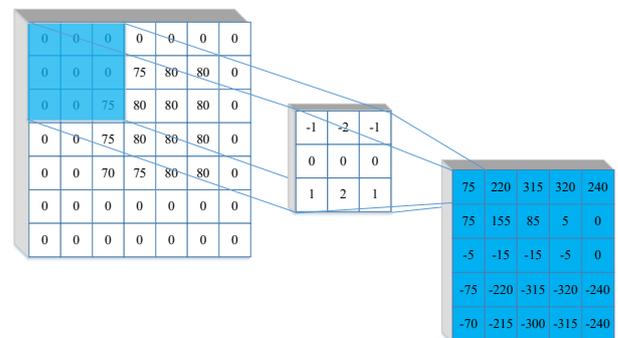


Figure 5. A simple graphical representation of the convolution operation.

This operation is repeated across the entire input to produce feature maps. In CNNs, the filters are learnable parameters updated during training through backpropagation. Each filter detects specific patterns or features in the input data. For example, some filters may specialize in detecting edges, while others may examine textures, corners, or other visual features. As the network learns, filters become increasingly specialized and optimized for the given task. Convolutional layers have a hierarchical organization. Early layers focus on extracting low-level features from the input data, such as edges and simple textures. These features serve as building blocks for more complex representations. As the layers progress, they learn higher-level features and patterns, eventually capturing abstract concepts related to the task. The feature map of the original size n will probably change after it passes through the convolutional layer with kernel size k . The following equation describes the process of 2D convolution:

$$W = \left\lfloor \frac{n-k+2p}{s} \right\rfloor + 1, \quad (9.)$$

where W is the size of the 2D post-convolution feature map, k is the size of the convolution kernel, s is the pooling unit move step, p is the number of laps of the expanded pixels.

Often, after the convolution operation, a pooling layer and an activation function (usually ReLU) follow. Pooling layers, such as max-pooling, downsample the spatial dimensions of the feature maps while retaining important information. Activation functions introduce non-linearity to the network, allowing it to learn complex relationships. ReLU, in particular, is widely used in convolutional layers to activate positive values and set negative values to zero, promoting sparsity and faster convergence.

6. ReLU Function

The ReLU activation function is fundamental to specific artificial neural networks. Deep neural networks are a representation of them. It plays a crucial role in introducing non-linearity to neural networks, enabling them to learn complex relationships in data. Here's an overview of the ReLU activation function in four paragraphs: ReLU is a simple mathematical function that operates on an input value x . As shown in Figure 6, it returns x if x is greater than or equal to zero and zero otherwise. The formula can describe this as follows:

$$y = \begin{cases} x, & x \in [0, +\infty) \\ 0, & x \in (-\infty, 0) \end{cases} \quad (10.)$$

where y and x are the output and input of ReLU, respectively. Eq. 10 denotes that ReLU replaces all negative input values with zero and leaves positive values unchanged. This piecewise linear behavior introduces non-linearity into the network, allowing it to approximate complex functions and learn intricate patterns.

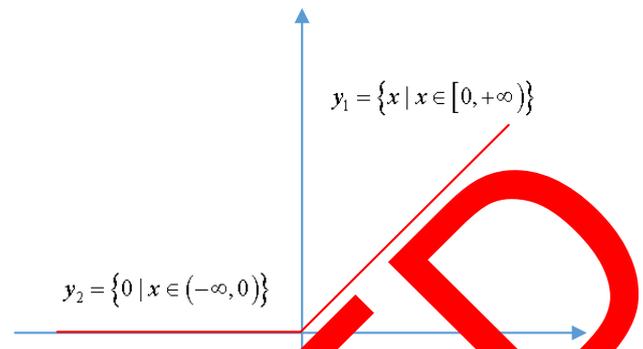


Figure 6. Graphical representation of ReLU.

ReLU has several advantages, making it a popular choice as an activation function in neural networks. Firstly, it is computationally efficient because it involves a simple thresholding operation. Secondly, ReLU helps alleviate the vanishing gradient problem, which can hinder the training of deep networks. In traditional activation functions like sigmoid and tanh, gradients can become very small in the presence of deep architectures, slowing down or preventing learning. ReLU mitigates this problem by maintaining a constant gradient for positive input values, ensuring faster convergence during training. Lastly, ReLU activation function sparsity—setting negative values to zero—can lead to more efficient representations, encouraging the network to focus on relevant features.

Over the years, several variants and modifications of ReLU appeared to address its limitations. One such variant is the Leaky ReLU, which allows a slight, non-zero gradient for negative input values to address the "dying ReLU" problem, where neurons can get stuck with zero gradients and cease to update. Another variant is the Parametric ReLU (PReLU) [22], which introduces a learnable parameter α to adjust the slope of the negative side of the function. This can be particularly useful when the optimal slope may vary across different neurons and layers. And the following is the equation of PReLU:

$$f_p(x) = \max(0, x) + \alpha \cdot \min(0, x), \alpha \in [0, 1], \quad (11.)$$

where, $f_p(x)$ is the output of the PReLU function, $f(x) \in (0, +\infty)$, x is its input.

There are also variants like Exponential Linear Unit (ELU) [23] and Swish [24], each designed to offer different advantages in terms of convergence and robustness. For the ELU function, the output $f_E(x)$ grows linearly for input values x not less than 0. Conversely, $f_E(x)$ decreases exponentially. This property makes ELU better than ReLU, especially when dealing with negative inputs. The ELU function has the following formula:

$$f_E(x) = \begin{cases} x, & x \geq 0 \\ e^x - 1, & x < 0 \end{cases} \quad (12.)$$

The main advantage of the Swish function is that it has a high-speed convergence rate and effectively avoids the problem of vanishing and exploding gradients. The formula for the Swish function is as follows:

$$f_S(x) = x * \text{sigmoid}(x), \quad (13.)$$

where sigmoid and * denotes the Sigmoid function and the element-wise multiplication operation, respectively. $f_S(x)$ and x are the output and input of the Swish function, respectively.

ReLU is an activation function in various neural network architectures, including CNNs for computer vision tasks, recurrent networks for natural language processing, and fully connected networks for multiple applications. It has been a critical factor in the success of DL, enabling networks to learn complex representations and achieve state-of-the-art results in tasks like image classification, object detection, and more. While ReLU is highly effective, choosing the appropriate activation function sometimes depends on the specific problem and architectural considerations. Researchers continue to explore alternatives and improvements in activation functions to address various challenges in DL.

7. Dropout

As shown in Figure 7, dropout works by randomly deactivating a fraction of neurons during each training iteration [25]. But dropout cannot be helpful with just a single hidden node [25]. During forward and backward passes, dropout randomly sets a subset of neuron activations to zero with a specified probability, typically between 0.2 and 0.5. The process means some neurons are "dropped out" or temporarily removed from the network for that specific iteration. Dropout introduces noise and uncertainty into the training process, forcing the network to rely on different neurons in each iteration.

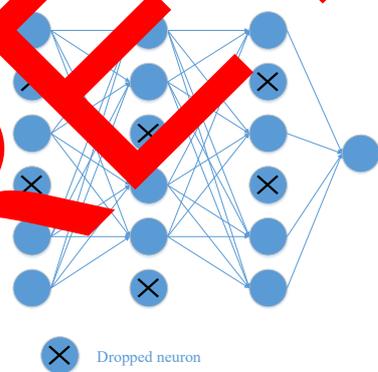


Figure 7. Simple graphical representation of dropout. The figure explicitly shows that dropped neurons have lost connections to neurons in neighboring layers.

Overfitting occurs when a neural network learns to memorize the training data rather than generalize. It leads to poor performance on unseen data. Dropout combats overfitting by preventing co-adaptation of neurons [26]. When randomly dropping out neurons during training, the network cannot rely on any single neuron or small group of neurons to make predictions. Instead, it must distribute the learning across a more diverse set of neurons, encouraging the network to learn robust and generalizable features.

Dropout is an ensemble learning technique to some extent. Multiple subnetworks with different neuron configurations are sampled from the whole network during training. By aggregating the predictions of these subnetworks during inference (usually by scaling the activations by the dropout probability at test time), dropout effectively creates an ensemble of networks. Ensemble learning tends to reduce model variance and improve overall performance.

Variants of dropout have been proposed, including dropout applied to different network layers, such as dropout in convolutional layers (DropBlock) [27], and more. These variants adapt dropout to specific network architectures and tasks. Dropout is a powerful and widely used regularization technique that helps neural networks generalize better to unseen data, making them more robust and less prone to overfitting [28]. For a linear neural network without dropout, its error E_N can be expressed as:

$$E_N = \frac{1}{2} \left(t - \sum_{i=1}^n w_i I_i \right)^2, \quad (14.)$$

where w is the set of weights of the neural network layer, t is the target value, I is the set of inputs to the neural network layer. After adding the dropout layer, the error will change as the following:

$$\begin{cases} E'_N = \frac{1}{2} \left(t - \sum_{i=1}^n w'_i I_i \right)^2, \\ w' = pw \end{cases}, \quad (15.)$$

where E'_N is the changed error, and p is the dropout rate. Dropout has been instrumental in the success of DL in various domains, from computer vision to natural language processing, and remains a valuable tool for improving model performance and reliability.

8. Variants of AlexNet

The classical structure of AlexNet laid the foundation for the field of DL, while some researchers have made some modifications and improvements to it, resulting in different variants. Some common variants of AlexNet, for example, the earlier appearing ZFNet, are improvements on AlexNet proposed by Zeiler, *et al.* [29]. It builds on AlexNet with some tweaks aimed at improving performance and a better

understanding of the hierarchical structure of images. Subsequently, Simonyan, *et al.* [30] proposed VGGNet with a deeper network structure. Compared to AlexNet, VGGNet has a characteristic that uses smaller convolutional kernels and deeper networks to increase the depth of the network. In the same year GoogLeNet, also known as Inception v1, was proposed by Szegedy, *et al.* [31]. It introduced the "Inception module" to increase the computational efficiency of the network by using different-sized convolutional kernels in parallel. Later, Microsoft Research proposed ResNet, which made it easier to train deep structures by introducing a residual learning mechanism. ResNet was designed to solve the problem of gradient vanishing and gradient explosion during deep neural network training. Also, SqueezeNet [32] is a lightweight neural network designed to reduce the model's size while maintaining good performance. It reduces the number of parameters by employing a 1×1 convolutional kernel, making it suitable for resource-limited environments. The later proposed DenseNet is a DL architecture proposed by Huang, *et al.* [33]. Compared to traditional CNNs, DenseNet utilizes a densely connected structure, which connects each layer's output to the outputs of all previous layers. This densely connected design helps to enhance gradient flow, facilitates information sharing, and effectively mitigates the gradient vanishing problem. All of these networks are based on CNN and consist of convolutional layers, pooling layers, etc.; all of them are deeper than earlier convolutional networks and, therefore, more effective in extracting features; all of them have been trained and validated on large-scale image datasets, and all of them have better performance. These networks differ in network structure, parameter utilization, depth, and application scenarios.

9. Applications

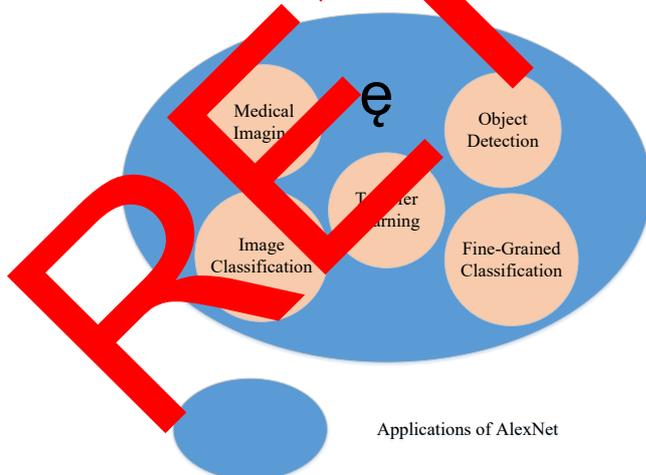


Figure 8. A schematic diagram of the application (not fully listed) of AlexNet.

As shown in Figure 8, we will know the application aspects of AlexNet, which include medical imaging, image classification, TL, object detection, and fine-grained classification. The following account addresses these aspects.

Image Classification: AlexNet's primary and most well-known application is image classification. It demonstrated remarkable performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [34], significantly outperforming previous methods. This success established the effectiveness of deep CNNs for image recognition tasks. Since then, similar architectures have broadly applied to various image classification tasks, including identifying objects in images, recognizing faces, and classifying medical images. For facial recognition, for example, Ghazal, *et al.* [35] developed an innovative transfer learning AlexNet model to detect the presence or absence of a smile through children's facial features. This model extracts image features using a convolutional layer of pre-trained AlexNet. These features to train a new FC layer. Comparison experiments on a dataset of 2940 images of children with autism from Kaggle show that their model has higher accuracy. Further, Balashanmugam, *et al.* [36] applied AlexNet to classify iris images of the eye, achieving a high correct recognition rate on 163,432 iris images, providing valuable insights into the development of biometrics. Kayadibi, *et al.* [37] applied pre-trained AlexNet to eye condition detection, and their experimental results outperformed the traditional pre-trained deep graph convolutional network.

Object Detection: AlexNet's architecture, particularly its convolutional layers, has been the foundation for developing object detection models. Researchers have adapted and extended the principles of AlexNet to create object detection frameworks like Fast R-CNN [38] and Faster R-CNN [39], which can classify objects and localize them within images. These advancements have applications in autonomous driving, surveillance, natural disaster warning [40], robotics [41, 42], etc.

Fine-Grained Classification: Fine-grained classification involves distinguishing between similar categories within a broader class. AlexNet's capability to learn intricate features has made it suitable for fine-grained classification tasks, such as identifying specific species of birds [43], plant varieties [44, 45], or classification of canine maturity and bone fracture time [46]. Researchers have built upon the principles of AlexNet to develop specialized models for these tasks.

Medical Imaging: AlexNet's CNN architecture was related to medical imaging for various diseases. It has been used for the classification of medical images, such as X-rays, MRIs, CT scans, and electroencephalograms, aiding in the diagnosis of diseases, including cancer [47], neurological disorders [48], and schizophrenia [49]. By adapting AlexNet to medical data, researchers and healthcare professionals have improved accuracy and efficiency in disease detection and diagnosis.

TL: One of the significant applications of AlexNet is in

TL. Researchers and practitioners leverage pre-trained AlexNet models, fine-tuning them for specific tasks or domains [50, 51]. This approach efficiently uses pre-trained feature extractors to solve new and related problems with smaller datasets. Sharma, *et al.* [52] demonstrated experimentally that the hierarchical fine-tuning method has potential for development. TL with AlexNet has enabled advances in various fields, including natural language processing [53, 54], where researchers used the network's features to extract meaningful representations from text data.

In video analytics, the CNN structure of AlexNet has broad applications for dynamic scene understanding. For video classification tasks, researchers use AlexNet to extract critical features in videos for efficient action recognition and event classification. For example, Peng, *et al.* [55] added additional convolutional and pooling layers to AlexNet and tuned the parameters to learn subpixel-level skeletal image features generated in a certain way. They carefully fine-tuned AlexNet, and their experimental results on the dataset outperformed even VGG16 and GoogLeNet. As another example, Rafiq, *et al.* [56] used a pre-trained AlexNet and added three additional fully-connected layers to it, applying it to a video scene classification task. They found that batch normalization after the convolutional layers did not improve model performance, so they did not use this technique. Their improvement is relatively simple, although it capitalizes on the benefits of the transfer learning approach and achieves better results on smaller datasets. However, for classification tasks with more video duration and frames, the pre-trained AlexNet, with fewer parameters, may be challenging to handle when more improvements are inexistent. However, simply adding subtracting layers or tuning parameters to the original model may not improve the model's generalization over larger datasets.

Further, Minhas, *et al.* [57] applied a DL scheme containing AlexNet to classify sports video footage. Through experiments, they proved that their method performs better than some machine learning methods and standard CNNs. The experiments they did were more detailed and specialized. However, since there is no significant improvement to AlexNet, it may be difficult for their method to perform more complex tasks. In addition, AlexNet's sparse temporal feature learning capability may perform well in video behavior analysis. These applications can suit application scenarios such as video surveillance and targeted placement of video advertisements. For example, Mumtaz, *et al.* [58] proposed a novel Deep Multi-Network (DMN) architecture based on AlexNet and GoogLeNet for video violence detection. They integrated two pre-trained networks to build a fast learning system with high video classification accuracy. Their experiment results demonstrated that the DMN performs better in optimal accuracy and trains faster than AlexNet and GoogLeNet. However, DMN is not as good as AlexNet and GoogLeNet in some result metrics. The shortcoming may indicate that DMN suffers from information loss or an increased risk of overfitting in fusing two CNN outputs.

Secondly, the MobileNet model proposed by Irfanullah, *et al.* [59] performs excellently on the field hockey game dataset. They used the Background subtraction method in the preprocessing stage. This method subtracts the current frame from the background reference model to compute the motion region efficiently.

Further, Khan, *et al.* [60] used a fine-tuned AlexNet for crowd anomaly detection in video frames. This fine-tuned AlexNet with one less convolutional layer extracts features on three anomaly detection public datasets and inputs them to six classifiers for comparative analysis. Among them, the softmax classifier has better performance. AlexNet also has expanded applications to video coding. For example, Imen, *et al.* [61] combined LeNet-5 and AlexNet. Their approach reduces the computational time used to examine block decision candidates. Improvements they made to LeNet-5 were to avoid gradient vanishing and to accommodate multi-label classification. While AlexNet did not change its parameters but downsampled its inputs. Their method reduces the complexity of the HEVC encoder. It also reduces the video processing time and improves the compression efficiency more than related work.

AlexNet uses real-time object recognition and tracking in virtual and augmented reality applications. By combining camera data and features extracted by AlexNet, virtual elements can interact more accurately with the actual scene. For example, the approach proposed by Hung, *et al.* [62] provides physicians with more effective machine-assisted treatment options. They tackle the challenge of capturing the most relevant visual details for better video evaluation through assisted supervision and attention mechanisms.

Although AlexNet is primarily designed for image classification, transfer learning opens up new possibilities for speech recognition tasks. By utilizing the high-level features learned by AlexNet in image recognition, researchers can improve the performance of speech recognition systems and increase the understanding and classification accuracy of speech signals. For example, Ziafat, *et al.* [63] fine-tuned a pre-trained AlexNet by removing the last three layers and did comparison experiments with DCNN and bidirectional Long Short-term Memory (LSTM). The experimental results show that AlexNet, with transfer learning, outperforms the other networks.

AlexNet's fine-grained classification capability in agriculture uses crop yield prediction, plant stress detection, weed and pest detection, disease detection, smart agriculture [64], and sugarcane quality detection [65]. By deploying image acquisition devices in farmland, AlexNet can recognize the health condition of plants [66] and help farmers take timely measures to improve the yield and quality of crops. Zhang, *et al.* [67] developed a targeted shake-and-catch apple harvesting technique to cope with the high cost of manual apple picking. Tarek, *et al.* [68] combined pre-trained AlexNet and an improved Gray Wolf optimization method for early diagnosis of plant diseases. Their model achieved better results on relevant datasets.

AlexNet and its derived models are critical in natural

disaster monitoring [69, 70]. By analyzing satellite images or data captured by drones, AlexNet can support disaster assessment [71] and emergency relief decisions [72]. The above shows that it has great potential for mitigating the impact of disasters and improving response capabilities.

10. Medical Image Classification Applications

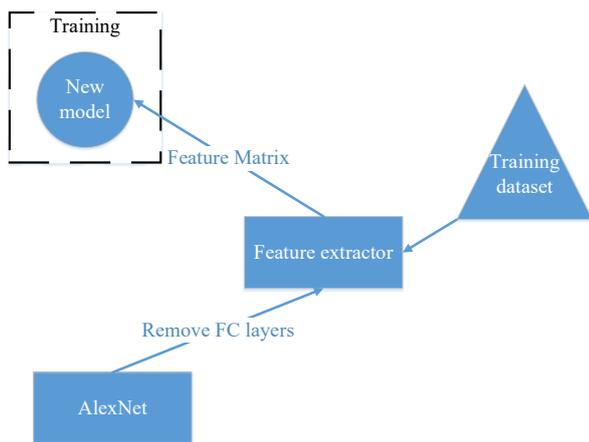


Figure 9. A simple illustration of using AlexNet as a feature extractor using TL.

At present, DL has broad applications in the field of medical disease diagnosis. AlexNet has applications in several tasks, but not just these. In these following applications, AlexNet is a feature extractor in a TL (Fig. 9) manner, and the extracted features are the input of new models.

Disease Diagnosis: AlexNet is used to diagnose various medical conditions based on imaging data. For example, it has been used in the analysis of radiological images such as X-rays, Magnetic Resonance Imaging (MRI) [73], and Computed Tomography (CT) scans to detect diseases like diseases of the abdomen [74], diseases of the chest [74], diseases of the sexual organs [74], breast cancer [75]. By training on large datasets of medical images, AlexNet can learn to identify patterns and abnormalities indicative of specific diseases, assisting radiologists and clinicians in making more accurate diagnoses. After a survey of 102 medical studies, Morid, *et al.* [76] stated that AlexNet is the most common model for brain MRI and mammography. AlexNet is also essential to the deep models applied to diagnose brain tumors. The fine-tuned AlexNet has produced good brain tumor dataset results [77]. However, AlexNet also has limitations in diagnosing brain tumors. For example, Ramya, *et al.* [78] proposed MIDNet18 CNN, which is more profound than AlexNet, and experimentally proved that this model performs better than AlexNet on specific brain tumor image datasets. Deeper layers allow the CNN to extract richer features, and AlexNet does not

serve as well as deeper CNNs on some datasets due to its shallow layers. AlexNet can be used as a feature extractor with several well-known machine-learning classifiers to improve brain tumor image classification [79].

Lesion Detection and Segmentation: AlexNet uses lesion detection and segmentation in medical images. For instance, it can assist in identifying and delineating skin lesions, including moles, tumors [80], and other dermatological conditions in dermatology. Similarly, in the context of medical imaging, AlexNet can aid in identifying and delineating lesions or anomalies within organs [81-83], tissues [84], or anatomical structures [85], which is essential for treatment planning and disease monitoring.

Ophthalmology: AlexNet has shown promise in ophthalmology, particularly in diagnosing and monitoring eye diseases [86]. It can analyze retinal images, such as fundus photographs [87], to detect signs of diabetic retinopathy [88], glaucoma [89], and age-related macular degeneration [90]. By automatically identifying abnormalities and disease-related features in these images, AlexNet assists ophthalmologists in diagnosing and tracking disease progression early.

Pathology and Histology: In pathology and histology, AlexNet can be applied to the analysis of tissue slides, assisting pathologists in diagnosing diseases like cancer [91]. It can help identify and classify cell types, tissue structures, and pathological conditions in biopsy samples.

Automating this process through DL models like AlexNet can improve diagnostic accuracy and reduce the workload of pathologists. For example, Abbas, *et al.* [92] proposed a deep-learning model for recognizing eye diseases using self-attention and dense layers. This model was a variant improved from AlexNet, with the main improvement being the addition of attention and dense layers. They used effective image enhancement techniques such as CLAHE and Bengram. These ways effectively preprocess the dataset. The methods combined with the improved pre-trained AlexNet proved to have high overall efficiency and accuracy, as demonstrated by the experimental results on the ocular disease dataset.

Further, Chan, *et al.* [93] extracted image features with pre-trained AlexNet, VGG16, and GoogleNet to classify ocular diseases [93]. They fused the last FC layer of AlexNet, the first FC layer of Vgg16, and the FC layer of GoogleNet to be used as a decision model. They also reduced the feature space using majority voting. The manipulation ignores some interrelated features and retains the most crucial information. However, this method also results in some data loss.

Overall, the application of AlexNet to medical imaging tasks has the potential to enhance the accuracy and efficiency of disease diagnosis, lesion detection, and image analysis in various medical specialties. Based on the fact that most of the studies on MRI images of the brain and mammography images using AlexNet as a DL model have achieved better results, Morid, *et al.* [76] argued that the best way to conduct such studies is to use shallow CNN models with large convolutional kernel sizes. As DL

techniques continue to advance, or with the help of expanding datasets, improved GPU performance, data preprocessing techniques, and growing medical image types to further enhance the performance of CNNs [94], all of these applications are likely to play an increasingly important role in improving medical outcomes and patient care.

11. Conclusions

In conclusion, AlexNet, with its pioneering DCNN architecture, has made significant strides in medical image classification. AlexNet models are very effective for faster model training and better feature extraction [95]. Thus, it has opened up new possibilities in disease diagnosis, lesion detection, and image analysis across various medical specialties. The applications of AlexNet in medical imaging are wide-ranging and impactful [77]. It has been successfully employed in the diagnosis of diseases such as brain tumors [77], neurological disorders [96], and ophthalmic conditions [97]. By analyzing MRI scan images [98], radiological images [99], histological samples, and dermatological images, AlexNet assists healthcare professionals in making more accurate and timely diagnoses.

Moreover, AlexNet's contribution to medical imaging extends beyond diagnosis. It has facilitated the automation [100, 101] of tasks such as lesion detection, segmentation, and identifying disease-related features within images. Kora, *et al.* [74] stated that AlexNet and others are the most widely used TL models for medical image analysis. Their word indicates that AlexNet improves diagnostic accuracy and streamlines healthcare practitioners' workflow, leading to more efficient patient care.

As the field of DL continues to advance and with the development of specialized architectures and the availability of larger medical imaging datasets, the impact of models like AlexNet on medical image classification is expected to grow. These advancements promise to enhance healthcare outcomes, enabling earlier disease detection and ultimately improving the well-being of patients. However, ensuring that these AI-powered tools are integrated into clinical practice and collaborating with medical professionals is crucial to achieving the best possible patient care.

Acknowledgments

We thank all the anonymous reviewers for their hard reviewing work.

References

- [1] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*: The MIT Press, 2017.
- [2] Y. Yan, R. Chen, Z. H. Yang, Y. Ma, J. L. Huang, L. Luo, *et al.*, "Application of back propagation neural network model optimized by particle swarm algorithm in predicting the risk of hypertension," *JOURNAL OF CLINICAL HYPERTENSION*, vol. 24, pp. 1606-1617, 2022.
- [3] M. Roder, L. A. Passos, G. H. de Rosa, V. H. C. de Albuquerque, and J. P. Papa, "Reinforcing learning in Deep Belief Networks through nature-inspired optimization," *APPLIED SOFT COMPUTING*, vol. 108, Article ID: 107466, 2021.
- [4] N. Yu, Z. Yu, F. Gu, T. Li, X. Tian, and Y. Pan, "Deep Learning in Genomic and Medical Image Data Analysis: Challenges and Approaches," *Journal of Information Processing Systems*, vol. 13, pp. 204-214, 2021.
- [5] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, pp. 100379, 2021.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE*, 2016.
- [7] C. Garbin, X. Zhu, and J. Marques, "Dropout and batch normalization: an empirical study on their impact on deep learning," *Multimedia Technology and Applications*, vol. 79, pp. 12777-12815, 2022.
- [8] F. Kong, F. Qian, K. D. Xu, and X. S. Shi, "Why does batch normalization reduce the model generalization on adversarial images?" *WORLD WIDE WEB-INTERNET AND WEB INFORMATION SYSTEMS*, vol. 26, pp. 1073-1091, 2023.
- [9] C. Y. Low, J. Park, and B. J. Teoh, "Stacking-Based Deep Neural Network: Deep Analytic Network for Pattern Classification," *IEEE TRANSACTIONS ON CYBERNETICS*, vol. 50, pp. 5021-5034, 2020.
- [10] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, pp. 504-507, 2006.
- [11] S. Ren, S. Wang, and Y. Zhang, "Weakly supervised machine learning," *CAAI Transactions on Intelligence Technology*, vol. 8, pp. 549-580, 2023.
- [12] A. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 23, pp. 33-47, 2022.
- [13] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [14] S. Lu, Z. Lu, and Y.-D. Zhang, "Pathological brain detection based on AlexNet and transfer learning," *Journal of Computational Science*, vol. 30, pp. 41-47, 2019.
- [15] L. Alzubaidi, J. L. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *JOURNAL OF BIG DATA*, vol. 8, Article ID: 53, 2021.
- [16] Y. D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S. H. Wang, "Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network," *INFORMATION PROCESSING & MANAGEMENT*, vol. 58, Article ID: 102439, 2021.
- [17] M. Varshney and P. Singh, "Optimizing nonlinear activation function for convolutional neural networks," *SIGNAL IMAGE AND VIDEO PROCESSING*, vol. 15, pp. 1323-1330, 2021.
- [18] D. Boob, S. S. Dey, and G. Lan, "Complexity of training ReLU neural network," *Discrete Optimization*, vol. 44, p. 10620, 2022.
- [19] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on*

- International Conference on Machine Learning - Volume 37*, Lille, France, 2015, pp. 448–456.
- [20] A. Bin Tufail, I. Ullah, A. U. Rehman, R. A. Khan, M. A. Khan, Y. K. Ma, *et al.*, "On Disharmony in Batch Normalization and Dropout Methods for Early Categorization of Alzheimer's Disease," *SUSTAINABILITY*, vol. 14, Article ID: 14695, 2022.
- [21] V. Jain, P. Gupta, A. Chaudhry, M. Batra, and D. J. Hemanth, "A Modified Deep Convolution Siamese Network for Writer-Independent Signature Verification," *INTERNATIONAL JOURNAL OF UNCERTAINTY FUZZINESS AND KNOWLEDGE-BASED SYSTEMS*, vol. 30, pp. 479-498, 2022.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 1026–1034.
- [23] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv: Learning*, 2015.
- [24] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: a Self-Gated Activation Function," *arXiv: Neural and Evolutionary Computing*, 2017.
- [25] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, "Impact of deep learning-based dropout on shallow neural networks applied to stream temperature modelling," *EARTH-SCIENCE REVIEWS*, vol. 201, Article ID: 103076, 2020.
- [26] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580*. doi: 10.48550/arXiv.1207.0580
- [27] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "DropBlock: regularization method for convolutional networks," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montreal, Canada, 2018, pp. 10750–10760.
- [28] I. Salehin and D. K. Kang, "A Review of Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *ELECTRONICS*, vol. 12, Article ID: 3106, 2023.
- [29] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *ArXiv*, vol. abs/1311.2901, 2013.
- [30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1558, 2014.
- [31] C. Szegedy, W. Liu, J. S. Jia, P. Sermanet, and A. Rabinovich, "Going deeper with convolutions," *IEEE Computer Society*, 2014.
- [32] N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," *ArXiv*, vol. abs/1602.07360, 2016.
- [33] S. Liu, L. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3200–3208.
- [34] Y. Zhang, V. Govindaraj, C. Tang, W. Zhu, and J. Sun, "High Performance Multiple Sclerosis Classification by Data Augmentation and AlexNet Transfer Learning Model," *J. Medical Imaging Health Informatics*, vol. 9, pp. 2012-2021, 2019.
- [35] T. M. Ghazal, S. Munir, S. Abbas, A. Athar, H. Alrababah, and M. A. Khan, "Early Detection of Autism in Children Using Transfer Learning," *INTELLIGENT AUTOMATION AND SOFT COMPUTING*, vol. 36, pp. 11-22, 2023.
- [36] T. Balashanmugam, K. Sengottaiyan, M. S. Kulandairaj, and H. Dang, "An effective model for the iris regional characteristics and classification using deep learning alex network," *IET IMAGE PROCESSING*, vol. 17, pp. 227-238, 2023.
- [37] I. Kayadibi, G. E. Güraksin, U. Ergün, and S. Güneş, "An Eye State Recognition System Using Transfer Learning: AlexNet-Based Deep Convolutional Neural Networks," *INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS*, vol. 5, Article ID: 49, 2022.
- [38] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 1440-1448.
- [39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137-1149, 2017.
- [40] S. Kizilirmak and E. Sert, "Human Eye Faster R-CNN-JS: Hurricane detection with faster R-CNN using artificial Jellyfish Search (JS) optimizer," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 81, pp. 37981-37999, 2022.
- [41] A. Singh, V. Kalanidhri, and R. Karthikeyan, "Performance analysis of object detection algorithms for robotic weld applications in planar environment," *INTERNATIONAL JOURNAL OF COMPUTER INTEGRATED MANUFACTURING*, vol. 36, pp. 1083-1108, 2023.
- [42] S. Y. Kim, S. S. Hu, M. Bagavathiannan, and D. Z. Song, "Toward Robotic Weed Control: Detection of Nutsedge Weed in Bermudagrass Turf Using Inaccurate and Inefficient Training Data," *IEEE ROBOTICS AND AUTOMATION LETTERS*, vol. 6, pp. 7365-7372, 2021.
- [43] G. M. Li, Y. B. Huang, Z. Q. Chen, G. D. Chesser, J. L. Purswell, J. Linhoss, *et al.*, "Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: A Review," *SENSORS*, vol. 21, Article ID: 1492, 2021.
- [44] C. S. Pereira, R. Morais, and M. Reis, "Deep Learning Techniques for Grape Plant Species Identification in Natural Images," *SENSORS*, vol. 19, Article ID: 4850, 2019.
- [45] Y. H. Bu, X. N. Jiang, J. P. Tian, X. J. Hu, L. P. Han, D. Huang, *et al.*, "Rapid nondestructive detecting of sorghum varieties based on hyperspectral imaging and convolutional neural network," *JOURNAL OF THE SCIENCE OF FOOD AND AGRICULTURE*, vol. 103, pp. 3970-3983, 2023.
- [46] G. B. Ergün and S. Güneş, "Classification of Canine Maturity and Bone Fracture Time Based on X-Ray Images of Long Bones," *IEEE ACCESS*, vol. 9, pp. 109004-109011, 2021.
- [47] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Convolutional neural networks for multi-class brain disease detection using MRI images," *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS*, vol. 78, Article ID: 101673, 2019.
- [48] M. Sharma, R. K. Patel, A. Garg, R. SanTan, and U. R. Acharya, "Automated detection of schizophrenia using deep learning: a review for the last decade," *PHYSIOLOGICAL MEASUREMENT*, vol. 44, Article ID: 03tr01, 2023.
- [49] A. Shalhaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," *PHYSICAL AND ENGINEERING SCIENCES IN MEDICINE*, vol. 43, pp. 1229-1239, 2020.

- [50] Mohebbanaaz, L. V. R. Kumar, and Y. P. Sai, "A new transfer learning approach to detect cardiac arrhythmia from ECG signals," *SIGNAL IMAGE AND VIDEO PROCESSING*, vol. 16, pp. 1945-1953, 2022.
- [51] M. A. Berwo, Y. Fang, J. Mahmood, N. Yang, Z. J. Liu, and Y. M. Li, "FAECCD-CNet: Fast Automotive Engine Components Crack Detection and Classification Using ConvNet on Images," *APPLIED SCIENCES-BASEL*, vol. 12, Article ID: 9713, 2022.
- [52] S. Sharma and R. Mehra, "Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image," *VISUAL COMPUTER*, vol. 36, pp. 1755-1769, 2020.
- [53] S. Y. Arafat, N. Ashraf, M. J. Iqbal, I. Ahmad, S. Khan, and J. Rodrigues, "Urdu signboard detection and recognition using deep learning," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 81, pp. 11965-11987, 2022.
- [54] Q. Xu and Z. Y. Wang, "A Data-Driven Model for Automated Chinese Word Segmentation and POS Tagging," *COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE*, vol. 2022, Article ID: 7622392, 2022.
- [55] B. Peng, H. B. Zhang, N. Yang, and J. M. Xie, "Vehicle Recognition from Unmanned Aerial Vehicle Videos Based on Fusion of Target Pre-Detection and Deep Learning," *SUSTAINABILITY*, vol. 14, Article ID: 7912, 2022.
- [56] M. Rafiq, G. Rafiq, R. Agyeman, G. S. Choi, and S. I. Jin, "Scene Classification for Sports Video Summarization Using Transfer Learning," *SENSORS*, vol. 20, Article ID: 1702, 2020.
- [57] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot Classification of Field Sports Videos Using AlexNet Convolutional Neural Network," *APPLIED SCIENCES-BASEL*, vol. 9, Article ID: 483, 2019.
- [58] A. Mumtaz, A. B. Sargano, and Z. Habib, "Fast Learning Through Deep Multi-Net CNN Model for Violence Recognition In Video Surveillance," *COMPUTER JOURNAL*, vol. 65, 2022.
- [59] Irfanullah, T. Hussain, A. Iqbal, B. Yang, and M. Hussain, "Real time violence detection in surveillance videos using Convolutional Neural Networks," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 71, pp. 38151-38173, 2022.
- [60] A. A. Khan, M. A. Nadeem, M. Shoaib, R. Changir, R. Alroobaea, M. Alsafar, *et al.*, "Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model," *ELECTRONICS*, vol. 11, Article ID: 4005, 2022.
- [61] W. Imen, A. Amna, B. Fatma, S. F. Ezahra, and N. Masmoudi, "Fast HVC intra-CU decision partition algorithm with modified LeNet-5 and AlexNet," *SIGNAL IMAGE AND VIDEO PROCESSING*, vol. 16, pp. 1811-1819, 2022.
- [62] Y. C. Hung, Y. Bao, I. Sunmola, D. A. Huang, J. H. Ngyu, and A. Alabdulkarim, "Capturing fine-grained details for video-based automation of suturing skills assessment," *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGICAL SURGERY*, vol. 18, pp. 545-552, 2023.
- [63] N. Ziafat, H. F. Ahmad, I. Fatima, M. Zia, A. Alhumam, and K. Rajpoot, "Correct Pronunciation Detection of the Arabic Alphabet Using Deep Learning," *APPLIED SCIENCES-BASEL*, vol. 11, Article ID: 2508, 2021.
- [64] I. Attri, L. K. Awasthi, T. P. Sharma, and P. Rathee, "A review of deep learning techniques used in agriculture," *ECOLOGICAL INFORMATICS*, vol. 77, Article ID: 102217, 2023.
- [65] M. Alencastre-Miranda, R. M. Johnson, and H. I. Krebs, "Convolutional Neural Networks and Transfer Learning for Quality Inspection of Different Sugarcane Varieties," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 787-794, 2021.
- [66] M. A. Khan, T. Akram, M. Sharif, M. Awais, K. Javed, H. Ali, *et al.*, "CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features," *COMPUTERS AND ELECTRONICS IN AGRICULTURE*, vol. 155, pp. 220-236, 2018.
- [67] J. Zhang, L. He, M. Karkee, Q. Zhang, X. Wang, and M. Gao, "Branch detection for apple trees trained in fruiting with a deep learning architecture using depth features and ResNeXt-Convolutional Neural Network (R-CNN)," *COMPUTERS AND ELECTRONICS IN AGRICULTURE*, vol. 155, pp. 386-393, 2018.
- [68] Z. Tarek, M. Elhoseny, I. I. Alghamdi, and M. El-Hasnony, "Leveraging deep-tier deep learning model for environmental cleaner plants production," *SCIENTIFIC REPORTS*, vol. 13, Article ID: 10999, 2023.
- [69] L. Wang, S. Li, C. Q. Teng, C. Jiang, J. Li, Z. Li, *et al.*, "Automated Detection Method for Mining Subsidence Basins Based on InSAR and CNN-AFSA-SVM," *SUSTAINABILITY*, vol. 14, Article ID: 13898, 2022.
- [70] J. Ren, S. Zhou, J. Wang, S. Yang, and C. Liu, "Research on Identification of Natural and Unnatural Earthquake Events Based on AlexNet Convolutional Neural Network," *Wireless Communications and Mobile Computing*, vol. 2022, p. 6782094, 2022.
- [71] L. Ichim and M. Popescu, "Segmentation of Vegetation and Flood from Aerial Images Based on Decision Fusion of Neural Networks," *REMOTE SENSING*, vol. 12, Article ID: 2490, 2020.
- [72] S. Manoj and C. Valliyammai, "Drone network for early detection of forest fire and dynamic fire quenching plan generation," *EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING*, vol. 2023, Article ID: 112, 2023.
- [73] S. Lu, S.-H. Wang, and Y.-D. Zhang, "Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm," *Neural Computing and Applications*, vol. 33, pp. 10799-10811, 2021.
- [74] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, *et al.*, "Transfer learning techniques for medical image analysis: A review," *BIOCYBERNETICS AND BIOMEDICAL ENGINEERING*, vol. 42, pp. 79-107, 2022.
- [75] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, "Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network," *Information Processing & Management*, vol. 58, p. 102439, 2021.
- [76] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *COMPUTERS IN BIOLOGY AND MEDICINE*, vol. 128, Article ID: 104115, 2021.
- [77] F. E. AlTahhan, G. A. Khouqeer, S. Saadi, A. Elgarayhi, and M. Sallah, "Refined Automatic Brain Tumor Classification Using Hybrid Convolutional Neural Networks for MRI Scans," *DIAGNOSTICS*, vol. 13, Article ID: 864, 2023.
- [78] M. Ramya, G. Kirupa, and A. Rama, "Brain tumor classification of magnetic resonance images using a novel CNN-based medical image analysis and detection network in comparison with AlexNet," *Journal of population therapeutics and clinical pharmacology = Journal de la therapeutique des populations et de la pharmacologie clinique*, vol. 29, pp. e97-e108, 2022.

- [79] A. Sarkar, M. Maniruzzaman, M. A. Alahe, and M. Ahmad, "An Effective and Novel Approach for Brain Tumor Classification Using AlexNet CNN Feature Extractor and Multiple Eminent Machine Learning Classifiers in MRIs," *JOURNAL OF SENSORS*, vol. 2023, Article ID: 1224619, 2023.
- [80] S. P. Singh, L. P. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D Deep Learning on Medical Images: A Review," *SENSORS*, vol. 20, Article ID: 5097, 2020.
- [81] J. Alyami, A. Rehman, T. Sadad, M. Alruwaythi, T. Saba, and S. A. Bahaj, "Automatic skin lesions detection from images through microscopic hybrid features set and machine learning classifiers," *MICROSCOPY RESEARCH AND TECHNIQUE*, vol. 85, pp. 3600-3607, 2022.
- [82] S. B. Melingi, R. K. Mojjada, V. K. Reddy, C. Kumar, and K. A. Kumar, "A bio-inspired AlexNet-DrpXLM archetype for an effective brain stroke lesion detection and classification," *CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE*, vol. 34, Article ID: e7100, 2022.
- [83] I. Shafi, M. Sajad, A. Fatima, D. G. Aray, V. Lipari, I. D. Diez, *et al.*, "Teeth Lesion Detection Using Deep Learning and the Internet of Things Post-COVID-19," *SENSORS*, vol. 23, Article ID: 6837, 2023.
- [84] V. K. Bairagi, P. P. Gumaste, S. H. Rajput, and K. S. Chethan, "Automatic brain tumor detection using CNN transfer learning approach," *MEDICAL & BIOLOGICAL ENGINEERING & COMPUTING*, vol. 61, pp. 1821-1836, 2023.
- [85] S. Petscharnig and K. Schöffmann, "Learning laparoscopic video shot classification for gynecological surgery," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 77, pp. 8061-8079, 2018.
- [86] Y. Kumar and S. Gupta, "Deep Transfer Learning Approaches to Predict Glaucoma, Contact, Choroidal Neovascularization, Diabetic Macular Edema, DRUSEN and Healthy Eyes: An Experimental Review," *ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING*, vol. 2023, pp. 521-541, 2023.
- [87] P. Kumar, D. Suganthi, M. Valarmathi, P. Swain, P. Vashistha, D. Buddhi, *et al.*, "A Multi-Thresholding-Based Discriminative Neural Classifier for Detection of Retinoblastoma Using CNN Models," *BioMed research international*, vol. 2023, pp. 580361-5803661, 2023.
- [88] M. J. Umeed, M. Sharif, M. Raza, and S. Kadry, "A deep feature fusion and selection-based neural eye disease detection from OCT images," *EXPERT SYSTEMS*, vol. 40, 2023.
- [89] R. Aziz, J. A. Taj, M. S. Al, and K. S. Karimov, "An ensemble framework based on Deep CNNs architecture for glaucoma classification using fundus photography," *MATHEMATICAL SCIENCES AND ENGINEERING*, vol. 13, pp. 5321-5341, 2021.
- [90] Y. Wang, W. T. Huang, W. H. Ho, and J. T. Tsai, "Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning," *BMC MEDICAL GENOMICS*, vol. 22, Article ID: 99, 2021.
- [91] A. Ben Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangere, F. Ghiringhelli, *et al.*, "Deep learning for colon cancer histopathological images analysis," *COMPUTERS IN BIOLOGY AND MEDICINE*, vol. 136, Article ID: 104730, 2021.
- [92] Q. Abbas, M. Albathan, A. Altameem, R. S. Almakki, and A. Hussain, "Deep-Ocular: Improved Transfer Learning Architecture Using Self-Attention and Dense Layers for Recognition of Ocular Diseases," *DIAGNOSTICS*, vol. 13, Article ID: 3165, 2023.
- [93] G. C. Y. Chan, R. Kamble, H. Muller, S. A. A. Shah, T. B. Tang, and F. Meriaudeau, "Fusing Results of Several Deep Learning Architectures for Automatic Classification of Normal and Diabetic Macular Edema in Optical Coherence Tomography," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2018, pp. 660-673, 2018.
- [94] X. Han, Z. Hu, S. Wang, and Y. Zhang, "A Survey on Deep Learning in COVID-19 Diagnosis," *Journal of Imaging*, vol. 9, 2022.
- [95] P. Sabitha and G. Meersandhi, "A dual stage AlexNet-HHO-DrpXLM archetype for an effective feature extraction, classification and prediction of liver cancer based on histopathology images," *BIOMEDICAL SIGNAL PROCESSING AND CONTROL*, vol. 77, Article ID: 103833, 2022.
- [96] A. Vijayarani, S. F. Ahamed, S. Ramakrishna, N. U. Kumar, and B. Raj, "CNSD-Net: joint brain-heart disorders identification using temporal optimization algorithm-based deep neural network," *SOFT COMPUTING*, vol. 27, pp. 12653-12668, 2023.
- [97] A. H. Al-Timemy, N. H. Ghaeb, Z. M. Mosa, and J. Escudero, "Deep Transfer Learning for Improved Detection of Keratoconus using Corneal Topographic Maps," *COGNITIVE COMPUTATION*, vol. 14, pp. 1627-1642, 2022.
- [98] M. H. Al-Jhaileh, "Diagnosis and classification of Alzheimer disease by using a convolution neural network algorithm," *SOFT COMPUTING*, vol. 26, pp. 7751-7762, 2022.
- [99] J. Cortés and S. Sánchez, "Deep Learning Transfer with AlexNet for chest X-ray COVID-19 recognition," *IEEE LATIN AMERICA TRANSACTIONS*, vol. 19, pp. 944-951, 2021.
- [100] M. Chen, P. Zhou, D. Wu, L. Hu, M. M. Hassan, and A. Alamri, "AI-Skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework," *INFORMATION FUSION*, vol. 54, pp. 1-9, 2020.
- [101] G. X. Wei, Y. Y. Liu, X. W. Ji, Q. X. Li, Y. Xing, Y. L. Xue, *et al.*, "Micro-morphological feature visualization, auto-classification, and evolution quantitative analysis of tumors by using SR-PCT," *CANCER MEDICINE*, vol. 10, pp. 2319-2331, 2021.