Deep Neural Network Concepts for Classification using Convolutional Neural Network: A Systematic Review and Evaluation

Mohammad Gouse Galety[1*], Firas Husham Al Mukthar[2], Rebaz Jamal Maaroof[3], Fanar Rofoo[4]

[1*] Department of Information Technology, Catholic University in Erbil, Erbil, Kurdistan Region, Iraq, m.galety@cue.edu.krd.

[2] Department of Information Technology, Catholic University in Erbil, Erbil, Kurdistan Region, Iraq, f.almukhtar@cue.edu.krd.

[3] Department of Information Technology, Catholic University in Erbil, Erbil, Kurdistan Region, Iraq, rebaz.maaroof@cue.edu.krd.

[4] Department of Information Technology, Catholic University in Erbil, Erbil, Kurdistan Region, Iraq, fanar.rofoo@cue.edu.krd.

Abstract. In recent years, artificial intelligence (AI) has piqued the curiosity of researchers. Convolutional Neural Networks (CNN) is a deep learning (DL) approach commonly utilized to solve problems. In standard machine learning tasks, biologically inspired computational models surpass prior types of artificial intelligence by a considerable margin. The Convolutional Neural Network (CNN) is one of the most stunning types of ANN architecture. The goal of this research is to provide information and expertise on many areas of CNN. Understanding the concepts, benefits, and limitations of CNN is critical for maximizing its potential to improve image categorization performance. This article has integrated the usage of a mathematical object called covering arrays to construct the set of ideal parameters for neural network design due to the complexity of the tuning process for the correct selection of the parameters used for this form of neural network.

Keywords: Convolutional Neural Network, Pooling, Rectified Linear Unit, Augmentation, Image Classification.

Introduction

"At the moment, one of the trendiest research disciplines is Computer Vision. It includes several academic disciplines, including Computer Science, Mathematics, Engineering, Physics, Biology, and Psychology. Since its cross-domain competence, many scientists feel that Computer Vision paves the path for Artificial General Intelligence because it represents a relative awareness of visual worlds and their contexts. The rapid involvement of image recognition systems has been substantially improved because of recent advances in neural networks and deep learning methodologies." [1] [2]

"Computer vision challenges aim to allow computers to automatically see, identify, and comprehend the visual environment in the same way that humans do. Computer vision researchers aimed to create algorithms for tasks like I object recognition, which determines whether image data contains a specific object, (ii) object detection, which locates instances of semantic objects of a given class; and (iii) scene understanding, which parses an image into meaningful segments for analysis. The challenges above in the computer vision field are exceedingly tough because of the large range of mathematics covered and the fundamentally difficult nature of recovering unknowns from insufficient information to characterize the solution adequately. Theoretically and practically, it is critical to investigate these issues. By combining well-designed features and feature descriptors with traditional machine learning methods, early efforts made a significant contribution to the philosophy of human vision and the core computational theory of computer vision. Despite decades of study into teaching robots to sight, the most advanced machine at the time could only sense common items and struggled to recognize a large variety of natural objects

with limitless shape variations, much like babies. Fortunately, experts hope that by teaching computer systems to observe trillions of photographs and videos created by the Internet, they can go beyond simple object recognition and learn to reveal subtleties and insights about the visual world. The largest image classification dataset, ImageNet, was created to feed the computer brain, containing 15 million images across 22,000 object classes, on which the well-known deep learning technology [3]has demonstrated its overwhelming superiority over traditional computer vision algorithms that treat objects as a collection of shape and color features." [4] [5]

"In 1956, John McCarthy created the term Artificial Intelligence (AI) during a symposium in Dartmouth, New Hampshire (summer research project authored by Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon). Scope of AI includes the development of Systems, Methods, Machines; which are capable of intelligent behavior like those which humans and animals exhibit with an ability to perceive, reason and act." [6] "The machine to behave like an intelligent human behavior is referred to as AI (Figure 1). Machine Learning (ML) is an area of artificial intelligence that allows computers to "learn" from data without having to be explicitly programmed. Deep Learning employs Artificial Neural Networks (ANNs), which are self-learning algorithms inspired by the structure and function of the brain (DL). ANNs are taught to "learn" models and patterns rather than being told how to solve a problem." [7]
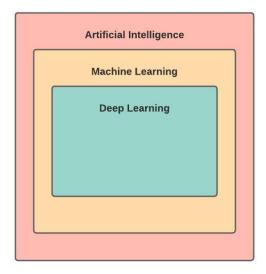

Figure 1 Architecture of AI, ML, and DL

I'll go over the five most important computer vision techniques I've come across, as well as the main deep learning models and applications for each of them, such as Image Classification, Object Detection, Object Tracking, Semantic Segmentation, and Instance Segmentation, in this work.

Image Classification
The technique of determining what an image depicts is known as image classification. The ability to discern between different sorts of images is taught to an image classification model. For example, you may teach a model to recognize photos of three other vehicles: cars, bikes, lorries, buses. Techniques of image classification are categorized as Artificial Neural Networks, Decision Trees, and Support Vector Machine. This learning object intends to introduce unsupervised and supervised image categorization algorithms. Supervised image classification is a method for recognizing spectrally comparable areas on an image by locating 'training' sites of known targets and extrapolating those spectral signatures to unknown target areas. Unsupervised image classification is the process of classifying each image in a dataset as a member of one of the intrinsic categories inherent in the image collection without the need for labelled training examples. The usage of labelled datasets is the difference between the two strategies. In other words, supervised learning algorithms make use of labelled input and output data, but unsupervised learning algorithms do not (Figure 2) [8].
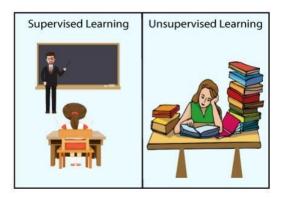
Figure 2 Differences between SL & USL

Convolutional Neural Network (CNN)
"Convolutional Neural Networks (CNNs) are a type of artificial neural networks (ANNs) that have shown to perform well on a variety of visual tasks, such as image classification, image segmentation, image retrieval, object detection, image captioning, face recognition, pose estimation, traffic sign recognition, speech processing, neural style transfer, and so on." **[9]** A Convolutional Neural Network (CNN) is a Deep Neural Network (DNN) used to analyze visual imagery in DL. A disadvantage of using ANN for image classification is too many computations, treats local pixels the same as pixels far apart, and sensitivity to the location of an object in an image. A CNN architecture comprises a series of discrete layers that use a differentiable function to turn the input volume into an output volume. Layers come in a variety of shapes and sizes. These are covered in more detail lower down (Figure 3)**[10]**.

CNN is a network that consists of an input layer, hidden layers, and an output layer. The activation function and final convolution of a feed-forward neural network hide the inputs and outputs of any middle layers. Convolutional layers are included in the hidden layers of a convolutional neural network. It's typical to utilize a layer that does a dot product of the convolution kernel and the layer's input matrix. The input to a CNN is a tensor with a form. The
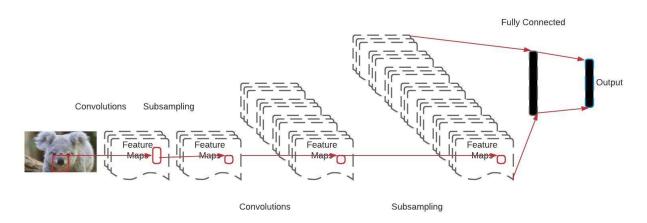

Figure 3 CNN Architecture

structure of the animal visual cortex is reflected in the connecting pattern between neurons. Biological activities influenced convolutional networks in the same way that the connecting pattern between neurons does. Individual cortical neurons respond solely to stimuli that fall inside the receptive field, a restricted section of the visual field. The receptive fields of different neurons partially overlap, allowing them to cover the whole visual field. In comparison to other image categorization algorithms, CNNs require extremely little pre-processing (Figure 4,5).
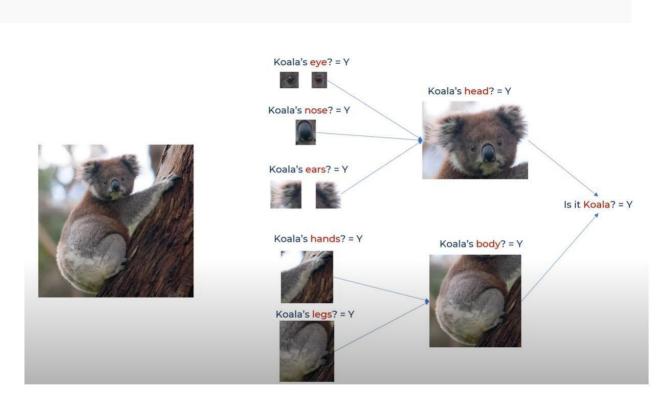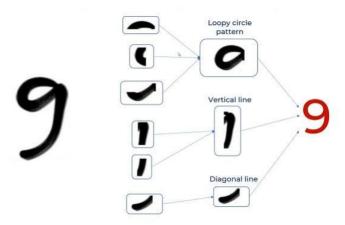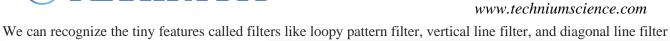
Figure 4 **Convolutional Neural Network**

Figure 5 CNN on Handwritten Numbers

The feature maps of a CNN capture the effect of applying the filters to an input image. In other words, each layer's output is the feature map. The purpose of inspecting a feature map for a specific input image is to understand better how our CNN locates features.

It means that, unlike previous methods, the network uses automated learning to enhance the filters (or kernels) (Figure 6). The fact that feature extraction does not rely on past knowledge or human interaction is a key advantage.

We can recognize the tiny features called filters like loopy pattern filter, vertical line filter, and diagonal line filter or filters are nothing but the feature detectors.
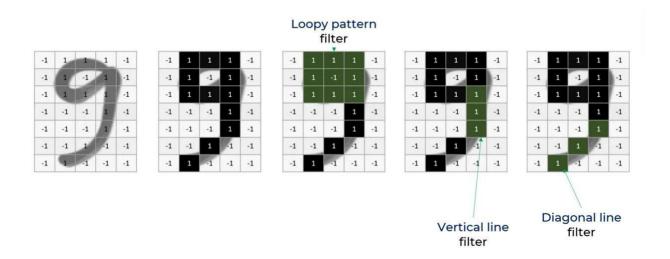

Figure 6 Concept of Filters

We will take the original image and apply the convolutional filter operation. Here we will take a 3x3 grid from the original image and multiply individual numbers with the loopy pattern filter and sum all the values and find the average (Figure 7).
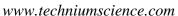

Figure 7 Original Image and Loopy Pattern Filter
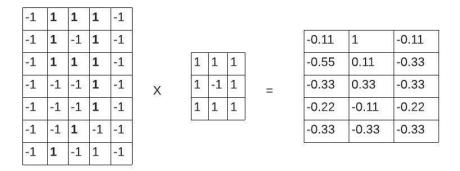
$$x = O1 * L1 + O2 * L2 + \cdots On * Ln$$

$$y = x/n$$

O is the original image, L is the Loopy pattern, and n is a number of elements in the grid.

y= -1+1+1-1-1-1-1+1+1 = -1 → -1/9=-0.11

By doing the above convolutional operation, you are creating a "feature map." Similarly, we apply the same convolutional process for the 2ⁿᵈ round of the 3x3 grid (using a 4x4 or 5x5 filter). Then you keep on doing this for whole numbers which are available in the original image grid, and at the end, we will get a "feature map" (Figure 8). In the feature map grid, wherever you find one or close to 1, It means you have a loopy circle pattern. The loopy circle will be available at the top (Figure 9).
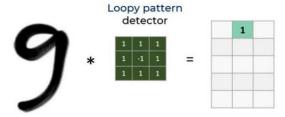

Figure 9 Loopy Circle Detection using 1

In the case of '9,' we need to apply three (3) filters. When we use those, we will get three (3) feature maps (Figure 10,11).
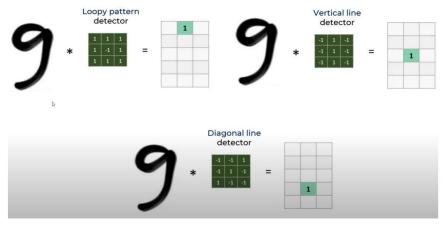

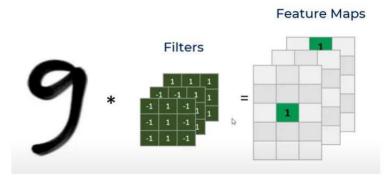Figure 10 Applying 3 filters on 9
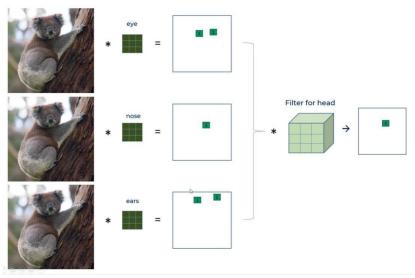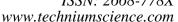
Figure 11 Three Featured Map

Figure 12 Three Features of Koala

As per the above figure 12, we are aggregating the results using the different filters for the head, and it gives the featured map of the Koala head detector. Similarly, the Koala body detector for body detection featured a map. Finally, we will flatten the one's (1) which are available in the featured maps of the head and body of the Koala, which means converting 2D array to 1D array [11]and join them together to get a fully connected dense neural network (Figure 13) for classification. In case the same Koala in the form of a different form, the neural networks are used to handle the variety in your inputs. Such that, it can generically classify that variety of inputs. In CNN, feature extraction and classification take place.
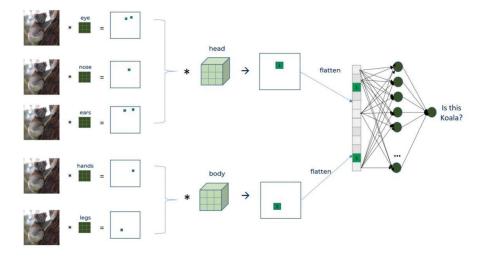
Figure 13 Fully Connected Dense Neural Network of Koala

ReLU (Rectified Linear Unit)
"In CNN, we also use ReLU; the activation function in a neural network is responsible for converting the node's summed weighted input into the node's activation or output for that input. The corrected linear activation function, or ReLU, is a piecewise linear function that outputs the input directly if the input is positive and 0 otherwise. Because a model that utilizes it is quicker to train and generally produces higher performance, it has become the default activation function for many types of neural networks" [12], or the negative values are replaced with zero (0). The values are more than zero; they will keep as it is. ReLU helps with making the model non-linear (Figure 14).

Figure 14 Applying ReLU on feaure map

Pooling
This article also shows the "Pooling" concept to reduce the size of the image. "Convolutional layers and pooling layers form a CNN. Each convolutional layer is programmed to provide representations (in the form of activation values) that reflect components of local spatial structures while accounting for a large number of channels. A convolution layer, for instance, generates "feature response maps" with several channels within a restricted geographic area. On the other hand, a pooling layer can only act in one channel simultaneously, "condensing" the activation levels in each spatially local section of the channel in question. There is an early mention of pooling procedures (albeit not explicitly using "pooling"). Modern visual recognition systems employ pooling approaches to build "downstream" representations that are more resistant to the effects of data variations while retaining major patterns. The specific selections of average pooling and max pooling are used in many CNN-like architectures; includes a theoretical analysis (although one based on assumptions that do not hold here)." [13] Pooling reduces the

dimensions and computations and reduces overfitting as fewer parameters and models tolerate variations and distortions.

Max Pooling

"Max Pooling is a convolution method in which the Kernel extracts the highest value from the area it convolves. Max Pooling tells the Convolutional Neural Network that information will only be carried forward if it is the greatest information available in terms of amplitude." [14]

You take windows of 2x2 from table 1 (Figure 15), and you pick the maximum number and put it into another 2x2 window. It is nothing but takes your feature map, applies max pooling, and generates a new feature map; that is, the new feature map is half of the original feature map, in a 2x2 filter with stride 2 (2 points forward).
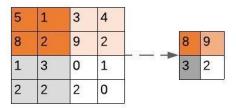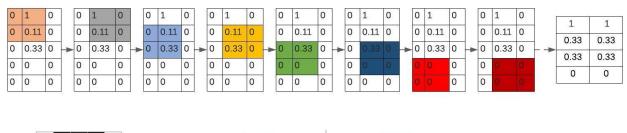

Figure 15 Max Pooling

In the "9" case, we will apply one stride and get the new feature (means 2x2 filter with one stride) (Figure 16).
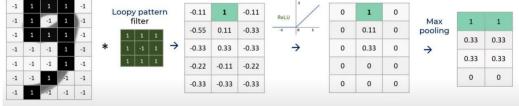

Figure 16 Max Pooling on 9

When the number "9" is shifted, you will get the below max-pooling map (Figure 17). Still, you are getting the loopy pattern at the top. Max pooling along with the convolution helps you with position invariant feature detection (Figure 17).
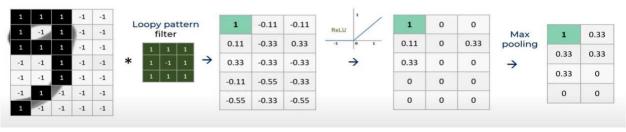
Figure 17 Max Pooling after the 9 position changes

Average Pooling
Downsampling is accomplished using an average pooling layer, which divides the input into rectangular pooling regions and computes the average values of each zone. Max pooling is more generally used.

$$y = \frac{x_1 + x_2 + x_3 + x_4}{n}$$

Y is the average value of each zone, x1, x2, x3, x4 are the values of each zone, and n is the number of values in each zone.
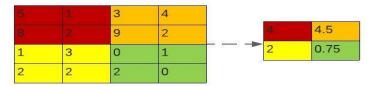

Figure 18 Average Pooling

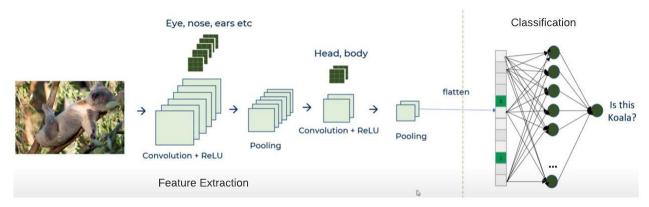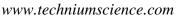The proposed convolutional neural network looks below (Figure 19).

Figure 19 Complete CNN

In this, you will typically have convolution and ReLU layer, pooling, another convolution and ReLU, n number of convolutional pooling's; at the end; there will be a fully connected dense neural network. The first convolution detects eyes, nose, ears, head, and body and then applies to flattening feature extraction. And the next one is classification; it is a simple artificial neural network. By this, we are detecting the features and reducing the dimensions.

Future Work

Although deep learning has recently made incredible strides, there are still obstacles to its implementation in the various imaging fields. Because no audit trail is left to justify its results, deep learning is considered a black box. As a result of this problem, researchers have invented a number of methods for revealing which features are identified in feature maps (feature visualization) and which input component is accountable for the corresponding prediction (feature visualization) (attribution). It's worth noting that adversarial instances have recently been discovered in deep neural networks, purposefully chosen inputs that affect the network's output to change without being obvious to a human. Even though the impact of negative occurrences in the medical field is unknown, this study demonstrates that artificial networks see and predict differently from humans. Research studying the susceptibility of deep neural networks in medical imaging is crucial compared to relatively simple non-medical tasks because the clinical application of deep learning requires extreme robustness for eventual use in patients. [15] [16]

Conclusion

Computer vision has been a challenging research subject that has gotten a lot of attention as a scientific discipline. Modern computer vision systems have been considerably modified by massive data, superior deep learning algorithms, and powerful hardware accelerators. This article investigated computer vision techniques in depth. The accomplishments of convolutional neural network techniques such as filters, pooling, and ReLU have been highlighted in particular in this article. CNN does not handle rotations or scaling on its own. We can rotate and scale using the trained dataset, and if you don't have one, you can produce fresh samples using data augmentation methods [17] [18]. In terms of algorithm research and hardware design, massive advancements in computer vision systems are predicted during the next five to ten years to solve the aforementioned significant issues. [19] [20]

Abbreviations

AI – Artificial Intelligence

ML – Machine Learning

DL – Deep Learning

ANN – Artificial Neural Network

CNN -Convolutional Neural Network

DNN - Deep Neural Network
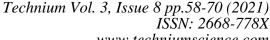
1D - One-Dimensional

2D - Two-Dimensional

3D - Three-Dimensional

ReLU- Rectified Linear Unit

References

[1] J. Le, "HeartBeat," 12 April 2018. [Online]. Available: https://heartbeat.fritz.ai/the-5-computer-vision-techniques-that-will-change-how-you-see-the-world-1ee19334354b. [Accessed 2018].

[2] F. H., F. B. A. a. K. G. Khushwant Rai, "Deep Learning for High-Impedance Fault Detection: Convolutional Autoencoders," *Energies,* vol. 14, p. 3623, 2021.

[3] e. a. Wejdan L. Alyoubi, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked,* 2020.

[4] Y. J., X. Y., M. D. X. L. Xin Feng, "Computer vision algorithms and hardware implementations: A survey," *Integration, the VLSI Journal,* vol. 69, pp. 309-320, 2019.

[5] J. P. W. S. X. G. Y. Z. X. Z. L. X. L. M. Zhixue Wang, "A Convolutional Neural Network-Based Classification and Decision-Making Model for Visible Defect Identification of High-Speed Train Images," *Journal of Sensors,* vol. 2021, no. https://doi.org/10.1155/2021/5554920, p. 17, 2021.

[6] S. L. Maskara, "Computer Vision, Artificial Intelligence and Robotics – A cursory Overview of Some of the Wonders of Modern Science and Technology," in *National Seminar on Computer Vision & Image Processing; September 8 -9, 2017*, Ahmedabad, 2017.

[7] A. GUPTA, "https://www.aiche.org/," June 2018. [Online]. Available: https://www.aiche.org/resources/publications/cep/2018/june/introduction-deep-learning-part-1?gclid=CjwKCAjwoNuGBhA8EiwAFxomA8ll-XqI13-laggRLX8fIRTXD20DUtdcx7ZxoZy2BDx9lq1KkQVBFRoCly8QAvD_BwE.

[8] Sonoo Jaiswal, "https://www.javatpoint.com/difference-between-supervised-and-unsupervised-learning," Javatpoint, 2011. [Online]. Available: https://www.javatpoint.com/difference-between-supervised-and-unsupervised-learning.

[9] N. B. D. Timea Bezdan, "CONVOLUTIONAL NEURAL NETWORK LAYERS AND ARCHITECTURES," *DATA SCIENCE & DIGITAL BROADCASTING SYSTEMS,* vol. 10, pp. 445-451, 2019.

[10] Wikipedia, Artist, *Typical CNN architecture.* [Art]. https://en.wikipedia.org/wiki/Convolutional_neural_network, 2021.

[11] A. M., T. R. J. R. R. Marcin KOŁODZIEJ, "A new method of cardiac sympathetic index estimation using a 1D-convolutional neural network," *BULLETIN OF THE POLISH ACADEMY OF SCIENCES TECHNICAL SCIENCES,* vol. 69, no. 3, 2021.

[12] J. Brownlee, "Machine Learning Mastery," 20 August 2020. [Online]. Available: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/.

[13] P. W. G. Z. T. Chen-Yu Lee, "Generalizing Pooling Functions in Convolutional Neural Networks," in *19th International Conference on Artificial Intelligence and Statistics (AISTATS) 2016*, Cadiz, Spain, 2016.

[14] ANALYTICS INDIA MAGAZINE PVT LTD, "Max Pooling in Convolutional Neural Network and Its Features," ANALYTICS INDIA MAGAZINE PVT LTD, 2020. [Online]. Available: https://analyticsindiamag.com/max-pooling-in-convolutional-neural-network-and-its-features/.

[15] R. N. M. D. R. e. a. Yamashita, "Convolutional neural networks: an overview and application in radiology. Insights Imaging," *Insights into Imaging,* vol. 9, pp. 611-629, 2018.

[16] M. G. Galety, "Data Security in Big Data using Parallel Data Generalization Algorithm," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 8, no. 1.2, pp. 75-79, 2019.

[17] T. Goyal, "COMPARATIVE STUDY & ENSEMBLE OF VARIOUS CONVOLUTIONAL NEURAL NETWORKS ON CIFAR-10," *International Journal of Scientific Research in Engineering and Management (IJSREM),* vol. 5, no. 6, 2021.

[18] G. M. M. A. Arulkumar N., "CPAODV: Classifying and Assigning 3 Level Preference to the Nodes in VANET Using AODV Based CBAODV Algorithm.," in *International Conference on Information, Communication and Computing Technology*, Istanbul City, Turkey, 2019.

[19] A. K. G. S. P. M. P. A. Sakshi Indoliaa, "Conceptual Understanding of Convolutional Neural Network-A," in *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*, Gurugram, Haryana, India, 2018.

[20] A. A. C. A. K. N. S. Galety M.G., "Improved Crypto Algorithm for High-Speed Internet of Things (IoT) Applications," in *Intelligent Computing Paradigm and Cutting-edge Technologies.*, Istanbul City, Turkey, 2019.