A Deep Learning Convolutional Neural Network Approach of Using ResNet-18 for 2D-Image Classification of Faces to Indian or Pakistani.

Hasnain Shaukat
200020213
Supervisor: Hazar Emre Tez
Data Science & Artificial Intelligence (Conversion)
Queen Mary University of London

Abstract—

Advancements in computer vision technology have resulted in the application of machine learning to a wide range of fields, including facial recognition, which has attracted a lot of attention. The use of machine learning and deep learning for ethnic/nationality classification using faces has been investigated in several studies, however, these studies have mostly utilized population with two very distinct facial features.

Using the architecture of residual neural networks, we present a deep convolutional neural network for the task of 2D gray-scale image-classification of two similar appearing faces of Indian and Pakistani. The experiment was conducted on a data-set using ResNet-18 and SoftMax as classifier we achieved an overall accuracy of 92%

Keywords—Deep Learning, Neural Network, Ethnicity/Nationality, Classification, Artificial Intelligence, Pakistani Face, Indian Face, Face Recognition

I. INTRODUCTION

Over the last few decades, numerous studies have been conducted in biology, psychology, and cognitive sciences to determine how the human brain observes, processes, and recognizes a face. A number of soft biometric features can be identified when examining a human face, including age, gender, expression, and mood, as well as the demographic location of the individual's background, namely his or her nationality/ethnicity (Lu, Lu and Jain, 2004).

There is evidence that the human population originated approximately 200,000 years ago (Stringer and Andrews, 1988) in Africa. From there, they began migrating to other continents approximately 60,000 years ago. As a result of this migration, a number of settlements in various geographical locations came into existence which faced a wide variety of climate, diseases, and genetic mutations which resulted in the diversity of human population (Mersha and Abebe, 2015).

There is also evidence that genes linked to the facial structure vary more than those linked to DNA in other parts of the body, indicating that evolution has selected for a variety of facial structures (National Geographic, 2014). Nevertheless, classification of human faces based on ethnicity and nationality remains a challenging task for humans and computers alike, especially for countries that appear visually similar, such as Chinese and Japanese.

It was in 1947 that the Indian subcontinent was partitioned, resulting in the formation of two independent countries, India and Pakistan (Lahore Resolution, 2003). While it may appear that Indians and Pakistanis share a similar facial appearance at first glance, these countries are populated by large interethnic groups. Pakistan's population is comprised of a variety of ethnic groups, including Saraiki, Pashtun, Balochi, Punjabi and Sindhi (*Ethnic Groups In Pakistan*, 2019), just as the population of India consists of a variety of ethnic groups, including Marathi, Gujratis, Rajasthanis, Punjabis, Bengalis, Bengalis etc. (Britannica, 2022). In these two countries, the inter-ethnic groups are diverse primarily because of regional and cultural differences, this results in a variety of different facial characteristics.

With the advancements in computer vision technology and the application of machine learning and deep learning in recent decades, several studies have focused on the classification of ethnicity/nationality using facial features. In the majority of these studies, there are two significant issues. First, the studies have concentrated on two very distinct ethnic groups, which are easily distinguished by an average person such as Face of Mongolian vs Negro (Masood *et al.*, 2018). Second, these studies have focused primarily on Caucasians, Mongolians, Japanese, Koreans, and Africans, resulting in a significant underrepresentation of Pakistani and Indian populations.

A major objective of this project is to address the limitations of existing studies by utilizing one of the most advanced and most advanced deep learning techniques, the residual neural network architecture (ResNet), to classify images of two populations whose faces appear very similar, Indians and Pakistanis, into two groups. ResNet-18 is used for this task. This results in a binary category problem, which is trained on a dataset of Indian and Pakistani ethnicities.

This paper has been designed in the following way in the first section of the paper an introduction to the topic is given, in section 2 of the paper a brief background about artificial intelligence is presented in section 3 a comprehensive literature review and related work section is presented and in section 4 of this paper a detailed methodology employed in this project is presented finally in section 5 a discussion on results is carried out and opportunities for future work is presented.

II. Background:

Artificial Intelligence is a field of computer science with the aim of creating a system that can perform tasks equivalent to human intelligence, The field of AI can be broken into different sub-sets, one such sub-set of AI is Machine Learning (Jakhar and Kaur, 2020).

Machine Learning is essentially giving computer systems the ability to learn and train using historical data to make future predictions, without being explicitly programmed for it. (Mitchell *et al.*, 1990).

There is also a subsection within Machine Learning called Deep Learning. The concept of deep learning is based on the idea of mimicking the functioning of neurons in the cerebral cortex of human brain (IBM, 2021). A human brain is made up of 86 billion neurons (Herculano-Houzel, 2009), these interconnected neurons create a network of nerves, In deep learning neurons are created artificially in computer and arranged in layers resulting in an Artificial Neural Network (ANN) (Panchal, 2018).

For the tasks related to visual representation a human brain uses the frontal lobe of brain called the visual cortex (Huff, Mahabadi and Tadi, 2022) similarly in deep learning a different type of neural network called Convolutional Neural Network (CNN) which has neurons arrange like a visual cortex is used (Meel, 2022).

An in-depth explanation of the underlying principles of CNN and other components of deep learning pipeline that are employed in this project is provided in the methodology section of this paper.

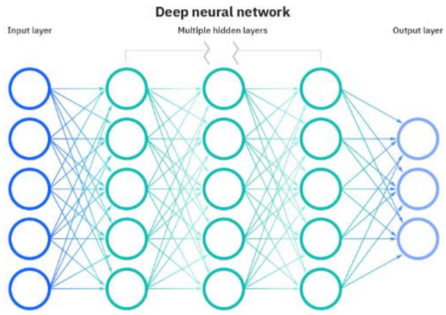

Figure 1 Deep Neural Network Structure (IBM, 2021)

III. Related Work: (Literature Review)

The research breakthrough in the field of deep learning came in 1990 when (LeCun, Bengio and Hinton, 1990) proposed hand written digits recognition using a back propagation network, The second breakthrough in deep learning came in 2012 when the advancements in GPU technology allowed the researcher (Krizhevsky, Sutskever and Hinton, 2012) to propose a deep convolutional network using GPU to increase training speed of neural networks (Foote, 2022). Following that Deep learning and image classification has become a hot topic in research, and deep learning has made many remarkable achievements in various computer vision fields (Al-Saffar, Tao and Talab, 2017).

In 2005 (Ou *et al.*, 2005) in their paper presented a program that could classify faces in a 2D image in real time, the program utilized traditional machine learning technique of principle component analysis (PCA) for feature extraction process and for the task of classification SVM (Support Vector Machine) combined with a novel combined with a novel algorithm '321' was used. The system achieved an accuracy of 82.5 % on the Facial Recognition Technology Database (FERET), however the program could only do a binary classification of faces as Asian or Non-Asian.

Ethnicity identification of two categories Asian vs Non-Asian was also presented by (Lu, Lu and Jain, 2004) using the traditional machine learning algorithms, in their paper they proposed a linear discriminant analysis an ensemble framework with integrated LDA for classification was used, Overall 2630 face images were used of 263 subjects with 132 subjects of Asian ethnicity and 131 subjects of Non-Asian ethnicity, hence resulting in a binary classification problem.

Facial classification between three ethnic groups of Asia vs European and African using traditional machine learning technique was done by (Hosoi, Takikawa and Kawade, 2004), They used combination of two algorithms Gabor Wavelet Transformation and Retina Sampling for the feature extraction process and for estimation of ethnicity a Support Vector Machine was used, the researchers reported an accuracy of 95.1%, 88.4% and 94.3 for Asian, European and African ethnicity the learning and evaluation was carried on grayscale images using HOIP and self-formed dataset.

(Guo and Mu, 2010) performed a study of ethnicity estimation on five categories; Black, White, Hispanic, Asian, Indian, and Other, ethnicities varied with gender and age, the researcher proposed using a biologically inspired feature (BIF) to extract the ethnic features from image and using Principal Component Analysis to reduce the dimensionality while using linear SVM (Support Vector Machine) as classifier.

From literature review we analyze that some work done in recent years have utilized the powerful concept of deep learning for facial recognition technology, as (Gudi, 2016) in their paper used a deep neural network on 2-D Images to identify the semantics of face, furthermore (Narang and Bourlai, 2016) in their paper have presented a gender and ethnicity classifier using deep learning, the proposed classifier uses slightly different type of images namely the Near infrared (NIR) long range night time face images, the images are captured outdoors at night time with standoff distances between the 30 to 120 meters, A Convolutional Neural Network of VGG architecture is used for classification, the researcher reported an accuracy of 78.98% and 61.83% for ethnicity classification of Asian and Caucasian groups at distance of 1.5m and 120m respectively.

Ethnic classification with deep learning of using 2-D grayscale images was recently done by (Wang, He and Zhao, 2016), three ethnical classifications were done in this paper Chinese vs Non-Chinese, Black vs White and Han Chinese vs Uyghurs people, the researcher used a popular neural network of CIFAR-10. Similarly using Convolutional Neural Network (Masood *et al.*, 2018) classified three distinct group of ethnicities Mongolian, Caucasian and Negro, images were taken from the FERET database, the paper proposed the model based on the Viola Jones machine learning algorithm to detect facial geometric features and color attributes followed by the classification using convolutional neural network based on the VGGNet 13 architecture, the network utilized a total of 357 images for training 120 for Caucasians and Mongolians and 117 for Negros, while for testing a total of 90 images were used, the model achieved a high accuracy of 82.4%.

One of the most recent works in ethnicity classification using deep learning was done by (Jilani *et al.*, 2019), the researcher conducted classification using deep learning architecture of ResNet-50, ResNet-101 and ResNet-152, a new novel database of 224 Pakistani and 239 Non-Pakistani participants was formed, the non-Pakistani participants belonged to different demographics such as White, Nigerian, Arab and Chinese, Egyptian and Poles (Jilani, 2020), the resulting problem was a binary classification for which 2D pre-processed images were used, the pre-processing involved transforming images to Gray-scale and removing any ornaments, the researcher conducted classification using only face feature of Eyes, Nose and Mouth reporting an accuracy of 85.9%, 91.5% and 94.8%, for classification using complete face accuracy of 99.2% was achieved.

(AlBdairi, Xiao and Alghaili, 2020) in their paper proposed a novel convolutional neural network to identify ethnicities, the researcher worked on three ethnicities of Chinese, Pakistani and Russian a dataset of 1000 images for each ethnicity were produced by them, the novel neural network proposed was inspired by the VGG Network and achieved a very high accuracy of 96.9% higher than the accuracy of VGG Network of 91.48%, however little to no pre-processing technique were applied on images.

Taking into account the comprehensive Literature review, we can identify two research problems: (1) under-representation of the subcontinent's population in ethnic classification using machine learning apart from Jilani, (2) lack of investigation using deep learning for ethnic/nationality classification of similar appearing facial populations.

In most of the work done above, the categories for ethnic classification have been handpicked, i.e., using two very distinctive ethnic groups such as Negro and Mongolian. This can be considered trivial from an Artificial Intelligence standpoint because a human brain is able to distinguish between these two groups very quickly and easily. Therefore, in this paper we present a novel classification of two very similar appearing nationalities Pakistani and Indian using deep learning.

IV. METHODOLOGY:

*A. **Anthropometry:***

Anthropometry is essentially a field of science that involves measurements of different human features such as size, form, functional capabilities etc. (*Anthropometry | NIOSH | CDC*, 2021). Similarly human faces are evaluated and measurements of different features of faces are recorded.

One of the most comprehensive work done in identifying the ethnicity of a person using facial features is done by (Farkas, Katic and Forrest, 2005), 14 anthropometric measurements were taken of 1470 subjects from four different regions namely European, Middle-East, Asia and Africa. The study concluded that nose height and width contrasted sharply in subjects, particularly the noses were extremely wide of Asian and Black ethnic group when compared to North American Whites, similarly Middle-Eastern nose was similar in width to North American Whites but significantly greater in height.

In order to identify the ethnicity of a face, the deep neural network will take advantage of the differences in the facial morphology that exist between the faces of Pakistani and Indian subjects.

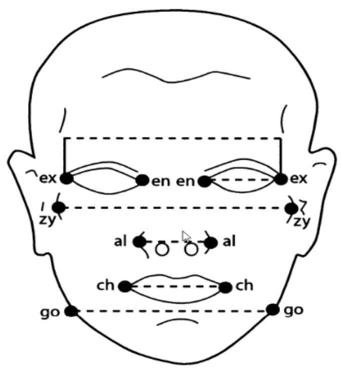

Figure 2: 14 Anthropometric Measurements Taken of Human Face *(Farkas, Katic and Forrest, 2005)*

B. *Image Database*:

This project used two-dimensional images from databases that were already available. For the Indian category, images were taken from the Chicago Faces Database (CFD-India) (Lakshmi et al., 2021). The database consists of 142 unique individuals with 144 photographs taken in RGB format and neutral facial expressions. This dataset was compiled in India and consists of individuals from a variety of inter-Indian ethnicities and geographical locations.

The 2-D images of the Pakistani category were also obtained from an existing database (AlBdairi, Xiao and Alghaili, 2020) which comprises three ethnic groups Pakistani, Chinese and Russian. From the database, a subset of images was taken for the Pakistani category.
The overall data set had to be split for training and validation, and in order to avoid an unbalancing in categories, the subset of 144 unique individuals were taken from Pakistani dataset and overall, 70% & 30% data splitting was applied for training and testing resulting in 100 pictures for training and 44 pictures for testing, furthermore the data augmentation processed increased the number of pictures to manifolds.

Figure 3: Raw Images of Pakistani Dataset

Figure 4: Raw Images of Indian Data-Set

C. *Image Processing:*

Image processing is one of the crucial steps within the deep learning framework. The first step in image pre-processing was done using Photoshop, firstly the background of all the Pakistani images was removed and a plain white background was installed. Secondly the Indian images were cropped to only include the facial area, finally a manual screening was performed on all images to remove any outlier images.

Image A **Image B** **Image C**

Figure 5: Image A: The Raw Pakistani Dataset Image, Image B Pre-Processed with Plain White Background Installed in Photoshop, Image C: Applied in Pytorch Conversion to Grayscale

The second step of image processing was done inside the Pytorch framework. The images were changed to a resolution of 100x100, transformed from colored images to Grayscale, furthermore a random rotation of 15 degrees was applied on all images to result an equivalent dataset, the final transformed images loaded into the Neural Network are shown below

Figure 6: Transformed Pakistani Images

Figure 7: Transformed Indian Images

*D. **Deep Learning Pipeline:***

A Deep Learning Pipeline for Image Classification is made up of 1) Convolutional Operation 2) Loss Function, 3) Optimizer; 4) Classifier 5) Model.

*1) **The Convolutional operation:***

As previously mentioned, neurons are the building blocks of neural networks. The fundamental component of a Convolutional Neural Network is the convolutional layer, which contains a kernel (or filter). Kernels consist of two trainable parameters, namely weights and biases. Upon receiving an image, a convolutional layer extracts features from the image and produces an output which is fed into the next layer (Mostafa and Wu, 2021). In the initial layers of the network, low level features are extracted, such as vertical and horizontal edges. As the layers become deeper, advanced features are extracted, such as noses, eyes, etc.

*2) **The Loss Function:***

A loss function is employed during the training phase of the network to calculate the error between the network's predicted outcome and the ground truth. In general, the goal of the neural network is to reduce the loss function to zero so that the predicted outcome is equal to the ground truth value (Brownlee, 2019).

In our network, the loss function used is the Cross Entropy Loss which is based on Information theory and is calculated using the equation,

$$H(p,q) = -\sum_{\forall x} p(x) \log(q(x)) \quad (1)$$

*3) **Optimizer: (Stochastic Gradient Descent):***

A Neural Network is trained using an optimizer which is an algorithm that updates the weight attributes of the neural network in order to reduce the loss value so that a more accurate prediction can be made (Gupta, 2021).

Stochastic Gradient Descents with Momentum (SGD) is used as the optimizer algorithm for our network. The term stochastic essentially means random hence each iteration uses a small batch of random images from dataset to calculate the loss value, which is then updating by computing its gradient; the momentum feature facilitates faster convergence of the loss value. The equation for calculating SGD is shown below (Sarwinda et al., 2021).

$$\Delta w_{ij} = -\eta \delta D/(\delta w_{ij}) \quad (2)$$

$$\Delta w_{ij}(t+1) = -\frac{\eta \delta E}{\delta w_{ij}} + \alpha \Delta w_{ij}(t) \quad (3)$$

*4) **Classifier: (SoftMax)***

Classifier is a final neural layer in the network, which transforms the output of previous neural layers to give us the predicted result, the classifier function used in our network is the SoftMax function. This function converts the outputs of previous neural layers into a simple probability distribution value between 0 to 1, where the sum of all probabilities is equal to 1 and the category with highest probability is the predicted outcome of network (Uniqtech, 2021) .

$$\boldsymbol{\sigma(z)_i} = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (4)$$

5) ***D: ResNet-18:***

The ResNet-18 also known as residual network, is a deep convolutional neural network architecture that was developed by (He *et al.*, 2016) and released by Microsoft Research team in 2016, The architecture is currently one of the latest deep learning techniques and is being implemented in majority of the new artificial intelligence technology (Maladkar, 2018), the network achieved state of the art accuracies on the ImageNet database competition (Krizhevsky, Sutskever and Hinton, 2012).

The genius of the network is based on something called 'skip connection' or shortcut, the network works by identifying a shortcut to update weights during the back propagation step allowing it to skip two convolutional layers,

this is achieved by adding the input feature to the output of ResNet block.

$$y = F(x, B + x) \quad (5)$$

As a result, the network is able to solve two critical issues with deep neural networks; reducing the large amount of time taken for training the network, and solving the issue of vanishing gradients for deep neural networks.
The number 18 of ResNet implies that model will have overall 18 neural network layers, the network consists of a stem, macro-module and a fully connected layer (Ruiz, 2019), the explanation of components and structure used is described below,

- A stem layer has a 3x3 Kernel and stride of 1, a Gray-scale image is used therefore input is a single channel and the total number of output channels by stem layer will be 62.
- One ResNet block consists of two 3x3 convolutional layers, two of these ResNet blocks are combined to form a macro-module.
- Overall, four macro-modules with increasing channel dimensions are used in the model,
- The output of the macro-modules is processed to one fully connected layer, hence the model results in total of 16 convolutional layers, 1 stem layer and 1 fully connected layer resulting in overall 18 layers.

The overall structure of the architecture is shown on the figure, and is described below

- A single colour represents one macro-module, all the convolution layers use a 3x3 kernel. The arrow represents the skip connection highlighting the addition of output from previous block.

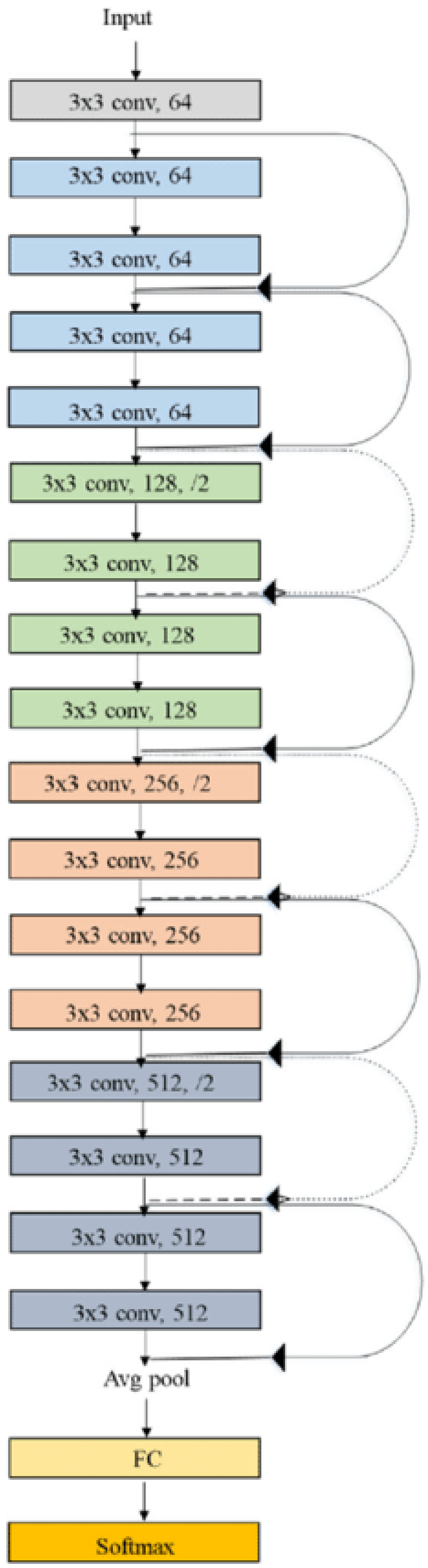

Figure 8: The ResNet 18 Structure Used (FC refers to Fully Connected Layer) Credits: *(Ramzan* et al., *2019)*

V. RESULTS & DISCUSSION:

A. *Evaluation Metrics:*

To quantify the quality of the deep learning program that has been developed, four parameters based on the confusion matrix are used to measure the program's performance. The evaluation metrics are measured for the testing dataset.

1) *Accuracy:*

The metric accuracy is measured based on the predictions made by the neural network and the ground truth value; accuracy measures the performance of the model across all classes. It is calculated using the following equation:

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

2) *Precision:*

In the classification problem of deep learning, precision is defined as the number of correct predictions made for one class divided by the total number of predictions made for that class; as shown in the equation below;

$$Precison = \frac{Correct\ Predictions\ For\ India}{Total\ Predictions\ Made\ OF\ India}$$

3) *Recall:*

Similar to precision we can define recall as the total number of correct predictions made for one class divided by the total number of possible predictions for that class; as shown in the equation below.

$$Recall = \frac{Correct\ Predictions\ For\ India}{Total\ For\ India\ Category}$$

4) *F1-Score:*

F1 Score is the harmonic average of precision and recall; and is calculated as

$$F1 = \frac{2x\ RecallxPrecision}{Recall + Precision}$$

Table 1: The Table Below Shows the Evaluation Values:

	Precision	Recall	F1-Score
Indian	0.93	0.89	0.91
Pakistani	0.89	0.93	0.91

Table 2: The Accuracy Achieved of Neural Network

Epoch No	Highest Training Accuracy	**Highest Testing Accuracy**
37		**92 %**
49	95.5%	

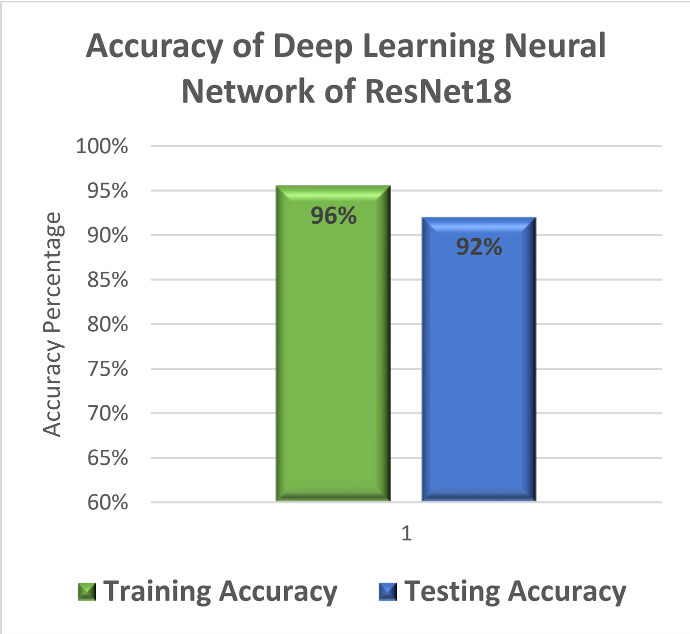

Figure 9: Graphical Representation of Training and Testing Accuracy

B. *Discussion:*

According to our experiment conducted for classifying Pakistani and Indian faces, the deep learning model performed very well as it achieved greater than 90% accuracy in testing. When comparing the results with the published literature of machine learning, we can conclude that the deep learning network outperformed the traditional machine learning techniques used, as it achieved a higher accuracy than (Ou et al., 2005) which used a feature extraction process based on traditional machine learning technique of PCA with a classifier of SVM and novel '321' algorithm and achieved an accuracy of 82.5% on classification of Asian vs Non-Asian.

However when we compare our results achieved with the published literature of using deep learning and neural networks (Jilani *et al.*, 2019) performed a binary ethnic classification of Pakistani and Non-Pakistani faces which consisted of ethnicities such as White, Nigerian, Arab and Chinese, Egyptian and Poles it achieved an almost perfect results with an accuracy of 99.2% on ResNet-101 with an

SVM classifier. Similarly (Masood *et al.*, 2018) who performed classification using CNN of three ethnicities Negroid, Caucasia and Mongolia also achieved a very high accuracy of 98.6% with 20 epochs, Similarly (AlBdairi, Xiao and Alghaili, 2020) performed three ethnicity classifications using a modified VGG network also achieved a higher accuracy of 96.2%.

It is critical to note, however, that all of the works above used two or three very distinct ethnicities, whereas our work is based on the classification of two faces that appear very similar. The ethnic classification of two very similar appearing faces does not have a lot of directly comparable literature. However, we will still be in agreement with the results of other researchers as our accuracy of 92% is only slightly less than theirs, despite using a more challenging classification task. Suggestions for improving the results are reported in the future work section.

VI. FUTURE WORK:

Based on the experiments carried out in this work, of classifying faces of two similar populations, and achieving an overall good accuracy, these are some of the proposals for future projects.

1) Based on the accuracy achieved in this project, future projects can also involve ethnic classification of two very similar populations of people, such as that of the British and French populations or that of the East Asian population of China, Japan, and Korea.
2) The Neural Network architecture used on this dataset was based on ResNet-18. The ResNet architecture consists of other variants such as ResNet-50, ResNet-101 and ResNet-152. These neural networks will have greater depth therefore a higher accurate result may be achieved.
3) In deep learning there are number of different advanced neural network architectures currently available for example the visual geometry group developed by oxford university also known as the VGG-16, or VGG-19 network; or the AlexNet and LeNet, A future project can employ these networks and accuracy can be compared to ResNet.
4) It is possible to add more classes of similar countries to the neural network in order to cover a greater geographical area. For example, the population of Bangladesh and Sri Lanka could be added.
5) Due to constraints, this project only utilized secondary datasets. In the future, for a project with a larger scope and timeframe, a dataset created by photographing a large number of people in Pakistan and India from different regions of both countries can be formed, this will result in coherent and consistent dataset hence overall improving the accuracy of project.

VII. CONCLUSION:

The paper was written as part of a project for the course Data Science & Artificial Intelligence Conversion Programme MSc. Using Deep Learning Convolutional Neural Networks, we conducted an experiment to answer the challenging problem of ethnicity classification using facial features from images.

Based on a comprehensive literature review, two research problems were identified: (1) lack of investigation into similar appearing facial populations in ethnic classification using Machine Learning, and (2) under-representation of the subcontinent's population in ethnic classification using Machine Learning.
A model of the ResNet-18 architecture was employed in this project and images of Pakistani and Indians were classified using the Python Pytorch framework. The raw images were obtained from two different databases and then transformed using Photoshop and Pytorch, including converting to grayscale, removing background, and rotating the images randomly by 15 degrees.

The network was then trained using 200 images, of which 100 images were of Pakistani faces and 100 images were of Indian faces, and tested it using 88 images, of which 44 images were of Pakistani faces and 44 images of Indian faces. The results achieved were very good, with the highest percentage achieved on the testing dataset being 92%.

References:

AlBdairi, A.J.A., Xiao, Z. and Alghaili, M. (2020) 'Identifying Ethnics of People through Face Recognition: A Deep CNN Approach', *Scientific Programming*, 2020, pp. 1–7. Available at: https://doi.org/10.1155/2020/6385281.

Al-Saffar, A.A.M., Tao, H. and Talab, M.A. (2017) 'Review of deep convolution neural network in image classification', in *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*. Jakarta: IEEE, pp. 26–31. Available at: https://doi.org/10.1109/ICRAMET.2017.8253139.

Anthropometry | NIOSH | CDC (2021). Available at: https://www.cdc.gov/niosh/topics/anthropometry/default.html (Accessed: 8 August 2022).

Britannica (2022) *India - People | Britannica*. Available at: https://www.britannica.com/place/India/People (Accessed: 14 August 2022).

Brownlee, J. (2019) 'A Gentle Introduction to Cross-Entropy for Machine Learning', *Machine Learning Mastery*, 20 October. Available at: https://machinelearningmastery.com/cross-entropy-for-machine-learning/ (Accessed: 14 August 2022).

Ethnic Groups In Pakistan (2019) *WorldAtlas*. Available at: https://www.worldatlas.com/articles/ethnic-groups-in-pakistan.html (Accessed: 14 August 2022).

Farkas, L.G., Katic, M.J. and Forrest, C.R. (2005) 'International Anthropometric Study of Facial Morphology in Various Ethnic Groups/Races', *Journal of Craniofacial Surgery*, 16(4), pp. 615–646. Available at: https://doi.org/10.1097/01.scs.0000171847.58031.9e.

Foote, K.D. (2022) 'A Brief History of Deep Learning', *DATAVERSITY*, 4 February. Available at: https://www.dataversity.net/brief-history-deep-learning/ (Accessed: 7 August 2022).

Gudi, A. (2016) 'Recognizing Semantic Features in Faces using Deep Learning'. arXiv. Available at: http://arxiv.org/abs/1512.00743 (Accessed: 7 August 2022).

Guo, G. and Mu, G. (2010) 'A study of large-scale ethnicity estimation with gender and age variations', in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 79–86. Available at: https://doi.org/10.1109/CVPRW.2010.5543608.

Gupta, A. (2021) 'A Comprehensive Guide on Deep Learning Optimizers', *Analytics Vidhya*, 7 October. Available at: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/ (Accessed: 14 August 2022).

He, K. *et al.* (2016) 'Deep Residual Learning for Image Recognition', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, pp. 770–778. Available at: https://doi.org/10.1109/CVPR.2016.90.

Herculano-Houzel, S. (2009) 'The Human Brain in Numbers: A Linearly Scaled-up Primate Brain', *Frontiers in Human Neuroscience*, 3, p. 31. Available at: https://doi.org/10.3389/neuro.09.031.2009.

Hosoi, S., Takikawa, E. and Kawade, M. (2004) 'Ethnicity estimation with facial images', in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pp. 195–200. Available at: https://doi.org/10.1109/AFGR.2004.1301530.

Huff, T., Mahabadi, N. and Tadi, P. (2022) 'Neuroanatomy, Visual Cortex', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK482504/ (Accessed: 7 August 2022).

IBM, I. (2021) *What are Neural Networks?* Available at: https://www.ibm.com/uk-en/cloud/learn/neural-networks (Accessed: 7 August 2022).

Jakhar, D. and Kaur, I. (2020) 'Artificial intelligence, machine learning and deep learning: definitions and differences', *Clinical and Experimental Dermatology*, 45(1), pp. 131–132. Available at: https://doi.org/10.1111/ced.14029.

Jilani, S.K. *et al.* (2019) 'On the Ethnic Classification of Pakistani Face using Deep Learning', in *2019 International Conference on Cyberworlds (CW)*. Kyoto, Japan: IEEE, pp. 191–198. Available at: https://doi.org/10.1109/CW.2019.00039.

Jilani, S.K. (2020) *An Investigation into the Performance of Ethnicity Verification Between Humans and Machine Learning Algorithms*.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (Accessed: 7 August 2022).

Lahore Resolution (2003) 'Lahore Resolution | The Lahore Resolution 23rd March 1940 by All India Muslim League.', *Story Of Pakistan*. Available at: https://storyofpakistan.com/lahore-resolution/ (Accessed: 14 August 2022).

Lakshmi, A. *et al.* (2021) 'The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category Membership', *Frontiers in Psychology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.627678 (Accessed: 9 August 2022).

LeCun, Y., Bengio, Y. and Hinton, G. (1990) 'Deep learning', *Nature*, 521(7553), pp. 436–444. Available at: https://doi.org/10.1038/nature14539.

Lu, X., Lu, and X. and Jain, A.K. (2004) 'Ethnicity Identification from Face Images', in *In: Proceedings of SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, pp. 114–123.

Maladkar, K. (2018) *Why ResNets Are A Major Breakthrough In Image Processing*, *Analytics India Magazine*. Available at: https://analyticsindiamag.com/why-resnets-are-a-major-breakthrough-in-image-processing/ (Accessed: 9 August 2022).

Masood, S. *et al.* (2018) 'Prediction of Human Ethnicity from Facial Images Using Neural Networks', in S.C. Satapathy et al. (eds) *Data Engineering and Intelligent Computing*. Singapore: Springer Singapore (Advances in Intelligent

Systems and Computing), pp. 217–226. Available at: https://doi.org/10.1007/978-981-10-3223-3_20.

Meel, V. (2022) *ANN and CNN: Analyzing Differences and Similarities*, *viso.ai*. Available at: https://viso.ai/deep-learning/ann-and-cnn-analyzing-differences-and-similarities/ (Accessed: 7 August 2022).

Mersha, T.B. and Abebe, T. (2015) 'Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities', *Human Genomics*, 9(1), p. 1. Available at: https://doi.org/10.1186/s40246-014-0023-x.

Mitchell, T. *et al.* (1990) *Machine Learning*.

Mostafa, S. and Wu, F.-X. (2021) 'Chapter 3 - Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images', in A.S. El-Baz and J.S. Suri (eds) *Neural Engineering Techniques for Autism Spectrum Disorder*. Academic Press, pp. 23–38. Available at: https://doi.org/10.1016/B978-0-12-822822-7.00003-X.

Narang, N. and Bourlai, T. (2016) 'Gender and ethnicity classification using deep learning in heterogeneous face recognition', in *2016 International Conference on Biometrics (ICB)*, pp. 1–8. Available at: https://doi.org/10.1109/ICB.2016.7550082.

National Geographic (2014) 'It's Thanks to Evolution That No Two Faces Are Alike, Study Finds', *Culture*. Available at: https://www.nationalgeographic.com/culture/article/140916-faces-evolution-cognition-social-recognition-genetics (Accessed: 11 August 2022).

Ou, Y. *et al.* (2005) 'A real time race classification system', in *2005 IEEE International Conference on Information Acquisition*, p. 6 pp.-. Available at: https://doi.org/10.1109/ICIA.2005.1635116.

Panchal, S. (2018) 'Artificial Neural Networks — Mapping the Human Brain', *Predict*, 21 October. Available at: https://medium.com/predict/artificial-neural-networks-mapping-the-human-brain-2e0bd4a93160 (Accessed: 7 August 2022).

Ruiz, P. (2019) *Understanding and visualizing ResNets*, *Medium*. Available at: https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8 (Accessed: 10 August 2022).

Sarwinda, D. *et al.* (2021) 'Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer', *Procedia Computer Science*, 179, pp. 423–431. Available at: https://doi.org/10.1016/j.procs.2021.01.025.

Stringer, C.B. and Andrews, P. (1988) 'Genetic and Fossil Evidence for the Origin of Modern Humans', *Science*, 239(4845), pp. 1263–1268. Available at: https://doi.org/10.1126/science.3125610.

Uniqtech (2021) 'Understand the Softmax Function in Minutes', *Data Science Bootcamp*, 17 October. Available at: https://medium.com/data-science-bootcamp/understand-the-softmax-function-in-minutes-f3a59641e86d (Accessed: 10 August 2022).

Wang, W., He, F. and Zhao, Q. (2016) 'Facial Ethnicity Classification with Deep Convolutional Neural Networks', in Z. You et al. (eds) *Biometric Recognition*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 176–185. Available at: https://doi.org/10.1007/978-3-319-46654-5_20.