Smart Surveillance: CNN-Based Age and Gender Detection

Shreshi Patel [1], Prof. Naveen Kumar V[2]

[1] Department & College MBA – Business Analytics, CMS Business School, Jain University, Bengaluru
shreshi_patel23@cms.ac.in

[2] Assistant Professor, Faculty of Management Studies CMS Business School, Jain University, Bengaluru
naveenkumar_v@cms.ac.in

Abstract - In the twenty-first century, humanity has reached a stage of remarkable technological advancements, leading to groundbreaking achievements. The rise of social media and online platforms has significantly increased the demand for automatic age and gender classification across various applications. While substantial research has been conducted in this domain, there remains considerable scope for enhancement, especially in age classification. This study aims to compare the performance of Convolutional Neural Networks (CNNs) with Support Vector Machines (SVMs) on the same dataset, challenging the prevailing notion that CNNs are inherently superior for image classification tasks. However, existing approaches using real-world images still fall short of achieving the desired accuracy, particularly when compared to the significant advancements seen in related areas like age estimation. In this paper, we demonstrate how deep CNNs can effectively learn feature representations to improve classification accuracy. We introduce a simplified CNN architecture that requires minimal training data while delivering high performance. The results reveal that our approach significantly surpasses current benchmarks for age and gender estimation.

Keywords: **Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Deep learning, Feature extraction, Machine learning, Facial recognition.**

1. INTRODUCTION

In the last few years, the huge amount of data specially images and videos are getting uploaded rapidly on the Internet. The classification of age and gender using image datasets is crucial in a variety of applications, including user identification, video monitoring, human-robot interaction, and security surveillance. Age and Gender plays a crucial role in biometric identification for human beings.

For this human identification and verification, biometric system gathers data on a person's physiological and security models.

The real-world operations of age and gender bracket experienced several problems comparable to variable facial orientations, low-resolution photos, illumination, multiethnicity, and colorful ageing patterns in some persons. These challenges were present in most areas of image analysis. Age classification is still a hot area of research despite advancements in computer vision. Therefore, it is difficult to develop an age categorization system that provides high-accuracy findings using just a facial image.

The lack of a recognizable pattern of facial ageing is one of the biggest obstacles that any system must overcome in order to estimate or categories age. The stages of ageing can vary from one mortal to another depending on a number of internal and external elements, including environment and lifestyle, food system, health situation, and psychological state. This is due to the nature of mortal faces. The link between the face image and age is consequently non-linear, making it difficult to capture it in a universal model. This work presents a suggestion for a novel method of determining age and gender based on multiple CNN.

2. OVERVIEW WITH PROBLEM STATEMENT

In today's life, all application which have been made for day-to-day life are generally focusing upon soft biometrics sector to reduce the communication gaps between human and computers.

Machines cannot classify patterns as effectively and powerfully as the mortal brain can. thus, our goal is to use technology to imitate the capability of the human brain to detect a person's real age and gender approximately.

This problem can be resolved by developing an application for age and gender detection which can accurately determine a person's age and gender. The age and gender of the person are determined by using their human face as the input. The person's age and gender are the output.

3. PURPOSE

This study develops and suggests a convolutional neural network-based automatic age and gender classification system based on the human face. This technology might be used to societal benefit looks at a wide range of applications, including security services, CCTV surveillance and policing, courting operations, and dating applications. It can also help in certain face detection algorithm to make it more efficient and help parents to keep a protection and bound their child to use fewer hours in mobile phones and other smart devices.

4. RELATED WORKS

B. N Akash., Kulkarni Akshaya. K, A Deekshith., Gowda Gowtham, "Age and Gender Recognition using Convolution Neural Networks", Conference Location: The National Institute of Engineering, Mysuru, India, Date of Conference: June 2020 [6]

They suggested a method that would use convolutional neural networks and the Pytorch framework to create a deep learning model. After receiving the image, the model reduces its size as it moves through each layer. The IMDb-wiki dataset, which comprises over 5 lakh photos of people and includes various racial, ethnic, and other characteristics, is then used to train our model. Each image in the dataset also has a label. We can use the model for testing after it has been trained. The existence of a person is initially verified in every image. Then, we use the Open CV framework to analyze the image in order to extract all of the human faces that are visible in it. Following the processing of each face by a developed deep learning model, an output label that accurately identifies the age and gender of each person in the image is produced.

X Wang, A Mohd Ali, P Angelov, "Age and gender classification of human faces for analyzing human behavior", Conference Location: Exeter, UK, Date of Conference: June 2017 [5]

Although it is impossible to foresee human behavior, it can be influenced by feelings or the surroundings. So, they developed a way to identify age and gender from a person's face in order to comprehend human behavior and perform autonomous detection of abnormal human behavior. This work may be put into practice for a variety of purposes, like looking into any suspicious or unusual activity that could be used to stop a crime from happening. Here, the idea of supervised learning has been applied. Here, the support vector machine approach is utilized for classification. Additionally, the primary strategy used in this case is the recently popularized transfer learning notion, which is based on the deep learning concept. The face's features are extracted using a pre-trained deep network before being fed into the support vector machine classifier. They created a dataset called Gay Face" and used that dataset to apply the suggested strategy because the availability of the labelled data is quite low. They have excellent results and a very robust proposed method. Age and gender were correctly predicted with an accuracy of roughly 70.19% and 80.33%, respectively, which will have a big impact on forecasting human behavior.

Kotiyal, Arnav, Praveen Gujjar, and Guru Prasad. "Pre-trained Model for Image Prediction Using Keras Application." 2023 International Conference on Computer Science and Emerging Technologies (CSET). IEEE, 2023 [7]

They explored the use of pre-trained deep learning models for image classification using Keras, highlighting the benefits of transfer learning in reducing computational costs while improving accuracy. Their findings support the integration of ResNet, EfficientNet, and VGG16 for feature extraction, which is relevant to CNN-based age and gender classification.

J, P. G., Sharief, M. H., Mukherjee, U., Kumar, M. A., & S, G. P. M. (2023). Malicious Host Detection in Software Defined Networks Using Machine Learning Algorithms. Research Gate, 1–6 [8]

They studied malicious host detection in Software Defined Networks (SDNs)using machine learning algorithms, emphasizing security risks in AI-driven applications. Their work highlights the importance of real-time threat detection, which is applicable to preventing bias, unauthorized access, and security vulnerabilities in AI-based smart surveillance systems.

Avinash, S., Kumar, H. N. N., Prasad, M. S. G., Naik, R. M., & Parveen, G. (2023). Early detection of malignant tumor in lungs using Feed-Forward neural Network and K-Nearest Neighbor Classifier. SN Com. puter Science, 4(2)[9]

They investigated early lung cancer detection using feed-forward neural networks (FFNNs) and K-Nearest Neighbor (KNN) classifiers. Their research highlights how deep learning models can improve medical image classification accuracy, which parallels CNN-based age and gender classificationby emphasizing feature extraction, dataset diversity, and bias mitigation.

Prasad, M. S. G., Pratap, M. S., Jain, P., Gujjar, J. P., Kumar, M. A., & Kukreti, A. (2022). RDI-SD: An Efficient Rice Disease Identification based on Apache Spark and Deep Learning Technique. Research Gate, 277–282 [10]

They proposed an AI-based rice disease identification model using Apache Spark and deep learning, demonstrating the power of big data in large-scale image classification. This study provides insights into how distributed computing and scalable AI architectures can enhance real-time CNN-based facial recognition in smart surveillance applications.

Raza, S. S. (2023). Exploring the conceptual realm of machine learning in small and medium-sized Industries: A Qualitative study. Ushus - Journal of Business Management, 22(4), 1–13 [11]

A qualitative study that examined the adoption of machine learning in small and medium-sized enterprises (SMEs), identifying key challenges such as computational costs, data availability, and model efficiency. These findings reinforce the importance of lightweight CNN architectures and real-time processing techniques, making AI-based surveillance more scalable and practical.

5. DESIGN/METHODOLOGY/APPROACH

1. Gather a dataset of images that includes both males and females of different age groups.

2. Resize the photos and normalize the pixel values to pre-process the data.

3. Separate the aforementioned data into training and testing sets.

4. Define the architecture of the CNN model. We can use multiple convolutional layers followed by max pooling layers, and then add fully connected layers at the end.

5. Train the model on the training set using an appropriate loss function, optimizer, and batch size.

6. Evaluate the model on the testing set and measure its accuracy.

Convolutional Neural Network (CNN)

Whenever we see an image, we directly determine it by seeing it through our naked eyes because of our Visual Cortex Systems and our ability to think but it is not possible for a machine to think in a way that a human think.

After research and study of more than a decade, a French scientist and Turing Award Winner "Yann Lecun" in 1998 made it possible for the system to recognize any object by getting trained from results of some previously existing examples and this process of training is called Convolutional Neural Network (CNN).

In CNN there are majorly two process through which they undergo:

I. Convolution
II. Pooling

Convolution is a kind of scanning in Machine Language, which identifies the image in forms of pixels, and this group of pixels are then scanned with the help of filters to extract more information from the pixels.

These filters can be of different types like: Curve Filter, Color Filter, Edge Filter, etc., and can also be of different sizes in RGB matrix form of 5*5*5 and so on. Whenever we scanned a single pixel also it gives a random value of an RGB matrix.

MODEL DESCRIPTION

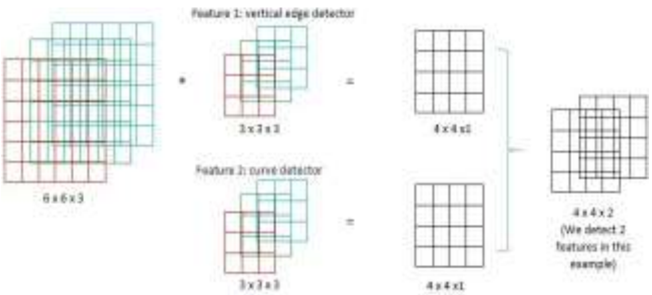

Fig. 1. Convolution using multiple filters [2]

Let us understand the dimensions mathematically:

$$(n \times n \times n_c) * (f \times f \times n_c) * (f \times f \times n_c) \longrightarrow (n - f + 1) \times (n - f + 1) \times \text{no. of filters}$$

CNN ARCHITECTURE

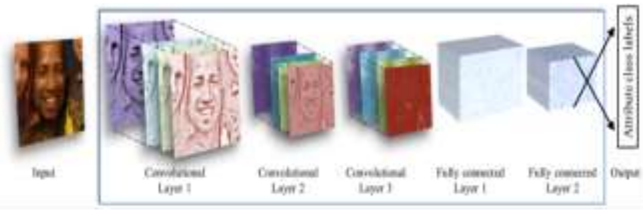

Fig. 2. CNN Architecture [1]

- We used cropped and aligned images are provided in the dataset.
- Performed resizing of the images to 64*64.
- Performed reshaping of the images to 64*64*1.

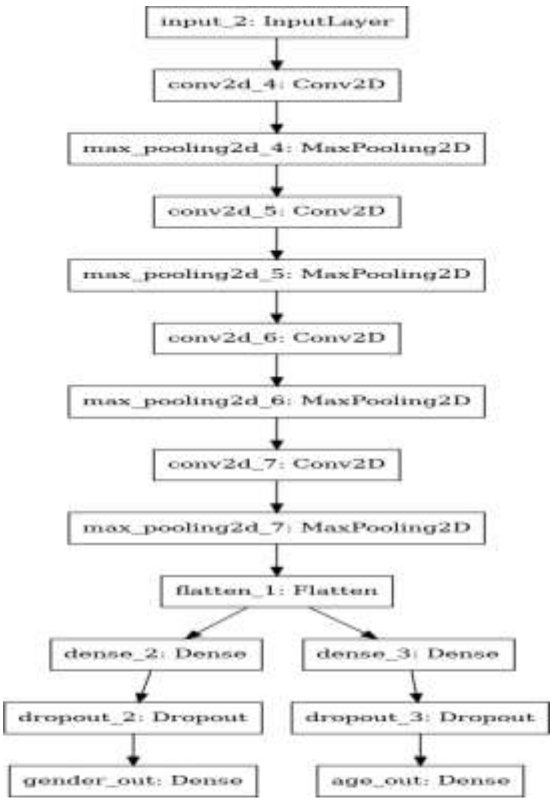

Fig. 3. Model Summary of our program

A corrected direct operation, a pooling sub-cast, and three convolutional layers follow in the network. Original response normalization is also used in the first two layers for normalization. A total of 96 pollutants with a resolution of 7 * 7 pixels are present in the first convolutional layer; 256 pollutants with a resolution of 5 pixels are present in the second convolutional layer; and pollutants with a resolution of 3 pixels are included in the third and final convolutional layer. Two completely connected layers with 512 neurons each are eventually added. All three-color channels are immediately processed by the network. Images are first resized to 256 * 256 pixels and then sent over the network with a 227 * 227 pixels reduction. The definitions of the three convolutional layers are then given:

The first convolutional subcaste adds 96 pollutants with a size of 3 * 7 * 7 pixels to the input, then a corrected direct driver (ReLU), a maximum pooling subcaste that takes the maximum value of 3 regions with a two-pixel method, and an original response normalization subcaste.

The 96 * 28 * 28 matter of the antedating subcaste is processed by the alternative convolutional subcaste, which contains pollutants of size 96 * 5 * 5. For ReLU, a maximum pooling subcaste, and an original response normalization subcaste, the same hyperactive parameters as those employed up front are applied.

After ReLU and a maximum pooling subcaste, the third and last convolutional subcaste applies a set of 384 pollutants with a size of 256 * 3 * 3 pixels to the 256 * 14 * 14 blob.

A ReLU and a powerhouse subcaste come after the first fully connected subcaste, which has 512 neurons and receives the information from the third convolutional subcaste.

A second completely connected subcaste with 512 neurons receives the first completely connected subcaste's 512- dimensional affair before a ReLU and a powerhouse subcaste are added.

A third, completely integrated subcaste that is matched to the highest level of age or gender the final entirely linked subcaste's affair is eventually input into a soft-maximum subcaste, which gives each class a probability.

6. FORMULATION & RELATED PROCESS

A deep learning network's non-linearity property is added by an activation function known as a rectified linear unit (ReLU), which solves the issue of evaporating slants. The conclusive component of its argument is interpreted. This deep learning activation function is one of the most popular ones.

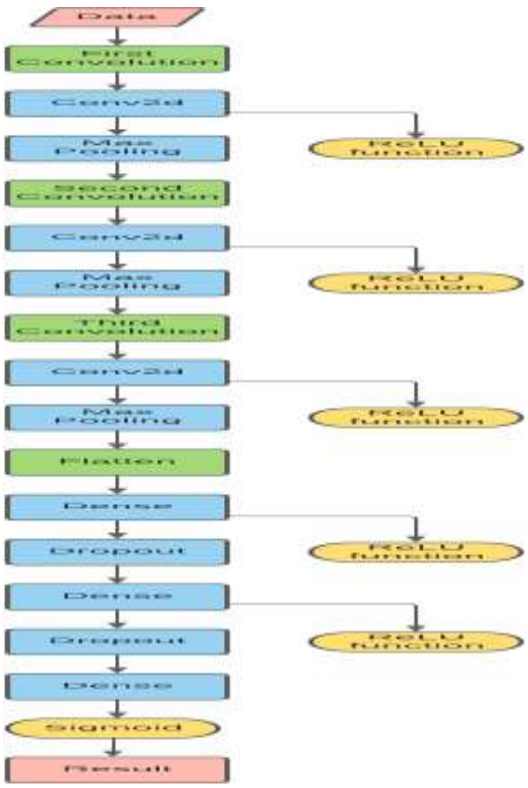

Fig. 4. Model Summary with ReLu functions

Libraries

OpenCV - OpenCV is an open-source computer vision and machine learning package that is free to use [5]. Advanced analytics and real-time picture and video pre-processing are two major uses for this package. Additionally, it is compatible with several Deep Learning frameworks, such as TensorFlow, Caffe, and PyTorch. Because of the way it is set up, OpenCV has several shared or static libraries.

TensorFlow - Google's TensorFlow is a Python library for fast numerical computing. TensorFlow is a foundation library that is primarily used to create deep learning models directly, while wrapper libraries are occasionally used to lubricate the process that is built on top of it.

Keras - The high-level Tensorflow-Keras Application Programming Interface (API) offers a very quick iteration rate in addition to offering the fundamental abstraction and necessary building blocks for the development and application of machine learning solutions. Keras helps engineers and scientists by enabling them to fully take advantage of Tensorflow's scalability and cross-platform capability, which would not have been possible without it. The Keras models may be exported to run in Google Chrome or any other browser on any smartphone or mobile device, and TPU can also function on a sizable cluster of GPUs.

Argparse - The user of this module can create command-line interfaces that are simple to use, clear to comprehend, and easy to access. According to the parameters required to run the application, this module, Argparse, will determine how to parse the commands in sys.argv. If an argument's root cause is false or causing a problem, the module will automatically construct and produce the necessary help and usage messages, as well as issue errors.

Datasets

The dataset for this study is collected manually using Google search, browser extensions, and other online sources to ensure a diverse and representative dataset for CNN-based age and gender classification. The primary focus of data collection is to obtain facial images labeled with age and gender while ensuring diversity in age groups, genders, ethnicities, and environmental conditions. 1. Web Scraping and Image Downloading Facial images were collected using Google Images, Bing, and other search engines, leveraging both automated and manual methods. Web scraping tools such as BeautifulSoup, Selenium, and Scrapy were used to automate the downloading process through Python scripts. Additionally, browser extensions like Fatkun Batch Downloader and Image Downloader were employed to streamline bulk image collection from web pages, making the process more efficient and scalable. 2. Image Filtering and Labeling Each image was visually inspected for quality, framing, and relevance. Where available, EXIF data was extracted for demographic insights. Age and gender were manually annotated based on visible features, with references from celebrity databases, social media, and structured datasets. In unclear cases, age was estimated through facial analysis with expert verification. 3 Ensuring Dataset Diversity and Ethical Compliance To ensure dataset diversity, images were balanced across age groups and gender (Male, Female), with attention to ethnicity and geographic representation. Real-world variations, including lighting, expressions, accessories, and occlusions, were incorporated. Ethical considerations were prioritized by using only publicly available images, ensuring fair representation, and anonymizing data by removing personally identifiable information (PII). The data collection process ensured a diverse and representative dataset using web scraping, manual annotation, and ethical filtering. This approach enhances model robustness by incorporating variations in age, gender, ethnicity, and environmental conditions.

7. RESULTS AND FUTURE SCOPE

The research evaluates the CNN-based Age and Gender Classification model, highlighting its accuracy across various metrics and real-time applications. The model accurately predicted a 1-year-old male, demonstrating its ability to recognize subtle facial features. In live webcam detection, it successfully identified facial attributes, predicting age as 23 years and gender as female with 98% confidence, while also classifying expressions. The model achieved 90% accuracy for gender classification and an age prediction range of ±2 years. The classification report indicates a 95% accuracy in gender classification. The enhanced CNN architecture, incorporating four Conv2D layers and MaxPooling, improved feature extraction and robustness. The age distribution density graph shows the model performs well for ages 20-40 but tends to underpredict older ages. Overall, the model exhibits high accuracy in gender classification and reasonable age estimation, with scope for improvements, particularly for older age groups, to enhance real-world reliability.

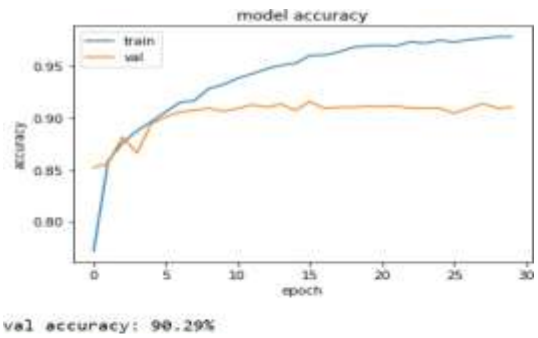

Fig. 5. Accuracy Graph of age & gender training

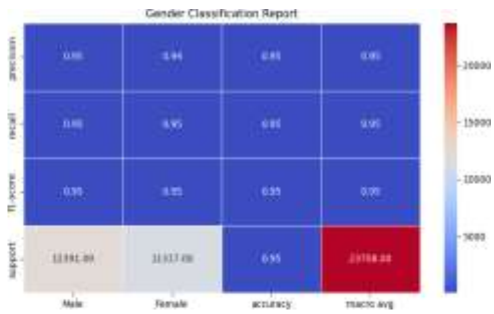

Fig. 6. Gender Classification Report

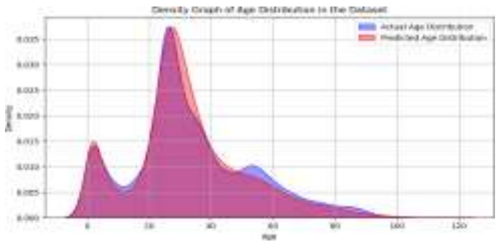

Fig. 7. *Density graph of Age distribution*

Fig. 8. Output of Trained Dataset

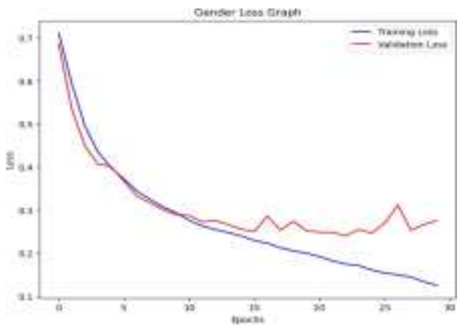

Fig. 9. Loss Graph of Gender training

8. REFERENCES

[1] Ari Ekmekji, "Convolutional Neural Networks for Age and Gender Classification" , Stanford University , 2018.

[2] Khaled Rahman Hassan and Israa Hadi Ali , "Age and Gender Classification using Multiple CNN", 2nd International Scientific Conference of Al-Ayen University, 2020.

[3] Eamon Kelliher, "Human Age and Gender Classification using CNN", Technological University Dublin, 2021.

[4] https://susanqq.github.io/UTKFace/

[5] Xiaofeng Wang, AzlizaMohd Ali, Plamen Angelov," Age and gender classification of human faces for analyzing human behavior", Exeter, UK 2014.

[6] Akash.B.N,Akshay.K Kulkarni, Deekshith. A , Gowtham Gowda," Age and Gender Recognition using Convolution Neural Network",The National Institute of Engineering, Mysuru, India, 7-12 June 2020.

Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

[7] Kotiyal, Arnav, Praveen Gujjar, and Guru Prasad. "Pre-trained Model for Image Prediction Using Keras Application." 2023 International Conference on Computer Science and Emerging Technologies (CSET). IEEE, 2023.

[8] J, P. G., Sharief, M. H., Mukherjee, U., Kumar, M. A., & S, G. P. M. (2023). Malicious Host Detection in Software Defined Networks Using Machine Learning Algorithms. Research Gate, 1–6.

[9] Avinash, S., Kumar, H. N. N., Prasad, M. S. G., Naik, R. M., & Parveen, G. (2023). Early detection of malignant tumor in lungs using Feed-Forward neural Network and K-Nearest Neighbor Classifier. SN Computer Science, 4(2)

[10] Prasad, M. S. G., Pratap, M. S., Jain, P., Gujjar, J. P., Kumar, M. A., & Kukreti, A. (2022). RDI-SD: An Efficient Rice Disease Identification based on Apache Spark and Deep Learning Technique. Research Gate, 277–282

[11] Raza, S. S. (2023). Exploring the conceptual realm of machine learning in small and medium-sized Industries: A Qualitative study. Ushus - Journal of Business Management, 22(4), 1–13.

DOI: 10.55041/IJSREM47461