Exploring Machine Learning Models for Age Recognition

Chathuranga Dharmarathne, Prof. Karim Ouazzane, Dr. Subeksha Shrestha, Dr Sandra Fernando

London Metropolitan University, 166-220 Holloway Rd, London N7 8DB, UK

cdchathuranga@gmail.com , k.ouazzane@londonmet.ac.uk, s.shrestha3@londonmet.ac.uk, s.fernando@londonmet.ac.uk,

Abstract: The continuous improvements of artificial intelligent models for classification and facial recognition have gotten a lot of attention in recent years, particularly in models for classifying and recognizing faces. These improvements have been instrumental in solving real-world challenges. Many studies focus on neural networks, sometimes called as "black boxes" due to their complex decision-making processes and why it forecasts a particular age. On the other hand, the accuracy of datasets, which are supposed to show a person's real age, might not be very accurate. Consequently, creating accurate and interpretable age recognition models becomes challenging. This journal about exploring a machine learning models that predicts a person's age based on facial images, using both traditional machine learning techniques and neural networks. The study utilizes a facial image dataset from the Imperial College of the United Kingdom. Image features are extracted using the Cannes edges method, and Principal Component Analysis reduces dataset dimensions. Random Forest, Support Vector Machine, Decision Tree, and Convolutional Neural Network (CNN), are applied to identify the most accurate model to predict age of the person. Performance metrics, confusion matrix, accuracy curve and loss curve were used to evaluate models. The model built using random forest algorithm was the highest accurate model of 60% and the SVM was recorded least accuracy of 49% compared to other models. On the other hand, CNN without convolution layer has performed 71% accuracy against 54% accuracy.

Keywords: Machine Learning Modes, feature extraction, Convolutional Neural Network, Canny edge, K-Means, SVM, PCA, Random Forest, AgeDB

1. Introduction

Artificial intelligence (AI) is currently a popular subject, as AI models are receiving significant recognition for their ability to solve complex problems with greater accuracy. The first automatic face recognition techniques were developed in the early 1960s and have since improved in a wide range of applications including security, surveillance, authentication, and human-computer interaction. This is a form of biometric identification that uses facial features to identify individuals. It works by analysing images taken from cameras or other sources and matching them against stored records in a database. The important is that it can accurately identify people with minimal effort and it's better than manual processes because people and machines don't have to touch each other. Hence, facial recognition systems are one step ahead than many existing technologies available today [1].

The accurate determination of an individual's age using visual data has substantial impact across various sectors, including marketing, retail, healthcare, and security. When considering the commercial intent and benefit to society of the proposed study, it is important to note that many of retail establishments in the United Kingdom engage in the sale of age-restricted products, including but not limited to tobacco, fireworks, alcohol, and knives. It is illegal to sell age-restricted products to minors in the United Kingdom. Retailers who are found selling such products to minors may be subject to legal consequences, including criminal charges and fines. Therefore, it is essential that retailers implement effective age verification procedures to ensure that age-restricted items are sold only to those of legal age. The present procedure involves verifying identification documents, such as a passport or driver's license. Hence implementing a machine learning model to meet this requirement, is expected to safeguard minors and ensuring a seamless, expedient shopping experience for customers by enhancing overall customer satisfaction. Moreover, this technology extends its impact to healthcare, where precise age information proves for tailoring personalized medical treatments, prescribing medications, and conducting age-specific health assessments. Furthermore, it enhances marketing and advertising initiatives through the comprehension of age demographics, thereby promoting equitable and fair utilisation [2][3].

This paper presents the results of experiments conducted to assess the effectiveness of incorporating convolutional layers into Convolutional Neural Networks (CNNs) for age detection of an individuals. We construct and evaluate two distinct CNN models, one incorporating convolutional layers and the other without convolutional layers. The performance is then validated and compared to that of other traditional machine learning algorithms. The outcomes of these experiments demonstrate that the predictive performance of the models is remarkably comparable. The feature extraction method's implementation performs identically to convolutional layers, showing that the feature extraction method is reliable and correct. Every machine learning model, nevertheless, has advantages and disadvantages. CNN uses a special technique to extract data from images. However, SVM and Random Forest perform well when handling tabular data and are robust to outliers. By comparing each of these models, it is possible to ascertain the most accurate approach to age recognition.

2. Related Work

Many research articles have been conducted regarding image recognition using convolutional neural network (CNN) to get better identification on real-world images. Accordingly, CNN can effectively classify unfiltered images using its good feature extraction methods. Large datasets made it possible to train CNN models that can learn small, distinguishable facial features. The main challenge they have emphasised was to detect faces in images with several visual differences such as lighting, posing and angle of the camera. A way to address this challenge is by training a model that can differentiate between the face and the background in an image using CNN with multiple prediction layers which can adjust the image's alignment and detect potential faces in images. Moreover, they further empathise that CNN network should consist a sufficient number of layers to obtain the best results and to avoid the risk of overfitting or underfitting the models [4].

On other hand, traditional machine learning algorithms have also been explored to predict age from facial images. In order to make age predictions based on facial images, the researchers used a support vector machine (SVM) that utilised a radial basis function (RBF) kernel. The accuracy rate on the FG-NET dataset, comprising facial images of individuals ranging from 0 to 69 years of age, was recorded at 65.6%.

One of the biggest challenges was finding a broad and comprehensive dataset for training the model after reading several research articles. It can take a lot of time and resources to gather a sizable dataset with variations in lighting, poses, expressions, and demographics. Data preprocessing presents another difficulty because it requires high computer resources as well as to address problems with image resolution, noise, occlusions, and alignment irregularities. These preprocessing steps are essential for improving the facial recognition system's robustness and accuracy. Another difficulty is selecting an appropriate model architecture and optimising it for performance. To get the best results, researchers must combine the appropriate deep learning models, tune the hyperparameters, and take care of the overfitting problem to ended with more accurate machine learning models. When addressing real-world scenarios, implementing the facial or age recognition system in practical settings poses difficulties with regard to computational effectiveness, hardware limitations, integration with current system and regulations of the company [20].

3. Material and Methods

Dataset

The AgeDB dataset was chosen due to its publication by the Imperial College London, United Kingdom, and its inclusion of accurate labelling for age identification. The AgeDB dataset originally comprised 16,488 images spanning ages 1 to 99. However, through cleaning, resizing, standardizing and transforming the image dataset in to consistent format, it was reduced to 14,748 images. The image features were extracted, and the dataset was then divided into training and test subsets using sklearn library available in python. A bar chart was plotted to understand the number of images for each age category and transformed same in to balanced dataset with five different age ranges [21].

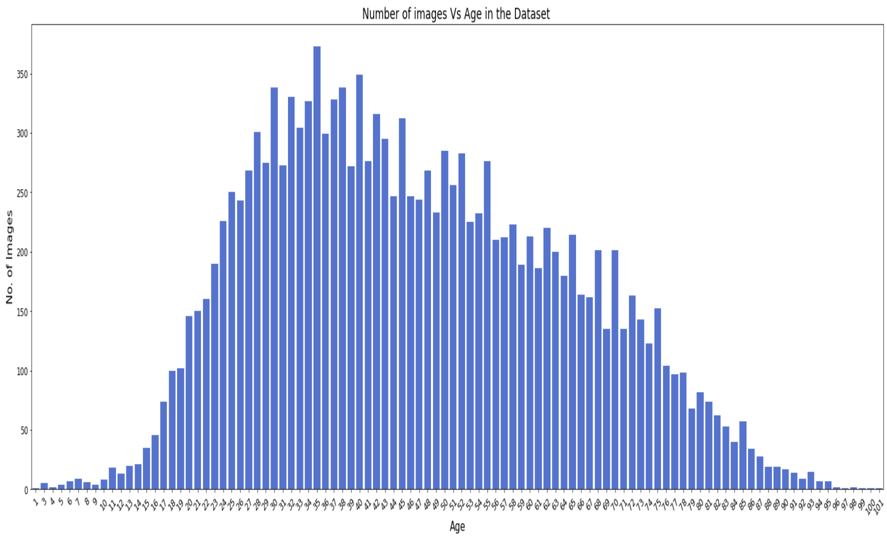


Figure 1 - Age Vs Number of Images

All images were segregated in to five different age groups as stated in the below table.

Class Label	Age Range	Number of images
0	0 to 18	373
1	19 to 40	5,842
2	41 to 60	5,042
3	61 to 80	3,028
4	Above 80	463

Table 1- Age Classes with number of Images

In this study all images were converted in to grayscale and applied the Canny edge detection filter. Then each image has been divided into 8x8 pixel sections and obtained the mean value of each section (256 unique scalar values). Then it has been turned into a data frame for build a machine learning model [19].

4. Building a Balance Dataset

After resizing and eliminating low-quality and unlabelled images, only 14,748 raw images remained in the AgeDB dataset. Then dataset have been divided into five distinct age groups, and the model will determine the age group rather than an individual's exact age. Thus, accurate age prediction is not feasible. Even after separating the images into five distinct classes, the number of images in each class does not correspond to the age group. Therefore, in order to rectify the imbalance between these groups, random oversampling techniques were used to balance the number of images in each age category [5] [16].

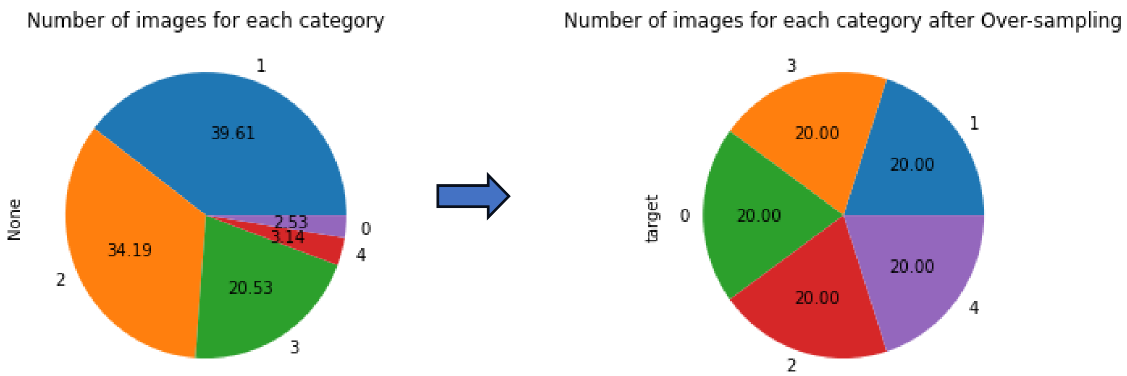


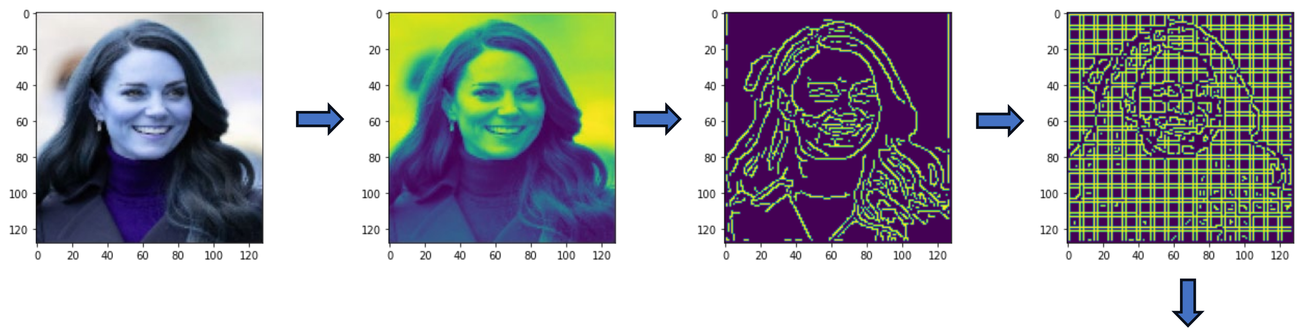
Figure 2 Transforming unbalance dataset in to balance dataset using SMOTE method.

After resampling, each age group now contains 5,842 images. As a result, model performance has increased and mitigated the risk of the model becoming biased towards the majority class. Furthermore, random oversampling method has reduced the likelihood of false negatives, and enhanced the model's capacity to accurately classify minority class instances, while addressing class imbalance issues [19].

Feature Extraction from the image

A critical stage in image processing and classification is feature extraction, which converts unprocessed image data into a set of useful features. There are several methods available for this, and each has advantages and disadvantages of its own. Common techniques include colour features, edge features, shape features, and texture features. For this project edge feature technique has been used which based on the edges in an image. Edges are the boundaries between different regions in an image. They can be detected using a variety of techniques, such as the Sobel operator or the Canny edge detector [6].

Colour images typically have three channels, specifically red, green, and blue. However, purpose of identifying age, image colour information is not significant. To ensure simplicity and ease of analysis, all images were transformed into grayscale. The process of converting data minimises the number of dimensions and improves computing performance in both the training and deployment stages of the machine learning model. After converting the images in to grayscale, the Canny edge detection technique is employed to extract the characteristics of the images. The algorithm successfully captures the fundamental facial features and defines the boundaries and contours of facial attributes, such as wrinkles, creases, and overall facial structure. As a result, it provides a comprehensive and informative framework of the face that can be used by machine learning algorithms to detect age related patterns [19] .



	sec1_mean	sec2_mean	sec3_mean	sec4_mean	sec5_mean	sec6_mean	sec7_mean	sec8_mean	sec9_mean	sec10_mean	...	sec253_mean	sec254_mean	sec255_mean	sec256_mean	age
0	0.203125	0.171875	0.171875	0.140625	0.000000	0.18750	0.234375	0.125000	0.125000	0.140625	...	0.031250	0.000000	0.000000	0.000000	7.0
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.093750	0.078125	0.078125	...	0.203125	0.125000	0.000000	0.093750	46.0
2	0.187500	0.140625	0.296875	0.281250	0.234375	0.28125	0.312500	0.250000	0.281250	0.281250	...	0.062500	0.265625	0.156250	0.109375	26.0
3	0.109375	0.125000	0.156250	0.328125	0.109375	0.00000	0.000000	0.000000	0.000000	0.000000	...	0.328125	0.125000	0.046875	0.156250	44.0
4	0.000000	0.203125	0.171875	0.296875	0.000000	0.18750	0.359375	0.203125	0.203125	0.171875	...	0.109375	0.203125	0.109375	0.000000	33.0

5 rows × 257 columns

Resized RGB Image (128x128)→ Gray Scale → Canny Edge filter → Divided image in to 8x8 → Calculate mean value of each block

Figure 3 - Image conversion in to Matrix values

After the successful implementation of the Canny edge detection technique, the images undergo a conversion process to transform them into scalar values, enabling their use as input for a machine learning classifier. To accomplish this task, the photos were resized to 128x128 (that is total 16,384 pixels) and then divided into 8x8 pixel sections as depicted in Figure 3. Subsequently, the mean value of each 64 pixels section was computed. The outcome of this process resulted in the generation of 256 unique scalar values for each image accompanied by age of the person. These values were subsequently organised into a data frame which consist of 257 columns, were divided in to train and test dataset using sklearn library in python with the intention of utilising them in to machine learning classifiers [7].

Convolutional Neural Networks (CNNs) are indeed highly proficient at extracting valuable information from colour images. These neural networks are inherently designed to work with multichannel data, such as the red, green, and blue channels in RGB colour images. They excel at learning hierarchical features, starting from low-level patterns like edges and textures in each channel and progressively combining this information to capture intricate relationships between colours and structures. Therefore, there is no requirement to convert images in to grey scale [8].

Based on the prevailing literature, it is widely recommended that the optimal outcome for Convolutional Neural Networks (CNNs) is achieved when images are resized to dimensions of 224x224 pixels. Consequently, Images were resized to conform to this standard. In the context of identifying and confirming individuals age based on the images, the salient aspects of a person's facial structure, encompassing the spatial arrangement of their features, facial lines, and distinctive facial attributes, assume paramount significance. Convolutional neural networks (CNNs) have a high level of proficiency due to its use of several convolutional layers. These layers enable the network to acquire hierarchical representations that extract significant features such as edges and textures [24].

5. Implementation Of Age Detection Models

After extracting features from the images, there were 256 distinct scalar values for each image, forming a comprehensive dataset that was organized into a structured data frame for an image. However, when consider entire volume of data points (256 features for each of the 29,210 images, totalling 7,477,760 data points), it is necessity of dimensionality reduction to enhance computational efficiency and model interpretability. Hence, Principal Component Analysis (PCA) was used to reduce the dimensionality of the data frame in to 10 features, which allowing to retain the most significant variance while significantly reducing the overall feature space. This transformation simplifies subsequent analysis while preserving the essential information needed for accurate age detection. After that, the K-means clustering technique was used and visualised in order to detect any hidden patterns and determine if the data points could be effectively divided into five distinct clusters based on their similarities. Then, the supervised machine learning algorithms of Random Forest, Support Vector Machine, Decision Tree, and CNN models were implemented to determine the most accurate model for detecting an individual's age [13].

The CNN architecture, defined using TensorFlow's Keras API, represents a convolutional neural network for image classification tasks. It begins with an input layer specifying colour image shape of (224, 224, 3). The network then comprises three convolutional layers with increasing filter sizes (8, 32, 64 and 128), each followed by a max-pooling layer for down-sampling. The "ReLU" activation function is applied throughout the convolutional layers to introduce non-linearity. Once the last max-pooling layer is completed, the output is reduced to a one-dimensional vector. There's a dense layer with 128 units and a ReLU activation function, followed by a dropout layer with a 20% dropout rate to prevent overfitting. The neural network concludes with a five-unit softmax layer, indicating that it was built specifically to handle a multi-class classification task involving five classes or image categories. After that, the model is compiled using the Categorical Cross-Entropy loss function, Adam optimizer with a learning rate of 0.001, and accuracy as the evaluation metric. It is then trained for 30 epochs using the training and testing data. The compile function configures the model for training, specifying the loss function, optimizer, and evaluation metric, while fit performs the training process for 30 number of epochs, collecting training history information to evaluate and visualize the performance of the model [14].

In addition to the normal CNN another CNN has been implemented without convolutional layer. The architecture comprising five connected dense layers, four hidden layers, and one output layer containing five neurons. The initial layer, including 256 neurons, is expected to function as either a feature extractor or an initial representation transformer. The successive levels exhibit a steady reduction in the number of neurons, namely with 256,128, 64, and 16 neurons in the second, third, and fourth layers, respectively. The gradual decrease in the number of neurons can facilitate the extraction of hierarchical and abstract characteristics from the dataset.

6. Evaluation And Model Performance

This study used both supervised and unsupervised machine learning algorithms for building machine learning models. The PCA plot displayed below was generated using the 10 principal components derived from the initial dataset, where the original 256 features were effectively reduced to just 10 dimensions [19].

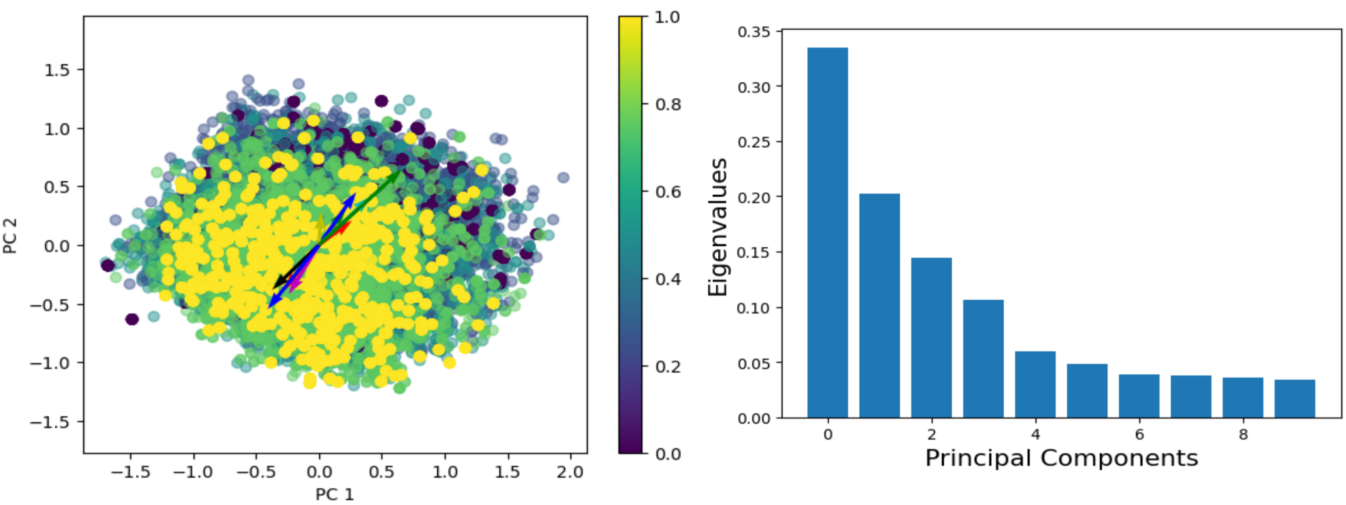


Figure 1 - Scatter plot with 10 with eigen vectors and Eigenvalues vs Principal components

Eigenvalues and Eigenvectors are crucial to Principal Component Analysis (PCA), which reduces dimensionality and extracts features. Eigenvalues indicate the importance of each principal component in PCA. The proportion of variance explained by each component is measured. Larger eigenvalues explain more variance and are more informative. This information is crucial for determining how many major components to keep to preserve a lot of the original data's variation. Eigenvalues measure how much each principal component accounts for a dataset's variation. The initial principal component explains 0.3346 of the dataset's variances, followed by the second component at 0.2023 and the tenth component at 0.1445.

After PCA dimensionality reduction revealed 10 principal components, the unsupervised K-Means approach was applied to discover data patterns and similarities. K was set to 5 to split the data into five age-group clusters, and the visualisation is shown below.

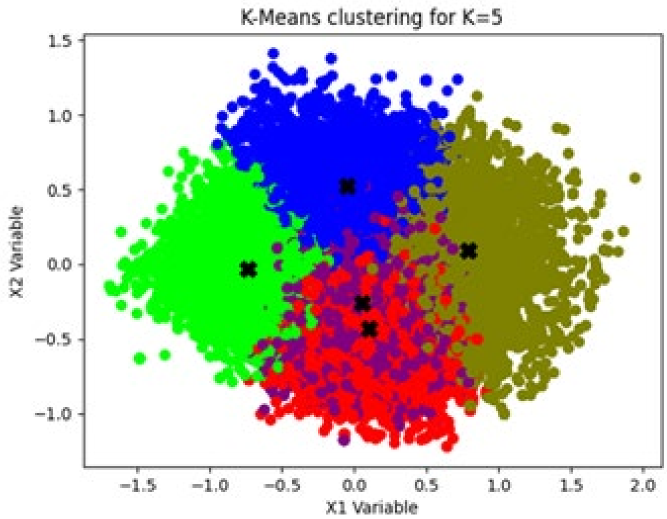


Figure 5 K-Mean Clustering for Image dataset k=5

Comparison using Confusion Matrix

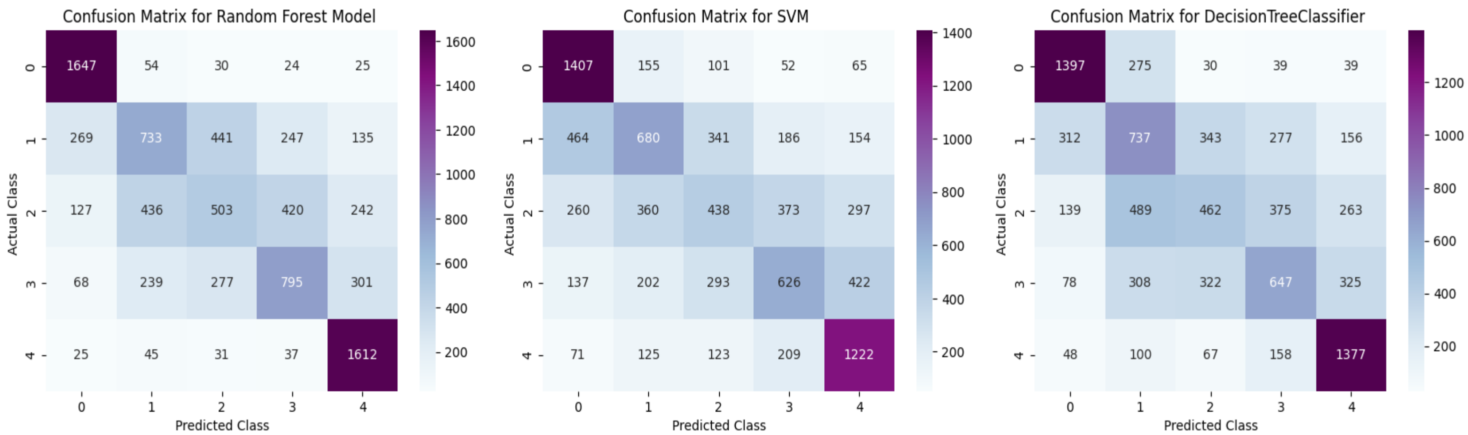


Figure 6 Confusion Matrixes of each model

The above confusion matrixes display how the model's predictions compare to the actual class labels across five distinct classes among all machine learning models. It seems to perform quite well in some classes.

Summary of the Performance of Three Models

	Random Forest			SCM			Decision Tree			
Class	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	support
0 (0 to 18)	0.77	0.93	0.84	0.60	0.79	0.68	0.71	0.78	0.74	1780
1 (19 to 40)	0.49	0.40	0.44	0.45	0.37	0.41	0.39	0.40	0.39	1825
2 (41 to 60)	0.39	0.29	0.33	0.34	0.25	0.29	0.38	0.27	0.31	1728
3 (61 to 80)	0.52	0.47	0.50	0.43	0.37	0.40	0.43	0.39	0.41	1680
4 (81<)	0.70	0.92	0.79	0.57	0.70	0.63	0.64	0.79	0.70	1750
accuracy			**0.60**			**0.50**			**0.53**	8763
macro avg	0.57	0.60	0.58	0.48	0.50	0.48	0.51	0.53	0.51	8763
weighted avg	0.57	0.60	0.58	0.48	0.50	0.48	0.51	0.53	0.51	8763

Table 2- Model Performance - Precision, Recall, and F1-Score

Above table shows the analysis of the three classification models for Random Forest, SCM (Support Vector Machine), and Decision Tree. In terms of overall accuracy, Random Forest emerges as the top-performing model with an accuracy of 0.60, surpassing both SCM and Decision Tree, which achieved accuracies of 0.50 and 0.53, respectively. Exploring deeper into the macro and weighted averages, Random Forest consistently outperforms its counterparts across precision, recall, and F1-score. This underscores its robustness in handling the complexities of the classification task. Interestingly, Decision Tree, while falling short of Random Forest, outshines SCM in most metrics, indicating its effectiveness in this specific context. However, Model accuracy is chosen not only for accuracy but also for the application's needs and limitations.

Performance of the convolutional neural networks

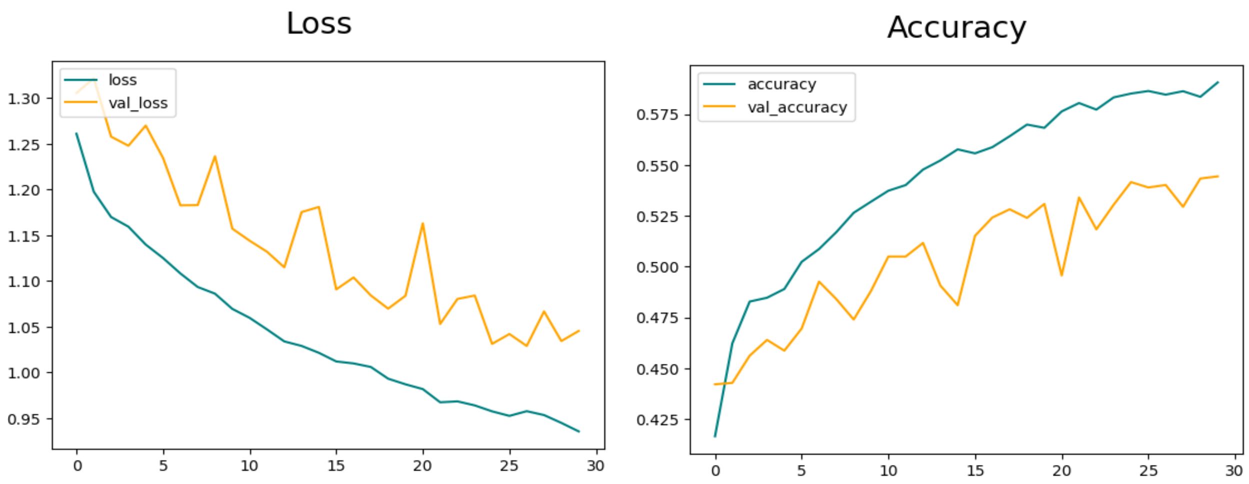


Figure 7 - Performance of the convolutional neural network with convolutional layer

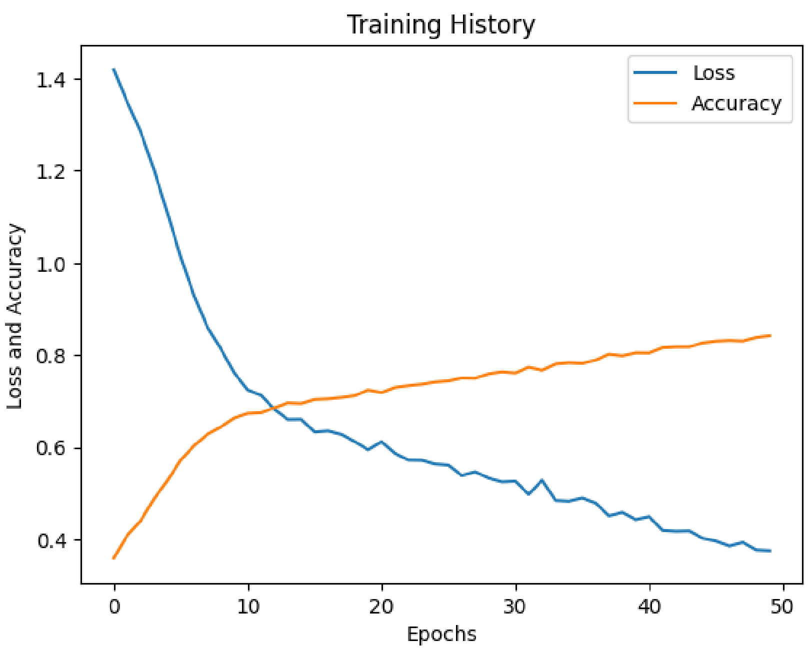


Figure 8 - Performance of the convolutional neural network without convolutional layer

According to the results indicates relating to the performance of the Convolutional Neural Network (CNN) with and without convolution layers. The CNN achieves an accuracy of 0.5463 and a loss of 1.011, while the modified version without convolution layers demonstrates a notable improvement with an accuracy of 0.7182 and a reduced loss of 0.8705. The stark contrast in accuracy and loss between the two models suggests that the convolutional layers in the standard CNN play a crucial role in enhancing its performance. The higher accuracy and lower loss in the CNN without convolution indicate that, in this specific scenario, the removal of convolutional operations has a positive impact on the model's ability to learn and generalize.

Summary of Supervise Learning Models

	Testing Accuracy	**Training Accuracy**
Random Forest Classifier	0.6037	0.7032
Support Vector Machine	0.4990	0.5325
Decision Tree Classifier	0.5272	0.5857
	Accuracy	**Loss**
CNN	0.5463	1.011
CNN - Without Convolution	0.7182	0.8705

Table 3- Performance Summary of the all models

7. Conclusion

This study has evaluated some key factors related to the prediction of age of a person using machine learning models. The main contribution of this paper is the feature extraction mechanism used to implement machine learning models and it was proved that the feature extraction method is success than CNN. Furthermore, this study aimed at evaluating the effectiveness of incorporating convolutional layers within Convolutional Neural Networks (CNNs) for Age detection of a person and validate by comparting with traditional machine learning models. The feature extraction technique used for this implementation shows a level of efficiency comparable to that of traditional convolutional layers, showcasing its potential as an innovative and effective approach for feature extraction for age detection of a person as discussed in section 5.

There is an exciting opportunity to further investigations by exploring different ways of feature extraction in images, beyond the use of Canny edge detection. This could involve trying out methods like Laplacian of Gaussian, Histogram of Oriented Gradients and Scale-Invariant Feature extraction. Additionally, changing the image size, using various colour filters, and trying out different methods to balance the dataset, such as SMOTE-ENN and Adaptive Synthetic Sampling, could help in creating a more balanced dataset for building machine learning models. The objective of this investigation is to ascertain the finest configurations that produce the most precise results in predicting of an individual's age. Moreover, the research could go beyond the calculations of pixel averages and include different statistical measures like the maximum value, standard deviation with variance to create a comprehensive dataset for developing machine learning models. This approach shows promise in improving the accuracy and effectiveness of age prediction models across various applications.

The potential commercial implementation of this project is quite beneficial, particularly in facilitating real-time age verification during self-checkout processes at retail shops in the UK. Enhancing the accuracy of the age detection model could significantly benefit the entire nation. Moreover, integrating this model into CCTV systems would make it more reliable in preventing younger individuals from accessing age-restricted locations which are operating during the nighttime and provide entertainment, such as music, dancing, and drinks.

8. Future Works and Recommendation

Regarding future work on this study, there are a number of exciting options and topics to pursue. Exploring cutting-edge feature extraction techniques, such as deep feature learning, has the potential to improve the model's ability to represent complex facial characteristics. Concurrently, machine learning models, particularly deep learning architectures, must be continuously fine-tuned through exhaustive experimentation with a variety of model configurations, hyperparameters, and optimisation techniques to achieve the highest level of accuracy and efficiency. Also, it is equally important to develop resource-efficient and real-time implementations for devices with limited computational resources, such as smartphones, self-vending machine.

The age detection using CNN without convolutional layer has shown good results. However, the study faced challenges, such as difficulty in finding a well-balanced real-world dataset, extracting features from images, and handling imbalanced data. One limitation was the inability to train the model with a larger number of dimensions (image features) unless a powerful computer processor with a GPU was available. Moreover, the study would likely experience improvements in the accuracy and robustness of the age detection models by incorporating more comprehensive combined, balanced datasets and testing with unknow new dataset to find out actual performance of the models.

9. References

[1] “VICE,” https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works, 2023. [Online]. Available: https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works.

[2] P. community, “face-recognition 1.3.0,” Python Software Foundation, [Online]. Available: https://pypi.org/project/face-recognition/. [Accessed 2023].

[3] “GOV.UK,” [Online]. Available: https://www.gov.uk/government/publications/code-of-practice-age-restricted-products. [Accessed 2023].

[4] J. Y. e. al, “Age Estimation by Multi-scale Convolutional Network,” no. 2015 IEEE International Conference on Image Processing.

[5] “Analytics Vidhya,” 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-networkcnn/.

[6] “https://www.researchgate.net,” Research Gate, 2023. [Online]. Available: https://www.researchgate.net/publication/269269817_Age-Group_Classification_of_Facial_Images.

[7] “How facial recognition technology works,” Ellucian, 2023. [Online]. Available: https://www.ellucian.com/emea-ap/blog/facial-recognition-can-give-students-better-service-and-security.

[8] “OpenCV,” 2023. [Online]. Available: https://opencv.org/.

[9] J. L. Lichun Yu's, “Face Recognition Based on Deep Learning of Small Data Set,” no. https://www.researchgate.net.

[10] “VGG Very Deep Convolutional Networks,” [Online]. Available: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/. [Accessed 2023].

[11] A. V. a. A. Zisserman, “VGG Convolutional Neural Networks Practical,” Oxford Visual Geometry Group - Oxford University, 2017. [Online]. Available: https://www.robots.ox.ac.uk/~vgg/practicals/cnn/index.html.

[12] P. N. S. S. C. a. D. R. Reddy, “Face Recognition Using Convolutional Neural Networks,” no. 2018 Second International Conference on Inventive Communication and Computational Technologies

[13] “3 Beginner-Friendly Techniques to Extract Features from Image Data using Python,” Analytics Vidhya, 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/3-techniques-extract-features-from-image-data-machine-learning-python/.

[14] “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” Springer Open, 2023. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8.

[27] “Convolutional Neural Networks,” paperswithcode, [Online]. Available: https://paperswithcode.com/methods/category/convolutional-neural-networks. [Accessed 2023].

[16] P. Agarwal, “Age Detection using Facial Images: traditional Machine Learning vs. Deep Learning,” [Online]. Available: https://towardsdatascience.com/age-detection-using-facial-images-traditional-machine-learning-vs-deep-learning-2437b2feeab2. [Accessed 2023].

[17] A. K. M. A. I. Rasha Atallah, “Face Recognition and Age Estimation Implications of Changes in Facial Features,” Research Gate, 2023. [Online]. Available: https://www.researchgate.net/publication/325236448_Face_Recognition_and_Age_Estimation_Implications_of_Changes_in_Facial_Features_A_Critical_Review_Study.

[18] S. Javaid, “Top 4 Facial Recognition Challenges & Solutions in 2023,” AI Multiple, [Online]. Available: https://research.aimultiple.com/facial-recognition-challenges/. [Accessed 2023].

[19]. G.C.M Dharmarathne, "Exploring Machine Learning Models for Age and Facial Recognitions," United Kingdom, 2023.

[20] M. Pils, “What are the current challenges and limitations of face detection and recognition research?,” LinkedIn community, [Online]. Available: https://www.linkedin.com/advice/0/what-current-challenges-limitations-face.

[21] S. a. P. A. a. S. C. a. D. J. a. K. I. a. Z. S. Moschoglou, “AgeDB,” the first manually collected, in-the-wild age database, no. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2017. https://ibug.doc.ic.ac.uk/resources/agedb/

[22] pathminds, “A Beginner's Guide to Neural Networks and Deep Learning,” pathminds, [Online]. Available: https://wiki.pathmind.com/neural-network. [Accessed 2023].

[23] G. Ognjanovski, “Everything you need to know about Neural Networks and Backpropagation — Machine Learning Easy and Fun,” medium, [Online]. Available: https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a. [Accessed 2023].

[24] "Artificial Intelligence stack exchange," stackoverflow, [Online]. Available: https://ai.stackexchange.com/. [Accessed 2023].

[25] J. Edstedt, "Towards Understanding Capsule Networks," Linköping University.