Applied Journal of Physical Science
Volume 4(4), pages 42-50, December 2022
Article Number: 1378CAD82
ISSN: 2756-6684
https://doi.org/10.31248/AJPS2022.085
https://integrityresjournals.org/journal/AJPS

Review Article

Convolutional Neural Network (CNN): The architecture and applications

Joshua Ayobami AYENI

Department of Computer Sciences, Ajayi Crowther University, Oyo, Nigeria.

Email: ja.ayeni@acu.edu.ng

Received 13th November 2022; Accepted 29th December 2022

ABSTRACT: The human brain is made up of several hundreds of billions of interconnected neurons that process information in parallel. Researchers in the field of artificial intelligence have successfully demonstrated a considerable level of intelligence on chips and this has been termed Neural Networks (NNs). Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning (ML) and they are at the heart of deep learning algorithms. These subsets of ML have their names and structures derived from the human brain and the way that biological neurons signal to one another. A class of NNs that are often used in processing digital data images is the Convolutional Neural Network (CNN or ConvNet). The human brain processes a huge amount of information with each neuron having its own receptive field connected to other neurons in a way that they cover the entire visual field. Mimicking the biological technique, where the neurons only respond to stimuli in the restricted region of the visual field referred to as the receptive field, each neuron in the CNN processes data only in its receptive field. In this review paper, the architecture and application of CNN are presented. Its evolution, concepts, and approaches to solving problems related to digital images, computer vision and are also examined.

Keywords: Artificial neurons, computer vision, deep learning, machine learning, visual data.

INTRODUCTION

Artificial neural networks (ANNs), often referred to as neural networks, are computing systems inspired by the biological neural networks that constitute animal brains (Wang, 2003). An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain (Wang, 2003). Bezdan and Džakula (2019) defined Convolutional Neural Networks (CNNs) as a specific type of artificial neural networks (ANNs), that has demonstrated high performance on various visual tasks, including image classification, image segmentation, image retrieval, object detection, image captioning, face recognition, pose estimation, traffic sign recognition, speech processing, neural style transfer (Bezdan and Džakula, 2019). Convolutional Neural Network (CNN or ConvNet) could also be described as a class of neural networks that is often used in processing digital data images. A digital image has a grid-like topology is a binary representation of visual data. Neural networks are either hardware or software programmed as neurons in the human brain. Convolution, a linear mathematical operation, is employed on CNN instead of general matrix multiplication in one of its layers.

The human brain processes a huge amount of information with each neuron having its own receptive field connected to other neurons in a way that covers the entire visual field (Mayank, 2020). Mimicking the biological technique, where the neurons only respond to stimuli in the restricted region of the visual field referred to as the receptive field, each neuron in the CNN processes data only in its receptive field (Mayank, 2020). CNN, as a class of ANN, has become dominant in various computer vision tasks and attracting interest across a variety of domains, including radiology. CNN is designed to automatically and adaptively learn spatial hierarchies of features through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers (Yamashita *et al*., 2018). In view of the emergence of a strong interest in deep learning in recent years, CNN has become the most established algorithm among deep learning models that have equally transformed into a dominant method in computer vision tasks, language processing, medical research, radiology and image recognition, reconstruction and analysis (Yamashita *et al*., 2018).

THE LAYERS OF CONVOLUTIONAL NEURAL NETWORK (CNN)

As depicted in Figure 1, a CNN is made up of four major layers and these are a convolutional layer, a rectified linear unit, a pooling layer and a fully connected layer (Mayank, 2020). However, some authors classify the input and output layers as independent on their own – therefore four layers. As found in other NNs, a CNN is composed of an input layer, an output layer, and many hidden layers in between (Anirudha *et al*., 2021). A CNN can have tens or hundreds of layers that each learns to detect different features of an image with filters applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer (Kabir, 2021)

The CNN layers are arranged in such a way as to detect simpler patterns first (lines, curves, etc.) and move further along for the detection of more complex patterns (faces, objects, etc.) (Mayank, 2020). It is worthy of note to emphasize the distinct structure of regular ANNs and CNNs (Figure 1). In the former, each layer consists of a set of neurons and is connected to all neurons in a previous layer while the latter (CNN) uses three dimensional layers in a width, height, and depth manner. All neurons in a particular layer of CNN are not connected to the neurons in the previous layer but are only connected to a small portion of neurons in the previous layer (flatworldsolutions, 2022; Bezdan and Džakula, 2019).

The convolution layer

According to Bezdan and Džakula (2019), the convolutional layer is essentially the top layer often perceived as the mathematical layer and deals with understanding the number pattern identified. The first position in this layer starts applying a filter known as a neuron or kernel around the top left corner of the image. It reads that part of the image and forms a conclusion of an array of numbers, multiplies the array, and deduces a single number out of this process (Bezdan and Džakula, 2019; Anirudha *et al.*, 2020). The receptive field is the section of the image scanned by the filter and the process is repeated by increasing by one (1) until the entire image is read and assigned a single number to each unit. This single number represents the top left corner that the convolutional layer has just read of the image. The part of the image that the filter scans over is the receptive field. The filter then moves right by 1 unit and starts the same process again. In this fashion, the convolutional layer reads the entire image and assigns a single number to each unit. This data gets stored in a 3D array. In essence, this entire process functions like the human brain. The receptive field in the world of CNNs represents the visual field in the world of human biology, the filter acts as the visual cortex containing small regions of cells targeting reading of specific areas of the visual field (Bezdan and Džakula, 2019; Anirudha *et al.*, 2020).

Daniel (2021) describes the essential function of the convolution layer as the processing of an input image and then attributes a label to the processed image. The convolution layer is the nucleus or rather the core building block of the CNN and carries the main portion of the network's computational load (Mayank, 2020). This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field (Mayank, 2020). The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels (Mayank, 2020). In Figure 2, the image is pushed into the NN neural network as the input image, and then undergoes several convolutional operations and the final output is now used by the NN neural network to predict the digit.

The rectified linear unit layer

The Rectified Linear Unit Layer (ReLU) which is the next layer of a CNN is responsible for the instantiation of the activation functions having been initially set to a zero threshold. The ReLU is the most used activation function in the world right now because it is used in almost all CNN or deep learning. ReLU combines non-linear and rectification layers on CNN. This does the threshold operation where negative values are converted to zero. However, ReLU does not change the size of the input. ReLU is important because it does not saturate; the gradient is always high (equal to 1) if the neuron activates. The activation gradient only functions at 0 and 1 and does not include intermediary gradients like its predecessors (Anirudha *et al.*, 2020; Kabir, 2021). The activation function in a multiple-layered NN is responsible

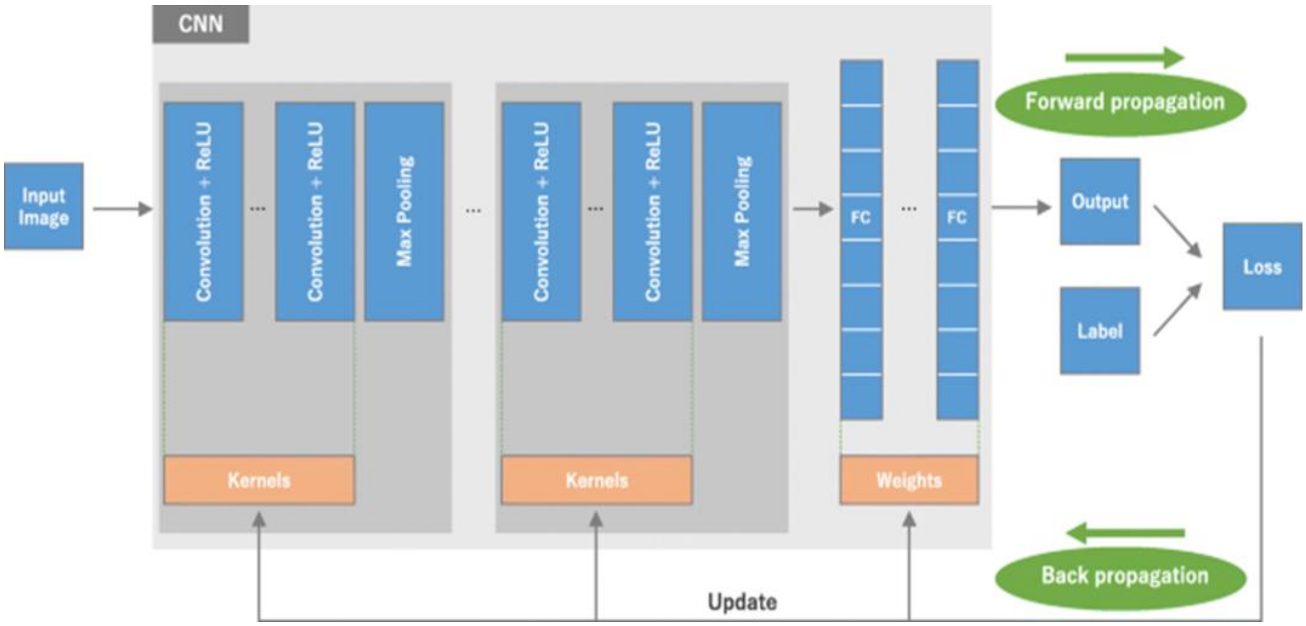

Figure 1. Convolutional neural network (CNN) architecture and the training process (Source: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9/figures/1).

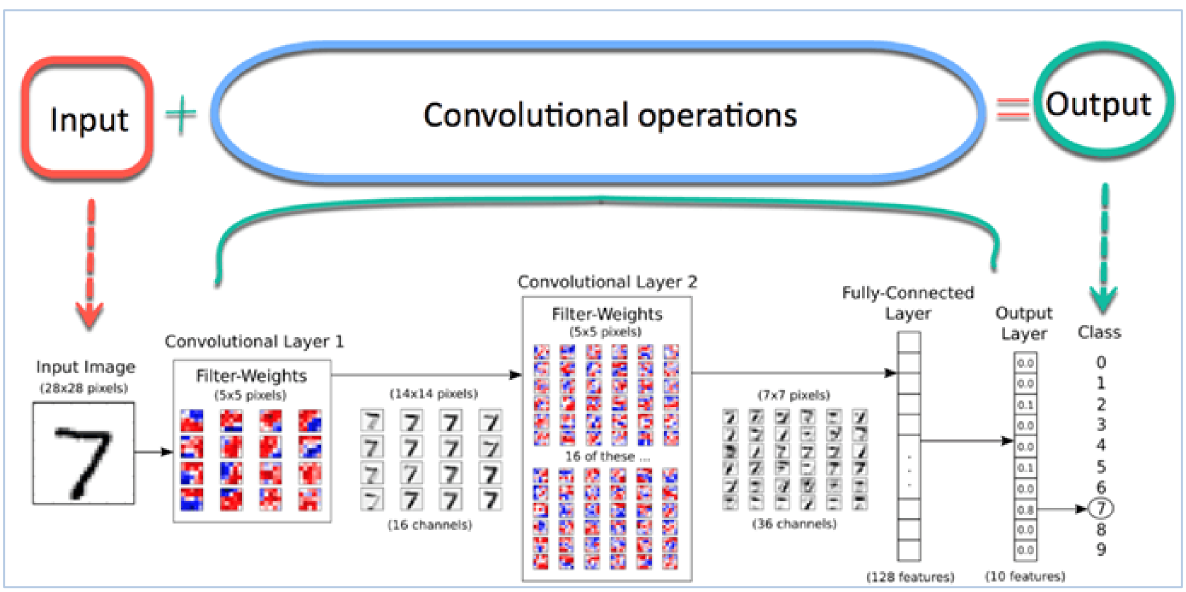

Figure 2. The architecture of a Convolutional Neural Network (CNN) (Source: Daniel, 2021).

for the transformation of the summed weighted input from the node into the node's activation or output for that input. ReLU is a rectified linear activation function, that outputs directly if the input is positive and outputs zero if the input is negative (Ohri, 2021). Most of the existing NNs have multi-layered structures and node layers that help the algorithm and network map learn the example outputs from the inputs. The summed activation of the nodes is obtained with the use of the summed-up weighted inputs. The activation function uses this summed/sigmoid activation function to define the node's activation, which then provides a specific output (Ohri, 2021).

According to Ajay (2021), nonlinear activation functions, like the ReLU, are preferred to train the learning nodes on

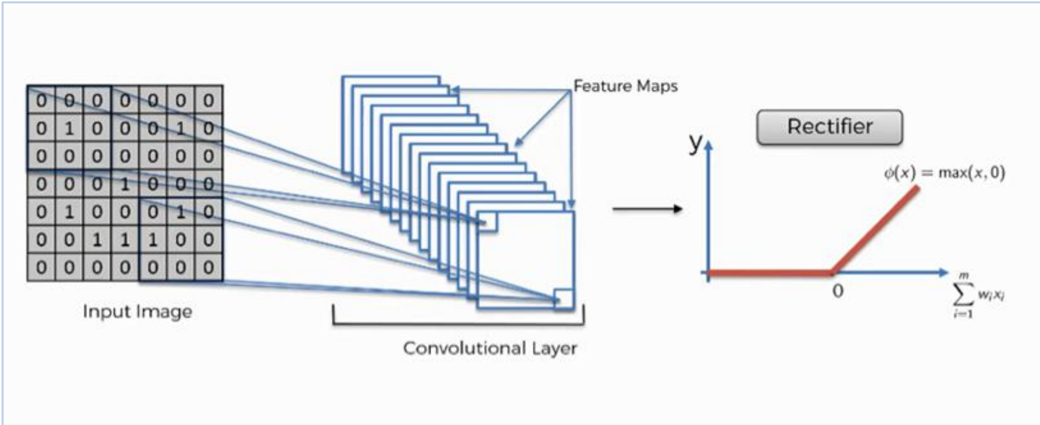

Figure 3. Image depicting CNN – ReLU (Source: Dhaoui, 2019).

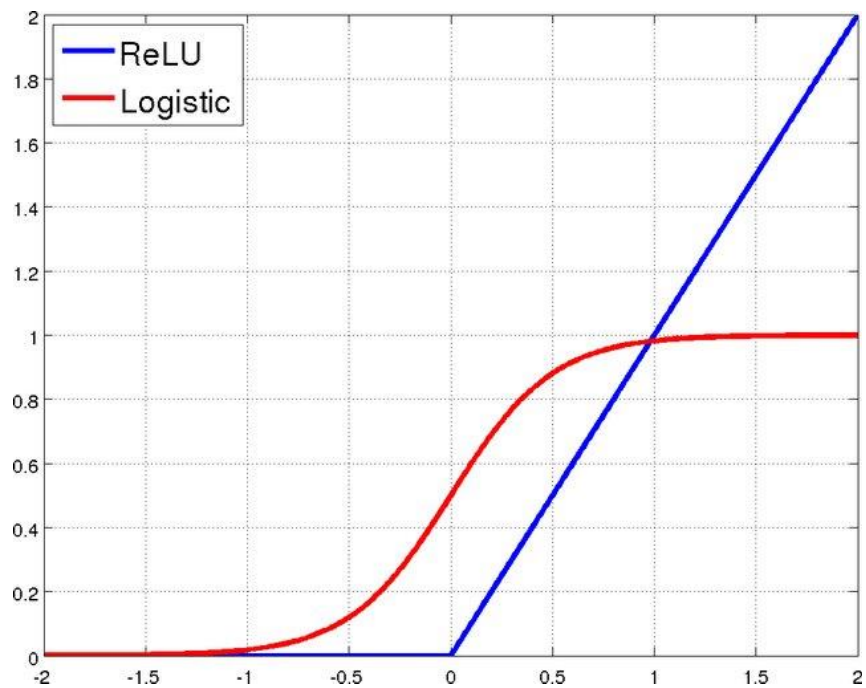

Figure 4. The ReLU f (x) = max (0, x) and the Logistic function f (x) (Source: https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQFDBvH84wwggSLWr2mGfliSWrkgNYpkOlRLziyWotdkQ&s).

the data's complex structures and the logistic sigmoid activation function causes the input's value to be transformed into values between one and zero. For inputs that are larger than one, it is transformed to one, and the small input values, are transformed to value zero. Ajay (2021) further stated that for output values lying between 1 and -1, the tanh function works well and produces a similar curve and is used because its predictive performance is better, and the model using it is easy to train. However, both these functions saturate and are responsive to change around the input middle values only. Again, when at the saturation stage, the algorithm often failed to adapt to the weights, and hence activation for the learning algorithm slows down (Ajay, 2021) as shown in (Figures 3 and 4).

In the convolutional layer, the data comes as input, then multiplied by the weight in the convolutional layer. The output is then processed by the activation function for nonlinearity. The usage of the ReLU function started to gain influence in image processing around 2012 with the

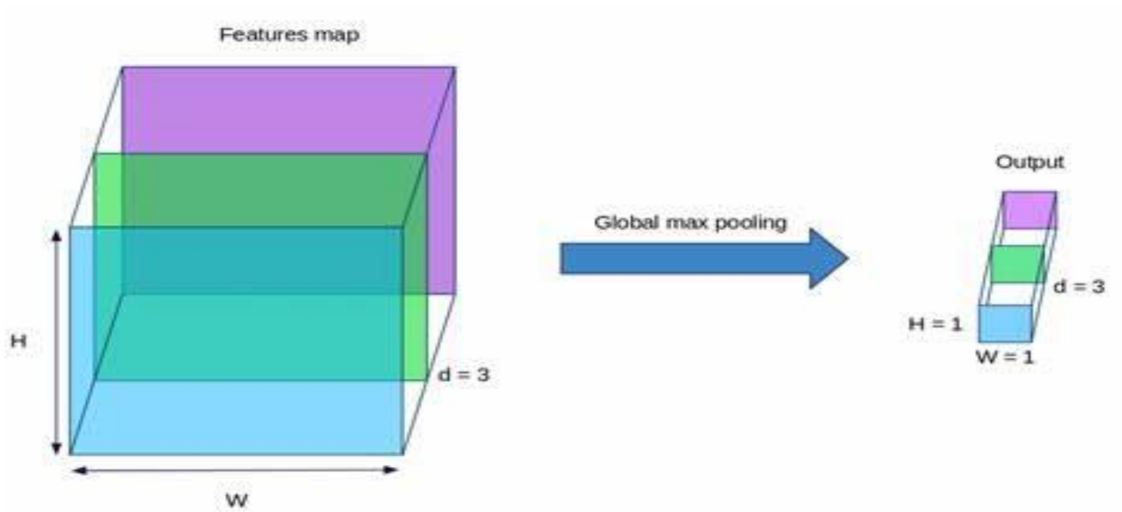

Figure 5. Feature Down Sampling at the Pooling Layer (Source: i2tutorials, 2019).

introduction of the AlexNet, the first major CNN that was able to do well on ImageNet and large-scale data. In a nutshell, ReLU is used for filtering information that propagates forward through the network. It takes an elementwise operation on your input and basically if your input is negative, it is going to put it to zero and then if it is positive, it is going to be just passed through its identity. This one is pretty commonly used because it does not saturate in the positive region. The sigmoid was not zero-centered tanh fixed this and now ReLU has this problem again and that is one of the issues of the ReLU. ReLU does not activate for negative inputs, it is possible to end up with "dead neurons" that never fire.

The pooling layer

The pooling layer which is responsible for reducing the size of activation maps is also referred to as the down sampling layer (Pedamkar, 2021). This is achieved by the application of a filter and stride of the same length to the input volume and in essence, while shrinking the feature maps it always preserves the most dominant features (or information) in each pool step. The layer ignores less significant data; hence image recognition is done in a smaller representation of the original input image (Anirudha *et al.,* 2020). The pooling layer reduces overfitting since the amount of parameters is reduced using the pooling layer and also reduction in cost (Pedamkar, 2021). The input (feature map) is divided into rectangular pooling regions and either the maximum or average is computed which consequently returns the maximum or average (Pedamkar, 2021) (see Figure 5).

According to Anirudha *et al.* (2020), the pooling operation is performed by specifying the pooled region size and the stride of the operation, similar to the convolution operation. There are different types of pooling techniques used in different pooling layers such as max pooling, min pooling, average pooling, gated pooling, tree pooling, etc. Max pooling is the most popular and mostly used pooling technique. However, despite the results often obtained with the use of this pooling technique, the main drawback of the pooling layer is that it sometimes decreases the overall performance of CNN (Anirudha *et al.,* 2020).

Brownlee (2019) identified the problem with the output feature maps as that of sensitivity to the location of the features in the input and suggested an approach to addressing this sensitivity by down sampling the feature maps. This has the effect of making the resulting down sampled feature maps more robust to changes in the position of the feature in the image, referred to by the technical phrase "*local translation invariance"* (Brownlee, 2019).

The pooling layer functions upon each feature map separately to create a new set of the same number of pooled feature maps and the process involves selecting a filter (Pooling function) to be applied to the feature maps (Brownlee, 2019). The size of the filter is smaller than the size of the feature map; precisely, it is almost always 2×2 pixels applied with a stride of 2 pixels consequently decreasing the size of each feature map by a factor of 2 (Brownlee, 2019). Brownlee (2019) asserted that, in all cases, pooling helps to make the representation approximately invariant to small translations of the input. Invariance to translation means that if we translate the input by a small amount, the values of most of the pooled outputs do not change (Brownlee, 2019). The pooling layer is the stage at which a lower-resolution version of an input signal is created that

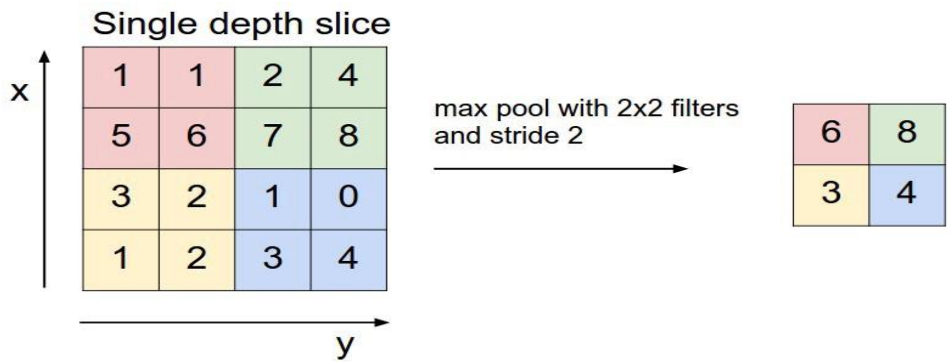

Figure 6. Representation of the Max Pooling Layer (Source: https://miro.medium.com/max/720/1*GksqN5XY8HPpIddm5wzm7A.webp).

still contains the large or important structural elements of the input image excluding the other elements that may not be useful to the task (Brownlee, 2019). The pooling operation is not learned but specified and there are two common functions used in the pooling operation; *Average Pooling* and *Max Pooling* that computes the average value on the feature map and the calculation of the maximum value for each patch of the feature map respectively (Figure 6).

The fully connected layer (FC)

A fully connected layer involves weights, biases, and neurons and connects neurons in one layer to neurons in another layer in classifying images between different categories by training. This is the final layer often referred to as the completion layer in a CNN. The input of this layer is the final output of the layer before it (be it a ReLU, pooling, or a convolutional layer) and provides an N-dimensional vector output (Anirudha *et al.*, 2020), where 'N' signifies the number of classes the program chooses from (flatworldsolutions, 2022). For example, if the program is looking at pictures of animals i.e. dogs, it will look at high-level features such as the 4 legs, the head, the tail, or muscle. This final layer referred to as a fully connected layer will look at the high-level features and connect that with the image thus giving the output of a classification of a dog (see Figure 7). In CNN, the output feature maps of the final convolution or pooling layer are transformed into a one-dimensional (1D) array of numbers (or vectors) and connected to one or more fully connected layers (Yamashita *et al*., 2018). It is also referred to as dense layers with every input connected to every output by a learnable weight. The final fully connected layer typically has the same number of output nodes as the number of classes followed by a nonlinear function, such as ReLU (Yamashita *et al*., 2018).

APPLICATIONS OF CONVOLUTIONAL NEURAL NETWORK

In today's digital world of applications and the requirement for high-performance applications to solve complex problems with the use of deep learning architecture, the application of CNN has gained a high level of usage compared to others of its type. The important application areas of CNN to be considered in this paper are:

Medical image analysis and classification process

CNN has been considered by researchers as the most performant tool because of its great learning capacity due to the utilization of many feature extraction stages that can learn representations from data automatically and with a higher percentage of accuracy (Sharma, 2021). Other image classification processes, CNN could be applied are Search Engines, Social-Media, and Recommender Systems. In medical computing, CNN has higher better accuracy in identifying tumours or other anomalies in X-ray and MRI images. CNN models have been applied in the analysis of human body part images, such as the lungs, and pinpoint where there might be a tumour and other anomalies like broken bones in X-ray images. Similarly, medical images like CT scans and mammograms can be used to diagnose cancer. In order to determine whether any indicators within a picture indicate malignancy or damage to cells owing to both hereditary and environmental factors, such as smoking habits, CNN models compare the image of a patient with database images that include comparable features (Gandharv, 2022).

Medical computing

Predictive Analytics, Healthcare Data Science, and Health Risk Assessment Using Predictive Analytics, CNN could

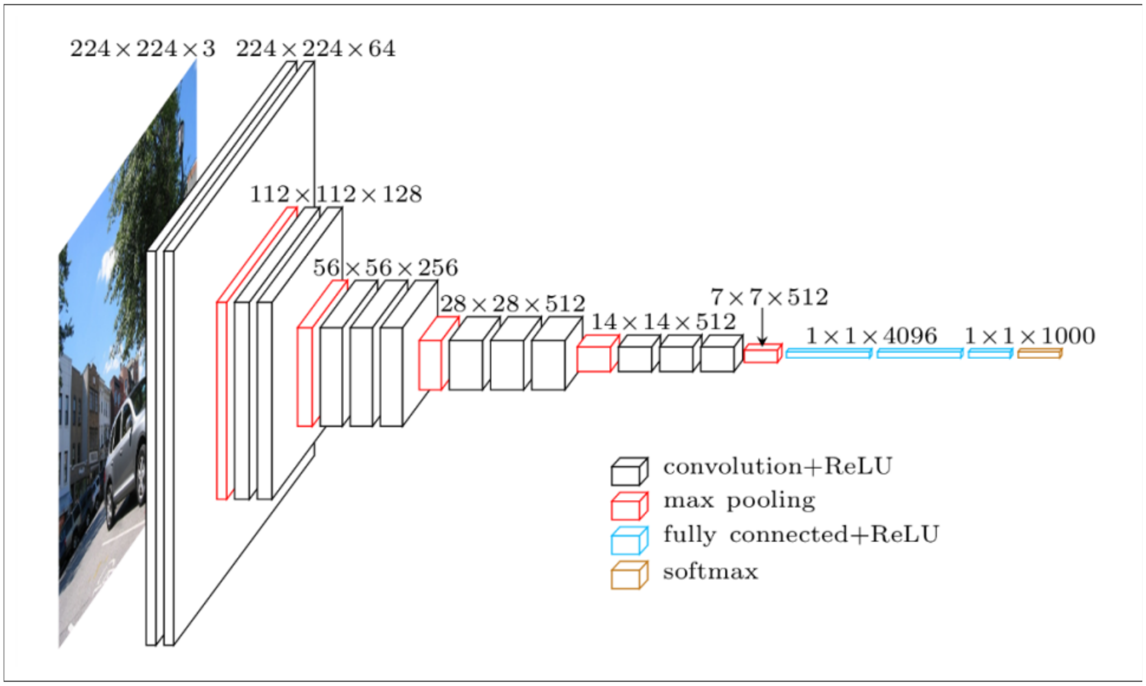

Figure 7. Convolutional Layers with Max Pooling (Source: https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529).

be applied to analyze medical images such as MRI (Magnetic Resonance Imaging, X-Ray, and CT scan). The result of the analysis and classification could be used by medical experts or radiologists for the purpose of predicting or establishing the severity of diseases such as Covid-19, Cough, brain/breast cancer, and tumor. Several researchers have carried out various types of work in this field (Li *et al*., 2014; Yadav and Jadhav, 2019; Kermany *et al.*, 2018; Afshar *et al*., 2019).

Face recognition RNN applications

CNN could also be applied in the development of Face Recognition Systems which is a special class of image recognition because it deals with complex images. This includes social media, identification, and surveillance. Such images could include human faces or other living beings such as animals, fish, and insects (Sharma, 2021).

Drug discovery using predictive analytics

Drug discovery using predictive analytics is another major healthcare field that makes extensive use of CNNs and is one of the most recognized inventive uses of convolutional neural networks (Sharma, 2021). According to Vamathevan *et al.* (2019), drug discovery and development pipelines are long, and complex and depend on numerous factors. Machine learning (ML) approaches provide a set of tools that can improve discovery and decision-making for well-specified questions with abundant, high-quality data (Vamathevan *et al.,* 2019). The process of developing a new drug is complex as there is a set of large datasets to be considered. CNNs have been used to extract information from various datasets of different dimensions has resulted in accurate interpretations in several subfields of biological research, like pharmacogenomics, addressing issues previously faced by other computational methods (Vaz and Balaji, 2021). The problem is that drug discovery and development is a time-consuming and costly process. In drug discovery, scalability and cost-effectiveness are critical. The process of developing new drugs lends itself well to the implementation of NNs. During the development of a new drug, there is a large amount of data to consider which could take a longer period with the current technique which could be controlled and reduced as a result of the application of CNN. The algorithm (CNN) is effective in optimizing and streamlining the drug discovery process at critical stages and allows for a reduction in the time required to develop cures for emerging diseases (Sharma, 2001).

Natural language processing

According to Wang and Gang (2018), natural language processing (NLP) is the processing of human language, is an important direction in the field of computer and artificial intelligence, and is automatic computing technology for human language analysis and representation. CNN has developed rapidly in the design and calculation of natural language processing (NLP). The principles, models, and applications of CNN in natural language processing tasks and methods in NLP task processing were also introduced in their work (Wang and Gang, 2018).

In Shamsaldin *et al.* (2019), CNNs have been used in solving NLP problems such as sentiment analysis, spam detection, or topic categorization. Even though it is less natural when it comes to processing such problems, it has accomplished a competitive outcome. In addition, CNNs have been used for the problems of speech recognition. Speech is an ethereal illustration of verbal words that includes hundreds of variables and usually encounters issues of overfitting when trained using fully connected feed-forward networks (Shamsaldin *et al.,* 2019).

CONCLUSION

There is a strong growing interest in deep learning in recent years and CNN has become the most established algorithm among deep learning models that have equally transformed into a dominant method in computer vision tasks, language processing, medical research, radiology and image recognition, reconstruction and analysis. CNNs uncover and describe hidden data in an accessible manner. The technique of recognizing images, discovery of new drugs, face recognition, medical images, etc. is astonishingly impressive. This impressive methodology will reshape the landscape in almost every applicable human field where it is applied not only by researchers but also radiologists, medical imaging analysts, and clinical radiologists as deep learning influence their practice in the future.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGEMENT

I deeply acknowledge the contributions of members of the Department of Computer Sciences of Ajayi Crowther University, Oyo, Nigeria for their contributions during group discussions and seminars on related topics in this field.

REFERENCES

Afshar, P., Plataniotis, K. N., & Mohammadi, A. (2019). Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2019. Pp. 1368-1372.

Ajay, K. (2021). Activation Functions. Retrieved from https://ajaykrish-krishnanrb.medium.com/non-linear-activation-functions-4b5e3ada8959.

Anirudha, G., Abu, S., Farhana, S., Amlan, C., & Debashis, D. (2020). Fundamental concepts of Convolutional Neural Network. In: *Recent trends and advances in artificial intelligence and internet of things*. Retrieved December 18 2022 from https://www.researchgate.net/publication/337401161_Fundamental_Concepts_of_Convolutional_Neural_Network.

Bezdan, T, & Džakula, N. B. (2019). Convolutional neural network layers and architectures. Retrieved December 18 2022 from https://www.researchgate.net/publication/333242381_Convolutional_Neural_Network_Layers_and_Architectures.

Brownlee, J. (2019). A gentle introduction to pooling layers for Convolutional Neural Networks. Retrieved 4th October 2022 from https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/.

Daniel, J. (2021). CNN image classification in tensor flow with steps & examples. Retrieved 7th December 2022 from https://www.guru99.com/convnet-tensorflow-image-classification.html.

Dhaoui, M. (2019). Convolutional Neural Networks (CNN)- Step 1b- The Rectified Linear Unit (RELU). Retrieved 9th November 2022 from https://mohameddhaoui.github.io/deeplearning/CNN_tuto11/.

flatworldsolutions (2022). 7 applications of convolutional neural networks. Retrieved 14th September 2022 from https://www.flatworldsolutions.com/data-science/articles/7-applications-of-convolutional-neural-networks.php.

Gandharv, K. (2022). Top 5 applications of Convolution Neural Network. Retrieved from https://indiaai.gov.in/article/top-5-applications-of-convolution-neural-network.

I2tutorials (2019). Explain concepts of Pooling in Convolutional Neural Network. Retrieved 24th September 2022 from https://www.i2tutorials.com/explain-concepts-of-pooling-in-convolutional-neural-network.

Kabir, H. (2021) Convolutional Neural Network. International Islamic University Chittagong. Retrieved December 17 2022 from https://www.researchgate.net/publication/348806368_Convolutional_Neural_Network.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122-1131.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen M. (2014). Medical image classification with convolutional neural network. *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. Pp. 844-848

Mayank, M. (2020). Convolutional neural networks, explained. Retrieved 13th January 2022 from https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939.

Ohri, A. (2021). Rectified Linear Unit (ReLU): An Important Introduction (2021). Retrieved from https://www.jigsawacademy.com/blogs/ai-ml/relu/

Pedamkar P. (2021). Convolutional Neural Networks. Retrieved 2nd November 2022 from https://www.educba.com/convolutional-neural-networks/

Shamsaldin, A. S., Fattah, P., Rashid, T. A., & Al-Salihi, N. K. (2019). A study of the applications of convolutional neural networks. *UKH Journal of Science and Engineering*, *3*(2), 31-39.

Sharma, P. (2021). Applications of Convolutional Neural Networks. Retrieved 20th October 2022 from https://www.analyticsvidhya.com/blog/2021/10/applications-of-convolutional-neural-networkscnn/.

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, *18*(6), 463-477.

Vaz, J. M., & Balaji, S. (2021). Convolutional neural networks (CNNs): Concepts and applications in pharmacogenomics. *Molecular Diversity*, *25*(3), 1569-1584.

Wang, S. C. (2003). Artificial neural network. In *Interdisciplinary computing in java programming* (pp. 81-100). Springer, Boston, MA.

Wang, W., & Gang, J. (2018, July). Application of convolutional neural network in natural language processing. In *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)* (pp. 64-70). IEEE.

Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, *6*(1), 1-18.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*(4), 611-629.