On Age Prediction from Facial Images in Presence of Facial Expressions

Abstract: Predicting age and gender from facial images is a fundamental research problem that has many applications in major research areas. The state-of-the-art online APIs can predict age and gender but their accuracy degrades when emotions are present. In this paper, we present feature-extraction based machine learning models that can predict ages with acceptable accuracy in presence of facial-expressions. After identifying 68 facial landmarks, different distances and ratios (that changes with age and expressions) are selected to predict the age that can overcome the impact of emotions with reasonable accuracy. The experimental results show that while neutralizing the effect of emotion, the proposed models can perform better on female images compared to the male image set. And images with disgust and contempt expressions deviate most during prediction. In contrast, predicted age is more accurate for angry expressions. Also for different ethnic groups, the predicted age deviates differently from the actual age.

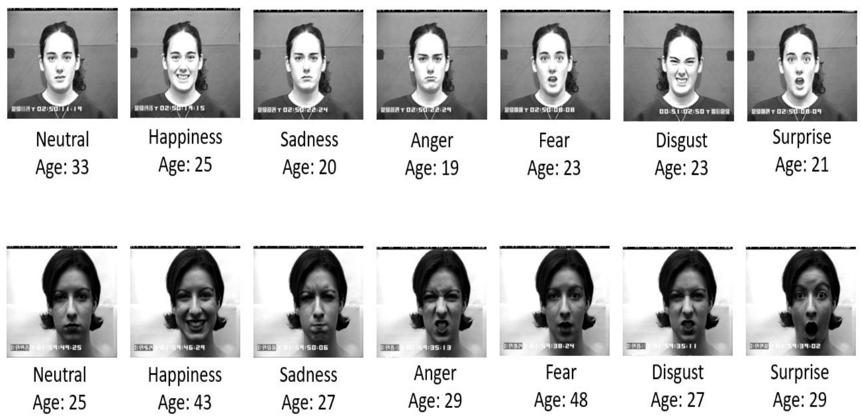
Keywords: Age estimation, Facial expression, API, Feature, Machine Learning, SVR, Facial Landmark, Feature Engineering, CK+

1 Introduction

Age estimation is the process of predicting a person's age with reasonable accuracy. Age prediction has potential for research avenues to support many real-life applications such as age-specific human-computer interaction, forensic art, access control, and surveillance monitoring, person identification, data mining and organization, and cosmetology, (Izadpanahi and Toygar (2014)), to name a few. For example, in many countries, minors are not allowed to enter nightclubs or purchase tobacco or alcohol. In those countries, an age estimator can facilitate the process of determining whether a person is really a minor or not. Organizations may be interested to know how old an employee is before claiming retirement benefits. Legal courts might want to estimate how old a crime victim is or want to determine whether an offender is really a minor or not. During the Beijing Olympics in 2008, several acquisitions were made against the members of many country's gymnastics teams of being below the threshold age limit of 16. Needless to say that in gymnastics, it is advantageous of being small and light (Age doping (https://www.vox.com/2014/10/20/6939271/age-test-aging-epigenetics-clockbiological- chronological-scandal-telomeres)). Therefore, it is desirable to predict age accurately.

Earlier studies (Age doping (https://www.vox.com/2014/10/20/6939271/age-test-aging-epigenetics-clockbiological- chronological-scandal-telomeres)) showed that despite many efforts, there is no scientific test to determine someone's exact age. Usually, the general appearance of a person can provide a rough estimation of age. With aging, the outer skin layer (epidermis) thins, and the skin's strength and elasticity get reduced. The graying of hair starts approximately at the age of 40 years. Thus, one can devise an age estimation algorithm based on a person's biometric features. Among all biometric traits, a person's facial image is readily available; hence in this work, we focus on age estimation that relies on biometric features that are extracted from a person's face. However, the so-called morphological appearance, i.e., the size, shape, and structure of a person can be misleading as they are highly variable and depends on many factors such as ethnicity, origin etc. The expressions or emotions present in the facial images make the estimation even more challenging.

Figure 1: Variation in age estimation under several facial expressions: age estimator fails to predict age accurately due to emotions present in faces.



To illustrate further, we use one of the online APIs available today to estimate age from facial images. More specifically, we run FACE++ API (https://www.faceplusplus.com/emotion-recognition/) on the images shown in Fig. 1 to see the age prediction accuracy under different facial expressions. As shown in the figure, the API predicts the age of the first person to be 33 on her neutral image but when she is happy, the API detects her age to be 25. As more emotions such as sadness, anger, fear, disgust, and surprise are explored, more deviations are observed. Analysis of the second person's image also yields a similar deviation. Her actual age is 25 (on her neutral image) but when she is afraid, the API detects her age to be 48 which is almost double the actual age.

There exist several challenges in estimating ages from facial images. There could be variations in environmental lights while capturing a facial image that might contribute to serious estimation errors. Many age prediction algorithms depend on distances between several elements of the face such as eye-to-eye or eye-to-mouth distances, and length of mouth, nose, and ears, etc. Faces that hold expressions or emotions might cause those distances and ratios to be changed rapidly. Males have beards and females often pose with their hair placed in front of their faces hiding the key features needed for accurate age estimation. Sometimes more than one emotion might be present in the face. Different ethnic groups' facial structures are not same also which contribute to different level of accuracy from one group to another. Some facial structures are so different that sometimes even expression detection becomes a daunting task. For example, a recent photo of North Korean president Kim Jong-Un with his subordinates is shown in Fig. 2. Although his expression

Figure 2: Confusing facial expressions: are they laughing or crying?

is clearly perceivable as a *happy* expression, the expression of his two other subordinates has been detected as *sad* despite the fact that they were happy too.

A number of research works have been conducted to predict a person's age from facial images. Unlike other works, we focus on estimating ages from those images which contain *facial expressions*. In particular, we consider seven emotions namely *happiness, sadness, anger, fear, disgust, contempt*, and *surprise*. To achieve our goal, we deploy an interesting blend of machine learning and image processing techniques. We also identify a number of suitable distances and ratios that can be used to create feature sets required for any machine learning algorithm for accurate age prediction.

The major contributions of the paper are as follows:

- To the best of our knowledge, no other previous work has been reported on modeling an age predictor by explicitly considering emotions/expressions present in the facial images.
- We conjecture that in presence of negative emotions (such as sadness, anger, fear, or disgust) the age predictors estimate higher age values than the actual while under the positive emotions (such as happiness, contempt, etc.) the age predictors estimate a lower age value than the actual.
- We introduce two feature sets (in section 3.2) that are derived from the previous research works aiming at the same research direction. Using those different feature sets, we trained our model and compared the performance of the proposed model with other APIs.
- We provide age estimation performance comparison between Face++ API and our model on the CK+ dataset (https://www.ri.cmu.edu/project/cohn-kanade-au-coded-facial-expression-database/). Our approach is better because we have reduced the deviation of predicted age from the actual age with emotions present in the face.
- Through experiments, we clearly show how age prediction accuracy varies rapidly based on ethnicity and gender.
- Emotion is eliminated from our feature set as it is implicitly inherited in those facial landmarks and ratios. So we need not determine which emotion/s is/are present in the image which reduces the complexity of computation.

The rest of the paper is organized as follows. Section 2 focuses on the challenges faced in the estimation of age from facial images considering facial expressions. In Section 3 we describe the proposed method of our work. Section 4 provides the findings and insights into the performance of our model. Section 5 contains an overview of the general methodology of most of the age estimation systems and previous works on age estimation systems. Finally Section 6 concludes the paper with pointers for future work.

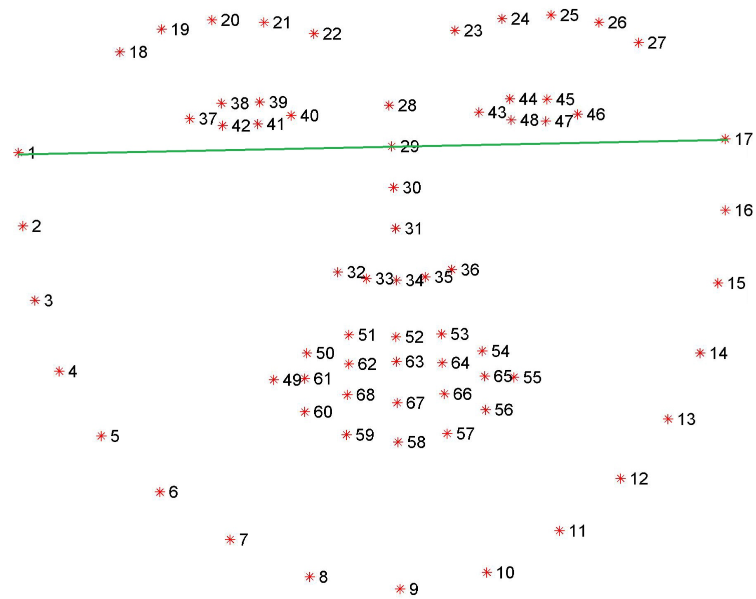
2 Challenges in Predicting age under different expressions

Estimating ages from facial images faces several challenges. Below we highlight some of the challenges.

- **Different ethnic groups:** Different ethnic groups may produce different accuracy rates in the predictor. Some ethnic groups might possess strong facial structures that do not change much with facial expressions. As a result, models built for prediction might show better results for that particular group compared to the other groups.
- **Presence of more than one emotion:** Sometimes more than one emotion might be present on a face. In such cases predicting age may become a daunting task.
- **Use of controlled laboratory environment:** Normally images in a data set to train machine learning models are taken under a controlled laboratory environment. But in practical scenarios, the situation might be different. In the laboratory, proper lighting is present and subjects face directly towards the camera, but in the real world, the subjects might be looking in different directions. Moreover, avoiding the problem of different types of posture can be more challenging.

  Again in the real world, lighting is not even everywhere. It might happen that one part of the face is getting more light than the other part or some portion of the face is not getting enough light at all. Such light variations may cause problems in identifying features. Normally, research works conducted in idealistic settings often exclude the problems faced in real-world situations.
- **Zooming of an image:** Facial distances used for predicting age may also vary due to the zooming effect. If the image is *zoomed in*, the distance values will be larger compared to those distances taken under *zoom out* condition. Two kinds of *scaling* can be done to tackle the issue. For example, scaling can be imposed *globally* by dividing all the distances by a global single distance or *locally* by dividing each distance with respect to a locally selected distance.
- **Creating feature set:** Creating a feature set is another challenging task. There are some facial distances that change with ages and there are some distances that change with expressions. Therefore, combining those and determining the minimal number of features producing the best results would be a daunting task.
- **Gender:** Sometimes subjects have facial hairs which makes it harder to detect facial points as well as predicting age. For example, male subjects often have beards, and female subjects often pose with their hair in front of their faces. Gender can also be a feature as it has an influence on age prediction. So, Proper labeling of gender is needed to increase the accuracy level.

Figure 3: 68 landmarks of a face image. The green straight line is the 'diameter of iris'.



3 Working Method

In this section, we discuss the working principle of our approach. We start with some definitions and then we describe our feature set along with a feature reduction algorithm. Finally, we discuss the overall methodology for predicting ages.

3.1 Facial Landmarks

Traditionally, *feature extraction based methods* extract features from facial images in two popular ways. In the first approach, facial landmarks are identified first and then several distances and/or ratios between landmarks are used as features because intuitively those distances and ratios are likely to vary with ages. The other approach uses wrinkles present on the face as features. The main motivation for using wrinkles as features is the number of wrinkles appearing in the human face grows with ages. Our approach falls into the first category where we have used 68 Facial landmarks (https://68landmarks.com/) as shown in Fig. 3. Those landmarks provide the outer contour of a face. Using those points, we can get various important features within a facial image. For example, we can determine the length/width of a mouth, length of a nose or an eyebrow, and so on upon which there is an aging effect. One example of our features is the "Diameter of the iris" which is defined to be the distance between landmarks numbered 1 and 17. It is marked by a green straight line in Fig.3.

3.2 Feature Set

The appearance of the face typically changes with age, so is with different expressions. With aging, people lose muscle tone and skin gets thinner which gives the face a loose, and fleshy appearance. The facial skin also dries out with aging and the underlying layer of fat shrinks in such a way that the face no longer has a rounded shape. Noses may also lengthen slightly. The ears may lengthen in some people (due to cartilage growth). In summary, the aging effect increases almost every facial attribute such as mouth width, mouth length, eye width, nose to mouth distance, diameter of iris (see Fig. 3), etc.

Besides aging, each emotion impacts different areas of a human face as people convey numerous nonverbal information in their faces through different expressions. For example, eyebrows can show distinctive emotional signals such as eyebrows can be raised when a person is surprised, or lowered and knit together meaning anger, sadness, or fear. Eyes might blink quickly (meaning distress or discomfort), or staring intensely to show attention or anger. Similarly, an open mouth may show fear, if one side of the mouth is raised it may be a sign of hate or contempt, raised corners of a mouth might indicate happiness and if both corners of a mouth are drawn down then it might convey sadness.

Once all the facial landmarks are identified, we can define two mathematical functions to derive necessary distance information from those facial landmarks which are described as follows:

Euclidean distance between two landmarks: Distance between two landmarks $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ can be expressed by $L(P_1, P_2)$ where,

$$L(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

Midpoint of two landmarks: Midpoint of two landmarks $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is determined by $M(P_1, P_2)$ where,

$$M(P_1, P_2) = (\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}) \tag{2}$$

Consider Fig. 3 where all 68 facial landmarks are shown. While smiling, as the eyes of a person get contracted, so the distance between eyebrows and corners of the eyes such as distance between landmark 22 and 40 or $L(22, 40)$ decreases. The width of the eye $L(37, 40)$ also decreases when someone smiles. At the same time, the width of the mouth $L(49, 55)$ increases but the distance from center of the nose to the corners of the mouth such as $L(34, 50)$ decreases. If someone is sad, the mouth corners go downward. So, the mouth widens resulting in an increase in $L(49, 55)$ and $L(34, 50)$ (i.e., the latter one is the distance from the center of the nose to a mouth corner).

For angry expressions, distance related to eyes play a major role. The eyes are contracted so $L(22, 40)$ decreases but the mouth width might increase but the distance between nose and mouth remains the same.

Surprise and fear expressions have many similarities. Mouth width is decreased but mouth height $L(52, 58)$ is increased. But comparatively mouth height for surprise is bigger than the height for fear. Eyes are wide open which causes $L(22, 40)$ to increase. The distance between the center of the nose and chin ($L(9, 34)$) as well as the distance between two maximum vertical points of a face $L(9, 28)$ also increase. For anger and disgust, the length of nose $L(28, 34)$ and inner eye corners $L(40, 43)$ change remarkably.

Using 68 facial landmarks and the factors (distances or ratios) that change with ages and emotions, we create a feature set as shown in Fig. 4. In this feature set, we consider 20 facial distances of which 9 distances (details are shown in Table 1) are affected mostly due to change in ages and the remaining 11 distances (details in Table 2) mostly change with emotions. We also choose five facial ratios which are dependent on age. They are elaborated in Table 3.

Table 1 9 Distance Features Dependent on Age

Feature Name	Feature	Determined by
D1	Distance between the outer corners of eye	L(37, 46)
D2	Distance between the inner corners of eye	L(40, 43)
D3	Distance between the pupils of eye	L(M(37, 40), M(43, 46))
D4	Length of the eye	L(M(38, 42), M(39, 41))
D5	Width of the nose	L(32, 36)
D6	Width of the mouth	L(49, 55)
D7	Length of the nose	L(28, 34)
D8	Distance between the two maximum vertical points of face	L(28, 9)
D9	Distance between nose and chin	L(34, 9)

*All the distances are scaled by double of L(1, 17)

Table 2 11 Distance Features Dependent on Emotion

Feature Name	Feature	Determined by	Normalizing factor
D10	Length related to left eyebrow 1	L(40, 22)	L(40, 19)
D11	Length related to left eyebrow 2	L(40, 21)	L(40, 19)
D12	Length related to left eyebrow 3	L(40, 20)	L(40, 19)
D13	Length related to right eyebrow 1	L(43, 23)	L(43, 26)
D14	Length related to right eyebrow 2	L(43, 24)	L(43, 26)
D15	Length related to right eyebrow 3	L(43, 25)	L(43, 26)
D16	Length related to left lip 1	L(34, 50)	L(34, 49)
D17	Length related to left lip 2	L(34, 51)	L(34, 49)
D18	Length related to right lip 1	L(34, 53)	L(34, 55)
D19	Length related to right lip 2	L(34, 54)	L(34, 55)
D20	Mouth height	L(52, 58)	L(49, 55)

Figure 4: Features in Combined Feature Set (a) Age Related Features. (b) Emotion Related Features. (c) Age Related Ratios.

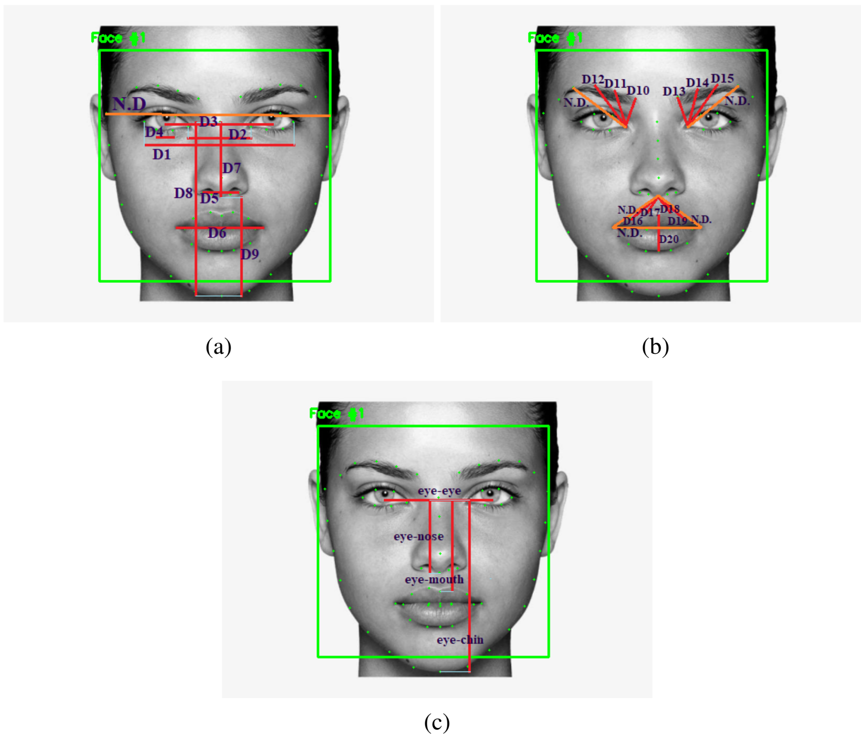


Table 3 5 Ratio Features Dependent on Age

Feature Name	Feature	Numerator	Denominator
R1	eye-eye / eye-nose	L(M(37, 40), M(43, 46))	L(M(M(37, 40), M(43, 46)), 34)
R2	eye-eye / eye-mouth	L(M(37, 40), M(43, 46))	L(M(M(37, 40), M(43, 46)), 52)
R3	eye-eye / eye-chin	L(M(37, 40), M(43, 46))	L(M(M(37, 40), M(43, 46)), 9)
R4	eye-nose / eye-mouth	L(M(M(37, 40), M(43, 46)), 34)	L(M(M(37, 40), M(43, 46)), 52)
R5	eye-mouth / eye-chin	L(M(M(37, 40), M(43, 46)), 52)	L(M(M(37, 40), M(43, 46)), 9)

3.3 Chosen Model

In this proposed work, we estimate the age of a person in years which is a discrete integer quantity. In some other works, the possible age range is divided into few groups, each group containing a sub-range of ages. For those cases, the problem becomes a classification problem where the model predicts the group where the predicted age belongs to. But our target model needs to predict a discrete numerical value for age estimation, not a class of age ranges. Therefore, in our approach, the age estimation problem becomes a regression problem. We find Support Vector Regression (SVR) suitable for this work which is actually a Support Vector Machine (SVM) model used for regression problems. SVR is better than linear regression model because it uses a method called kernel trick to perform more complex predictions for non-linear classifications. Kernel is basically a function which maps a lower dimensional plane to a higher dimensional space. It also tells us that given two data points in the original feature space what the similarity is between the points in the newly transformed feature space. Linear, polynomial, radial basis function (RBF), and sigmoid are some commonly used kernels. In our work, we select (Radial Basis Function) RBF Kernel (http://www.saedsayad.com/support_vector_machine_reg.htm) for the proposed SVR model as the value of the RBF kernel decreases with distance and ranges between zero and one. And also the feature space of the kernel has an infinite number of dimensions. The used Gaussian Radial basis Function is as follows,

$$k(x_i, x_j) = exp(-\frac{|x_i - x_j|^2}{2\sigma^2})$$

After the kernel trick, SVR finds a hyper-plane in the transformed feature space to predict the continuous output.

3.4 Feature Set Reduction

The presence of too many features in the proposed feature vector drastically affects the execution time of our model's data processing. Therefore, we devise a **Recursive Feature Deletion** algorithm that carefully analyzes the feature vector and removes features affecting the prediction accuracy of the model. The algorithm iteratively removes features one-by-one from the initial feature set and thereby determines removal of which feature (from the initial feature set) provides us the best accuracy. This process continues as long as the removal of features generates a performance improvement.

The pseudocode of the Recursive Feature Elimination (RFE) algorithm is shown in Algorithm 1. At first, the algorithm starts with the entire feature set. The RMSE value of the feature vector containing all the features using a machine learning model is stored in $current\_RMSE$ variable. The value of min_RMSE is initially set to $infinity$. Then, the algorithm deletes each feature from the feature vector one by one and calculates the corresponding RMSE value. Upon deletion, if this new_RMSE is less than the min_RMSE, the min_RMSE is updated to this iteration's RMSE value and the deleted feature is recorded. Once the current iteration is complete, the deleted feature is added back again to the feature vector and the next feature from the vector is removed and the same procedure is followed. At the end of each iteration, $current\_RMSE$ is always updated to be of the least value and the corresponding feature is permanently deleted from the feature set. Iteration continues until min_RMSE improves at each iteration.

3.5 Putting All Pieces Together

After describing and justifying the feature set, and presenting the feature reduction algorithm, we are now in a position to describe the overall methodology.

Algorithm 1: Recursive Feature Elimination (RFE) Algorithm

```
Input: Initial Feature Set F={F_1, F_2, ..., F_N}
Output: Updated Feature Set F_updated = {F_1, F_2, ..., F_M} with lowest RMSE
  where M≤N
current_RMSE ← RMSE(F);
while current_RMSE decreases do
    min_RMSE ← inf;
    f_del ← Null;
    for each f in F={F_1, F_2, ..., F_N} do
        F ← F - {f};
        new_RMSE ← RMSE(F);
        if new_RMSE < min_RMSE then
            min_RMSE ← new_RMSE ;
            f_del ← f ;
        end
        F ← F ∪{f}
    end
    if min_RMSE < current_RMSE then
        F ← F-{f_del};
        current_RMSE ← min_RMSE;
    else
        return F_updated={F_1, F_2, ..., F_M};
    end
end
```

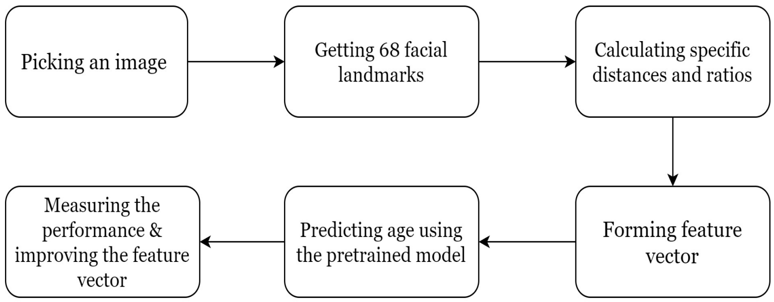
At first, we pick a facial image of a person where emotion is present. We use dlib and openCV library package to extract the 68 facial landmarks from this image. First, we convert the input image into a grayscale image. Then, we use a method from dlib package to detect the frontal face within that grayscale image. Then we use another function of dlib library to predict its shape. This function gives us our final desired 68 landmarks.

Using those 68 facial landmarks, we calculate specific distances and ratios that are needed for calculating values of the features within the feature vector. All distances are normalized so that each feature is of the same importance. Also, it reduces the issues related to posture present in an image and variation in lighting conditions in the image. Thus, for every input image, we get a feature vector for estimating the age. Due to different age ranges and the existence of emotion in the facial image, each feature vector is different from the other.

The feature vector from the input image is then fed into a machine learning model to predict the age of the person in the image. The model is chosen to predict the age as a discontinuous value i.e. we treat the problem as a regression problem. From the predicted age and the actual age, we calculate the estimation error of the model.

For training and testing purposes, we split the dataset into 80-20 split. Once the model is trained batch-wise with 80 percent images, we test each image either separately or batch-wise. To compute the overall estimation error of our model, we use the Root Mean Squared

Figure 5: Working methodology of the age prediction from facial image containing emotion



Error function, which is:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_i' - y_i)^2}{N}}$$

here y_i' is the predicted value of i-th occurrence and y_i is the actual value of the i-th occurrence. The flow diagram of the proposed methodology is shown in Fig. 5.

4 Experiment and Result Analysis

In this section, we provide the performance of the proposed method with the proposed feature set and the feature reduction algorithm. Two data sets were chosen for experimentation:–
(i) The Extended Cohn-Kanade Dataset CK+ dataset (https://www.ri.cmu.edu/project/cohn-kanade-au-coded-facial-expression-database/), and
(ii) The AGE, GENDER AND ETHNICITY (FACE DATA) dataset (https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv).
For a particular dataset, at first, we choose an image within the data set and then we calculate different facial features from 68 facial landmarks to form the feature vector. Next, we import SVR from the Machine Learning library scikit-learn library (https://scikit-learn.org/stable/) and feed the feature vector to the model. To measure the performance of the model, the whole dataset is split into two sets where the train set contains 80% of the original dataset and the remaining 20% form the test set. After that, we perform some feature reductions based on the performance of the model to find the most suitable feature vector.

4.1 Experimental Results on CK+ Dataset

Extended Cohn-Kanade Dataset CK+ dataset (https://www.ri.cmu.edu/project/cohn-kanade-au-coded-facial-expression-database/) consists of facial images with seven different expressions namely *happiness*, *sadness*, *anger*, *fear*, *disgust*, *contempt*, and *surprise*. This dataset provides a total of 10,709 images of 123 subjects consisting of 54 male subjects and 69 female subjects. For a person, there may be several images for a specific expression, but an image for each expression of every person is not available. In the data set, 5690 images belong to 69 female persons and 5019 images belong to 54 male persons. There are 1491 angry faces, 408 contempt faces, 1338 disgust faces, 1259 feared faces, 2252 happy

Table 4 Distribution of emotions with respect to gender in CK+ Data set

Emotion	Total Number of Images	Female		Male	
		Number of Images	Percentage	Number of Images	Percentage
Angry	1491	721	48.357%	770	51.643%
Contempt	408	194	47.549%	214	52.451%
Disgust	1338	715	53.438%	623	46.562%
Fear	1259	656	52.105%	603	47.895%
Happy	2252	1276	56.661%	976	43.339%
Sad	1891	1080	57.113%	811	42.887%
Surprised	2070	1048	50.628%	1022	49.372%

Table 5 Different combination of feature groups and corresponding age prediction performance

Comb. No	Age Related Distances (D1-D9)	Emotion Related Distances (D10-D20)	Age Related Ratios (R1-R5)	RMSE (without Gender) Without feature reduction	RMSE (with Gender) Without feature reduction	RMSE (with Gender) With feature reduction
1	✓			4.73	4.25	4.19
2		✓		5.22	5.21	5.21
3			✓	5.30	5.01	4.93
4	✓	✓		4.12	3.75	3.64
5		✓	✓	4.21	4.01	3.90
6	✓		✓	3.86	3.62	3.38
7	✓	✓	✓	3.63	3.41	3.22

faces, 1891 sad faces, and 2070 surprised faces in total. Expression wise image distribution between male and female is given in Table 4. Notably, the images within each expression category constitute a more or less similar percentage.

In CK+ dataset, images are already labeled with a particular expression and 68 facial landmarks of each image are also available, However, the age of the persons within the images is not available in the dataset. To work around the issue, we use FACE++ API (https://www.faceplusplus.com/emotion-recognition/) to predict the age, and then we use this retrieved age as the benchmark for comparing our model's performance.

We have introduced three types of features in our feature set in Table 1, Table 2, and Table 3. We conduct experimentation with various combinations of those features, with or without the feature - "Gender", and last but not the least with or without the feature reduction algorithm (i.e., the RFE presented in Algorithm 1). Results are presented in Table 5. It is evident from the result that, "Gender" plays an important role in determining age irrespective of emotion. And also feature reduction algorithm (RFE algorithm) is effective in improving results. From the result, it is evident that incorporating all of the proposed features in the feature vector along with the gender feature provides the best accuracy (i.e., lowest RMSE value) after applying the feature reduction algorithm (RFE).

Table 6 Emotion wise age prediction error analysis

Emotion	**Type**	**Numbers of images considered**	**RMSE before feature deletion**	**RMSE after feature deletion**
Angry	All	287	2.704	2.494
	Male	162	2.65	2.172
	Female	180	1.947	1.75
Sad	All	380	2.878	2.812
	Male	148	2.66	2.382
	Female	203	2.793	2.466
Fear	All	261	3.296	3.125
	Male	110	2.846	2.876
	Female	131	2.784	2.656
Surprised	All	445	3.39	3.145
	Male	216	2.882	2.883
	Female	184	3.308	3.3
Happy	All	460	3.677	3.549
	Male	198	2.768	2.403
	Female	251	2.249	2.215
Contempt	All	71	3.757	3.617
	Male	46	3.677	3.753
	Female	46	3.612	2.96
Disgust	All	238	4.192	3.974
	Male	123	3.909	3.485
	Female	143	3.449	3.464

Next, we focus on determining which emotion affects age the most. So, we divide our test cases emotion-wise and calculate the error. The result is shown in Table 6. Also, we separate the male and female category images and find the error before and after applying the feature reduction algorithm (RFE). The result shows that *angry* expression generates less error on the age prediction accuracy. On the other hand, the presence of *disgust* emotion causes most inaccuracy on predicted ages. Moreover, predicting ages for male faces with emotions present are more error-prone than that on female faces as female faces are more expressive than the male ones.

4.2 AGE, GENDER AND ETHNICITY (FACE DATA) Dataset and Experiment Results

AGE, GENDER AND ETHNICITY (FACE DATA) dataset (https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv) is a modified version of UTKFace dataset (https://susanqq.github.io/UTKFace/) with 23705 images from five different ethnic groups which are: White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern). In the dataset 11314 images belong to female subjects and 12391 images belong to male subjects. There are in total 10078 White faces, 4526 Black faces, 3434 Asian faces, 3975 Indian faces and 1692 Others (Hispanic, Latino, Middle Eastern) faces. Ethnicity-wise image distribution between male and female

Table 7 Distribution of ethnicity with respect to gender in Dataset

Ethnicity	Total Number of Images	Female		Male	
		Number of Images	Percentage	Number of Images	Percentage
White	10078	4601	45.654%	5477	54.346%
Black	4526	2208	48.785%	2318	51.215%
Asian	3434	1859	54.135%	1575	45.865%
Indian	3975	1714	43.119%	2261	56.881%
Others (Hispanic, Latino, Middle Eastern)	1692	932	55.083%	760	44.917%

Table 8 Ethnic group wise age prediction error analysis

Ethnicity	Type	Numbers of images considered	RMSE
Others (Hispanic, Latino, Middle Eastern)	All	325	10.255
	Male	196	10.336
	Female	118	9.551
Indian	All	733	11.908
	Male	307	10.026
	Female	429	13.08
Asian	All	612	12.236
	Male	339	11.687
	Female	283	14.618
African	All	784	13.071
	Male	378	12.093
	Female	400	13.574
White	All	1783	16.134
	Male	833	17.553
	Female	953	15.227

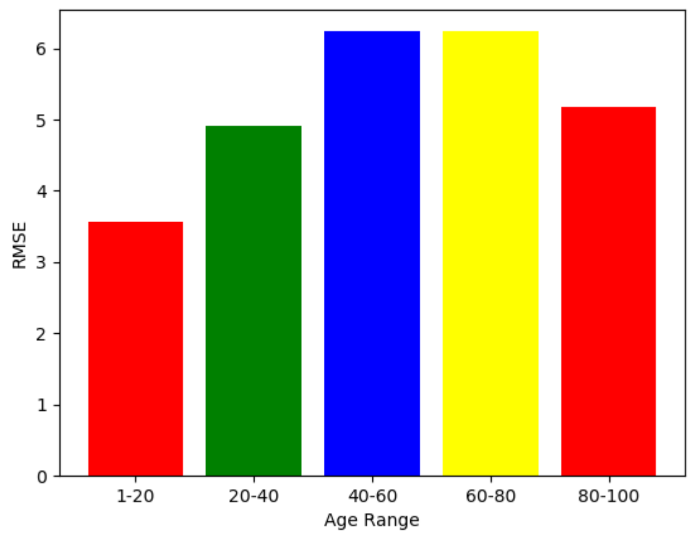
subjects is shown in Table 7 which also shows more or less even distribution of images among different ethnic groups.

The age and gender are already labeled on each image within the data set. However, we had to use OpenCV and dlib library packages to extract 68 facial landmarks for each image.

The main reason for exploring this ethnic dataset is to find out which ethnic group results in the most prediction error. We also conduct the same experiments with male and female images separately. The result of ethnicity-wise prediction error is shown in Table 8. From the results, we can conclude that the White ethnic group produces the most prediction error. Although not surprising, we observe that age of females subjects are harder to predict as their expressions make their face more distorted compared to male subjects.

We also explored which age group is more prone. For that purpose, we divide all images into five age groups, each group consisting of 20 years of age span. Thus, the first age

Figure 6: Errors on different age groups



group consists of faces from 1 year to 20 years, the second group 21-40, the third one to be 41-60, the fourth one 61-80, and the rest in 81-100 years of age range. The RMSE errors on ages of different age groups are shown in Fig 6. The age group 40-60 shows the highest prediction error as their facial structure impacts most on the features that are used for capturing expressions present on the face.

4.3 Exploring Feature Reduction Algorithm

In this section, we show the effect of applying the "Recursive Feature Elimination Algorithm" (presented in Algorithm 1 of Section 3.4) on the feature set in detail. Remember, we remove each feature present in the current feature set one at a time and look for the best accuracy in an iterative way. At first, we arbitrarily pick an image from the CK+ dataset. We can visualize the first two iterations (iteration 1 and 2) in Fig. 7 of the algorithm applied to the image. The lowest points of the curves of both iterations are marked as "Descend" in the curve. On the first iteration, the lowest prediction error of 3.363 years is achieved after removing the feature D17 which is the length related to left lip 2 (see Table 2). On the second iteration, the lowest prediction error of 3.322 years is found after removal of D9 which is the distance between nose and chin (See Table 1). The last two iterations (iteration 8 and 9) of the algorithm are shown in Fig. 8 where the removal of D4 (length of the eye) and the removal of R3 (the ratio between eye to eye and eye to chin) produces the best results of 3.231 and 3.221 years respectively.

Table 9 shows all iterations of the algorithm in detail by showing which feature gets eliminated in which iteration along with the RMSE value before and after the removal of that feature. As it can be seen from the table, the algorithm continues to remove features up to ninth iterations as it gets better accuracy, but at the 10th iteration, the error does not decrease and the algorithm terminates with the final feature set

Figure 7: First two iterations of recursive removal of features from the feature set

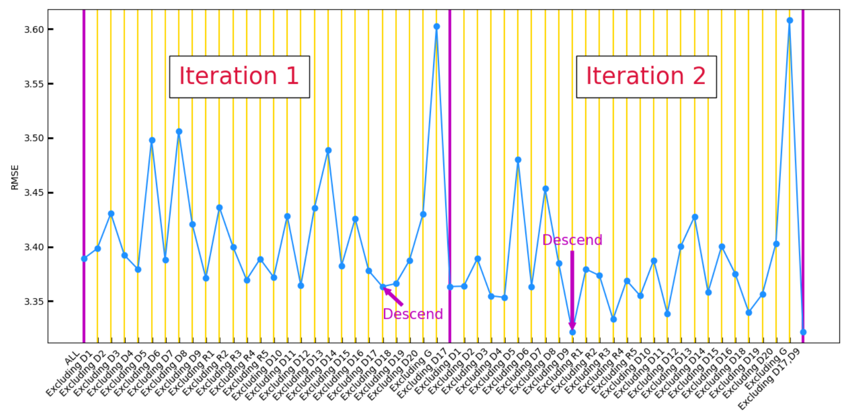
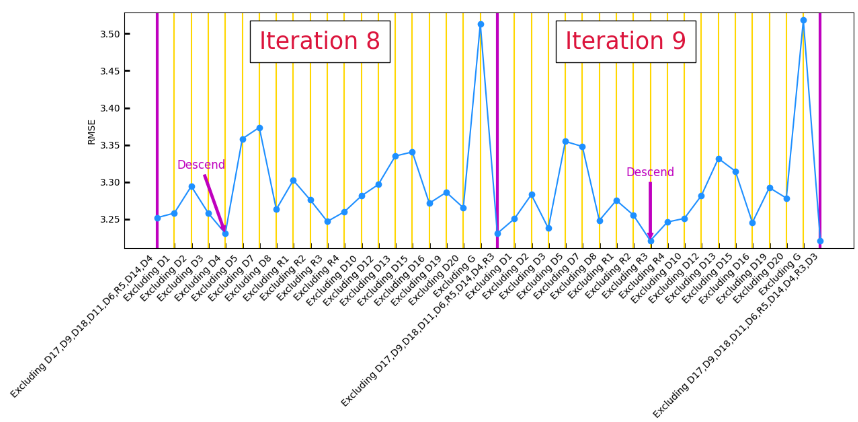


Figure 8: Recursively delete features from the feature set. RFE - iteration 8 & 9



$\{D1, D2, D3, D4, D5, D7, D8, R1, R2, R4, D10, D12, D13, D15, D16, D19, D20, G\}$.

Table 10 and 11 demonstrates the effect of the feature reduction algorithm on male and female subjects separately to visualize which gender is most benefited using the proposed feature reduction algorithm. We notice that female images require less feature reduction but they can generate better results compared to the male images as female faces are more expressive than male ones.

4.4 Comparison with Other Prediction Models

To compare the performance of the proposed model with other machine learning models found in the literature, we choose the work presented by Perak Agarwal - Age detection using facial images (https://towardsdatascience.com/age-detection-using-facial-images-traditional-machine-learning-vs-deep-learning-2437b2feeab2). The author combined Facial age dataset (https://www.kaggle.com/frabbisw/facial-age) and UTKFace dataset (https://susanqq.github.io/UTKFace/) for his experiments. Unlike our work, they

Table 9 Iterations of Feature Reduction Algorithm

Iteration No.	**Deleted Feature**	**RMSE (before removal)**	**RMSE (after removal)**
1	D17	3.389	3.363
2	D9	3.363	3.322
3	D18	3.322	3.312
4	D11	3.312	3.298
5	D6	3.298	3.289
6	R5	3.289	3.267
7	D14	3.267	3.252
8	D4	3.252	3.231
9	R3	3.231	3.221
10	D3	3.221	3.223
Reduced feature set	**D1, D2, D3, D4, D5, D7, D8, R1, R2, R4, D10, D12, D13, D15, D16, D19, D20, G**		

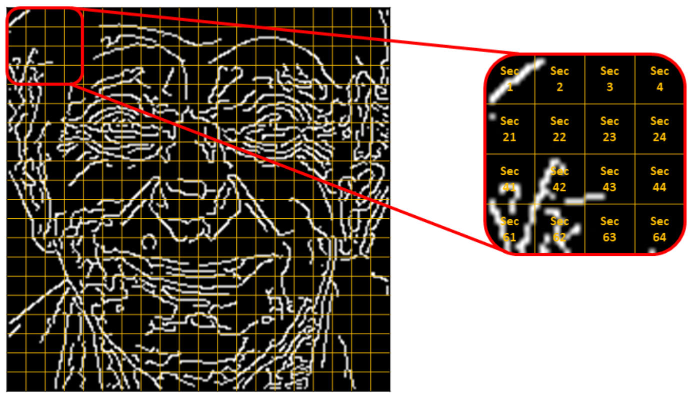
Table 10 Feature Reduction Order for Male Faces

Iteration No.	**Deleted Feature**	**RMSE before**	**RMSE after**
1	D17	2.977	2.937
2	D19	2.937	2.902
3	D6	2.902	2.855
4	D11	2.855	2.832
5	D16	2.832	2.826
6	R5	2.826	2.804
7	D1	2.804	2.784
8	R3	2.784	2.752
9	D9	2.752	2.753
Reduced feature set	**D2,D3,D4,D5,D7,D8,D9,R1,R2,R4, D10,D12,D13,D14,D15,D18,D20,G**		

Table 11 Feature Reduction Order for Female Faces

Iteration No.	**Deleted Feature**	**RMSE before**	**RMSE after**
1	D17	2.791	2.766
2	D11	2.766	2.739
3	R5	2.739	2.722
4	D14	2.722	2.688
5	R4	2.688	2.683
6	R2	2.683	2.662
7	D10	2.662	2.667
Reduced feature set	**D1,D2,D3,D4,D5,D7,D6,D8,D9,R1,R3,R5, D10,D12,D13,D15,D16,D18,D19,D20,G**		

Figure 9: Breakdown of image showing Canny Edges into multiple sections for feature extraction Age detection using facial images (https://towardsdatascience.com/age-detection-using-facial-images-traditional-machine-learning-vs-deep-learning-2437b2feeab2)



used wrinkles for predicting ages. It requires image processing prior to training the model such as detecting *canny edges* from the images. In their approach, at first, an image of 200x200 pixels is divided into 10x10 pixels which results in 400 sections. Mean and standard deviation of pixel values for each section are then used as features for the classifier. Finally, the author proposes the following two classifiers that classify different age values under ten categories:

- Random Forest Classifier
- Support Vector Classifier

We apply their technique for age prediction on AGE, GENDER AND ETHNICITY (FACE DATA) dataset (https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv) (See Section 4.2 for details on this dataset). We modify the target of their classifiers to discrete age values instead of only ten classes (i.e., the way they proposed) so that we can use their model as a regression model (to make it compatible with our model). After getting the predicted age values, we calculate the error rate (Root Mean Squared Error) and directly compare their results with ours. The dataset is split into 70/30 for training and testing and RMSE value is calculated for the test dataset. Table 12 shows the result. The result of our proposed approach is better compared to their random forest classifier (with canny edges) but is slightly worse compared to their support vector classifier (SVC). However, the major drawback of a canny edge based approach is the canny edges are not always clearly detectable in all images (especially on real-life images) and it always requires computationally expensive image processing techniques.

4.5 Findings

From our experiments and analysis, we identify the following important observations:

- Gender plays an important role in determining ages.

Table 12 Comparison between two approaches

Approaches	Regression/ Classification Model	Accuracy Converted to RMSE
Our Approach	SVR	13.978
Approach proposed in the article	RFC	16.705
Approach proposed in the article	SVC	12.122

- Female images need less feature reduction than males but they provide better results.
- In terms of error rates we can order the seven emotions (decreasing to increasing) as follows:
  Angry < Sad < Fear < Surprised < Happy < Contempt < Disgust

  For male subjects the order becomes:
  Angry < Sad < Happy < Fear < Surprised < Disgust < Contempt

  For female subject the ordering is as follows:
  Angry < Happy < Fear < Sad < Surprised < Disgust < Contempt
- Based on the error rates the ethnic groups can be ordered as follows (decreasing to increasing):
  Others (Hispanic, Latino, Middle Eastern) < Indian < Asian < African < White

  For male subjects, the ordering of ethnic groups becomes:
  Indian < Others (Hispanic, Latino, Middle Eastern) < Asian < African < White

  For female subjects, the ordering of ethnic groups is as follows:
  Others (Hispanic, Latino, Middle Eastern) < Indian < African < Asian < White

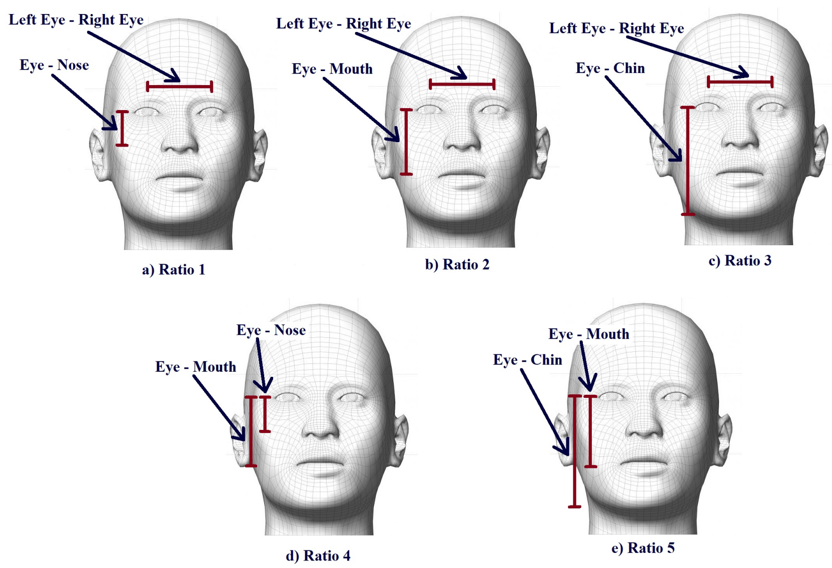
5 Related Works

Though no previous work has been reported on modeling an *emotion invariant* age predictor, estimating age from *neutral* facial images is not a new topic. Starting from the first work by Kwon and Lobo (1999), where they described theory to classify input images into babies, young adults, and senior adults, many works have been proposed regarding the age classification problem. Some of the important works in this research domain are described next. Note that, some of the proposed features in our work have been directly inherited from some of these works.

5.1 Age Estimation from Facial Landmarks

Kwon and Lobo (1999) developed the first algorithm to classify input images into different ages. The algorithm is able to classify input images into one of three age groups: babies, young adults and senior adults. They used some ratios found from primary features to distinguish an image into different age groups.They described eyes, nose, mouth, chin, virtual-top of the head, and the sides of the face as primary features. In secondary feature analysis, they used a wrinkle geometry map to distinguish senior adults from the other two groups. Using primary features, they proposed six ratios that change with ages.

Figure 10: Five ratios proposed by Kwon and Lobo (1999) to predict age



The major drawback of their work is finding the virtual top of the head from a face image is not practical. The other five ratios are shown in Fig. 10. They are able to classify baby images from adult images with an accuracy of 92% using ratio 3. For other ratios, accuracy varied from 38 to 78 percent. In their work, they showed that ratios from different facial features can be used for age classification which paved the path for further research. Also, their model cannot produce a numeric value for predicted ages, it can only classify into one of those three categories.

Dehshibi and Bastanfard (2010) propose an algorithm that classifies subjects into four different age categories. They identify eight landmarks and six facial measurements which they use to classify the subjects in the images.

They also perform wrinkle analysis in their work. They use ANN to classify the face into age groups using computed facial feature ratios and wrinkle densities. Their algorithm can classify age groups with an accuracy of 86.64%. Their work also fails to predict the age as a single numeric quantity, rather they simply classify an image into one of those four age groups.

Izadpanahi and Toygar (2014) also propose a different algorithm using different geometric ratios and wrinkle analysis. Using 22 distances shown in Fig. 12 they define sixteen ratios that have been used along with the wrinkle analysis to classify images. Their algorithm can classify images into one of seven age groups (three age groups for children and four age groups for adults). They evaluate their method using Support Vector Classifier (SVC). Their method classifies images with an accuracy of 90%. Their work is also limited to classifying into age groups. Predicting a single age value is not possible using their approach.

Machado et al. (2017) define 10 distances in their work. In their work, they showed average relative growth of these distances as a function of ages for five different age groups. They showed that this growth can range from -14.83% to 9.41% for different distances

Figure 11: Landmark points and ratios proposed by Dehshibi and Bastanfard (2010) to estimate age

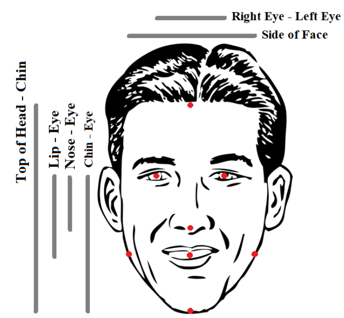
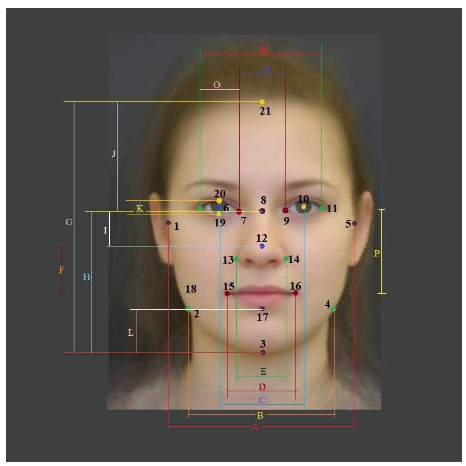


Figure 12: Landmark points and distances proposed by . Izadpanahi and Toygar (2014)



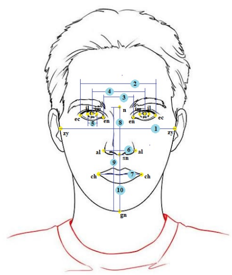
for different age groups. So, with the change of ages, average relative growths for some distances are positive and some are negative.

5.2 Expression Recognition from Facial Landmarks

Munasinghe (2018) first shows that expression recognition from facial images can be achieved using facial landmarks. In their work, he consider eight landmarks related to eyebrow movements and six landmarks related to mouth movement to detect expressions. He used normalized distances to create the feature set. Using Random Forest Classifier he built a model which could detect four basic expressions (Anger, Happiness, Sadness, Surprise) with accuracy ranging from 79% to 96%.

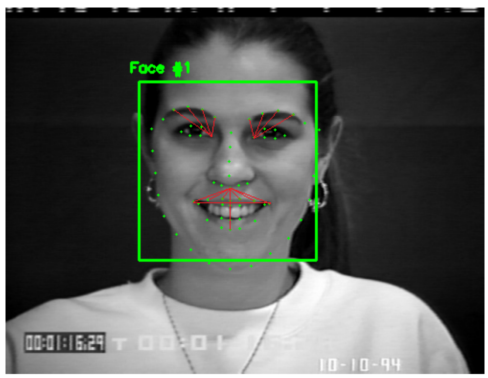
One thing that is obvious from the above literature review is that to predict age from facial images we need to calculate various facial ratios found from the facial landmarks. So the challenge is to find a proper set of distances and ratios which tend to change promisingly with the change of ages. In our work, we used different combinations of the above-mentioned

Figure 13: 10 distances proposed by Machado et al. (2017)



Number	Measurement
1	zy-zy
2	ec-ec
3	en-en
4	pu-pu
5	iris*
6	al-al
7	ch-ch
8	n-sn
9	n-gn
10	sn-gn

Figure 14: 14 distances proposed by Munasinghe (2018)



ratios and distances to find the best combination. And we worked to develop a model which can predict a single age value for an image and most importantly irrespective of expressions of that person.

6 Conclusion and Future work

Accurate age prediction is an important research problem that has many applications. Expression or emotion of a facial image affects age estimation accuracy to a great extent which has been overlooked by the research community although an inverse work i.e., "facial expression recognition influenced by human aging" Guo et al. (2013) has been proposed in the literature. To the best of our knowledge, this is, by far, the first system to eradicate the effect of emotion on age estimation. Still, there is plenty of room for improvement. Exploring with new methods or using various models to fit the dataset could be some possibilities. Also, someone can use bigger dataset to evaluate the performance of the proposed model rigorously.

Here are some pointers for future works. The age prediction can be made hierarchical. At first, we can divide the entire age range into a smaller number of subgroups. At the top level, we can use a decision tree to classify the image into one of the age groups. Then within an age group, the image can be fed into an SVR model to predict the age at a more granular level. Also, the problem can be transformed from a regression problem into a classification problem by generating 100 classes, one class for each age value ranging from 1 to 100.

References

Dehshibi, M. M. and Bastanfard, A. (2010) 'A new algorithm for age recognition from facial images', *Signal Processing*, Vol. 90, No. 8, pp.2431–2444. https://doi.org/10.1016/j.sigpro.2010.02.015

Guo, G., Guo, R. and Li, X. (2013) 'Influenced by Human Aging', *IEEE Transactions on Affective Computing*, Vol. 4, No. 3, pp.291–298.

Izadpanahi, S. and Toygar, Ö. (2014) 'Human age classification with optimal geometric ratios and wrinkle analysis', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 28, No. 2, pp.1–17. https://doi.org/10.1142/S0218001414560035

Kwon, Young H and da Vitoria Lobo, Niels (1999) 'Age classification from Facial Images', *Computer vision and image understanding*, Vol. 74, No. 1, pp.1–21.

Machado, C. E. P., Flores, M. R. P., Lima, L. N. C., Tinoco, R. L. R., Franco, A., Bezerra, A. C. B., Evison, M. P. and Guimarães, M. A. (2017) 'A new approach for the analysis of facial growth and age estimation: Iris ratio', *PLoS ONE*, Vol. 12, No. 7, pp.1–19. https://doi.org/10.1371/journal.pone.0180330

Munasinghe, M. I. N. P. (2018) 'Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier' in ICIS 2018 : Proceedings of the *17th IEEE/ACIS International Conference on Computer and Information Science*, IEEE, pp.423–427. https://doi.org/10.1109/ICIS.2018.8466510

Websites

68 Facial landmarks. [online] https://68landmarks.com/ (Accessed 2 Jun 2019).

Age detection using facial images. [online] https://towardsdatascience.com/age-detection-using-facial-images-traditional-machine-learning-vs-deep-learning-2437b2feeab2 (Accessed 14 Dec 2020).

Age doping . [online] https://www.vox.com/2014/10/20/6939271/age-test-aging-epigenetics-clockbiological- chronological-scandal-telomeres (Accessed 19 March 2021).

AGE, GENDER AND ETHNICITY (FACE DATA) dataset. [online] https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv (Accessed 14 Dec 2020).

CK+ dataset. [online] https://www.ri.cmu.edu/project/cohn-kanade-au-coded-facial-expression-database/ (Accessed 10 Aug 2019).

FACE++ API. [online] https://www.faceplusplus.com/emotion-recognition/ (Accessed 25 Jun 2019).

Facial age dataset. [online] https://www.kaggle.com/frabbisw/facial-age (Accessed 14 Dec 2020).

RBF Kernel. [online] http://www.saedsayad.com/support_vector_machine_reg.htm (Accessed 19 Oct 2020).

scikit-learn library. [online] https://scikit-learn.org/stable/ (Accessed 27 July 2019).

UTKFace dataset. [online] https://susanqq.github.io/UTKFace/ (Accessed 14 Dec 2020).