International Journal of Multidisciplinary Research and Growth Evaluation.

# Age estimation in social network using deep learning algorithm

**Abinaya A** [1*], **Baskar. A** [2]
[1-2] Computer Science and Engineering, E.G.S. Pillay Engineering College Nagapattinam, Tamil Nadu, India

* Corresponding Author: **Abinaya A**

## Article Info

**ISSN (online):** 2582-7138
**Volume:** 04
**Issue:** 03
**May-June** 2023
**Received:** 20-03-2023;
**Accepted:** 03-04-2023
**Page No:** 03-09

**Abstract**
Real-world human-to-human communication is made possible by the important visual signals provided by human faces, which also carry a great deal of nonverbal information. Modern intelligent systems should therefore be able to correctly recognize and comprehend human faces in real time. In applications of real-world facial image studies, such as multimedia communication, human- computer interaction (HCI), and security, identity, age, gender, appearance, and ethnic origin are all significant variables. Face mug photo retrieval is a tool that law enforcement agencies can use to locate potential criminal suspects. Only a small amount of research has been done on how to accurately evaluate and use demographic information like age, gender, and society that is present in facial photographs, despite the extensive study on human identification from facial photos. Estimating human ages from face photos is still a challenging subject, even though automatic image-based age valuation is a major method complex in many real-world applications. Numerous practical uses in online social networking apps can be derived from mechanically estimating human age using facial picture analysis. Online social networks (OSN) have played a major role in connecting people and facilitating information sharing for many years. OSN are essential platforms for (among other things) content and opinion transmission, and they are currently used by billions of machine workers to interact. Each social network also has an age limit for signing in. It is still a challenging issue to resolve in the present, though. This research examines age estimation from face datasets using a variety of face feature extraction algorithms.

**Keywords:** Social network, Facial image analysis, Facial feature points, Age estimation, Face detection

## 1. Introduction

One of the trickier parts of facial analysis has always been determining age from photographs. Some of the causes include the uncontrollable nature of the aging process and the intense precision to particular patterns. Similar to how it is for maximizing image popularity jobs, teaching the classifier properly requires a significant and extensive amount of data/pictures. Additionally, supervised classifiers need data and/or images to be tagged, in this instance with an accurate age. The formerly accessible datasets, however, were constrained and severely biased. This is particularly troublesome in video surveillance and forensics, where it's common to encounter unidentified people who are uncooperative. A lot of information about identification, age, gender, mood, and ethnicity can be found on a person's face. It is a crucial population and a straightforward biometric characteristic for identifying people. In person-to- person verbal conversation, a person's age is also significant. Face characteristics have an impact on a man or woman's appeal to another. They can provide you with exercise and reproductive hints. These factors can therefore increase a person's productivity and achievement. A lot of information about identification, age, gender, mood, and ethnicity can be found on a person's face. It is a crucial population and a straightforward biometric characteristic for identifying people. In person-to-person verbal conversation, a person's age is also significant. Face characteristics have an impact on a man or woman's appeal to another. They can provide you with exercise and reproductive hints. These factors can therefore increase a person's productivity and achievement.

The subject of Automatic Age Estimation (AAE) utilizing face pictures is challenging due to the huge variety of facial looks. It is brought on by a combination of external and internal influences. The extrinsic aspects are influenced by a variety of extrinsic circumstances, including the living environment, health, way of life, and others. On the other hand, intrinsic factors include physiological elements like genes. In strong AAE systems that solely rely on facial images, face expressions and variations in appearance must be taken into account. Only a few of the programs offered by AAE Systems include electronic customer dating management, monitoring, and content filtering for the internet. (E-CRM). They are particularly sought after because it is difficult for humans to determine their ages. As a result, developing AAE systems that outperform human performance is critical. The fundamental method for estimating the age of a person's face is shown in Figure 1.

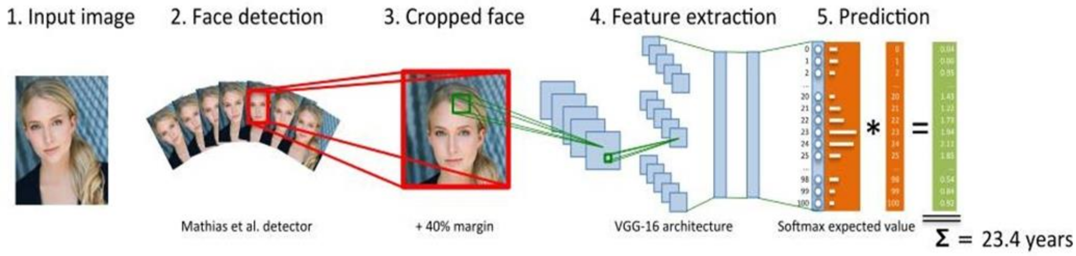


**Fig 1:** Face age estimation step

## 2. Related Work

O. Agbo-Ajala and S. Viriri, *et.al*,…[1] provides a version that makes use of CNN architecture to predict the gender and age group of human faces in situations that are not filtered in real life. This novel CNN method trains classifiers that predict the human's age and gender by treating the age and gender labels as discrete annotations. Additionally, we created a powerful and high-quality image preprocessing method to combine and prepare the raw images for the CNN model, which has a significant impact on the precision of our age and gender classifiers. Then, we demonstrate how effective teaching of our age and gender CNN version is made possible by retraining on large-scale datasets, enabling the classifiers to generalize on the images and avoid over fitting. Last but not least, the OIU-Audience benchmark is used to assess the performance of our novel CNN version. Despite the extremely challenging nature of the dataset's images, our technique produces full-size improvements in age institution and gender type accuracy over existing techniques, which can satisfy the needs of a number of real-world programmes.

V. Badrinarayanan, *et.al*…[2] A stack of encoders is followed by a reliable decoder stack in the Seg Net algorithm, which feeds into a smooth-max type layer. The decoders can map low resolution function maps on the encoder stack output to full input image length characteristic maps. This solves a major problem with current deep learning systems that use item classification networks for pixel smart labelling. There is no mechanism to map deep layer feature maps to input dimensions in these strategies. To enhance sample capacity, they employ ad hoc techniques like replication. As a result, there are fewer pooling layers accessible, which prevents excessive up sampling and reduces regional context by producing noisy forecasts. SegNet overcomes these difficulties by mastering the mapping of encoder outputs to image pixel labels. SegNet has a "flat" architecture, which means that each layer's set of capabilities (64 in our case) stays the same while yet having full connectivity. There are two factors that influence this desire. Secondly, it prevents parameter expansion, in contrast to a rising deep encoder network with complete function connectivity (identical for decoder). Second, since the usual map decision is smaller and convolutions are faster, the training time is constant (in our experiments, it really slightly lowers) for each additional/deeper encoder-decoder pair.

Ali Maina Bukar, *et.al*,…[3] One of the suggested contributions is to improve the traditional AAM by applying partial least-squares (PLS) regression in the PCA region. The dimensionality discounting method PLS (supervised appearance model) maximises the covariance between the predictor and the response variable, producing latent scores with each reduced dimension and higher predictive energy (SAM). The difficulty of estimating a ge and categorizing gender is then addressed using the feature extraction version. Finally, we look at how the classifications are applied to the FGNET-AD benchmark database (DB). In order to categories gender types, attributes must first be extracted. There are two categories that can be applied to the geometric and look-based feature extraction methods that have been discussed in the literature. Neighborhood functions are no longer sufficient for particular age estimation since geometric functions describe the most effective form modifications that happen in early life and local textures are restricted to wrinkles, which happen in adulthood.

K.-Y. Chang, *et.al*,… [4] Employ use relative positions of age labels since it provides more reliable information than specific age figures. Record retrieval frequently uses ranking, which has been formalized as a "learning to rank" technique that converts input files into ranked lists. Ranking based on regression or type was successfully accomplished with several early systems. Other well-known methods include paired selections and factor-smart ordinal regression. Also, they put forth the Ranking SVM system, which is totally based on hinge loss and the SVM technique. The learning process uses the difference between two characteristic vectors, and the higher-ranking vector is transferred to better test results. Rank Boost and RankNet use exponential loss and pass entropy loss, respectively, to study pairwise ranking algorithms. And come up with a cost-effective answer to any issue with the subs. The idea of fee-sensitive assets has lately gained attention in the community of gadget learners as a practical tool to highlight the severity of misclassification problems. Because the value of misclassification typically varies among various pairs of labels, value-sensitive learning aims to reduce overall cost rather than total errors.

K.-Y. Chang, C.-S. Chen, *et.al*,… [5] proposes a single ranking-based strategy that concentrates on the more accurate use of relative-order data for age estimation. Our approach breaks down the inference problem into a series of straightforward binary questions, which are then combined to predict age. The trials' findings show that the suggested method for tackling this issue by viewing it as a multi-magnitude or regression problem surpasses earlier methods.

Nonlinear regression techniques like the Gaussian Process (GP) or Support Vector Regression (SVR) were used to solve the age estimation problem. Yet, a human face can age in different bureaucracy, sha pe, and texture variations throughout the course of a very long period of time. This trait types the random procedure shaped by non-stationary ageing patterns because the kernel functions used to evaluate the pair-wise similarities between a lengthy period will be shift- or time-varying. Yet, it might be difficult to solve a regression problem using non-stationary kernels because doing so can quickly result in over-fitting during the learning phase.

C. Chen, A. Dantcheva, *et.al*,...[6] Early results on how facial cosmetics affect automated systems forpredicting gender and age were presented. This study proposes that makeup should be taken into account because automated gender and age estimate methodologies are used in many commercial applications. While a person cannot purposefully trick the system with cosmetics, it is simple to conceive scenarios where a hostile user may trick the system with conventional makeup. Automatic gender classification systems are impacted by gender spoofing brought on by makeup. Conversions from male to female and female to male are both feasible. It was discovered that the male-to male transition was slightly more challenging than the female-to-male change. In order to study makeup-induced age alteration, we also constructed another dataset (MIAA-Makeup Induced Age Alteration) made up of images downloaded from the Internet. One shot was taken prior to each participant applying makeup, and there are a total of 53 subjects in these pictures. Although we don't know the subjects' specific ages, we assume that they are older than 30 and that cosmetics are being utilized to make them look younger.

A. Das, A. Dantcheva, *et.al*,… [7] To lessen inter-class bias, a gender, age, and race class method was suggested. The proposed multi-undertaking CNN technique uses joint dynamic loss and performed well on the UTK Face and the Bias Estimation in Face Analytics (BEFA) challenge datasets. In 2018, the BEFA competition was won by the recommended algorithm at the European Conference on Computer Vision (ECCV). In further work, we intend to broaden the current style to include more facial features. We also intend to look into the method described in this paper in terms of reducing biases in face recognition. The widespread commercial adoption of automatic face evaluation systems-using face popularity as a trustworthy authentication strategy-has aroused interest from the therapeutic community. The age, ethnicity, and gender of faces can be identified, evaluated, and categorized using face photos using the device learning algorithms that are currently available.

E. Eidinger, R. Enbar, *et.al*,…[8] deliver two contributions: a class pipeline created to make the most of the limited data available, as well as a brand new, significant records set and benchmark for age and gender estimates. We also provide a novel, reliable face alignment method based on iterative evaluation of the localization uncertainty of facial functions. Last but not least, we offer thorough tests that illustrate the improved capabilities of our method and the raised level of complexity of our new benchmark. But to a lesser extent than the linked issue of face popularity, determining a person's age from their facial features in a photograph has been studied in the past. Also, give a thorough analysis, contrasting each prospective benchmark with the difficulty levels that correspond to them as well as the accuracy of automated age and gender estimation methods. We demonstrate that our personal device outperforms competitors on all of the evaluated benchmarks by significant margins.

X. Geng, C. Yin, *et.al*,… [9] IIS-LLD is a set of iterative optimization rules based on the greatest entropy version that are suggested for learning from label distributions. Results from experiments demonstrate IIS-advantages LLD's over conventional learning techniques that rely on single-classified data. LLD can be helpful for several learning difficulties in addition to achieving the proper overall performance on face age estimation. Usually speaking, LLD will be helpful in at least three of the following situations: The instances are initially given elegance distribution labels. The distributions of elegance may originate from facts or expert information. Certain learning and training have a strong correlation. A label distribution may be clearly generated in accordance with the correlation between the different classes. It is questionable how unique sources should be labelled. It could be better to create a label distribution that contains the data from all assets rather than selecting a single winning label.

G. Guo and G. Mu, *et.al*,… [10] gave a novel insight that estimating age, gender, and ethnicity need the greatest three talents. An extensive database containing more than 55 face pictures was used for the experimental validations. We examined the impact of feature dimensionality on the use of the rank idea for CCA-based algorithms. The behaviours of the CCA and PLS-based approaches have been thoroughly compared, along with any inaccuracies or inconsistencies in the dimensions and walking duration. In trials, the rCCA walks at a similar speed to the CCA and PLS but with fewer mistakes. The regularized CCA is suggested for practical purposes based on a typical attention due to its quick pace and surprisingly small errors. Recent studies have shown that PLS and CCA-based approaches perform exceptionally well in resolving computer vision issues. In order to better understand the behaviors of PLS and CCA-based techniques in both broad and precise vision applications, it is crucial to test and assess them in a variety of vision challenges.

## 3. Age Estimation Techniques

The determination of someone's age based on biometric capabilities is known as age estimation. Although unique biometric developments can be utilised to estimate age, this article focuses on facial age estimation, which uses biometric information taken from a person's face. The article's main points cover common packages that can be used for facial age estimate, the hassle and obstacles involved with facial age estimation, common methodologies described in the literature, and future study instructions. The purpose of automatic facial age estimation is to use dedicated algorithms to approximate a person's age using functions derived from his or her face image. The facial age estimation problem is comparable to other common face image interpretation tasks in that the execution level includes the face detection technique, area of facial features, function vector components, and type. The output of the class degree can be an estimate of a person's precise age or the age institution of someone, or a binary result indicating whether or not the age of a subject is within a positive age range, depending on the software for which an age estimation device is intended to be used. The age-group category is the most widely utilised of the three types stated above, as it is significantly easier to obtain a rough estimate of a topic's age rather than his or her true age in most packages. Another essential aspect of the age estimation problem is the temporal span that is taken into account. This parameter is a critical aspect of the problem since distinct ageing features appear in different age groups; hence, a device capable of dealing with a specific age range may not be applicable to other age groups. The difficulty of face age estimation is comparable to the problem of age progression.

Age progression is the forecast of a subject's future face look based solely on images of his or her prior a ppearance. Age estimate and age development must both be taken into account when explaining age-related face deformations that occur throughout the course of a person's life. In some cases, In perspective, the age estimation issue is resolved, however in the opinions of others, the issues of age estimate and age progression are both dealt with in a comparable ways. A face age estimation method is composed of the two main terms as well as function extraction and class techniques. Many algorithms are covered in the ensuing chapters.

### 3.1 Features Extraction

Age estimation can be calculated based on facial feature points and feature points based landmark values of each face images. Figure 2 shows the main facial feature points.

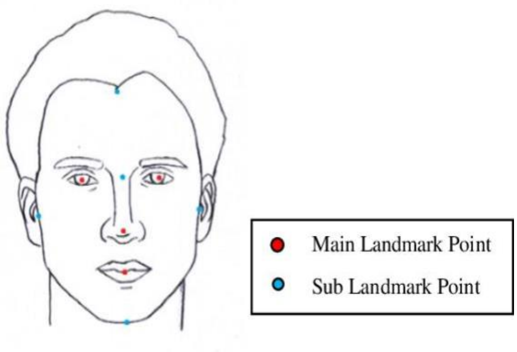


**Fig 2:** Facial landmark points

### Principal Component Analysis (PCA)

PCA is an uncontrolled characteristic extraction method that produces the most important additives of facts that are ordered according to disagreement standards. In PCA-assisted characteristic removal, the mechanism involving the higher authority is kept, while the mechanism involving the lower authority is eliminated. As a result, PCA may not perform well in categorization due to the fact that it ignores the facts of division labels and training intolerance. Most important element evaluation is one of the most widely used unsupervised characteristic elimination strategies. PCA does not use the instruction restricted inside the elegance labels for supervised categorization problems. PCA is the eigen value disintegration of records in addition to the statistical instructions with the greatest variety. Because PCA capabilities show an increased amount of electrical compaction, it is a favoured characteristic region. Uncorrelated production functions are generated by PCA.

With the help of PCA, a lower dimensional portrayal of unique information is generated during the time of photographing the statistics course to facilitate has the greatest difference. Through connecting the Eigen standards with eigenvectors of covariance environment of unique information, this unsupervised characteristic removal technique tasks a statistics locate to a new synchronize scheme. They're then sorted in a sliding order after the eigen values and eigenvectors have been worked out. By merging the eigenvector matrix, the important additives may be built as a linear refurbishment of statistics. The PCA model is described as follows:

1) Let have N data samples $S_1$, $S_2$, $S_3$,….. ,Sn in M dimensional space. Each Si is a M X 1 vector. Let $\bar{S}$denotes the mean vector of the contribution data and preserve be represented as

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} Si$$

2) C which represents the covariance matrix is distinct as

$$C = \frac{1}{N}\sum_{i=1}^{N} (Si - \bar{S})(Si - \bar{S})^T$$

3) Let $\Psi_1, \Psi_2, \ldots, \Psi_N$ are the n eigen vectors corresponding n largest eigen values of C. appearance the vector $W=[\Psi_1, \Psi_2, \ldots, \Psi_N]$ Find the characteristic vector as

$$Y_i = W^T((Si - \bar{S})\forall i = 1,2,\ldots,N$$

Significant records are specifically managed with the use of only a few number of critical additives, and it is commonly a linear accumulation of records from multiple instructions.

### Linear Discriminant Analysis

To identify a linear mixture of functions, LDA algorithms are utilized in information, model identification, and machine learning. One dependent variable is expressed as a linear combination of several features or measurements using LDA. In the same way that PCA and thing analysis look for a linear aggregate of variables that best explain the facts, LDA does as well. LDA makes an explicit attempt to version the change in statistics training. PCA, on the other hand, ignores any distinction in class, relying instead on aspect evaluation to construct the characteristic.

Combination is entirely based on contrasts rather than similarities. LDA looks for vectors in the fundamental liberty that are easily distinguishable across curriculum.

LDA develops a linear aggregate of these, which yields the primary mean differences among the favored teachings, after an official agreement on a number of impartial qualities against which the facts are characterized.

We describe two events: 1) Solitary is called inside- class scatter environment as given by

$$S_w = \sum_{j=1}^{c}\sum_{i=1}^{N_j}(x_i^j - \mu_j)\ (x_i^j - \mu_j)^T$$

Where$x_i^j$ is the ith example set j, $\mu_j$is the indicate of set j, c is the amount of program and $\mu j$ is the number of samples in set j and I linking set scatter matrix

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T$$

Where $\mu$ represent the indicate of all program.

$ST$ is the total scatter matrix. Because its extracted capabilities exploit the elegance information, LDA is a more efficient function extraction approach than PCA in supervised learning. The distributions of samples in each phase, however, are considered to be normal and homoscedastic. As a result, if this assumption is broken, finding a proper and illustrated characteristic area can be challenging.

### 3.2 Age Classification

Cluster face values based on features and classify the findings to determine a person's age. This is how the classification can be defined:

- The distance between two points of interest (for example, the separation of the eyes)
- The distance calculated along an axis between two landmarks (such as the vertical height of the nose)
- The distance calculated along a surface between two landmarks (such as the upper lip boundary arc length)
- Inclination angle relative to an axis (for example, the slope of the nose bridge)
- The angle between the two face positions (such as the angle formed at the tip of the nose)

### Support Vector Machine

Support vector machines (SVM) are supervised learning models with related learning algorithms for classification and regression analysis in machine learning. It's primarily used to solve categorization challenges. Each data item is plotted as a point in n-dimensional space (where n is the number of features), with the value of each feature equal to the value of a certain coordinate. The hyper-plane that best distinguishes the two classes is then used to classify the data. SVMs may also conduct non-linear classification, implicitly translating their inputs into high-dimensional feature spaces, in addition to linear classification. A Support Vector Machine (SVM) is a discriminative classifier using a separating hyper plane as its formal definition. In other words, the algorithm produces an ideal hyper plane that categorizes fresh cases given labelled training data (supervised learning).

## 4. Probabilistic Neural Network Classifiers

Age estimation is very important issue in the field of computer vision or HCI that is Human Computer Interaction from last several years. Thus, study age estimation system is to be carried out which will take human facial image as input and system will classify that in specific age year or in a different age group. Age estimation from face images is still very challenging compared to other cognition problems. This is primarily because many external factors influence the aging progress. The aging process can be accelerated or slowed down by physical condition, living style. The most important parts for the process of automatic age estimation are the internal parts of the face and in particular the area around the eyes. As a result, different people with the same age can have quite different appearances due to different rates of facial aging. Many methods have been proposed to detect faces such as neutral networks, skin locus, and color analysis. Usually, the image sequence has the face in frontal view. Once the face is detected from the image sequence, the next step is to extract the information about age of given images. Because of high variability in the types of faces, it is very difficult for the machine to extract facial features. Variations in lighting conditions, head movements, non-frontal views, various distractions like glasses, facial hair makes the problem more difficult. Finally, we have to classify the extracted facial information into a particular age or age group. Principal Components Analysis (PCA) is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. And we can implement the system in real time social network environments. The proposed framework is shown in fig 3.

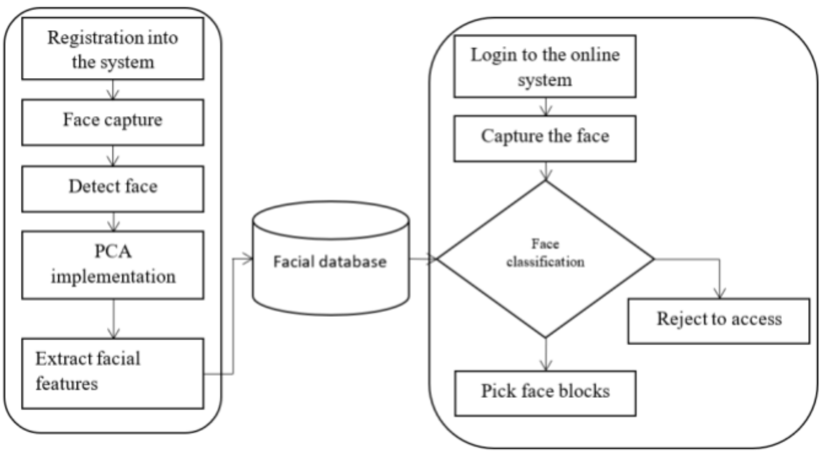


**Fig 3:** Proposed work

The Probabilistic Neural Network (PNN) is a network-based approach to "probability density estimation." It is a competitive learning paradigm with a "winner takes all" mentality and a basic notion based on multivariate probability estimation. The computational load from the training phase is shifted to the evaluation phase, which makes PNN unique.

The fundamental advantage of PNN over back propagation networks is that training is instantaneous, simple, and fast. The Parzen window notion of multivariate probabilities was used to construct PNN. The Bayes technique for decision-making is combined with a nonparametric estimator for getting the probability density function in the PNN, which is a classifier version. An input layer, a pattern layer, a summation layer, and an output layer make up the PNN architecture. The essential framework of the PNN method is shown in Figure 4.

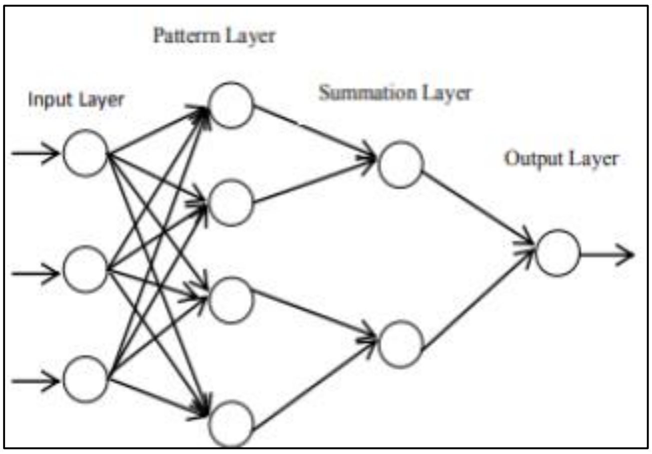


**Fig 4:** PNN approach

The neuron $xij$ of the pattern layer receives a pattern x from the input layer and computes its output as given by the equation below

$$\emptyset_{ij}(x) = \frac{1}{(2\pi)^{d/2}\sigma^{d}} exp\left[-\frac{(x-x_{ij})^{T}(x-x_{ij})}{2\sigma^{2}}\right]$$

Where $\sigma$ denotes the smoothing parameter

$xij$ denotes the neuron vector and d denotes the dimensions of the pattern vector x The summation layer neurons compute the maximum likelihood of pattern x being classified into C, by summarizing and averaging the output of all neurons that belong to the same class using equation given below

$$\rho_i(x) = \frac{1}{(2\pi)^{d/2}\sigma^{d}} \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left[\frac{(x-x_{ij})^{T}(x-x_{ij})}{2\sigma^{2}}\right]$$

Where Ni is the total number of samples in Class Ck. The decision layer unit classifies the pattern x in accordance with bayes decision rule based on the output of all summation layer neurons by

C(x)=argmax{pi(x)} for i=1, 2,…,m

Here C(x) denotes the estimated class of the pattern x and m is the total number of classes in the training samples. Excessive categorization features raise both compute time and storage memory requirements. They can make categorization more difficult at times. A reduction in the number of features is required. Reduced dimension refers to a smaller collection of features that are fed into the PNN during the training and testing phases.

## 5. Conclusion

The suggested architecture is tested with real-time database images, the most effective of which is a single face. This is the simplest method if you wish to keep local community statistics that allow you to determine a characteristic that changes over time. The current device discretely classifies the worry in the photograph, then guesses the approximate age of the man or woman near to the age referenced to in the database and displays a range within which the person's age could fall. The age estimation technique with PCA and Euclidian distance classifier is discussed in this project. As previously said, age or age group identification is broken into three sub-issues: face detection, characteristic extraction, and type. Because there are a few issues such as head rotation, the effect of ageing, variations in illumination due to minor influences, and so on, this strategy isn't always suitable for all of the challenges. As a result, if the neural community is exploited, greater impacts can be obtained. Also, if a first-class camera is employed to capture images for database creation, results can be advanced. Increased database length and the usage of multiple classifiers could be used to further refine the suggested system.

## 6. References


1. O Agbo-Ajala, S Viriri, Deeply learned classifiers for age and gender predictions of unfiltered faces, Sci. World J. 2020, 1-12. doi: 10.1155/2020/1289408.
2. V Badrinarayanan, A Handa, R Cipolla. Seg Net: A deep convolutional encoder-decoder architecture for robust semantic pixel wise labelling, 2015, arXiv: 1505.07293. [Online]. Available: http:// arxiv.org/abs/1505.07293
3. AM Bukar, H Ugail, D Connah. Automatic age and gender classification using supervised appearance model, J Electron. Image. 2016; 25(6):061605, doi: 10.1117/1.JEI.25.6. 061605.
4. KY Chang, CS Chen, YP Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation, in Proc. CVPR, 2011, 585-592.
5. KY Chang, CS Chen, YP Hung. A ranking approach for human ages estimation based on face images, in Proc. 20th Int. Conf. Pattern Recognit, 2010, 3396-3399.
6. C Chen, A Dantcheva, A Ross. Impact of facial cosmetics on automatic gender and age estimation algorithms, in Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP). 2014; 2:182-190.
7. A Das, A Dantcheva, F Bremond. Mitigating bias in gender, age, and ethnicity cla ssification: A multi-task convolution neural network approach, in Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, 2018, 1-13.
8. E Eidinger, R Enbar, T Hassner. Age and gender estimation of unfiltered faces, IEEE Trans. Inf. Forensics Security. 2014; 9(12):2170-2179.
9. X Geng, C Yin, ZH Zhou. Facial age estimation by learning from label distributions, IEEE Trans. Pattern Anal. Mach. Intell. 2013; 35(10):2401-2412.
10. G Guo, G Mu. Joint estimation of age, gender and ethnicity: CCA vs. PLS, in Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG), 2013, 1-6.
11. G Guo, Y Fu, CR Dyer, TS Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 2008;

17(7):1178-1188.
12. D Huang, L Han, F De la Torre. Soft-margin mixture of regressions, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, 6532-6540.
13. G Levi, T Hassncer. Age and gender classification using convolutional neural networks, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2015, 34-42.
14. Z Niu, M Zhou, L Wang, X Gao, G Hua. Ordinal regression with multiple output CNN for age estimation, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, 4920-4928.
15. K Li, J Xing, W Hu, SJ Maybank. D2C: Deep cumulatively and comparatively learning for human age estimation, Pattern Recognit. 2017; 66:95-105.