Editions Ishango - UAC
Revue Internationale Multidisciplinaire Etincelle

Article

Classification of Human Faces with or without Mask Based on the CNN Model in the Context of the COVID-19 Pandemic

Alfred Kambale Syatsukwa[1,*], Echello Katya Muhambya[2] & Noëlla Kavira Waziwazi[3]

1 Faculty of Sciences Appliquées, Université de l'Assomption au Congo, Butembo, P.O. Box 104, The Democratic Republic of Congo
2 Faculty of Computer applications, Ganpat University, Ganpat Vidyanagar, Mehsana-Gandhinagar Highway, P.O. 384012, North Gujarat, India
3 Faculty of Sciences économiques et de gestion, Université de l'Assomption au Congo, Butembo P.O. Box 104, The Democratic Republic of Congo

* Corresponding author: alfredsyatsukwa@gmail.com

Abstract: Even though vision for humans is intuitive, computer vision has been a major problem for several decades. Fortunately, the technological developments of recent years are contributing to the rise of Artificial Intelligence and in particular Deep Learning, which has risen since the 2010s. Human's face analysis includes a variety of specific problems such as face detection, facial recognition, gender recognition, age estimation to name just the most common. Significant research efforts have been devoted to deep learning problems. The explosion of the Deep Learning paradigm, which determines a dramatic increase in performance, is now well known to the public; therefore, the number of approaches based on deep learning increases impressively. In the current context of the COVID-19 pandemic where wearing a mask is recommended in public places especially, this paper aims to implement a convolutional neural network model capable of classifying faces with or without masks.

Citation: Syatsukwa, A.K.; Muhambya, E.K.; Waziwazi N.K. Classification of human faces with or without mask based on the CNN model in the context of the COVID-19 pandemic. *Etincelle*, 2022, Vol. 24, no. 1. https://doi.org/10.61532/rime241113

Received: 03/03/2022
Accepted: 27/07/2022
Published: 15/10/2022

Publisher's note: Ishango-uac remains neutral in terms of juris-dictional revendications in published geographic maps and the authors' institutional affiliations.

Keywords: Faces classification, Deep Learning, Convolutional Neural Network, Masks, COVID-19.

1. Introduction

Actually, there is unprecedented progress in image classification, thanks to advances in artificial intelligence research, and in particular in machine learning and deep learning. This progress is due to the large amount of research in the field, but also to the availability of huge international image databases, such as ImageNet (Deng, et al., 2009), which allow researchers to delve as deep as possible into the field. However, the world of face recognition still lacks large public datasets, and largely due to this factor, most of the community's recent progress remains limited to internet giants such as Facebook, Google, Baidu, etc. (Parkhi, et al., 2015). Face recognition is one of the most important tasks that has been extensively studied in the field of computer vision. Human faces provide largely better features for recognizing the identity of the person compared to other common biometric-based approaches such as iris and fingerprints (Parkhi, et al., 2015). Therefore, many recognition systems have used facial recognition (FR) features for forensic and security purposes. However, the performance of FR algorithms is negatively affected by the presence of facial disturbances such as occlusion and variations in illumination and facial expression(Alzu'bi, et al., 2019).

For this paper, different datasets (images of faces with masks and without masks) are built from scratch. But with the lack of images of masked and unmasked human faces but also the heaviness of the task, the technique of increasing data was used to expand the datasets. This paper has used the TensorFlow and Keras platforms. TensorFlow is an open-source library developed by the Google Brain team to conduct research on machine learning and deep learning. It implements machine learning methods based on the principle of deep neural networks (deep learning) (Moolayil, 2019). Keras is an open-source library written in python (Moolayil, 2019). Python programming language were chosen as programming language to implement the model. To evaluate the performance of the classification model, the confusion matrix was used.

The following sections of this manuscript are: 2) Literature review, where we are understanding some concepts such as Deep learning and illustrating some related work; 3) Methodology. In this section we present all the used materials to realize this study: 4) We illustrate the implementation process and results of the study and discussion; 5) Conclusion and recommendations.

2. Literature review

2.1. Theoretical review

2.1.1. Deep learning

Deep learning is one of the techniques used in machine learning (Mishra, et al., 2021). So, to understand deep learning, one must first understand machine learning. Machine learning is a set of tools and computer algorithms that automate the construction of a prediction function *f* from a set of observations called the learning set (Pirmin, et al., 2015). Unlike classical programming, Machine learning is the discipline giving computers the ability to learn without them being explicitly programmed. So, rather than manually writing code with a specific set of instructions to accomplish a specific task, the machine is trained using data and algorithms that give it the ability to perform the task without being explicitly told how to do it (Moolayil, 2019). Deep Learning, on the other hand, is a subfield of machine learning that uses algorithms inspired by the structure and function of neural networks in the brain. It is based on what has been called, by analogy, «artificial neural networks», the latter is a computer system composed of a set of connected units called neurons, organized into layers (Mpia, et al., 2022). The following is a simplistic representation of a basic neural network:

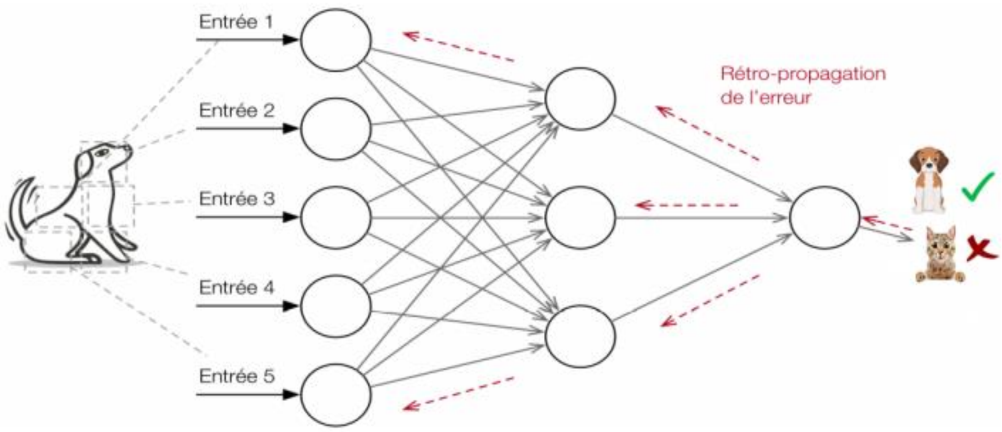

Figure 1. Simple neural network illustration (Mpia & Inipaivudu, 2021)

The results of a first layer of «neurons» serve as input for the calculations of a second layer and so on. The words «deep» or «deep» are used in reference to the number of layers of neurons that make up these networks: the greater the number of layers, the deeper the network and the more complex problems it can be treated (Argotic, 2018). There are three types of layers in every artificial neural network: Input Layer, Hidden Layer, and Output Layer. Different layers perform different types of transformations on their inputs. Data flows through the network starting with the input layer and passing through the hidden layers until the output layer is reached. This is called a direct pass through the network. The layers between the input and output layers are called hidden layers. Deep learning works on the principle of extracting features from the raw data by using multiple layers for identifying different aspects relevant to input data. Deep learning techniques include convolutional network, recurrent neural network, and deep neural network. Deep learning uses artificial neural network, especially the convolutional network. In the past, machine learning use was limited due to its inability to process the raw input data. Deep learning has helped in overcoming this limitation, as it has the ability to operate on large volumes of data and thus has been an effective and useful technique of machine learning. Deep learning has also picked up the pace due to the hardware advancements of computers. A deep experience in feature extraction was necessary to convert the raw data into a suitable form so that the subsystem of the machine can recognize and classify the raw data (Mishra, et al., 2021). The main difference between machine learning algorithms and deep learning is the stage of extracting features. For machine learning, the characteristics to be identified depend on human expertise. This practice is very difficult and requires a specialist in the field, sometimes the distinctive characteristics are not even identifiable by human. Deep learning makes it possible to overcome this problem by using several layers of neural network. The first layers will extract simple characteristics that the following layers will combine to form increasingly complex and abstract concepts.

2.1.2. Convolutional Neural Networks

Convolutional neural network (CNN) is currently the most efficient model for classifying images. The first versions were developed by Yann LeCun around the 1990s, it is inspired by the biological processes of the visual cortex of animals and make it possible to transform a global problem of recognition into a succession of steps easier to solve (Ketkar, 2017). CNNs are particularly useful for applications based on object recognition and computer vision. The layers present in a CNN are of the order of several tens or even several hundred layers. All of these layers perform operations that modify the data in order to learn the specific characteristics of that data. Thus, the architecture of the CNNs includes: An alternation of processing layers (layers of convolution, activation and simplification or pooling) that make it possible to extract characteristics. Then layers similar to that of a classic multi-layer perceptron (MLP) (Fully Connected layers) which, after flattening the «maps» from the extraction of characteristics, perform the final classification(Chen, et al., 2023).

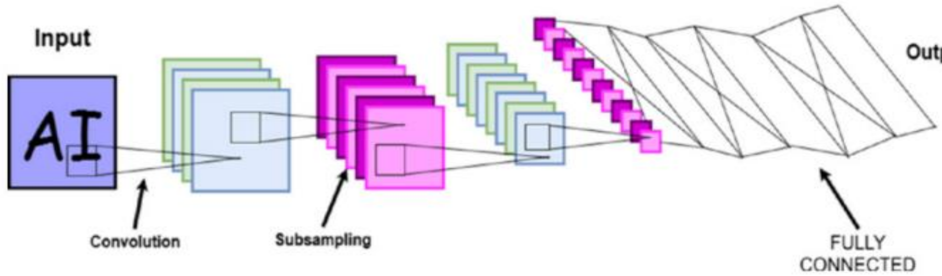

Figure 2. Simplified architecture of CNN (Ghosh, et al., 2020)

2.1.3. Confusion matrix

A Confusion Matrix is a better method for evaluating the performance of a classification model. It represents a summary of the results of predictions about a classification problem. Correct and incorrect predictions are highlighted and broken down by class. A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making. For a binary classification problem, we would have a 2 x 2 matrix as shown below with 4 values (Ivo & Günther, 2019).

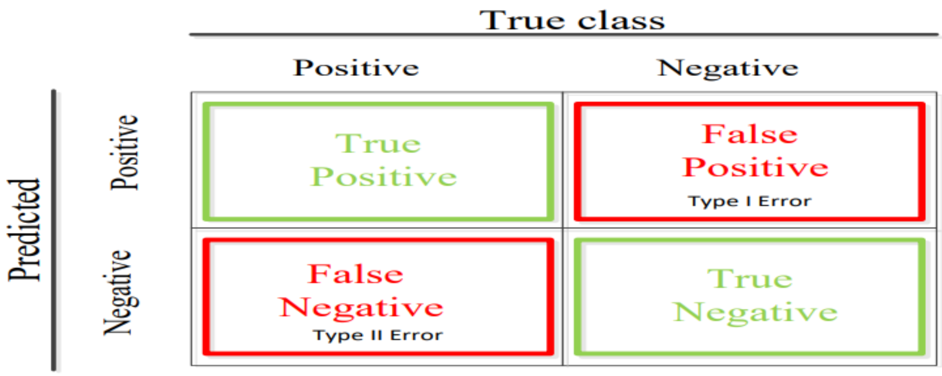

Figure 3. Main terminologies of the confusion matrix (Ivo & Günther, 2019)

Where True Positive (TP represents cases where the prediction is positive, and where the real value is actually positive. E.g. The model predicts that the face is masked and that in reality the face is indeed masked; True Negative (TN) is cases where the prediction is negative, and the real value is actually negative. E.g. The model predicts that the face is not masked and that in reality the face is not masked. False Positive (FP) stands for cases where the prediction is positive, but the actual value is negative. E.g. the model predicts that the face is masked when in reality the face is not masked. False Negative (FN) represents cases where the prediction is negative, but the real value is positive. E.g. The doctor tells you that you are not pregnant, but you are pregnant (Markoulidakis, et al., 2021).

2.2. Related works

There are many research papers about face recognition. This paper reviews some of them. The World Health Organization (WHO) has made the use of a face mask a mandatory biosafety measure, which raises the issue of facial recognition. This context motivated Talahua, et al. (2021) to propose a facial recognition system. Their study describes the development of a system for recognizing people from photographs, even when they are wearing a face mask. A classification model based on the MobileNet V2 architecture and the OpenCV face detector is used. The FaceNet model is used as a feature extractor and a feedforward multilayer perceptron to perform face recognition. To train the face recognition models, a set of 13,359 observations is generated; 52.9% images with a face mask and 47.1% images without a face mask. The experimental results show that there is an accuracy of 99.65% in determining whether a person is wearing a mask or not. An accuracy of 99.52% is achieved for face recognition of 10 people with masks, while an accuracy of 99.96% is achieved for face recognition without masks (Talahua, et al., 2021). The present work differs from the previous one in that it uses a CNN instead of MobileNet V2 and the OpenCV platform. The similarity is only in the problem statement. Both papers are interested in face recognition in the context of COVID-19.

The research of Alzu'bi, et al. (2019) aimed at implementing a face recognition (FR) to overcome the challenges of face masking on the existing security and authentication systems that already rely on FR. Their study has presented a comprehensive review of the recent work on masked face recognition (MFR) based on deep learning techniques. Their study has discussed the generic MFR pipeline that has been adopted in recent years and has identified the recent advances that have contributed to improving the performance of MFR methods. Many important issues that directly affect MFR systems have been discussed, including image pre-processing, feature extraction, face detection and localization, face unmasking and restoration, and identity matching and verification. In addition, some recent interesting and promising techniques have been presented, which are expected to motivate further research efforts to address the existing MFR challenges.

The study of Mbunge, et al. (2021) aimed at providing a comprehensive review of artificial intelligence models that have been used to detect face masks. The paper has revealed that deeper and wider deep learning architectures with increased training parameters, such as inception-v4, Mask R-CNN, Faster R-CNN, YOLOv3, Xception, and DenseNet are not yet implemented to detect face masks.

3. Methods and materials

3.1. Choice of development tools

Google Colaboratory, often shortened to «COLAB» which offered us free access to GPUs for training the model (Poornima, et al., 2021) was used. We used python as language to implement the proposed prototype. We used Keras version 2.13.1 and Tensorflow 1.13.0.

3.2. Data preparation

The image database contained 1000 images and was subdivided into two classes, images of faces with masks (500 images) and images without masks (500 images). We were in a case of supervised learning where the objective of the algorithm is to learn from annotated example data and create a general set of rules to match an input to an exact output. The following figure illustrates the sample of images with masks:

Figure 4. Sample of images with masks

The following figure illustrates the sample of used images without masks:

Figure 5. Sample of images without masks

3.3. Data augmentation

Data augmentation is a technique of creating new data based on changes to existing data (Chollet, 2018; Géron, 2019) to address the problem of data insufficiency but also to overcome the problem of overfitting. Essentially, new data is created by making reasonable changes to existing data. For example, we could increase the image data by flipping the images, horizontally or vertically. We could rotate the images, zoom in or out, crop or even vary the color of the images. These are all common data augmentation techniques (Mpia, et al., 2023). We set then the rotation_range parameter to 10, the width_shift_range parameter to 0.1, height_shift_range was set 0.1, the attribute shear_ range was given the value of 0.15, zoom_range parameter was 0.1, the channel_shift_range was set to 10, and horizontal_flip was set True.

Thus, from 500 images with masks and 500 images without masks, we were able to create 10,000 images (5000 with masks and 5000 without masks) more than enough for our model. The final dataset organization was illustrated in the following figure:

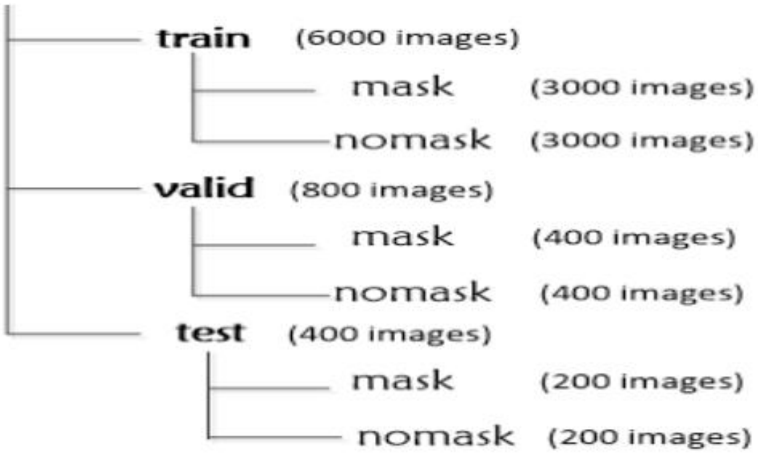

Figure 6. Dataset organization

3.4. *Proposed CNN model architecture*

We used Sequential() model to build the proposed CNN model. The proposed model had 4 convolutional layers (Conv2D). The first Conv2D was set with filters equal to 32, kernel_size of (3,3), ReLU as the activation function, padding was set to same and input_shape of (224, 224, 3). The scond Conv2D layer was set as the first one except the inprut_shape parameter which was excluded. The third and fourth Conv2D layers were set as the second. However, their filters parameters were set to 64 respectively. After the two first and the two last Conv2D, we added max pool layers (MaxPool2D) with both set as follows: pool_size = (2,2) and strides = 2. At the end of every MaxPool2D layer, we defined Dropout layers having a rate of 0.4 for all in other to address the overfitting problems (Mpia, et al., 2023).

Before starting the fully connected phase, we put a Flatten layer. The first fully connected layer had 125 unit and ReLU as function of activation. The output Dense layer contained 2 units and Softmax as activation function. Between the first Dense layer and the output Dense layers, we set a Fropout layer with a rate of 0.25. The model was trained with Adam as optimizer function, the learning_rate parameter was set to 0.0001. The authors used categorical_crossentropy as loss function to minimize the error. The metric used to evaluate the model were accuracy and confusion matrix and 20 epochs were considered.

4. Results and discussion

4.1. *Model evaluation results*

Figure 7 illustrates the performance results of the proposed CNN model. In the first epoch, the train loss was 2.0591 and validation loss was 0.4831. While validation accuracy scored 0.8188 and train accuracy reached 0.6890. At epoch 20, the validation loss and train loss decreased significantly to 0.2840 and 0.0551 respectively and validation and training accuracies increased exponentially 0.9250 and 0.9781 respectively.

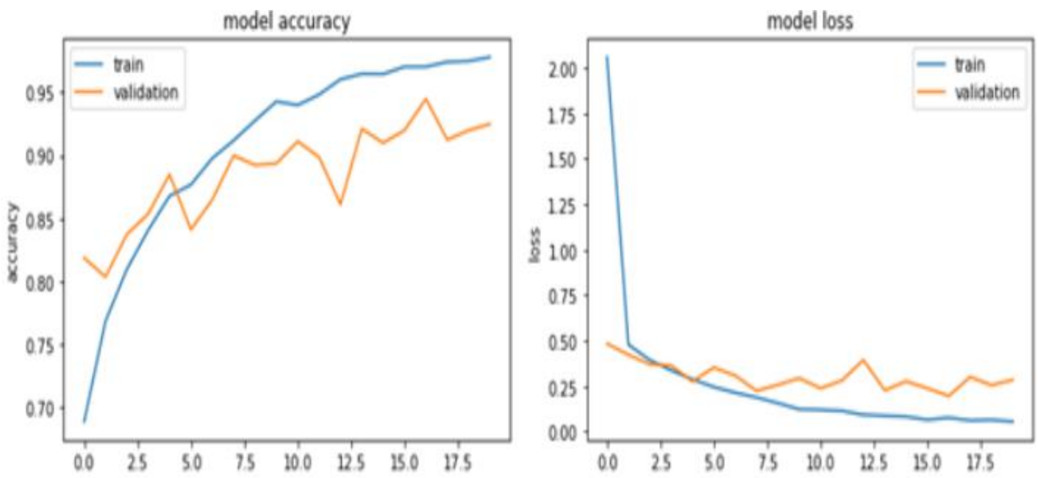

Figure 7. Results of the proposed CNN model validation and training for loss and accuracy

Beyond accuracy, the authors used also confusion matrix to evaluate the proposed model. It was observed in figure 8 that on a number of samples of 200 images with masks in our test dataset, the model predicted 192 accurately, and made the mistake on 8 samples. Even, out of 200 samples without masks, there are 180 correct predictions and 20 that are incorrect. In short, a validation accuracy of 93%.

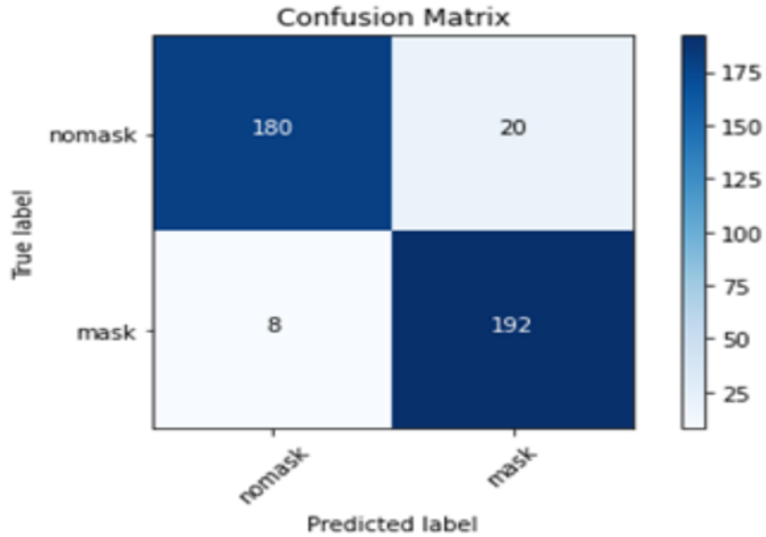

Figure 8. Validation set performance results of the proposed CNN model using Confusion matrix

4.2. Comparison with the VGG16 model

The proposed model was compare with the VGG16 model. VGG16 is a CNN model proposed by Simonyan and Zisserman (2015) from the University of Oxford in the article Very Deep Convolutional Networks for Large-Scale Image Recognition. The model achieves a test accuracy of 92.7% in the top 5 in ImageNet. The results of this research allowed this model to win the 2014 ImageNet challenge. So we made a model based on this pre-trained model and replacing the last layer to predict on our two classes.

4.3. Experiment: Use of the model in real pictures

To test the perfomance of the proposed model in real world, we used two different pictures. The first picture was from an individual without mask and the second from the same individual with mask. The results for the first picture is given as follows:

Figure 9. Prediction outcomes for a sample of image without mask

From the above figure, we have observed that the model has predicted that at 98.72% the individual does not wear mask and at 1.28% that he has put mask. This shows that the proposed model is performing better. The below figure shows the results of the same individual wearing mask. As we can see, still the model has predicted well by detecting that at 6.73% the individual does not put mask while at 93.2% the model has detected that the individual has put mask.

```
<matplotlib.image.AxesImage at 0x24abbb14790>
```


```
# predict the class
img=load_image(image_path)
result = model.predict(img)
s_msk, msk =float(result[0][0]*100), float(result[0][1]*100)
```

```
print('Sans masque:{} %\nAvec masque:{}%'.format(round(s_msk,2), round(msk,2)))
```

```
Sans masque:6.73 %
Avec masque:93.27%
```

Figure 10. Prediction outcomes for a sample of image with mask

5. Conclusions

In this study, we made an approach to implement a CNN model for the classification of human faces with or without masks. With the model built from scratch, we were able to achieve 97.81% of accuracy on training data and a loss of 5.51%. On the other side, with the data that was not used during training, we had 92.5% accuracy and 28.4% loss. To achieve these precisions, we used the dropout technique to limit overfitting. It is possible to achieve better accuracy by increasing the training dataset but also by playing on the abandonment rate of the different layers. Much better results are obtained with transfer learning using the VGG16 model: 99.95% accuracy on training data and 100% on validation data.

Contributions: Conceptualization, S.A.K.; methodology, S.A.K. and M.E.K.; validation, M.E.K.; investigation, W.N.K.; resources, S.A.K.; data processing, W.N.K.; manuscript writing, W.N.K.; visualization, S.A.K.; supervision, W.N.K.; manuscript proofreading, M.E.K. The authors have read and approved the published version of this manuscript.

Funding: This research has received no financial support.

Data availability: Available by contacting the authors.

Acknowledgments: Not applicable.

Conflicts of interest: The authors declare no conflict of interest.

References

1. Alzu'bi A., et al., "Masked Face Recognition Using Deep Learning: A Review", Electronics, Vol. 10, 2666, 2021. https://doi.org/10.3390/electronics10212666.
2. Argotic C., Deep Learning et agriculture, Comprendre le potentiel et les défis à relever, 2018.
3. Chen Y.N. et al., Special Issue Review: Artificial Intelligence and Machine Learning Applications in Remote Sensing, Vol. 15, 2023. https://doi.org/10.3390/rs15030569.
4. Chollet F., Deep Learning with Python, Shelter Island , Manning Publications Co., 2018.

5. Deng J., et al., "ImageNet: A large-scale hierarchical image database", 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009. https://doi.org/10.1109/cvpr.2009.5206848.
6. Géron A., Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd Edition, O'Reilly Media, 2019.
7. Ghosh A., et al., "Fundamental Concepts of Convolutional Neural Network", in Balas, V., Kumar, R., Srivastava, R. (eds) Recent Trends and Advances in Artificial Intelligence and Internet of Things. Intelligent Systems Reference Library, Vol. 172, Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-32644-9_36.
8. Ivo D. & Günther G., Confusion Matrices and Rough Set Data Analysis, J. Phys.: Conf. Ser., Vol., 1229, 012055, 2019. https://doi.org/10.1088/1742-6596/1229/1/012055.
9. Ketkar N., Deep Learning with Python: A Hands-on Introduction, Apress, Berkeley, CA, 2017. https://doi.org/10.1007/978-1-4842-2766-4.
10. Markoulidakis I., et al., Multiclass Confusion Matrix Reduction Method and its Application on Net Promoter Score Classification Problem, Technologies, Vol. 9, no. 4, 81, 2021. https://doi.org/10.3390/technologies9040081.
11. Mbunge E., et al., "Application of deep learning and machine learning models to detect COVID-19 face masks - A review", Sustainable Operations and Computers, Vol. 2, 2021. https://doi.org/10.1016/j.susoc.2021.08.001.
12. Mishra R.K., et al., "The Understanding of Deep Learning : A Comprehensive Review", 2021. https://doi.org/10.1155/2021/5548884.
13. Moolayil J., Learn Keras for Deep Neural Networks, Vancouver, Apress, 2019.
14. Mpia H.N., Mburu L.W., Mwendia S.N., "CoBERT: A Contextual BERT model for recommending employability profiles of information technology students in unstable developing countries", Engineering Applications of Artificial Intelligence, Vol. 125, 2023. https://doi.org/10.1016/j.engappai.2023.106728.
15. Mpia N.H. & Inipaivudu B.N., "Gradient Back-Propagation Algorithm in the Multi-layer Perceptron: Foundations and Case Study", International Journal of Innovation and Applied Studies, Vol. 32, no. 2, 2021.
16. Mpia H.N., Mwendia S.N., Mburu, L.W., "Predicting Employability of Congolese Information Technology Graduates Using Contextual Factors: Towards Sustainable Employability", Sustainability, 2022, Vol. 14, no. 20, 13001. https://doi.org/10.3390/su142013001.
17. Parkhi, et al., Deep Face Recognition, University of Oxford, 2015.
18. Pirmin L., et al., Big Data et Machine Learning – Manuel du data scientist, Paris, Dunod, 2015.
19. Poornima G.N., et al., Conceptualizing Python in Google COLAB, Shashwat, Bilaspur, 2021.
20. Simonyan K. & Zisserman A., Very Deep Convolutional Networks for Large-Scale Image Recognition, University of Oxford, 2015.
21. Talahua J.S., et al., "Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic", Sustainability, Vol. 13, 6900, 2021. https://doi.org/10.3390/su13126900.