Deep Learning Models for Image Classification: Comparison and Applications

2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-6654-3789-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICACITE53722.2022.9823516

Shagun Sharma
Chitkara University Institute of Engineering & Technology,
Chitkara University, Punjab, India
shagun.sharma@chitkara.edu.in

Kalpna Guleria
Chitkara University Institute of Engineering & Technology,
Chitkara University, Punjab, India
*kalpna@chitkara.edu.in

Abstract— **Deep learning is the subfield of machine learning which performs data interpretation and integrates several layers of features to produce prediction outcomes. It has a significant performance in a wide range of sectors, specifically in the realm of image classification, object identification and segmentation. Deep learning algorithms have significantly enhanced the effectiveness of fine-grained classification tasks, which aims to distinguish among the sub-classes. In this review, a detailed analysis of the various deep learning models, comparative analysis and their frameworks, as well as model descriptions have been presented. Convolutional Neural Networks, have been found as the standard method for object recognition, computer vision, image classification, and other applications. However, as input data becomes more intricate, traditional convolutional neural network is no longer capable of delivering adequate results. As an outcome, the goal of this review article is to put several deep learning models along with their methodologies back to prominence and to present their findings on a wide range of popular databases.**

Keywords— Machine Learning, Deep Learning, Convolutional Neural Network, AlexNet, GoogleNet

I. INTRODUCTION

Artificial intelligence (AI) is a word defined by the machine intelligence. In computer engineering, AI study is described as the analysis of "intelligent agents," or devices that sense their surroundings and conduct activities to improve their chances of achieving an objective. AI is used, when a machine duplicates "cognitive" capabilities that humans identify with some other human minds, such as" problem solving," and "learning" the phrase [1]. Knowledge, reasoning, planning, Natural Language Processing (NLP), learning, and perception, along with the capability of moving and manipulating the objects are fundamental objectives of AI [1][2]. One of the AI objectives includes rudimentary intelligence. Statistical methodologies, intelligent systems, and classical symbolic AI are among the techniques of AI. It employs a wide variety of tools, including mathematical optimization, logic, economics and variations. Computer science, psychology, mathematics, philosophy, linguistics, artificial psychology and neuroscience are all used in the AI field. AI has many advantages in real world applications including e-commerce, fraud detection, voice assistants, autonomous vehicles along with face identification and classification.

Machine Learning (ML) is a branch of AI [2][3]. Since the 1950s, computer researchers have been working in the field of ML to understand various concepts of prediction and forecasting. It has been developed tremendously during the past several decades. As a result, machines are expected to perform better in various fields from security to health prediction. The application of emerging fields is constantly an ongoing endeavor in the scientific community, as research learning and understanding is brought presented across many new domains.

Intelligent Machinery term has been coined in the 1950s which introduced a new field, where machines were aiming to get intelligent as humans. It was an initiative step in entering a new period. In 1948, Two scientists, Turing and Champernowne invented a chess game called 'paper & pencil' [2]. It was one of the first software program. The software was written using a paper and pencil, with Turing and Champernowne practically and physically performing the calculations every step would take approximately 30 or more minutes to determine [4].

Deep Learning also called subfield of ML, is an example of this approach [3]. ML is the process of understanding the principles from a tremendous number of past information using comparable techniques, testing hypotheses or assessments on new test data, and understanding the concepts like humans. In the realm of ML, DL is a relatively recent concept where it goals to mimic and create the working of human brain network for evaluation and retraining. It simulates the biological human brain's mechanism for processing and understanding the data such as voice, text, and images. It's a type of self-training, based on Artificial Neural Network (ANN) model. A DL structure is a multi-layer perceptron with various hidden layers to identify the dispersed features of data, these features are stated layer by layer, and more feature extraction is elevated which generates semantics by the combination of reduced features to reflect attribute classes or characteristics [3].

DL focuses in depth for building the model depth which contains four five or more hidden layers for highlighting the relevance of feature extraction [3]. Extracting the features from images, is the most important aspect of a pattern matching systems in the context of image classification. The accuracy of extracting features has a direct impact on the recognition rate. DL obtains the best representation of characteristics using layer by layer feature translation; it earns features from text and images automatically

The ability to learn vast volumes of information is among the advantages of DL [5]. In recent years, DL has grown rapidly, and it's been effectively applied to a plethora of different applications. More significantly, throughout many sectors, including NLP, cybersecurity, text mining, image classification, video recommendation, robotics, and bioinformatics, DL has surpassed various ML based methodological models. DL is also characterized as

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.

Representation Learning (RL). It is based on a traditional neural network, but somehow it performs better than its predecessors significantly. Furthermore, in establishing multi-layer learning framework, DL applies both graph technology and transformations. DL is the method of training many levels of interpretations and abstractions of the fundamental data distribution to be modelled efficiently. A DL algorithm automatically retrieves the features required for classification [5]. A characteristic that is hierarchical dependent on certain features is referred to as a top/high level feature extraction. As an example, the computer vision is a DL classifier which learns low level features from the raw image data and then builds representations based on the extracted features, and again repeats the same process for high level feature extraction. There are various advantages of using DL because it can be used in autonomous voice translation in the field of electronics [5]. As an example of it home assistant devices are used to recognize the human voice and respond to that by performing computational DL based algorithm. Furthermore, medical researchers also use DL models to detect various health issues present in the human body based on the past data [2][3].

This review is divided into IV sections, where section II describes various DL models including various architecture of CNN models for image categorization. Further, Section III, shows a comparative analysis of various existing CNN frameworks and their applications on various kinds of datasets. Lastly, the results and future scope of the review is mentioned in section IV.

II. Deep Learning Models for Image Classifications

There are various DL models used for image classification including: Artificial Neural Network (ANN), Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN). These models are different from each other in case of various parameters such as applications and input data as shown in Table I.

ANN is a biological human brain which established the architecture of the human working procedure, and ANN is derived from biological NNs [3]. It works on the same mechanism as human brain, consisting of various neurons, which are coupled to each other in multiple layers in such a network where these nodes are called as neurons and the connection among them are called as edges. It has activation functions, input, output and hidden layers in the architecture, and the input to the next layer is the output from the previous layer [6]. It has only one input and output layers but hidden layers may vary as per the requirements of complexity of the problem. Every neuron and connection in the architecture has some value and weights respectively, which are further multiplied with each and summed up for substituting to the next layer.

RNN architecture is determined not just by the present inputs, but also by assigning new inputs to it [6]. These networks produce results consisting of a combination of previous and present knowledge. This type of NN works best with the time series data where the data values may vary with the increase in time. These networks are widely used in various applications such as Siri, Google translate and voice assistant.

CNN is a DL model designed to process visual input like images, animations and videos [6]. It includes a variety of layer to perform various functions. Convolution layer, pooling layer, fully connected layer, and dropout layer are the names of these layers. It also contains activation functions including sigmoid, ReLu and SoftMax function having different ranges [6][7][8]. Further, CNN is a model which is most oftenly used DL framework and utilized in a wide range of tasks and real time applications, including NLP, computer vision, image classification and many more. It is a type of multi-layer NN that is intended to analyze visual features directly from image patches with little or no preprocessing. In recent years, CNN have utterly outplayed the computer vision sector. An input, output, and hidden layers comprise a CNN [9]. This model takes the dataset that has already been handed to it for training and testing purposes and forecasts the probable labels that will be provided in future. CNN can employ any type of data because of its capabilities to overcome the complexity constraint.

Image categorization, identification, object recognition, and image captioning are some of the areas where CNN models are extensively used [9][10]. To capture generic data characteristics, CNN directly applies various convolution operations on the images for effective prediction outcomes [9][11]. This model can represent and organize image data in a distributed way for quickly gathering the visual features from a massive amount of data. CNNs have a framework that allows them to handle difficult regression equations and non-linear challenges successfully [11]. Weight-sharing, spatial subsampling and sparse connections are features of CNNs, which result in a flexible framework. There are various layers such as pooling layers, normalizing layers, convolutional layers, fully connected layers, and some common hidden layers along with activation functions. For increasingly sophisticated simulations, additional layers can also be utilized. CNNs have attained a revolution performance in the field of image categorization [9]. In the convolution layer, N filters can be applied based on requirement. Feature maps can be created by convolving them with a source images and then processing the outcome with a nonlinear activation function. Conventional CNNs are used for traditional image categorization tasks instead of hyperspectral image recognition, which need an effective utilization of both spectral and spatial associations. The essential benefits of using CNN over its counterparts would be that it discovers essential traits without the need for human intervention. AlexNet, ResNet, GoogLeNet, and VGG are successful implementations and examples of CNN models [12]. Table I. Shows various advantages of using CNN network over other DL models.

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.

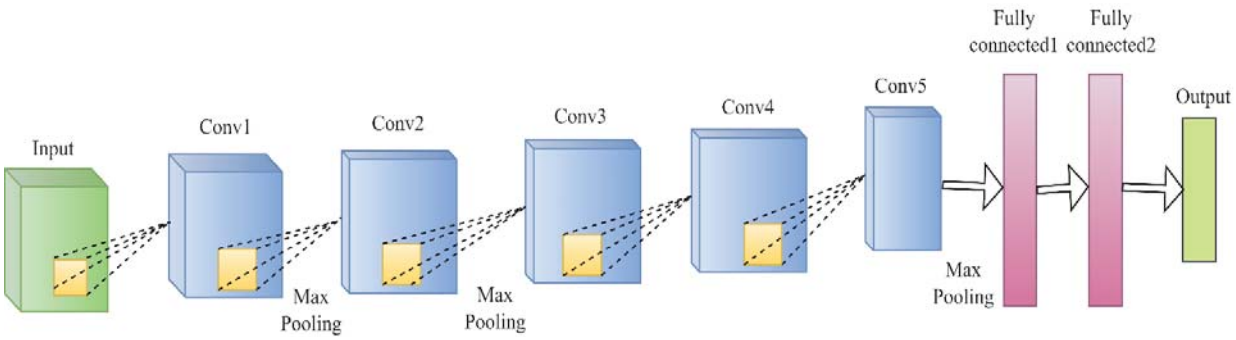

Fig. 1. Architecture of AlexNet model [13]

A. AlexNet

Alex Krizhevsky was the main designer of AlexNet. In the year of 2012, an image classification competition has been held namely ImageNet, in which AlexNet won by a large margin of 11% over the second place winner [13][12]. The AlexNet algorithm first used DL to image recognition and segmentation in 2012, resulting in major advancements in CNNs and propelling the evolution of CNN significantly [13]. When compared to standard neural networks, the CNN dramatically reduces number of parameters, network complexity, and efficiently eliminates the overfitting challenges. AlexNet includes three max-pooling layers, five convolutional layers, two fully connected layers, one softmax/output layer, and two normalization layers in its design as shown in Fig. 1. A nonlinear activation function ReLU and Convolutional filters are used in every convolutional layer. Max or average pooling is done by using pooling layer. It is primarily utilized to minimize the size of the number of parameters and representation [7]. Attributed to the prevalence of fully connected layers, the input size is always fixed which is usually stated as 224x224x3, however, it is actually 227x227x3 due to padding. AlexNet has overall 60 million parameters and 650,000 neurons [13][12].

B. ResNet

ResNet is a deep residual network built by Kaiming He, Shaoqing Ren, Xiangyu Zhang and Jian Sun in the year of 2016 [14][15]. The DL model training takes much longer and is restricted to a specific amount of layers, hence ResNet framework is designed to address difficulties in training the DL models. The rationale for ResNet's complexity is to use it for skipping a connectivity or create a shortcut. When ResNet is compared to other frameworks, it retrieves various advantages of organization to operate even as the design becomes more complicated. DL models have an inclination to include a lot of layers in order to obtain essential components and features from multiple pattern-based images. As a result, the initial layers may recognize edges, while the latter layers may identify.

However, if the network has above than 30 layers, the accuracy falls and efficiency drops [15]. This contradicts the popular belief that adding layers to a neural network improves its performance. It is not because of overfitting issues, but in that scenario, regularization and dropout methods may be utilized to eliminate the problem entirely.

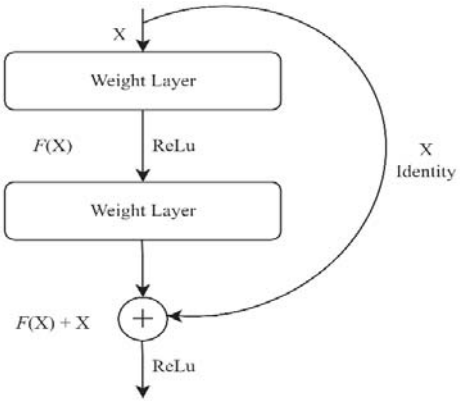

Fig. 2. ResNet Residual Block

It's mostly there because of the well-known vanishing gradient issue. Hence, the concept of ResNet comes into play to resolve this issue by skipping shortcuts and connections, to jump through certain layers [8]. The majority of ResNet algorithm use double- or sometimes triple-layer bypasses with batch normalization and nonlinearities (ReLU) in between as shown in Fig. 2.

C. GoogleNet

In 2014, GoogLeNet led the pack in the ImageNet contest. It has been invented by, Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, and Alexander A. Alemi with certain adjustments in network width and depth, the architecture derives a framework from AlexNet and LeNet [16][17]. The GoogleNet have been implemented to reduce the problem of vanishing gradient, in which the The values of the loss become zero if the number of layers are increased by feeding any activation functions to NNs, and then makes the network difficult to train. For instance, if the weight is too small in initial stage and backtracks to update the weight value, and almost becomes negligible which leads to the problem for a NNs to understand the value resulting in vanishing gradient issue [17]. To avoid this problem, the GoogleNet has introduced the concept of inception module as shown in Fig. 3, which has fixed convolution size for every layer in the architecture. GoogleNet consists of 22 layers and solves various problems including overfitting. In contrast, deepening the network to improve generalisation ability will have a slew of undesirable consequences, including gradient disappearance, gradient explosion, and overfitting. GoogLeNet increases training accuracy by making better use

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.

of computational resources, i.e. obtaining additional characteristics for the same time spent computing [16][18].

The Inception module, parallel performs 1*1, 3*3, 5*5 convolution filter and a 3*3 max pooling then concatenates them for further giving the output to next layers. In this, the convolution filters have different sizes to handle the objects of numerous scales. Furthermore, this inception module also has some complexity in case of dimensions, hence, another module with dimensionality reduction has been created named as inception module with dimensionality reduction. This module is used to enable more effective and deeper computation by reducing dimensionality with layered 1*1 convolution operations. Where, the modules were created to address difficulties such as computational expense and overfitting, among others.

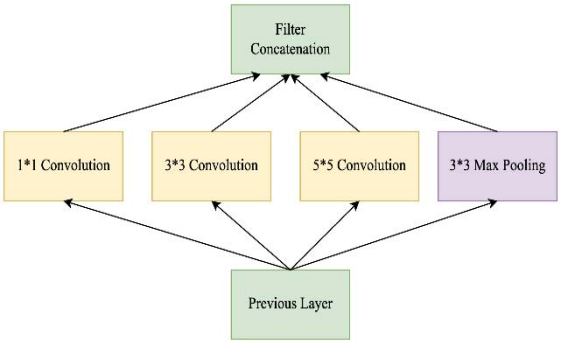

Fig. 3. Inception Module

D. VGG

VGG refers for Visual Geometry Group, and that is a multilayer deep CNN framework. The term "deep" represents the total numbers of layers in VGG-19 or VGG-16, which have 19 and 16 convolutional layers respectively [18][19][20]. It is framework serves as the foundation for cutting-edge object detection methods. VGG outperforms baselines on a variety of datasets in addition to ImageNet. Furthermore, it has been found as one of the most widely used image classification architectures today [20]. The Google DeepMind researchers and Oxford University visual geometry group collaborated on VGGNet which uses a multi-layer functions and is predicated on the CNN paradigm. Because of its simple structure and high performance, it is widely employed in various application including healthcare and many more [6].

III. COMPARATIVE ANALYSIS AND DISCUSSION

This section describes the state of art of various existing articles on different CNN architectures.

TABLE I. COMPARATIVE ANALYSIS OF VARIOUS IMAGE CLASSIFICATIONS MODELS USING DIFFERENT DATASETS ALONG WITH THEIR APPLICATIONS.

Article No.	Model	Image Size	Layers	Optimizer	Accuracy	Application
[7]	AlexNet	**227×227**	**7 Layers**	**-**	**98%**	Phonocardiogram signals classification
[8]	ResNet	**28×28**	**22 Layers**	**Adam**	**99.3%**	Handwritten digits classification
[12]	Combined AlexNet, chaotic bat algorithm and Extreme Learning Machine (ELM) model	**-**	**-**	**-**	**85.71%**	Abnormal and Normal brain classification
[13]	AlexNet	**32×32**	**8 Layers**	**-**	**87.2%**	Object Classification (automobile, bird, cat, airplane, deer, frog, dog, horse, truck and ship)
[14]	ResNet	**567×430 - 775×522**	**-**	**Stochastics Gradient Descent's (SGD)**	**88%**	Colorectal Cancer detection
[15]	ResNet-18 ResNet-34 ResNet-50	**-**	**-**	**-**	92.01% 92.55% 93.50%	Thangka Image Classification
[16]	GoogleNet	**-**	**15 Layers**	**Adam**	**98.82%**	Circular Vegetable and Fruit Classification
[17]	ResNet	**-**	**-**	**-**	**-**	Hyperspectral Image Classification
[18]	AlexNet and GoogleNet	**-**	**11 Layers**	**SGD**	**94.5% (GoogleNet)**	Heart Diseases Classification
[19]	VGG	**224x224**	**16 Layers**	**Adam**	**98.4% (Fruit Leaf) 95.71% (Vegetable**	Leaf Disease Classification

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.

					Leaf)	
[20]	VGG-16	**66 x 64**	**16 Layers**	**-**	**79%**	Remote Sensing Image Categorization
[21]	VGG-16	**32x32**	**16 Layers**	**-**	**-**	Object Classification (automobile, bird, cat, airplane, deer, frog, dog, horse, truck and ship)
[22]	VGG-16	**512×512**	**16 Layers**	**-**	**97.87%**	Pneumonia Classification

Wang et. al [12] proposed four variants of CNN model: BN-AlexNet, T-AlexNet, BN-AlexNet-ELM-CBA and BN-AlexNet-ELM for identifying brain diagnosis methods for brain MRI. The results of the experiment demonstrated that BN-AlexNet-ELM-CBA with the overall sensitivity, specificity and accuracy as 97.14%, 95.71% and 96.43% respectively.

On the basis of the accuracy parameter, Table. II displays the performance of several image classification models. The datasets employed in these research are diverse, and they have yielded a variety of applications for classifying images in various fields. In a comparison of the accuracy of state-of-the-art models, the ResNet model was found to be the best, with a classification accuracy of 99.3% using the MNIST dataset. The entire set of models has been applied to various sized images with varying numbers of layers applied. Using a dataset including multiple photos of different vegetables and fruits, GoogleNet was likewise determined to be the highest performing model for categorizing vegetable and fruit groups. Afterwards VGG with 16 layers has been also applied for the identification of disease in a fruit leaf which resulted an accuracy of 98.4%. Lastly, the VGG accuracy has been found in VGG-16 for classifying objects having an input image size as 32*32 with a performance accuracy of 79%.

IV. CONCLUSION

CNNs have got a huge amount of attention in last many decades. These have significant influence on vision-related tasks and image classification, which has apprehended the interest of academia. Many researchers have made significant contributions to this field, such as modifying the CNN architecture to increase its effectiveness and performance. Researchers have achieved breakthroughs in CNN via changing activation functions, creating or modifying loss functions, application-specific adjustments in architecture, architectural innovations, regularization, and designing various learning algorithms. Various CNN architectures are discussed in this study. CNN architectures like as AlexNet, GoogleNet, ResNet, and VGG have been discovered. This review includes a comparison of these architectures based on different datasets, optimizer, number of layers used and accuracy findings. It also demonstrates a large number of uses for CNN architectural variants. When dealing with image classification challenges, it is also discussed why DL plays an attention seeking role than ML. When a comparison of the state of art models has been done, a ResNet model has been found as the best model having 99.3% accuracy on MNIST dataset with a 22 layered architecture along with an input image size as 28*28 pixels. The VGG-16 has been found as the least performing CNN architecture on image classification tasks with an accuracy of 79%. Furthermore, this work can assist both academia and researchers in determining the optimum CNN model for conducting image classification tasks in various areas/fields.

REFERENCES

[1] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," Int. Conf. ICT Knowl. Eng., pp. 1–6, 2018, doi: 10.1109/ICTKE.2017.8259629.

[2] A. Sharma, K. Guleria, and N. Goyal, "Prediction of Diabetes Disease Using Machine Learning Model," Lect. Notes Electr. Eng., vol. 733 LNEE, no. March, pp. 683–692, 2021, doi: 10.1007/978-981-33-4909-4_53.

[3] A. Kaur, K. Guleria, and N. Kumar Trivedi, "Feature selection in machine learning: Methods and comparison," *in 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 789-795, 2021, doi: 10.1109/ICACITE51222.2021.9404623.

[4] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, pp. 1–6, 2018, doi: 10.1109/ICCUBEA.2018.8697857.

[5] P. K. Sarangi, K. Guleria, D. Prasad and D.K. Verma, "Stock movement prediction using neuro genetic hybrid approach and impact on growth trend due to COVID-19", International Journal of Networking and Virtual Organisations, vol. 25, no. 3-4, 2021.

[6] S. Li, L. Wang, J. Li, and Y. Yao, "Image Classification Algorithm Based on Improved AlexNet Image Classification Algorithm Based on Improved AlexNet," 2021, doi: 10.1088/1742-6596/1813/1/012051.

[7] P. Dhar, S. Dutta, and V. Mukherjee, "Cross-wavelet assisted convolution neural network (AlexNet) approach for phonocardiogram signals classification," Biomed. Signal Process. Control, vol. 63, no. September 2020, p. 102142, 2021, doi: 10.1016/j.bspc.2020.102142.

[8] D. Sarwinda, R. Hilya, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," in 5th International Conference on Computer Science and Computational Intelligence 2020 Deep, 2021, vol. 179, no. 2019, pp. 423–431, doi: 10.1016/j.procs.2021.01.025.

[9] S. L. S. Wang and Y. Zhang, "Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm," Neural Comput. Appl., vol. 4, 2020, doi: 10.1007/s00521-020-05082-4.

[10] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2019 Int. Interdiscip. PhD Work. IIPhDW 2019, pp. 117–122, 2019.

[11] E. Limonova, F. R. C. C. S. C. Ras, D. Alfonso, J. S. C. Mcst, V. V Arlazarov, and F. R. C. C. S. C. Ras, "ResNet-like Architecture with Low Hardware Requirements," pp. 6204–6211, 2021.

[12] F. U. Yuesheng et al., "Circular Fruit and Vegetable Classification Based on Optimized GoogLeNet," IEEE Access, vol. 9, pp. 113599–113611, 2021, doi: 10.1109/ACCESS.2021.3105112.

[13] D. M. Luong, Y. Takayama, S. Kuang, Q. Xiao, and S. Song, "Imbalanced Thangka Image Classification research Based on the ResNet Network Imbalanced Thangka Image Classification research Based on the ResNet Network," doi: 10.1088/1742-6596/1748/4/042054.

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.

[14] J. Song, S. Gao, Y. Zhu, and C. Ma, "A survey of remote sensing image classification based on CNNs," Big Earth Data, vol. 00, no. 00, pp. 1–23, 2019, doi: 10.1080/20964471.2019.1657720.

[15] Y. Wang, K. Li, L. Xu, Q. Wei, F. Wang, and Y. Chen, "A Depthwise Separable Fully Convolutional ResNet With ConvCRF for Semisupervised Hyperspectral Image Classification," vol. 14, pp. 4621–4632, 2021, doi: 10.1109/JSTARS.2021.3073661.

[16] K. Saito, Y. Zhao, and J. Zhong, "Heart Diseases Image Classification Based on Convolutional Neural Network," 2019 Int. Conf. Comput. Sci. Comput. Intell. Hear., pp. 930–935, 2019, doi: 10.1109/CSCI49370.2019.00177.

[17] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classi fi cation using Convolutional Neural Network VGG," Artif. Intell. Agric., vol. 6, 2022, doi: 10.1016/j.aiia.2021.12.002.

[18] M. Ye et al., "A Lightweight Model of VGG-16 for Remote Sensing Image Classification," vol. 14, pp. 6916–6922, 2021, doi: 10.1109/JSTARS.2021.3090085.

[19] S. Liu and W. Deng, "Very Deep Convolutional Neural Network Based Image Classification Using Small Training Sample Size," 2015.

[20] W. Tan et al., "Classification of COVID - 19 pneumonia from chest CT images based on reconstructed super - resolution images and VGG neural network," Heal. Inf. Sci. Syst., vol. 9, no. 1, pp. 1–12, 2021, doi: 10.1007/s13755-021-00140-0.

[21] V. Göreke, S. Vekil, and K. Serdar., "A novel classifier architecture based on deep neural network for COVID-19 detection using laboratory findings." Applied Soft Computing, vol. 106, 2021, pp. 1-8.

[22] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 9, pp. 5408–5423, 2018, doi: 10.1109/TGRS.2018.2815613.

Authorized licensed use limited to: University of Ottawa. Downloaded on October 24,2023 at 00:35:19 UTC from IEEE Xplore. Restrictions apply.