Live Pothole Detection: A Machine Learning Based Approach

Vedant Kangde, Mittul Daswani, Shubham Jaiswal, Tanay Kamath, Nabanita Mandal

Department of Computer Engineering, Thadomal Shahani Engineering College, Bandra, Mumbai, India
vedant.kangde@gmail.com, mmdaswani9009@gmail.com, shubhamjaiswal6501@gmail.com, 2tanaykamath@gmail.com, nabanita.mandal@thadomal.org

Abstract—**In India, around 5000 people die due to accidents caused by potholes. Time and again we see distressing news about how a pothole caused an accident and claimed the lives of some commuters. Potholes not only cause accidents but also make transport slow and clumsy, damage vehicles causing their owners precious repairs and overall are an imminence to our country's logistics and hence is of utmost significance to bring it to our government's attention and fix them as soon as possible. This composition proposes a system where images are taken from CCTV cameras as well as commuter reports. These images are given as input to the model which identifies roads grounded on their quality and creates a detailed report and as a result displays pothole areas on a chart to commuters which aids in avoiding accidents.**

Keywords—Potholes, images, roads, accidents, Convolutional Neural Network (CNN)

I. INTRODUCTION

Pothole detection is critical to accident prevention worldwide. Although many studies have been conducted, there are certain biases or tools for obtaining detector data. We provide a way to apply pothole detection using live feeds from government-installed security cameras and commuter travel reports via an app, and the brackets are done before literacy. After parentheses, individual agencies are notified of road conditions at separate locations.

Rapid advances in technology in recent years have had a major impact on the safety of transportation systems.

Smart solutions for transportation systems that aim to improve transportation systems are becoming popular. For business safety, passengers often feel uncomfortable driving on rough roads, especially over potholes in the road. According to statistics from Taiwan's Ministry of Justice, from 2008 to 2011, government compensation is about $240 million. Potholes in the road are detrimental to driver safety. Therefore, a live pothole detection system can be built to improve safety of pedestrians, drivers, etc.[3].

Also, further and further widgets include sensors, compass, gyroscope, GPS, cameras, etc. Several functions use such detectors in mobile. Thus, using mobiles with their sensors and cameras to describe potholes is suitable and accessible.[1] This study proposes a pothole discovery system grounded on mobile sensors, cameras and shares the pothole information with commuters and government. For this purpose, the mobile device should be equipped with g- detectors and gps to collect data and position information as well as a camera to take photos and include it in the report.

II. LITERATURE REVIEW

Some pothole detection solutions have been proposed that largely fall into two categories: image recognition systems, mobile surveillance systems.

A. Image Recognition Technology

The pothole detection approach proposed by Yu and Salari is based on a ray-based road information collection method as well as artificial neural network (ANN) algorithms for analyzing road information and decoding potholes[6]. It is inefficient for mobile displacement as processing the ray image requires a lot of computing power. Lin and Liu's approach used a Support Vector Machine (SVM) to analyze traffic images for pothole detection.

This approach provides high accuracy, but image recognition requires a lot of computing power and as a result this approach is not at all suitable for mobile devices.[4]

B. Mobile Detection Method

An inexpensive model for evaluating 3D images using a low-cost Kinect detector that reduces computational cost by providing direct depth measurement is proposed[5]. It uses a G-sensor and GPS that collects and analyzes accelerometer data for pothole detection[2].
Still, this approach requires the mobile to be at a specific angle. Likewise, this design only considers assaying *z*- axis accelerometer data with high misstep.
For the BusNet design, the on board unit (OBU) in the machine is fitted with a G-sensor and GPS to get data from sensors such as an accelerometer and position information. This data can be sent to data processing centers using wireless networks and then data processing centers can go through this data to check if the vectors of data from the accelerometer exceed the thresholds set for pothole discovery. This approach is dependent on the batch accelerometer data being transferred when the machine enters the machine station. Thus, this approach isn't capable of giving real time pothole discovery information.
A pothole management system proposed by a group of developers at the Massachusetts Institute of Technology combines a G-detector and GPS. (2) high-pass filter, (3) *z*-peak, (4) *xz*-rate, and (5) velocity versus *z*-rate.

III. DATASET

Dataset consists of 1200 images of Indian roads in various conditions, weather conditions, sharpness, and size. These images generally fall into three categories: normal, dusty, and potholes. As our proposed method mainly focuses on distinguishing potholes from other types of deformations and road conditions, this data set perfectly matches the parameters of our goal. The number of images is large enough to achieve very good accuracy and small enough not to consume excessive processing power and time.

IV. PROPOSED METHOD

A. Diagram

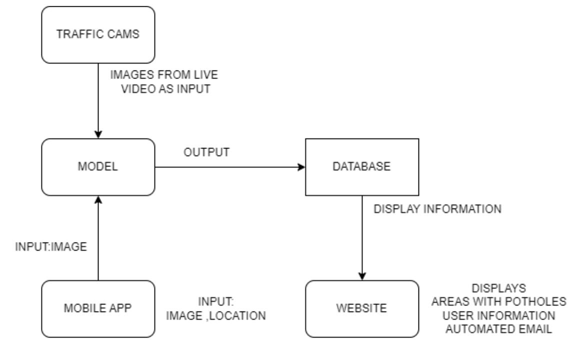

Fig. 1. Proposed system block diagram

B. Description

The system we propose is expressed as the above block diagram. The system consists of four major subsystems.
They are listed as:

Input: Input is captured in two ways. Inputs can be taken from images uploaded using the mobile app. Images are also captured using live video from road cameras.

Classification Models: CNN-based image classification models are used to segregate images into respective categories. Dataset used to train this model is merged from three existing datasets and consists of approximately 1200 images in different categories.

Database: Data consists of the output of classification models along with user information.

Output: The web application is used to mark potholes using the Google Maps API. An automated email is sent to the appropriate authority with information about the output of the classification model and its location.

C. Algorithm

Convolutional Neural Network(CNN)
CNNs are a particular kind of artificial neural network that are used for classifying images into different categories.They are similar to Recurrent neural networks used for sequence words.CNN is one the most popular algorithm for carrying out image recognition and classification tasks .A CNN model needs an image as an input after carrying out various preprocessing methods to remove unwanted information without comprising the key features of an image.A CNN model consists of different layers which are as follows:

Layers in CNN
Convolutional Layer

CNN's fundamental building blocks are convolutional layers. Contrary to the other layers, this layer is in charge of extracting features from the input images, and as a result, it requires more processing power. The process of feature extraction is done by sliding a window of a fixed size over the input image pixel by pixel.In case of RGB images or in other words for an image with more than one channel, the convolution process is done channel by channel.

Pooling Layer

The major purpose of using max pooling layers in a CNN model is to minimize the number of parameters, which significantly decreases the computing power.Hence, pooling layers are generally paired with convolution layers.

Fully Connected Layer

It is known as the output layer of the model.It takes input from the preceding layer and in return computes an uni directional array as an output which has the size as the number of classes.The above process is carried out with the help of matrix multiplication.

D. Model

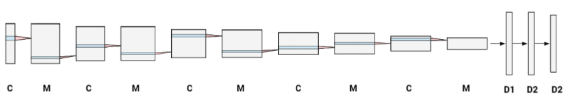

Fig. 2. Architecture of our proposed model

V. COMPARATIVE STUDY

NAME	ACCURACY(%)	LOSS(%)	AUC(%)	PRECISION	RECALL
XCEPTION	95.4	98.17	29.39	96.51	96.51
VGG16	93.02	98.55	98.44	96.67	96.67
VGG19	98.52	98.44	65.41	97.46	97.46
RESNET50	97.04	98.8	19.89	95.56	95.56
RESNET101	98.52	99.24	15.49	88.69	87.14
RESNET152	97.99	99.32	10.46	84.73	83.65
DENSENET121	95.03	98.16	23.01	92.98	92.54
DENSENET169	94.92	97.54	41.83	95.05	94.44
DENSENET201	97.25	99.11	14.98	96.51	96.51
INCEPTIONV3	91.11	97.19	36.85	87.94	87.94
INCEPTION RESNETV2	82.33	88.65	67.14	82.56	82.56

Table. 1. Table stating various performance measures and their values for all the models used for this comparative study

NAME	ACCURACY(%)	AUC(%)	LOSS (%)	PRECISION(%)	RECALL(%)
1 CONVOLUTION 1 MAX POOLING	74.07	73.81	91.97	61.11	61.11
2 CONVOLUTION 2 MAX POOLING	85.24	90.11	1.8028	84.94	84.13
3 CONVOLUTION 3 MAX POOLING	89.74	87.44	18.70	94.91	94.76
4 CONVOLUTION 4 MAX POOLING	94.92	98.53	23.05	92.38	92.38
5 CONVOLUTION 5 MAX POOLING	96.56	98.87	18.70	94.91	94.76

Table. 2. Table stating various convolution and max pooling layers affect the performance metrics

Xception

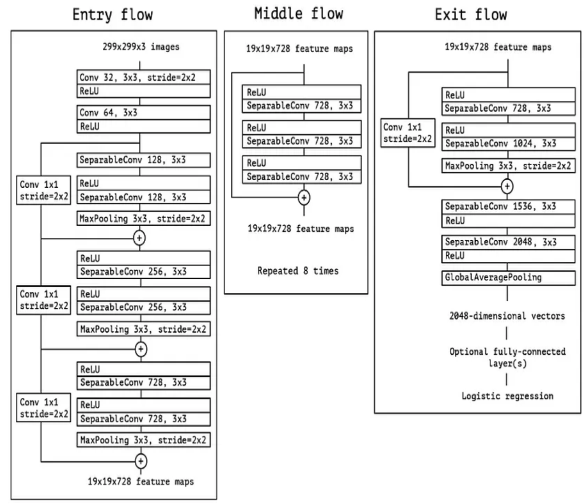

Fig. 3. Architecture of Xception [8]

Xception follows deep CNN architecture that consists of depthwise separable convolutions layers. Early modules in convolutional networks were introduced by Google as steps between regular convolution and depth segmentation convolution operations (deep convolution followed by point convolution).Hence, a deeply separable convolution can be viewed as an initial module with multiple layers. Based on this observation, they developed a new deep convolutional neural network architecture with a deeply separated convolutional module in place of the Inception one.[9]

Vgg16

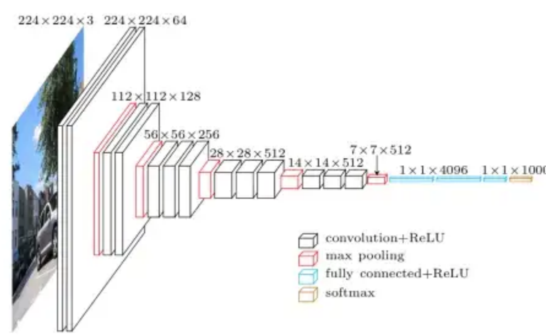

Fig. 4. Architecture of Vgg16 [10]

VGG16 is known for image classification and object detection algorithm. With approx 92% accuracy, VGG16 algorithm detects and classifies 1000 images into respective(1000) categories[11]. VGG16 is another type of Convolutional Network which is currently considered to be the best model. This architecture is achieved by evaluating the prior CNN model and adding up the layers with 3x3 convolution filter that showed a significant improvement. Approximately 138 trainable parameters have been introduced by increasing the depth to 16–19 weight layers.

Vgg19

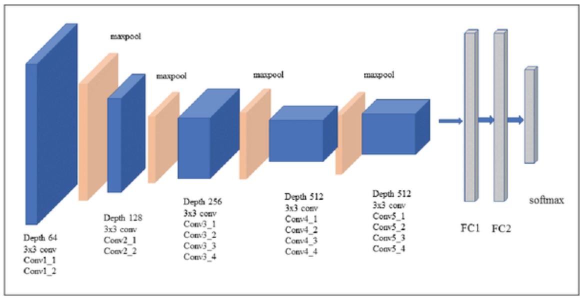

Fig. 5. Architecture of Vgg19 [12]

The VGG19 is an advanced CNN with layers that are pre-trained and understands what makes up an image in terms of shape, color, and structure. It is trained on millions of images[13]. Although VGG was introduced for ILSVRC, but still it can be modified and used for other classification purposes as well or for face recognition tasks i.e. the transfer learning. We can use keras framework to use and customize the weights of VGG-19[14].

Resnet50

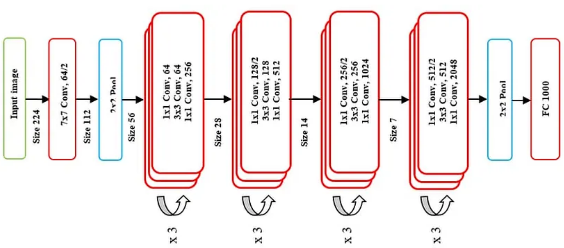

Fig. 6. Architecture of Resnet50 [15]

ResNet50 is a fifty layer residual network proposed by He et al [16], consisting of 48 convolutional layers, 1 MaxPool layer and 1 intermediate pool layer. A residual network is a type of an ANN that overlays blocks of residuals to form a network. The 50-level ResNet uses a building block bottleneck design.
The bottlenecked residual block reduces the number of parameter and matrix multiplications using 1×1 convolution known as the "bottleneck". This uses a stack of 3 layers instead of 2, so each layer can be trained much faster [17]. As in,
VGG network, the dimensions of the convolutional layer is 3x3 filter, the size of the input to this model is fixed at 224x224, and it follows some simple constructs: Output for layers with the equal number of filters. The number of filters becomes two times by halving the folded output so that the complexity of each layer is maintained.
The architecture has intermediate pooling layer and a thousand path fully connected layer with softmax. This model has fewer filters and is less complex than the VGG networks[18].

Resnet101

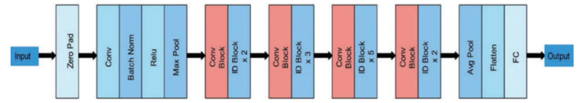

Fig. 7. Architecture of Resnet101[26]

ResNet101 has 104 convolutional layers that are divided into 33 layers-per-block. This model was primarily trained using the ImageNet dataset[31]. This diagram shows the flow right from the input images to residual blocks and the final output.

Resnet152

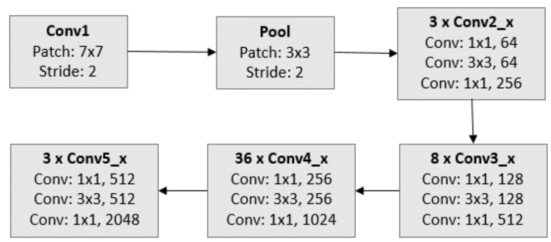

Fig. 8. Architecture of Resnet152 [33]

ResNet-152 is CNN architecture that was given by Microsoft Research Asia in 2015. It consists of 152 layers that is stated in paper "Deep Residual Learning for Image Recognition" by He et al[33]. This network uses residual blocks, that skip some layers, and enabling the network to better propagate gradients during training. Resnet introduces a residual learning unit structure to slow down deep neural network deterioration. The advantage of the unit is gives better accuracy without making the model complex [33].

Densenet121

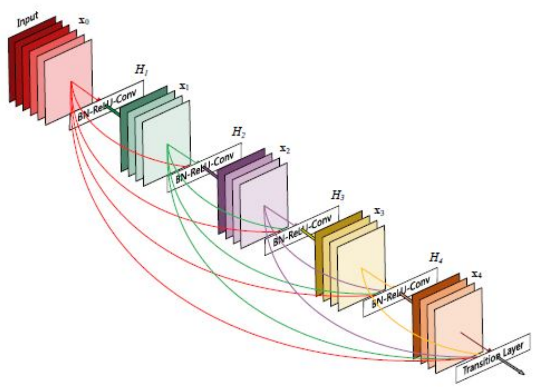

Fig. 9. Architecture of Densenet121 [20]

DenseNet is a model in which every layer is connected to next layers which are deeper in the network i.e the first layer with second, third, and so on, and then the same is repeated with the second layer. This method enables the highest possible flow of features among the layers[21]. DenseNet is specifically designed to improve accuracy degradation due to vanishing gradients in high-level neural networks. In simpler terms, information that is sent is lost before it reaches its destination and the reason is the longer path between the input and output layers [22]..

Densenet169

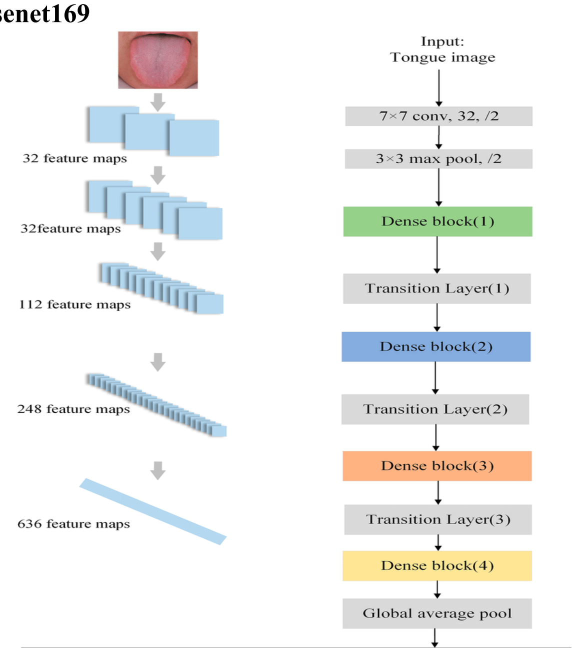

Fig. 10. Architecture of Densenet169 [24]

The DenseNet group of models, which are intended to do image classification, includes the Densenet-169 model. The size and accuracy of the densenet-169 model are the most obvious differences as compared to densenet-121 model. This model is significantly more storage intensive by about 55MB in size whereas the Densenet-121 model tajes up roughly 31MB size. It is primarily trained on Torch, then it was converted into Caffe format. All the models under the DenseNet umbrella have been pretrained on the ImageNet image database.[25].

Densenet201

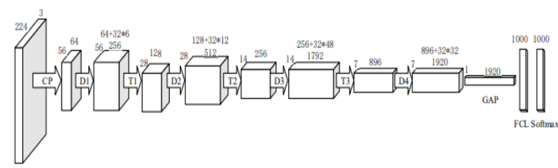

Fig. 11. Architecture of Densenet201 [28]

The Densenet-201 model too is designed to perform image classification. DenseNet-201 is a deeper version of the DenseNet architecture compared to DenseNet-121 and DenseNet-169. The main difference is the number of layers, with DenseNet-201 having 201 layers, compared to 121 and 169 layers in the other models. It also uses a higher growth rate than the other models, meaning that each layer in the dense blocks produces more feature maps, which can be reused and accuracy can further be increased. Originally trained in Torch, the authors converted it to Caffe format.

InceptionV3

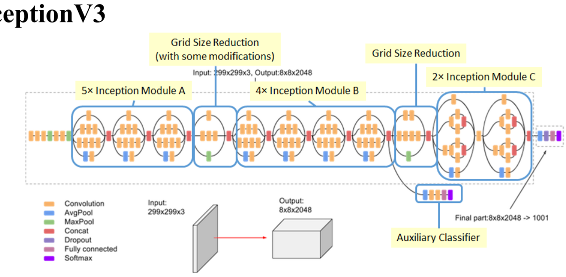

Fig. 12. Architecture of InceptionV3 [19]

Inception V3 is the third installment in a series of Inception frameworks which consists of Inception V1 And Inception V2.The first version of inception v3 model was trained on ImageNet dataset, which has one million images from 1001 classes. The purpose of Inception V3 model was for competing in ILSVRC (ImageNet Large Scale Visual Recognition Challenge), where it won first place. Inception V3 is just an improved and optimized interpretation of the Inception V1 model.[22] Several network optimization methods were used in the Inception V3 model to better fit the model.Improved efficiency and deeper network compared to v1 and v2. Also, V3 has less computational value and uses an additional classifier as a regularizer. Inception V3 is used as a parenthesis image to illustrate the parenthesis of ancient artifacts for each dynasty [27].

InceptionResNetV2

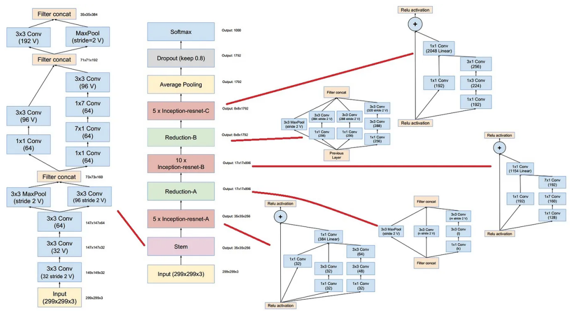

Fig. 13. Architecture of InceptionResNetV2 [29]

Inception-ResNet-V2 is type of CNN which is trained on a lot of images.This network can classify images into sequences of 1000 objects such as keyboards, mice, pencils and countless creatures because of the network's 164 layers of architecture.As a result, the network learned extended dot representations for various images. The web has input images that are 299X299. It is formed with the combination of structure and residual connections.In this network block, convolutional contaminants of different types are merged with residual links. The operation of the residual connection reduces training time and also solves the declination problem caused with deep structures

VI. DISCUSSION

The best accuracy achieved out of the pre-trained models with the Vgg19 model and the worst with the InceptionResnetV2 model.One of the reasons for difference between the accuracy achieved with these models may be due the different structure of architectures.The Vgg19 model contains 19 layers which includes 16 convolution layers and 5 max pooling layers whereas in InceptionResnetV2 which contains 164 layers and more amount convolution and max pooling layers than vgg19.This shows that the architecture of vgg19 which consists of 19 layers (16 convolution , 5 max pooling) is optimal for the given dataset.

VII. CONCLUSION

This study proposes a real- time pothole discovery system. This system uses Convolutional Neural Network to classify the image into potholes, unpaved, normal road and position of the pothole is brought. All these inputs are maintained in a database and potholes data is transferred to separate authorities of that particular area using automation. The proposed approach can help road conservation authorities to formulate rapid-fire and optimized conduct for road structure repairs. A more sophisticated result with the help of the global position system(GPS) can describe and point out the position of pavement failures. Future Work of Pothole Discovery includes making Android operation and syncing data using API. Another important point is Google Map integration to incorporate potholes in the route. Getting video feeds from road cams. This work can further be extended to describe other pavement torments, road depressions, classify roads as per quality, and depth estimation of potholes.

References

1. Sukhad Anand ,Gunjan Chugh, Divya Bansal and Sanjeev Sofat, “Road Condition Detection Using Smartphone Sensor: A Survey”, International Journal of Electronic and Electrical Engineering
2. Jin Lin and Yayu Liu Ajit Danti, Jyoti Y. Kulkarni, and P. S. Hiremath, “An Image Processing Approach to Detect Lanes, Potholes and recognize road Signs in Indian Roads”, International Journal of Modeling and Optimization, Vol. 2, No. 6, December2012
3. Kiran Kumar Vupparaboina and Roopak R. Tamboli "Overtaking & receding vehicle detection for driver assistance and naturalistic driving studies." In Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, pp. 697-702. IEEE, 2014
4. Zhajian Li I. Kolmanovsky, E. Atkins, J. Lu, D. Filev, and J. Michelini, “Cloud aided safety- based route planning,” in 2014 IEEE International Conference on Systems,Man, and Cybernetics, pp. 2525–2530, October2014
5. I. Moazzam, K. Kamal, S. Mathavan, S. Usman, and M. Rahman, “Metrology and visualization of potholes using the microsoft Kinect sensor,” in Proc. 16th Int. IEEE Conf. Intell.Transp. Syst., Oct. 2013,pp. 1284–1291
6. Sachin Bharadwaj Sundra Murthy and Golla Varaprasad, “Automatic Detection and Notification of Potholes and Humps on Roads to Aid Drivers”, IEEE Sensors J., Vol. 15, No. 8, August 2015
7. S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186
8. https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568
9. https://maelfabien.github.io/deeplearning/xception/
10. https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c#:~:text=VGG16%20is%20a%20convolution%20neural,vision%20model%20architecture%20till%20date
11. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918
12. https://blog.techcraft.org/vgg-19-convolutional-neural-network
13. https://medium.com/mlearning-ai/image-detection-using-convolutional-neural-networks-89c9e21fffa3
14. https://iq.opengenus.org/vgg19-architecture/
15. https://ijeecs.iaescore.com/index.php/IJEECS/article/view/20562
16. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016
17. https://datagen.tech/guides/computer-vision/resnet-50/#:~:text=ResNet%2D50%20is%20a%2050,networks%20by%20stacking%20residual%20blocks
18. Ridha Ilyas Bendjillali, Mohammed Beladgham, Khaled Merit, Abdelmalik Taleb-Ahmed, “Illumination-robust face recognition based on deep convolutional neural networks architectures”, Indonesian Journal of Electrical Engineering and Computer Science Vol. 18, No. 2, May 2020
19. https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c
20. https://iq.opengenus.org/architecture-of-densenet121/#:~:text=In%20short%2C%20DenseNet%2D121%20has%20120%20Convolutions%20and%204%20AvgPool,use%20features%20extracted%20early%20on

21. https://towardsdatascience.com/creating-densenet-121-with-tensorflow-edbc08a956d8#:~:text=DenseNet%20is%20a%20convolutional%20neural,4th%2C%205th%20and%20so%20on.
22. https://www.intel.com/content/www/us/en/developer/articles/technical/inception-v3-deep-convolutional-architecture-for-classifying-acute-myeloidlymphoblastic.html
23. https://www.pluralsight.com/guides/introduction-to-densenet-with-tensorflow
24. Tao jiang, xiaojuan hu, xinghua yao, liping tu, jingbin huang, xuxiang ma, ji cui, qingfeng wu and jiatuo xu, "Tongue image quality assessment based onadeep convolutional neural network" in Springer.
25. https://docs.openvino.ai/2021.3/omz_models_model_densenet_169.html
26. Awais Khan, Muhammad Attique Khan, Muhammad Younus Javed, Majed Alhaisoni,Usman Tariq, Seifedine Kadry, Jung-In Choi, Yunyoung Nam, "Human Gait Recognition Using Deep Learning and Improved Ant ColonyOptimization", CMC, 2022, vol.70, no.2
27. https://iq.opengenus.org/inception-v3-model-architecture/
28. Shui-Hua Wang, Yu-Dong Zhang, "DenseNet-201 based deep neural network with composite learning factor and precomputation for multiple sclerosis classification", School of Architecture Building and Civil engineering, Loughborough University, Loughborough, LE11 3TU, UK, School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000 Henan, P R China.
29. https://medium.com/the-owl/building-inception-resnet-v2-in-keras-from-scratch-a3546c4d93f0
30. Y. Rao, L. He and J. Zhu, "A residual convolutional neural network for pan-shaprening," in 2017 Int. Workshop on Remote Sensing with Intelligent Processing, Toranto, Canada, pp. 1–4, 2017.
31. https://medium.com/@zahraelhamraoui1997/inceptionresnetv2-simple-introduction-9a2000edcdb6
32. Long D. Nguyen, Dongyun Lin, Zhiping Lin, "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation", School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore.