Archiving 4.0: Dataset Generation and Facial Recognition of DRC Political Figures Using Machine Learning

Ferdinand Kahenga Ngongo[1,2](✉), Antoine Bagula[2], and Olasupo Ajayi[2]

[1] Ecole Supérieure d'Informatique SALAMA, Lubumbashi, Democratic Republic of Congo
ferdinandkahenga@esisalama.org

[2] University of Western Cape, Cape Town, South Africa

Abstract. It is widely recognized that digital archiving represents many advantages compared to manually preserved documents. These include i) reduced risk of losing data, ii) eco-friendliness, iii) data security, iv) faster access to data, v) simple data management, vi) overall costs saving, and vii) potential for data recovery. However, in contrast to developed nations, which have experienced a steady maturity in the field, digital archiving is still in its infancy in developing countries, especially in Africa, where years of slavery, colonization, both economic and political issues, and wars have deprived nations of their history. Building around techniques of the fourth industrial revolution (4IR), this paper addresses this gap by proposing a digital archiving model called "Archiving 4.0". The model curates a dataset of images of political figures in the Democratic Republic of Congo (DRC) from different sources and classifies these images using machine learning techniques to achieve facial recognition that will help recognize historical people in photographs and videos. To the best of our knowledge, the proposed dataset constitutes the first attempt at digital archiving in the political space of the DRC and provides an example that can be emulated to spin out related works on the African continent where digital archiving is needed for political research studies and also for the preservation of the history of nations. Through performance evaluation, accuracy, precision, recall, and loss; the paper reveals that Transfer learning outperforms traditional machine learning on different metrics of interest when using the generated dataset: 93% of validation accuracy.

Keywords: Archiving · Deep Learning · Democratic Republic of Congo · Transfer learning · Face recognition

1 Introduction

It is widely recognized that our past shapes our future by defining and redefining our perception of the world. This is a result of our memories being storages of blocks of experiences of the past upon which we construct visions and plan the future. Therefore, a better understanding of the world we live in can be attained from the knowledge of our

M. Masinde and A. Bagula (Eds.): AFRICATEK 2022, LNICST 503, pp. 53–64, 2023.
https://doi.org/10.1007/978-3-031-35883-8_4

past which shapes our present and ineluctably influences our ways of doing things, acting in society, and how identifying ourselves. Archiving is a key process for achieving social and economic development and administrative efficiency, as it encompasses the entire gamut of significant governmental activity from past to present experiences, including its successes and failures. While developed nations have experienced a steady maturity in the field, digital archiving is still in its infancy in developing countries, especially in Africa where years of slavery, colonization, economic and political instabilities, and wars have robbed nations of their history. As largely reported, there are several national archives and records of African countries being held in foreign countries abroad. This is evidenced by the recent repatriation of large amounts of African archival records to western metropolitan cities.

Like in many other African countries, the history of archives in the Democratic Republic of Congo (DRC), has gone through very difficult times. Delphin Bateko Moyikoli in [1] reported that there was the first destruction and looting of the archives of the Belgian colony (Belgian Congo) by Leopold II in the year 1908. This was followed by the looting and destruction of archives around the 1960s period of the country's independence. Added to these, is the looting of the presidential archives by the Alliance of Democratic Forces for the Liberation of Congo and its allies in 1997, archives that had already been neglected during the Second Republic. These repeated archive lootings, theft, and destruction have made it difficult to maintain a proper archiving system of the DRC, which could help in keeping track of the country's history. Today's key archiving issue in the DRC consists of not only building adequate storage and digital repository but also in the reconstruction and documentation of these archives.

This paper contributes to the field of digital archiving by proposing a model called "Archiving 4.0" that relies on Fourth Industrial Revolution (4IR) techniques to i) generate a dataset of political figures of the DRC from different sources and ii) detect and recognize the faces of these personalities by relying on machine learning techniques. The second contribution of this article is to determine, based on experiments carried out, the appropriate machine learning models for the detection and recognition of figures given the small quantity of data and the degraded quality of certain pictures, with some dating as far back as the early 1960s. Several approaches have been used and their results are compared to select the best models and paradigms to be used when dealing with the generated and/or similar archiving datasets. Performance evaluation reveals that a pretrained model using Transfer learning outperforms traditional machine learning models on different evaluation parameters.

The remainder of this paper is organized as follows: a background and literature review on the topic being studied is presented in Sect. 2, while Sect. 3 reports on the research methodology used and the data generation. A performance evaluation with a comparison of models used is presented in Sect. 4. Section 5 concludes the paper and provides suggestions for future work.

2 Background and Literature

2.1 Face Detection and Face Recognition

There are multiple applications of facial recognition in the fields of security, access control, transport, and healthcare. A significant number of scientific papers have been devoted to face detection and face recognition systems that make it possible to identify or verify the identity of an individual in just a few seconds based on his facial characteristics: e.g. distance between the eyes, bridges of the nose, corners of the lips, ears, chin, etc.

Many of the techniques used in facial recognition are based on models of convolutional neural networks trained on several images containing figures. Several frameworks and results of these works are accessible. Haar [2], DeepFace [3] from Facebook, FaceNet [4] from Google, Dlib [5], Multi-task Cascaded Convolutional Networks (MTCNN) [6], Retina-Face [7] are some of the most popular ones. Face recognition can be divided into 3 steps: (1) face detection, (2) feature extraction, and (3) classification. Dlib CNN, MTCNN, and Retina-Face are networks used for face detection and facial landmarks. Feature extraction in the case of face recognition can be done in different ways. FaceNet, for example, is a deep neural network that takes an image of the person's face as input and outputs a vector of 128 numbers, called embedding, which represents the most important features of a face [4].

As part of the "Archiving 4.0" framework, machine learning was used for facial detection on Robben Island Mayibuye Archives in [8]. Dlib HOG, Dlib CNN, OpenCV's Haar cascade, OpenCV's Caffe, and TensorFlow were compared based on the Average Precision, Precision, and Recall metrics with CNN performing the best [8]. Similarly, in [9] Dlib and MTCNN as face detectors were also used to index the faces on the historical photographs of the Stockholm city museum with MTCNN showing better performance.

2.2 Transfer Learning

Motivation. "Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned" [10]. The fundamental motivation for Transfer learning was discussed in NIPS-95 workshop on "Learning to Learn" [11] which focussed on the need for lifelong machine learning methods that retain and reuse previously learned knowledge. This has been recently reinforced by Andrew Ng in the tutorial called "Nuts and bolts of building AI applications using Deep Learning" where he stated that Transfer learning will be the next driver of machine learning commercial success [12].

Transfer Learning in Deep Learning. In CNN, the first few layers are used to extract high-level general features, while class-specific features are captured by the last couple of layers. The learned features from the first layers appear not to be specific to a particular dataset or task, but general in that they are applicable to many datasets and tasks [13].

The CNN Transfer learning strategy consists of freezing the first layer's weights of the pre-trained model and only training the newly added dense layers with randomly initialized weights. Figure 1 shows how the parameters learned in Task 1 of the pre-trained model are transferred to Task 2.

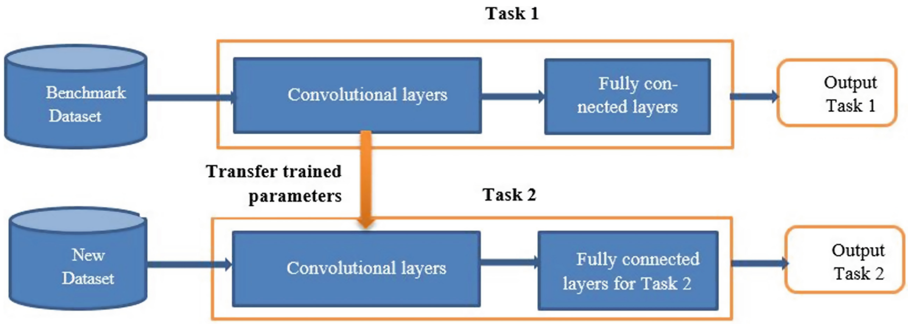

Fig. 1. Transfer of trained parameters in the context of deep learning

2.3 Evaluation Metrics

In machine learning, the accuracy (1), precision (2), and recall (3) metrics respectively indicate what proportions of all positive predictions are actually positive and how many positives the model identifies. In the context of object detection, the intersection over union (IoU), defined in (4) helps to know how accurate the predicted bounding boxes of objects in the image are. It gives the overlap between the ground truth bounding box and the predicted bounding box. For instance, Fig. 2 shows the ground truth and the predicted bounding box of president Mobutu's image.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Number\ of\ total\ predictions} \tag{1}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

$$IoU = \frac{area\ of\ intersection}{area\ of\ union} \tag{4}$$

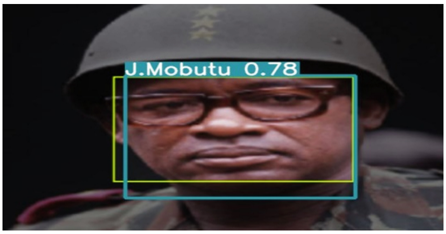

Fig. 2. The ground truth bounding box is in yellow and the predicted bounding box is in blue. The area of intersection is the intersection between the yellow and the blue bounding boxes. The detected face at 78% in the image is former president Joseph Mobutu (Color figure online)

3 Methodology and Data Acquisition

3.1 Methodology

The bulk of our methodology is summarized in Fig. 3 and consists of two phases. Phase 1 consists of the dataset generation while Phase 2 brings together the different stages of training machine learning models for the detection, classification, and recognition of personalities in images.

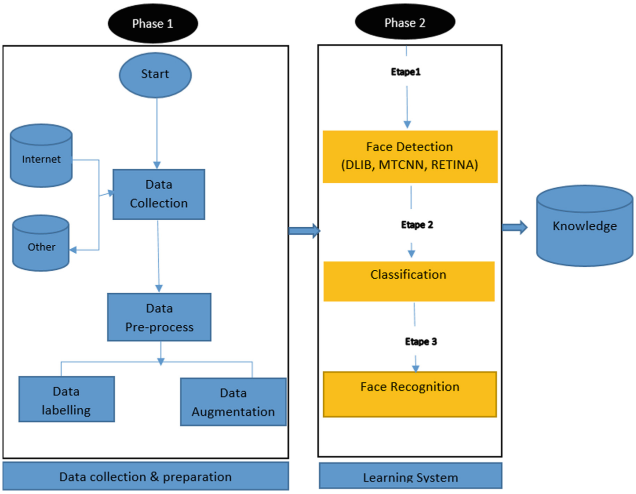

Fig. 3. Main phases of the methodology

An image sent to phase 2 is subjected to face detection step 1, which extracts all the faces present in the image using models such as DLIB, MTCNN, or RETINA-FACE. The faces are then classified using the classification model previously trained with the set of data collected in phase 1. Once classified, the annotation of each face is placed on the initial image.

3.2 Data Collection and Preparation

Our dataset is a collection of old and recent images of well-known political figures from the DRC [17]. This dataset is composed of 3431 images with 4113 annotations of the faces of 19 politicians, including 5 presidents of the republic, 1 vice-president, 7 prime ministers, 1 governor, 2 presidents of the national assembly, a president of the senate, and the others. 30% of the images are old black and white photographs of politicians who fought for the independence of the DRC in the 1960s.

The acquisition of these images was done through the manual and automatic collection (web-scrapping). At least 70% of the data was acquired from the Internet, Google, Facebook, YouTube, and Twitter. The other images were obtained from old newspaper publications and old videos with black and white images. Figure 4 shows the breakdown of the Dataset in terms of images per person.

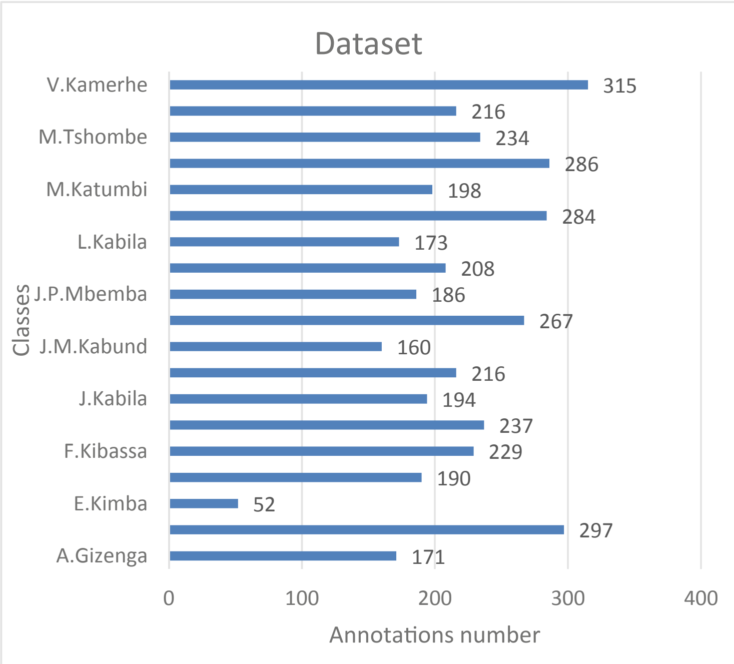

Fig. 4. Dataset balance

The approach taken for data preparation consisted of a minimum of 50 images for people with fewer images, around 150 images for the second category of politicians, and 300 images for those who had more images available through the different media that were considered. The quality of the images was balanced between those of high resolutions and others with very low resolutions. All images were then set to a size of 300×300 pixels. Finally, manual labeling was done using Roboflow [14].

The richness of the curated dataset lies not only in the diversity of the quality of the images but also in the fact that we have images of certain people from their youth to old age. Similarly, the images were captured using scanners and cellphone cameras from newspapers (Fig. 5).

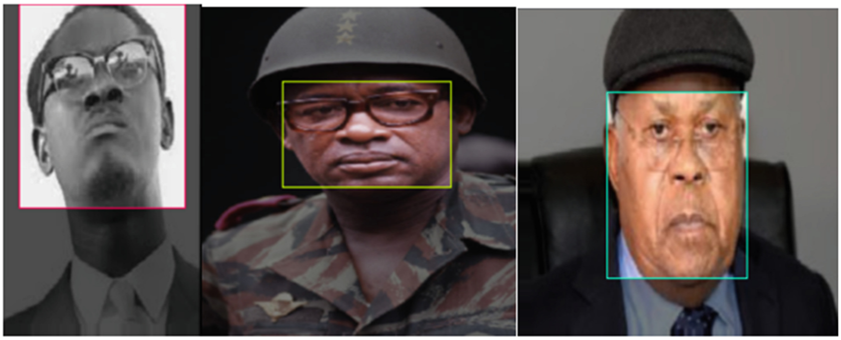

Fig. 5. Labelling process

4 Performance Evaluation

4.1 Face Detection

Three different models (MTCNN, RETINA-FACE and DLIB) were used for the face detection. As shown in Figure 6, each of these models gave us the ability to:

- align and detect faces by flipping through 4 points representing facial-area and landmarks
- extract features
- extract facial FaceNet embeddings which are vectors of the characteristics of a face

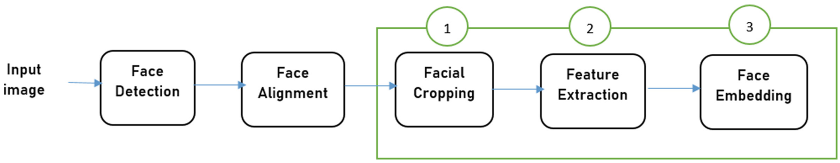

Fig. 6. Steps of face detection

Table 1 summarizes the detection accuracies of each of the three models considered.

Table 1. Accuracy of face detection by model

	Model	Detection accuracy (%)
1	MTCNN	60.5
2	DLIB	75.4
3	RETINA-FACE	92.1

RETINA-FACE gave the highest accuracy and hence was considered for the rest of this work. As mentioned earlier, the dataset is composed of images with various parameters in terms of quality, size, the physical appearance of people, and so on. All models showed a low success rate against very old images.

4.2 Classification

After the face detection step, we placed all the detected faces in sub-folders labeled with the class names. For instance, the Lumumba sub-folder contained all the faces extracted for Lumumba. Three classification approaches were tested to find the best performance on our data. These include:

- a CNN called "Model 1" which takes as input cropped faces as shown in point (1) of Fig. 6.
- A classifier referred to as "Model 2" which takes as input embedding vectors, and
- a deep neural network MobileNetV2 model called "Model 3" which was pre-trained on imageNet [15] and thereafter trained by Transfer Learning on cropped faces.

Model 1. This model is a convolutional network of 3 convolutional layers followed each time by a max pooling and then by a Fully connected layer. The model takes as input a small image of the face resulting from the face detection step. The validation accuracy is around 35%.

Model 2. We compared the results of 4 different classifiers, namely, BoostedTreesClassifier, RandomForestClassifier, DecisionTreesClassifier, and LogisticClassifier. The results obtained after validation allowed us to set our sight on LogisticClassifier as shown in Fig. 7.

Model 3. Obtained after freezing 100 layers of MobileNetV2 model [16] pre-trained on imageNet to allow the training of only the remaining 54 layers. This technique allowed us to improve the results from 71% to 93.7%. MobileNetV2 model was trained using the Adam optimization function and the sparse categorical cross-entropy loss function for 20 epochs.

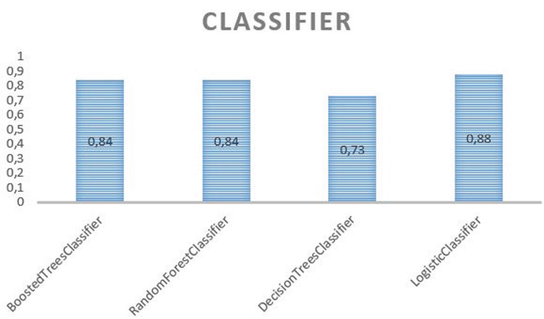

Fig. 7. Classifier validation accuracy comparison

Table 2 compares the results of the three models obtained for three different performance metrics. They reveal that Model 3 outperforms the other two models across all three metrics.

Table 2. Evaluation metrics comparison on the test set

	Model	Accuracy (%)	Precision (%)	Recall (%)
1	Model 1	35.7	32.2	31.4
2	Model 2	88.4	86.3	86.4
3	Model 3	93.7	93.1	92.9

4.2.1 Facial Recognition

Facial recognition is the phase that consists of placing each detected and classified face in the context of its initial image and telling which class the person(s) in the image belongs to. In this work, facial recognition was performed by combining the best performances of the previous steps 1 and 2.

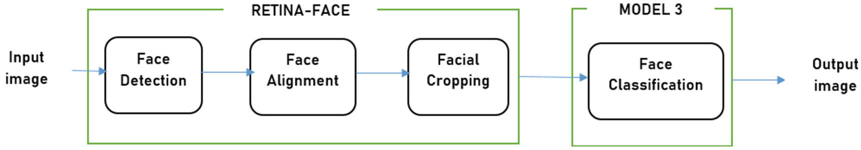

Fig. 8. Facial recognition steps

As shown in Fig. 8, a new input image containing a personality under study undergoes the following operations: face detection, face alignment, and facial cropping with RETINA-FACE; then the extracted face from the image is then used as input to the CNN model Model3 which classifies the face. If the image contains several faces, the classification step will be repeated for each of them. The Output image is the Input image on which bounding boxes are added to identify each known personality.

4.3 Discussion of Results

In this paper, we extracted from our original dataset a portion that allowed us to constitute 10 classes of training and test sets composed of respectively 507 and 127 images. According to the results obtained, it appears that the image type (black and white or color) does not influence face detection. On the other hand, the quality of the pixels and the position of the face on the image have an impact on the detection. Images taken from old newspapers have a low detection rate, as evidenced by the detection accuracy of 79.8% for Kasa-vubu and 90% for Lumumba, the lowest and for the oldest pictures in our illustration study as shown in Fig. 9.

It is also worth noting from this study that photographs with faces in a non-frontal position also presented some difficulties during the face detection step. The classification using cropped faces obtained from the face detection gave a poor result on the small amount of data, however when for the same number of images, the classification was done using the embedding vectors, the results were much better. Ultimately, using Transfer

Fig. 9. Non-detected faces of Kasa-Vubu and Lumumba.

learning with adequate frozen layers yielded the best result of all. Figure 10 shows the benefit of freezing and fine-tuning certain layers of Model 3 network which significantly boosted the accuracy of the model from 70% to 93.7% of accuracy.

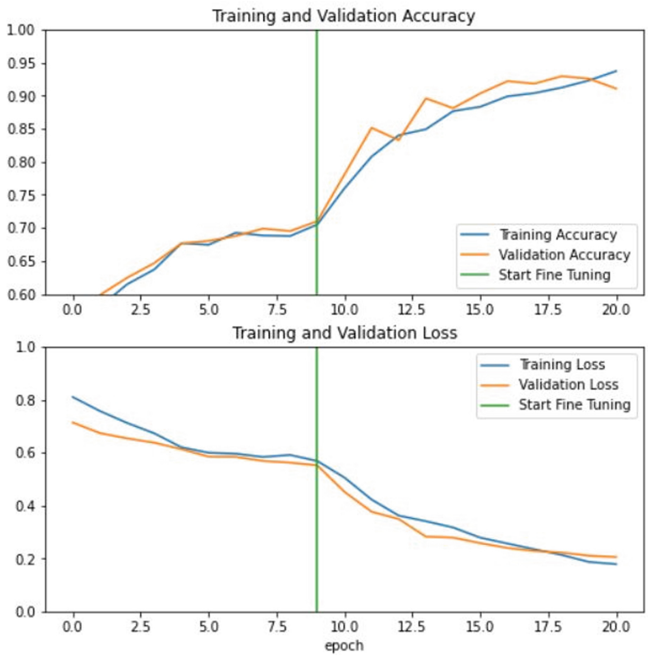

Fig. 10. Training and validation Accuracy and Loss.

All of these executions were performed on Google Colab Pro using the Tesla 100-PCIE GPU Accelerator.

5 Conclusion

In this paper, a curated dataset of several images of different generations of political figures who have played key roles in the history of DRC from her independence to the present day was presented. The particularity of the dataset is that it is made of images from several sources and different qualities. Once curated, the images were passed through different machine-learning models for facial classification and recognition. The obtained results revealed that the models performed better with more recent pictures for image classification, while models that leveraged Transfer learning gave better results in terms of face detection compared to those based on basic machine learning only. There is scope for future extensions of the presented work in the field of deep learning optimization by using automatic layer selection as well as the deployment of the transfer learning model into real-life federated learning infrastructures for museums. Such federated learning infrastructures will benefit from extensions the networking techniques borrowed from [18–20] for the interconnection of participating museums.

References

1. ADIAC: Journée internationale des archives: la RDC en marge de l'événement. Agence d'information d'Afrique Centrale, 14 June 2018. https://www.adiac-congo.com/content/journee-internationale-des-archives-la-rdc-en-marge-de-levenement-84714. Accessed 04 Apr 2022
2. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Conference on Computer Vision and Pattern Recognition, p. 1 (2001)
3. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
4. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015)
5. Dlib c++ Library. http://dlib.net/. Accessed 09 May 2022
6. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. **23**(10), 1499–1503 (2016)
7. Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., Zafeiriou, S.: RetinaFace: single-shot multi-level face localisation in the wild. arXiv:1905.00641 (2019)
8. Kern, D., Zweng, M., Sello, S., Bagula, A., Klauck, U.: Archiving 4.0: application of image processing and machine learning for the Robben Island Mayibuye Archives. In: 2020 International SAUPEC/RobMech/PRASA Conference, pp. 1–6 (2020)
9. Poudel, A.: Face recognition on historical photographs (2021). https://uu.diva-portal.org/smash/get/diva2:1622968/FULLTEXT01.pdf. Accessed 09 May 2022
10. Soria, E., Martín, J., Martinez-Sober, M., Magdalena-Benedito, J.R., et al.: Handbook of Research on Machine Learning Applications and Trends: Algorithms, methods, and techniques: Algorithms, methods, and techniques. IGI Global, Hershey (2009)
11. Learning to Learn: Knowledge Consolidation. http://socrates.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/. Accessed 01 June 2022
12. Ng, A.: Nuts and bolts of building AI applications using Deep Learning. NIPS Keynote Talk. (2016)

13. Yosinskim, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems 27 (NIPS 2014) (2014)
14. "Roboflow. https://docs.roboflow.com/annotate. Accessed 10 June 2022
15. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
16. Chiu, Y.C., Tsai, C.Y., Ruan, M.-D., Shen, G.Y., Lee, T.T.: Mobilenet SSDv2: an improved object detection model for embedded systems. In: 2020 International Conference on System Science and Engineering (ICSSE), pp. 1–5 (2020)
17. Kahenga Ngongo, F., Bagula, A.: J. YOTE, "DRC_POLITICIANS_DATASET," *Mendeley Data,* vol. V1 (2022). https://doi.org/10.17632/dz47c9pnd5.1
18. Bagula, A.B.: Hybrid traffic engineering: the least path interference algorithm. In: Proceedings SAICT 2004, ACM International Conference Proceedings Series, pp. 89–96, October 2004
19. Bagula, A.B.: Hybrid routing in next generation IP networks. Elsevier Comput. Commun. **29**(7), 879–892 (2006)
20. Bagula, A.B.: On achieving bandwidth-aware LSP/LambdaSP multiplexing/separation in multi-layer networks. IEEE J. Sel. Areas Commun. (JSAC) **25**(5), 987–1000 (2007). Special issue on Traffic Engineering for Multi-Layer Net-works