



Emotion recognition to support personalized therapy in the elderly: an exploratory study based on CNNs

Ariane Sarmento Torcate¹ · Maíra Araújo de Santana¹ · Wellington Pinheiro dos Santos²

Received: 25 July 2023 / Accepted: 17 June 2024

© The Author(s), under exclusive licence to The Brazilian Society of Biomedical Engineering 2024

Abstract

Purpose Emotions are pivotal in human life, and the ability to recognize and interpret emotions accurately holds significant value, particularly for populations such as the elderly who face challenges in expressing their emotions. Emotion recognition systems serve as valuable tools in various scenarios, including the customization of therapies, with the potential to enhance their effectiveness and precision.

Objective To contribute in this context, we present a approach based on traditional Convolutional Neural Networks for emotion recognition in facial expressions. **Methods:** The proposed CNN model is composed of six convolutional layers and was trained, validated and tested with a comprehensive and robust dataset, formed by combining and joining the FER-2013, Chicago Face, Yale Face and KDEF databases. After model training, we apply Haar Cascade to static images of elderly people to identify the face and our CNN model to classify emotions.

Results The result obtained during the training/validation stage was 90.77% accuracy, demonstrating robustness with a diverse data set. However, accuracy dropped to 69.05% during the testing stage, aligning with the literature on generalization challenges. Despite this, sensitivity and specificity remained strong at 0.8882 and 0.9823 respectively, indicating reliable detection of true positives and negatives. When comparing this performance with studies in the literature, our model stands out for having better results. Furthermore, the proposed model successfully classified emotions in images of elderly subjects.

Conclusion The outcomes of our study are highly promising. The development of intelligent emotion recognition systems for the elderly represents a viable alternative for enhancing the quality of life for this population and their support networks.

Keywords Emotion Recognition · Personalized Therapies · Elderly · CNNs · Facial Expressions

Introduction

Motivation and problem characterization

Emotions constitute a vital component in the human experience, exerting direct influence on various aspects, encompassing communication, daily activities, and personal as well as social advancement (Khateeb et al. 2021). Fundamental emotions can be comprehended as involuntary physiological responses that are visually discernible and evolve over time (Bomfim et al. 2019). It is important to acknowledge that emotion is not confined to a single variable; rather, it represents a multifaceted process comprising diverse interconnected elements (Abdullah et al. 2021). In this regard, individuals possess multiple avenues for expressing their emotions, encompassing facial expressions, voice, galvanic skin response, electroencephalographic signals, among others (Abdullah et al. 2021). Among these channels, facial

Ariane Sarmento Torcate, Maíra Araújo de Santana, and Wellington Pinheiro dos Santos are contributed equally to this work

✉ Wellington Pinheiro dos Santos
wellington.santos@ufpe.br

Ariane Sarmento Torcate
ast@ecomp.poli.br

Maíra Araújo de Santana
mas2@ecomp.poli.br

¹ Núcleo de Engenharia da Computação, Escola Politécnica da Universidade de Pernambuco, Recife, PE, Brazil

² Departamento de Engenharia Biomédica, Universidade Federal de Pernambuco, Recife, PE, Brazil

expressions and speech prevail as the primary conduits through which emotional information is conveyed, accounting for approximately 93% of emotional content in human communication (Pan et al. 2023).

Alreshidi and Ullah (2020) explain that the importance of considering human feelings in the development of technological devices is undeniable, since emotions are present in everyday life and are essential for human communication. In order to improve the interaction between human and machine, in recent years the focus of Affective Computing has been to investigate efficient methods for the recognition of emotions, seeking to make the existing emotion in communication between people, is also present in the interaction between humans and computers.

In addition to emotion recognition systems, medical devices and personalized medicine have gained prominence in recent years due to their ability to leverage the benefits of medical practice (Santana et al. 2018; de Freitas Barbosa et al. 2021; Espinola et al. 2021; Oliveira et al. 2020; Nunes et al. 2023; de Santana et al. 2022; Shirahige et al. 2022; de Souza et al. 2021). According to Motadi et al. (2023), personalized medicine is understood as an innovative method to change the diagnosis, prevention and treatment of diseases considering the differences and individualization of each individual. Through individual patient data, it is possible to devise strategies to individualize your own care, from diagnosis, selection and monitoring of treatment, to therapeutic practices (Ho et al. 2020).

With regard to therapeutic practices, it is worth noting that emotion recognition systems through facial expressions can significantly contribute to personalized and assertive therapy sessions, especially for certain audiences who have difficulty expressing emotions, such as children with autism spectrum disorder (ASD) (Teh et al. 2018) and the elderly (Ferreira and Torro-Alves 2016). Making a specific cut for the elderly, it is important to understand that along with the natural aging process, changes in perception and cognition can generate damage that imply that the elderly have difficulties both in expressing emotions through the face and also in recognizing facial emotions in another individual (Ferreira and Torro-Alves 2016; Torcate et al. 2023). Therefore, perceptions of emotional facial expressions develop in childhood, reach an ideal state in youth, and begin to decline in old age.

Ferreira et al. (2021) explain that the elderly have impairments in the recognition of emotions, regardless of whether they are healthy or have some cognitive impairment. Considering all the limitations of this public, in 2014, Castillo et al. (2014) already discussed the importance of developing personalized solutions for the elderly, where one of the main tools pointed out were emotion recognition systems. In 2020, Boateng and Kowatsch (2020) already pointed to emotion recognition systems as

important to assess the emotions of the elderly, especially in nursing homes, in order to improve their mental health. Jiang et al. (2022) point out that emotional assessment/recognition systems provide a continuous and non-invasive assessment of the emotional state of individuals and that, in the case of the elderly, it can be used to assess their emotions in therapies, as well as to inform activities and interventions to improve their mental health. Considering this context, according to Ferreira et al. (2021), research that seeks to develop interventions or contribute with new scientific evidence in relation to this public is totally necessary and opportune.

Considering that the ability to express and recognize emotions through the face is a fundamental stage of basic communication, not being able to signal emotions such as anger, sadness or disgust can result in social isolation (Grondhuis et al. 2021) or negatively affect the verbal or non-verbal communication of individuals. As a result, older adults may have difficulties communicating important messages, such as the discomfort associated with treatments. In the context of therapies, identifying the emotions of this audience at the time of therapy helps the therapist to change/customize his/her approach, depending on the biofeedback returned (Santana et al. 2021).

Despite this fact, Ma et al. (2019) reinforce that studies in the literature are entirely dedicated to the recognition of emotions in young people and adults, and to a lesser extent, in children and adolescents. Little research addresses emotion recognition in the elderly. The authors state that aging causes many changes in the shape and appearance of the face and can alter non-verbal behavior patterns, therefore, it is important that systems developed for the automatic recognition of emotions, specifically for this audience, consider all their respective limitations (Ma et al. 2019). However, it is worth mentioning that there are few databases built for emotion recognition in the elderly. This data gap is even greater in relation to emotion recognition in elderly people with dementia, such as Alzheimer's disease.

Knowing this problem and with the aim of contributing in this context, we present a traditional CNN (Convolutional Neural Networks) approach to recognize emotions in facial expressions. It is worth mentioning that as a differential, our model was trained/validated and tested with a complete database, created from the combining and junction of four important databases available in the literature, which are: FER-2013 (Goodfellow et al. 2013), Chicago Face Database (Ma et al. 2015), KDEF (Lundqvist et al. 1998) and Yale Face (Belhumeur et al. 1997). After training and validation of the model, we applied the Haar Cascade Frontal Face (Viola and Jones 2001) in static images of elderly people to detect the face and, subsequently, we used our CNN to perform the classification of emotions.

It is worth mentioning that this research is fully exploratory, but that it aims to contribute to the development of a more robust system, capable of positively impacting real applications that target the elderly, and that considers the limitations and changes caused by aging (such as wrinkles, folds, wear of facial muscles, atrophy of the facial skeleton, loss of soft tissue, among others) (Grondhuis et al. 2021). As a novelty and contribution, the model we are proposing was trained/validated and tested considering the heterogeneity of the four databases combined, creating a more robust data set. Finally, we hope that our research will also provide insights for other researchers, helping to identify limitations, needs and opportunities in developing work in this field.

Related works

Several studies in the literature use CNNs to perform computer vision and image processing tasks. Some of the applications of this model are to perform emotion recognition in static or dynamic images. Bodapati et al. (2022) developed a CNN-based model to perform emotion recognition, using the FER-2013 database. The model has a sequence of blocks, each composed of several convolutional layers and subsampling layers. In addition, the authors made adjustments to important hyperparameters, such as dropout, batch normalization, max pooling, among others. As a result, the developed model presented an accuracy of 69.57%.

Also using the FER-2013 database, Khopkar et al. (2021) proposes a deep CNN implemented using TensorFlow and Keras. The number of epochs defined for model training was 15. The result obtained in terms of accuracy was 66.7%. In addition, the model performed better to classify the emotions of Happy, Sad and Surprised. Which is understandable, as both are the emotions that are best represented in the said database. Sadeghi and Raie (2022) developed a CNN that uses a histogram calculation layer to provide statistical description of feature maps in the output of convolutional layers. In the histogram space, a graspable matrix is introduced into the chi-square distance equation. Then, the modified equation is used in the loss function. The accuracy related to the performance of the CNN proposed for recognizing facial expressions of seven classes of emotions in the CK+, MMI, SFEW and RAF-DB databases are 98.47%, 83.41%, 61.01% and 89.28%, respectively.

Agrawal and Mittal (2020) present two new CNN models, both trained with the FER-2013 database. The authors sought to investigate the effects of kernel number and filters on the CNN architecture. Model 1 architecture is unique in that it not only uses fixed kernel size, but also fixes the number of filters in the network depth. In the Model 2 architecture, the number of filters decreases with the depth of the network. Both architectures use a kernel size of 8. In the model implementation, dropout and fully connected layers

were not used. The results show that model 2 achieved 65% accuracy, in addition to having a smaller number of parameters.

Zahara et al. (2020) presented a CNN-based system design for facial emotion recognition implemented on the Raspberry Pi. The system consists of three main processes, which are: i) face detection, ii) feature extraction and iii) emotion classification. For training the CNN model, the FER-2013 database was used. As a result, an accuracy of 65.97% was achieved. Another study developed by Borgalli and Surve (2022) presents a customized CNN to perform the recognition of facial emotions in static images. To train the model, a k-fold cross-validation method was used. The developed CNN was trained using the FER-2013, CK+ and JAFFE database separately, with an accuracy of 86.78%, 92.27% and 91.58%, respectively.

In order to analyze the exact emotion of a given user's facial expressions in real time, John et al. (2020) developed a CNN with built-in facial landmarking and HOG (pre-processing methods). Initially, the facial images were pre-processed after capture, from which the features and emotion detected by the CNN model were extracted. For training the proposed model, the JAFFE and FER-2013 databases were used. As a result, for the JAFFE base an accuracy of 91.2% was achieved. As for the FER-2013 database, the accuracy obtained was 74.4%. Another interesting study was developed by Mounq et al. (2022), where a personalized CNN was proposed, composed of three classification models, namely: i) CNN; ii) ResNet50 and iii) InceptionV3. The model mean set classifier method is used to group the predictions of the three models. Subsequently, the proposed model is trained and tested with the FER-2013 dataset. The obtained results demonstrate that the personalized CNN reached 72.3% of accuracy in the classifications. But the proposed model performed poorly in detecting negative emotions compared to a single classifier.

Gautam and Seeja (2023) also contribute by developing an effective framework for emotion recognition. The strategy is to use HOG and SIFT to extract the explicit attributes of the images and use a CNN to classify the emotions. The experimental results of the proposed model show that it is able to recognize emotions with an accuracy of 98.48% for HOG-CNN and 97.96% for SIFT-CNN for the CK+ dataset. For the JAFFE dataset, the accuracies found were 91.43% for HOG-CNN while 82.85% for SIFT-CNN.

Recent research developed by Gahlan and Sethia (2024) presents a Deep Convolutional Recurrent Attention Network (DCRAN) with the potential to perform facial emotion recognition. The effectiveness of the DCRAN model proposed by the authors is demonstrated through validation using the FER-2013 database, achieving a test accuracy of 81.1%. Also using deep convolution neural networks, Almulla (2024) proposes a system to perform emotion recognition

based on text, audio and facial expressions. To train and test the model, the FER-2013, RAVDESS and ETD datasets were used. The results show a recognition accuracy of 100% for audio, 69% for faces and 64% for text. When applying decision-level fusion, the accuracy rate recorded is 80%.

The study developed by Maghari and Telbani (2024) used Vision Trans- former (ViT) to accurately recognize facial emotions in masked faces. The dataset used to train and test the model was AFFECTNET. To insert facial masks into the images, the authors used the "Mask the Face" script. As a result, the model achieved an accuracy of 81% in classifying facial emotions. Zakieldin et al. (2024) also contribute in this context by presenting the ViTCN model, composed of a combination of Vision Transformer (ViT) and Temporal Convolution Network (TCN).

The proposed architecture was validated on DFEW, AFFWild2, MMI and DAiSEE datasets. The experimental results demonstrated that ViTCN achieved 70.14%, 95%, 99.2% and 83.42% accuracy, respectively.

A hybrid model can be understood as one that uses two different algorithms to perform different functions, but which together complement each other to conceive the final objective of a given system. Knowing this, the work developed by Hosgurmah et al. (2022) presents a hybrid model for the recognition of emotions using facial expressions, using the ORL and YALE databases. The CNN algorithm was applied for attribute extraction and the classification was performed by the linear collaborative model of discriminant regression (LCDRC). The results indicated that the proposed model (CNN-LCDRC) reached an accuracy of 93.10% for the ENT database and 87.60% for the YALE database. While only the traditional LCDRC reached 83.35% and 77.70% accuracy in the respective bases. Using the FER-2013 base, Wahab et al. (2021) proposes a hybrid model composed of a CNN responsible for extracting attributes from the images and a KNN algorithm (k-nearest neighbors) is used to perform the classification task. As a result, the CNN-KNN hybrid model achieved an overall accuracy of 75.26%.

Ruiz-Garcia et al. (2018) developed a hybrid model to be integrated into the humanoid robot to perform real-time emotion recognition through facial expressions. The model consists of a CNN used for attribute extraction and an SVM (Support Vector Machine) responsible for classification. When applied to the KDEF database, the model achieved 96.26% accuracy, but when integrated into the robot, the CNN- SVM model achieved a lower accuracy rate of 68.75%. In order to contribute in the context of personalized therapies, Torcate et al. (2023) propose a hybrid model to perform emotion recognition in facial expressions. The proposed model is composed of a LeNet network and the Random Forest algorithm. For training, testing and validation of the model, a complete database (consisting of the FER- 2013, Chicago Face Database, KDEF and

Yale Face databases) was used. As a result, 70.52% accuracy was obtained in the training/validation stage and 82.92% in tests, despite all the variability between images from different databases.

In the context of recognizing emotions in the elderly through facial expressions, Lopes et al. (2018) developed a model with the aim of classifying emotional facial expressions both in the elderly and in other age groups, with the aim of making a comparison between both. The proposed model used the Lifespan database. In the face detection step, Haar Features and the Gabor filter were applied to extract facial features. These characteristics were classified by a Multiclass SVM. The results demonstrate an accuracy of 90.32%, 84.61% and 66.6%, when detecting the state of neutrality, happiness and sadness, respectively in the elderly. In young people and adults, the correct answers were 95.24%, 88.57% and 80%, when detecting neutral, happy and sad states.

In order to understand why it is difficult to identify an emotional expression displayed by an older face than a younger one, Grondhuis et al. (2021) carried out experiments investigating the influence of two variables that can cause this difficulty, they are: (i) wrinkles/folds or (ii) facial muscles. For this, a database with images of 28 individuals was used. A Generative Adversarial Networks (GANs) was applied to make the faces look older or younger artificially. The obtained results demonstrate that the model was able to correctly classify 80% of the cases. Where, younger faces when artificially wrinkled are 16.2% less likely to be identified. However, emotions on faces of naturally elderly people were 50.9% less likely to be correct. In contrast to this, older, artificially rejuvenated faces were 74.8% less likely to be correct compared to natural youth.

For better understanding and analysis, in the Table 1 we present a brief summary of the works cited throughout this section, containing the method used, database and main results. The last row of the table contains the model we are proposing.

Material and methods

Datasets

The database of facial expressions, called "Complete", used to carry out the experiment (see [Experimental setup](#) section) reported in this research, is composed of the combining and junction of four databases, which are:

1. **Facial Expression Recognition 2013 (FER-2013):** Database of facial expressions introduced in Challenges in Representation Learning (ICML 2013) (Goodfellow et al. 2013). All images that make up the FER-2013 base have gray scale and 48×48 pixels resizing. This database is composed of 35.887 images, distributed in the

Table 1 Summary of related works

Work	Method	Database	Main Results
Ruiz-Garcia et al. (2018)	Hybrid model composed of a CNN and an SVM	KDEF	The CNN-SVM model achieved 96.26% accuracy. When the model was integrated into a robot, the accuracy dropped to 68.75%
Lopes et al. (2018)	Haar Features and Gabor filter to extract attributes and SVM to perform classifications	Lifespan	Accuracy of 90.32%, 84.61% and 66.6%, when detecting neutral, happiness and sadness in the elderly. In young people and adults, the accuracy was 95.24%, 88.57% and 80%
Agrawal and Mittal (2020)	CNN	FER-2013	Accuracy of 65%
Zahara et al. (2020)	CNN built into the Raspberry Pi	FER-2013	Accuracy of 65.97%
John et al. (2020)	CNN with face frame and HOG (pre-processing methods) built-in	JAFFE and FER-2013	For JAFFE an accuracy of 91.2% was achieved. As for FER-2013, the accuracy was 74.4%
Khopkar and Adholiya (2021)	CNN	FER-2013	Accuracy of 66.7%
Wahab et al. (2021)	Hybrid model of CNN and KNN	FER-2013	The CNN-KNN model obtained 75.26% accuracy
Bodapati et al. (2022)	CNN	FER-2013	Accuracy of 69.57%
Sadeghi and Raie (2022)	CNN with a histogram calculation layer	CK+, RAF-DB, MMI and SFEW	The accuracy for the CK+, MMI, SFEW and RAF-DB bases were 98.47%, 83.41%, 61.01% and 89.28%, respectively
Borgalli and Surve (2022)	CNN	FER-2013, CK+ and JAFFE	Accuracy of 86.78%, 92.27% and 91.58%, for the FER-2013, CK+ and JAFFE bases
Moung et al. (2022)	Ensemble composed of CNN, ResNet50 and InceptionV3	FER-2013	Accuracy of 72.3%
Hosgurmath et al. (2022)	Hybrid model composed of CNN and LCDRC	ORL and Yale	The CNN-LCDRC model reached an accuracy of 93.10% for ENT and 87.60% for Yale
Grondhuis et al. (2021)	Using GANs networks to recognize emotions in the elderly	Own	The model was able to correctly classify 80% of cases. But artificially wrinkled faces are 16.2% less likely to be identified. On faces of naturally elderly people, 50.9% less likely to be hit
Gautam and Seeja (2023)	HOG and SIFT techniques for feature extraction and CNN for classification	CK+ and JAFFE	The HOG-CNN model obtained an accuracy of 98.48% for the CK+ base and 91.43% for JAFFE. SIFT-CNN obtained 97.96% accuracy in CK+ and 82.85% for JAFFE
Torcate et al. (2023)	Hybrid model composed of a LeNet and a Random Forest	Complete database	70.52% accuracy in the training/validation stage and 82.92% in tests
Gahlan and Sethia (2024)	DCRAN	FER-2013	Accuracy of 81.1%
Almulla (2024)	Deep Convolution Neural Networks	FER-2013	Accuracy of 69%
Maghari and Telbani (2024)	Vision Transformer (ViT)	AFFECTNET	Accuracy of 81%
Zakieldin et al. (2024)	Vision Transformer (ViT) and Temporal Convolution Network (TCN)	DFEW, MMI, AFFWild2 and DAISEE	The experimental results demonstrated that ViT achieved 70.14%, 95%, 99.2% and 83.42% accuracy, respectively

Table 1 (continued)

Work	Method	Database	Main Results
Our approach	Combining and junction four databases for training/validation of elderly people	FER-2013, Yale, Chicago and KDEF	The proposed model performed well during the training/validation stage in terms of accuracy (90.77%), kappa (0.8873), sensitivity (0.9872), specificity (0.9979) and AUC (0.9898). In addition, it classified the emotions correctly in the images of the elderly

- following seven emotion classes: Anger (4.953), Disgust (547), Fear (5.121), Happiness (8.989), Sad (6.077), Surprised (4.002) and Neutral (6.198).
2. **Chicago Face Database (CFD):** The CFD (Ma et al. 2015) database is a high-resolution, demographically diverse facial expression image dataset. The CFD has 1,204 images and covers five classes of emotions, distributed as follows: Neutral (594), Happy with open mouth (161), Happy with closed mouth (146), Anger (154) and Surprised (149). Since publication of the CFD base in 2015, two extensions have been released. The first was the Multiracial Chicago Face Database (CFD-MR) (Ma et al. 2021), consisting of 88 images, covering only the neutral emotion class. The second was the Chicago Face Database–India (CFD-INDIA) (Lakshmi et al. 2021), consisting of 142 images referring to the neutral emotion class. In order to carry out this research, we chose to join the CFD-MR and CFD-INDIA bases to the main base, which is the CFD. In addition, we also merged the classes “Happy with mouth open” and “Happy with mouth closed”, thus constituting the class of “Happy” and forming only four classes. After the organization of this database, it was composed of 1.434 images.
 3. **Karolinska Directed Emotional Faces (KDEF):** The KDEF (Lundqvist et al. 1998) database aims to provide standardized and quality material for research that focuses on emotion recognition. KDEF is composed of 4.900 images of emotional facial expressions and covers seven classes of emotions, which are: Neutral, Happy, Anger, Fear, Disgust, Sad and Surprised. In all, each facial expression was captured from 5 different angles (full left profile, half left profile, straight, half right profile, full right profile). It is important to clarify that the KDEF base is fully balanced, having 700 images for each emotion class.
 4. **Yale Face Database:** The Yale (Belhumeur et al. 1997) base is composed of 165 images in grayscale, covering four classes of emotions, which are: Neutral, Sad, Surprised and Happy, the rest of the categories refer to different configurations of accessories or gestures facials. The distribution of images by emotion class is as follows: Happy (15), Neutral (118), Surprised (15) and Sad (15). Two images from the neutral class were in GIF format, for this reason they were excluded and the final number of the base was 163 images.

As you can see in Fig. 1, the database we call “complete” is composed of the combination of the FER-2013, CFD, KDEF and Yale Face databases. The idea of building the complete base was precisely to take advantage of the advantages and contributions of each base for the development of a more robust one. For example, the CFD database has

high resolution images and participants of different races and genders, on the other hand, it does not contain diversity in relation to positions, which is the case of the KDEF database that presents images of facial expressions from five different angles, but does not consider the diversity of participants. However, both did not admit participants with accessories, so Yale Face and FER-2013 also add images that have different configurations, positions and accessories.

It is important to highlight that to carry out the experiment reported in this research, we chose not to use the Disgust emotion class, from the FER-2013 and KDEF bases, because it has few images. Knowing this, overall, 41.137 images and six classes of emotions make up the complete database, where the final file size of the database was equivalent to 5.27GB. The distribution of images by emotion classes is as follows: Happy (10.011), Sad (6.792), Neutral (7.840), Anger (5.807), Surprised (4.866) and Fear (5.821).

Experimental setup

As shown in the Fig. 2, in the first part of the experiment, the complete database went through the pre-processing step, where all images were converted to grayscale and resized to 48x48 pixels. This resizing was carried out because smaller images require less processing power and memory, which allows the model to be trained more quickly and with fewer computational resources. In this experiment, the attributes used are the list of pixels that build the images. Still in the pre-processing stage, it was necessary to transform the list of pixels into an array, for which we used the numpy library (Oliphant et al. 2006). Subsequently, the pixels were on a scale of 0 to 255. For better processing of our classification model, we normalized the pixels to a scale of 0 to 1.

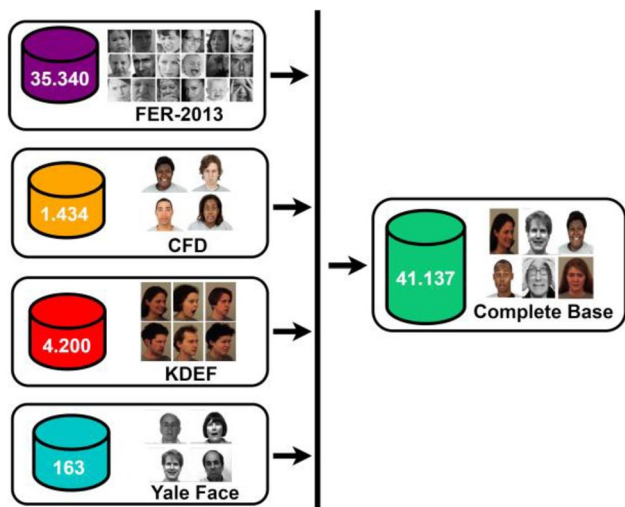


Fig. 1 Base junction process and general quantitative

For the classification stage, a Convolutional Neural Network was used. We chose this type of network because it is the most used model in the literature to work with images (Albawi et al. 2017). More details about the architecture of said CNN can be viewed in "Structure of CNN" section.

To train the model, we split the data into 70% for training/validation and 30% for testing. At the end of each epoch, we validate the model in real time. We only use the test set after the model has been fully trained. To evaluate the performance of the model, we used five metrics (described in Metrics section), which are: Accuracy, Kappa Index, Sensitivity, Specificity and area under the ROC curve (AUC).

After the training stage, in order to explore the performance of our model in the context of the elderly, we carried out part 2 of the experiment. In this step, we used static images of elderly people chosen randomly and without copyright on Pixabay (Pixabay 2023), Pexels (Pexels 2023) and Unsplash (Unsplash 2023) platforms. After obtaining the test images, we apply Haarcascade Frontal Face (Viola and Jones 2001) to detect the face in the images. Subsequently, we use our CNN classifier to predict emotions. It is noteworthy that to carry out this experiment we used the Python programming language and important libraries such as Keras and TensorFlow (Géron 2022).

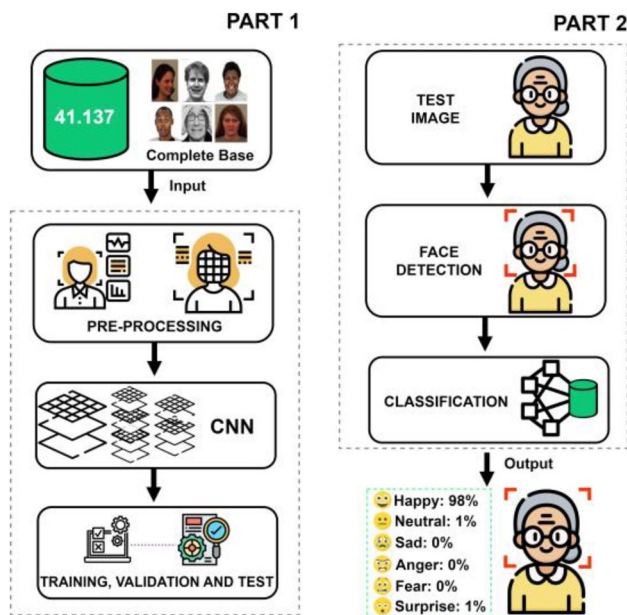


Fig. 2 In part 1 of the experiment, pre-processing was performed on the complete database. Subsequently, our CNN model went through the training/validation and testing stage. In part 2, our model together with frontal face haarcascade was applied to static images of elderly people for emotion recognition

Structure of CNN

In deep learning, CNNs are the most common networks used for image classification. The CNN architecture used in this work can be seen in Fig. 3. In summary, in all convolutional layers of our model, we use the 2D convolution operation (Conv2D). The difference between layers was the number of filters we defined. For example, on layer 1 and 2, we apply 64 filters. For layers 3 and 4, 128 filters were used. On layers 5 and 6, we define 256 filters. The 3x3 kernel size has been applied to all Conv2D layers, as well as the Relu activation function. After two layers of 2D convolution, we apply Batch Normalization. The Pooling layer was applied to reduce the input data size (resource map) and regularize the network. This helps to reduce memory cost and improve processing. For this, we used the Max Pooling technique with a pool size of 2x2 and Strides of 2x2. To avoid overfitting, we apply a dropout of 0.25 after two layers. We also use a Flatten layer to transform the matrix resulting from the other convolution layers into a linear, one-dimensional (1D) matrix. A dense layer with 512 units and the last dense layer, together with a softmax activation function, with six 6 units (equivalent to the six emotion classes: Anger, Fear, Happy, Sad, Surprised and Neutral.), were defined. In addition to what was described, the batch size number was equivalent to 16 and the number of epochs was 100.

Metrics

The performance of our CNN model was evaluated based on the metrics of accuracy, kappa index, sensitivity, specificity and area under the ROC curve. Accuracy indicates the general average of hits that the classifier obtained in relation to the total number of predictions. That is, it is the probability equivalent to the rates of true positives (True Positive - TP) and true negatives (True Negative - TN) among all results.

The Kappa statistic indicates how much the data agree with the classification performed through variables that are qualitative. Sensitivity (also known as Recall) is the rate of true positives (TPR) and seeks to assess the ability of the algorithm to successfully detect results classified as positive. Specificity (known as the True Negative Rate - TNR) aims to evaluate the performance in identifying true negatives. The area under the ROC curve (AUC) seeks to identify how good the created classifier is to distinguish between relation to classes. The Table 2 presents the metrics used along with their respective mathematical expressions.

In the Table 2, ρ_o is the observed agreement rate, also called accuracy. The expected concordance rate is defined by ρ_e in Equation (1) below.

$$\rho_e = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + FN + TN)2} \tag{1}$$

Fig. 3 Details about the CNN architecture used to carry out this experiment

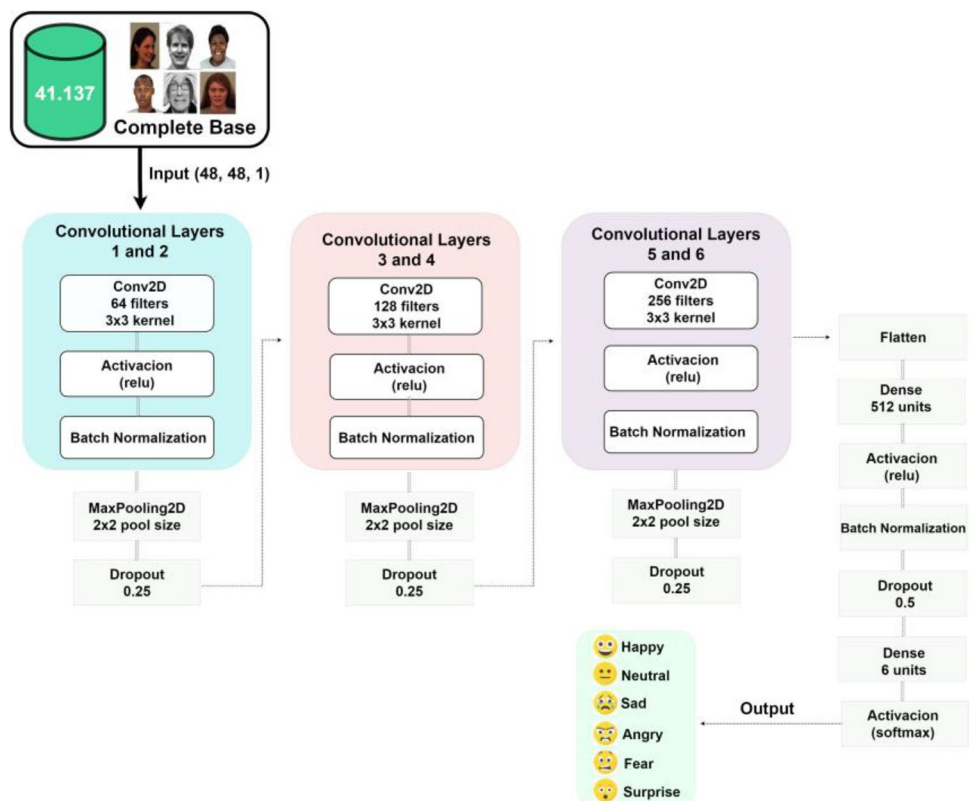


Table 2 Metrics used to evaluate the classifiers along with their respective mathematical expressions

Metric	Mathematical Expression
Accuracy	$\rho o = \frac{TP+TN}{TP+TN+FP+FN}$
Kappa	$\kappa = \frac{(\rho o - p e)}{(1 - p e)}$
Sensitivity	$\rho TP = TPR = \frac{TP}{TP+FN}$
Specificity	$\rho TN = TNR = \frac{TN}{TN+FP}$
AUC	$AUC = \int_0^1 \rho TP d \rho TN$

Table 3 Result obtained by the model during the training and validation stage

Metric	Result
Accuracy	90.77%
Kappa Index	0.8873
Sensitivity	0.9872
Specificity	0.9979
AUC	0.9898

The metrics used to evaluate the proposed model were chosen based on its wide acceptance and use in the state of the art in emotion recognition and in healthcare. These metrics not only facilitate comparison with previous work, but also provide a comprehensive and accurate assessment of model performance.

Results

As shown earlier in the Fig. 2, in part 1 of the experiment the CNN model was trained/validated. The results obtained in this step are presented in the Table 3. It is worth noting that, as this is an exploratory study, this result refers to only one round of the experiment (which is different from the number of epochs). As you can see, the CNN model we propose obtained good results in terms of accuracy (90.77%), kappa index (0.8873), sensitivity (0.9872), specificity (0.9979) and AUC (0.9898).

The results obtained during the test step can be seen in the Table 4. Regarding accuracy (69.05%) and kappa (0.6725), the performance clearly declined. However, the proposed model maintained good results regarding sensitivity (0.8882), specificity (0.9823) and AUC (0.8981).

For a better analysis and understanding of the result obtained, the Fig. 4 shows the confusion matrix, generated from the prediction of the test data set (corresponding to 12.341 images). The X (horizontal) axis of the matrix represents predicted emotions and the Y (vertical) axis represents the correct classification of emotions.

Table 4 Result obtained by the model during the test stage

Metric	Result
Accuracy	69.05%
Kappa Index	0.6725
Sensitivity	0.8882
Specificity	0.9823
AUC	0.8981

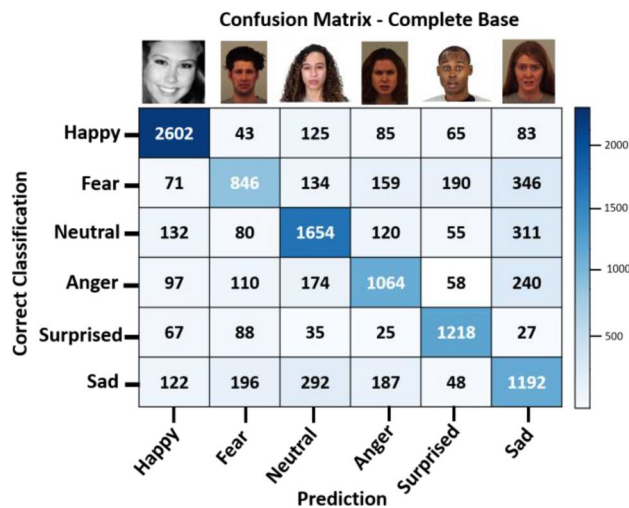


Fig. 4 Result referring to the test stage presented through the Confusion Matrix

Overall, of the 12.341 images that make up the test set, the model was able to correctly classify 8.576 images into their respective emotion classes, and 3.765 were classified incorrectly. It is noticed that the Happy, Sad and Surprised class were the ones that obtained the most correct classifications. It is believed that this occurs because both are majority classes and consequently are well represented in the database in relation to the other classes. With regard to the Fear, Neutral and Anger classes (seen as minority classes), there is a significant amount of misclassifications, where the model is erroneously classifying the images in the sad class.

After model training/validation and testing, we performed the second part of the experiment (see Fig. 2). We used our model trained with the complete database to perform the classification of emotions in static images of elderly people. The results obtained reinforce what was exposed in the confusion matrix and can be seen in the Fig. 5.

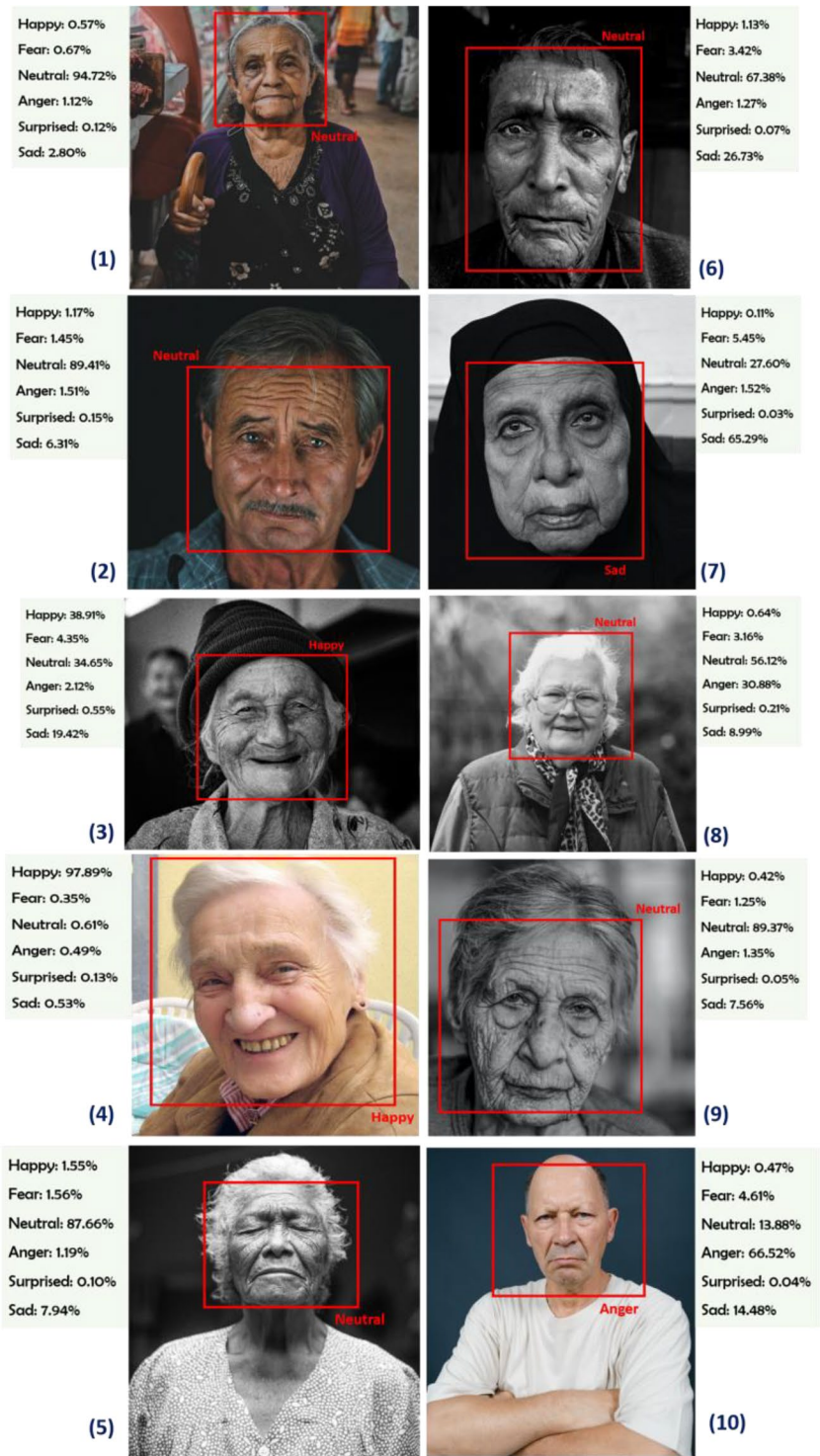
The model was able to classify the images well in the Happy classes, despite having a low probability of also classifying them in the Neutral class. This fact can be identified mainly in the subimages (3), (4) and (8) of the Fig. 5. In addition, it is clear that the model was able to better classify the images that belong to the neutral class (see subimages

(1), (2), (5), (6), (7) and (9)) and, despite that there is a probability of images of this class being confused with the sad class, is minimal. The model was able to correctly classify the subimage (10) in the anger class, where the probability of this image being confused with neutral and sad also exists.

As can be seen in the Fig. 6, during the emotion recognition stage in the images of the elderly, there were errors in face detection and classification of emotions.

It is visible that face detection errors are concentrated in regions where there is a significant amount of wrinkles and folds on the face of the elderly. So, although our model obtained good results, errors in both face detection and

Fig. 5 Results of emotion recognition in static images of elderly people



emotion classification are expected, since we do not work with a specific database of elderly people that considers all the limitations of this public.

Discussion

Upon examining the cited literature within this study, particularly in the "Related works" section, it becomes apparent that contending with the FER-2013 database presents formidable obstacles owing to multiple factors, encompassing variability, noise, and image ambiguity, along with their corresponding annotations. Noteworthy endeavors utilizing this database for model training have yielded accuracies below 70% (Bodapati et al. 2022; Khopkar and Adholiya 2021; Agrawal and Mittal 2020; Zahara et al. 2020) in some instances, whereas others have attained accuracies ranging from 72% to 75% (Zahara et al. 2020; John et al. 2020; Mounq et al. 2022; Ab Wahab et al. 2021). It is worth acknowledging that, despite incorporating the FER-2013 dataset as an integral component of our comprehensive training database, our model achieved an impressive accuracy of 90.77%. Nonetheless, it is crucial to recognize that the inherent challenges posed by the FER-2013 database may have exerted an influence on our model's performance during testing, culminating in an accuracy of 69.05%.

Studies employing conventional CNN approaches (Sadeghi and Raie 2022; Borgalli and Surve 2023; Borgalli and Surve 2022; John et al. 2020; Lakshmi et al. 2021) or hybrid models (Torcate et al. 2023; Hosgurmath et al. 2022; Ab Wahab et al. 2021; Ruiz-Garcia et al. 2018) for emotion recognition have demonstrated promising outcomes, albeit

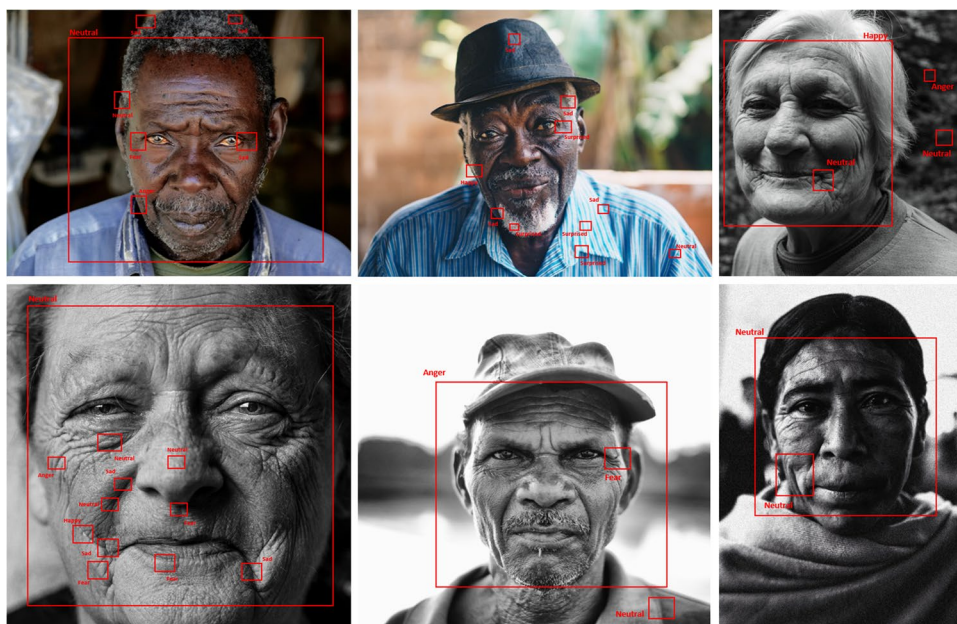
employing databases of limited scale. It is pertinent to highlight that these investigations solely trained their models on individual databases in isolation. Conversely, the novel CNN architecture we have devised exhibits the capacity to effectively address the heterogeneity and peculiarity inherent in each utilized database, thereby facilitating the construction of a more resilient and robust dataset.

Other recent studies (Gahlan and Sethia 2024; Almulla 2024) used deep convolution neural networks and demonstrated good results, between 69% and 81.1%. Potential approaches using innovative Vision Transformer techniques were also used in works (Maghari and Telbani 2024; Zakiedin et al. 2024) and managed to achieve significant results in different databases considered small, with accuracy varying between 70.14% and 99.2%.

The limited body of research (Grondhuis et al. 2021; Lopes et al. 2018) concentrating on emotion recognition in the elderly has revealed the impact of aging on facial expression recognition tasks. The findings indicate that the decline in expressive muscle capacity and age-related physical changes in the face collectively contribute to the challenges in recognizing emotional expressions from older faces.

Our approach of constructing and utilizing a comprehensive database for training our emotion classification model on static images of the elderly yielded favorable outcomes, with the model accurately classifying emotions in the elderly subjects' images. However, we acknowledge that this approach has its limitations concerning this specific population. While commendable results have been attained, we anticipate that the potential errors might be magnified when testing this initial application in the context of real therapies for the elderly. This anticipation stems from considering

Fig. 6 Face detection errors and classification of emotions in images of elderly people



the challenges posed by impairments in the brain and facial structures crucial for emotion processing and expression in this population. It is plausible that the classifier could encounter heightened difficulties in accurately recognizing emotions, along with the face detector's ability to identify faces with precision.

Conclusion

It is undeniable the importance of developing emotion recognition systems for the elderly public, thinking about specific applications for therapies, in order to personalize them and make them increasingly assertive. In order to contribute in this context, we decided to take the first step. In this exploratory study, we present a potential model for recognizing emotions through facial expressions. Our model achieved an accuracy of 90.77% in the training/validation stage, using a database considered complete, composed of the merging and merging of the FER-2013 base, KDEF, Chicago Face Database, Yale Face. Despite the simplicity and traditional nature of our approach, wherein we exclusively employ image pixels as attributes, we effectively harness the inherent variability and peculiarities within each database to optimize our model's performance.

Regarding the testing stage, the result of our model in relation to accuracy decreased to 69.05%, in line with the literature on generalization challenges when using the FER-2013 dataset, which was one of the databases used to construct the complete base. Despite this, the sensitivity and specificity of our model remained high, with 0.8882 and 0.9823 respectively, indicating reliable detection of true positives and negatives. We highlight that despite the challenges inherent in recognizing facial emotions in the elderly due to factors such as wrinkles and folds, the proposed model presented promising results when dealing with a diverse and heterogeneous database. Given this, we believe that the present study contributes significantly to affective computing and personalized therapy, providing a basis for future advances in automatic emotion recognition systems in the elderly.

Despite the success of our model in accurately classifying emotions in static images of the elderly, it is imperative to underscore that its application in the intended motivational context of this research may yield divergent outcomes. This discrepancy arises due to the model's development without accounting for the specific limitations inherent in this particular demographic, attributed to the unavailability of appropriate data. As we reflect on the initial motivation driving this work, we must also confront the encountered challenges and identified gaps during its execution, culminating in the delineation of promising avenues for future

research. These include: i) Constructing a dedicated database of emotional facial expressions exclusively focusing on the elderly population; ii) Evaluating deep and hybrid architectures to enhance model performance; iii) Exploring additional pre-processing techniques for data balancing and feature selection; and iv) Devising and implementing an Artificial Neural Network architecture capable of real-time emotion recognition in the elderly population.

Author contributions A. T.: Conceptualization, methodology, investigation, writing.

M. S.: Conceptualization, methodology, investigation, writing.

W. P.: Conceptualization, methodology, investigation, writing, scientific supervision.

Funding This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES)—Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico—Brazil (CNPq).

Data availability Not applicable.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

- Ab Wahab MN, Nazir A, Ren ATZ, Noor MHM, Akbar MF, Mohamed ASA. Efficientnet-lite and hybrid cnn-knn implementation for facial expression recognition on raspberry pi. *IEEE Access*. 2021;9:134065–80.
- Abdullah SMSA, Ameen SYA, Sadeeq MA, Zeebaree S. Multimodal emotion recognition using deep learning. *J Appl Sci Technol Trends*. 2021;2(02):52–8.
- Agrawal A, Mittal N. Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *Vis Comput*. 2020;36(2):405–12.
- Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 International conference on engineering and technology (ICET); 2017. pp 1–6. Ieee.
- Almulla MA. A multimodal emotion recognition system using deep convolutional neural networks. *J Eng Res*. 2024.
- Alreshidi A, Ullah M. Facial emotion recognition using hybrid features. In: *Informatics*. 2020;7:6. MDPI.
- Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell*. 1997;19(7):711–20.
- Boateng G, Kowatsch T. Speech emotion recognition among elderly individuals using multimodal fusion and transfer learning. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. 2020. pp 12–16.
- Bodapati JD, Srilakshmi U, Veeranjanyulu N. Fernet: a deep cnn architecture for facial expression recognition in the wild. *J Inst Eng (India): Series B*. 2022;103(2):439–448.
- Bomfim AJDL, Ribeiro RADS, Chagas MHN. Recognition of dynamic and static facial expressions of emotion among older adults with major depression. *Trends Psychiatry Psychother*. 2019;41:159–66.

- Borgalli MRA, Surve S. Deep learning for facial emotion recognition using custom cnn architecture. In: *Journal of Physics: Conference Series*. 2022;2236:012004. IOP Publishing.
- Castillo, J.C., Fernández-Caballero, A., Castro-González, Á., Salichs, M.A., López, M.T.: A framework for recognizing and regulating emotions in the elderly. In: *Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, December 2–5, 2014. Proceedings 6; 2014.* pp 320–327. Springer.
- de Freitas Barbosa VA, Gomes JC, de Santana MA, Albuquerque JEDA, de Souza RG, de Souza RE, dos Santos WP. Heg. ia: an intelligent system to support diagnosis of covid-19 based on blood tests. *Res Biomed Eng*. 2021;1–18.
- De Oliveira APS, De Santana MA, Andrade MKS, Gomes JC, Rodrigues MC, dos Santos WP. Early diagnosis of parkinson's disease using eeg, machine learning and partial directed coherence. *Res Biomed Eng*. 2020;36:311–31.
- de Santana MA, de Lima CL, Torcate AS, Fonseca FS, dos Santos WP. Affective computing in the context of music therapy: a systematic review. *Res Soc Dev*. 2021;10(15):392101522844–392101522844.
- de Santana MA, de FreitasBarbosa VA, de Cássia Fernandes de Lima R, dos Santos WP. Combining deep-wavelet neural networks and support- vector machines to classify breast lesions in thermography images. *Health Technol*. 2022;12(6):1183–1195.
- de Souza RG, dos Santos Lucas e Silva G, dos Santos WP, de Lima ME, Initiative ADN. Computer-aided diagnosis of alzheimer's disease by mri analysis and evolutionary computing. *Res Biomed Eng*. 2021;37:455–483.
- Espinola CW, Gomes JC, Pereira JMS, dos Santos WP. Detection of major depressive disorder using vocal acoustic analysis and machine learn- ing—an exploratory study. *Res Biomed Eng*. 2021;37:53–64.
- Ferreira CD, Torro-Alves N. Reconhecimento de emoções faciais no envel- hecimento: uma revisão sistemática. *Universitas Psychologica*. 2016;15(SPE5):1–12.
- Ferreira BLC, de MoraisFabrício D, Chagas MHN. Are facial emotion recognition tasks adequate for assessing social cognition in older people? a review of the literature. *Arch Gerontol Geriatrics*. 2021;92:104277.
- Gahlan N, Sethia D. Emotion recognition from facial expressions using deep recurrent attention network. In: *2024 16th International conference on communication systems & NETWORKS (COMSNETS); 2024.* pp 86–91. IEEE.
- Gautam C, Seeja K. Facial emotion recognition using handcrafted features and cnn. *Procedia Computer Science*. 2023;218:1295–303.
- Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media; 2022.
- Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee D-H, et al. Challenges in representation learning: A report on three machine learning contests. In: *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3–7, 2013. Proceedings, Part III 20; 2013.* pp. 117–124. Springer.
- Grondhuis SN, Jimmy A, Teague C, Brunet NM. Having difficulties read- ing the facial expression of older individuals? blame it on the facial muscles, not the wrinkles. *Front Psychol*. 2021;12:620768.
- Ho D, Quake SR, McCabe ER, Chng WJ, Chow EK, Ding X, Gelb BD, Ginsburg GS, Hassenstab J, Ho C-M, et al. Enabling technologies for personalized and precision medicine. *Trends Biotechnol*. 2020;38(5):497–518.
- Hosgurmath S, Mallappa VV, Patil NB, Petli V. A face recognition system using convolutional feature extraction with linear collabora- tive discrim- inant regression classification. *Int J Electr Comput Eng*. 2022;12(2):1468.
- Jiang L, Siriaraya P, Choi D, Kuwahara N. Emotion recognition using elec- troencephalography signals of older people for reminiscence therapy. *Front Physiol*. 2022;12:2468.
- John A, Abhishek M, Ajayan AS, Sanoop S, Kumar VR. Real-time facial emotion recognition system with improved preprocessing and feature extrac- tion. In: *2020 Third international conference on smart systems and inventive technology (ICSSIT); 2020.* pp 1328–1333. IEEE.
- Khateeb M, Anwar SM, Alnowami M. Multi-domain feature fusion for emotion classification using deap dataset. *Ieee Access*. 2021;9:12134–42.
- Khopkar A, Adholiya A. Facial expression recognition using CNN with Keras. *Biosci Biotechnol Res Commun*. 2021;14(5):47–50.
- Lakshmi A, Wittenbrink B, Correll J, Ma DS. The india face set: Inter- national and cultural boundaries impact face impres- sions and perceptions of category membership. *Front Psychol*. 2021;12:627678.
- LopesN, Silva A, Khanal SR, Reis A, Barroso J, Filipe V, Sampaio J. Facial emotion recognition in the elderly using a svm classifier. In: *2018 2nd International conference on technology and innovation in sports, health and wellbeing (TISHW); 2018.* pp 1–5. IEEE.
- Lundqvist D, Flykt A, Öhman A. *Karolinska directed emotional faces*. Cogn Emot. 1998.
- Ma DS, Correll J, Wittenbrink B. The chicago face database: A free stim- ulus set of faces and norming data. *Behav Res Methods*. 2015;47:1122–35.
- Ma DS, Kantner J, Wittenbrink B. Chicago face database: Multiracial expansion. *Behav Res Methods*. 2021;53:1289–300.
- MaK, Wang X, Yang X, Zhang M, Girard JM, Morency L-P. Elderre- act: a multimodal dataset for recognizing emotional response in aging adults. In: *2019 International conference on multimodal interaction*. 2019. pp 349–357.
- Maghari A, Telbani A. Masked facial emotion recognition using vision transformer. Available at SSRN 4831518. 2024.
- Motadi L, Mabongo M, Demetriou D, Mathebela P, Dlamini Z. Ai as a novel approach for exploring ccfnas in personalized clinical diagnosis and prognosis: Providing insight into the decision-making in precision oncology. In: *Artificial Intelligence and Precision Oncology: Bridging Cancer Research and Clinical Decision Support*, pp. 73–91. Springer, ???, 2023.
- Moung EG, Wooi CC, Sufian MM, On CK, Dargham JA. Ensemble-based face expression recognition approach for image sentiment analysis. *Int J Electr Comput Eng*. 2022;12(3):2588–600.
- Nunes IB, de Santana MA, Gomes JC, Torcate AS, Charron N, de Brito NC, Moreno GMM, de Gusmão CMG, dos Santos WP. Music recommendation systems to support music therapy in patients with dementia: an exploratory study. *Res Biomed Eng*. 2023;39(3):1–11.
- Oliphant TE. In: *Guide to NumPy, vol. 1.* USA: Trelgol Publishing; 2006.
- Pan J, Fang W, Zhang Z, Chen B, Zhang Z, Wang S. Multimodal emotion recognition based on facial expressions, speech, and EEG. *IEEE Open J Eng Med Biol*. 2023.
- Pexels. *As Melhores Fotos Profissionais Gratuitas e Imagens e Vídeos Livres de Royalties Que Os Criadores Compartilharam*. 2023. Last accessed 1 June 2023. <https://www.pexels.com/pt-br/>.
- Pixabay. *Use O Nosso Banco de Imagens Royalty Free. Mais de 1 Milhão de Imagens, Fotos e Vídeos em Alta Qualidade Para Seus Projetos*. 2023. Last accessed 1 June 2023. <https://pixabay.com/pt/>.
- Ruiz-Garcia A, Elshaw M, Altahhan A, Palade V. A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots. *Neural Comput Appl*. 2018;29:359–73.

- Sadeghi H, Raie A-A. Histnet: Histogram-based convolutional neural net- work with chi-squared deep metric learning for facial expression recognition. *Inf Sci.* 2022;608:472–88.
- Santana MAD, Pereira JMS, Silva FLD, Lima NMD, Sousa FND, Arruda GMSD, Lima RDCFD, Silva WWAD, Santos WPD. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Res Biomed Eng.* 2018;34:45–53.
- Shirahige L, Leimig B, Baltar A, Bezerra A, de Brito CVF, do Nascimento YSO, Gomes JC, Teo, W-P, Dos Santos WP, Cairrão M, et al. Classification of parkinson's disease motor phenotype: a machine learning approach. *J Neural Transm.* 2022;129(12):1447–146.
- Teh EJ, Yap MJ, Rickard Liow SJ. Emotional processing in autism spec- trum disorders: Effects of age, emotional valence, and social engagement on emotional language use. *J Autism Dev Disord.* 2018;48:4138–54.
- Torcate AS, De Santana MA, Dos Santos WP. Emotion recognition to support personalized therapy: An approach based on a hybrid architecture of cnn and random forest. In: 2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI). 2023. pp. 1–6. IEEE.
- Unsplash. Unsplash. Fonte de Recursos Visuais da Internet. Fornecidos Por Cri- adores de Todo O Mundo. 2023. Last accessed 1 June 2023. <https://unsplash.com/pt-br>
- Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simplefeatures. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. 2001. vol 1. Ieee.
- Zahara L, Musa P, Wibowo EP, Karim I, Musa SB. The facial emotion recognition (fer-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (cnn) algorithm based raspberry pi. In: 2020 Fifth International Conference on Informatics and Computing (ICIC); 2020. pp 1–9. IEEE.
- Zakiieldin K, Khattab R, Ibrahim E, Arafat E, Ahmed N, Hemayed E. Vitcn: Hybrid vision transformer with temporal convolution for multi-emotion recognition. *Int J Comput Intell Syst.* 2024;17(1):64.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.