IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

Age and Gender Identification Through Advanced Deep Learning Technique

Prof. S. V. More, Giram Samarth, Jawale Ujwala, Muluk Nikita, Takale Pallavi
Department of Computer Engineering
Adsul Technical Campus Faculty of Engineering, Chas, Ahmednagar, India

Abstract: *Gender equality initiatives are often faced with a problem: In order to determine whether initiatives are successful the gender of individuals in the target group must be known. As self- identification inherently has the problems that individuals have to respond and results may, therefore, be biased and incomplete, the temptation to use automated gender identification methods is evident. In the scientific literature, multiple sources ranging from the individual's name, their social media choices, biological features (e.g., brain scans or fingerprints), to texts attributed to the individual are used for automated gender identification with varying success. In this paper, we systematically inspect scientific publications for gender prediction based on textual data which are published between January 2017 and January 2019 in order to determine if such approaches may supply viable means to reliably determine an author's gender. However, we find that the best approach in the current state-of-the art works with an accuracy of only 93.4possible harm that gender identification systems might entail due to their inaccuracy and also given that they are assuming a binary gender model. We conclude that gender identification based on textual data is currently no reliable substitute for self- identification.*

Keywords: Age identification, gender recognition, deep learning, CNN, real-time performance

I. INTRODUCTION

Gender identification (or sometimes gender detection) is an active field of research in machine learning and artificial intelligence. The idea is that based on an artifact (e.g., text, voice recording, image) a trained machine can predict the gender of the person behind that artifact. The results of such a prediction is already used on websites to target advertisement based on gender. This may create perception biases and may proliferate gender stereotypes. Gender-API.com — a paid service for gender identification based on the name and location of a person — is reporting over 20 million queries per month by March 2018. There are multiple services like this available which have easy-to-use support for multiple languages. These services are based on name databases which have been constructed using automated learning steps as well as manual work. Karimi et al. have inspected these services and databases and found their accuracies ranging from 74that even the best approach is incorrectly predicting a person's gen der in 9when combining three online services they were able to reach an ac curacy of 98unclassified. Another problem we (and others [8, 13, 21]) see is the implicit assumption of a binary gender model which discriminates against persons of non-binary genders. Moreover, the given name of a person might not be available. Various collaborative platforms on the Internet (e.g., GitHub) do not require the actual name but rather rely on pseudonyms. In order to make gender identification more robust against these issues, style-based analysis of a user's provided textual artifacts is considered as an alternative. In this paper, we systematically review the scientific literature on gender identification and com pare the accuracy of machine learning approaches for gender recognition relying on textual artifacts. The idea of such machine learning based approaches is to take textual input the user provides over time to query a trained machine for the gender. There are other approaches that aim to determine gender by still images, video, or audio data, which we are not inspecting in this paper.

II. MOTIVATION OF THE PROJECT

Personalization:

Age and gender identification helps in providing personalized content and services, particularly in e- commerce, digital marketing, and social media platforms. Security and Surveillance: In surveillance systems, age and gender recognition can be used to enhance facial recognition systems, allowing for better tracking and identification of individuals in real-time.

Healthcare:

Such systems may help in applications related to health and wellness, particularly in managing services for different demographic groups. Human- Computer Interaction: Understanding a user's age and gender can help in creating more adaptive and intuitive user interfaces, such as virtual assistants or interactive robots.

III. LITERATURE SURVEY

[1]. COVID-19 detection based on Computer Vision and Big Data- Mark Wu1, University of Wisconsin- Madison, Madison, United States, hwu378@wisc.edu – For detecting COVID-19 and checking the severity of the patient's condition, CT examination of the lungs is significant. However, the current manual viewing of CT images requires professionalism. In order to improve the inspection efficiency of the huge number of CT images, it is necessary to develop an intelligent detection algorithm to perform CT inspections. This paper proposes a COVID-19 detection algorithm based on EfficientDet. EfficientDet leverages a faster and easier multi-scale fusion approach, which is more suitable for COVID 19 detection tasks with finer feature granularity. In addition, data augmentation is also significant in COVID-19 detection tasks. This paper verifies the effectiveness of EfficientDet on the SIIM FISABIO-RSNA COVID-19 Detection dataset provided by Kaggle plat form. Experimental results show that EfficientDet has achieved better performance than other detection algorithm Taking MAP@0.5 as an indicator, EfficientDet reaches 0.545, which is 7.9and YOLO-V5. Index Terms—COVID-19 Detection, EfficientDet, data augmentation, Big Data**.**

[2]. Review Paper for Detection of COVID-19 from Medical Images and/ or Symptoms of Patient using Machine Learning Approaches- The new type1of corona virus COVID-190 virus was first detected in Wuhan China- A COVID-19 certified patient is characterized by fever, fatigue, and dry cough. The coronavirus (COVID-19) epidemic is spreading worldwide. In this re view paper, a database of X- ray, CT-Scan images from 0patients with common1bacterial pneumonia, 0confirmed Covid- 191infection, and common cases, were used to automatically detect Coronavirus infection. The purpose of the study was to evaluate the effectiveness of COVID-19 acquisition. During the COVID-19 scenario, the number of infected cases rises in huge number globally. Due to this fact, a vital decision had been taken by medical experts and infected patients to adopt various medical facilities within a reasonable amount of time. Keywords: COVID-19, X-ray, Deep learning, Artificial intelligence (AI), Healthcare, Machine learning (ML).

[3]. Five Strategies for Bias Estimation in Artificial Intelligence-based Hybrid Deep Learning for Acute Respiratory Distress Syndrome COVID-19 Lung Infected Patients using AP (ai)Bias 2.0: A Systematic Review- Jasjit S. Suri, Fellow IEEE, Sushant Agarwal- Corona virus 2019 (COVID-19) has led to a global pandemic infecting 224 million people and has caused 4.6 million deaths. Nearly 80 Artificial Intelligence (AI) articles have been published on COVID-19 diagnosis. The first system atic review on the Deep Learning (DL)- based paradigm for COVID-19 diagnosis was recently published by Suri et al. [IEEE J Biomed Health Inform. 2021]. The above study used AtheroPoint's "AP (ai)Bias 1.0" using 10 AI attributes in the DL framework. The proposed study uses "AP (ai)Bias 2.0" as part of the three quantitative paradigms for Risk-of-Bias quantification by using the best 40 dedicated Hybrid DL (HDL) studies and utilizing 39 AI attributes. In the first method, the radial bias map (RBM) was computed for each AI study, followed by the computation of bias value. In the second method, the regional-bias area (RBA) was computed by the area difference between the best and the worst AI performing attributes. In the third method, ranking-bias score (RBS) was computed, where AI-based cumulative scores were computed for all the 40 studies. These studies were ranked, and the cut off was determined, categorizing the HDL studies into three bins: low, moderate, and high. Using the Venn diagram, these three quantitative methods were benchmarked against the two qualitative non-randomized-based AI trial methods (ROBINS-I and PROBAST). Using the analytically derived moderate-high and low moderate cutoff of 2.9 and 3.6, respectively, we observed 4027.5RBM, RBA, RBS, ROBINS-I, and PROBAST, respectively. We present an eight- point recommendation for AP (ai)Bias 2.0 minimization. Index Terms—COVID-19 diagnosis, HDL, risk-of-bias, radial regional- ranking, PROBAST-ROBINS-I, AP (ai)Bias 2.0.

[4]. COVID-19 Identification from Chest X- Rays- -Iosif Mporas School of Physics Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom- Artificial Intelligence and Data Science community has contributed to the global response against the new coronavirus, COVID-19. Significant attention has been given to detection and diagnosis tools with rapid diagnostic tools based on X-rays using deep learning being proposed. In this paper we present an evaluation of several well-known pretrained deep CNN models in a transfer learning setup for COVID-19 detection from chest X-ray images. Two different publicly available datasets were employed and different setups were tested using each of them separately of mixing them. The best performing models among the evaluated ones were the DenseNet, ResNet and Xception models, with the results indicating the possibility of identifying COVID-19 positive cases from chest X-ray images. Keywords— COVID-19, X-rays, transfer learning.

IV. PROBLEM DEFINITION AND OBJECTIVE

The ability to automatically identify age and gender from facial images has become increasingly important due to its wide range of applications across various industries such as security, marketing, healthcare, and entertainment. The task of age and gender identification presents several challenges due to the variability of human faces in terms of age, gender, lighting, facial expressions, and occlusions (e.g., glasses, hats, etc.). Traditional methods for age and gender prediction were largely rule-based or statistical, relying on simple features extracted from statistical, relying on simple features extracted from images. However, these methods often fail to account for the complex and high- dimensional nature of human faces. With the advent of deep learning, especially convolutional neural networks (CNNs), the performance of age and gender classification has dramatically improved, enabling more accurate predictions.

Objective:

- To develop a deep learning-based system that can accurately identify the age and gender of individuals based on facial images using CNNs
- To enhance the robustness and generalization of the system by training it on large, diverse datasets that cover a wide range of ages, genders, ethnicities, and lighting conditions.
- To achieve high accuracy in age and gender classification, reducing error rates for real-world applications.
- To enable real-time performance, with an emphasis on processing speed and minimal latency in applications such as security or customer analytics.

V. SYSTEM OVERVIEW

In this sub-section, we show the overall system of our proposed gender classification. Figure 1 shows the flow chart of the proposed gender classification. As can be shown in Figure 1, before classifying gender, the age estimation is used, and we extract different feature for gender classification, based on this age group division.

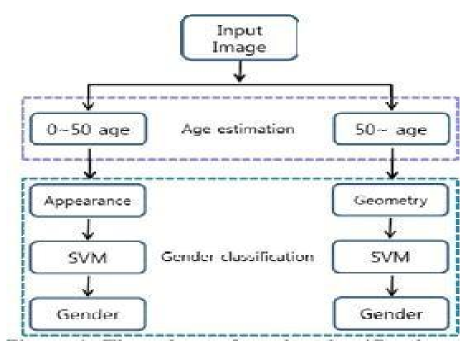

Fig. Flow Chart of Gender Classification

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

VI. FUNCTIONAL REQUIREMENTS:

System Feature

The system must allow the collection of facial image data with labeled information for both age and gender.

The system must calculate and output performance metrics for gender classification, such as accuracy, precision, recall.

The system should allow real-time or near-real- time predictions of age and gender from live camera feeds or new input images.

The system should comply with relevant data privacy regulations when handling user data.

VII. NON-FUNCTIONAL REQUIREMENTS:

Performance Requirements:

The system must achieve high classification accuracy for both age and gender prediction. Common performance metrics include precision, recall, F1 score, and overall classification accuracy.

The model should be able to scale effectively with the increasing size and diversity of datasets, potentially using transfer learning or fine-tuning techniques.

The system should perform well under various real-world conditions such as varying lighting, occlusions (e.g., glasses, facial hair), different poses, and diverse ethnic backgrounds.

Safety Requirements:

Age and gender identification systems typically involve personal data (e.g., facial images), which must be handled in compliance with privacy regulations such as GDPR or CCPA. Appropriate safeguards, such as anonymization and encryption, should be implemented to protect user data

The model should be trained and tested on diverse datasets to avoid bias towards certain demographic groups. It must avoid producing discriminatory results based on race, age, gender, or ethnicity. Regular audits for fairness should be conducted.

Safety mechanisms should be in place to handle scenarios where the model might fail, such as clear fallback mechanisms or manual verification options in high-stakes use cases (e.g., security).

Security Requirements:

The system should be secure from both data breaches and adversarial attacks. This includes protecting against spoofing (where attackers use images of faces that do not belong to them) and ensuring the integrity of the underlying model.

The system should only collect and store the minimal amount of data necessary for age and gender prediction. This reduces the risk associated with storing and processing unnecessary information.

The system should maintain detailed audit logs of user activities, including image uploads, predictions made, and system access. These logs should be stored securely and be accessible only to authorized personnel for troubleshooting or auditing purposes.

The system should ensure that only authorized users can access it. The system must require multi-factor authentication for admin users.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, April 2025 Impact Factor: 7.67

VIII. SYSTEM ARCHITECTURE

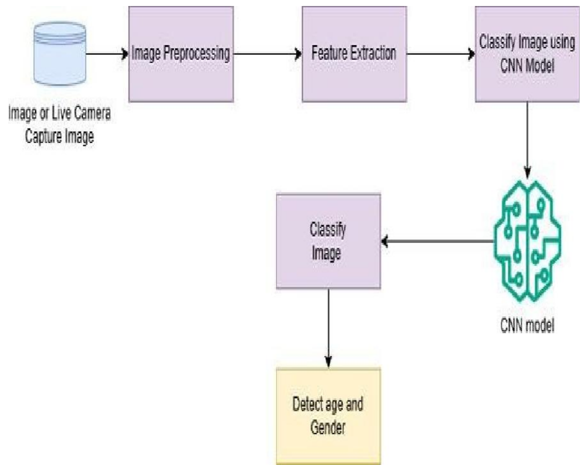

Fig. System Architecture

IX. PROJECT PLAN

we are going to have an overview about how much time does it took to complete each task like- Preliminary Survey Introduction and Problem Statement, Literature Survey, Project Statement, Software Requirement and system Design, Specification, Partial Report Submission, Architecture Design, Implementation

Deployment, Testing, Paper Publish, Report Submission and etcetera. This chapter also gives focus on stakeholder list which gives information about project type, customer of the proposed system, user and project member who developed the system.

IMPLEMENTATION STEPS: -

Project Planning: To develop a deep learning model for age and gender identification from images using advanced techniques.

Data Collection & Preprocessing: Identify publicly available datasets for age and gender classification. Check if additional data needs to be collected or curated. Ensure data diversity in terms of age, gender, ethnicity, and lighting conditions.

Model Development: Design a deep learning model that can predict both age and gender. Add two separate output layers: one for age classification (categorical or regression) and one for gender classification (binary).

Model Evaluation and Optimization: Evaluate the model on the test set and analyze metrics like accuracy, confusion matrix, F1 score, and precision/recall. If the model performs poorly on age or gender classification, analyze error cases and consider model adjustments (e.g., adding more layers, using a different architecture). Implement model compression (e.g., pruning, quantization) for faster inference, especially for mobile or web deployment.

Deployment: Develop a simple API using Flask or Fast API that accepts image inputs and returns the predicted age and gender. Perform integration testing to ensure the model works with the frontend and other system components. Monitoring & Maintenance: Continuously monitor the model's performance in the production environment. As new data becomes available, periodically retrain the model to improve its accuracy and generalization.

Risks and Mitigation: Low-quality or biased data could affect model accuracy. Solution: Ensure balanced, representative datasets and perform thorough data augmentation. The model might overfit on the training data. Solution: Use regularization methods, early stopping, and cross-

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

validation. The model might overfit on the training data. Solution: Use regularization methods, early stopping, and cross-validation.

X. RESULTS:

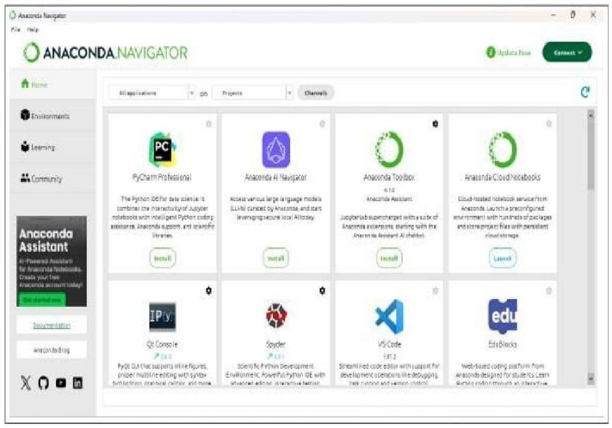

Fig. a. Gui Main page

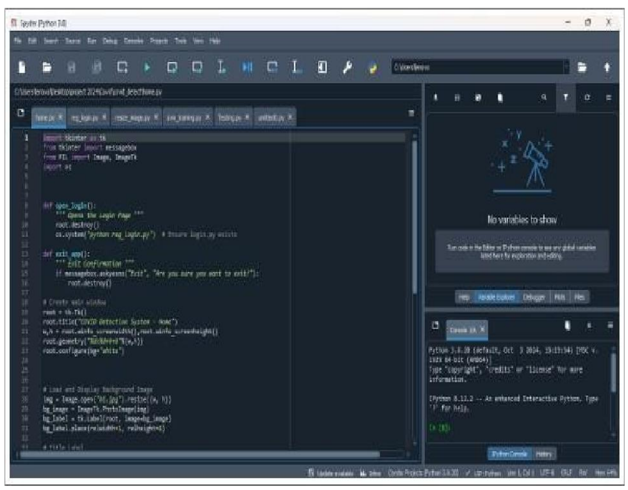

Fig. b: login page

Fig. c. register page

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

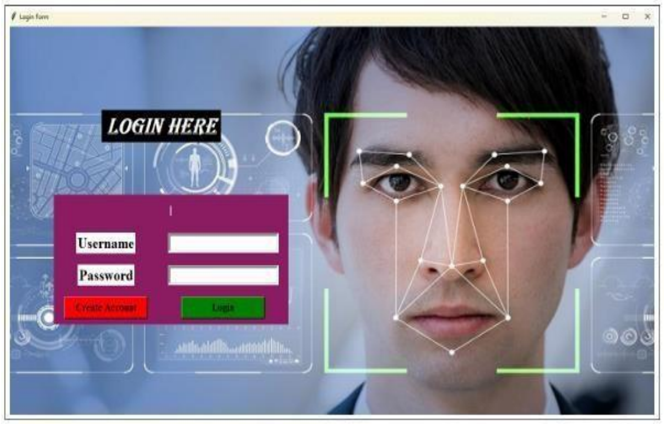

Fig. d: output

Fig. e: output

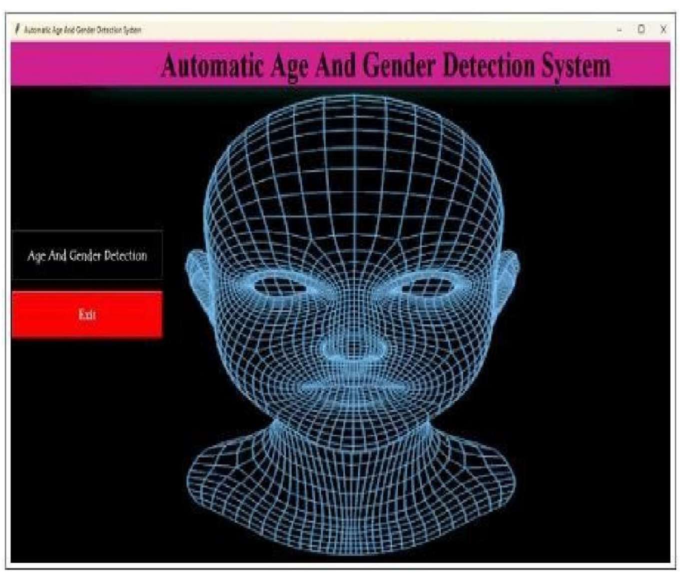

Fig. f: output

DOI: 10.48175/568

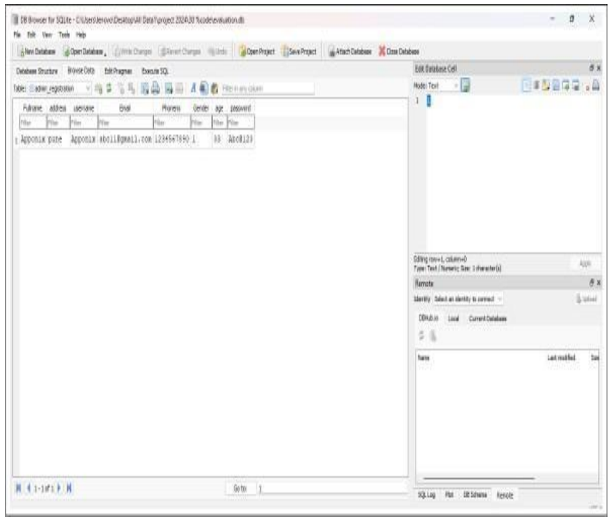

Fig. g: Database

XI. CONCLUSION

The development of a robust, accurate, and scalable system for age and gender identification from facial images using advanced deep learning techniques, particularly Convolutional Neural Networks (CNNs), is a highly impactful and relevant task with wide-ranging applications in areas such as security, personalization, marketing, and human-computer interaction. However, several challenges must be overcome to create a system that is both technically sound and ethically responsible. Key objectives include ensuring high classification accuracy, minimizing biases, handling diverse real-world conditions (such as varying lighting, camera angles, and facial features), and achieving real- time processing for practical deployments. additionally, the system must be designed with user privacy and data security in mind, ensuring compliance with privacy regulations and providing mechanisms to safe guard sensitive information. While there are numerous challenges in terms of model performance, fairness, and scalability, deep learning techniques particularly CNNs have shown immense potential in this domain. Through thoughtful design and continuous training with diverse and representative datasets, it is possible to build a system that generalizes well, operates efficiently, and adapts to different use cases across industries. In conclusion, the successful deployment of an age and gender identification system can provide significant value, ranging from improved user experiences in personalized services to enhanced security in surveillance applications. However, attention must be paid to bias mitigation, real-time performance, and privacy concerns to ensure that such technologies are used responsibly and equitably across different de myographic. The future of this technology lies in refining these models, ensuring fairness, and making them adaptable to an increasingly diverse and dynamic world.

REFERENCES

[1]. Francesco Barbieri. 2017. Shared Task on Stance and Gender Detection in Tweets on Catalan Independence-LaSTUS System Description. In Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017. 217-221.

[2]. Rupal Bhargava, Gunjan Goel, Anjali Shah, and Yashvardhan Sharma. 2017. Gender Identification in Russian Texts. In Working notes of FIRE 2017- Fo rum for Information Retrieval Evaluation, Bangalore, India, December 8-10, 2017.13-16.

[3]. Bassem Bsir and Mounir Zrigui. 2018. Bidirectional LSTM for Author Gen der Identification. In Computational Collective Intelligence- 10th International Conference, ICCCI2018, Bristol, UK, September 5-7, 2018, Proceedings, Parti. 393-402.

[4]. Bassem Bsir and Mounir Zrigui. 2018. Enhancing Deep Learning Gender Identification with Gated Recurrent Units Architecture in Social Text. Computation y Sistemas 22, 3 (2018). http://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/ view/3036

[5]. Juan Soler Company and Leo Wanner. 2018. On the role of syntactic dependencies and discourse relations for author and gender identification. Pattern Recognition Letters 105 (2018), 87-95.

[6]. Saman Daneshvar and Diana Inkpen. 2018. Gender Identification in Twitter using N grams and LSA: Notebook for PAN at CLEF 2018. In Working Notes of CLEF 2018- Conference and Labs of the Evaluation Forum, Avignon, France, Septem.