ISSN 2714-7533

Impact the Classes' Number on the Convolutional Neural Networks Performance for Image Classification

Amna Kadhim Ali [a,1,*], Abdulhussein Mohsin Abdullah [b,2], Sabreen Fawzi Raheem [c,3]

[a] Veterinary Public Health Branch, College of Veterinary Medicine, University of Basrah, Basrah, Iraq
[b] Computer Technology Engineering Department, Alkunooze University College, Basrah, Iraq.
[c] Basrah technical Institute, Southern Technical University, Basrah, Iraq.

[1] amna.kadhim@uobasrah.edu.iq, [2] abdulhussein.mohsin@kunoozu.edu.iq, [3] sabreen.fawzi@stu.edu.iq
* corresponding author

ARTICLE INFO

Article history
Received June 30, 2023
Revised July 10, 2023
Accepted July 20, 2023

Keywords
CNN
Deep learning
Facial expression
Image classification

ABSTRACT

Deep learning was developed as a realistic artificial intelligence technique that takes in numerous layers of information and produces the best results in various classes. Deep learning has demonstrated excellent performance in several areas, particularly picture grouping, division, and recognition. The convolutional neural network (CNN) is one of the algorithms that relies on deep learning in its work. It has proven its effectiveness in classifying images with high efficiency in medical images and their diagnoses, face recognition, and other different fields. In this paper, the focus was on images to alert new researchers to their effects on the performance of CNN in terms of the number of classes that existed within the database, in addition to the impact of incorrect classification of images by the source on the classification result and the necessity of adopting reliable and correct sources of data to avoid inaccurate results. A group of face images has been used, and three experiments on them were conducted using all existing classes with reduction. The results showed a significant improvement in the performance of the algorithm whenever the number of classes was reduced. The best result was when only two classes were chosen for classification, reaching a validation accuracy of 85%.

1. Introduction

Since the 1950s, a small portion of artificial intelligence, known as machine learning, has revolutionized various areas. A neural network is a subfield of machine learning, and it was from this topic that Deep Learning (DL) began [1]. Since its beginnings, DL has caused ever-increasing disruptions, demonstrating exceptional performance in practically every application sector [2].

Convolutional Neural Network (CNN) is one of the DL algorithms that typically analyzes visual input like images and movies. CNNs have transformed the area of computer vision by achieving cutting-edge performance in a variety of applications [3].

The key idea behind CNN is to exploit the spatial structure of the input data. Traditional neural networks process the entire input data as a flat vector, ignoring any spatial relationships. CNN, on the other hand, takes advantage of the local correlations present in the data by using a special type of layer called a convolutional layer [4].

It is widely used in various real-life problem-solving scenarios, particularly in the field of computer vision to solve real-world problems like image classification [5], where CNN excels at image classification tasks because it can accurately identify and categorize objects within images. It has been used in applications such as autonomous driving (to detect pedestrians, traffic signs, and other vehicles) [6], medical imaging (to diagnose diseases from scans) [7], [8], [9] and facial recognition systems [10].

Also, it can perform pixel-level segmentation for semantic segmentation, where each pixel in an image is classified into specific classes or categories. This technique is useful in medical imaging (for tumor segmentation) [11], autonomous vehicles (to identify drivable areas and obstacles) [12], and image editing tools (for background removal and image manipulation) [13].

While CNN is primarily used in computer vision, it can also be applied to Natural Language Processing (NLP) tasks such as text classification and sentiment analysis, especially for tasks involving text that has a grid-like structure, such as character recognition in handwriting or document analysis [14], [15], in addition to other fields.

The CNN algorithm is applied to databases that contain a different number of classes. In research [16], the researcher employed deep learning technology to identify melanoma on a three class dataset, Melanoma is a form of skin malignancy. The suggested model differentiates between benign lesions, superficial spread, and nodular melanoma. This enables early viral detection and the prompt isolation and treatment required to prevent future spread. Deep learning (DL) and the non-standard machine learning technique are expressed in the neural network algorithms' deep layer structure of the CNN, and the findings demonstrated the usefulness of the performance of the CNN algorithm with 88% accuracy when compared to other algorithms utilized.

In paper [17], a database of brain tumors consisting of four classes was used, namely (GLIOMA, MENINGIOMA, NO-TUMOR and PILUITARY).A three-step preprocessing method is presented, as well as a novel Deep Convolutional Neural Network (DCNN) architecture for glioma, meningioma, and pituitary tumor identification. The approach employs (batch normalization) to allow for faster training with a greater learning rate and to simplify layer weight initialization. A few convolutional layers, max-pooling layers, and training iterations are included in the suggested design. The proposed designs were compared to the other models discussed in this paper. The overall competitive accuracy is 98.22% when evaluated on a dataset of (3394) MRI pictures, with 99% recognizing glioma, 99.13% detecting meningioma, 97.3% detecting pituitary, and 97.14% detecting normal images.

In research [18] , the researchers used two facial expression databases and described a unique technique based on hierarchical DL, with the first base (CK+) consisting of seven expressions and six emotions (anger, disgust, fear, happiness, sorrow, and surprise) being used as experiment data. Furthermore, the (JAFFE) collection includes gray-scale frontal facial expression images of ten women, each with a distinct face emotion (anger, disgust, fear, joy, sadness, surprise, and neutral). The feature produced from the appearance feature-based network is fused with the geometric feature in a hierarchical structure. The proposed method was compared to other current algorithms for the 2 datasets, and the ten-fold-cross validation results show that the CK+ dataset is 96.46% accurate.

In study [19] researchers created a computationally efficient and scalable deep learning model utilizing CNN for autonomously identifying diabetic retinopathy (DR). It is a diabetic eye complication that causes impaired vision or blindness. For autonomously diagnosing DR, the researcher employed a computationally efficient and scalable deep learning model based on CNN. To boost accuracy, several preprocessing methods are used, and a transfer learning strategy is used to speed up the process. The investigation made use of the online fundus picture collection. Kaggle datasets are divided into five categories (none, mild, moderate, severe, or proliferative). The computer simulation generated a comparatively high F1 score of 93.2% for stage-based DR categorization as the final conclusion of relevant performance criteria.

In our paper, the effect of the number of classes in the database on the classification accuracy of CNN algorithm was studied. So a database that contains a large number of classes, which are specific to facial expressions, was chosen, and the algorithm was applied to all classes, which are 8, then reduced the number of classes to 4, then 2, and the results showed a clear difference in the performance of the algorithm.

2. Research Method

The proposed method includes obtaining a database of images containing a large number of classes to achieve the aim of this research, the second step is processing these images for ease of dealing with them, and then classifying them using the CNN algorithm. The working method is shown in (Fig. 1), and the steps below explain the mechanism in detail.

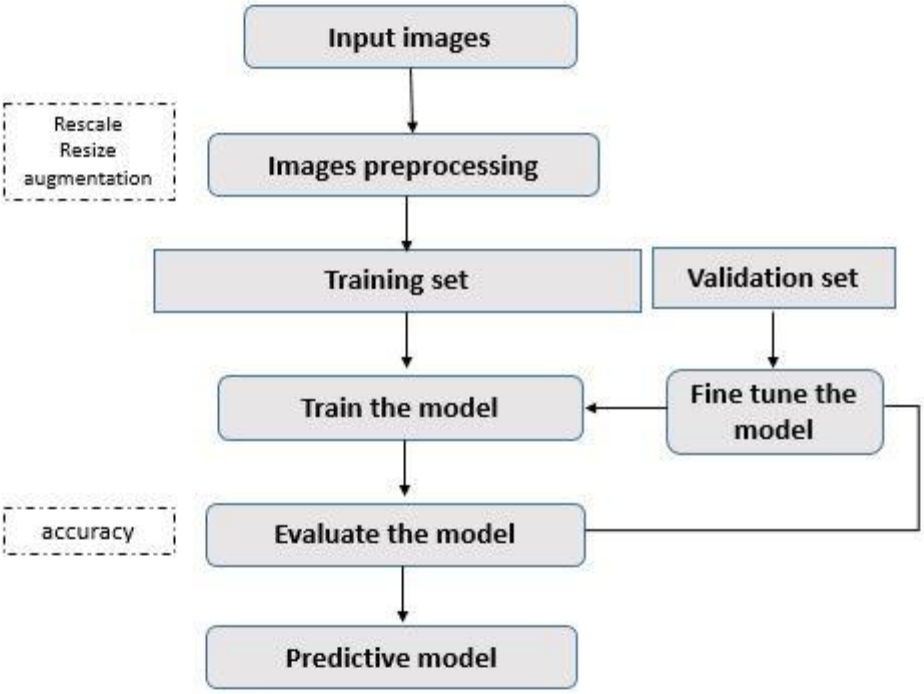

Figure 1. Flowchart of the proposed work

2.1 Dataset

The dataset obtained from AffectNet [20], The AffectNet collection comprises images gathered from the Internet by searching three search engines with 1250 emotion-related keywords in 6 languages. The existence of 8 distinct facial expressions (categorical model) and the strength of valence and arousal (dimensional model) are manually annotated in about half of the recovered pictures. The remaining images are automatically labeled with an average accuracy of 65% using a ResNext Neural Network trained on all hand annotated training set samples.

The data consists of 8 labels, they are 1: Neutral, 2: Happy, 3: Sad, 4: Surprise, 5: Fear, 6: Disgust, 7: Anger, 8: Contempt. And the total of these pictures is approximately 31,000 divided into categories according to the following (Table 1), and (Fig. 2) shows a sample of the categories.

Table 1. Categories of face expressions

Facial Expression	**Number of images**
Neutral	5132
Happy	5043
Sad	3430
Surprise	4296
Fear	3622
Disgust	2660
Anger	3638
Contempt	3179

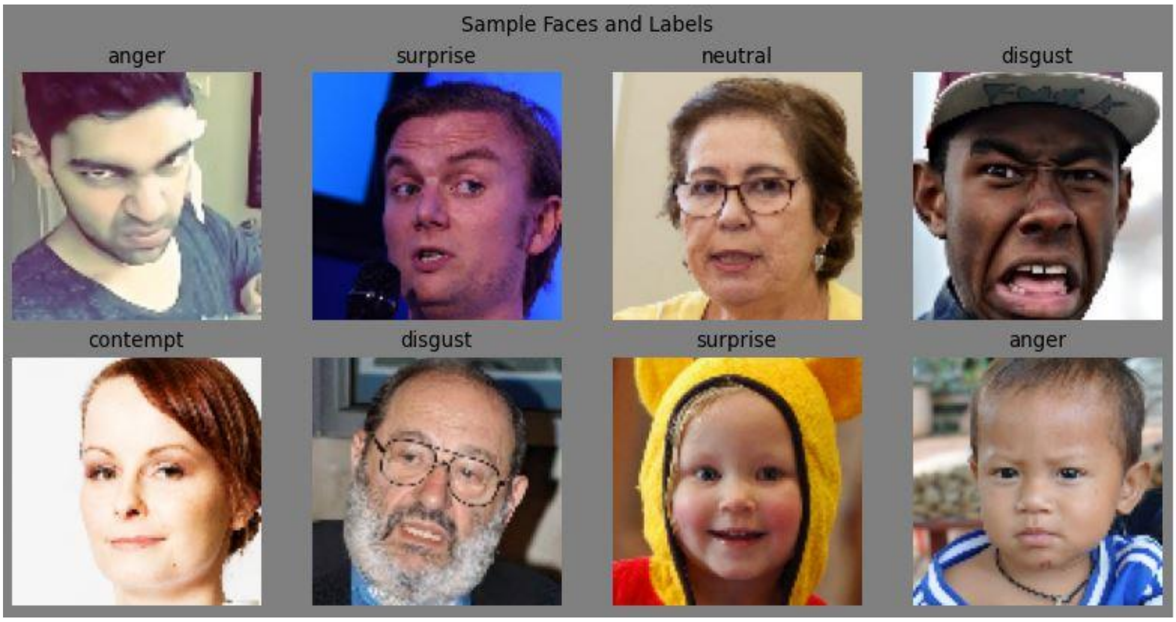

Figure 2. Sample of the face categories.

2.2 Image pre-processing

Image pre-processing improves the quality of input data and provides more useful information for training models. Consequently, it can improve the performance of deep learning models and reduce training time [21].

It enables several operations to be performed on an image before it is fed into a deep learning model. Among the main benefits of pre-processing are: noise reduction, dimensionality reduction to reduce the image size and therefore reduce training time and increase analysis speed, edge enhancement to improve the quality of the image's edges and improve the differentiation of objects and features within the image, colour conversion to improve the appearance of the image, Increased data diversity by making changes to the input images, such as rotating, flipping, resizing, adjusting contrast, and distorting them in a known way[22] . This helps improve the model's ability to recognize different objects and features in the images [23].

In the proposed work, the train data object was configured to apply various transformations to the training images, including rescaling, rotation, horizontal and vertical shifts, shearing, and zooming. These transformations can help the model learn to be more robust to variations in the training data.

The validation data object was configured only to rescale the validation images since data augmentation is not typically applied to the validation set.

2.3 Convolutional Neural Network (CNN)

CNN is a type of artificial neural network that is particularly well-suited for image processing and analysis tasks, it is based on the idea of using a series of convolutional layers to extract and transform features from input images, which are then used to perform classification or other tasks [24].

The basic architecture of a CNN consists of several layers shown in (Fig. 3), including convolutional layers, pooling layers, and fully connected layers. The input to the network is a 2-dimensional array of pixel values that represents an image, and the output is a prediction or classification of the image based on its features [25].

A convolutional layer is often the first layer in a CNN, and it applies a collection of filters to the input picture to extract important information. Each filter performs a convolution operation on the input image, sliding over it in a process known as "convolution". The output of this layer is a set of feature maps, which represent different aspects of the image [26].

The next layer in a CNN is typically a pooling layer, which decreases the spatial dimensions of the feature maps by subsampling them. This reduces the amount of parameters in the network and can help to prevent overfitting [27].

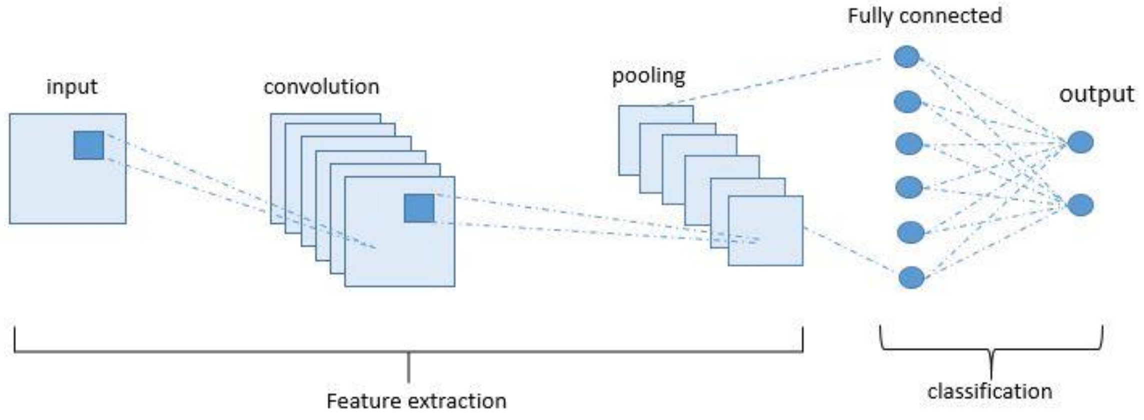

Figure 3. The basic architecture of CNN

During training, the weights of the network are adjusted to minimize the error between the predicted output and the true output. This is typically done using a technique called backpropagation, which computes the gradients of the loss function with respect to the weights of the network and updates them accordingly [28], [29].

In this paper, 2 convolutional layers, 2 max pooling layers, and 2 dense layers was used as shown in (Fig. 4), and trains it using the ImageDataGenerators train_generator and validation_generator created previously.

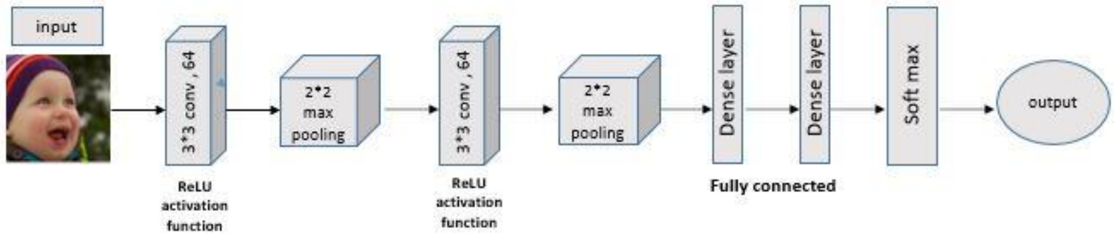

Figure 4. Architecture of the proposed CNN

The architecture of the model is defined using the Sequential API of Keras. The first 2 layers are convolutional layers with 64 filters each, a filter size of (3x3), and a ReLU activation function. The input shape of the first layer is specified as (224, 224, 3). After each convolutional layer, a max pooling layer with a 2x2 pool size is added.

The output of the max pooling layer is flattened and fed into a fully connected (dense) layer with 128 neurons and a ReLU activation function. Finally, a dense layer with output neurons and a softmax activation function is implemented to output the predicted class probabilities.

The Adam optimizer with a learning rate of 0.0001 and a categorical cross-entropy loss function is used to build the model. The accuracy metric is also supplied to track the model's performance during training.

The model is trained using the fit method of the model object with 100 steps per epoch, 30 epochs, and a batch size of 64. The train_generator and validation_generator are used as input data for the model, and the dataset was divided into a training and validation set of 75% for training and 25% for validation.

To achieve the goal of the research and to demonstrate the effect of increasing the number of classes on the resulting accuracy of the CNN algorithm, three experiments with the same settings were conducted. The first experiment included the use and classification of all 8 classes in the database, which are (Happy, Sad, Neutral, Surprise, Fear, Disgust, Anger, Contempt), and extracted the accuracy, while in the second experiment 4 classes (Anger, Surprise, Sad, Happy) were chosen, and in the last experiment only 2 classes (Happy, Sad).

2.4 Evaluation Stage

Accuracy is a performance metric used to measure the effectiveness of a machine learning model [30]. It reflects the proportion of correct predictions made to total predictions produced by the model. In other words, accuracy is the fraction of properly identified occurrences in the dataset out of all instances.

3. Result and Discussion

This part discusses the research results and provides a thorough discussion. Google Colab was used to run all experiments, which is a Python-based integrated development environment used for deep learning, data analysis, and scientific programming based on the Jupyter Notebook environment. The results showed a significant difference in accuracy between all experiments in terms of training and evaluation. As shown in (Table 2) and (Fig. 5), all experiments were done on the same settings related to image processing and algorithm using only 30 epochs.

Table 2. Performance of CNN model with different classes

Dataset	Training loss	Training accuracy	Validation loss	Validation accuracy
Dataset with 8 classes	1.784	35%	1.884	30%
Dataset with 4 classes	1.019	55%	1.319	44%
Dataset with 2 classes	0.3047	87%	0.3262	85%

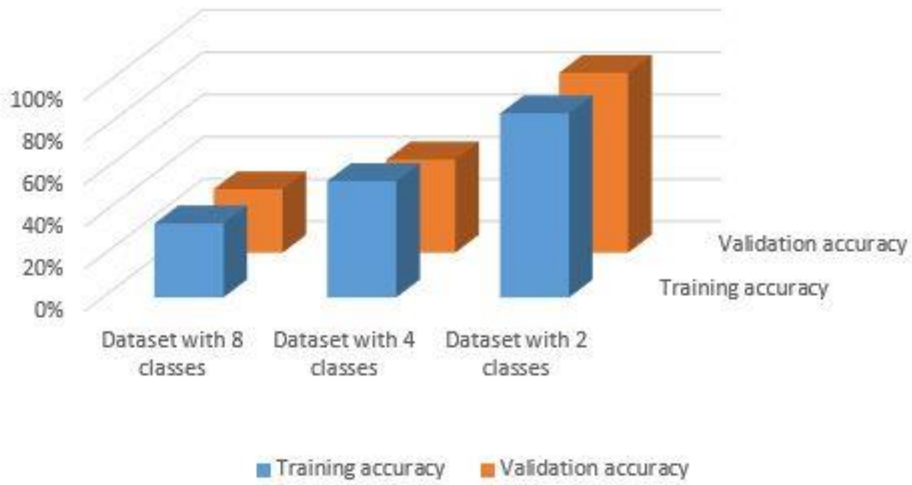

Figure 5. Visualization of the accuracy between the three experiments

When the model is trained on a dataset with 8 classes, the training accuracy and validation accuracy are relatively low at 35% and 30%, respectively. This suggests that the model is struggling to learn and distinguish between the various classes.

However, when the number of classes in the dataset is reduced to 4, the accuracy of the model improves significantly, with a training accuracy of 55% and a validation accuracy of 44%. This suggests that reducing the number of classes in the dataset makes it easier for the model to distinguish between them.

Furthermore, when the number of classes in the dataset is reduced to 2, the accuracy of the model improves even further, with a training accuracy of 87% and a validation accuracy of 85%. This suggests that the model is able to perform better when there are fewer classes to distinguish between.

It is also possible to note the training loss and validation loss which are metrics used to evaluate the performance of the CNN for the three groups, where the difference was clear as shown in (Fig. 6).

For the first dataset with 8 classes, the training loss was 1.784 and the validation loss was 1.884. This indicates that the model struggled to learn the patterns in the data, and the validation loss was higher than the training loss, which suggests overfitting.

For the second dataset with 4 classes, the training loss was 1.019 and the validation loss was 1.319. These values were lower than those of the first dataset, indicating that the model improved in learning the patterns in the data, but the validation loss was still higher than the training loss, suggesting overfitting.

For the third dataset with 2 classes, the training loss was 0.3047 and the validation loss was 0.3262. These values were the lowest among the three datasets, indicating that the model effectively learned the patterns in the data and generalized well to new, unseen data.

Overall, these results suggest that the number of classes in the dataset plays a crucial role in determining the accuracy of the CNN model, with a smaller number of classes generally leading to better performance.

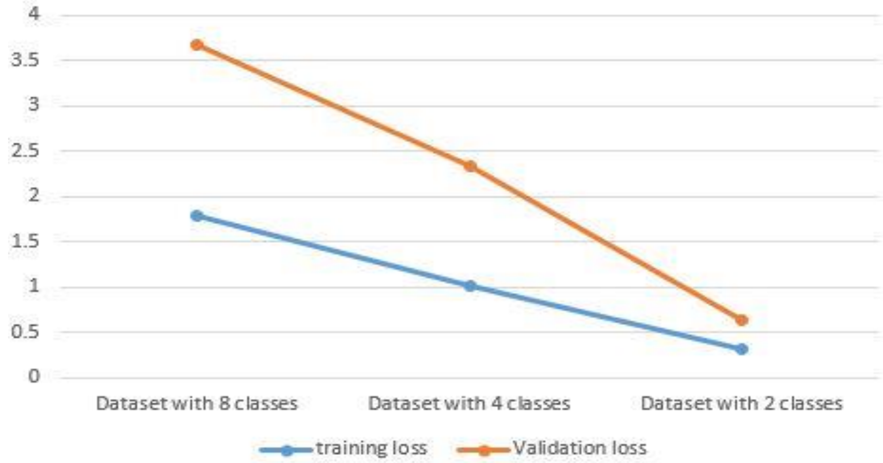

Figure 6. Visualization of the training and validation loss between the three experiments

In addition to the number of classes, the quality of the dataset itself can be a reason for decreasing accuracy. For example, when looking at a sample of the images of surprise expressions, as in (Fig. 7), we notice that not all expressions are clear, and the algorithm may misunderstand them, leading to confusion in their performance. When there are more classes in the dataset, it can become harder for the CNN model to effectively learn and distinguish between them, especially if the classes are similar or have overlapping features. This can lead to increased confusion and misclassification, resulting in lower accuracy scores. On the other hand, when the number of classes in the dataset is reduced, it can become easier for the CNN model to learn and distinguish between the classes, leading to higher accuracy scores. A high-quality and diverse dataset with well-labeled and representative samples can help the model better learn and generalize patterns across different classes, leading to higher accuracy scores.

Figure 7. Samples of surprise expression

4. Conclusion

Convolutional neural networks are one of the most widely used deep learning algorithms in different fields and have high efficiency. Most of the previous research focused on image processing methods and increasing or decreasing the convolutional layers of the algorithm to improve its performance. In this research, the effect of the database itself on the performance of the algorithm was studied in terms of the number of classes within the database. A database with many classes was used, and in three stages, the number of classes was reduced to note how it could affect its performance. The results demonstrate that reducing the number of classes in a dataset can lead to improved accuracy in the CNN algorithm, as the model is better able to distinguish between smaller numbers of classes.

References

[1] C. Zuo et al., "Deep learning in optical metrology: a review,", Light Sci Appl, vol. 11, no. 1, p. 39, 2022, doi: 10.1038/s41377-022-00714-x.

[2] M. Z. Alom et al., "A state-of-the-art survey on deep learning theory and architectures,", Electronics, vol. 8, no. 3, p. 292, 2019, doi: 10.3390/electronics8030292.

[3] T. Kattenborn, J. Leitloff, F. Schiefer, S. J. I. j. o. p. Hinz, and r. sensing, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,", J ISPRS, vol. 173, pp. 24-49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[4] J. Gu et al., "Recent advances in convolutional neural networks,", Pattern Recognition, vol. 77, pp. 354-377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[5] P. T. Q. Anh, D. Q. Thuyet, Y. J. P. B. Kobayashi, and Technology, "Image classification of root-trimmed garlic using multi-label and multi-class classification with deep convolutional neural network,", Postharvest Biology and Technology, vol. 190, p. 111956, 2022, doi: 10.1016/j.postharvbio.2022.111956.

[6] M. M. Yusro, R. Ali, and M. S. J. B. S. J. Hitam, "Comparison of Faster R-CNN and YOLOv5 for Overlapping Objects Recognition,", BSJ, vol. 20, no. 3, pp. 0893-0893, 2023, doi: 10.21123/bsj.2022.7243.

[7] A. K. Sharma et al., "Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network,", IEEE Access, vol. 10, pp. 17920-17932, 2022, doi: 10.1109/ACCESS.2022.3149824.

[8] F. S. Hanoon, A. H. H. J. I. J. o. E. Alasadi, and C. Engineering, "A modified residual network for detection and classification of Alzheimer's disease,", IJECE, vol. 12, no. 4, pp. 4400-4407, 2022, doi: 10.11591/ijece.v12i4.pp4400-4407.

[9] H. S. Das, A. Das, A. Neog, S. Mallik, K. Bora, and Z. J. F. i. G. Zhao, "Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach,", Frontiers, vol. 13, p. 1097207, 2023.

[10] H. O. Ikromovich and B. B. J. O. A. R. Mamatkulovich, "FACIAL RECOGNITION USING TRANSFER LEARNING IN THE DEEP CNN", Open Access Repository, vol. 4, no. 3, pp. 502-507, 2023, doi: 10.17605/OSF.IO/NRMK2.

[11] Y. Cao et al., "MBANet: A 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images,", Biomedical Signal Processing and Control, vol. 80, p. 104296, 2023, doi: 10.1016/j.bspc.2022.104296.

[12] J. Alfred Daniel et al., "Fully convolutional neural networks for LIDAR–camera fusion for pedestrian detection in autonomous vehicle,", Multimedia Tools and Applications, pp. 1-24, 2023.

[13] M. Ahmadi, A. Sharifi, M. Jafarian Fard, and N. J. I. j. o. n. Soleimani, "Detection of brain lesion location in MRI images using convolutional neural network and robust PCA,", International Journal of Neuroscience, vol. 133, no. 1, pp. 55-66, 2023, doi: 10.1080/00207454.2021.1883602.

[14] H. Chen, L. Wu, J. Chen, W. Lu, J. J. I. P. Ding, and Management, "A comparative study of automated legal text classification using random forests and deep learning,", Information Processing & Management, vol. 59, no. 2, p. 102798, 2022, doi: 10.1016/j.ipm.2021.102798.

[15] S. Soni, S. S. Chouhan, and S. S. J. A. I. Rathore, "TextConvoNet: A convolutional neural network based architecture for text classification,", Applied Intelligence, vol. 53, no. 11, pp. 14249-14268, 2023.

[16] V. R. J. I. J. o. C. Allugunti, Programming and D. Management, "A machine learning model for skin disease classification using convolution neural network,", International Journal of Computing, Programming and Database Management, vol. 3, no. 1, pp. 141-147, 2022.

[17] A. S. Musallam, A. S. Sherif, and M. K. J. I. a. Hussein, "A new convolutional neural network architecture for automatic detection of brain tumors in magnetic resonance imaging images,", IEEE Access, vol. 10, pp. 2775-2782, 2022, doi: 10.1109/ACCESS.2022.3140289.

[18] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. J. I. a. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure,", IEEE access, vol. 7, pp. 41273-41285, 2019, doi: 10.1109/ACCESS.2019.2907327.

[19] N. M. A.-M. M. Al and R. S. J. I. Khudeyer, "ResNet-34/DR: a residual convolutional neural network for the diagnosis of diabetic retinopathy,", Informatica, vol. 45, no. 7, 2021.

[20] A. Mollahosseini, B. Hasani, and M. H. J. I. T. o. A. C. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild,", IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 2017, doi: 10.1109/TAFFC.2017.2740923.

[21] A. M. Hasan, A. F. Qasim, H. A. Jalab, and R. W. J. B. S. J. Ibrahim, "Breast cancer MRI classification based on fractional entropy image enhancement and deep feature extraction,", BSJ, vol. 20, no. 1, pp. 0221-0221, 2023, doi: 10.21123/bsj.2022.6782.

[22] R. Sarki et al., "Image preprocessing in classification and identification of diabetic eye diseases,", Data Science Engineering, vol. 6, no. 4, pp. 455-471, 2021, doi: 10.1007/s41019-021-00167-z.

[23] D. G. J. J. o. I. I. P. Ranganathan, "A study to find facts behind preprocessing on deep learning algorithms,", Journal of Innovative Image Processing, vol. 3, no. 1, pp. 66-74, 2021, doi: 10.36548/jiip.2021.1.006.

[24] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. J. P. o. t. I. Müller, "Explaining deep neural networks and beyond: A review of methods and applications,", IEEE, vol. 109, no. 3, pp. 247-278, 2021, doi: 10.1109/JPROC.2021.3060483.

[25] Z. Li, F. Liu, W. Yang, S. Peng, J. J. I. t. o. n. n. Zhou, and l. systems, "A survey of convolutional neural networks: analysis, applications, and prospects,", IEEE, vol. 33, no. 12, pp. 6999 – 7019, 2021, doi: 10.1109/TNNLS.2021.3084827.

[26] S. Cong and Y. J. A. I. R. Zhou, "A review of convolutional neural network architectures and their optimizations,", Artificial Intelligence Review, vol. 56, no. 3, pp. 1905-1969, 2023.

[27] A. Khan, A. Sohail, U. Zahoora, and A. S. J. A. i. r. Qureshi, "A survey of the recent architectures of deep convolutional neural networks,", Artificial Intelligence Review, vol. 53, pp. 5455-5516, 2020.

[28] D. Negrov, I. Karandashev, V. Shakirov, Y. Matveyev, W. Dunin-Barkowski, and A. J. N. Zenkevich, "An approximate backpropagation learning rule for memristor based neural networks using synaptic plasticity,", Neurocomputing, vol. 237, pp. 193-199, 2017, doi: 10.1016/j.neucom.2016.10.061.

[29] S. Kiranyaz et al., "1D convolutional neural networks and applications: A survey,", Mechanical Systems and Signal Processing, vol. 151, p. 107398, 2021, doi: 10.1016/j.ymssp.2020.107398.

[30] V. H. Phung and E. J. J. A. S. Rhee, "A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets,", Applied Sciences, vol. 9, no. 21, p. 4500, 2019, doi: 10.3390/app9214500.