OPEN ACCESS

Masked and unmasked face recognition model using deep learning techniques. A case of black race

***Mabirizi Vicent, Ampaire Ray Brooks and Muhoza B. Gloria**

Kabale University, Uganda

**Corresponding Author*
vmabirizi@kab.ac.ug

KURJ
ISSN 2790-1394

pp. 22 - 31
Vol 2. Issue 1.
May 2023

ABSTRACT

Currently, many institutions of higher learning in Uganda are faced with major security threats ranging from burglary to cyber threats. Consequently, the institutions have recruited and deployed several trained personnel to offer the desired security. As human beings, these personnel can make errors either by commission or omission. To overcome the limitation of trained security personnel, a number of face recognition models that detect masked and unmasked faces automatically for allowing access to sensitive premises have been developed. However, the state -of -the art of these models are not generalizable across populations and probably will not work in the Ugandan context because they have not been implemented with capabilities to eliminate racial discrimination in face recognition. This study therefore developed a deep learning model for masked and unmasked face recognition based on local context. The model was trained and tested on 1000 images taken from students of Kabale University using Nikon d850 camera. Machine learning techniques such as Principal Component Analysis, Geometric Feature Based Methods and double threshold techniques were used in the development phase while results were classified using CNN pre-trained models. From results obtained, VGG19 achieved the higher accuracy of 91.2% followed by Inception V 3 at 90.3% and VGG16 with 89.69% whereas the developed model achieved 90.32%.

Keywords: Deep learning, Face detection, Face recognition, Masked face, Unmasked face

Introduction

Human face recognition is widely used to authenticate identities and prevent unauthorized access to organization's physical facilities, networks and database networks due to it's because it is not intrusive nature which creates an opportunity for distant authentication checks without the one's attention [1]. Although there is registered progress in the domain of face detection and recognition for security, there are still existing issues hindering the progress to reach or surpass human level accuracy [2]. These issues are variations in human facial appearance such as; varying lighting condition, noise in face images, scale, pose, masked face [1] and eliminating racial discrimination in face recognition.

Recently, methods that do not necessarily require structured image data, termed as "deep learning algorithms" have been developed. The models based on deep learning algorithms can extract and learn image features automatically, overcoming the limitation of manual feature extraction in classical machine learning-based models. The deep learning paradigm tackles problems on which shallow architectures for example, support vector machine (SVM) are affected by the curse of dimensionality [3]. Deep learning is a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively and classify higher level features from data [4]. Convolutional Neural Networks (CNNs), is one of the deep learning algorithms that has been widely used for identification of objects such as bones, handwritten digits, and traffic sign and faces [5]. The algorithms have various architectures such as VGG-16, VGG-19 ResNet50, Inception-V3 and EfficientNet which are pretrained on ImageNet dataset.

Motivated by the advantages of deep learning algorithms over classical machine learning algorithms, the authors in [6] proposed a method that combines deep learning models and Local Binary Pattern (LBP) features into a unified framework to recognize the masked face. The study achieved a result of 87% f1-score. In [7], the authors presented a framework for recognizing individuals with and without face mask using ResNet-50 deep learning architecture. The study obtained F1-score of 0.897 and 0.447 for the masked and unmasked respectively. Sanjaya et at. [8] built MobileNetV2 deep learning model for image classification to detect people who are wearing a face mask and not wearing it. The authors achieved a classification accuracy of 96.85%.

Much as deep learning methods have achieved state-of-the-art performance in the recognition of masked and non-masked faces, the reported models do not have local context and probably will not perform well in our local context. This is because they have not been implemented with capabilities to eliminate racial discrimination in face recognition. Therefore, there was need to improve the state -of -the art models by developing and testing the models on local dataset. The developed model was trained and tested using CNN pre-trained models, Principal Component Analysis (PCA) and Radial Basis Function Networks using locally collected dataset with the aim of improving the recognition accuracies of the models

Related Work

In [9], the authors proposed a combination of two CNN architecture embedded with a similarity-based descriptor to recognize faces. The experimental results of their study out-performed 6 state-of-the-art models by at least 15.6%.

The study by Wu, GuiLing [10] proposed a masked face recognition algorithm based on attention mechanism. First, the authors separated masked face image using local con-strained dictionary learning method and applied dilated convolution to reduce resolution in the subsampling process. In the final stage, the attention mechanism neural network was used to reduce the information loss in the subsampling process.

Hariri [11] proposed a method based on occlusion removal and deep learning-based features. In the study, the masked face region was first removed, deep features extracted using Convolution Neural Networks (VGG-16, AlexNet, and ResNet-50) and Multilayer Perceptron (MLP) applied for the classification process. The Experimental results on Real-World-Masked-Face-Dataset revealed that ResNet-50 attained highest recognition rate of 88.9%

Soni et al. [12] developed a model that detects whether a person is wearing a helmet in real time thereby, detecting any violations. The project was also implemented with the help of TensorFlow, Keras and OpenCV. Their proposed model showed major improvements when compared to some previous models that gave wrong predictions whenever a rider wears clothes over their face. They achieved an overall accuracy of 98% when tested.

In [13], the authors developed a CNN model for detecting face mask and for detecting violation of social distancing rule. The authors used You Only Look Once (YOLO) object detection model for detecting people in a frame and social distancing was calculated based on Euclidean distance between the centroids of the detected boxes. The model attained an accuracy of 96.35%

Work in [14] used smart image sensor with local gradients to design and implement smart pixels and digital coprocessor for human face recognition. Their method achieved 96.5% classification accuracy on a database of infrared face images. Authors in [15] used k-Nearest Neighbor (k-NN) for enhancing and recognizing people faces in employee attendance system. Their study achieved a classification accuracy of 92.5%.

In [16], the authors suggested a facial recognition technique using SURF features and Support Vector Machine (SVM) classifier on Yalefaces and UMIST face databases. Their experimental results yielded a recognition accuracy of 97.78% on Yalefaces and 97.87% on UMIST database.

Although studies in [9]–[16] achieved great successes, the authors did not explore implemented with capabilities to eliminate racial discrimination in face recognition. Addition-ally, the authors used classical machine learning algorithms which requires human crafted features for training. The process of manual feature extraction is tedious and requires deep understanding of the image [17].

Methodology

In this study, we adopted image processing technique as illustrated in figure 1 to guide research development.

A. Image acquisition

First, a dataset was obtained at step one. This dataset was primary data containing a total of 500 images both masked and masked images of scale magnification x200, x400 and x650 taken using a Nikon d850 camera in a non-controlled light intensity. These images were obtained from students and staff members of Kabale University who red and signed a consent form proving their voluntary acceptance to participate in this study. To cater for ethical consideration, participants were briefed about this study and how their identity will be kept confidential throughput the entire research process.

The target size of the dataset was 1000 images for model training and testing however, due to scarcity of research participants, we did not meet the target size. To cater for this limitation, augmentation techniques [18] were applied to manipulate the datasets to meet the target size.

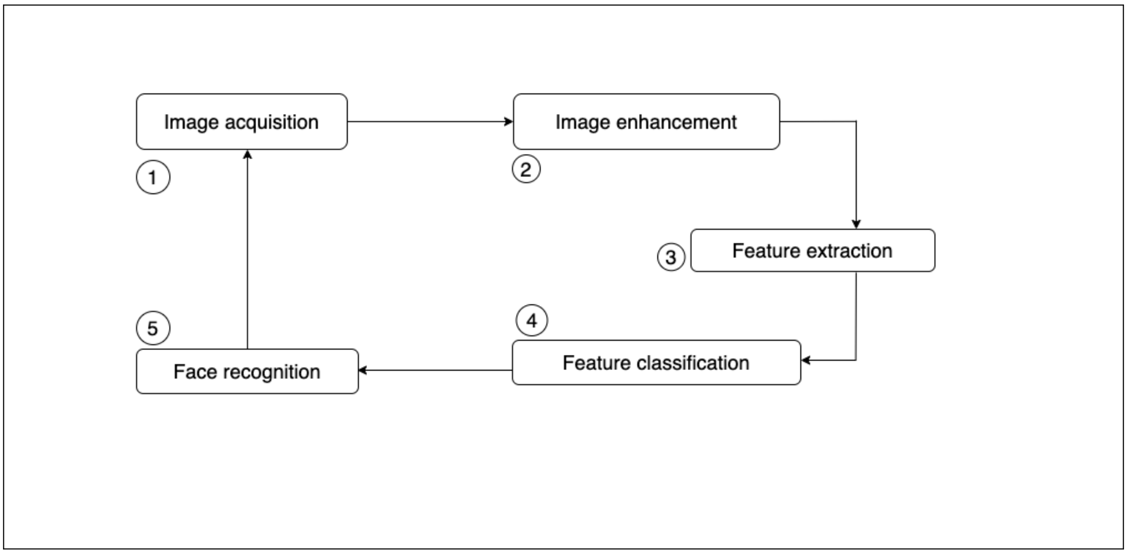


Figure 1: Methodology steps

B. Data preprocessing

The images taken had noise such as unwanted back-ground, too much light intensity since were taken from a non-controlled environment. Thus, data pre-processing was done to eliminate this noise. To achieve this, gray scale conversion and threshold data cleaning techniques were applied. Gray scale conversion process involves representing the intensity of information of the light in form of pixels. Figure 2 shows a processed masked and unmasked image.

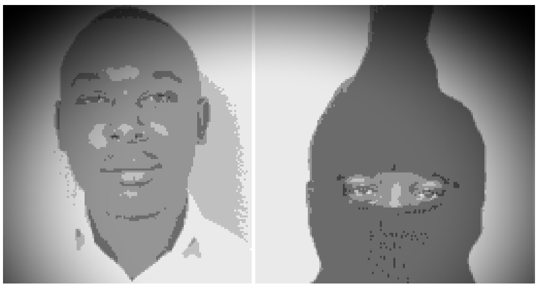


Figure 2: Processed masked and unmasked image

C. Feature extraction

After eliminating the unwanted regions of the dataset, the next step was extracting biomarkers needed in masked face recognition. To achieve this, Principal Component Analysis (PCA) feature extraction technique was utilized. The PCA was invented by Karl Pearson in 1901 [19], 64 years later it was modified to cater for pattern recognition [20] which was then recognized as a prominent technique for extract features specifically for face recognition in early 90's [20]. While PCA guided the process of feature extraction, the actual process of extracting features was done by using Geometric Feature Based Methods (GFBM) [21]. In this case a set of geometrical face features of the eyes, a mouth, and a nose were computed while the mask was eliminated. Thereafter, outline of the face and position of the different facial features from a feature vector were represented. To compute the geometrical relationships, the location of those points was used. The output of this step is an unmasked image extracted from a processed masked image shown in figure 1 below.

D. Feature classification

The extracted features were then classified to whether there matching patterns between the processed image and the images stored in the database. Feature classification was done using VGG16, VGG19, Inception V 3 and scratch model. We opted for these models because they were well suited for ensuring that optimum set of features are selected during classification. The search for optimum set was guided by a fitness value. When simulated annealing is finished, all the different subsets of features are compared and the fittest (one that performs the best) was selected. The fitness value search is obtained with a wrapper and K-fold cross validation is used to calculate the error on the classification algorithm.

E. Face recognition

Finally, using the classified features, the model was able match masked or non-masked by using geometric approach [22] also known as template-based algorithms. This approach compares the input image with a set of templates constructed using Support Vector Machine (SVM) statistical tools. In this case local facial features were analyzed using their geometric relationships.

Proposed Model

The proposed model is divided into three layers namely; image loading, data processing and face recognition shown in figure 3.

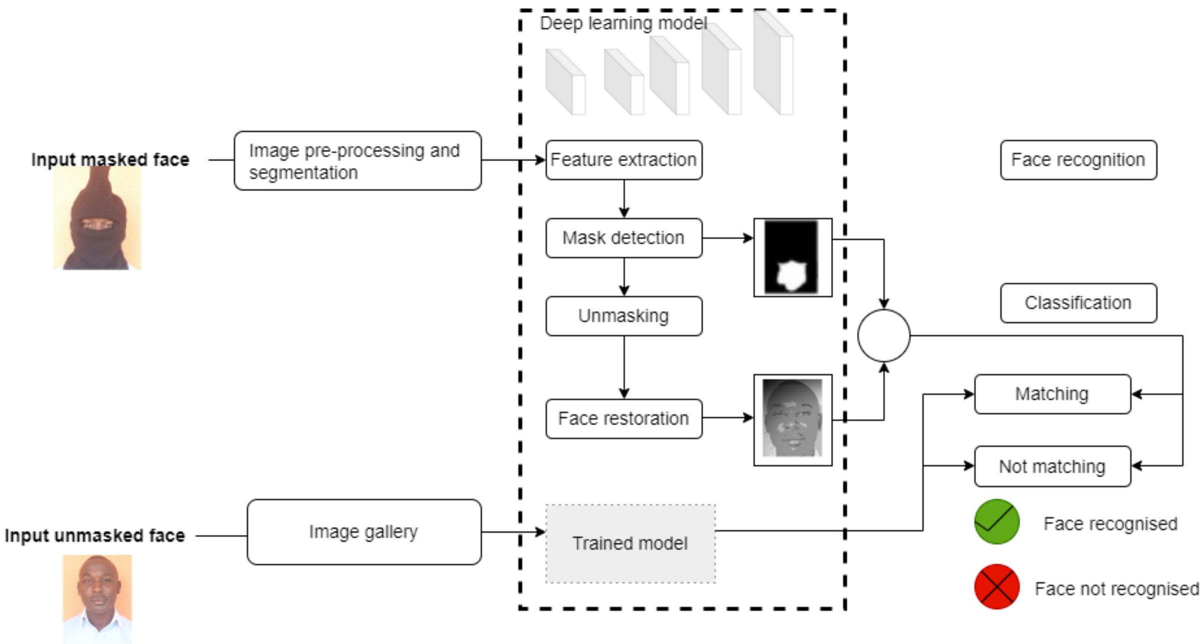


Figure 3: A deep learning model for masked and unmasked face detection.

Model Description

The captured image is subjected to the image loading layer implemented by TensorFlow library to detect whether the input image is of human face and a black race else the image is trashed. Once the input meets the criteria is then subjected to CNN pre-processing model to eliminate the unwanted regions including; the background and unwanted regions.

The processed image is then converted to gray scale for accurate feature extractions. Thereafter water threshold technique is applied to detect the mask and separate it from the original face. Separation of the mask from the original image causes some distortion in the original image thus, Principal Component Analysis (PCA) feature extraction technique is be utilized to restore the composition of the original image.

The restored image is them compared to the original image stored in the photo gallery to establish matching patterns. The result from image comparison is the classification results confirming whether the images are matching. To achieve this, double-strategy random forest algorithm was used. In case the masked face is successfully recognized a green light is triggered indicating to grant access whereas failure in face recognition is communicated by a red light.

Model Training And Testing Results

Results obtained were classified based on whether an image was correctly classified masked and the model was able to match it with its corresponding unmasked image in the gallery or incorrectly classified as unmasked and the model was unable to match it with its corresponding masked image stored in the image gallery. To achieve this, VGG16, VGG19, and Inception V 3 deep learning classification models were used. These classifiers were used one at a time in a sequence of VGG16, VGG19, and Inception V 3.

Model performance

Model performance for the different proposed pre-trained models and training from scratch method was evaluated and compared using four performance metrics, including precision, recall, accuracy and F1-socre, as shown in equation (1-4).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \ldots\ldots\ldots\ldots\ldots(\text{i})$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{ii})$$

$$\text{Recall} = \frac{\text{TP}}{\text{TN+FN}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{iii})$$

$$\text{F1\_Score} = 2\left(\frac{\text{Precision}*\text{Recall}}{\text{Precision+Recall}}\right) \ldots\ldots\ldots(\text{iii})$$

Where; TP is true positive value, TN is true negative value, FP is false positive value, FN is false negative value.

Table 1: Model performance on training dataset

Models	Precision	Recall	F1- score	Accuracy	Training loss
VGG19	0.940	0.910	0.940	0.942	0.086
VGG16	0.891	0.892	0.871	0.872	0.240
Inception v3	0.930	0.934	0.931	0.931	0.132
Scratch model	0.970	0.972	0.972	0.973	0.031

Table 2: Model performance on testing dataset

Models	Precision	Recall	F1-score	Accuracy	Training loss
VGG19	0.910	0.910	0.910	0.910	0.290
VGG16	0.890	0.89	0.880	0.870	0.240
Inception v3	0.903	0.903	0.903	0.903	0.100
Scratch model	0.901	0.902	0.902	0.90	0.351

The curves in figures (5-10) represents model performance on both training and testing dataset for VGG16, VGG19, Inception V3, and scratch model respectively.

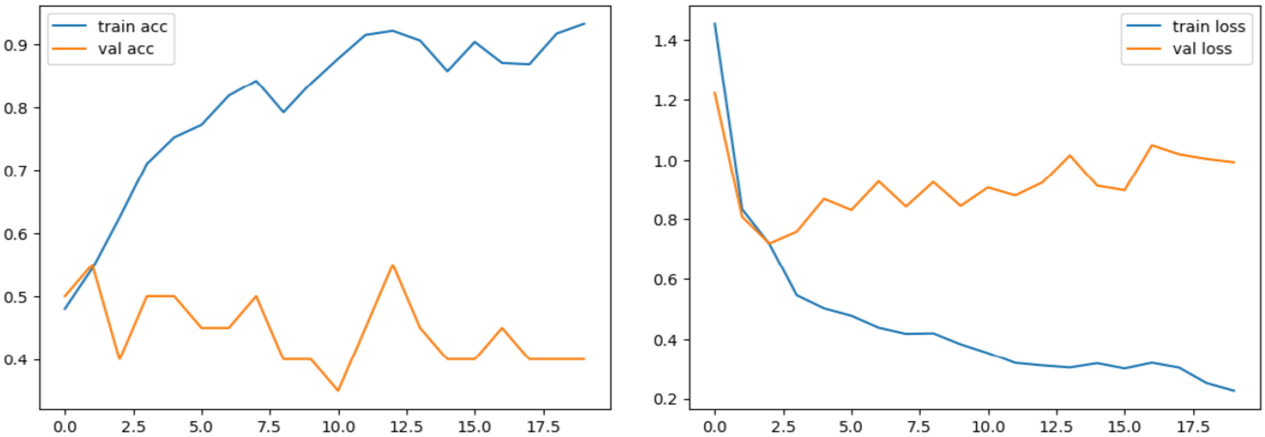


Figure 4: vgg16 model accuracy vs loss

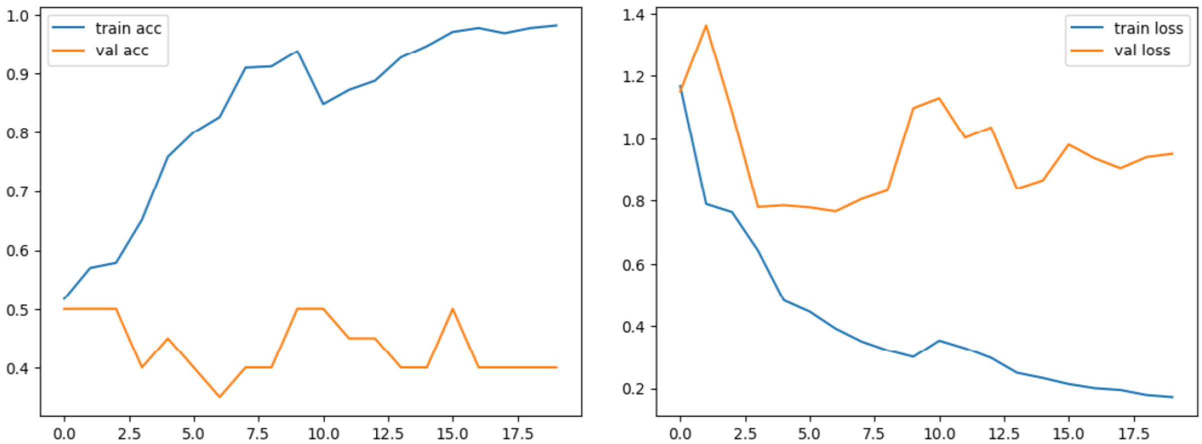


Figure 5: vgg19 model accuracy vs loss

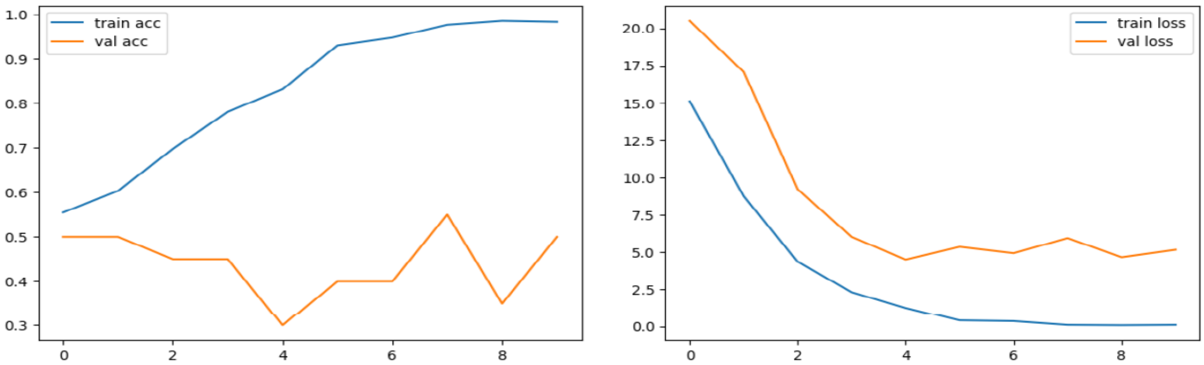
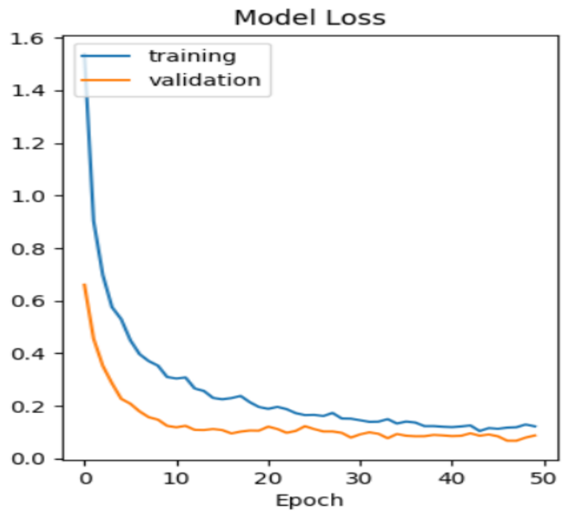
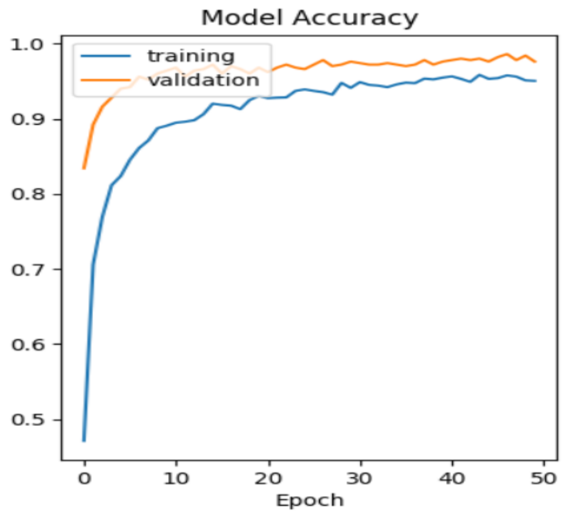


Figure 6: Inception V3 model accuracy vs loss





Results and Discussion

In Table 1 we present model performance metrics using testing dataset. As observed, vgg19 yielded the highest performance with 0.94 accuracy and 0.086 training loss. Inception v3 yielded a closer performance compared to vgg19 with 0.93 accuracy and 0.13 loss using a similar dataset while vgg16 had the worst performance among the pre-trained models with 0.89 accuracy and 0.24 model loss on training dataset. Our proposed model achieved better performance compared to the pre-trained CNN models on testing dataset i.e. the model achieved 0.97 accuracy with 0.031 training loss, this was achieved by increasing the dataset and epochs which eliminated overfitting from during model training.

Table 2 presents results obtained using both pre-trained models and the scratch model using testing dataset. As observed in results table 2, vgg19 yielded the highest performance compared to other models with 0.91 accuracy and 0.29 testing loss however, the model achieved better results during training as compared to testing. Inception V3 followed vgg19 in terms of performance with 0.903 accuracy and 0.10 model loss. As it was the case during model training, vgg16 yielded the worst performance during model testing with 0.89 testing accuracy and 0.24 model loss. Unlike in training case, the scratch model performance dropped from 0.97 to 0.90 and 0.031 to 0.351 testing accuracy and model loss respectively. This negative change is due to limited dataset used test the model as compared the dataset used for training.

Conclusion

This study mainly focused on recognizing both masked and unmasked faces to improve security checks in institutions of higher learning using images obtained from black Africans. This was achieved through developing masked and unmasked face recognition model. The model was trained and tested on a total of 1000 images taken from students and staff of Kabale University. Principal Component Analysis, Geometric Feature Based Methods and double threshold techniques were used to develop the model whereas classified was done using CNN pre trained models namely; VGG19, VGGG16 and Inception_V3. Model testing performance using vgg19, Inception_V3, vgg16 and scratch model on training was 94%, 93%, 89%, and 97% respectively where as the model performance on testing dataset was 91%, 90.3%, 82.43% and 90% respectively.

From literature, a number of face recognition machine learning models have been developed and trained using transfer learning models for example Wu, GuiLing [10] developed an algorithm for recognizing a masked face based on attention mechanism, Soniet et al. [12] developed a model that detects a person wearing an element, Kumar G and Shetty D [13] proposed a model detecting face mask using You Only Look Once (YOLO) object detection mechanism. However, these models were trained on images of white race which affects their implementation in African setting. Thus, in this study we have developed a scratch model, trained and tested it on images of black race making it suitable for African setting.

Acknowledgment

The authors acknowledges Kabale University for providing a computer laboratory where the study experiments were carried out and the Internet service used in conducting research. In a special way the authors acknowledge students from Faculty of Computing, Library and Information Sciences and staff members from computing services department of Kabale University for their voluntary participation in this study.

References

[1] H. Rathod, Y. Ware, S. Sane, S. Raulo, V. Pakhare, I. A. Rizvi, Automated attendance system using machine learning approach, in: 2017 International Conference on Nascent Tech-nologies in Engineering (ICNTE), IEEE, 2017, pp. 1–5.
[2] M. Sharif, M. Y. Javed, S. Mohsin, Face recognition based on facial features, Research Journal of Applied Sciences, Engineering and Technology 4 (17) (2012) 2879–2886.
[3] L. Arnold, S. Rebecchi, S. Chevallier, H. Paugam-Moisy, An introduction to deep learning, in: European Symposium on Artificial Neural Networks (ESANN), 2011.
[4] A. Mathew, P. Amudha, S. Sivakumari, Deep learning tech-niques: an overview, Advanced Machine Learning Technolo-gies and Applications: Proceedings of AMLTA 2020 (2021) 599–608.
[5] V. S. Dhaka, S. V. Meena, G. Rani, D. Sinwar, M. F. Ijaz, M. Wozniak,´ A survey of deep convolutional neural networks applied for prediction of plant leaf diseases, Sensors 21 (14) (2021) 4749.
[6] H. N. Vu, M. H. Nguyen, C. Pham, Masked face recognition with convolutional neural networks and local binary patterns, Applied Intelligence 52 (5) (2022) 5497–5512.
[7] B. Mandal, A. Okeukwu, Y. Theis, Masked face recognition using resnet-50, arXiv preprint arXiv:2104.08997 (2021).
[8] S. A. Sanjaya, S. A. Rakhmawan, Face mask detection using mobilenetv2 in the era of covid-19 pandemic, in: 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), IEEE, 2020, pp. 1–5.
[9] S. Ge, J. Li, Q. Ye, Z. Luo, Detecting masked faces in the wild with lle-cnns, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2682– 2690.
[10] G. Wu, Masked face recognition algorithm for a contactless distribution cabinet, Mathematical problems in engineering 2021 (2021) 1–11.
[11] W. Hariri, Efficient masked face recognition method during the covid-19 pandemic, Signal, image and video processing 16 (3) (2022) 605–612.
[12] A. Soni, A. P. Singh, Automatic motorcyclist helmet rule violation detection using tensorflow & keras in opencv, in: 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), IEEE, 2020, pp. 1–5.
[13] G. S. Kumar, S. D. Shetty, Application development for mask detection and social distancing violation detection using convolutional neural networks., in: ICEIS (1), 2021, pp. 760– 767.
[14] W. Valenzuela, J. E. Soto, P. Zarkesh-Ha, M. Figueroa, Face recognition on a smart image sensor using local gradients, Sensors 21 (9) (2021) 2901.
[15] T. X. Tee, H. K. Khoo, Facial recognition using enhanced facial features k-nearest neighbor (k-nn) for attendance sys-tem, in: Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications, 2020, pp. 14–18.

[16] B. Anand, P. K. Shah, Face recognition using surf features and svm classifier, International Journal of Electronics Engi-neering Research 8 (1) (2016) 1–8.
[17] M. Harris, A. Qi, L. Jeagal, N. Torabi, D. Menzies, A. Ko-robitsyn, M. Pai, R. R. Nathavitharana, F. Ahmad Khan, A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis, PloS one 14 (9) (2019) e0221339.
[18] R. M. Haralick, L. G. Shapiro, Image segmentation tech-niques, Computer vision, graphics, and image processing 29 (1) (1985) 100–132.
[19] M. Sewell, Principal component analysis, University College London: London, UK (2008).
[20] T. Kurita, Principal component analysis (pca), Computer Vision: A Reference Guide (2019) 1–4.
[21] D. Ghimire, J. Lee, Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines, Sensors 13 (6) (2013) 7714– 7734.
[22] V. V. Starovoitov, D. Samal, A geometric approach to face recognition. in: NSIP, 1999, pp. 210–213.