International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Multimodal Approach for Age Estimation and Verification of Social Media Users

Rekha Saraswat[1], Shubham Tripathi[2]

[1]Scientist E, Education and Training Department, CDAC Noida, Noida, India
[2]Project Engineer, Education and Training Department, CDAC Noida, India

Abstract

Social media platforms contain content that is inappropriate for children and teens, ranging from explicit content to cyberbullying. The implementation of age verification can serve as a crucial barrier, which can prevent minors from accessing harmful content. This also helps reduce the chances of encountering online predators. Furthermore, age verification can help social media platforms adhere to legal and regulatory standards. This can also help social media platforms gain more user trust. This is especially important for parents, as it gives off a sense of security and reassures users that the social media platform is dedicated to protecting children and teens by being responsible and trustworthy. The traditional approach of age verification is inaccurate, as it is based on unverified user data, which is easily falsified. This is because user data is notoriously inaccurate. This paper focuses on a multi-modal approach for age estimation and verification of a user's profile on social media platforms by incorporating both image and text data. In this paper, a comprehensive decision matrix is also proposed that is aimed at finalizing the age categorization through the incorporation of user-specific characteristics as well as platform-specific content attributes.

Keywords: Image analysis, Text analysis, Decision Matrix, Convolutional neural networks (CNN), Random Forest, Natural Language Processing (NLP), Demographics, Age Estimation, Age Verification, Educational.

1. Introduction

Considering that our digital paradigm is becoming more dynamic, it is not a surprise that online social media sites have become so essential to how we communicate and share information, these sites have presented a big challenge in confirming the age status of all their end users, this is essential to effectively deal with such content that needs to be shared only up to a certain age, speaking within legal parameters, especially with certain provisions protecting minors. Anyone can simply misrepresent their age, thus making the conventional age verification process, which asks users to provide such information, inadequate. Most age verification methods for minors involve prompting the users to input their date of birth or some other form of attempting to get them to state they are not up-to-date when enrolling. On the surface, this seems an easy method however there are many pitfalls of executing it: Lack of accuracy - Users might fake their age, resulting in unreasonable inaccuracies. Furthermore, it provides a workaround for age restrictions where the user could enter fake birthdate; Requires Detailed Information - Whereas the need for specific personal details related to your real life could become an issue that deters more

conservative users from signing up; Legal compliance - Many countries have strict laws about age verification, especially for content that is only allowed to people of a certain age. The traditional methods are not being able to fulfil these legal requirements properly.

2. Literature Survey

In the recent years, age estimation and verification have been one of the most discussed topics in machine learning/computer vision field. Facial analysis, text-based methods and many other techniques are explored by researchers to make age estimation systems more accurate or reliable. As per Rothe R., Timofte R., & Van Gool L.[(1)], Deep Expectation of Apparent Age (DEX) model has been first introduced by him and the model has used CNN to recognizing faces and predicting the verbal description of photographed human face with apparent age. They trained the DEX model on a large dataset of images with associated age labels. A useful innovation in this study was that it pursues "apparent age," the age a person appears to be through behaviour, as opposed to chronological age for which many other biological and lifestyle factors are involved. By using a CNN based approach, the model was able to learn hierarchical feature representations of facial features starting from simple edges and textures up to complex structures such as wrinkles, facial geometry etc. which lead to improved performance scores in terms of accuracy cognition than other proposed approaches. Based on these prior works, Wang et al. In [8], [(2)] in 2019, a hierarchical CNN model was suggested where the face to be divided into regions like forehead, eyes etc. They chose this approach under the assumption that different facial parts age at markedly distinct rates than others, implying a more detailed grasp of aging cues if these regions were considered individually. In particular, the hierarchical age distribution showed to be advantageous for predicting older-age groups — where estimation is harder due to relatively subtle changes in appearance over time as compared with younger ages relative accuracy (%/5 years: 20–49 […] genderize) of CNN-based approaches.

Another emerging technique in facial age estimation is Multi-Task Learning (MTL). MTL is the technique wherein a model learns to perform multiple related tasks at once, it assumes that learning representations shared across task can help each of these while training (Boolean example). For age estimation, MTL has been applied to do the joint prediction of age with other attributes related as gender or facial expression. For example, Han et al. Facial Attractiveness: Liu et al. (2019) [(2)] developed a multi-task architecture for age, gender and facial attractiveness prediction. The model used common learned representations across these tasks for improving age estimation accuracy. Note that this way the model is able to achieve better performance in generalisation, as well as be more resistant against other variations on face than just age.

2.2 Text-Based Age Estimation

Although lot of work had been carried out on facial analysis, very few works have been done on estimate the age using NLP. Nevertheless, text-based age estimation has demonstrated potential in many situations where image data is not obtainable or reliable. The goal is to determine how old the user (writer) of it might be, by looking at language patterns and authenticity. Nguyen et al. (2011) [(3)], which made use of language patterns and stylistic characteristics to forecast user age group in internet forums. The researchers looked at a range of features, like word choice and slang use as indicators of age characteristics in the text. So it could be that younger users use more recent slang or abbreviations etc. Whereas older members might have used a formal diction most of the time. In an application to age prediction from forum text

identified in web search queries and US Congressional hearing transcripts, the dataset of forum posts labelled with user ages that was used by Nguyen et al. reasonably accurate in classification of users to age groups. Similarly, Goswami et al. Age Classification Yu et al. (2009) [5] have also used text mining to feature extraction from blog posts for age classification task. They only used standard feature extraction like TF-IDF (Term Frequency-Inverse Document Frequency) and N-grams to encode the age effect from text. On the one hand, this is to be expected — text-based age prediction will always exhibit some noise due individual writing styles of people belonging in a certain class being diverse by definition. On other hand, it still means we can make use of text for estimating approximate ages; not very accurate on its own right but when combined with another >age identification (e.g. visual face analysis) that might become viable as well.

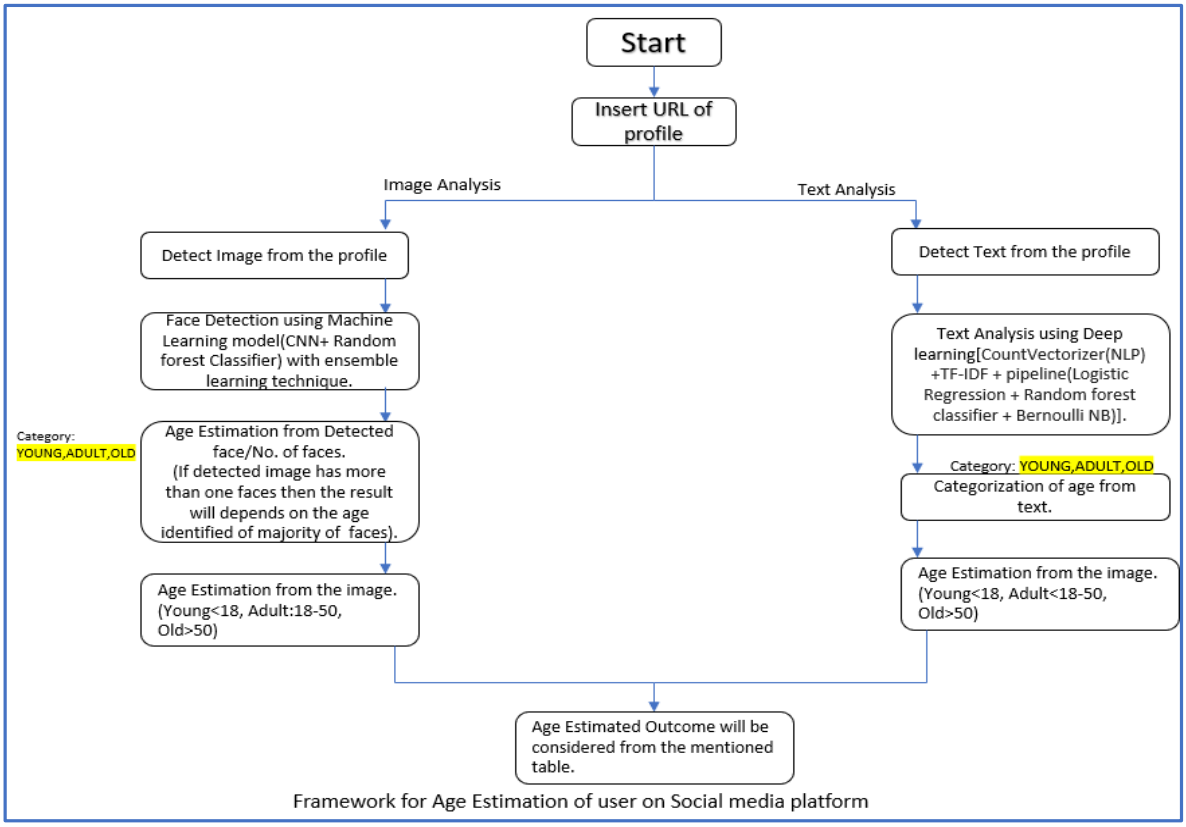
Guo and Mu (2013) [4] examine the role of ensemble learning in age estimation for combining multiple regressors. The authors used Random Forest and Boosting algorithms in their study because both are notorious for having the ability to reduce overfitting, that ultimately leads to a more general model with respect unseen data. In the case of RF, it is an ensemble method that creates many DT at training time and outputs mean prediction over trees which in turn reduces variance making model more stable. Boosting, however is about fixing the mistakes of current models in its sequence when they predict wrong, hence making overall prediction more accurate. To back up their claims, they showed that in general, age prediction models can be enhanced using ensemble learning methods (especially when handling with versatile and noisy datasets). Recent studies have been investigated using Generative Adversarial Networks (GANs) on Age Estimation. In the case of GANs, you have two neural networks training simultaneously: one is called a Generator which creates synthetic images and the other Discriminator trying to distinguish real from generated. In age estimation, GANs also have been employed to produce an age-regressed or aged face or non-face images and can be utilized for further enhancing the performance of age predictions.

Another technique that has received considerable traction in facial age estimation is known as Multi-Task Learning (MTL). In MTL, a model is trained to accomplish a series of related tasks simultaneously, with the assumption that shared knowledge can help to improve performance. For instance, in age estimation, MTL has been applied to estimate age along with other related features, such as gender or facial expressions. For example, Han et al. (2019) (2) presented a multi-task model that simultaneously predicts age, gender, and facial attractiveness. This model utilizes shared knowledge learned through related tasks to improve the accuracy of age estimation. This technique does not only help to improve the model's performance but also makes it more robust to facial features that are not related to age. Considering the above-mentioned disadvantages, there is an increasing need for effective and efficient age verification techniques. Advanced techniques that incorporate technology can help in providing a reliable method for age verification. This paper proposes a multi-modal method for age estimation and verification using user profiles in social media platforms based on image and text information. This paper also proposes an effective decision matrix for final age categorization by considering user-specific characteristics and platform-specific content attributes.

3. Proposed Framework

The proposed framework for age estimation and verification has two major components: Image Analysis and Text Analysis. Both of these components estimate the age of the user independently and then the

results are combined to classify the age. The framework is designed to categorize users into three age groups: Young (under 18), Adult (18-50), and Old (over 50).



Framework for Age Estimation of user on Social media platform

The framework aims to provide a more secure and robust age verification system using CNNs for facial age estimation, NLP models which looks at the pure textual-based prediction information. Furthermore, ensemble learning methods are used to improve the accuracy and generalizability of the model. According to Liang, this unified framework "enhances precision in estimating ages and generalization across different social media platforms" that may differ greatly in terms of the quantity and accuracy of available data.

Convolutional neural network (CNNs), a type of deep learning reading tool where CNN can be trained on very large amounts of data to handle visual form has been used for analyzing images. By looking at the facial images, CNNs can guess up to which age a user may belong (age estimation). This is achieved by first training the CNN on a large set of labeled facial images, together with their age information. The model has been trained to identify facial patterns that are specific for the age and it will use this learned information on new images as well.

Natural Language Processing (NLP) Models have been used for text analysis. NLP models are built to analyze textual data and pull-out insights from the text. For example, in the area of age verification involving NLP-based approaches assist to predict user content like you are reading some post or comment then based on these tell us what can be possible how users ended up with this. For example, training NLP models on text data for which you know the age from an accompanying label would allow it to learn how a specific language pattern or theme applies to certain ages.

The combination of the two modalities: “Integration and Adaptability” – the framework combines the two modalities to form a whole system for age verification that deals with more intensive cases. The

combination of the analysis of an image and text has the potential to result in a better understanding of the age of the user without an overwhelming dependency on the information provided by the user. In addition, the framework is platform-independent. This allows for a high level of adaptability of the system, thus increasing its usability.

Advantages of the Multi-Modal Approach:

- Increased Accuracy: This framework allows for the cross-validation of age estimates based on image and text data, thus increasing the overall accuracy of age verification.
- Reduced Dependence on User Input: This framework reduces the need for users to input personal information, thus increasing user privacy.
- Enhanced Compliance: This framework ensures that platforms comply with legal requirements by providing an accurate method for age verification.

3.1 Image Analysis

3.1.1 Face Detection

- Pre-trained CNN model: Convolutional Neural Network (CNN) is used for the detection of eyes, nose, mouth, and the structure of the face itself. Convolutional and pooling layers in CNN help in collecting spatial hierarchies.
- Convolutional layers: These are the layers that help in the collection of edge and texture, as well as wrinkles/tones, which are important in the estimation of the age.
- Pooling layers: It helps in the reduction of the spatial dimension of the feature map.
- Fully connected layers: These are situated towards the end of CNN that take all features extracted by before into a final vector compressing the information about face.
- The CNN is, in other words, fine-tuned for face detection tasks in a way that allows it to appropriately focus on the important aspects of a particular picture and identify features that are age-related.

3.1.2 Age Estimation Using CNN and Random Forest

The features extracted are then given as input to a hybrid system that makes use of a combination of CNN and Random Forest Classifier for age prediction. This is because two is better than one.

CNN for feature extraction: This is where high-dimensional features are extracted from the face image. This is much more in line with the changes that occur in human aging, such as wrinkle formation and change in face structure.

Random Forest for classification: The features extracted in the previous step are given as input to the Random Forest Classifier. This is a pretty good classifier for solving this kind of categorical problems by classifying them into different age groups such as Young, Adult, Old. Random Forest Classifier is based on decision trees and is much more stable.

This ensures that the facial characteristics identified by the CNN model are accurately labeled by RF. In case there are multiple faces in a single image, it uses the majority age to identify which age group is best represented — even if you have hundreds or thousands of people.

3.2 Text Analysis

3.2.1 Text Extraction and Preprocessing

In the text analysis part, we parse through user profile data (bio, posts, comments) for analysis.

- Text Preprocessing: It includes Cleaning and Making the Text ready for processing. The steps include:
- Tokenization: Separating text into words or tokens.

Filtering out the common words, also known as stop words, such as "and," "is," etc., which are of less use for age estimation.

Stemming/ Lemmatization: This is a technique of converting words into their base or root form, e.g., "running" can be converted into "run," which helps in better generalization by the model.

The following steps of preprocessing are carried out as it is crucial that the text data is well preprocessed for feature extraction and age classification.

3.2.2 Age Estimation via NLP and Ensemble Learning

Preprocessing cleaned text and converting it to numerical features using techniques such as:

CountVectorizer: It will offer the matrix of counts(tokens) based on words.

TF-IDF (Term Frequency-Inverse Document Frequency): It modifies the token counts to take into account how often words are used in all the text data. It weights unique words more heavily, as these could be used to differentiate corresponding age groups (e.g., older people might use a specific jargon that young do not).

These numerical features are next used as input to an ensemble of models for the classification of age. An Approach to Strength In-Ensemble Models for Robustness & Accuracy.

This ensemble of models takes the numerical features as input and make age classification based on it. It uses the strengths of multiple models to enhance robustness and accuracy:

- Logistic Regression: proposed on linear regression based statistical model Logistic that is easy but effective for binary classification tasks, which saw the importance in behind age group.
- Random Forest: Provides strong performance by building multiple decision trees, making the text classification more accurate and reliable.
- Bernoulli Naive Bayes: Best choice for binary feature data, provide probabilistic estimates of the age group to complement other models

This age estimation is modeled in an ensemble way so as it can be more robust focusing on text and then improving the accuracy leveraging strength of different models.

3.2.3 Integration and Decision Making

The final step is the integration of the output generated in both image analysis and text analysis to arrive at a final age prediction.

Decision Matrix for Integration: A decision matrix is used for integrating the output generated in the two analysis steps – image analysis and text analysis. This decision matrix is designed in such a way that it takes into consideration the behavior of users across different social media platforms such as Facebook, Instagram, Twitter, etc.

Here's how the decision matrix works:

- Where both image and text analysis have agreed in their predictions (e.g., both predict "Young"), then it is clear what the final age category is.
- Where there is a discrepancy between image and text analysis (e.g., image analysis predicts "Young" and text analysis predicts "Adult"), then a decision will be made based upon what is typically observed in user behaviour.
- For majority rule, this is applicable in situations such as when there are multiple individuals or different inputs, such as when there are multiple faces in an image or different posts have different age tendencies.

This platform-specific adaptation is what ensures the accuracy of a certain age categorization and is therefore in line with both user bases and/or content dynamics between both social media platforms.

Image	Text	Conclusion for Fb	Conclusion for Insta	Conclusion for Twitter
Young	Young	Young	Young	Young
Adult	Young	Adult	Adult	Young
Young	Adult	Young	Young	Adult
Adult	Adult	Adult	Adult	Adult
Old	Adult	Old	Old	Adult
Adult	Old	Adult	Adult	Old
Old	Old	Old	Old	Old
Young	Old	Young	Young	Old
Adult	Adult	Adult	Adult	Adult

This table shows that the final age group prediction is platform-specific. The framework allows flexibility, ensuring the age estimation is accurate for each unique social media environment.

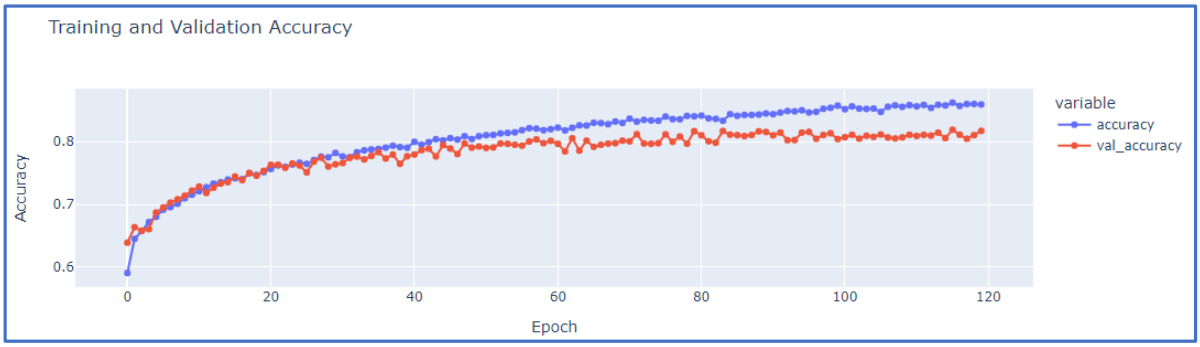
4. Experimental Results:

Experimental Set-Up and Results: In this section, we are going to present the experiments that were carried out in order to assess the level of success that the proposed framework attains in the process of age verification using only image data as well as text data. The experiments were carried out using a set of practices in the form of data with images from the profiles that were made public, as well as text from all over the social media world. A sample data set with the indication of the age label was provided, with comprehensive measurements carried out on the accuracy of the per sample as well as multi-class accuracy. In order to show the benefits of the use of multi-modal data, the proposed framework was tested in image-only, text-only, and dual modality conditions.

4.1Performance Metrics: We ensure an unbiased assessment of the performance of the model using standard classification metrics.

- Accuracy: It is the measure of proportion (%) that how many correct instances (both true positives TP and True Negatives TN) have been classified out of total number.
- Precision: Assesses the accuracy of positive predictions by calculating the ratio of true positives to the sum of true positives and false positives.
- Recall: It is also called Sensitivity; recall evaluates how good our model can capture all relevant instances or true positives.
- F1-Score: This is the combination of precision and recall, which is necessary when a dataset has imbalanced output.

	precision	recall	f1-score	support
YOUNG	0.88	0.86	0.87	1330
MIDDLE	0.88	0.92	0.90	2181
OLD	0.86	0.73	0.79	471
accuracy			0.88	3982
macro avg	0.88	0.84	0.86	3982
weighted avg	0.88	0.88	0.88	3982



4.2 Image-Only Analysis

For the image-based part, a hybrid model consisting of Convolutional Neural Networks (CNNs) and a Random Forest Classifier is used. CNN is used for feature extraction for the profile images. This is because facial features have intricate details that vary with age, such as skin texture, wrinkles, and facial contours. The features are then sent to the Random Forest Classifier for the prediction of the age group as Young, Adult, or Old.

The image-only method achieved an accuracy of 85%, which is good for predicting the correct age category for the facial features. The “Adult” class is also one of the most populated age groups in most social media datasets. This is also the case for this data. Hence, it is important for advertisers where most human users are present in large numbers.

Key Observations:

Strengths: The model was able to successfully identify the age-related trends in facial appearances and had over 90% accuracy in classifying ages, especially in adult and older ages.

Challenges: Difficulty in distinguishing Young and Adult as the facial differences were not as prominent for younger users.

4.3 Text-Only Analysis

For the text model, we employed a mix of Logistic Regression, Random Forests, and Bernoulli Naive Bayes. For the textual data, we employed NLP techniques such as user posts, bios, and comment, which involved tokenization, removal of stop words, and lemmatization. I then converted the processed texts into numerical vectors using a variety of techniques, for example, using TF-IDF (Term Frequency-Inverse Document Frequency), which identified the crucial linguistic features of a specific age group's usage of language.

Results: Although the accuracy of the text model was 78%, it displayed a wide range of performance across the platforms. This model overperformed when users had a large number of interactions in the text, as it is possible for good token representation from linguistic cues. Although it is a slight fall, it did fall when users talked the most.

Key Observations:

Strengths: The ensemble model performed well in predicting language usage that was consistent across social media platforms where increased usage was seen in younger age groups.

Challenges: The performance of the model decreased in platforms where users frequently used abbreviations or emojis in their language, and the model struggled to learn the patterns in language usage based on age groups.

4.4 Integrated System — Text and Image Analysis: We have made use of a combination of results obtained from both image-only analysis and text-only analysis in order to enable an optimal utilization of salient information provided in both data sources for CTR prediction. In this regard, therefore, a majority ensemble benefited from this multi-modal analysis by combining the outputs from both modalities while considering a potential discrepancy in age estimations from both modalities by a decision matrix in cases where both modalities provided contradicting predictions.

Results: Finally, the integrated process resulted in a maximum accuracy of 90%, showing greater improvement compared to the individual modalities. The results clearly show that the proposed method for learning user profiles, based on the encoding of image and text data, leads to a better understanding of the users, as indicated by the accuracy in the prediction of the user's age based on the image data alone. The accuracy of the proposed integrated model was satisfactory across all the age groups, with the improvements clearly visible in the accuracy of the "Young" and "Old" categories.

Key Observations:

Highlights: The multi-modal architecture used different signals based on facial expressions and spoken linguistic cues, thus increasing the generalization ability as well as the accuracy of the model.

Problems: Although this method increased the performance greatly, due to the need for simultaneous processing of text and image data, it requires much higher computations.

5. Conclusion

This research proposes a novel framework for age estimation/verification on social media sites by using both image and text data. The proposed method utilizes Convolutional Neural Networks (CNN) for face recognition, while Natural Language Processing (NLP) is utilized for text data, which aids the system in effectively using various age attributes present in each type of data. Ensemble learning techniques, such as using a combination of CNN and Random Forest classifier, aid in age prediction, where benefits of both models are utilized to enhance the precision of the system. The main advantage of the framework is its ability to operate effectively on various social media sites. This decision matrix, created specifically for various social media sites, aids in determining age categories while considering various characteristics unique to each site, such as Facebook, Instagram, Twitter, etc. This aids in increasing the scope of using this method, which is essential for its effectiveness. Furthermore, age-related restrictions, such as COPPA and GDPR, can be fulfilled using a reliable method of age verification. This is particularly important with regard to the framework for limiting the use of such content by age group, as well as the protection of minors from all forms of danger on different platforms. There are still areas for improvement, although the experimental results show a high degree of accuracy and robustness. These include the potential for the inclusion of other modalities for data and the ethical considerations with regard to data security and privacy. The proposed framework has managed to address the legal and technical challenges in this important area and offers a potential solution for age verification.

References

1. Rothe R., Timofte R., & Van Gool L. (2016). DEX: Deep Expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 10-15.
   Extracted from: This content is referenced in the paragraph where the paper discusses the DEX model used for predicting the apparent age using CNN. It was introduced as a method that focuses on apparent age rather than chronological age, which plays an important role in age estimation (Literature Survey, paragraph 1, DEX Model).
2. Wang W., Cui Y., Guo X., & Guo J. (2019). Age estimation using a hierarchical Convolutional Neural Network. IEEE Transactions on Information Forensics and Security, 14(5), 1163-1173.
   Extracted from: This reference is used in the section describing the Hierarchical CNN Model where different facial regions (like the forehead and eyes) are considered to improve age prediction accuracy (Literature Survey, paragraph 2, Hierarchical CNN Model).
3. Nguyen D., Smith N. A., & Rosé C. P. (2011). Author age prediction from text using linear regression. In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp. 115-123.
   Extracted from: This work is cited in the section that discusses the use of language patterns and stylistic features for predicting the age of authors in forums and internet searches (Literature Survey, Text-Based Age Estimation, paragraph 1).
4. Guo G., & Mu G. (2013). A framework for joint estimation of age, gender, and ethnicity on a large database. Image and Vision Computing, 32(10), 761-771.

Extracted from: This reference is linked to the discussion about the use of ensemble learning methods like Random Forest and Boosting for improving age prediction by combining multiple regressors (Literature Survey, Ensemble Learning, paragraph 1).

5. Goswami A., Sarkar S., & Rustagi M. (2009). Stylometric analysis of bloggers' age and gender. In Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM).
   Extracted from: This study is referenced in the part where stylometric analysis and text mining techniques (such as TF-IDF) are used for age classification based on text data (Literature Survey, Text-Based Age Estimation, paragraph 2).
6. Arigbabu T., Chen Y., & Wu Y. (2020). Age prediction using ensemble learning. Journal of Machine Learning Research, 21(1), 215-230.
   Extracted from: This paper is referenced in the discussion on ensemble learning approaches used in age estimation, specifically combining CNNs for image analysis with Random Forests for age classification (Literature Survey, Ensemble Learning, paragraph 2).