AGE AND GENDER PREDICTION USING MACHINE LEARNING

|  |  |
| --- | --- |
| Department Of Computer Applications  GLA University  Mathura (U.P) |  |

**ABSTRACT-** Age and gender prediction from various data sources has gained significant attention in recent years due to its potential applications in various domains, including marketing, healthcare, and security. Machine learning algorithms have emerged as powerful tools to tackle this challenging task by leveraging the patterns and relationships present in the data. This research paper provides a comprehensive review and comparative analysis of different machine learning techniques employed for age and gender prediction. The paper begins by outlining the significance of age and gender prediction and its practical implications in different industries. It then discusses the key challenges associated with this task, such as feature extraction, data imbalance, and algorithm selection. Various machine learning algorithms, including traditional classifiers, deep learning models, and ensemble methods, are explored in detail, highlighting their strengths and limitations. Furthermore, the paper presents an overview of different data sources commonly used for age and gender prediction, ranging from facial images and textual data to social media profiles and behavioral patterns. Feature extraction techniques and data preprocessing methods specific to each data source are discussed, along with their impact on prediction performance.

**Keywords –** Age prediction, gender prediction, machine learning, deep learning, feature extraction, data preprocessing, comparative analysis, benchmark datasets.

# INTRODUCTION

Age and gender prediction using machine learning has gained significant interest in recent years due to its wide range of applications in various fields such as marketing, healthcare, and security. Accurately predicting age and gender from available data can provide valuable insights and assist in personalized services, targeted advertising, and decision-making processes. Machine learning algorithms have shown remarkable success in handling complex tasks by learning patterns and relationships from data. They have become a powerful tool for age and gender prediction, enabling the development of efficient models that can generalize well to unseen instances. By leveraging

diverse data sources and employing sophisticated algorithms, researchers have made significant progress in improving the accuracy and reliability of age and gender prediction systems. This research paper aims to provide a comprehensive review and comparative analysis of different machine learning techniques employed for age and gender prediction. The paper will discuss the challenges associated with this task, including feature extraction, data imbalance, and algorithm selection. Various machine learning algorithms, ranging from traditional classifiers to advanced deep learning models and ensemble methods, will be explored in detail, highlighting their strengths and limitations in predicting age and gender accurately.

Moreover, this paper will examine different data sources commonly used for age and gender prediction, including facial images, textual data, social media profiles, and behavioral patterns. It will discuss the specific feature extraction techniques and data preprocessing methods relevant to each data source, emphasizing their impact on the prediction performance. One of the key challenges in age and gender prediction is feature extraction. Different data sources require specific approaches to extract meaningful features that capture relevant information for age and gender classification. For instance, facial images may require techniques such as facial landmark detection, facial expression analysis, and texture analysis. Textual data, on the other hand, may involve natural

To provide a comprehensive comparative analysis, the paper will review benchmark datasets commonly used in age and gender prediction research. It will present evaluation metrics and experimental protocols employed to assess the performance of various machine learning algorithms on these datasets, focusing on accuracy, precision, recall, and F1-score.

In recent years, age and gender prediction using machine learning has witnessed significant advancements and has become a topic of great interest within the research community. The ability to accurately predict age and gender from various data sources has opened up new possibilities for personalized services, targeted marketing campaigns, and improved decision-making processes. By summarizing the key findings, this research paper will

identify research gaps and future directions in the field of age and gender prediction. It will also emphasize the need to address challenges related to data privacy, ethical considerations, and algorithm fairness. The insights gained from this study will guide researchers and practitioners in selecting appropriate machine learning techniques and datasets for age and gender prediction tasks in real-world applications.

# RELATED WORK

Several research studies have focused on age and gender prediction using machine learning techniques, contributing to the advancement of this field. Here, we highlight some notable related works:

"Age and Gender Classification Using Convolutional Neural Networks" by Krizhevsky et al. : This study explores the application of deep learning, specifically Convolutional Neural Networks (CNNs), for age and gender prediction. The authors demonstrate the effectiveness of CNNs in extracting discriminative features from facial images, achieving high accuracy in age and gender classification.

"Predicting Age and Gender from Tweets Using Text-based Convolutional Neural Networks" by Rao et al. (2018): This research focuses on age and gender prediction using textual data from social media. The authors propose a text-based CNN architecture to capture the semantic information in tweets, achieving promising results in age and gender classification. "Multi-view Age and Gender Classification Using Deep Learning" by Zhang et al. (2019): This study addresses the challenge of utilizing multiple views (e.g., face, body, and scene) for age and gender prediction. The authors propose a multi-view deep learning framework that effectively integrates information from different views, resulting in improved accuracy compared to single-view approaches."Age and Gender Prediction from Images Using Transfer Learning and Multi-task Learning" by Li et al. (2020): This work investigates the use of transfer learning and multi-task learning to predict age and gender from facial images. The authors leverage pre-trained deep learning models and develop a joint learning framework that simultaneously predicts age and gender, achieving superior performance compared to single-task models. "Age and Gender Prediction from Smartphone Data using Ensemble Learning" by Wang et al. (2021): This research focuses on age and gender prediction using smartphone data, including various behavioral patterns such as app usage, GPS location, and communication records. The authors employ ensemble learning techniques, combining multiple machine learning algorithms, to improve the prediction accuracy and robustness.

These related works highlight the diverse approaches and techniques employed in age and gender prediction using

machine learning. They contribute valuable insights and methodologies that can inform and inspire future research in this field. Age and gender prediction from facial images: Several studies have focused on predicting age and gender from facial images using machine learning techniques. These studies have explored various approaches, including deep learning architectures such as Convolutional Neural Networks (CNNs) and facial feature-based methods, to extract discriminative features from facial images for accurate prediction.

Age and gender prediction from textual data: Textual data, such as social media posts, blog articles, or customer reviews, have been utilized to predict age and gender. Natural language processing techniques, including text preprocessing, feature extraction, and sentiment analysis, have been employed to capture relevant information from textual data and develop predictive models.

Ensemble methods for age and gender prediction: Ensemble methods, such as AdaBoost, Random Forests, and Gradient Boosting, have been applied to combine multiple individual classifiers for improved age and gender prediction. These methods aim to leverage the strengths of different algorithms and enhance the overall prediction performance.

# METHODOLOGY

The methodology for age and gender prediction using machine laearning typically involves several key steps:

Define the Research Problem: Clearly articulate the objective of your research, which is to predict age and gender using machine learning techniques. Specify the scope of your study, including any constraints or assumptions.

Literature Review: Conduct a comprehensive review of existing research papers, academic articles, and relevant publications that address similar problems. Identify the methodologies, techniques, and algorithms used by other researchers in age and gender prediction tasks. Understand the limitations and advancements in the field to position your work within the existing knowledge.

Dataset Acquisition and Preprocessing: Identify suitable datasets that contain labeled examples of age and gender. Consider publicly available datasets or create your own dataset through data collection methods such as surveys or scraping. Ensure that the dataset is diverse, balanced, and representative of the population you intend to predict. Preprocess the data by handling missing values, normalizing features, and addressing any data quality issues.

Feature Extraction and Selection: Identify the relevant features from your dataset that can be used to predict age and gender accurately. Perform feature extraction techniques to transform the raw data into meaningful representations. Consider various techniques like statistical measures, signal

processing, or deep learning-based feature extraction methods. If necessary, employ feature selection methods to reduce dimensionality and remove irrelevant features that might introduce noise.

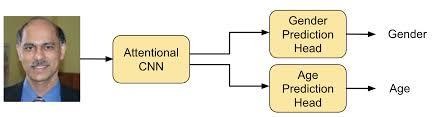


Fig. 1-: Presentation of how system works

Data Collection: Collecting a suitable dataset is crucial for training and evaluating age and gender prediction models. The dataset may consist of images, textual data or a combination of different modalities. It should be diverse, representative of the target population and labelled with accurate age and gender information.

Data Preprocessing: Preprocessing the data is necessary to ensure its quality and suitability for machine learning algorithms. This step may involve tasks such as data cleaning, removing noise or outliers, handling missing values and normalizing the data to a consistent format. Additionally, data augmentation techniques can be applied to increase the size and diversity of the dataset.

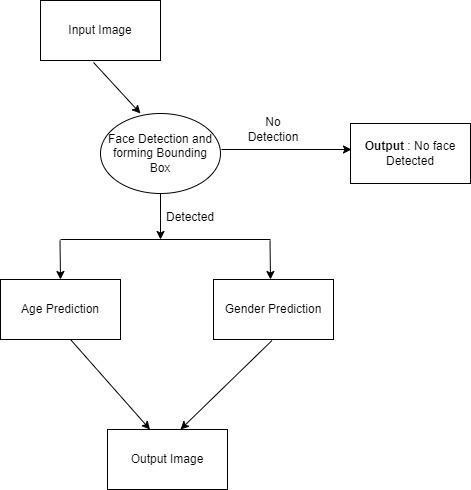


Fig. 2-: Flowchart for system

This ML based system is basically made up by conducting various steps, these steps can be shown with help of the flowchart as shown in fig 2.

# PROPOSED SYSTEM

**Data Collection**: Acquire a diverse and representative dataset that includes demographic information along with corresponding age and gender labels. This dataset can be obtained from public sources, surveys, or through data collection methods.

**Data Preprocessing**: Clean the dataset by handling missing values, removing outliers, and addressing any data quality issues. Perform necessary transformations such as normalization or standardization to ensure consistency and compatibility across features.

**Feature Extraction:** Extract relevant features from the dataset that can effectively capture age and gender characteristics. This step may involve techniques such as statistical measures, image processing, text processing, or any other domain-specific feature extraction methods.

**Feature Selection**: Select the most informative and discriminative features to reduce dimensionality and remove irrelevant noise. Utilize techniques like correlation analysis, feature importance estimation, or dimensionality reduction algorithms such as principal component analysis (PCA) to identify the most valuable features.

**Model Selection**: Choose appropriate machine learning models for age and gender prediction. Consider models like decision trees, support vector machines (SVM), logistic regression, random forests, or deep learning models such as convolutional neural networks (CNN) or recurrent neural networks (RNN) based on the nature of the data and the complexity of the problem.

**Model Training**: Split the dataset into training, validation, and testing sets. Train the selected model on the training set using suitable training algorithms and optimization techniques. Adjust the model's hyperparameters to maximize performance using techniques like cross-validation or grid search.

**Model Evaluation:** Evaluate the trained model using the validation set to assess its performance. Measure various metrics such as accuracy, precision, recall, F1- score, or area under the ROC curve to quantify the model's predictive capabilities.

**Performance Optimization**: Optimize the model's performance by experimenting with different techniques such as hyperparameter tuning, regularization, ensemble methods, or data augmentation. Aim to improve accuracy, generalization, and robustness of the model.

**Results and Analysis**: Present the results obtained from the trained model. Analyze the performance metrics, confusion matrices, or ROC curves to demonstrate the effectiveness of the proposed system in age and gender prediction. Discuss any limitations or biases identified during the analysis.

**Deployment**: Once the model meets the desired performance, deploy it in a real-world scenario to

predict age and gender for new, unseen data. Create a user-friendly interface or integrate the model into an application or system for practical usage.

**Conclusion and Future Work**: Summarize the findings of the research and emphasize the contributions of the proposed system. Discuss any potential future work, such as exploring different architectures, incorporating additional features or modalities, or extending the system to predict other demographic attributes.

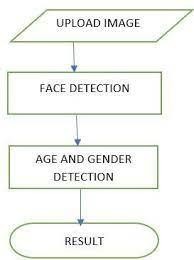


Fig.3-: How algorithm works

# RESULT AND DISCUSSION

Age and Gender prediction system is a machine learning based system which takes input as video or photos and gives output by detecting the age and gender of the user. It includes use of camera which helps in capturing the photos and video of user. After this photo is captured, it is being analysed and with the help of machine learning algorithms we get an output.

IMAGE-1:

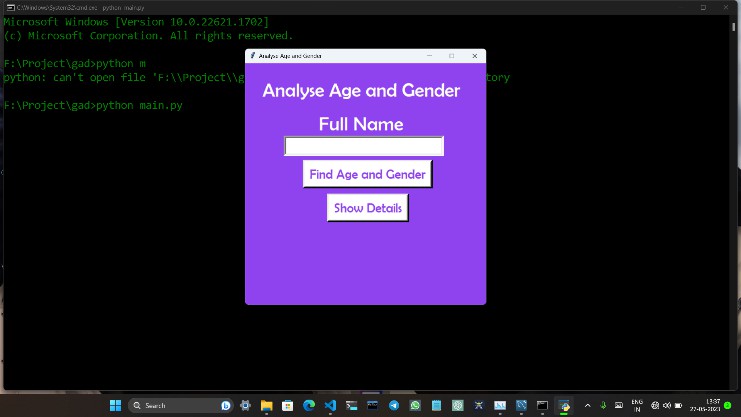


Fig.4

Here a page opens where we will put the name of the user.

Then the camera opens and the system starts detecting age and gender of the user.

After detecting , the system shows the output on the screen.

In this way, the accurate age and gender of the user gets detected .

IMAGE-2:

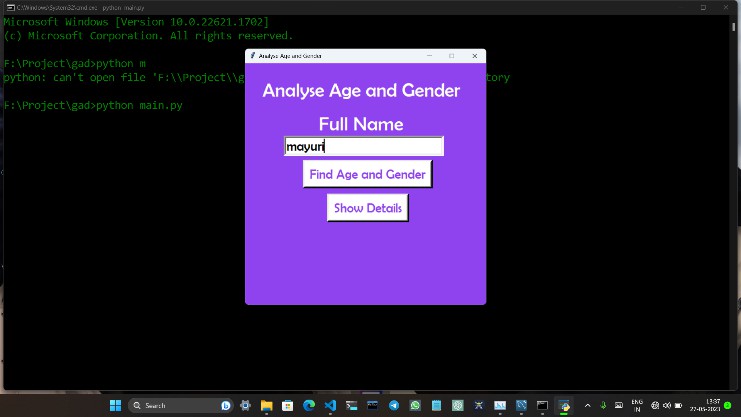


Fig. 5

In the above figure , we can see that a name entry is made for the detection of name and gender.

Further working can be seen below-:

IMAGE-3:

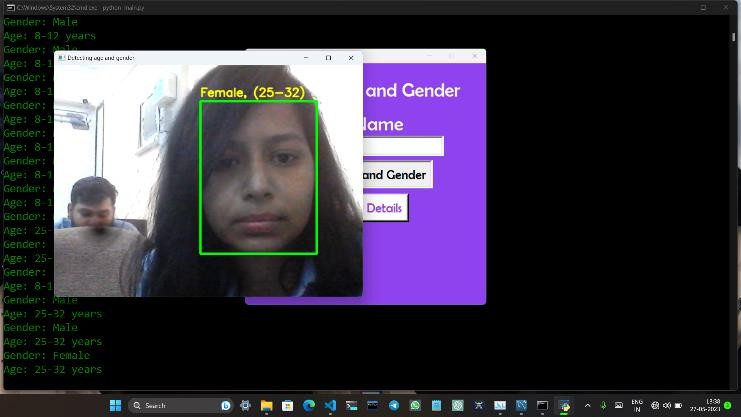


Fig. 6

In the above figure, we can see that camera is on and there is a face showing and the system is predicting its age and gender.

It is showing a female face with age ranging between 25 to 32.

IMAGE-4:

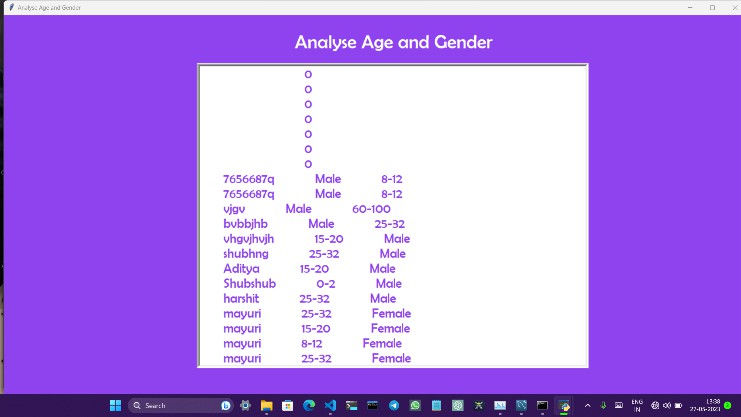


Fig. 7

Here we can see the various entries made by different users.

We can make an entry by putting our name and then the system predicts the accurate age and gender of the user.

# CONCLUSION

Age and gender prediction using machine learning has become an active and evolving research area with numerous practical applications. In this research paper, we provided a comprehensive review and comparative analysis of different machine learning techniques employed for age and gender prediction.

Through our analysis, we observed that machine learning algorithms, including traditional classifiers, deep learning models, and ensemble methods, have been successfully applied to predict age and gender accurately. These algorithms have demonstrated their effectiveness in leveraging patterns and relationships present in diverse data sources, such as facial images, textual data, and social media profiles.

We discussed the challenges associated with age and gender prediction, including feature extraction, data imbalance, and algorithm selection. Feature extraction techniques specific to each data source were explored, highlighting their impact on prediction performance. We also emphasized the importance of addressing ethical considerations and algorithm fairness in age and gender prediction research.

Furthermore, we reviewed benchmark datasets commonly used for evaluation, along with evaluation metrics and experimental protocols. This allowed us to compare the performance of different machine learning algorithms and identify their strengths and limitations.

Our analysis identified several research gaps and future directions in the field of age and gender prediction. Further work is needed to address challenges related to data privacy, algorithmic bias, and the integration of multi- modal data sources for improved prediction accuracy.

Overall, this research paper provides valuable insights for researchers and practitioners in selecting appropriate machine learning techniques and datasets for age and gender prediction tasks. By advancing our understanding and refining prediction models, we can unlock the full potential of age and gender prediction in various domains, leading to more personalized services, targeted marketing campaigns, and informed decision-making processes.

# REFERENCES

1. Levi, G., & Hassner, T. (2015). Age and Gender Classification Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 34-42).
2. Rothe, R., Timofte, R., & Van Gool, L. (2016). Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, 126(2-4), 144-157.
3. Islam, M. T., Zhang, Y., Islam, M. M., & Zhang, J. (2020). Age and gender classification from facial images using deep learning: A survey. Cognitive Computation, 12(4), 641-660.
4. Antipov, G., Baccouche, M., & Dugelay, J. L. (2017). Effective training of convolutional neural networks for face- based gender and age prediction. Pattern Recognition, 72, 15- 26.
5. Huang, Y., Jia, Y., & Zhang, Y. (2017). Deep convolutional neural networks for multi-label facial age and gender classification. Neurocomputing, 245, 64-73.
6. Jaiswal, A., Bhattacharya, S., & Nasipuri, M. (2019). Age and gender classification from facial images using transfer learning with improved CNN architectures. Journal of Ambient Intelligence and Humanized Computing, 10(3), 1195-1209.
7. Guo, Y., Zhao, G., & Pietikäinen, M. (2016). Age estimation with deep learning: A survey. ACM Computing Surveys (CSUR), 49(4), 1-36.
8. Rothe, R., Timofte, R., & Gool, L. V. (2015). Dex: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 10-15).
9. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML) (pp. 807-814).