HIC-DEEP: A Hierarchical Clustered Deep Learning Model for Face Mask Detection

Olugbenga S. Olukumoro[1], Folurera A. Ajayi[1], Adedeji A. Adebayo[1], Al-Amin B. Usman[1], Femi Johnson [2]

[1]*Computer Technology Department, Yaba College of Technology, Yaba, Lagos, Nigeria*

[2]*Computer Science Department, Federal University of Agriculture, Abeokuta, Ogun, Nigeria*

***Abstract*: The use of face masks is apparently not strange in these present days as conceptualized in the past due to the emergence of the Pandemic Covid-19 Corona virus. As part of the non-clinical preventive measures for the spread of this virus is the prescription and proper usage of face mask by the World health organization (WHO). In lieu of this, heads of organizations, directors of industries and individuals have adopted the "*No facemask, no entry" policy* in varieties of designs placed at their door posts. The state of the arts technologies has also been developed to help detect face mask non-compliant users. Whereas, the use of non-supervised machine learning approach for classifying and detecting Covid-19 facemask compliant users is not widespread. In this paper, HIC-DEEP (an un-supervised machine learning) model is proposed using a pre-trained InceptionV3 network for Kaggle database Image features vector extraction for subsequent computations of Euclidian, Spearman, and Pearson distance matrixes. The Hierarchical clustering method is then activated to identify face mask wearing faces from defiant faces. The distance algorithms all returned a perfect precision rate of 100% for the identification of faces with no face masks while an accuracy of 60%, 78% and 85% are achieved by Spearman, Pearson and Euclidian respectively for the cluster with full face mask compliance. However, the Euclidian distance algorithm returned the best overall accuracy in terms of the distance matrix with data points grouped along close proximities with unique clusters**

***Keywords*:Face mask, InceptionV3, Deep learning, Euclidian distance, Spearman distance, Pearson distance, Covid-19**

I. INTRODUCTION

In the era of pandemics, efforts are geared towards flattening the curve as a proactive measure to either avoid community spread or taming the tide. The advent of the Covid-19 pandemic moreover has influenced several of such measures to reduce the vulnerability of people and as well reduce the stretch already placed on medical facilities across the globe [13, 15]. One of such non-pharmaceutical approaches towards safeguarding the lives of predisposed citizens is the mandatory use of face masks [18] by citizens of the world. Despite widespread compliance rate, non-compliance is likewise recorded which greatly slowed down global efforts towards flattening the curve of the virus. While non-compliance gains popularity, the entire world continues to convulse like a child under the cold hands of the ravaging pandemic. The complying public also had its fears [12] but the rate of non-compliance defies any logical reason as to why such a preventive measure could be discouraged.

Consequent upon the foregoing, several studies have made attempts at detecting non-compliance of face mask wearing using different approaches thereby returning varying degrees of state-of-the-arts. It is pertinent to state that the artificial intelligent approach is popular either as an expert system [5] or a real time system [11]. In all, the deep learning variant of machine learning is popularly employed either through the transfer learning approach [10] by a pre-trained model, or through base learners and ensemble learning [11, 6, 16].

However, the aforementioned and several others are identified as supervised learning approaches of machine learning. Relatively few studies have adopted the non-supervised approach of machine learning hence the need to avail the research domain the functionalities of clustering. Therefore, this study proposes HIC-DEEP, a hierarchical clustering approach for the identification of faces with zero compliance with face mask usage regulation. The public data set acquired from Kaggle is subjected to feature extraction through transfer learning from pre-trained InceptionV3 deep learner. The Euclidian distance, Spearman distance and Pearson distance measures were computed on the feature vectors returned from the image embedding effort of InceptionV3 to aid the clustering effort. The rest of the study is presented in the following. Section II reviewed existing literature on the detection of face mask compliance while section III presents the methodological framework of the study. The result is discussed in section IV while conclusion and recommendation for future studies are presented in section V.

II. RELATED WORK

Several studies have returned state-of-the-arts in a concerted effort to fight the dreaded Covid-19 pandemic. Depending on the research questions, studies have targeted different aspects of the fight against the pandemic either as detecting fake Covid-19 tweets [12], or developing Artificial Intelligence and machine learning models to aid in flattening the curve [18, 4, 9].

Other attempts include meta-analysis of supposed myths or misconceptions about the dreaded virus. However, this section reviews existing methodologies on the detection of face mask usage and their state-of-the-arts. In [3], a transfer learning model is employed towards identification of face mask usage. The pre-trained InceptionV3 is fine-tuned and tested on the

simulated masked face dataset. The model outperformed other models with a 99.9% and 100% accuracy rates at training and testing phase. Some researchers [3] developed a convolutional neural network-based architecture for detecting improper usage of face masks with a two-stage architecture. The study deploys NASNetMobile, Dense Net121, and Mobile NetV2 models on a training set size of 3440 (masked) and 4415 (no mask) classes. Experimental result returns NASNetMobile with the highest training and validation accuracies of 99.82% and 99.45% respectively. The Dense Net121 however returned highest accuracy for test phase with 99.49 weighted average.

A real time face mask detection system with OpenCV Deep Neural Network was the main thrust of the study performed by researchers [11]. TensorFlow, MobileNetV2, and Keras architectures were employed as image classifiers. The proposed model outperforms popular deep learners including LeNet-5, VGG-16, AlexNet, and ResNet-50 with 0.9264 weighted average accuracy and a F1 score of 0.93. Loey [10] proposed deep hybrid and predictive analytics for face mask detection. Two components of feature extraction using ResNet50 and classification using Support Vector machine (SVM), Decision trees, and an ensemble model. Three training sets of Labeled Faces in the Wild (LFW), Real-World Masked Face Dataset (RMFD) and the Simulated Masked Face Dataset were employed in the study. The SVM achieved 99.64% accuracy with RMFD while achieving 100% testing accuracy with LFW.

A single stage detector of face mask towards assisting the control of the dreaded Covid-19 was implemented in another study [5]. The study used the MAsked Faces (MAFA) dataset by re-annotating it to produce Masked Faces for Face Mask Detection (MAFA-FMD). The new set contains three classes of 'no mask wearing', 'correct mask wearing', and 'incorrect mask wearing'. A Context Attention Module (CAM) is likewise introduced by the study for feature extraction purposes, and then employed transfer learning for the modelling.

An optimistic convolutional neural network was used to detect face mask compliance [16]. The Kaggle dataset was fitted by the deep learner with a training accuracy/loss curve performance evaluation approach, while YoloV5 was proposed as a deep learning model for face mask detection [6]. A performance accuracy of 96.5% was achieved on the deep learning model created with 300 epochs over the 20, 50, 100, and 500 initial attempts.

III. METHODOLOGY

Proposed methodological framework for this study, as depicted in Fig.1 comprises four stages of image acquisition, image embedding, calculation of image vector distance, and the hierarchical clustering. The four stages would return the clustering of input images into the two classes of 'used nose mask' and 'does not use nose mask'. The evaluation of the distance algorithms**,** including the Euclidian, Spearman, and Pearson will be computed on the feature vectors extracted from each image by InceptionV3 of the image embedding phase.

Image Acquisition

Images of faces of people wearing face mask and those not wearing face masks makes up the input images for this study. The dataset is made up of four folders comprising faces of people that are fully covered with face mask according to WHO standard, not-covered, not face, and partially covered. The public data set is made up of 1451 facial images with a usability rate of 9.38 and samples of fully covered and not covered as presented in Fig. 2. The dataset nets 3478 views with the highest rate of download in the first quarter of 2022. The distribution of the input images across the two classes is presented in Table 1.

Inception-v3 Network Based Feature Extraction

Images are stored in the computer in the form of numbers and the vector space describes each image by its image vector. The image vectors are features that uniquely identifies each image input which is achieved by word embedding. Pre-trained convolutional deep learners are employed through transfer learning for the image embedding phase, which is achieved by the InceptionV3 architecture. InceptionV3 is Google's deep neural network reputable for image recognition and pre-trained on ImageNet data. The activations of the penultimate layer of the model is used to represent the input images as vectors. InceptionV3 is actually an improvement on V1 and V2 which incorporates batch normalization layers to streamline training, including the addition of factoring convolutions with larger spatial filters. Computational efficiency is also improved by the network [7] which will eventually return 2048 features vectors on each input image for further analysis. There are Inception A, Inception B, and Inception C inception modules in the InceptionV3 network as shown in Table 2, from left to right. They are well designed modules that generate discriminative features and thereby reduce parameter numbers. Each of the three modules is made up of convolutional and pooling layers which are parallel to one another. The 3x3, 1x3, 3x1, and 1x1 small convolutional layers ensures reductions in the number of parameters. The Conv. Refers to the convolutional layer, the pool refers to the pooling layer, and Fc means a fully connected layer. The layers prior to Fc forms the base network of the InceptionV3 while the patch size is the kernel size of the connected layer, convolutional and the pooling layers. The gap amidst two operations is the stride and the input size denote the size of the feature map input into the layer, while the output size of individual layer is the input size of the next layer.

Table 1: Distribution of input image faces across the two classes

Image description	Image Input ID	Class category	
		Fully Covered	Not Covered
Face mask ON	fc	fc1	-

Face mask ON	fc	fc2	-
Face mask ON	fc	fc3	-
Face mask ON	fc	fc4	-
Face mask ON	fc	fc5	-
Face mask NOT ON	nc	-	nc1
Face mask NOT ON	nc	-	nc2
Face mask NOT ON	nc	-	nc3
Face mask NOT ON	nc	-	nc4
Face mask NOT ON	nc	-	nc5

A. Distance Measure Computation

The 2048-no feature vectors extracted from each input image from the previous phase uniquely represents each image inputs of both fully covered or not covered faces. The ultimate identification of each image input, and their classification into either of fully covered or not covered classes would require some methodology for computing the distance, or in some cases, the dissimilarity between each pair of observations. The result will yield the distance matrix or information. The distance matrix computations is achieved in this study by a comparative analysis of three distance measures including the Euclidian measure, the Spearman measure and the Pearson measure. The distance measures determine how the similarity or dissimilarity of two distinct input images is calculates thereby influencing the shape of the resulting cluster. The input images already represented by vectors, through InceptionV3 embedding, thereby constitute a point in an n-dimensional space.

1) The Euclidian Distance Matrix

The Euclidean distance represents the distance between any two points within a n-dimensional space hence could compute the distance between two points of different images. The Euclidian distance is therefore the length of a line segment between two points and is computed by the Cartesian coordinates of the various points using the Pythagorean theorem. The concept shows that the distance between any two points is the absolute value of the numerical difference of the coordinates. Therefore, if IMx and IMy are two points within the 2048 vector space, the distance between is computed by:

$$d_{euc}(IMx, IMy) = \sqrt{(IMx - IMy)^2} \quad (1)$$

2) The Spearman distance matrix

The spearman method computes the correlation between the rank of IMx and the rank of IMy variables thus:

$$d_{spear}(IMx, IMy) = 1 - \frac{\sum_{i=1-n}((IMx_i' - IMx')(IMy_i' - IMy'))}{\sum_{i=1-n}((IMx_i' - IMx')^2 \sum_{i=1-n}(IMy_i' - IMy')^2)}$$

$$Mx_i' = rank(IMx_i) \text{ and } IMy_i' = rank(IMy)$$

Pearson distance matrix

The Pearson correlation measures the extent of a linear relationship between two data points within a vector space. It is a correlation-based distance measurement technique which is reputable for gene expression for data analytics. Correlation-based approach is often computed by subtracting a computed correlation coefficient from 1 and is computed as:

$$D_{pears}(IMx, IMy) = 1 - \frac{\sum_{i=1-n}((IMx_i - IMx)(IMy_i - IMy))}{\sqrt{\sum_{i=1-n}((IMx_i - IMx')^2 \sum_{i=1-n}(IMy_i - IMy)^2)}} \quad (3)$$

$$D_{pears}(IMx, IMy) = 1 - \frac{\sum_{i=1-n}((IMx_i - IMx)(IMy_i - IMy))}{\sqrt{\sum_{i=1-n}((IMx_i - IMx')^2 \sum_{i=1-n}(IMy_i - IMy)^2)}} \quad (3)$$

$$D_{pears}(IMx, IMy) = 1 - \frac{\sum_{i=1-n}((IMx_i - IMx)(IMy_i - IMy))}{\sqrt{\sum_{i=1-n}((IMx_i - IMx')^2 \sum_{i=1-n}(IMy_i - IMy)^2)}} \quad (3)$$

Hierarchical clustering

Hierarchical cluster analysis is an algorithm that groups similar objects into clusters where each cluster is distinct from each group such that each cluster are similar to each other. The algorithm is an unsupervised machine learning technique that identifies distance patterns from data points. In this study, the distances as computed by Euclidian, Spearman, and Pearson are employed by the dendrogram of the Orange data mining tool to cluster each image input into their respective class of either 'fully covered' or 'not fully covered'. The average attribute is set as the linkage while image name identifies each input. A maximum depth of 10 is set for the dendrogram pruning, with a 75.0% height ratio for selection.

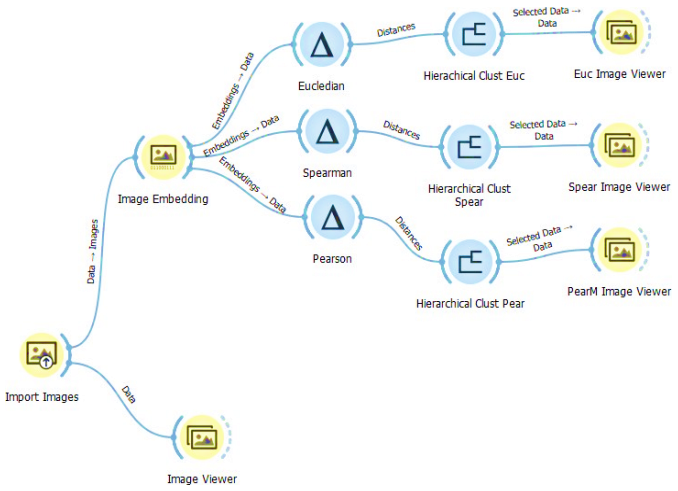

Fig. 1: Study framework as implemented in Orange data mining toolkit

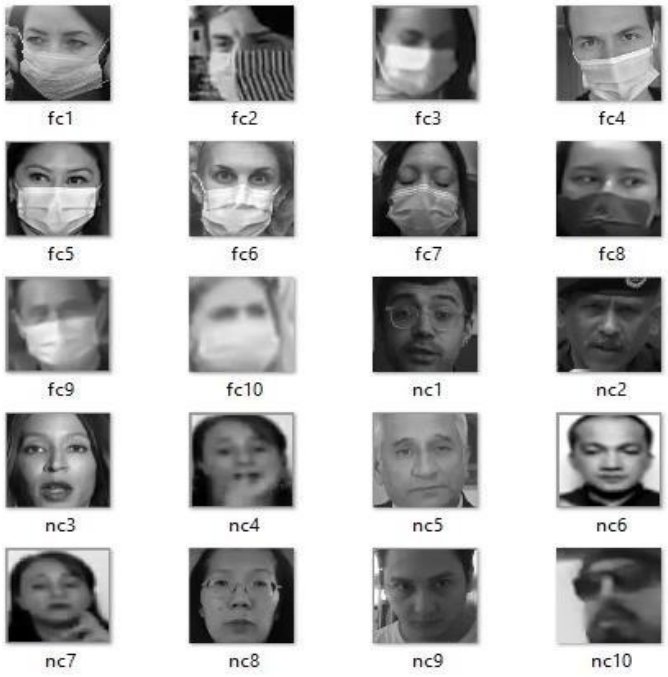

Fig. 2: Sample dataset from the Kaggle repository

Table 2: InceptionV3 Network Features

Layer	Patch size/stride	Input size
Conv	3x3/2	224x224x3
Conv	3x3/1	111x111x32
Conv padded	3x3/1	109x109x32
Pool	3x3/2	109x109x64
Conv	3x3/1	54x54x64
Conv	3x3/2	52x52x80
Conv	3x3/1	25x25x192
Inception Ax3	-	25x25x288
Inception Bx5	-	12x12x768
Fc	51,200x1,024	5x5x2,048
Fc	1,024x1,024	1,024
Softmaz	Classifier	4

IV. RESULT AND DISCUSSION

A total of 2048 feature vector attributes are extracted from each 20-no input images by the InceptionV3 convolutional network. The dendrogram representing the hierarchical clustering of the image inputs, with respect to the computed distance matrix by Euclidian distance, Spearman distance Pearson distance. As observed from the experimental results, three clusters of C1, C2, and C3 are identified by the hierarchical clustering for Euclidian and Spearman respectively whereas the Pearson distance matrix returned two clusters of C1 and C2. The first two biggest clusters of C2 and C3 for Euclidian and Spearman are presented in Fig. 3 and Fig. 6 respectively. The third unitary cluster C1 is observable from the aforementioned figures. The dendrogram showcasing the bigger cluster in C2 and C3 for Pearson is presented in Fig. 9. The corresponding image clusters for Euclidian C2 and C3 is presented in Fig. 4 and Fig. 5, while the image clusters for Spearman C2 and C3 is likewise presented in Fig. 7 and Fig. 5. The Pearson image clusters for C2 and C3 is as presented in Fig. 7 and Fig. 10 respectively. As could be observed from Table 3, experimental result returns input image *fc2* as the unitary C1 cluster that is mutual to both Euclidian and Spearman, representing the second input image with fully covered face mask. The input images fc1, fc4, fc5, fc6, and fc7 are mutual to Euclidian and Spearman as belonging to the same cluster C2, with close proximity to *fc2* in C1, and which are correctly grouped together as inputs with fully covered faces. The proximity of cluster C2 to C1 for Euclidian and Spearman, in terms of distance matrix, is ascertained by the encapsulation of *fc2* with the C2 as computed by Pearson distance matrix algorithm.

In essence, the three distance measurement matrixes correctly clusters fc1, fc2, fc4, fc5, fc6, and fc7 as inputs with fully covered faces representing a 60% recognition rate for fully covered faces with face masks. The outstanding fc3, fc8, fc9, and fc10 are clustered alongside the 'not covered' faces by the three distance measures with their image clusters presented in Fig. 11 for Euclidian, Spearman and Pearson respectively. For Euclidian, and as noticed in Fig. 11, the dark background exclusive to fc8, nc1, nc10, nc2 may be responsible for their close proximities in the vector space as computed by InceptionV3. The Spearman's distance matrix clustering follows the same pattern with that of the Euclidian, as may be observed on the image cluster in the second row presented in Fig. 11. However, the Pearson distance matrix exactly shares similarity with Spearman in the categorization of the outstanding fc family. The input 'nc' image family are however correctly clustered into the 'not covered' category with 100% precision rate across the Euclidian, Spearman and Pearson distance matrix indices.

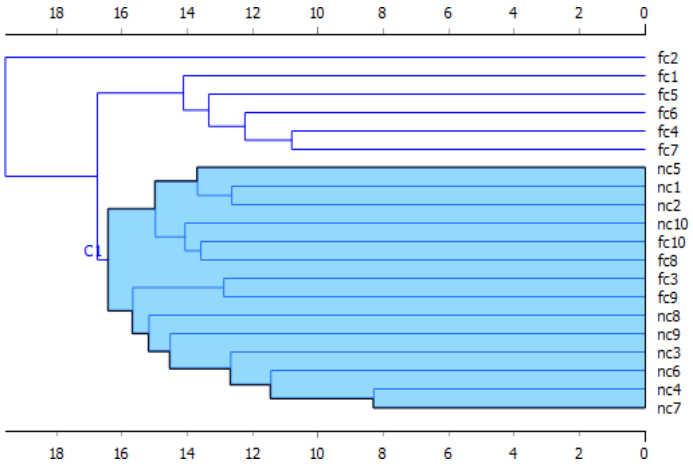

Fig. 3 Dendrogram showing Euclidian biggest cluster

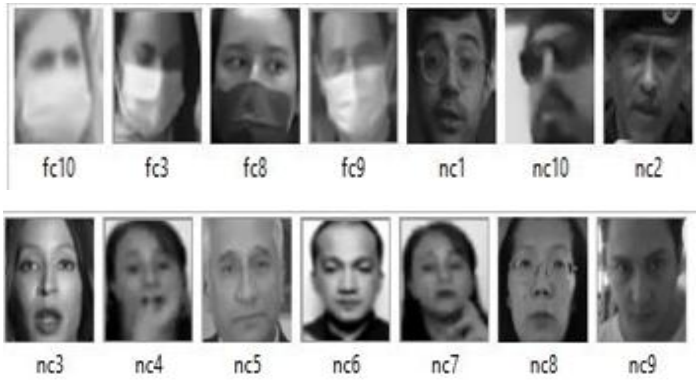

Fig. 4 Image inputs belonging to the Not_Covered Euclidian cluster

Fig. 5: Image inputs belonging to the Face_Covered Euclidian and Spearman clusters

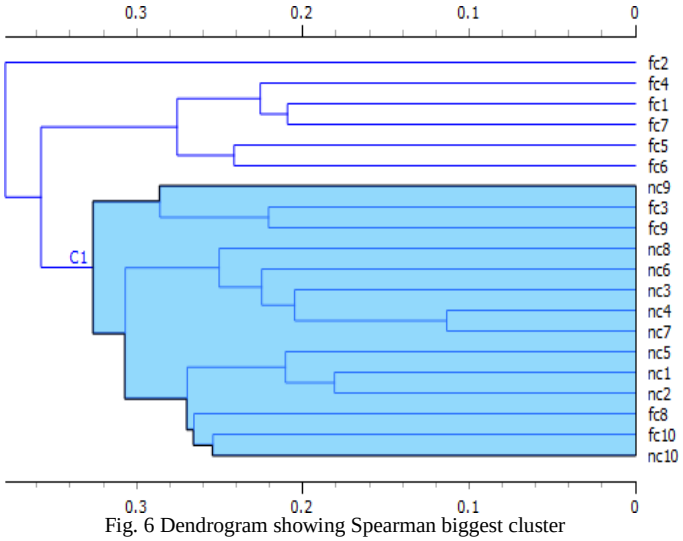

Fig. 6 Dendrogram showing Spearman biggest cluster

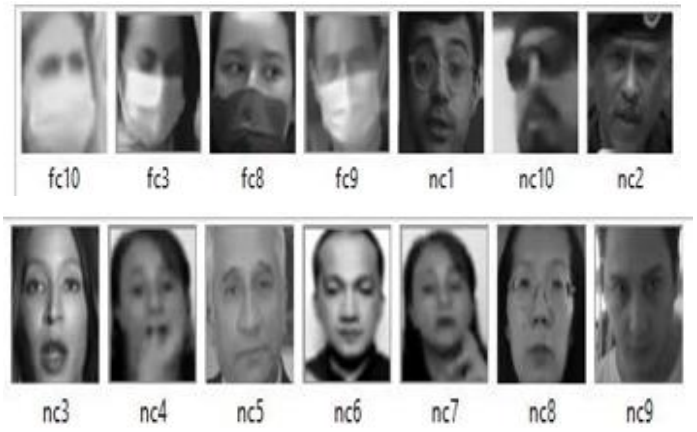

Fig. 7 Image inputs belonging to the Not_Covered Spearman and Pearson cluster

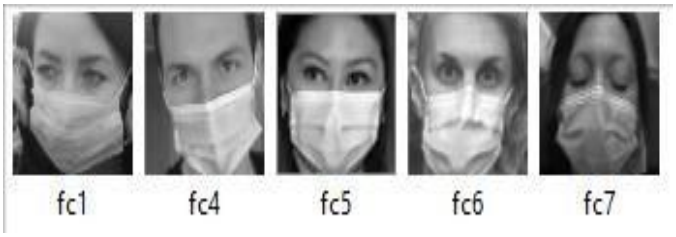

Fig. 8. Image inputs belonging to the Face_Corvered Spearman cluster

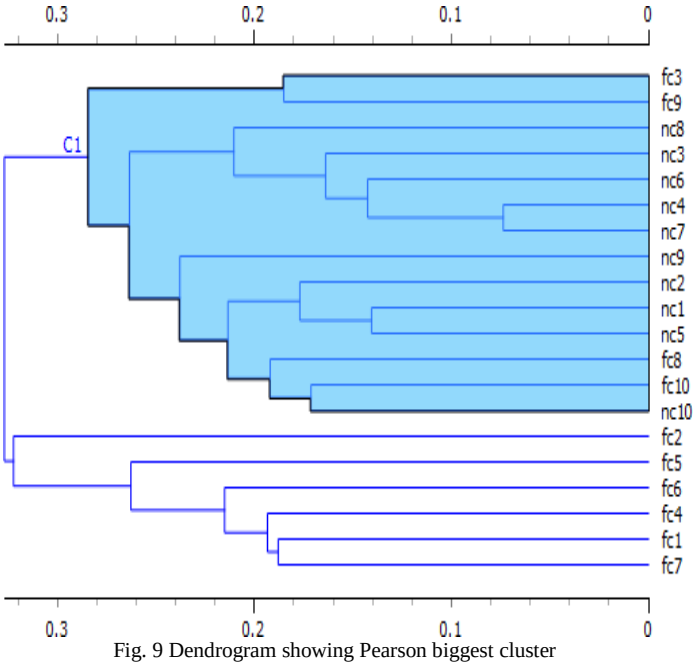

Fig. 9 Dendrogram showing Pearson biggest cluster

Fig. 10. Image inputs belonging to the Face_Covered Pearson cluster

Table 3: Clustering results across the distance matrix

Clusters	Euclidian	Spearman	Pearson
C1	fc2	fc2	-
C2	fc1	fc4	fc2
	fc5	fc1	fc5
	fc6	fc7	fc6
	fc4	fc5	fc4
	fc7	fc6	fc1
	-	-	fc7
C3	nc5	nc9	**fc3**
	nc1	**fc3**	**fc9**
	nc2	**fc9**	nc8
	nc10	nc8	nc3
	fc10	nc6	nc6
	fc8	nc3	nc4
	fc3	nc4	nc7
	fc9	nc7	nc9
	nc8	nc5	nc2
	nc9	nc1	nc1
	nc3	nc2	nc5
	nc6	**fc8**	**fc8**
	nc4	**fc10**	**fc10**
	nc7	nc10	nc10

Fig. 11 Row-wise misclassification of not covered faces by Euclidean, Spearman and Pearson's metrics.

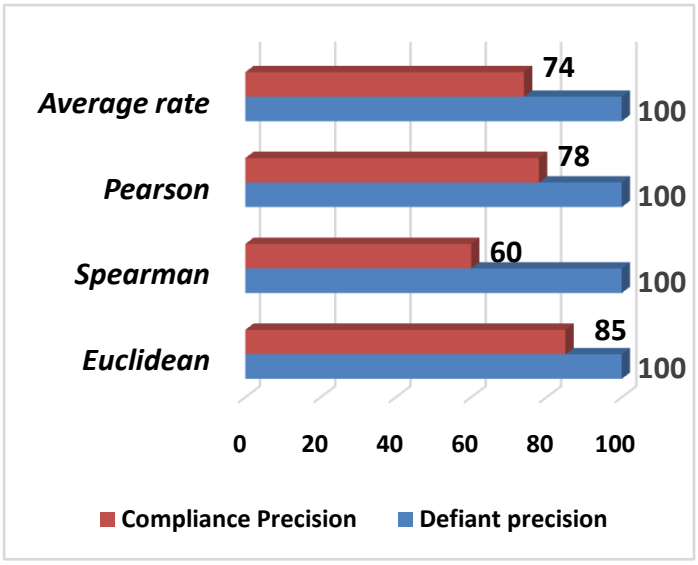

Fig. 12 Comparison chart of HIC-DEEP distance – Metrics Algorithms

V. CONCLUSION

The study seeks to identify faces of individuals without the use of face mask as obligated by the WHO towards the fight against the dreadful Covid-19. An unsupervised machine learning model (HIC-DEEP) is proposed through the hierarchical clustering of input images acquired from the public domain containing faces of compliant and defiant instances. The pre-trained InceptionV3 convolutional network is used to extract 2048 feature vectors from each input image, and the vector space is passed through three distance computing metrics (the Euclidian, Spearman and Pearson algorithms) for the subsequent clustering into the two categories of either face mask compliant or defiant. Experimental result shows the each distance metric could accurately identify faces of face mask defiant with 100% precision while an average of 74% precision rate is achieved on the compliant human faces. The result likewise shows Spearman and Pearson adopts similar pattern recognition approach while the Euclidian distance measure indeed clustered faces with distance measures into close proximities. Future studies could employ a bigger image sets as well as compare its result with a supervised approach.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

REFERENCES

[1]. Bussan, D.D.; Snaychuk, L.; Bartzas, G.; Douvris, C., (2021). Quantification of trace elements in surgical and KN95 face masks widely used during the SARS- COVID-19 pandemic. Sci. Total Environ., 814, 151924.

[2]. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning, Vieena, Austria, 12–18 July 2020; pp. 1597–1607.

[3]. Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). Face Mask Detection Using Transfer Learning of

InceptionV3. International Conference on Big Data Analytics (pp. 81-90). Springer.

[4]. De Sio, L.; Ding, B.; Focsan, M.; Kogermann, K.; Pascoal-Faria, P.; Petronella, F.; Mitchell, G.; Zussman, E.; Pierini, F. Personalized Reusable Face Masks with Smart Nano-Assisted Destruction of Pathogens for COVID-19: A Visionary Road. Chem. A Eur. J. 2020, 27, 6112–6130.

[5]. Fan, X., & Jiang, M. (2020). RetinaFaceMask: A Single Stage Face Mask Detector.arXiv:2005.03950v3, PP. 1-6.

[6]. Ieamsaard, J., Charoensook, S. N., & Yammen, S. (2021). Deep learning-based face mask detection using yoloV5. 2021, 9th International Electrical Engineering Congress (iEECON) , pp. 428-431). IEEE.

[7]. Jan, B.; Farman, H.; Khan, M.; Imran, M.; Islam, I.U.; Ahmad, A.; Ali, S.; Jeon, G. Deep learning in big data analytics: A comparative study. Comput. Electr. Eng. 2019, 75, 275–287

[8]. Jiang, M.; Fan, X.; Yan, H. Retinamask: A face mask detector. arXiv 2020, arXiv:2005.03950.

[9]. Latif, S.; Usman, M.; Manzoor, S.; Iqbal,W.; Qadir, J.; Tyson, G.; Castro, I.; Razi, A.; Boulos, M.N.K.;Weller, A. Leveraging data science to combat COVID-19: A comprehensive review. IEEE Trans. Artif. Intell. 2020, 1, 85–103.

[10]. Loey, M., Manogaran, G., N.Taha, M. H., & M.Khalifa, N. E. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Elsevier Journal of Measurement, 1-11.

[11]. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., & Hemanth, J. (2021). SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable Cities and Society.

[12]. Olaleye, T., Abayomi-Alli, A., Adesemowo, K., Arogundade, O. T., Misra, S., & Kose, U. (2022). SCLAVOEM: hyper parameter optimization approach to predictive modelling of COVID-19 infodemic tweets using smote and classifier vote ensemble. Springer Berlin Heidelberg, 1-20.

[13]. Peishu Wu, Han Li, Nianyin Zeng, Fengping Li,: FMD-Yolo (2022): An efficient face mask detection method for COVID-19 prevention and control in public, Image and Vision Computing, Volume 117, 104341,ISSN 0262-8856,https://doi.org/10.1016/j.imavis.2021.104341.

[14]. Qin, B.; Li, D. Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors 2020, 20, 5236

[15]. Shahzad, Y.; Javed, H.; Farman, H.; Ahmad, J.; Jan, B.; Nassani, A.A (2021). Optimized Predictive Framework for Healthcare through Deep Learning. Comput. Mater. Contin, 67, 2463–2480.

[16]. Suresh, K., Palangappa, M., & Bhuvan, S. (2021). Face Mask Detection by using Optimistic Convolutional Neural Network. 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1084-1089). IEEE.

[17]. Yar, H.; Abbas, N.; Sadad, T.; Iqbal, S.(2021). Lung Nodule Detection and Classification using 2D and 3D Convolution Neural Networks (CNNs). Artif. Intell. Internet Things 2021, 365–386.

[18]. Yadav, S. Deep learning based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence. Int. J. Res. Appl. Sci. Eng. Technol. 2020, 8, 1368–1375.

[19]. Zhu, N.; Zhang, D.; Wang,W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi,W.; Lu, R. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 2020, 382, 727–733.