International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

The Evolution of Deep Learning: A Performance Analysis of CNNs in Image Recognition

Mohit Mittal

Uttar Pradesh Technical University, Lucknow, Uttar Pradesh, India

ABSTRACT: Computer vision, or image recognition, analyses and interprets visual data in real-world scenarios like images and videos. AI and ML research focusses on object, scene, action, and feature identification because of its usefulness in image processing. Neural networks and deep learning have improved image recognition systems significantly in recent years.

Early image recognition used template matching to identify objects. A photo is compared to a stored template using similarity measures like correlation to get the best match. There are several constraints, especially with distorted, scaled, or noisy images. Template matching is computationally expensive. CNNs are the most common deep learning architecture for computer vision. CNNs automatically learn visual input hierarchies. They include pooling, convolutional, and fully connected layers. Regional patterns like shapes, textures, and edges in the input image are filtered by convolutional layers. Pooling layers reduce feature map spatial dimensions, making feature extraction consistent and robust. Thick layers, or fully linked layers, classify or regress using learning features. CNN training requires a big labelled dataset. This research article introduces a deep learning model for the classification and recognition of images. The input data for this model is the Images Data Set. The GenNet Algorithm is employed to derive critical features. The accuracy of classification results is enhanced through the use of preprocessing and feature extraction. The classification model is generated using Convolution Neural Network, AlexNet, ResNet, and VGG.

KEYWORDS: CNN, AlexNet, VGG, ResNet, Hyperparameter Tuning, Image Recognition

I. INTRODUCTION

The study and application of visual data analysis and interpretation to real-world contexts, such as still photos and moving films, is known as computer vision, or image recognition [1]. The significance of this discipline in image processing has made object, scene, action, and feature recognition a central focus of AI and ML research. In the past few years, there have been significant advancements in image recognition systems, particularly in the field of deep learning and neural networks.

In the early stages of image recognition, template matching was a fundamental approach to object identification. This method compares a picture to a stored template and identifies the most suitable match by employing similarity metrics such as correlation. Nevertheless, there are numerous limitations, particularly when dealing with images that are distorted, scaled, or chaotic. The computational cost of template matching is also substantial [2].

The discipline of image recognition was truly revolutionized by Convolutional Neural Networks (CNNs). CNNs, which were introduced by YannLeCun and others in the late 1980s, gained widespread popularity after the AlexNet architecture emerged victorious in the 2012 ImageNet competition by a significant margin. CNNs utilize a series of layers, such as convolutional, pooling, and fully connected layers, to extract hierarchical information from images. The deep learning strategy exhibited superior scalability and accuracy in comparison to more conventional feature-based approaches [3].

Computer vision is capable of managing a diverse array of tasks that fall under the following categories due to its foundation in image processing methods. 1. Image registration. 2. Image processing. 3. The process of extracting characteristics from images. Fourth, an examination of the attributes of the image. The initial step is perceived as being associated with the image input method and/or the creation of the image in the system, which utilizes specific data frameworks, such as scalar or vector values. For instance, the second task's solution offers the requisite picture transformation for noise reduction, scaling, rotation, and contrast correction, dependent on the geometric and/or color features provided. This appears to be more significant during the third image resolution task, when images are depicted as characteristics, from the perspective of the activities related to the concrete systems. For example, the semantic

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

contents of the image analysis are retrieved, object parameters are established, and integrations are implemented, all of which are included in the final operation [4].

Although machine learning techniques have been available for some time, it has only been in the past few years that scientists have been able to effectively apply complex mathematical calculations to vast datasets. The processing power and storage capacities of computers have constantly increased over the years. As a result, machine learning algorithms are now capable of analyzing a greater number of datasets and larger ones, which enables them to construct a more comprehensive knowledge base [5]. The application of deep learning has revolutionized the manner in which problems are addressed when confronted with vast quantities of unstructured, chaotic data. The term "universal function approximator" is frequently used to refer to artificial neural networks, as they are capable of learning any function with a single hidden layer, even if the function is ambiguous. This is due to the fact that artificial neural networks are capable of learning any function, regardless of its vagueness, in contrast to real-life neural networks. For instance, individuals who possess sufficient computing power and storage capacity may be capable of developing deep neural networks. These neural networks are composed of numerous interconnected layers. The deep neural network (DNN) is a type of artificial neural network that employs multiple simulation layers to replicate more intricate neural networks [6] [7].

One subfield of machine learning, "deep learning," is dedicated to the development and application of artificial neural networks (ANNs) to address complex problems. Deep learning has revolutionized computer vision by enabling computers to recognize and interpret visual information. It involves the interpretation of visual data, such as photographs or videos, through analysis. Historically, this would have been accomplished by employing human specialists who engaged in feature engineering by hand. Nevertheless, these techniques encountered obstacles due to the complexity and unpredictability of visual input. In contrast, deep learning pursues a data-driven methodology that generates representations from the data automatically. Artificial neural networks with numerous layers of linked neurons are implemented to replicate the human brain's functionality. The term "deep neural networks" is derived from the substantial number of layers that constitute these networks.

Convolutional neural networks (CNNs) are the most prevalent deep learning architecture for computer vision issues. Convolutional neural networks (CNNs) are intended to automatically acquire hierarchical perspectives on visual input. Pooling, convolutional, and fully linked layers are among the components that constitute them. Convolutional layers apply filters to the input image by separating regional patterns such as forms, textures, and edges. The extraction of consistent and robust features is facilitated by the aggregating layers, which reduce the spatial dimensions of the feature maps. The final classification or regression task is performed by fully connected layers, which are also referred to as dense layers, using the learnt features. A large tagged dataset is necessary for CNN training. The network minimizes a loss function that quantifies the discrepancy between the projected outputs and the ground truth labels to determine the optimal values for its internal parameters, such as its biases and weights, during training. Back-propagation is frequently employed in conjunction with gradient descent techniques, such as stochastic gradient descent (SGD), to facilitate the computation of gradients in this optimization.

II. LITERATURE SURVEY

Deep learning has significantly transformed computer vision, enabling the rapid and accurate classification of digital images. Due to the abundance of extensive tagged image datasets and improvements in algorithms and hardware, deep learning models provide cutting-edge technology for many image classification applications. The term "image classification" refers to the process of categorizing images into predefined groups. Convolutional Neural Networks have shown exceptional efficacy in this field. Convolutional neural networks (CNNs) use a network of interconnected neurons to autonomously acquire a hierarchical representation of an image. These networks excel in classifying pictures because to their capacity to discern intricate patterns and features at different scales. The standard stages in a deep learning classification process include data preparation, model development, training, evaluation, optimization, fine-tuning, and deployment.

To recognize images in the ImageNet competition, the authors of this paper present a CNN architecture called AlexNet. The network uses ReLU activations and has 5 convolutional layers with 3 fully connected layers. They demonstrate the significance of regularization techniques such as dropout and data augmentation in enhancing model performance. This technique represented a significant advancement over previous approaches, achieving a top-5 error rate of 16.4%, marking a pivotal point in the evolution of deep learning in computer vision [1]. This study introduces the VGG

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

network, which significantly increases the network's depth for image classification. The authors show that networks with up to 19 layers exhibit superior performance when trained with smaller 3x3 filters to enhance depth, using the ImageNet dataset. Furthermore, they demonstrate that deeper models exhibit superior generalization compared to shallower models and underscore the relationship between depth and accuracy [2].

The authors analyze several deep learning methodologies for image recognition, focusing on CNNs and DNNs. Furthermore, they conduct a comprehensive analysis of several architectures and methodologies, including deep Boltzmann machines, autoencoders, and stacked autoencoders, and their impact on the efficiency and accuracy of image recognition tasks [3]. This study primarily focusses on using convolutional neural networks (CNNs) to develop mid-level representations of pictures applicable to various recognition tasks. To enhance performance in tasks like as item identification and action recognition, the authors demonstrate that features acquired from large datasets may be used in smaller datasets [4].

The authors provide an architecture for a multi-column deep neural network (MCDNN) that independently trains several CNNs on the same dataset prior to consolidating their outcomes. This technique improves classification performance by using the many attributes acquired from each column, demonstrating outstanding results on the ImageNet dataset [5]. This study introduces Residual Networks (ResNets), which use shortcut connections to train incredibly deep neural networks of up to 152 layers. To train deeper networks and enhance performance on the ImageNet challenge, the authors demonstrate that residual connections alleviate the vanishing gradient problem. This approach, after pioneering advancements in image categorization, became fundamental to several computer vision applications [6].

The objective of this research is to assist readers in understanding and visualizing convolutional networks. The authors illustrate the functionality of convolutional neural networks (CNNs) by generating images that fully use diverse filters and layers. Their research provides resources for diagnosing and enhancing network topologies and elucidates the mechanisms by which CNNs acquire hierarchical features from unprocessed image data [7]. YannLeCun and YoshuaBengio's seminal research elucidates convolutional networks (CNNs) comprehensively, including its architecture, training methodologies, and practical applications. They elucidate the use of CNNs in image recognition and establish the foundational principles of modern CNNs' efficacy in image processing by presenting the concepts of weight sharing and local receptive fields [8].

This paper introduces DeCAF, a method for acquiring deep convolutional features amongst various picture recognition problems. The authors explore the use of deep convolutional features acquired from extensive datasets to improve the performance of other tasks, even when the target task has little data. Due to their efforts, CNNs have shown efficacy as a technique for transfer learning [9].

The authors introduce the Network in Network (NiN) architecture, which facilitates more intricate feature learning by replacing traditional convolutional layers with small neural networks, such as multilayer perceptron (MLPs). In several image recognition tasks, NiN surpasses traditional networks and improves the representational capacity of CNNs [10]. Eliminating eye problems that result in blindness will help both individuals and society collectively. A significant proportion of the population in developing countries endures untreated ocular disorders. The socioeconomic situation of patients contributes to the postponement or absence of therapy for some diseases. Contemporary research primarily focusses on preventing vision impairment and establishing precise scientific methodologies for the identification and treatment of ocular diseases. Optical Coherence Tomography (OCT), as a non-invasive optical medical diagnostic imaging technique, is essential in ophthalmology [11, 12]. It may generate two-dimensional or three-dimensional images by quantifying the light that is reflected or backscattered and measuring the duration for the echo to return. The first evaluation of a human retina further substantiates the significant impact and rapid advancement of optical coherence tomography (OCT) imaging on clinical diagnosis. It provides in-vivo cross-sectional imaging of microstructures inside biological systems and allows the visualization of retinal architecture that other non-invasive diagnostic methods cannot achieve. One of the most sophisticated therapeutic applications of optical coherence tomography imaging is in ophthalmic treatment. Ocular diagnostics has garnered significant worldwide interest since the introduction of fourth-generation devices, with several companies expanding this technology internationally. Its modifications in textual and morphological properties make it advantageous for early illness identification [13]. Conversely, artificial neural networks (ANNs) have several applications in domains like medical analysis, computer vision, and speech processing, among others. Artificial neural networks (ANNs) are the foundation of the majority of deep learning models and frameworks. These technologies have been embraced by several fields because to their

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

promising results that surpass human analysis. A deep learning approach was proposed to distinguish between normal OCT retinal images and those affected by Age-related Macular Degeneration (AMD or ARMD). An automated deep learning segmentation method for macular OCT images was also proposed to detect intraretinal fluid. A deep learning-based automated technique has been created to identify and quantify macular fluid for the diagnosis of retinal diseases in ophthalmology. The results are accurate and authentic. Transfer learning is a prevalent technique in deep learning that enables the use of a pre-trained model to address a novel problem. Medical experts often use transfer learning to analyze retinal OCT pictures and diagnose ocular disorders. Retinal OCT images are analyzed via a pre-trained convolutional neural network (CNN), namely GoogleNet, and then categorized as dry AMD, no pathology, or Diabetic Macular Oedema (DME).

The three ocular conditions under investigation are diabetic macular oedema (DME), drusen, and choroidal neovascularization (CNV). The capacity of these illnesses to induce irreversible blindness and visual impairment in both affluent and underdeveloped countries was a determining factor in their selection [14] [15]. Timely detection of sickness would certainly enhance treatment methods, perhaps alleviating the suffering associated with blindness. Researchers globally are now categorizing OCT pictures of DME, AMD, CNV, and Drusen with deep learning and other CNN methodologies. A computer-aided diagnostic model was proposed for distinguishing diabetic macular oedema (DME), age-related macular degeneration (AMD), and healthy macula, utilizing the Correlation-based Feature Subset (CFS) selection method and the linear configuration pattern (LCP) features of optical coherence tomography (OCT) images. An automated technique using feature learning with color retinal fundus images was developed to facilitate the early detection and treatment of DME. A machine learning technique was proposed for the automated assessment of AMD severity using OCT images. The application will use Cohen's statistics and receiver operating characteristic (ROC) analysis. Automated segmentation of CNV in OCT images was used to enhance the treatment of CNV illnesses. An algorithm was proposed to find CNV regions in individuals with AMD. The U-Net CNN architecture enables the automatic distinction of Drusen from fundus photographs and the detection of early or advanced AMD.

III. METHODS

This section presents deep learning model for image classification and recognition. This model is presented in figure 1. In this model, Images Data Set is used as input data. Important features are extracted using GenNet Algorithm. Preprocessing and feature extraction helps in achieving more accurate classification results. Classification model is prepared by Convolution Neural Network, AlexNet, ResNet and VGG.

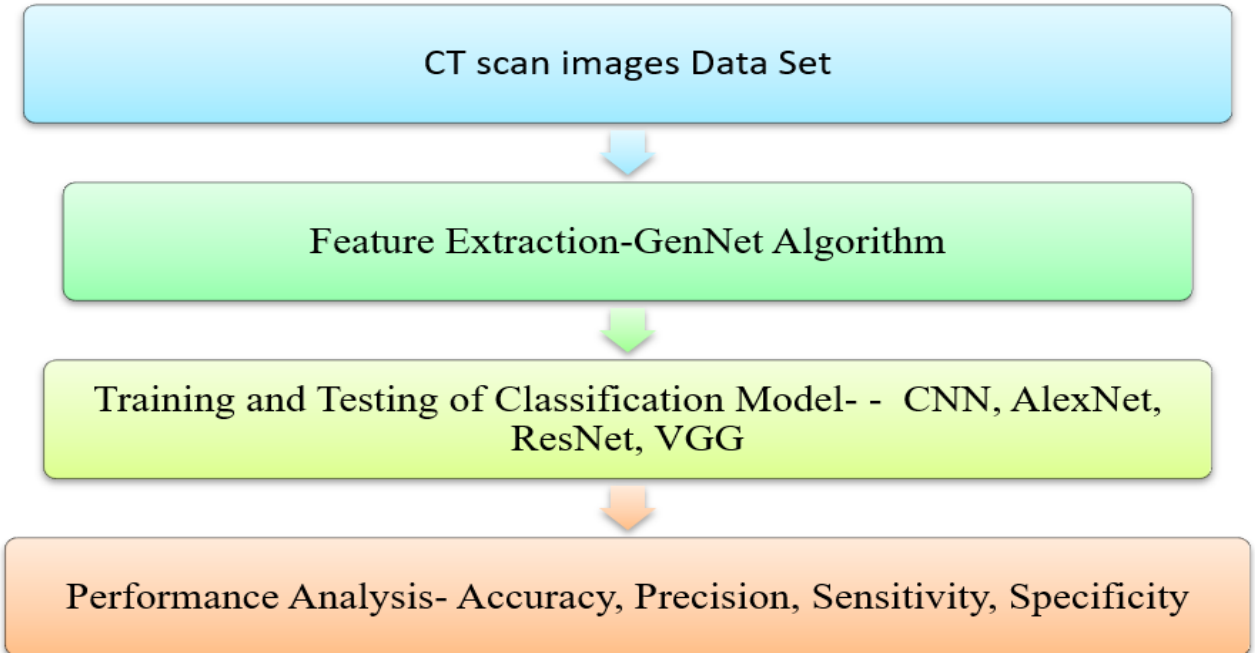

Figure 1: Deep Learning enabled methodology for Image Classification and Recognition

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

The GenNet algorithm employs an evolutionary approach based on SAGenetics to choose the most effective features. The process of feature selection starts by constructing a population using various subsets of the potential attributes. To achieve the objective, a predictive model is employed to evaluate certain segments of the population. Subsequently, a competition is organized to ascertain which subcategories will endure till the subsequent generation, considering every element of the populace. The next generation consists of the competition's prizewinners, together with a little amount of genetic diversity and cross-pollination. The GenNet methodology can be understood using the following algorithm.

Algorithm: GenNet Modeling

Step 1: An initial number of (population: EEG) are created.
Step 2: Individuals of the population are given a numerical rating.
Step 3: A competition is performed to choose a subset for further replication.
Step 4: Decide which parts of the EEG to pass on.
Step 5: Make changes to the EEG feature modelling.
Step 6: For generations to come, repeat the process.

Genetic algorithms are iterative methods that seek for a solution by using a population of individuals, where each individual is represented by a finite string of symbols known as the genome. To implement a genetic algorithm for next-generation sequencing, you should follow these fundamental steps: The starting population is generated randomly or through the use of heuristics. Every step in the evolutionary process (where the next generation is determined) requires deciphering and organizing the current population members based on a fitness function that defines the optimization problem in the search area. This approach utilizes concepts from evolutionary processes such as mutation, selection, and crossover to develop solutions for optimization problems in a manner analogous to natural selection. This study employs the GenNet technique for extracting features, which relies on the qualities, characteristics, or measurements that are utilized to describe each pattern that is inputted into the classifier. The elements employed to delineate the patterns in a pattern language provide the language's definition without directly articulating it. The language's failure to effectively explain the necessary information will restrict the precision of the acquired classification function, regardless of the employed learning technique. The objective of a feature subset selection task is to identify a practical subset of attributes that accurately represent patterns while limiting the expenses and potential risks associated with measurement.

Convolution neural networks (CNNs) are a type of neural network that consist of multiple layers, including a pooling or ReLU layer and a fully connected or wholly connected layer. CNN is mostly utilized for the identification of visual features in images, such as the border and contour of the image [16].

- **Convolution layer**

Convolutional neural networks do not adhere to a basic first-come, first-served methodology while being formed. A Convolutional Neural Network (CNN) typically expects an input in the form of a MxNx1 matrix. This matrix represents a two-dimensional picture with varying dimensions of M rows and N columns, where M is the image's row count and N is its column count. CNN employs filters of equivalent depth to the input picture with the purpose of generating a filtered image. The filter requires the input photographs to adhere to a certain form or curve in order to function properly. Currently, there is an appropriate level of contrast between the input picture and the curved filter shape. An equation can be used to represent the convection process.

$$(t)=(x^{*}w)(t)$$

- **Pooling layer**

By implementing this layer, we have the ability to decrease the amount of data being processed. The metric data is reduced by splitting the matrix data and replacing each sector with a single value. Arrays in a bucket may be allocated either the average or maximum value inside that bucket by using the well recognized pooling features.

- **Fully Connected layer**

These layers are altered to adhere to the network's architecture. A computational process that connects all input and output elements is referred to as a completely connected layer. Similar to how traditional artificial neural networks use this layer to establish connections between preceding and succeeding activities, this one functions in a similar manner.

- **Soft-max layer**

This level utilizes the data from the previous level to compute class probabilities using the Soft-max approach. Given the high probability of the projected output class for certain inputs at this level, it has a significant effect on the final

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

result. A wide variety of deep neural networks may be used for image categorization. Although neural networks have been taught to detect previously uploaded photos, improving the current classification task still requires transfer learning. Every target network still requires an equivalent amount of training. There are a maximum of twenty-five potential methods to decompose dates. Mini-batch instances alter the internal components of an ideal. The experimental sessions constantly used a learning rate of 0.0001 and a mini-batch size of 7.

In 1996, YannLeCun and YoshuaBengio developed a class of deep neural network algorithms known as Convolutional Neural Networks (CNNs) or ConvNets. The field of computer vision has lately seen rapid developments because to CNN technology. Both animals and humans use grid patterns for picture recognition; Convolutional Neural Networks (CNN) emulate this process by autonomously and adaptively learning hierarchies of spatial features, progressing from intricate patterns to fundamental ones. The Convolutional Layer, the Pooling Layer, and the Fully Connected Layer are the three essential components of a Convolutional Neural Network (CNN). It is seen in Figure 2 below. The convolutional and pooling layers perform feature extraction, whereas the fully connected layer integrates the features mapped by the preceding two layers into the output [17]. AlexNet, created by Alex Krizhevsky and his associates, is a convolutional neural network (CNN) and the first CNN model to use a graphics processing unit (GPU) to enhance training efficiency. In 2012, AlexNet triumphed in the ImageNet competition, leading to a rise in interest and research about the mechanisms of CNN and DL. Following its participation in the 2012 ImageNet (ILSVRC) competition, AlexNet triumphed with an error rate of just 15.3%, far lower than the prior record of almost 25%. AlexNet has eleven layers: five convolutional, three max-pooling, two normalization, two fully connected, and one softmax. Convolutional layers use convolutional filters with ReLU as the activation function. Max Pooling is performed in the pooling layers. AlexNet supports an input size of 227x227x3. AlexNet comprises 650,000 neurons and 60 million parameters.

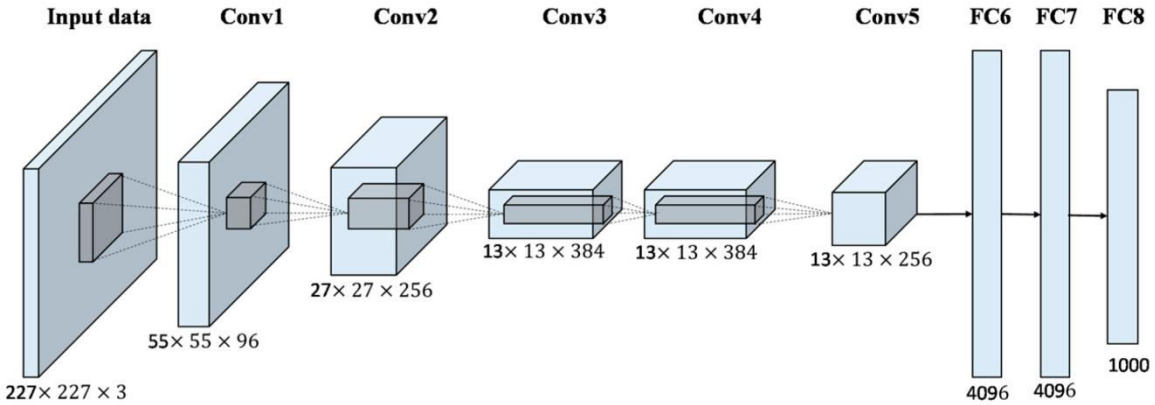

Figure 2: Architecture of AlexNet

Kaiming introduced the Residual Neural Network, known as ResNet. This design introduces a concept called as skip connections. The input matrix undergoes two linear transformations using the ReLU activation function. Increasing the network's depth enhances its accuracy rate; however, this also leads to the problem of overfitting. The use of negligible learning results in a diminishing gradient, as the challenge of increasing layer depth necessitates increasingly complex weight adjustments from the output layer. The second issue is that training a deeper network leads to an accelerated training rate, complicating optimization within an extensive parameter space. The ResNet model facilitates the training of deep networks by constructing them using components referred to as residual models [17].

The Visual Geometry Group, located in Oxford, developed this architecture, thus the designation VGG. It enhances the AlexNet architecture by including 16 layers of learning parameters, with the first two convolutional layers use a 3x3 filter instead of kernel sizes of 11 and 5, respectively. VGG-D is regarded as a simplistic architectural model due to its absence of hyperparameters. For convolutional layers, a 3x3 kernel size with a stride of 1 is commonly used, whereas pooling layers adopt SAME padding with a 2x2 stride. A pooling layer succeeds two convolutional layers in the construction of the network. Blocks of two convolutional layers and a pooling layer are iteratively repeated. To extract more complex image information, these blocks are constructed with analogous filter sizes that are used frequently. Following VGG, the concept of constructing networks based on blocks gained popularity. The execution of VGG is very time-consuming and resource-intensive, notwithstanding its attainment of superior accuracy on the ImageNet

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

dataset. The extensive range of the convolutional layers renders it inefficient[18].

IV.RESULTS ANALYSIS

In this experimental set up, ImageNet data set [19] is used. 250 images are used in the study. 200 images are used in training of model and 50 images are used for testing of model. Then important features are extracted using GenNet Algorithm. Preprocessing and feature extraction helps in achieving more accurate classification results. Classification model is prepared by Convolution Neural Network, AlexNet, ResNet and VGG. In this study, the performance of a number of different algorithms is analyzed and compared based on three criteria: accuracy, sensitivity, and specificity. The execution value for the performance metrics is calculated using then following equations.

$$Accurcay = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

The AleXNet and SVM models were trained using a subset of the dataset known as the training set. The training set is a subset of the dataset that is intended for training purposes. An epoch is a hyperparameter that, when used with an AlexNet model, determines the total number of data iterations utilised for training. This variable's values may vary from 0 to infinity. Running the training data through a certain number of iterations at each process epoch improves the model's performance. Over multiple cycles, the model will decide which portions of the dataset are critical and use those sections to label each input. After the classification process is completed, the loss will be determined, and the model's parameters will be fine-tuned as necessary. It is possible to ensure that the model's accuracy improves over time by repeating this procedure after each epoch.

By comparing it to the validation set, the model is both tested and improved. We utilize this set of hyperparameters to guarantee that the model is successfully trained. Using these hyperparameters, the model fails to give any analytically meaningful new insights. At the start of each epoch, the validation set is tested simultaneously with the AlexNet model being trained. It is expected that this tendency will continue till the end of the term. To apply the characteristics gained from the training set, a classification technique identical to the one described above is used to each individual input from the validation set. To account for the calculated loss, we will not change the model's weights until we perform the validation. If our model produces results that are too similar to the data, we may detect this by comparing them to the validation set. A model is termed overfit if its performance on training data is much better than on test data. If the model performs well on the validation set, we may be confident in its predictions thus far; here is where the model is put to use. The model's high level of data similarity suggests that it is erroneous.

Results are shown in table 1, figure 3 and figure 4. Table 1 shows accuracy, precision , recall and F1 score of CNN, AlexNet, ResNet and VGG. CNN is performing better than other deep learning models in terms of precision, recall and F1 score. In figure 4, it is clear that the accuracy of CNN is 96.42 percent. Accuracy of CNN is 4 percent more than the accuracy of AlexNet. Similarly, specificity, precision, F score of DWT based CNN is in the range of 95-96 percent. It is higher than the Similarly, specificity, precision, F score of AlexNet, ResNet and VGG.

Table 1- Performance Analysis of CNN

Parameter	VGG	Resnet	AlexNet	CNN
Accuracy	89.24	92.34	94.11	97.58
	89.53	92.56	93.91	96.78

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

Precision				
Recall	**88.78**	**93.67**	**91.85**	**97.52**
F1	**88.84**	**96.78**	**95.25**	**97.78**

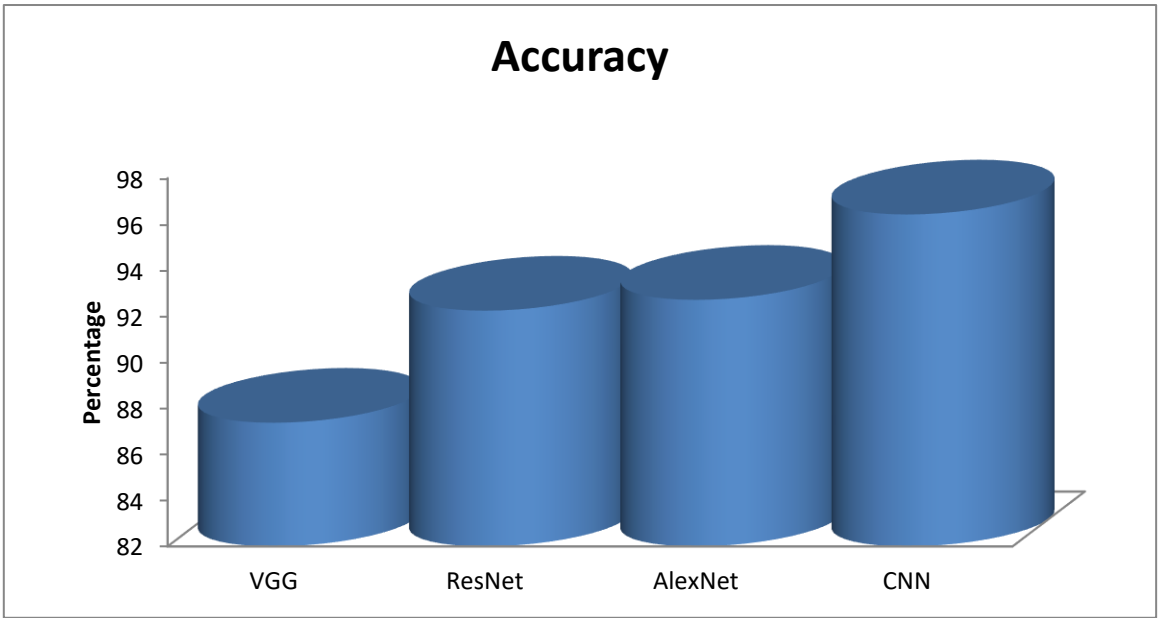

Figure 3: Graphical representation of average accuracy performance of classifiers for Image Classification and Recognition

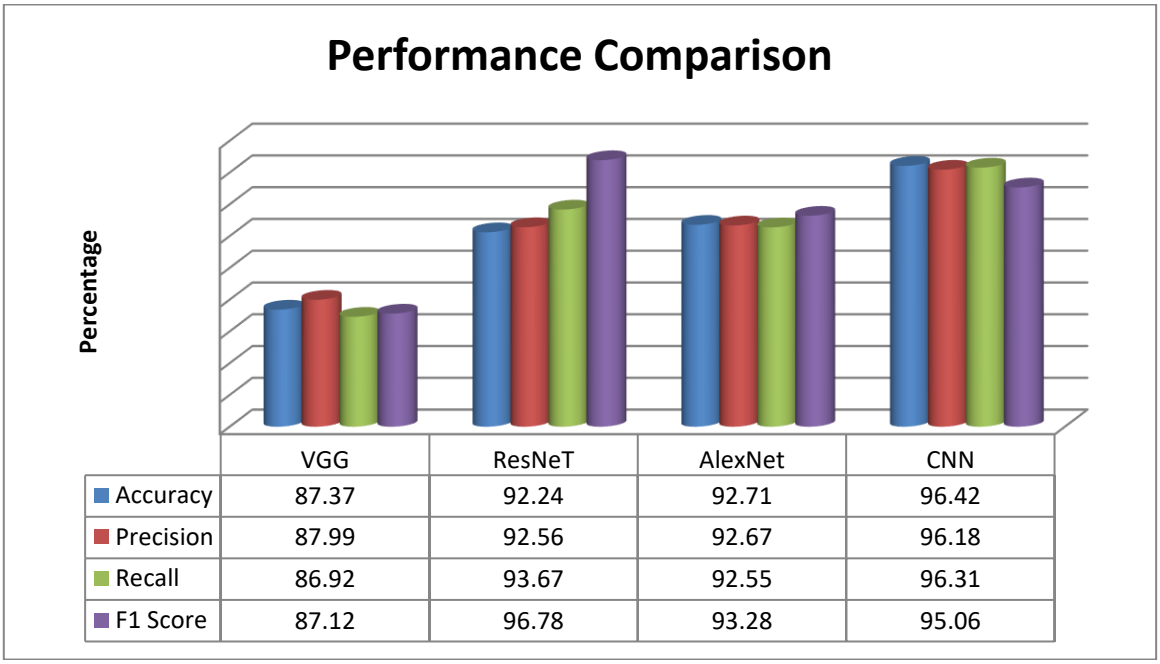

Figure 4: Overall Average Performance Analysis of CNN for Image Classification and Recognition

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

V. CONCLUSION

Machine learning, sometimes known as "deep learning," is the area that develops and applies artificial neural networks (ANNs) to complicated problems. Deep learning has transformed computer vision by allowing computers to identify and interpret pictures. Visual data, such as images and videos, are examined for interpretation. Previously, feature engineering was done manually by human specialists. The intricacy and unpredictability of visual input made these approaches difficult to implement. In contrast, deep learning uses data to construct representations automatically. Artificial neural networks, which are made up of several layers of connected neurons, simulate human brain function. Because of their several layers, these networks are referred regarded as "deep neural networks". This research paper describes a deep learning model for picture categorization and recognition. The Images Data Set is used as input data for this model. The GenNet Algorithm is used to extract critical characteristics. Preprocessing and feature extraction are used to enhance classification results. The classification model is generated using the Convolution Neural Network, AlexNet, ResNet, and VGG algorithms. CNN has an accuracy rating of 96.42 percent. CNN's accuracy is 4% greater than that of AlexNet. Similarly, DWT-based CNNs have F scores, accuracy, and specificity around the 95-96 percentile. It is more particular, precise, and has a higher F score than AlexNet, ResNet, and VGG.

REFERENCES

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks.*IEEE Transactions on Neural Networks and Learning Systems, 25*(12), 1-9. doi:10.1109/TNNLS.2014.2326367

[2] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: a review and new perspectives,” IEEE TPAMI, vol. 35, no. 8, pp. 1798–1828, 2013

[3] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” IEEE TPAMI, vol. 35, no. 8, pp. 1872–1886, 2013

[4] Oquab, M., Bottou, L., Laptev, I., &Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(9), 1717-1724. doi:10.1109/TPAMI.2014.223

[5] Ciresan, D. C., Meier, U., &Schmidhuber, J. (2012). Multi-column deep neural networks for image classification.*IEEE Transactions on Neural Networks and Learning Systems, 23*(2), 3642-3649. doi:10.1109/CVPR.2012.6248110

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.*IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(9), 1-9. doi:10.1109/CVPR.2016.90

[7] J. Lu, Y.-P. Tan, and G. Wang, “Discriminative multi-manifold analysis for face recognition from a single training sample per person,” IEEE TPAMI, vol. 35, no. 1, pp. 39–51, 2013.

[8] N.-S. Vu, “Exploring patterns of gradient orientations and magnitudes for face recognition,” IEEE Trans. Information Forensics and Security, vol. 8, no. 2, pp. 295–304, 2013

[9] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., &Tzeng, E. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(10), 2279-2287. doi:10.1109/TPAMI.2014.2345905

[10] Lin, M., Chen, Q., & Yan, S. (2014). Network in network.*IEEE Transactions on Neural Networks and Learning Systems, 25*(10), 2148-2156. doi:10.1109/TNNLS.2014.2300421

[11] T. Hassan, M. U. Akram, B. Hassan, A. Nasim and S. A. Bazaz, “Review of OCT and fundus images for detection of Macular Edema,” in Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST-2015), Macau, pp. 1-4, 2015.

[12] A. M. Bagci, R. Ansari and M. Shahidi, "A method for detection of retinal layers by optical coherence tomography image segmentation," in Proceedings of IEEE / NIH Life Science Systems and Applications Workshop, Bethesda, MD, pp. 144-147, 2007.

[13] W. Drexler and J. G. Fujimoto, “State-of-the-art retinal optical coherence tomography,” Prog.Retin. Eye Res. vol. 27, pp. 45–88, 2008.

[14] H. Fan, Z. Cao, Y. Jiang, and Q. Yin, “Learning deep face representation,” in arXiv: 1403.2802v1, 2014.

[15]C. B. Rickman, S. Farsiu, C. A. Toth and M. Klingeborn, "Dry age-related macular degeneration: Mechanisms, therapeutic targets, and imaging dry AMD mechanisms, targets, and imaging," Investigative Ophthalmol. Vis.Sci. vol. 54, ORSF68, 2013.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 4.468| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 3, Issue 6, November - December 2016 ||

DOI:10.15680/IJARETY.2016.0306016

[16] N. Sengar, M. K. Dutta, R. Burget and L. Povoda, "Detection of diabetic macular edema in retinal images using a region based method," in Proc. of 38th International Conference on Telecommunications and Signal Processing (TSP), Prague, pp. 412-415, 2015.
[17] Yim, J., Ju, J., Jung, H., Kim, J. (2015). Image Classification Using Convolutional Neural Networks With Multi-stage Feature. In: Kim, JH., Yang, W., Jo, J., Sincak, P., Myung, H. (eds) Robot Intelligence Technology and Applications 3. Advances in Intelligent Systems and Computing, vol 345.Springer, Cham.https://doi.org/10.1007/978-3-319-16841-8_52
[18] J. Dai, Y. Lu, and Y.-N. Wu, "Generative Modeling of Convolutional Neural Networks," arXiv (Cornell University), Jan. 2014, doi: https://doi.org/10.48550/arxiv.1412.6296.
[19] https://www.kaggle.com/datasets/hmendonca/imagenet-1k-tfrecords-ilsvrc2012-part-0