Field Overview of Object Detection and Classification in the Field of Radiography

Robin Sulg
Univeristy of Tartu, Estonia
November 23, 2021

Abstract— **The aim of this report is to provide a field overview of object detection and classification in the field of radiography. The overview consists of analyzing papers that have used various computer vision and machine learning techniques in detecting medical anomalies from X-ray and magnetic resonance images. This report focuses mostly on knee health, however other medical conditions are covered as well.**

I. INTRODUCTION

The nature of many medical diagnoses is that they tend to worsen as time goes on. Furthermore, if the medical condition is left untreated it may lead way to new health concerns. Early detection of these anomalies may provide a larger option pool of treatments or prevent a serious medical issue stemming from leaving the underlying medical condition untreated. For example, knee osteoarthritis (OA), which is a degenerative type of arthritis, is a significant reason for disability in people [1]. To make matters worse, there are limited treatment options for late-stage osteoarthritis. Thus, detecting osteoarthritis early is critical to the patient's long-term wellbeing [1]. Computer vision (CV) has been applied in several fields and has proven its efficiency. For example, it is well known that CV has played an important role in the development of the automotive industry. By detecting traffic signs, lanes, etc. many modern cars have become safer and more autonomous. Many of the algorithms behind the successes in the automotive industry use techniques from the field of machine learning (ML). That is why it is critical to try and apply CV and ML in the field of medicine. This paper focuses on object detection and classification problems in the field of radiography and reviews five papers that have applied CV and/or ML for detection and classification of various medical conditions from X-ray and magnetic resonance images. The objective is to demonstrate that by applying the forementioned techniques it is possible to automate the diagnoses process of many health conditions. Furthermore, by doing so it would lead to an increased throughput of patients receiving their diagnoses in time as well as provide a second opinion that is to be considered next to the opinion of the medical expert in that respective field.

II. USING DEEP LEARNING TO DETECT AND CLASSIFY KNEE OSTEOARTHRITIS

Three research papers on automating the detection of knee joints and classifying the severity of knee OA have had positive results using convolutional neural networks (CNN). A convolutional neural network is a deep learning neural network that is used for processing data such as images. These images can be represented as structured arrays. The convolutional layers of the network can detect patterns (features) like lines and gradients which has made it a state-of-the-art technology in computer vision for the problem of image classification [6].

A. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative

The research article "**An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative**" by Yifan Wang, Xianan Wang, Tianning Gao, Le Du and Wei Liu claim to have reached **95%** accuracy on detecting the region of interest (ROI), that is the knee joints, and **69.19%** accuracy when classifying the severity of knee OA based on the Kellgren–Lawrence (KL) grading system. The KL grading system has the lowest grade of zero, meaning no knee OA to all the way up to four - indicating severe knee OA [1].

The approach presented in the paper uses four steps to assess the severity of knee OA:

1) An object detection CNN extracts the bounding boxes of the knees from the X-ray images.
   In addition, some of the detected objects are filtered out as their quality is not sufficient. This is based on the confidence parameter and intersection over union parameter. Note that the used CNN in this step is a modified version of You Only Look Once (YOLO) [4].
2) A CNN backbone extracts the spatial feature maps for the cropped knee images. The CNN backbone is based on the ResNet50 architecture, where the last pooling and dense layers are removed [1].
3) The extracted spatial feature maps from the convolutional layers are flattened and recomposed as a sequence.
4) The visual transformer exploits the correlations between different local regions for the final classification task.
5) Finally, the visual transformer layers outputs features for the final classification task.

The first four steps of the proposed method are illustrated on figure 1.

The IOU score convergence is shown on figure 2.

1) Object detection convolutional neural network:

The CNN that is used for object detection is based on the popular YOLO object detection system. The researches annotated images from the Osteoarthritis Initiative (OAI) database, which is available for public access at [1]. They labeled 200 samples, which accounts for 4.43% of the dataset. With this data, they trained a modified version of the YOLO architecture. As mentioned above, the KL knee OA grading system has in total five states, ranging from zero to four. The bounding box regression from YOLO requires five outputs from the model per prediction per anchor [1]. Anchors are the prior knowledge of the object size (height, width) determined by clustering the sizes of all training objects [1]. Anchors allow one grid cell to detect multiple objects. Hence, the total output channels can be calculated as $B*(C+5)$ where B is the number of anchors and C is the number of object class labels. In the case of the proposed method the number of labels is one, as the only label to predict in the knee joint detection task is "knee joint". This means that we can determine the number of channels in the last layer with the following formula: $5*(1+5) = 30$.

As for the training, the authors faced two challenges:

1) the number of annotated samples is limited
2) the model validation is difficult as there is no ground truth for the testing data.

The researchers overcame this limitation by using transfer learning [8]. Transfer learning is the idea that if a model is trained on a large and general enough dataset, the model will effectively serve as a generic model of the visual world [9]. The initial weights, excepts for the final layer, were obtained from the pretraining on the COCO dataset [1]. During training they monitored intersection over the union (IOU) score defined by the formula:

$$IOU = \frac{A \cap B}{A \cup B}$$

where A denotes the predicted bounding box and B denotes the ground truth. When the IOU scores converge the training is terminated to avoid overfitting. With this method, the YOLO model predicts the bounding boxes for the rest of the 95.57% of images. In addition, the validation process had to be altered as well. The segmentation results are verified by statistical analysis. The analysis consists of four forms of measurements:

1) Detection count per image. All images that were used consist of two knees. Therefore, two detections per image are expected.
2) Detection size: knee size is similar in humans
3) Detection location: a pair of knee joints should be located at the same height and on both sides of the image.
4) Object confidence: since all of the images contain the knee joints, the model should give a high confidence score.

[1]http://www.oai.ucsf. edu/

2) Classification:

The next step following knee detection is classification. The region of interests (ROIs) are cropped according to the predicted bounding boxes from the original image. In addition, the cropped images are resized to $224*224$ and normalized by the mean and variance computed from the training data. The reason for resizing and normalization is to adjust the statistical properties of the knee images to match the data that is used for the pretraining. As mentioned above the classification task is from steps two to five, that were listed above. The CNN backbone that extracts the spatial features has the last pooling and dense layers removed. The output is then denoted as O^{C*H*W}, where C, H, W are the number of channels, height and width of the image. When given an image with dimensions of $224*224$, the output of the ResNet50 (CNN backbone) is $O^{1024*14*14}$. The spatial features are flattened and combined with a class label token and position embedding. The size of each feature map is $14*14$. This means that it corresponds to 196 local regions of the original image. The flattening reshapes the output into a matrix M that has the dimensions of $196*1024$. Each row in that matrix contains features from the same region. The next step is for the visual transformer the exploit the relationship of the features from different regions. The visual transformer block follows twelve parallel self-attention layers that are connected to the dense layer for the OA severity classification.

In figure 3 the classification result of the proposed method of the researches are compared to other models. We can see that the proposed method has the highest accuracy. On figure 4, the label distribution of the cropped bounding boxes are displayed.

B. Knee Osteoarthritis Classification Using 3D CNN and MRI

Another paper **„Knee Osteoarthritis Classification Using 3D CNN and MRI"** by Carmine Guida, Ming Zhang, and Juan Shan also tackled the issue of diagnosing knee osteoarthritis from magnetic resonance (MR) images and compared the results to a model that detected knee OA from X-ray images. It should be noted that researches used the same OAI public dataset used in the paper discussed above [1]. However, in this paper they were more focused on detecting either that there is some form of osteoarthritis or not. Furthermore, they still conducted experiments diagnosing knee OA from the Kellgren and Lawrence (KL) grade from zero up to four. To implement the 2-class classification solution, they divided the KL grade of severity in the following way: non-OA as everything equal or less than one and OA as everything equal or bigger than two. The reasoning behind it is that, clinically, diagnosing a patient with knee OA is the ultimate goal of assigning a KL grade [7]. As for the results, the 3D model with MR images

Fig. 1. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative, Proposed Method[1]

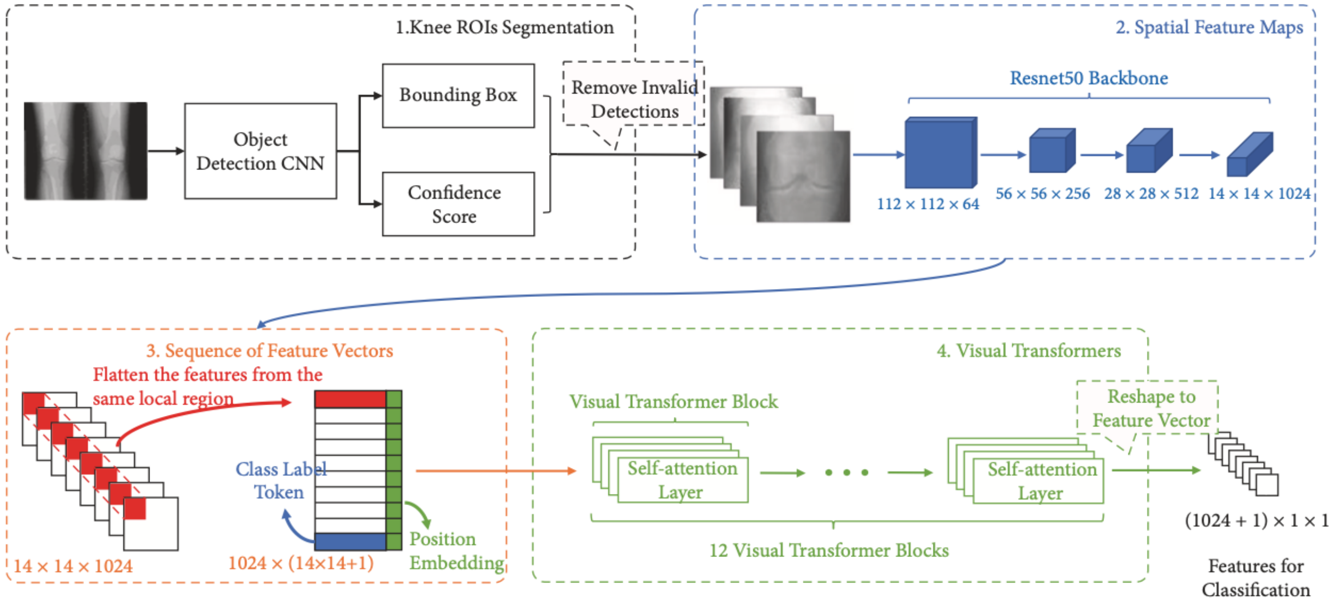


Fig. 2. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative, IOU score [1]

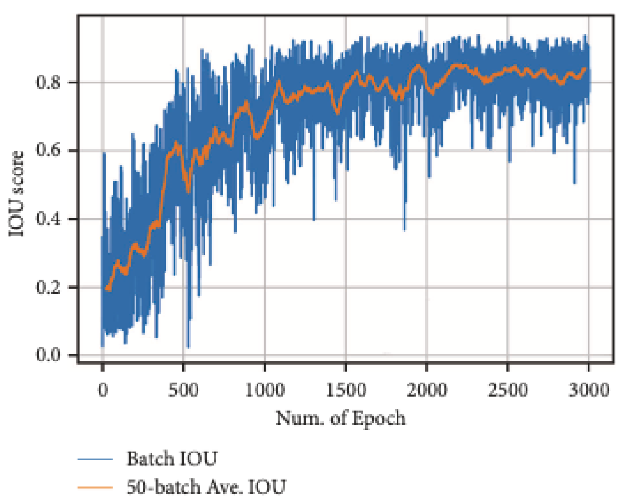


Fig. 3. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative, Classification results [1]

Model	Accuracy (%)
VGG-19 [25]	53.4
ResNet50	66.68
ResNet101	66.70
Ordinal loss (ResNet50) [22]	66.2
Ordinal loss (ResNet101) [22]	65.5
Siamese net [15]	66.71
The proposed method	69.18

achieved higher accuracy in both the 5-category classification and the 2-category classification problem. In the case of the 2-category classification, the accuracy for the 3D model was **83%** compared to the 2D model using X-ray images that achieved the accuracy of **77%**. When comparing the accuracies of the 5-category classification problem, then the 3D model reported **54%** accuracy and the 2D model **50%**.

Fig. 4. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative, Label distribution of the cropped images [1]

	KL-0	KL-1	KL-2	KL-3	KL-4
Total ROI	3234	1475	2186	1141	266
Training set	2587	1180	1749	913	213
Validation set	647	295	437	228	53

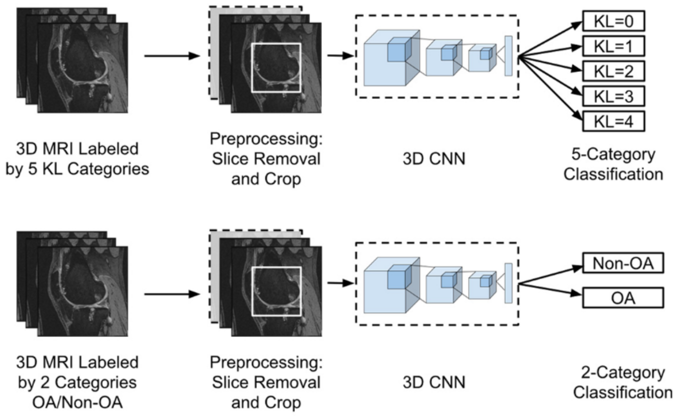
The results from this paper indicate that an application that utilizes MR images has greater potential to improve the diagnosis accuracy for knee OA clinically than the 2D model. Keep in mind that in the paper discussed above [1], the authors achieved a greater accuracy using the 2D model than reported in this study. In fact the accuracy for the 5-category classification problem was higher for both the 3D and 2D models presented in this paper. **69.19%** vs. **54%** and **50%**.

1) 3D CNN Method Overview:

„Knee MR imaging produces a 3D representation of the knee joint, utilizing a sequence of 2D images taken laterally across the knee. Given the 3D nature of MR images, 3D CNN can be advantageous in evaluating the whole sequence of images as one unit. Through the implementation of 3D kernels, information from adjacent slices could be integrated. Therefore, 3D features that may not be detectable using 2D CNN could be potentially captured„ [7]. In this study, the researchers built a machine-learning model capable of analyzing sequences of MR images. The output of the model, as noted above, is a class ranging from zero to four or in the case of the 2-classification problem non-OA (0) and OA (1). The overview of both of the models can be found on the figure 5.

In total, there are three stages in the proposed 3D CNN. The first stage begins with convolutional layer containing 32 kernels of size $7 * 7 * 7$ with a stride of $2 * 2 * 2$. This

Fig. 5. Knee Osteoarthritis Classification Using 3D CNN and MRI, Overview of models [7]

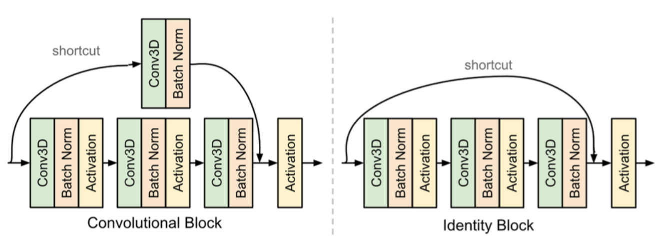


is followed by batch normalization and a ReLU activation layer. To reduce dimensionality, a max pooling layer is used. Finally, a dropout layer is added to help the model to avoid overfitting. Every dropout layer uses the dropout rate of 0.5 meaning that each node has a 50% chance that it is set to 0.

The second stage consists of a sequence of six residual blocks. Each such block features a shortcut connection from the input to output. There are two types of such feature blocks: the convolutional block and an identity block. These blocks are present on figure 6. The convolutional block uses a convolutional layer in the shortcut path. This layer was used when the input dimensions were changed. The identity block however did not have any layers in the shortcut and was used when the input and output dimensions matched.

The third and the final stage of the model uses a global max pooling, followed by a fully connected layer and a dropout layer. The last layer is a softmax function that gives us the probabilities that the input belongs to any given class.

Fig. 6. Knee Osteoarthritis Classification Using 3D CNN and MRI, Convolutional Block and Identity block [7]



2) The 2D CNN:

Lastly, I want to discuss the results that the authors of this paper obtained with the 2D model. The researchers compared many architectures to solve the knee OA diagnoses problem. Based on their experiments they concluded that Inception-ResNetV2 performed the best. The results are presented on the table in figure 7.

Fig. 7. Knee Osteoarthritis Classification Using 3D CNN and MRI, Performance of 2D CNN architecture using X-ray data [7]

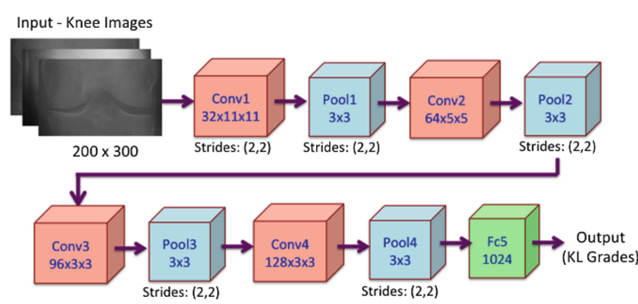
Architecture	5-Category Accuracy	2-Category Accuracy
VGG16	20.0%	60.0%
VGG19	28.0%	60.0%
ResNet50V2	49.5%	69.0%
ResNet101V2	42.5%	77.5%
ResNet152V2	20.0%	60.5%
InceptionV3	54.0%	80.5%
InceptionResNetV2	55.5%	80.0%
DenseNet121	55.0%	70.0%
DenseNet169	45.0%	78.0%
DenseNet201	45.0%	81.0%

C. Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks

In the paper, "**Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks**" by Joseph Antony, Kevin McGuinness, Kieran Moran and Noel E. O'Connor present how to automatically localize the knee joints and classify the knee OA. The localization of the knee joints is based on a fully convolutional neural network (FCN). To automatically classify the knee joint they propose either to train a CNN from scratch for multi-class classification or to train a CNN to optimize a weighted ratio of the loss functions: categorical cross-entropy and mean-squared error. The researchers found that detecting the knee joint with FCN is highly accurate (**99.6%**). The CNN implemented from scratch gave better results than the multi-purpose image classifier WND-CHARM and reached the accuracy of **60.3%** [2].

The architecture of the implemented CNN can be seen on the figure 8:

Fig. 8. Architecture of the Implemented CNN [2]

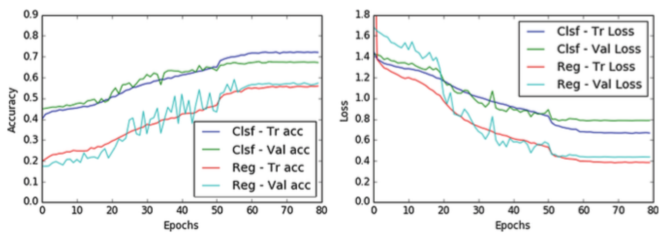


In addition, they also compared the classification results when knee joint localization was automated versus when the localization was manual. The classification accuracies are comparable and presented in the figure 9.

D. Conclusion: Knee Osteoarthritis Diagnoses Using CNN

In this chapter I have reviewed three studies that have proposed two different approaches in detecting knee osteoarthritis from both MR and X-ray images. The papers reported positive results in utilizing computer vision and

Fig. 9. Manual vs. Automatic Localization [2]



(a) Accuracy Curves. (b) Loss Curves.

Fig. 10. Manual vs. Automatic Localization Results [2]

Method	Classification-Acc	Classification-MSE	Regression-MSE
Manual localization	**63.6%**	**0.706**	**0.503**
Automatic localization	61.9%	0.781	0.541

machine learning techniques in diagnosing knee osteoarthritis. The best result for 5-category knee osteoarthritis was obtained by the study “**An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative**” [1] that reported the accuracy of **69.19%**. However, good results were also reported from **„Knee Osteoarthritis Classification Using 3D CNN and MRI“** [7] that proposed using MR images instead of X-ray images. The authors reported **83%** accuracy on 2-category knee osteoarthritis diagnoses and **54%** on 5-category diagnoses. This implies that:

1) Diagnosing simply either knee osteoarthritis is present or not can increase the accuracies obtained by papers that focused on 5-category classification. Furthermore, this is clinically more significant.
2) 3D models using MR images show promise of improvement in knee osteoarthritis diagnoses over using 2D models

III. Neural Network Using Chest X-Rays

Another field where neural networks can be used is to detect various lung diseases. One such paper that demonstrated that is “**Pneumonia Detection from Chest X-Rays using Neural Networks**” by Ashish Jain, Raman Vashisth, Ravi Karan Yadav, Unnati Ahuja. The researchers used a public Kaggle dataset of X-ray images that they pre-processed. The processed images were input to a CNN that consisted of a series of convolutional layers, filters and a fully connected layer and ends with the Sigmoid activation function to get a result ranging from zero to one. Zero indicating that there is an inflammation and one indicating then everything is normal. They claim that the accuracy of their CNN is **91.5%** of detecting pneumonia from the X-ray images [3].

IV. Caries Detection in Panoramic Dental X-ray Images

Finally, a thesis "**Caries Detection in Panoramic Dental X-ray Image**" by João Paulo Ribeiro de Oliveira carried out detecting dental caries from X-ray images taken of the teeth. He has provided results with many different methods under different circumstances. One interesting finding is that the result with perfect tooth segmentation (PTS) yields much higher accuracies than with imperfect tooth segmentation (IPTS) that is deduced in the thesis. This suggests that by building a better automatic system to segment teeth, one could expect a fairly accurate automated product for detecting dental caries. The accuracies are reported below:

PTS Accuracies [5]:

1) Artificial Neural Network (ANN) 98.7%
2) SVM 97.54%
3) Naive Bayes 95.72%

IPTS Accuracies [5]:

1) ANN 65.39%
2) SVM 62.02%
3) Naive Bayes 56.83%

V. Conclusion

The aim of this report was to review how computer vision and machine learning techniques have been applied in the field of medicine to detect and classify health conditions. To further narrow the subject, only research papers in the field of radiography were considered. In this paper I have reviewed five articles focusing on detecting medical conditions from X-ray and MR images. The conditions diagnosed were knee osteoarthritis, pneumonia and dental caries.

In this report, literature focusing on detecting knee osteoarthritis was most prevalent. Therefore, this study focuses on three papers on knee osteoarthritis and one for both pneumonia and dental caries. From the papers reviewed it is apparent that knee osteoarthritis is benefitting the most from computer vision and machine learning techniques. The best result was obtained by deploying a 2D CNN for 5-category knee osteoarthritis detection and 3D CNN for 2-category knee osteoarthritis detection with accuracies of **69.19%** and **83%**.

In conclusion, this report gave an overview of some of the many possibilities of object detection and classification using computer vision and machine learning in the field of radiography. Doing so, this paper has achieved the goal of understanding the current state of computer vision and machine learning in medicine. The techniques presented in this paper can be used to build upon on and hopefully stem new ideas how to further improve medical diagnoses in the field of radiography by using object detection and classification.

References

[1] Yifan Wang, Xianan Wang, Tianning Gao, Le Du, Wei Liu, "An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative", Journal of Healthcare Engineering, vol. 2021, Article ID 5586529, 10 pages, 2021. https://doi.org/10.1155/2021/5586529.

[2] Antony J., McGuinness K., Moran K., O’Connor N.E. (2017) Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks. In: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2017. Lecture Notes in Computer Science, vol 10358. Springer, Cham. https://doi.org/10.1007/978-3-319-62416-7_27.

[3] Ashish Jain, Raman Vashisth, Ravi Karan Yadav, Unnati Ahuja, Pneumonia Detection from Chest X-Rays using Neural Networks. http://dx.doi.org/10.22214/ijraset.2021.36489.
[4] YOLO. https://pjreddie.com/darknet/yolo/.
[5] João Paulo Ribeiro de Oliveira, "Caries Detection in Panoramic Dental X-ray Image". 2008. http://www.di.ubi.pt/ hugomcp/doc/joao_oliveira_2008.pdf
[6] Thomas Wood, "Convolutional Neural Network". https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network.
[7] Guida, Carmine Shan, Juan Zhang, Ming. (2021). Knee Osteoarthritis Classification Using 3D CNN and MRI. Applied Sciences. 11. 5196. 10.3390/app11115196.
[8] "Transfer Learning". https://deepai.org/machine-learning-glossary-and-terms/transfer-learning
[9] "Transfer learning and fine-tuning". https://www.tensorflow.org/tutorials/images/transfer_learning