Summary:

There are many studies focusing on plant classification using different machine learning and deep learning approaches.

About 3 million species of plants have been named and classified. It is impossible for a botanist to know the
total number of named species and to identify all the plant species on earth. One way to identify and classify the plant is by their leaves. Each plant leaf serves as a tool to plant biologists and botanists for distinguishing plant species. Due to the unstoppable growth in human population and varying climate, there is an alarming threat to many plants. Many of the plants are on the verge of extinction. They are easily identified based on flowers and fruits. However these are three dimensional objects and increases complexity but plant leaves are two-dimensional in nature. Thus, they are most suitable for machine processing. Plant classification based on leaves is the easiest way to identify plant since they are the easily visible and they can be easily found and collected at all seasons, while flowers can only be obtained at blooming season. The advancement in image processing has made this a quick and easy process and the advancement of Information Technology, systems with more functionality such as automatic labeling and flexible searching are required and it is achieved through image processing and machine learning techniques. It is easy to transfer the leaf image to a computer and then the computer can extract necessary features automatically using image processing techniques and subsequently can recognize the plant / leaf using machine learning techniques.

One of the famous work on plant classification is PL@NTNET. In their paper, they used a weighted average of the softmax probability vectors. The weight of each plant picture only depends on its view type (e.g. the pictures tagged as flower are more weighted than the pictures tagged as leaf because flowers are much more discriminant than leaves). In their paper, the values of the weights have been optimized empirically.

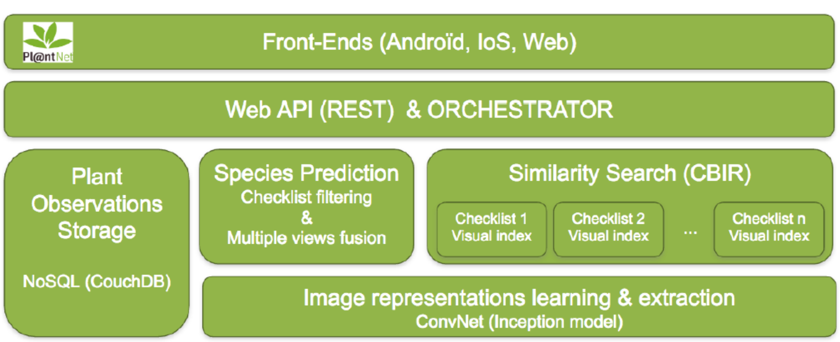


Figure 1: Pl@ntNet system architecture

Here, I summarized some of the paper on plant classification using different machine learning and deep learning approaches.

In the first paper, "**Large-Scale Plant Classification with Deep Neural Networks**" by Ignacio Heredia 2015. Their dataset is consisting around 250K images belonging to more than 6000 plant species of Western Europe. These species are distributed in 1500 genera and 200 families. Most images have resolutions ranging from 200K to 600K pixels and aspect ratios ranging from 0.5 to 2. Their dataset is highly unbalanced because most labels contain very few images.

They used convolutional neural network architecture as the ResNet model. This architecture consists of a stack of similar (so-called residual) blocks, each block being in turn a stack of convolutional layers. The innovation is that the output of a block is also connected with its own input through an identity-mapping path. This alleviates the vanishing gradient problem, improving the gradient backward in the network and allowing training much deeper networks. Their model to have 50 convolutional layers. They initialize the weights of the model with the pre-trained weights on the ImageNet dataset provided in the Lasagne Model Zoo.

During their training, they applied standard data augmentation (as sheer, translation,mirror, etc) so that the network never sees the same image. They did not apply rotation or upside down mirroring to the images tagged as 'habit', as it does not make much sense to have a tree or a landscape upside down. After applying the transformations, we downscale the image to the ResNet standard 224*224 input size.

IN their paper, they suggested to increase the training dataset size. It should be noted that iNaturalist contains

even more images than PlantNet so when training a net for deployment one should combine both datasets to increase the predictions accuracy.

They mentioned that Resnet50 is a quite space-consuming architecture so if we were to implement plant recognition app in embedded devices, so that the user could identify without connecting to the Net, one could use some recent modifications of shallower architectures that almost as good performance with much less memory consumption.

In the second paper, “**Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review**” by Yu Jiang and Changying Li, 2020.

CNNs are artificial neural networks that combine a set of mathematical operations (e.g., convolution, pooling, and activation) using various connection schemes (plain stacking, inception, and residual connection), and the operational parameters (e.g., convolutional kernels) can be learned from annotated images to predict image class labels (image classification in Figure 2). The development of modern CNNs for image classification can be grouped into three stages: (1) emergence of modern CNNs (2012 to 2014); (2) intensive development and improvement of CNN architecture (2014 to 2017); and (3) reinforcement learning for CNN architectural design (i.e., the concept of using artificial intelligence (AI) to improve AI, 2017 to present). In the graph below, they showed comparison of image classification, object detection, and semantic segmentation. In their paper, they mostly they focused on this three approach.

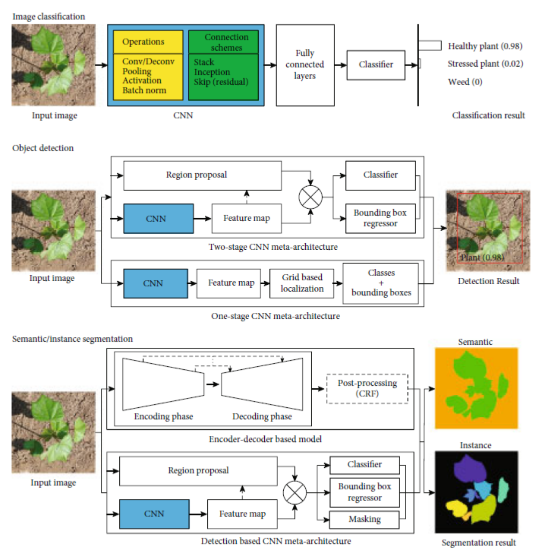


Figure 2: Diagrams of CNN architecture mechanisms for image classification, object detection, and semantic and instance segmentation.

In third paper, “**Use of deep learning techniques for identification of plant leaf stresses: A review**” by Karim Noon.

Their paper mostly focused on the plant leaf stress. General deep learning based technique for plant leaf disease recognition whereas Table A1 gives explanation of different abbreviations used throughout this paper. The input dataset usually undergoes some preprocessing e.g., contrast enhancement, normalization or data augmentation etc. The final dataset containing typically thousands of images is then split into training and testing portions. The training data (usually 70% to 80%) of the total dataset is used to train millions of parameters of a deep network. This time consuming step is typically performed with the help of an easily accessible parallel computing hardware called Graphical Processing Unit (GPU).

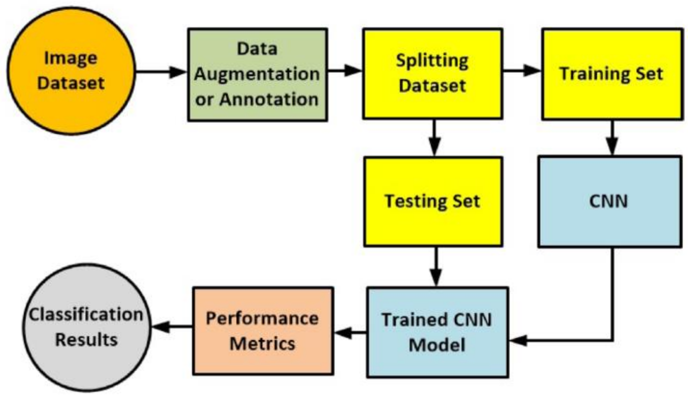


Fig. 1. Block diagram of a deep learning based plant leaf stress recognition model.

The forth paper is "**Plant species classification using deep Convolutional neural network**" by Dyrmann et al., 2016. This paper presents a method that is capable of recognizing plant species in color images by using a convolutional neural network. Here it is the list of some papers on plant recognition, the number of species and classification accuracy.

Source	#Species	% Correct	Approach
Søgaard (2005)	3	65%–93%	Active shape models
Åstrand and Baerveldt (2002)	2	91%	Colour features
Giselsson (2010)	8	94.8%	Shape and colour features
Dyrmann and Christiansen (2014)	7	95.8%	Shape and colour features
Kazmi (2014)	2	99%	Local features
Golzarian and Fricka (2011)	3	82.4%–88.2%	Shape and colour features

The methods that they used are:

1. **Data augmentation:**

A convolutional neural network is translation invariant but not rotation invariant. The plants, however, are photographed

vertically towards the ground and therefore had no fixed orientation. We can therefore generate more training data by rotating the original training data. As plants were assumed to be symmetric, even more

training data can be generated by mirroring the training data. The training set was thereby increased eight-fold by mirroring the images and rotating them in 90_ increments. After the augmentation of the training data, there were 50,864 training samples.
After the model architecture and convolutional layers, they did Batch normalization.

2. Batch normalization

Batch normalisation ensures that the inputs to layers always fall in the same range even thought the earlier layers are
updated. Batch normalization results in a significant reduction in the required number of training iterations, but the same results are obtained
as the network without normalisation. Batch normalization is applied by normalising the output of a given layer to its standard deviation.

3. Activations functions

An activation function introduces non-linear decision boundaries to the network. The rectified linear unit (ReLU) is an often used activation
function in deep learning applications, as it is considerably faster to calculate than alternatives such as the sigmoid function, while still providing good results.

4. Max-pooling layers

Max-pooling is a process, which reduces the spatial size of a feature map and provides translation invariance to the network. This is done by only keeping the maximum value within a k neighborhood in the feature map.

5. Fully connected layers

In a fully connected layer all inputs are connected to all outputs of the previous layer. This is known from traditional neural networks. Thus, a fully connected layer causes the spatial information to be removed. In convolutional neural networks, fully connected layers are usually used as a way of mapping spatial features to image labels.

6. Residual layers

Here, they introduced Residual capacity. The filter capacity is a measure of how well a filter is able to detect complex structures in images. If the capacity is small, it means that only local features in the image are mapped to the next layer. On the other hand, if the capacity is large, it means that the filter is able to find complex structures of elements that are not neighbours in the input image. For example, in the case of small plants, there will often be two leaves opposite each other.
The filter capacity is calculated as the ratio between the real filter size and the receptive field (Cao, 2015).

$$Capacity = \frac{\text{real filter size}}{\text{receptive field}}$$

In the fifth paper, "**Deep Learning for Plant Species Classification"** by Gawli 2020.

He talked about the challenges of automatic plant recognition.

1. Image quality:

Quality of the leaf image captured plays a very important role in the accuracy of plant identification. In general, the quality of leaf images is affected by three factors, namely, contrast, blur and noise. The presence of these degradation factors has a direct impact on the performance of the automation process.

2. Leaf variation:

In general scenario, a leaf image can take a great number of biological variations. These variations produce more than one representation of the same leaf. Accurate and efficient feature extraction techniques that best distinguish these similar leaves are required for successful design of automated system. Further, the availability of huge number of leaf features and selecting a subset that best enhances the process of identification is challenging.

3. Lack of standard leaf datasets:

The design and implementation of a consistent automatic plant species identification system from leaf images requires a representative database that can be used by the machine learning algorithms to identify plants accurately. There is a

lack of standard leaf image database that can be used for plant classification and therefore, the database is normally constructed by the researchers. Assembling such a database is time consuming and complex.
In the sixth paper, "**On Identifying Leaves: A Comparison of CNN with Classical ML Methods**" by Hedjazi et al., 2017.

They show that a pre-trained CNN model on a large dataset (ImageNet) can be used to train a model from a small training set (ImageCLEF2013 Plant Identification). The resulting model outperforms the classical machine learning methods using local binary patterns (LBPs), a well-utilized feature in the field.

Transfer learning is a machine learning technique that transfers knowledge learned from a source domain to a target domain. It is a very useful method to avoid over-fitting when the task related data is small. There have been successful attempts in the literature to apply transfer learning and the fine-tuning approaches in classification tasks. They also employ transfer learning and fine-tuning to training our CNN.

In general, we don't usually train an entire CNN from scratch with random initialization, because it does not only take a lot of times but it is also relatively rare to have a dataset of sufficient size that is required for the depth of network required, for example AlextNet used to recognize more than 1000 categories of generic objects that are part of the ImageNet dataset which contains more than a million image, while our dataset just contains 41970 images from 250 categories. In detail, they explained the two approach:

A. Replace and re-train the classifier on top of the CNN

Take a CNN pre-trained model on ImageNet like AlexNet or GoogleNet, ignore the output layer (this layer's outputs are the 1000 class in ImageNet dataset however our dataset has just 250 output class), train the pre-trained model from the scratch, then copy the weights to the rest of the CNN layers (we do not need to do the back-propagation).

B. Fine-tuning CNN The second strategy is not only to replace and re-train the classifier on top of the CNN on our new dataset, but to also fine-tune the weights of the pre-trained network by calculating the back-propagation. It is possible to fine-tune all the layers of the CNN, or keep some of the earlier layers fixed and only fine-tune the last layers.

When we learn models from scratch, the network parameters are initialized randomly from Gaussian distributions. For transfer learning, they used AlexNet model pre-trained on ImageNet for generic object recognition.

In the next paper, "**Identification of Plants using Deep learning: A Review**" by Sk and Wadhawan 2021. Here is summary of the paper that used deep learning for plant classification, with dataset and the accuracy.

Table 1
Previous pattern recognition methods

Author	Technique used	Dataset Size	Results
Sapna Sharma, Dr. Chitvan Gupta (2015) [1]	GLCM (grey level co-occurrence matrices: GLCM is a histogram at a given offset over an image with co-occurring greyscale values.) and Principal component analysis (PCA): the method measuring the principal component and to perform on the data basis of change, Super Vector Machine (SVM): uses for classification problem methods.	16 different classes of the leaf.	Review
T. Gaber (2015) [2]	Bagging classifier,2D based technique	using Flavia a dataset which 1907 colored images.	Accuracy: 95%
T. J. Jassmann (2015) [3]	Rectified Linear Unit (ReLU): is an activation functions and ReLu is the most used activation function in the neural network, moreover in the CNNs.	Flavia dataset contains leaves of 32 plants.	Accuracy: 60%
Hulya Yalcin (2016) [4]	Convolutional Neural Network (CNN) model.	16 plant species.	Accuracy: 97.47%
Amala Sabu (2017) [5]	K-Nearest Neighbor (KNN).	Review	Review
Lee, S.H (2017) [6]	CNN	Using 113,205 images	Accuracy: 96.3%
Ghazi, M.M (2017) [7]	Transfer Learning using AlexNet GoogLeNet and VGGN: VGGN is an object-oriented Model and supports 19 layers and VGGN is still the most popular used architecture for image recognition.	using 91,758 images.	Accuracy: 80.18%
Barbedo, J.G (2018) [8]	Using deep learning concepts	almost 50,000 images use from plantvillage dataset.	Accuracy: 81%
Barbedo, J.G.A (2018) [9]	CNN	Containing 12 different plants and 1383 images were used.	Accuracy: 84%
Zhu, X (2018) [10]	RCNN	For each species, 180 images are taken, of which 150 are taken for	Accuracy:99%
Garcia-Garcia (2018) [11]	semantic segmentation using deep learning techniques.	In this paper they use 2D and 3D datasets.	Review
Kaya, A., Keceli (2019) [12]	Transfer learning method use on deep learning.	Using total number of 54,306 images.	Accuracy: 98.70%
Noon, S.K., Amjad, M., Qureshi, M.A., Mannan, A(2020)[13]	deep learning-based	Review	Review

Based on them, their Process of Leaf Identification are:

1. **Image Acquisition**

Image acquisition is the process of collecting datasets for identification leaves. Infected leaf photographs of collected in managed settings and are processed in JPEG format. Against a white backdrop, the contaminated leaf is put smooth, the light source was mounted on either side of leaves at a temperature of 45 degrees to remove each reflection and provide even light wherever thereby increasing illumination and clarity. This crop is zoomed such that the photograph captured includes just the crop and white backdrop.

2. **Image Pre-processing**

Image pre-processing procedures are essentially used to expose information that is hidden or basically to show any features in such an image. Such methods are largely contextual and are structured to alter a picture and taking advantages of the psycho-cultural dimensions of the human sensory system. *Equalization of the histograms* and electronic filtration methods were used.

2A. Histogram equalization:

it's some of those strategies for improving images. A certain approach allocates image intensities. Among this process, the contrast between the fields rises through local contrast to greater intensity. The equalization of histograms is used to enhance computational complexity, clarity, and image consistency.

2B. Grayscale conversion:

Gray scale conversion is used for converting images into grayscale. The grayscaled conversion used the method of contrast feature and intensity enhancement techniques for converting the images and then placed them as pieces together for further processing.

2C. Binary conversion:

Create binary images on a gray image scale to use the threshold method. The Binary picture is a visual image that contains just two potential meanings for each pixel. The two shades that a contrasting picture uses are typically white and black.

2D. Noise Removal:

Digital photographs are susceptible to a plethora of Noise levels. Noise emerges from errors inthe virtual method of image acquisition which results in pixels values They can be used to eliminate linear filtering those values noise Styles. Few filters are ideal for this purpose, such as Gaussian filters or low pass filters, averaging. An ordinary filter, for example, is useful for having grain noise off the picture. A median filter and averaging filter are used for salt and paper noise removal from an image.

3. Feature extraction

Mainly the characteristics of leaf color and form. The Specific plant leaf is generally identical in color and form are considered for classification and so a specific function alone cannot achieve anticipated results.

4. Classification

Common statistical identification is the method of defining based on the previous information such as a training dataset a group of groups, or classes to which a new phenomenon belongs. More precisely, classification in this work is the method used to attribute a picture to a certain plant genus, based on its collection of features. It is a subclass of more general statistics and deep learning identification problems,including supervised learning.

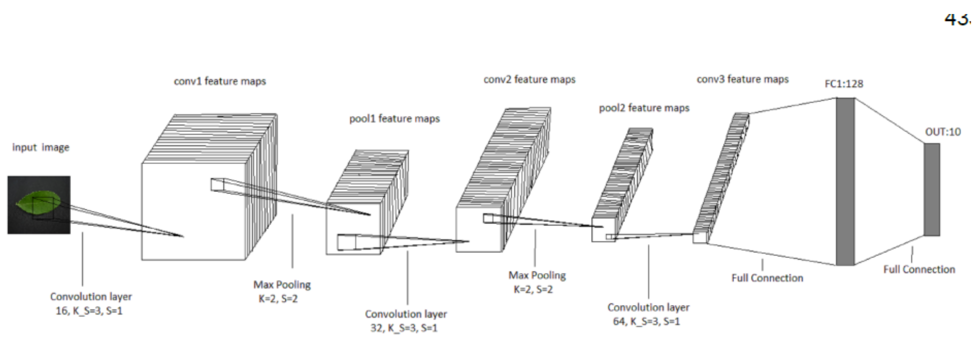


Figure 4: The architecture of our model (CNN)

Table 2
For identification, Some Layer is used for our model

Layer	No of filters	Filter size	Stride Value
1st	16	2*2	1
2nd	16	2*2	1
3rd	32	2*2	1
4th	32	2*2	1
5th	64	2*2	1
7th	64	2*2	1

In the next paper, "**Analysis of transfer learning for deep neural network based plant classification models**" by Kaya et al., 2019.

In their paper, they mentioned that Barré et al. (2017) proposed a CNN architecture for the leaf identification. They applied the Foliage, LeafSnap, and Flavia datasets and built a CNN model for the

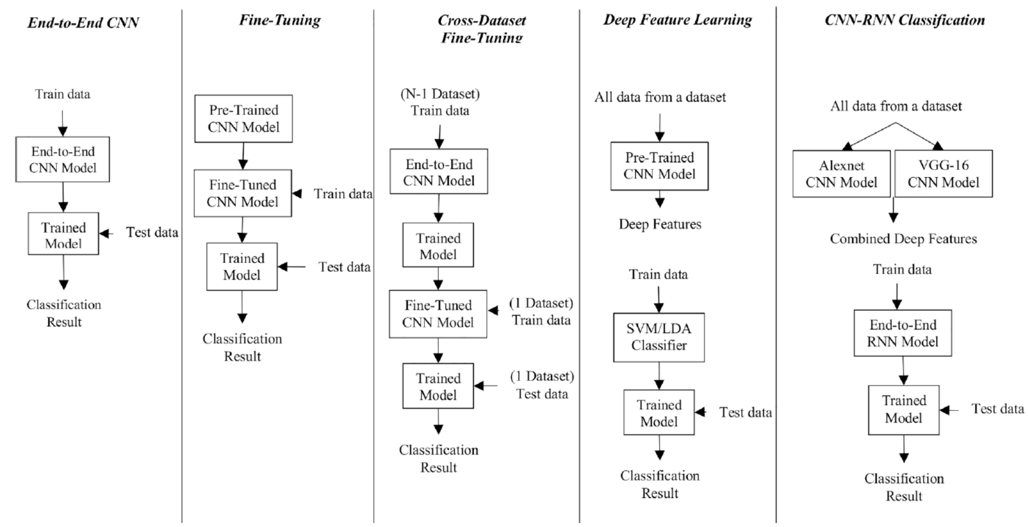
classification. An architecture similar to well-known CNNmodels like AlexNet and CifarNet was proposed in their study. Jeon and Rhee (2017) used the transfer learning from pre-trained CNN models. Pre-trained GoogleNet was used for the feature extraction. Leaf images in different scales were given to GoogleNet as input and the activation values of different layers were stored as features.
The table below shows the transfer learning that have been used in plant classification:

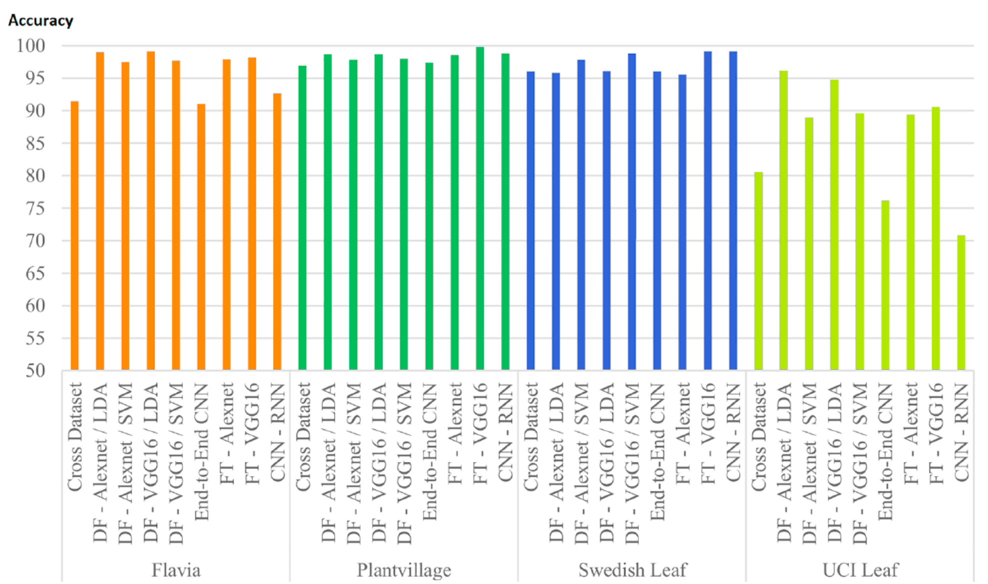
Related studies summary.

	Datasets	Best results	Method
Our Study	Flavia	99.00	DF - VGG16/LDA
	Swedish	98.80	CNN - RNN
	UCI Leaf	96.20	DF - Alexnet/LDA
	Plantvillage	99.80	FT - VGG16
Lee et al. (2017)	Flavia	99.40	CNN, Fine-Tuning
	MK	97.47	
Jeon and Rhee (2017)	Flavia	99.60	CNN
Caglayan et al. (2013)	Flavia	96.30	Hand Crafted Shape & Color Features
Yalcin and Razavi (2016)	TARBIL	97.47	CNN, SVM
Murat et al. (2017)	Flavia	95.25	HOG, Moments, ANN, RF, SVM
	Swedish	99.89	
Yousefi et al. (2017)	Flavia	97.50	Fourier & Wavelet Descriptors, MLP

Based on this paper, the two common transfer learning strategies in deep learning are deep feature extraction and fine-tuning. In the deep feature extraction, the input data is provided to the pre-trained network and activation values of various layers are stored and used as features. In fine-tuning, deep neural network is trained for a similar task, in which labeling is relatively easier. While the first layers of the pre-trained network can be fixed, fine-tuning can be done on the final layers of the model to learn the properties of the new dataset. The pre-trained model is re-trained with the new small dataset and weight values of the model are updated.

They present five Deep Neural Networks for the plant classification problem and the accuracy. These models are shown in Fig below.





In addition, here is the name of CNN model, number of parameters, and details of layers in different CNN model.

TABLE 1. Comparison of some CNN models

Year	Name of CNN	No of Parameters	Details of Layers
2012	AlexNet	60 million	8 layer
2014	GoogLeNet	4 million	22 Layers
2014	VGGNet	138 million	VGG16 - 16 layers VGG19 - 19 layers
2015	ResNet	ResNet50 - 25.6M ResNet101 - 44.5M	152 layers
2017	MobileNet	MobileNet V1 - 4.24M MobileNet V2 - 3.47M	28 layers