IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

A Survey on CNN-driven Architectures for Medical Image Analysis: Current Trends, Challenges, And Innovations.

Moksha Patel[1], Anuradha Desai[2] and Happy Patel[3]

P.G. Student, Department of Computer Engineering, Silver Oak University, Ahmadabad, India.[1]

Assistant Professor, Department of Computer Engineering, Silver Oak University, Ahmadabad, India.[2]

P.G. Student, Department of Computer Engineering, Silver Oak University, Ahmadabad, India.[3]

Abstract: Convolutional neural networks, or CNNs, are now the backbone of medical image processing and have revolutionized the interpretation and application of different medical data in clinical image, video,- decision-making in different classifications With a focus on significant advancements, cutting-edge trends, and enduring difficulties in the area, this survey study describes the investigation of CNN-based architectures for medical image processing.
The study began with basic models of CNN, such as LeNet, AlexNet, VGG, and ResNet, before heading to the advanced architectures used in DenseNet, U-Net, and Vision Transformers (ViTs). From these architectures, the discussion reflects their applications to medical image tasks such as disease classification, organ and lesion segmentation, and anomaly detection that cut across imaging modalities like Pathology, Colonoscopy MRI, CT scans.
The survey article provides a broad overview of Convolutional Neural Networks (CNNs), focusing on their applications in medical imaging. It demonstrates how various forms of CNN architectures are used for the interpretation of different types of medical imaging data such as x-ray, CT, MRI and ultrasound images.
The paper covers the developments in CNN methods and their capability in analyzing complex medical data sets and performing tasks such as disease identification, organ delineation and abnormality recognition. In this regard, the survey gives an explanation of the use of CNNs in medical images, and those features provide possibilities for predicting changes in the course of the disease and improve the results of treatment.

Keywords: Deep -learning, Medical Image, CNN Architectures, Image Classification.

I. INTRODUCTION

There is increased advancement within the medical sector due to advancement within computer science and artificial intelligence (AI). Deep learning stands out as one of the most useful AI techniques and in particular, Convolutional Neural Networks (CNNs) have changed the entire landscape of medical imaging as it enabled accurate and self-sufficient diagnosis to be carried out. Medical imaging includes the use of X-rays, MRI, CT, ultrasound, and microscopic images and these forms of imaging greatly supports decision making in the clinic by providing a visual assessment of the anatomy and any pathologies.

Recently, it appears that CNNs have taken over as the best formed deep learning technique for tackling the analysis of medical pictures thanks to their capability to learn the spatial aspects of patterns in a hierarchical manner. This has led to scientists putting forth CNN-based structures aimed at disease recognition, lesion detection, organ recognition and even tissues that are not normal in appearance. This has seen improved understanding and detection of ailments such as cancer, brain disorders, and issues related to the heart.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14318

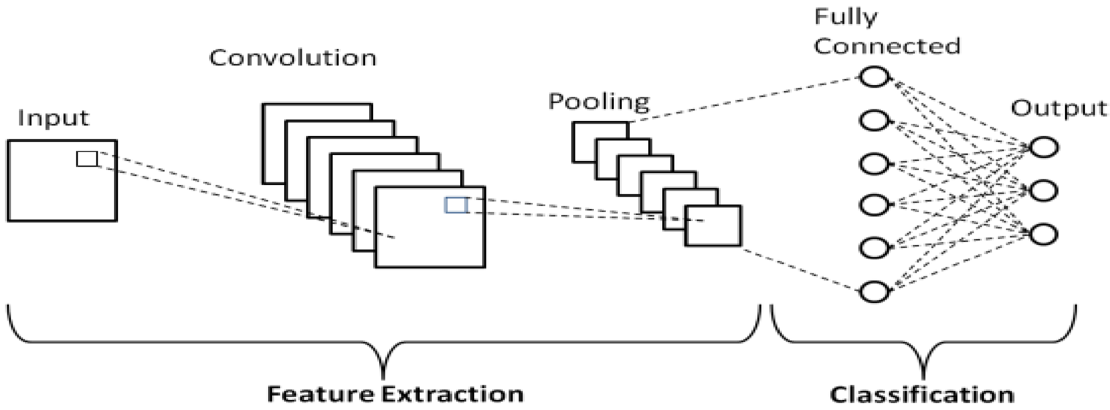

Fig. 1. Architecture of CNN [28]

The interdisciplinary blend between computer science and the field of medical imaging is very hot. It is interesting to note that CNNs and other learning techniques are being designed to address such domain specific issues as lack of data, uneven distribution of classes and the explainable artificial intelligence paradigm for clinical use. Also, more advanced concepts like federated learning for secure data sharing, compact models for inferences, and three-dimensional CNNs for in-depth computation are further increasing the prospect for AI in medicine.

This paper describes an exhaustive synthesis of literature regarding applications of CNNs for paramedical imaging. Recent developments are outlined, issues such as model risk management and data availability are addressed, and relevant good practices to mitigate such issues are reported. CNNs are advancing the connection between computer engineering and medicine and diagnosis, allowing this procedure to be automatic, accurate, and widespread that ultimately improves patient care and clinical efficacy.

The term convolutional, referring to the use of convolution, applies to the neural network topology in question. In a very broad sense, the overlapping of portions of the input data enables the network to create new features from existing ones. This also allows the network to generalize and recognize patterns different from those explicitly trained by the algorithm. It follows that in most applications, convolutional nets perform better than regular feedforward nets when asked to recognize and classify objects[1].

This section concentrates and demonstrates the above mentioned basic concepts related to development of a neural network based on CNN. The architecture of CNN includes a number of proliferation building blocks like convolution layers, pooling layers, and fully connected layers. It is commonplace to design a number of convolution layers and a layer for pooling repeatedly, followed by a few fully linked layers. Forward pass is the term for using these levels to convert the input data into the output[1].

II. THEORETICAL BACKGROUND

Researchers have created methods for automatic interpretation of medical images fed into a computer, beginning from the 1970s to 1990s. Medical image interpretation during that period involved cut-and-try approaches that included image enhancement techniques involving edge and line detector filters and mathematical modeling involving the fitting of lines, circles and ellipses so as to build up compound rule based systems which solved specific problems. There's a similarity with those expert systems that were constructed using a multitude of if-then-else, which were quite popular in artificial intelligence at that time. These expert systems have been characterized as GOFAI (good old-fashioned artificial intelligence) [1] of which many tended to be brittle; very much like rule-based image processing systems [2].

In the late 1990s, there was an uptrend in the use of supervised techniques in medical imaging in which a model is constructed from training data. Examples include active shape models (for segmentation), atlas methods (where the atlases that are fit to new data form the training data), and the concept of feature extraction and use of statistical classifiers (for computer aided detection and diagnosis). Such pattern recognition or machine learning approach is still very much in vogue and forms the foundation for many medical image analysis systems in the market today. So, we have moved from completely human-driven systems to systems that employ supervised computer training using sample data to form feature templates. The algorithms employed select the best possible hyperplane in the high dimensional feature space. A crucial step in the design of such systems is the extraction of discriminant features from the images. This process is still done by human researchers and, as such, one speaks of systems with handcrafted features. The next step would be for machines to determine the features appropriate for optimal representation of the relevant problem.Many deep learning algorithms rely on this principle, based on models or networks that consist of multiple layers which change the given data such as pictures.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⵋ Peer-reviewed & Refereed journal ⵋ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

This performance includes outputs (e.g. disease present/absent) while learning increasingly more abstraction level features. So far, the best models for image analysis are Convolutional neural networks (CNNs) [2].
The interpretation of medical images is crucial for the diagnosis and monitoring of diseases and disorders. X-ray, MRI, CT and ultrasound images are the examples of medical images.Since medical data is extensive, intricate, and multi-faceted, it is difficult to interpret manually, which has led to quick growth of automated methods based on deep learning architectures, such as CNNs [3].
Due to its unique architecture, CNNs have emerged as the primary framework through which deep learning is able to deliver image analysis. This is more critical in the medical realm where the images have subtle features and patterns or even anomalies that the clinicians might not be able to see.
A CNN has convolutional, pooling and fully connected layers. Feature extraction is performed in the convolution process. Each convolutional layer applies a filter (or kernel) to the input image (or feature map from the previous layer), producing feature maps that highlight edges, textures and shapes of the images [3].

III. EXISTING TECHNIQUES

A. LeNet-5

Yann LeCun and his fellow researchers created a convolutional neural network called LeNet-5 in the late 80s which was aimed at recognizing numbers but was able to be used for a range of different image processing tasks including images of lesions (LeCun et al, 1989). Comprising two convolutional layers, pooling layers and fully connected layers, its application in medical imaging was not broad as it could only allow for basic imaging and pathology lesions detection via images such X-rays and MRI[11].

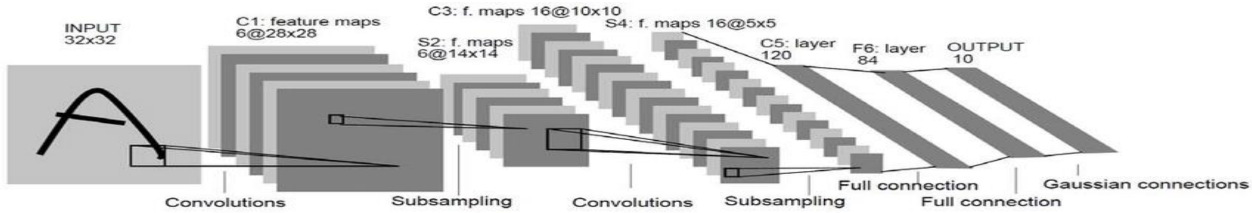

Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, for digits`digits`recognition An important milestone was achieved in 2012 as the team from Toronto (Alex Krizhevsky et al.) won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[31].

B. AlexNet

Another known neural network further assisting medical imaging is the AlexNet. In the year 2012, the neural network gained attention after clinching number one in the ImageNet challenge of the same year, hence demonstrating the capabilities of deep learning networks in modern medical practice(Krizhevsky et al. 2012). An application of the AlexNet in medical practice includes identifying diseases embedded in chest x-rays, MRI and mammograms[12].

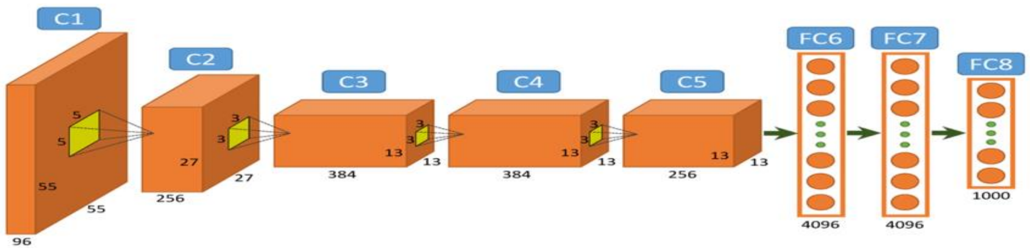

Fig 3. Architecture of Alexnet. From left to right (input to output) five convolutional layers with Max Pooling after layers 1,2, and 5, followed by a three layer fully connected classifier (layers 6-8). The number of neurons in the output layer is equal to the designed number of output classes[32].

C. VGGNet (VGG16, VGG19)

VGGNet, which was presented by Simonyan and Zisserman (2014), is famous for its elegant and care-free design which employs uniform use of multiple 3x3 convolution filters. The depth of the model (which can be either 16 or 19 layers) alongside the small size of the filters help the model in acquiring very detailed spatial information, making it robust for medical images. It has been employed in the localization of tumors and in disease classification based on different forms of medical images such as MRI, CT scans, X-rays[13].

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

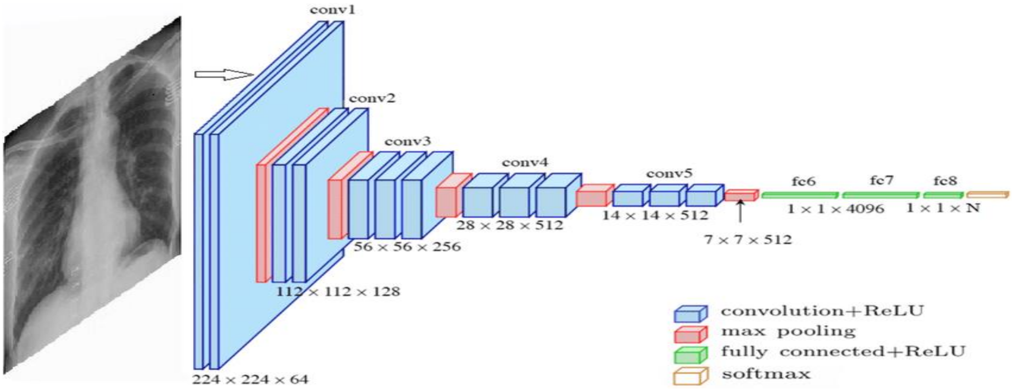

Fig 4. The structure of the VGG16 model[33].

D. ResNet (Residual Networks)

ResNet, proposed by He et al. (2015), adding skip connections alongside residual learning allows these networks to be very deep and solves the issue of vanishing gradients. This explains the growing use of the ResNet model for medical images; even complex tasks such as training on the model for the purposes of deeper radiology scans can be done accurately[14].

ResNet is broadly applied for tasks such as detailed analysis of radiology scans to identify diseases in brain MRIs or CT scans. Therefore, it ensures accurate outcomes for very deep architectures in critical medical applications.

Along with other very important scientific advances, the ResNet (Residual Network) introduced a new paradigm of deep learning with the concept of residual learning with skip connection, which enables the training of very deep networks without the problem of vanishing gradients. The working principle of ResNet commences from an input image into the network which is processed in the initial stage involving some convolution and pooling which are intended to capture low-level features[29].

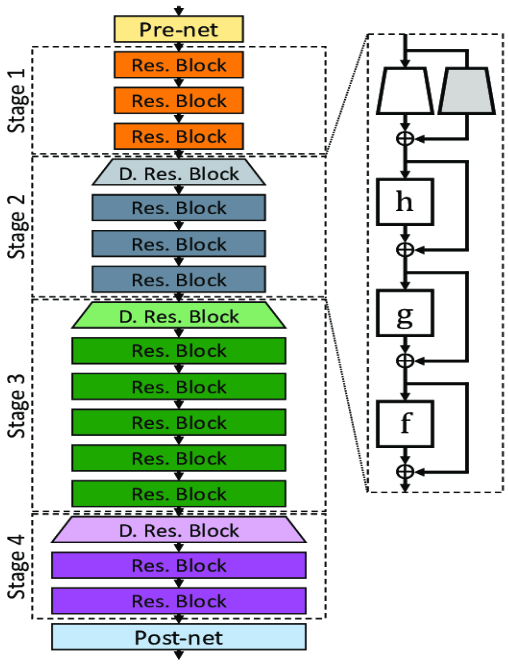

Fig 5. The residual network architecture. Specifically, ResNet-50, which has 3, 4, 6, and 3 blocks in each stage, from input to output[30].

What makes ResNet stand out is its use of residual blocks through which each block receives its input composed of one or more convolutional layers that have been skipped. As a feature, this "shortcut" enables the maintenance of important low level information while also guaranteeing effective gradient flow in backpropagation, thereby allowing the network to be able to learn higher levels of representation[29].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

Every two to three convolutional layers within a residual block, batch normalization and ReLU activation are applied before the residual block is completed. These are particularly effective in learning the residual or difference between the input and output, thus making the learning process easier. The multiple residual blocks are compiled then processed, the output feature maps are then pooled globally to form a vector of a set length. That vector history is sent through dense layers with the intent of classifying the images or performing other functions. The construction of the network is modular, which allows for efficient scaling, and such networks like ResNet-50, ResNet-101, and ResNet-152 of increasing depth are suitable for large tasks[29].

E. U-Net

Developed by Ronneberger et al. (2015), the U-Net is a type of convolutional neural network which features a U Shaped architecture. The use of spatial encoder-decoder architecture with skip connections makes this neural network very effective in basic tasks such as segregation of anatomy. This Neural net has also been widely used for medical purposes for tumor segregation, organs that need outlining and recognition of anatomy lesions on ultrasound scans, CT, and so forth[15].

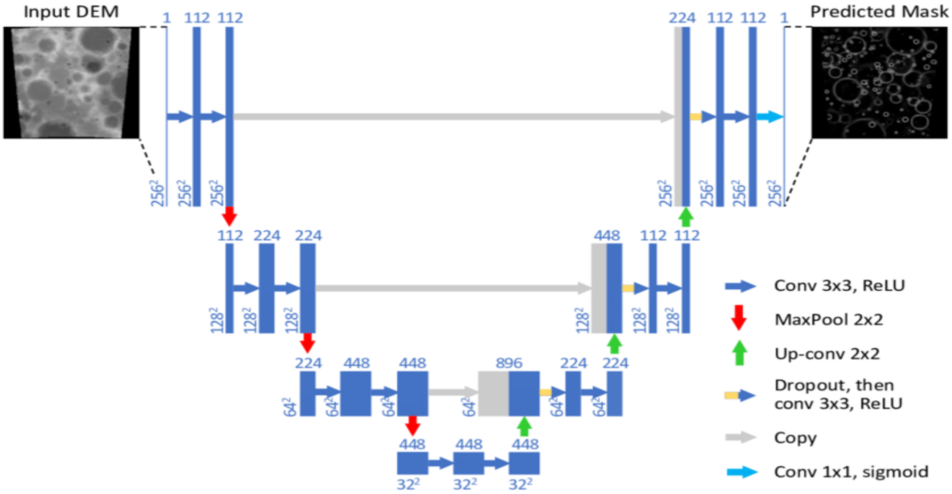

Fig 6. Convolutional neural network (CNN) architecture, based on UNET. Boxes represent cross-sections of square feature maps.[27].

U-Net is characterized by its “U” shaped deep neural network architecture, consisting of an encoder and a decoder. The encoder downsampled the input image in order to retrieve its features, whereas the decoder upsamples and recovers the spatial dimensions of the previous image by making a prediction of a segmentation map in which every pixel of the image is associated with a certain label, like a tissue, an organ or a background[26].

MaxPooling Convolutional Neural Network, which is an encoder, is an imaging sequence of convolutional layers with max-pooling which downsamples the image and still maintains high-level features. These layers are responsible for important features regarding the structure of the image. The decoder on the other hand uses transposed convolution layers to upsample the feature maps step by step. This combines features for certain spatial dimensions of higher resolution while providing necessary detail for segmentation[26].

A significant feature of U-Net is the use of skip connections. Connections allow direct forwarding of the feature maps from the encoder to the decoder on the same scaling levels. This prevents the loss of fine details during downsampling from being recovered in the respective upsampling resulting from the U-Net, improving the segmentation accuracy[26].

F. FCN (Fully Convolutional Networks)

According to Long et al. (2015), Fully convolutional networks as the name suggests, have done away with the fully connected layers that traditional neural networks contain and replaced them with convolution layers for each pixel, making them apt for classification and segmentation tasks. In the fields of medical diagnosis, FNCs have found wide application, especially in segmenting tissues, lesions, and organs such as in CT scans for lungs and brain tumors[16].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

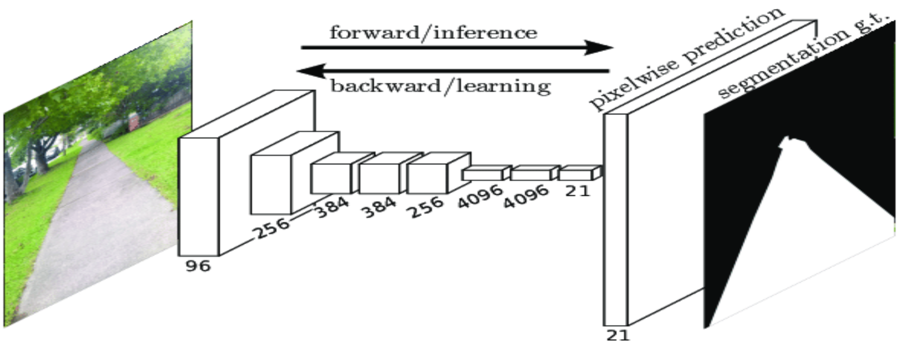

Fig 7. Architecture of FCN[25].

An FCN needs a standard convolution and pooling layers as its first step in the working process in order to get necessary features from the input image. These layers preserve the vital semantic high-level information while lowering the spatial resolution. In order to do this the FCN has to perform deconvolution layers (or also known as transposed convolutions) so as to up sample the feature maps to the original image dimensions. This up sampling assists the network in producing a segmentation map in which every pixel has been assigned a class such as tissue organ, and background tissue to name but a few.

In FCN architectures, the u-net contains skip connections which are used to link earlier layers to later ones. By doing this, low level information from shallow layers and high level information from deep layers are fused which increases the accuracy of the segmentation. They are applicable in CT scan and MRI images as well as ultrasound data to segment and locate organs, analyze tissues and tumors.

FCNs can offer a high degree of accuracy on pixel-level classification however one important aspect they consider is the way they handle the boundaries in order to provide very good smooth segmentation. This notwithstanding owing to the quality and accuracy of the predictions made by FCN they are fast turning out to be the norm in the medical imaging field.

G. Inception Network (GoogLeNet)

GoogLeNet, introduced by Szegedy et al. in 2015, features a unique architecture known as the "Inception module." This module employs various convolution filters of different sizes at each layer, enabling the model to effectively capture multi-scale features. Inception has found applications in medical image analysis, especially for tasks such as disease classification and organ detection in CT and MRI scans[17].

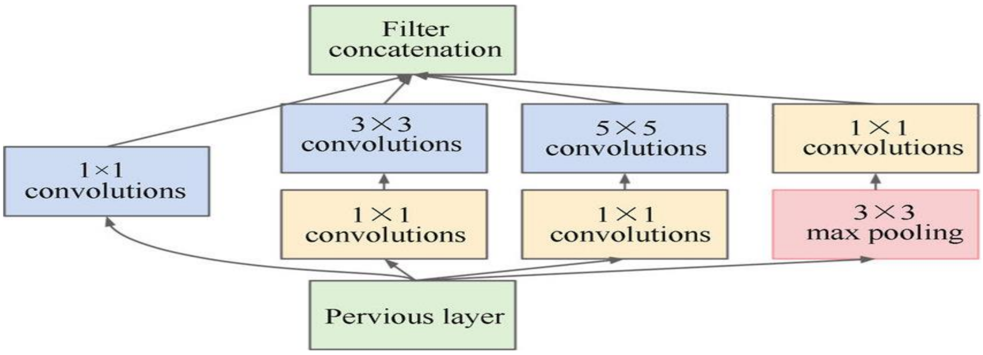

Fig 8. Basic 3D CNN architecture[23].

The Inception Network, known as GoogLeNet, works on a specific design called the Inception module, which processes an image at multiple scales simultaneously. When an image is fed into the network, the Inception module applies filters of different sizes: 1×1, 3×3, and 5×5.

5×5, pool operation. This helps allow the network to look not only at small details within the image but also much larger patterns, ensuring that the network doesn't miss very important features. After processing, the outputs coming from these filters are taken together, creating a strong and detailed representation of what's going on in that image.

To make it efficient, GoogLeNet adopts 1×1 convolutions to reduce the number of calculations before applying larger filters.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

This approach lowers the computational cost without reducing accuracy. Another feature of the network is global average pooling that replaces traditional fully connected layers at the end. This reduces the size of the network and also minimizes overfitting.

The GoogLeNet also includes auxiliary classifiers in the middle of the network. They are more miniature networks that guide learning, making sure even deeper layers of the network will be well-trained. Additionally, this avoids problems, including vanishing gradients, where gradients vanish in very deep networks. As for these features, GoogLeNet is able to analyze complex images efficiently with a great number of its applications, such as finding diseases in medical imaging and also recognising objects in real time.

H. Dense Net

According to Huang et al. (2017), DenseNet is a new architectural design that connects every layer to all the other layers in a feedforward manner in order to increase the rate of feature reuse and improve the efficiency of the network in both training time and performance. Images in dense networks have shown good performance in various medical image classification tasks, such as chest X-ray abnormalities, brain scans, and mammograms [18].

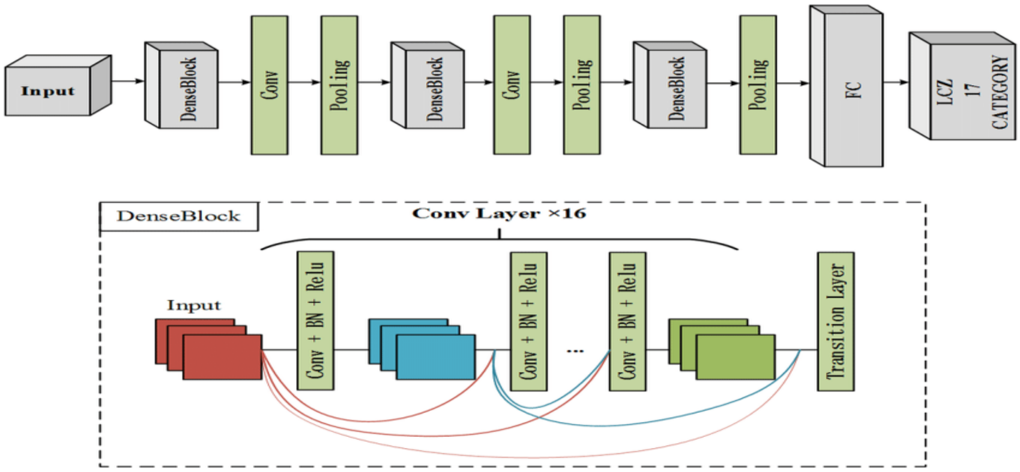

Fig 9. Framework of Densenet, which contains 3 Dense blocks, and each block contains 16 convolution layers[24].

This design encourages feature reuse, in that each layer can utilize the outputs of all previous layers without needing to relearn redundant information. The network is therefore more efficient both in terms of computation and memory usage.

When an image is processed in DenseNet, each layer receives the combined features from all the earlier layers and adds its own learned features. These combined outputs are passed to the next layer. This dense connectivity helps the network extract both low-level and high-level features, which improves its performance in tasks like image classification and segmentation.

Another important feature of DenseNet is that it applies batch normalization, ReLU activation, and 1×1 convolutions to reduce the size of feature maps before adding new layers, making sure the network remains computationally efficient. The network finally applies global average pooling and fully connected layers for prediction.

I. 3D U-Net

A 3D U-Net referred to as Cicek et al (2016), is a variant of U-Net that allows for volumetric image segmentation. It employs 3D convolution which makes it appropriate for medical image segmentation in volumetric imaging data sets like mri or CT. It is considered a standard method to segment organs or cavities for nuclear imaging diagnostic tasks[19].

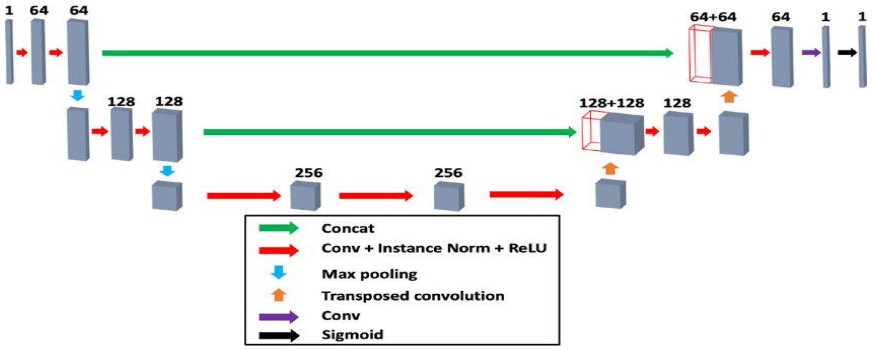

Fig 10. The architecture of the standard 3D U-Net[22].

The 3D U-Net performs a much needed task in the medical field as it segregates target organs in a medical image. Broadly speaking, the task is performed with the aid of MRI and CT imaging, and is a form of segmentation. Essentially a 3D U-Net is simply a broader and more complex 2D U-Net in the sense that it takes 3 dimensional volume data while being able to work on organ targeting and solving a multitude of other tasks.

First of all, it's important to remember that a 3D U-Net consists of an encoder-decoder architecture. Because this again is a neural network that aims to target depth information in medical imaging.

1) ***Starting with the encoder:*** Since the task is based on depth, volumetric images get passed through multiple 3D convolutions along with a max pooling layer. Such steps as these also allow for spatial dimension reduction to aid targeting. Eventually the encoder captures all relevant features from the initial input, starting from the most basic to the most complex structures.

2) ***Bottleneck layer: From a higher level perspective:*** when looking at the entire architecture at once, this can be thought of as the most condensed version of the image. Most parameters are lost and only the significant features of the image are captured.

3) ***The Decoder:*** As the aim is to reach a complex volumetric organ targeting tasks, this is why the image gets progressively upsampled through 3D up-convolutions. Such methods allow for the integration of the detailed features captured by the encoder upon reconstruction.

The skip connections are fundamental components of 3D U-Net. They transport some features from the encoder directly to the decoder which makes it possible for the network to keep in memory the original shape of the image. This guarantees that the last result is complete and precise.

Eventually, the output returned by the network is a three-dimensional one, for instance a segmented volume within which it is possible to delineate the regions of interest such as tumors, organs, or cavities.

Due to its ability to deal with large volumes of data with good segmentation outputs, the 3D U-Net is increasingly becoming the go to approach in medical imaging applications. It has achieved relatively good results in tasks such as brain tumor segmentation, lung nodule detection and cardiac structures analysis in nuclear imaging.

J. 3D CNN

The 3D CNNs are the traditional networks that are characterized by 2D convolutions but extended to include the processing of three-dimensional medical images scans such as MRI or CT. These types of models employ 3D convolutions to model slice spatial relationships and have been used for a wide range of applications such as tumor detection and organ segmentation. This has a high application in volumetric data of imaging such as in brain tumor detection or lung segmentation[20].

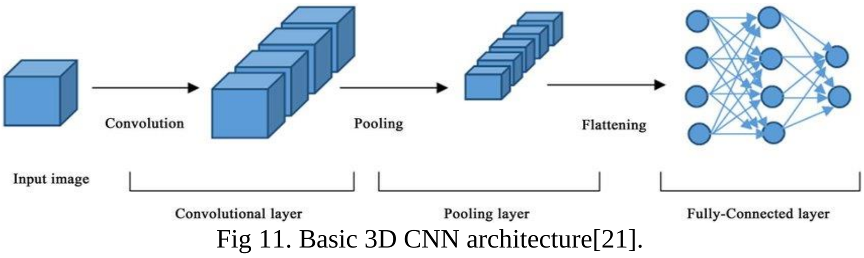

Fig 11. Basic 3D CNN architecture[21].

The 3D CNN (Three-Dimensional Convolutional Neural Network) uses volumetric data such as those obtained from MRIs and CT scans by extending the consideration of the spatial relationships to height, width and depth.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14318

A volumetric image, which can be viewed as a stack of multiple two-dimensional images, is represented in the form of a three-dimensional array which serves as the input to 3D CNN. The network starts off using 3D convolutional filters that span all 3 dimensions and perform feature extraction such as texture, shape, and spatial context. With each convolution a ReLU activation function is added to the output in order to inject some non-linearity and hence enable the network to learn more complicated patterns.
The network goes on to employ 3D pooling layers such as max-pooling or average pooling for down sampling of the feature maps. This step contributes in terms of making the network less computational intensive. The convolution and pooling operations are further carried out in deeper layers to get more higher level features of the data. In the end, the structured feature maps are converted to 1D arrays and through global pooling layers are used to classify or create a vector for some other descriptive measure like regression.
A 3D CNN may eventually classify a tumor as a yes / no depending on its designed output. So it can tell the probability score for the examples that are meant for classifying the subjects (ex: tumor vs no tumor) or for 3D segmentation tasks, it may create a volumetric map where every voxel is assigned a number as an index of a category. This systematic representation of the 3d cnn model is suitable for processing volumetric errors and reconstruction patterns along with the context in the era of medical application for recognizing the existence of tumors, performing organ partitioning as well as classifying certain ailments.

K. Deep CNNs and Their Evolution

The beginning CNN architectures, like AlexNet, VGG, and ResNet, made primary blocks and initiated the journey toward developing enthralling networks focusing on specific domains; that is, for medical purposes. Contemporary trends drift deeper toward networks with much more intricate architectures for attempting to capture hierarchy in features embedded in more complex medical images.

L. Transfer Learning

Transfer learning for medical image application is being proved as popular by pre trained CNNs such as ResNet, DenseNet, and Inception models. Transfer learning uses pretrained models to adapt a very small but more domain-related dataset of medical images since these pretrained models are usually developed using a different, much larger dataset.

M. Attention Mechanisms:

CNNs have been combined with attention mechanisms to highlight an image's most critical features. This is particularly true for certain areas in a medical image, such as a tumor or lesion, when an image is to be accurately evaluated.

N. Hybrid Models

More 'hybrid' architectures where CNNs are used with other approaches such as reinforcement learning, generative adversarial networks (GANs), and transformers are increasing. This encourages even better, more versatile and efficient solutions.

IV. ANALYSIS OF DEEP LEARNING TECHNIQUES

TABLE I. Analysis of Medical Images Detection Techniques

Sr. No.	Research Title	Publication & Year	Learning Paradigm	Method	Challenges
1	Optimized Deep Learning Model for Comprehensive Medical Image Analysis Across Multiple Modalities	Elsevier - 2024	Supervised Learning, Transfer Learning, Multimodal Learning	CNNs, Multi-Task Learning, Hybrid Models, Attention Mechanisms, GANs [4]	Data Imbalance, Data Privacy, Multimodal Fusion, Model Generalization, Interpretability, Computational Resources

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

2	Enhancing Early Detection of Brain Aneurysms: A CNN-Driven, Real-Time Approach with Angiography Imaging	IEEE - 2024	Supervised Learning	CNNs, Real-Time Imaging Processing, Feature Extraction from Angiography Images[5]	Real-Time Processing, Data Imbalance, Need for High Accuracy, Model Generalization, Limited Dataset
3	Medical Internet of Things Using Deep Learning Techniques for Skin Cancer Detection	IEEE - 2022	Supervised Learning, IoT Integration	CNNs, Deep Learning-based Skin Cancer Detection, Internet of Things Integration [6]	Data Privacy, Hardware Integration, Generalization across Devices, Data Imbalance
4	Revolutionizing diabetic retinopathy diagnosis through advanced deep learning techniques: Harnessing the power of GAN model with transfer learning and the DiaGAN-CNN model	Elsevier - 2024	Deep Learning, Transfer Learning, Generative Adversarial Networks (GANs)	GAN-based model combined with transfer learning and DiaGAN-CNN architecture [7]	Data imbalance, high computational cost, model interpretability, and achieving high accuracy in complex retinal images
5	Deep Learning-Based Diagnosis of Alzheimer's Disease Using FDG-PET Images	Elsevier - 2023	Supervised Learning	CNNs, FDG-PET Image Analysis, Feature Extraction for Alzheimer's Detection [8]	Data Imbalance, Image Quality Variability, Data Privacy, Lack of Large Annotated Datasets
6	Brain Tumor Segmentation from MRI Images Using Hybrid Convolutional Neural Networks	Elsevier - 2020	Supervised Learning	Hybrid CNNs for Tumor Segmentation, Data Augmentation Techniques [9]	Data Imbalance, Variability in MRI Quality, Generalization across Different MRI Devices, Accurate Segmentation
7	TBConvL-Net: A Hybrid Deep Learning Architecture for Robust Medical Image Segmentation	Elsevier - 2024	Supervised Learning	Hybrid CNNs, ConvL-Net Architecture for Robust Image Segmentation [10]	Data Imbalance, High Computational Cost, Lack of Large Datasets, Model Generalization, Accuracy in Segmentation across Modalities

V. CONCLUSION

CNNs have revolutionized the analysis of medical images, which in turn has helped to significantly improve the accuracy of diagnosis, detection of diseases and automatic interpretation. This survey that was conducted sought to assess the changes in the application of CNN in this field and the developments, problems and new ideas around there.
Hybrid models such as those removing the limitations of supervised training with recurrent networks have been used in many cases where both spatial and temporal features are essential while U-Net architectures have been used for segmentation purposes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

The increasing proportion of these specialized architectures reinforce the rationale to treat complex tasks as specified problems. Deep supervision and multi-scale learning are examples of approaches aimed at tackling the multi-scale feature nature of medical images in order to enhance robustness and accuracy of the models being developed.
There is good potential for CNN based models in the field of medicine especially in the analysis of images. Still, there are several hurdles such as data scarcity, explain ability, and legal endorsement that must be addressed in order to make this a reality. From a clinical point of view, considering a wide variety of tasks, it is also necessary to develop interpretable models that will work with confidence across the healthcare system. The fact that researchers, clinicians, and regulatory agencies will collaborate is the key to the development of the CNN based analysis of medical images so that these systems will assist in availing better healthcare and to have better results in dealing with patients. As technology continues to change so will the medical field, including CNN’s integration into everyday practice.

ACKNOWLEDGMENT

I would like to thank **Dr. Vikas Tulshyan**, PG Co-ordinator at Silver Oak University, for his guidance, encouragement and support offered to me during my research. He is a studious person with great concern for details and has provided many valuable suggestions during the course of this work.
I am also thankful to my guide**, Dr. Anuradha Desai**, whose collaborations were instrumental throughout this study in providing support, expert guidance and feedback. Her detailed understanding of this topic and tireless efforts to help the students were key factors for the successful completion of this research project.
For their supervision and advices which were a source of encouragement to me, I appreciate their efforts in this work. This overrides my opinion expressed earlier regarding their contributions to this work.

REFERENCES

[1] Rguibi, Z., Hajami, A., Zitouni, D., Elqaraoui, A., & Bedraoui, A. (2022). "CXAI: Explaining Convolutional Neural Networks for Medical Imaging Diagnostic," Electronics, vol. 11, no. 11, pp. 1775. https://doi.org/10.3390/electronics11111775.

[2] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … & Sánchez, C. I. (2017). "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88. https://doi.org/10.1016/j.media.2017.07.005.

[3] Tajbakhsh, N., Shin, J. Y., Wu, J., & Hurst, R. T. (2016). "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299-1312. https://doi.org/10.1109/TMI.2016.2535302.

[4] Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Li, X. (2024). "Optimized deep learning model for comprehensive medical image analysis across multiple modalities," Neurocomputing, vol. 619, pp. 129182. https://doi.org/10.1016/j.neucom.2024.129182.

[5] B, A. K., Gudodagi, R., T, M. H., D, K. D., & M, M. H. (2024). "Enhancing Early detection of brain aneurysms: A CNN-Driven, Real-Time approach with angiography imaging," IEEE, pp. 1389–1394. https://doi.org/10.1109/icicnis64247.2024.10823359.

[6] Veeraiah, V., Ravikaumar, G. K., Kalpana, R., Sreenivasulu, K., Singh, Y., & Shukla, S. K. (2022). "Medical Internet of Things using Deep Learning Techniques for Skin Cancer Detection," IEEE, vol. 18, pp. 317–321. https://doi.org/10.1109/ic3i56241.2022.10073052.

[7] Shoaib, M. R., Emara, H. M., Mubarak, A. S., Omer, O. A., El-Samie, F. E. A., & Esmaiel, H. (2024). "Revolutionizing diabetic retinopathy diagnosis through advanced deep learning techniques: Harnessing the power of GAN model with transfer learning and the DiaGAN-CNN model," Biomedical Signal Processing and Control, vol. 99, p. 106790. https://doi.org/10.1016/j.bspc.2024.106790.

[8] Kishore, N., & Goel, N. (2023). "Deep learning based diagnosis of Alzheimer’s disease using FDG-PET images," Neuroscience Letters, vol. 817, p. 137530. https://doi.org/10.1016/j.neulet.2023.137530.

[9] Daimary, D., Bora, M. B., Amitab, K., & Kandar, D. (2020). "Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks," Procedia Computer Science, vol. 167, pp. 2419–2428. https://doi.org/10.1016/j.procs.2020.03.295.

[10] Iqbal, S., Khan, T. M., Naqvi, S. S., Naveed, A., & Meijering, E. (2024). "TBConvL-Net: A hybrid deep learning architecture for robust medical image segmentation," Pattern Recognition, p. 111028. https://doi.org/10.1016/j.patcog.2024.111028.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324. https://doi.org/10.1109/5.726791.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14318

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105. https://doi.org/10.1145/3065386.
[13] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556.
[14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90.
[15] Ronneberger, O., Fischer, P., & Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.
[16] Long, J., Shelhamer, E., & Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965.
[17] Szegedy, C., et al. (2015). "Going deeper with convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9. https://doi.org/10.1109/CVPR.2015.7298594.
[18] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). "Densely connected convolutional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708. https://doi.org/10.1109/CVPR.2017.243.
[19] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). "3D U-Net: Learning dense volumetric segmentation from sparse annotation," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424-432. https://doi.org/10.1007/978-3-319-46723-8_49.
[20] Zhu, W., et al. (2016). "Deep learning-based 3D segmentation of volumetric medical images," arXiv preprint arXiv:1606.04306. https://doi.org/10.48550/arXiv.1606.04306.
[21] "Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Basic-3D-CNN-architecture_fig1_339685457.
[22] "CycleGAN With a Blur Kernel for Deconvolution Microscopy: Optimal Transport Geometry," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/A-modified-3D-U-net-architecture-for-our-high-resolution-image-generator_fig2_342794245.
[23] "A survey on deep learning-based fine-grained object classification and semantic segmentation," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/nception-module-of-GoogLeNet-This-figure-is-from-the-original-paper-10_fig3_312515254.
[24] "Embranchment CNN based Local Climate Zone Classification using SAR and Multispectral Remote Sensing Data," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Framework-of-Densenet-which-contains-3-Dense-blocks-and-each-block-contains-16_fig3_332522654.
[25] "Pedestrian Lane Detection for Assistive Navigation of Vision-Impaired People: Survey and Experimental Evaluation," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Architecture-of-FCN-Figure-is-adapted-from-33_fig5_363729848.
[26] Ronneberger, O., Fischer, P., & Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
[27] "Lunar Crater Identification via Deep Learning," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Convolutional-neural-network-CNN-architecture-based-on-UNET-Ronneberger-et-al_fig2_323597886.
[28] "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26_fig1_336805909.
[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/cvpr.2016.90.
[30] "Recurrent Residual Networks Contain Stronger Lottery Tickets," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-residual-network-architecture-Specifically-ResNet-50-which-has-3-4-6-and-3_fig2_36858174.
[31] "DEEP LEARNING IN INDUSTRY 4.0 – BRIEF," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Architecture-of-LeNet-5-a-Convolutional-Neural-Network-for-digits-digits-recognition-An_fig1_329891470.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14318

[32] "Toward aircraft recognition with convolutional neural networks," Scientific Figure on ResearchGate. Available from:https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers_fig2_312303454.

[33] "Detection and analysis of COVID-19 in medical images using deep learning techniques," Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-structure-of-VGG16-modelThis-figure-was-created-with-Image-onlineco-and-exported_fig2_355049790.

[34] C. En Guo, S.-C. Zhu and Y. N. Wu, "Primal Sketch: Integrating Structure and Texture", Computer Vision and Image Understanding, vol. 106, no. 1, pp. 5-19, 2007.