WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

REVIEW ON CONVOLUTIONAL NEURAL NETWORK

[1]Onwe Festus Chijioke, [2]Otuonye Anthony Ifeanyi, [3]Nwanga Mathew Emeka

[1]Open Distance and e-Learning Centre,University of Port Harcourt,Rivers State. festus.onwe@uniport.edu.ng
[2]Department of Information Technology,Federal University of Technology, Owerri. anthony.otuonye@futo.edu.ng
[3]Department of Information Technology,Federal University of Technology, Owerri. Mathew.nwanga@futo.edu.ng

INTRODUCTION:

Convolutional Neural Network (CNN), also known as ConvNet, has a deep feed-forward architecture and a remarkable capacity for generalization. Unlike other networks with FC layers, it can learn highly abstracted features of objects, particularly spatial data, and identify them more effectively (Ghosh et al, 2020). Convolutional neural networks are artificial intelligence systems based on multi-layer neural networks that can identify and categorize items, detect their presence, and segment them, according to Todt and Krinski (2019). They also learn important features from photos. According to (Sufian et al, 2019), a finite number of processing layers make up a deep CNN model, which can learn different input data features (such images) at several levels of abstraction. Higher abstraction is used by the deeper layers to learn and extract low abstraction features, whereas lower abstraction is used by the initiatory levels to learn and extract high level characteristics Convolutional neural network (or CNN) is the state-of-the-art algorithm for issues with pattern recognition, computer vision, and natural language processing. It is a unique kind of multilayer neural network or deep learning architecture modelled after the visual system of living entities and it is very suitable for fields of natural language processing and computer vision. They use a specific type of deep neural network to analyze input data that has some spatial organization (Goodfellow et al., 2016). These networks are mostly used to solve computer vision problems (e.g., self-driving cars, robotics, drones, security, medical diagnostics, blind therapy, etc.) utilizing photos as input, as it makes use of the data's grid structure. One or more blocks of convolution and pooling layers, one or more fully connected (FC) layers, and an output layer comprise a traditional convolutional neural network. According to Wang et al. (2019), convolutional neural networks are deep neural networks that optimize the utilization of the enormous amounts of data and computing power that are currently accessible in a variety of industries. CNNs eliminate the need to specifically specify which independent variables (inputs) to include or leave out of the analysis because they optimize the entire process from start to finish, mapping data samples to outputs that match the massive, labeled training sets of a deep neural network. Based on empirical data, CNNs are an effective tool for multi-layer filter image processing. Furthermore, the accuracy gains over the past few years have been so remarkable that they have altered the direction of this field's study. When the training data set is large enough, CNNs typically outperform other machine learning algorithms. Subsequent layers collect increasingly complex combinations of earlier information, while the initial filters capture low-level picture features (e.g., recognizing horizontal or vertical borders, bright points, or color variations) (Wang et al. 2019). The ability of CNNs to accurately classify new examples with a limited number of training examples and to learn the best features to represent the objects in the images gives them an advantage over other classification algorithms (SVM, K-NN, Random-Forest, etc.) (Todt and Krinski, 2019).

CONVOLUTIONAL NEURAL NETWORK LAYERS

i. Convolutional: A convolutional layer consists of filters (kernels) that slide over input data. Kernels use width, height, and width-by-height weights to extract features from input data. The kernel learns from the training data and starts with random weights during training. One fundamental element of a convolutional neural network is the convolution operation. Learnable filters, or kernels, make up the parameters of the convolutional layer. Filters are modest in size yet cover the entire input volume. The convolution operation is a basic component of a convolutional neural network. The parameters of the convolutional layer are learnable filters, or kernels. The size of each filter is tiny. (along height and width), but it encompasses the input volume's whole depth (Bezdan and Dzakula, 2019).

ii. Pooling: The pooling layer, also known as down-sampling, reduces the dimensionality of feature maps, saving only essential information. The pooling layer filters input data and performs pooling operations (max, min, avg). The max pooling algorithm is the most commonly utilized in the literature.

iii. Relu: The ReLU (Rectified Linear Units) layer, coupled after a convolutional layer, creates non-linearity in the network. ReLu enables the network to learn complex decision functions and reduces overfitting. ReLu uses the function $y = \max(x, 0)$.

iv. Fully Connected: CNNs are divided into two parts: convolutional and dense steps. The first learns which traits to extract from photos, while the second classifies them into categories. The CNN-generated features are fed into the input layer. The buried layer consists of neurons with weights that are learned during training. A MLP consists of one or more hidden layers. The output layer is also composed of neurons. However, the activation function is different. Every category in the issue scope has a probability generated by the softmax function. The CNN usually ends with a number of fully connected layers following various convolution and pooling levels. After these layers convert the output tensor into a vector, numerous neural network layers are added. The final few layers of the architecture are often the ones that are fully connected.

OVERVIEW ON CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

i. LeNet -5: This is a convolutional Neural Network Architecture introduced by (LeCun et al, 1998), for the purpose of classifying handwritten digits. It is a simple convolutional network, and are a type of feed-forward neural network that excels in large-scale image processing because their artificial neurons can react to a portion of the surrounding cells within the coverage range. It featured a number of significant advances and features that are now conventional in modern deep learning. The efficiency of CNNs for image identification tasks was illustrated, and important ideas like convolution, pooling, and hierarchical feature extraction—which form the basis for current deep learning models were presented (Geeksforgeeks, 2024). It is clear that the overall goal of the setup is to do various convolutions with maximum pooling between two workouts and interface the yield layer to the remaining convolutional layer through fully connected layers.

ii. AlexNet: On the ImageNet classification experiment, this convolutional neural network design performed admirably. Approximately 15 million high-resolution, labeled photographs across roughly 22,000 categories make up the visual dataset known as ImageNet. Scholars are interested in testing their picture classification algorithms on the ImageNet Dataset because of its high quality. The structure comprises eight weighted layers that are learnable: the first five levels are convolutional layers, while the latter three layers are fully connected layers. When it comes to object detection and image classification, AlexNet has shown to be the best design (Singh et al, 2022). With more than 60 million parameters, deep CNNs like AlexNet provide a strong and efficient deep learning solution for issues involving a lot of datasets. However, these models have a tendency to overfit when dealing with sparse datasets, like the Google Jigsaw Dataset, which leads to subpar model results (Alex et al, 2012). Alzubaidi et al, (2021) in his research stated that AlexNet played a major role in redirecting the focus of deep learning research. Its architecture is comparable to LeNet-5's, but it has more layers and filters, which improves the extent and increases the number of learning variables. The work showed that deep learning frameworks with deep domain awareness could be used to learn features instead of producing them by hand.

iii. VGGNet: In 2014, Simonyan and Zisserman announced VGGNet, one of the most well-known convolutional neural network architectures (Simonyan and Zisserman, 2014). VGG, or Visual Geometry Group, is a common deep Convolutional Neural Network (CNN) architecture that consists of many layers. When referring to VGG-16 or VGG-19, which consist of 16 and 19 convolutional layers, "deep" means the number of layers. Boesch (2021). The VGG architecture's main goal is to use uniform, simple convolutional layers to increase the network's depth (Grigoryan, 2023). Within the VGG architecture, a modular unit called a VGG block contains a number of convolutional and pooling layers. Together, these layers capture elements with different levels of complexity, ranging from simple forms and edges to intricate textures and patterns. VGG realized that by building up these pieces, the network may be trained to identify high-level characteristics that characterize an object's identity in an image. It has various architectures, including VGG11, VGG13, VGG16, and VGG19, where the numbers represent the total layers in each architecture. The efficacy of deeper structures for image classification is demonstrated by the VGG architecture, which outperform the current state-of-the-art models on a number of benchmark datasets, including the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) dataset (Boesch, 2021). VGG was able to classify a variety of objects within photos more correctly than previous models. Its higher accuracy is attributed to the systematic layering of convolutional blocks, which made it possible to capture complex information. Nevertheless, VGGs are computationally demanding and their scalability is limited by the need for processing resources, which makes their deployment impractical on devices with limited resources.

iv. GoogLeNet: Also known and called Inception-VI was developed to achieve high level accuracy with decreased computational cost. It was proposed by (Szegedy et al, 2015) in their research. The GoogleNet concepts of divide, merge, and transform were applied, with the assistance of addressing a problem associated with various learning types of variants included in a class of multiple photos that was similar. The goal of GoogLeNet was to increase both the learning capacity and the effectiveness of CNN parameters. Furthermore, a 1×1 convolutional filter is inserted as a bottleneck layer to control the computation before large-size kernels are used (Alzubaidi et al, 2021).

v. DenseNet: This CNN otherwise known as Densely Connected Convolutional Networks architecture was introduced by Huang and his team. The DenseNet takes the same output feature map and concatenates it with all previous input feature maps, in contrast to a typical CNN, which only utilizes the current output feature map as the input for the following layer (Aramendia, 2024). It extends on the concept of residual mapping by propagating each block's output to all blocks inside each dense block in the network. The vanishing gradient problem is resolved and feature propagation ability is strengthened by propagating the data both forward and backward during the model's training (Huang et al, 2016). According to (Huang et al , 2020) possibly counter-intuitive effect of this dense connectivity pattern is that it requires fewer parameters than traditional convolutional networks, as there is no need to relearn redundant feature-maps. Traditional feed-forward architectures can be viewed as algorithms with a state, which is passed on from layer to layer. Each layer reads the state from its preceding layer and writes to the subsequent layer. It changes the state but also passes on information that needs to be preserved.

vi. ResNet: A feature of the artificial neural network ResNet is known as the "identity shortcut connection," which enables the model to bypass one or more layers. With this method, the network can be trained on thousands of layers without seeing any degradation in performance. For a variety of computer vision tasks, it has emerged as one of the most widely used architectures (Feng, 2022). Deep convolution neural network has a series of major breakthroughs in image classification (Chollet, 2017). But when deep learning advances and we begin to think about the convergence of deeper networks, we encounter a degradation issue. This means that as a neural network gets deeper, accuracy increases initially before reaching saturation. The accuracy will drop as the depth increases. It is established that overfitting is not the cause of the influence when the error rises in both the training and test sets. Through cross-layer feature fusion, RESNET improves its capacity to extract network features, and as the network becomes deeper, network performance progressively gets better. The study team analyzed other deep learning models and tested the deeper RESNET in a reasonable amount of time. The results showed that RESNET performs better for classification than other models and that it can increase accuracy by going deeper. The Kaiming team at Microsoft Research Institute proposed RESNET (residual neural network), also known as residual network in Chinese. The primary goal of RESNET design is to address the neural network degradation issue, which is that training error rates increase with deeper neural networks (Zhang et al., 2018). To solve this problem, the team proposed a residual structure. The function of network layer is reprogrammed as residual function of input of each layer. In mathematical statistics, the concept of residual is the difference between the actual observation value and the estimated value (fitting value).

APPLICATION AREAS OF CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks have been applied in some major areas to achieve state of-the-art performance and they include image classification, human pose estimation, action recognition, text recognition, object detection, image captioning and others (Gosh et al, 2020)

CONVOLUTIONAL NEURAL NETWORKS IN IMAGE CLASSIFICATION:

The rapid expansion of digital content in recent times has made automatic picture classification one of the most important problems for visual information indexing and retrieval systems. (Ramprasath and Hariharan, 2022) in their research noted that for decades, image classification has been a major issue in computer vision. Understanding and classifying images is a relatively simple task for humans, but it is a tremendously costly task for computers. Generally speaking, every image is made up of a collection of pixels, and each pixel has a unique value. The computer will now require additional storage space in order to store an image. It needs to make more calculations in order to classify photos. Systems with greater configuration and processing power are needed for this. Making judgments in real time based solely on input is not feasible due to the lengthy processing times involved in completing the numerous calculations needed to produce a conclusion. In computer vision, non-textual forms of information, including images, are typically used to represent data (Pranav *et al*, 2017). One of the core issues in computer vision is image classification, which is the process of classifying images into one of several predetermined classes. Other computer vision tasks including segmentation, detection, and localization are based on it (Kabassi, 2016). The picture databases grew as a result of the enormous quantity of photos. Compared to other approaches, CNN has better classification accuracy because of a number of features, such as weight sharing and many feature extraction levels, such classifiers. This is especially true for large datasets. There are several layers in CNN, and they are organized hierarchically. Every layer picks up unique aspects of the picture. The convolutional neural network structure is an advancement over the conventional artificial neural network and consists of convolutional layers, pooling layers, dropout layers, and an output layer. Abdelhafiz et al, (2019) asserts that although the form and purpose of the layers have changed, the convolutional neural network is still a hierarchical network, just like the artificial neural network (ANN). It is separated into two sections: fully linked layers for classification and convolution and pooling layers for feature extraction. Prior to being sent through fully connected layers for classification, the image is initially sent through a sequence of convolution and pooling layers for feature extraction. CNN is a useful tool for classifying photos, performing object recognition inside scenes, clustering images based on similarity (photo search), and identifying faces, individuals, street signs, cancers, platypuses, and a variety of other visual data features (Hossain and Sajib, 2018) . Since then, researchers have made a number of improvements to the CNN model, which has made CNN the go-to option for image classification problems. Chen *et al* (2016) focused on the Convolutional Neural Network (CNN) deep learning approach for feature extraction from Hyper Spectral Images (HSI). It makes use of the different pooling layers in CNN to extract the feature (nonlinear, invariant) from the HIS that is required for accurate image classification and target identification. The general problems with the features of the HSI photos are also addressed. From an engineering standpoint, it looks to automate processes that the visual system of a human can perform. It deals with the automatic extraction, processing, and comprehension of meaningful information from photographs. Krizhevsky *et al* (2012) in their research studied, deep convolutional neural networks were utilized to classify high-resolution photos from the 15 million tagged images in the ImageNet data set. Three layers make up CNN: the input layer, the hidden layers, and the output layer. Images are often built as matrices of pixels, and the pixel values, together with weights and biases (for non-linearity), are supplied as input to the input layer. Typically, the output layer is a fully linked layer that is used to classify the image according to its class. The hidden layer could be completely connected, pooling, or convolutional. A research by (Dong and Izquierdo, 2024) talked about utilizing a biologically stimulated model to classify natural photos.

It makes use of established similar advancements in the visual information system and the process of inferring how the human brain functions. The main applications of this paradigm are natural categorization and image analysis. The knowledge structuring unit, the clustering of visual information unit, and the biologically inspired visual selective attention unit are the three main parts of this system. Using the low-level information found in the images, it automatically extracts meaningful correlations between them. The technology imitates the limits of the human visual system in order to achieve higher classification accuracy.

CONVOLUTIONAL NEURAL NETWORKS IN TEXT RECOGNITION

Textual information extraction from natural photographs is a difficult subject with a wide range of useful applications. Contrary to character identification for scanned documents, a variety of differences in backgrounds, textures, typefaces, and lighting conditions make it difficult to recognize text in unconstrained photographs (Mishra *et al*, 2012). Text identification and detection inside images have long been the subject of extensive research. Text on an image is frequently a significant source of information and transmits high level semantics directly; as a result, it has gained a lot of attention from researchers. Numerous research have demonstrated the effectiveness and accuracy of CNN-based neural networks for text recognition picture categorization. Deep learning-based technologies are being used to train Artificial Neural Networks (ANNs) to extract information from images. Several convolution layers are used in CNN architectures such as VGG, ResNet, MobileNet, GoogleNet, Xception, and DenseNet in order to extract features from the images (Li et al, 2018). Reading handwritten texts is challenging, as is identifying text from photos or something that has received a lot of attention. The two main parts of most systems are text recognition (ii) and text detection (i). Using a technique called text detection, text instances from the photos can be localized and forecasted. The autoencoder performs text recognition by decoding the text into a machine-readable format. With LeNet-5, CNN made its first noteworthy contribution to this field by correctly identifying the data in the MNIST dataset. According to Wang et al. (2012), CNN has made significant advancements in recent years and is now able to identify text in images, including letters, numbers, and symbols from many languages. Coates *et al* (2011) used a straightforward and scalable feature learning architecture that incorporates very little hand-engineering and prior knowledge to obtain competitive results in both text detection and character recognition. CNN was used by (Pranav *et al*, 2017*)*

to extract characteristics from Malayalam characters. According to (Acharya *et al*, 2015), text may be detected and classified using deep convolutional neural networks and image datasets. Test accuracy can be increased by using the dataset increment strategy and dropout technique.

CONVOLUTIONAL NEURAL NETWORKS IN HUMAN ACTION RECOGNITION

The foundation of feature extraction is the convolution operation, which is the basis of CNN, a form of artificial neural network. 1D, 2D, and 3D convolution are examples of convolution techniques. Among these, feature extraction in action recognition can frequently be done using 2D and 3D convolution. The automatic analysis and detection of one or more targets' motions in unknown video using deep learning techniques is known as action recognition based on deep learning techniques. Numerous scholars have been engaged in computer vision research, particularly in the areas of robot navigation, surveillance systems, human activity detection to help hospital patients and physically challenged individuals receive medical care, and many other areas. Poulose et al (2022) presented a HAR approach that utilizes the deep learning models in the human image threshing machine. A face image threshing machine was used to crop and resize the photographs; a DL model is utilized for image classification; and a region-based CNN model is used for HAR. For HAR, this technology has an extremely high accuracy rate. A unique approach to multimodal gesture identification using Deep Dynamic Neural Networks was presented by Luwe et al. in 2022. A semi-supervised hierarchical dynamic framework based on a Hidden Markov Model is developed for the simultaneous segmentation and gesture detection. The observations of multimodal input consist of RGB (red, green, and blue) pictures, depth, and skeletal joint data. This approach mainly used deep neural networks to create high-level spatiotemporal representations tailored to the modality of input data. The Gaussian-Bernoulli Deep Belief Network in this case handles skeletal dynamics, whereas a three-dimensional CNN handles and fuses the sets of RGB and depth picturesA method for the automatic extraction of discriminative characteristics for various HARs is presented by Zeng et al, (2014). This method is based on CNN, which has been demonstrated in the fields of speech and picture recognition to obtain scale invariance and local dependency of a signal. In order to attain even greater improvement, partial weight sharing is also suggested and provided in relation to the accelerometer's readings. This methodology's accuracy rate is higher than that of the most advanced method. An algorithm for simultaneous face detection, gender recognition, pose estimation, and landmark localization using deep CNN was proposed by (Ranjan, et al, 2019). Known as HyperFace, this algorithm is described. An algorithm of multi-task learning operates on the feature of the fused layer after the deep CNN intermediate layer is fused using a separate CNN. This method has important implications for the HAR field. Ding et al, (2022) developed HAR, a system for individualized nighttime monitoring around parked aircraft, based on thermal infrared vision. The suggested algorithm flawlessly combines all of the functions of this work, including the preprocessing module, which creates a new data structure to introduce information about human actions; the spatial feature extraction module, which uses CNN; the temporal feature extraction module, which uses a triple layer convolutional LSTM (Long Short-Term Memory) network; and the classification module, which uses two fully connected layers. The research project's output yields a higher than 96% recognition rate. Gupta and associates. A HAR approach combining several gyroscopes and accelerometer sensors was introduced by (Ha and Choi, 2016) using CNN. This study builds CNN-based models that can learn from multi-sensor data by utilizing both partial and complete weight sharing. The attributes unique to each modality and common characteristics among the approaches that make up the multi-sensor data in this study are collected in the higher layers. Additionally, the CNN-model's common properties are trained through modalities using the multi-modal data. Comparing this suggested CNN model to the conventional HAR system technique, it performs higher and better. Utilizing data from wearable sensors, (Gupta, 2021) presents a HAR model based on the DL algorithm. A novel hybrid deep neural network model was proposed by this research project. For HAR, the Gated Recurrent Unit (GRU) and CNN are coupled with the hybrid model. With greater accuracy than previous deep neural network-based HAR models, the suggested model has been successfully trained and validated. It is evident that the CNN-based HAR approach significantly and impressively affects activity recognition. Therefore, in order to improve the HAR's performance, accuracy, and efficiency, more CNN-based HAR techniques are needed. The creation of synthetic data may be one of the HAR research trends of the future. A large dataset must be used for model training in the majority of data mining methodologies and approaches in order to effectively learn the HAR model. A class imbalance and intra-class variability may be present in the data due to the fact that certain activities take place over longer periods of time than others. Currently, a number of efficient CNN base techniques may accurately predict the action or behavior of human subjects based on the visual appearance and motion dynamics of any human body. In terms of AI, this takes CNN to the next level. Action recognition from still photos or from a video sequence is part of it.

SUMMARY

In this research, we reviewed the importance of convolutional neural network in computing. Convolutional neural networks (CCNs) is a machine learning model, and a special type of deep learning algorithm that is designed for task that requires object recognition, image classification, detection and segmentation. A generalization of the neural cognitive machine is the convolutional neural network. Convolutional neural networks are multilayer networks used for training that are made up of several single-layer convolutional neural networks Nonlinear transformation, down sampling, and convolution are all included in a single-layer convolution neural network. Every layer has a feature map as its input and output made up of a collection of vectors (the first layer's initial input signal can be thought of as a high-dimensional feature map with high sparsity). Large-scale picture feature representation and categorization have proven successful for convolutional neural networks. As CNNs are made up of neurons with learnable weights and bias constants, they resemble artificial neural networks. Additionally, CNNs are feedforward artificial neural networks that enable the introduction of weight-sharing methods and the encoding of particular qualities into the network structure, increasing the efficiency of the feed forward function by lowering the number of parameters.

REFERENCES:

1. Acharya, M. S., Armaan, A., & Antony, A. S. (2015). "A comparison of regression models for prediction of graduate admissions," in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) pp. 1-5.

2. Alex, K., Ilya, S., & Geoffrey, E. H., (2012). ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - (NIPS'12), 1, 1097–1105. 10.1145/3065386.
3. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A, Al-Amidie, M., Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data volume 8, Article number: 53.
4. Aramendia, A. I. (2024). DenseNet: A Complete Guide: Extending the ResNet to improve performance. Available online and retrieved on 17/03/2024
5. Bezdan, T. & Džakula, N.B. (2019). Convolutional Neural Network Layers and Architectures,. Data Science & Digital Broadcasting Systems. Conference Paper
6. Boesch, G. (2021). Very Deep Convolutional Networks (VGG) Essential Guide, available online at https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks. Retrieved on 15th June, 2024.
7. Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2016). Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sensing 2021, Vol. 13, Page 4712, 13(22), 4712. https://doi.org/10.3390/RS13224712
8. Chollet, F., (2017) Deep learning with depthwise separable convolutions. In Conferenceon Computer Vision and Pattern Recognition, 1800-1807.
9. Coates, A. Carpenter, B. Case, C. Satheesh, S. Suresh, B. Wang, T. Wu, D. J. & Ng, A. Y. (2011). Text detection and character recognition in scene images with unsupervised feature learning. In ICDAR.
10. Ding, M., Ding, Y., Wei, L., Xu, Y. & Cao, Y.(2022) Individual surveillance around parked aircraft at nighttime: Thermal infrared vision-based human action recognition. IEEE Trans. Syst. Man Cybern. Syst. https://doi.org/10.1109/TSMC.2022.3192017.
11. Feng, Z. (2022). An Overview of ResNet Architecture and Its Variants Available online at https://builtin.com/artificial-intelligence/resnet-architecture and retrieved on 15/07/2024
12. Ghosh, A. Sufian, A. Naskar, F. & Sultana. B. (2020). Handwritten numeral digit recognition based on densely connected convolutional neural networks. CoRR, abs/1906.03786.
13. Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MAMIT Press, Cambridge
14. Grigoryan, A.A. (2023). Understanding VGG Neural Networks: Architecture and Implementation avalaible online at https://thegrigorian.medium.com/understanding-vgg-neural-networks-architecture-and-implementation.
15. Ha, S. and Choi, S. (2016). Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 381–388. https://doi.org/10.1109/IJCNN.2016.7727224 .
16. Hossin, M., & Sulaiman. (2015). (). International Journal of Data Mining & Knowledge Management Process (IJDKP), 5(2). https://doi.org/10.5121/ijdkp.2015.5201
17. Huang, G., Liu, G., K. Q. & Weinberger, K.Q (2016). Densely connected convolutional networks. CoRR, abs/1608.06993.
18. Kabassi, K., Dragonas, L., Ntouzevits, A., Pomonis, T., Papastathopoulos, G. (2016). Evaluating a learning management system for blended learning in Greek higher education. Springerplus, 5, 101.
19. Krizhevsky, A., Sutskever, I.,.Hinton, G.E., (2012). ImageNet Classification with Deep Convolutional Neural Network. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
20. Li,G., Liu, F., Wang,Y.G.,Xiao,L., Zhu, L., (2018). A Convolutional Neural Network (CNN) Based Approach for the Recognition and Evaluation of Classroom Teaching Behavior. Scientific Programming Volume 2021, Article ID 6336773, 8 pages
21. Luwe, Y. J., Lee, C. P. & Lim, K. M. (2022). Wearable sensor-based human activity recognition with hybrid deep learning model. Informatics 9, 56. https://doi.org/10.3390/informatics9030056
22. Mishra, A. Alahari, K. & Jawahar, C. V. (2012) Top-down and bottom-up cues for scene text recognition.
23. Poulose, A., Kim, J. H. & Han, D. S. (2022). Human image threshing machine for humanactivity recognition using deep learning models. Computat. Intell. Neurosci. https://doi.org/10.1155/2022/1808990.
24. Pranav, P. N., Ajay, J. Saravanan, C.(2017).Malayalam Handwritten Character Recognition Using Convolutional Neural Networks‖ , International Conference on inventive Communication and Computational Technologies.
25. Pranav, P. N., Ajay, J. Saravanan, C.(2017).Malayalam Handwritten Character Recognition Using Convolutional Neural Networks‖ , International Conference on inventive Communication and Computational Technologies.
26. Ramprasath, M.M. and ,Hariharan,S. (2022). Image Classification using Convolutional Neural Networks.International Journal of Pure and Applied Mathematics. Volume 119 No. 17, 1307-1319
27. Ranjan, R., Patel, V. M. & Chellappa, R. (2019). HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern Anal. Mach 41.(1), 121-135.
28. Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556.
29. Singh, I., Goyal, G., & Chandel, A. (2022). AlexNet architecture based convolutional neural network for toxic comments classification. Journal of King Saud University - Computer and Information Sciences. Volume 34, Issue 9, Pages 7547-7558.
30. Sufian, A., Ghosh, A.. Naskar, F. & Sultana. B. (2020). Handwritten numeral digit recognition based on densely connected convolutional neural networks. CoRR, abs/1906.03786.
31. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In The IEEE Conference on Computer Vision.
32. Todt, E., & Krinski, B.A. (2019). Convolutional neural networks-CNN. VRI Group - Vision Robotic and Images Federal University of Parana.

33. Wang, X. Xuan, H., Evers, B., Shrestha, S., Pless, R., Poland, J. (2019) High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat. GigaScience 8:1–11
34. Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P. & Zhang, J. (2014). Convolutional Neural Networks for human activity recognition using mobile sensors. 6th International Conference on Mobile Computing, Applications and Services, 197–205. https://doi.org/10.4108/icst.mobicase.2014.257786
35. Zhang, X., Zhou, X., & Lin, M., (2012). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Conference on Computer Vision and Pattern Recognition, 6848-6856.
36. Zhang, X., Zhou, X., & Lin, M., (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Conference on Computer Vision and Pattern Recognition, 6848-6856.