Computers in Human Behavior 174 (2026) 108806

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Facial emotion recognition from feature loss media: Human versus machine learning algorithms

Diwakar Y. Dube [a], Mathy Vandhana Sannasi [b,*], Markos Kyritsis [a], Stephen R. Gulliver [a]

[a] Henley Business School, Digital, Marketing, and Entrepreneurship, University of Reading, RG6 6UD, United Kingdom
[b] School of Business and Management, Royal Holloway University of London, Egham, Surrey TW20 0EX, United Kingdom

ARTICLE INFO

Handling Editor: Dr. Bjorn de Koning

Keywords:
Artificial emotion recognition (AER)
Convolutional neural networks (CNN)
Machine learning (ML)
Feature extraction
Singular vectors
Low resolution images

ABSTRACT

The automatic identification of human emotion, from low-resolution cameras is important for remote monitoring, interactive software, pro-active marketing, and dynamic customer experience management. Even though facial identification and emotion classification are active fields of research, no studies, to the best of our knowledge, have compared the performance of humans and Machine Learning Algorithms (MLAs) when classifying facial emotions from media suffering from systematic feature loss. In this study, we used singular value decomposition to systematically reduce the number of features contained within facial emotion images. Human participants were then asked to identify the facial emotion contained within the onscreen images, where image granularity was varied in a stepwise manner (from low to high). By clicking a button, participants added feature vectors until they were confident that they could categorise the emotion. The results of the human performance trials were compared against those of a Convolutional Neural Network (CNN), which classified facial emotions from the same media images. Findings showed that human participants were able to cope with significantly greater levels of granularity, achieving 85 % accuracy with only three singular image vectors. Humans were also more rapid when classifying happy faces. CNNs are as accurate as humans when given mid- and high-resolution images; with 80 % accuracy at twelve singular image vectors or above. The authors believe that this comparison concerning the differences and limitations of human and MLAs is critical to (i) the effective use of CNN with lower-resolution video, and (ii) the development of useable facial recognition heuristics.

1. Introduction

70 % of all human communication is non-verbal (Hull, 2016; Jaiswal & Nandi, 2020), and therefore consideration of emotions is critical to understanding human cognition and intention (Kim & McGill, 2025; Li et al., 2024). Human faces have key features (i.e., eyes, nose, eyebrows, and mouth) that can be identified (McKone & Robbins, 2011, pp. 149–176), and facial recognition can recognise the presence of a specific human face (Melinte & Vladareanu, 2020); facilitating faster and more secure authentication (e.g. multi-factor authentication), personalisation of service provision, marketing and analytics, and enhanced surveillance (Tekkök et al., 2021). Furthermore, the analysis of the relative position and shape of features, can potentially be used to classify human expressions, thus facilitating Facial Emotion Recognition (FER)(Ye & Kovashka, 2021). FER focuses on analysing human sentiment by processing images or video inputs, and is the first step towards the realm of 'affective computing' (Andalibi & Buss, 2020). FER has been incorporated in a plethora of disciplines including psychology (Banskota et al., 2023), linguistics (Lau et al., 2022), computer science (Wang et al., 2022), anthropology (Park et al., 2022), artificial intelligence application (Wright, 2023), and Human-Computer Interaction (AlEisa et al., 2023). Despite common use, FER file storage (e.g. large image files), and transmission and computation load have been identified as practical issues for artificial intelligence FER applications, specifically in cases of smart devices, edge locations and wearable devices (Molas & Nowak, 2021; Sabry et al., 2022). Various image compression techniques have been used to mitigate concerns increasing memory and computational processing requirements (Dantas et al., 2024), however such compression techniques suffer from algorithmic fairness (Stoychev & Gunes, 2022) and model performance (Pascual et al., 2022)

Research shows that humans are able to identify emotional expressions within milliseconds (200 ms) (Derntl et al., 2009), which indicates that there are certain human facial features play a pivotal role in the heuristic processing and classification of emotion classification (Maratos

* Corresponding author.
E-mail address: mathy.sannasi@rhul.ac.uk (M.V. Sannasi).

https://doi.org/10.1016/j.chb.2025.108806
Received 2 May 2025; Received in revised form 18 September 2025; Accepted 21 September 2025
Available online 22 September 2025

et al., 2009). For Paul Ekman's basic (or universal) emotion types, the occurrence of an invariant set of facial configurations was widely accepted amongst the scientific community (Barrett et al., 2019) - and emotional expressions can be identified by humans despite information loss (Smith & Rossit, 2018). Information loss can be simulated experimentally using matrix factorisation techniques, such as Singular value decomposition - where the image matrix can be deconstructed and reconstructed using singular vectors and values - and different levels of information loss (or henceforth referred to as granularity) can be generated by incrementally reducing the number of singular vectors (and values) used to generate the image. Hence this study uses a novel approach of systematically comparing human and artificial neural networks' performance across varying levels of image granularity. This will enable us to comprehend the amount of information required by humans and neural networks and inform researchers and practitioners on the least amount of information required by the neural networks to mimic human performance in an FER task. Moreover, this study will enhance the current understanding concerning the impact of low-resolution on emotion recognition - especially in cases of real-world application surveillance, telehealth, or mobile communication (Lo et al., 2023; Menaka & Yogameena, 2021) - thus enabling algorithmic fairness (Stoychev & Gunes, 2022).

1.1. Aims and objectives

This study aims to explore human emotion detection capabilities using facial images that have been systematically degraded using a matrix factorisation technique (i.e., singular value decomposition). An online experiment was conducted to find the participants' emotion detection success rates amongst different emotions, whilst determining the level of feature loss in the image (also called the level of granularity) that mediated the success of their classifications. The results of this experimentation were then compared to the performance of a standard convolutional neural network, to understand and critically analyse the differences amongst both these systems.

Objectives:

- To review and analyse the relevant literature to understand and evidence factors that play an imperative role in human emotion recognition
- To identify and confirm factors that influence the accuracy of emotion classification by humans amidst systematic feature loss
- Explore how the machines can perform under systematic feature loss
- Compare human and machine emotion recognition performance when subjected to systematic feature loss

2. Introducing Human and MLA FER capabilities

2.1. The Human response to emotional faces

The importance of non-verbal cues, such as facial expressions in human social interaction, plays a major contribution to human social intelligence and human intellectual capabilities. Detecting and processing human emotion is highly important, therefore, for survival (Adolphs, 2008), social behaviour, and communication (Bayle et al., 2011). Identifying facial emotions provides humans with information concerning their context and surroundings. For example, the presence of fear in someone's face implies the existence of perceived danger (Bayle et al., 2011), which in turn serves as a stimulus for behavioural adaptation (Gratch & Marsella, 2013).

The ability to understand emotional expressions from facial cues initiates at an early age, and human infants can differentiate between anger and happiness at the age of four months. FER is fully developed by the age of eleven and exists as a core part of human social competence (Chronaki et al., 2015). Humans do not, however, appear to use all parts of a face to support emotional recognition (Holland et al., 2019; Maratos et al., 2009). In fact, humans unconsciously ignore redundant information, i.e. focusing on critical features such as a smile in case of happiness, and frown in case of sorrow (Grauman & Leibe, 2011; Jacintha et al., 2019), allowing more efficient identification of emotional expressions despite considerable information loss (Smith & Rossit, 2018).

Scientific consideration of facial expressions has resulted in an ontology of "basic emotions" (Russell & Fernandez-Dols, 1997). Plutchik (1982) proposed Plutchik's emotional model, which, using eight neighbouring emotions (i.e., joy, trust, fear, surprise, sadness, disgust, anger, and anticipation) supported the identification of up to 48 different 'mixed' emotional expressions. Ekman (1999), however, stated that basic human emotions need to be universal in nature, regardless of culture. Six basic emotions (i.e., disgust, fear, joy, surprise, sadness, and anger) were identified as being consistent in all cross-cultural studies (Bavelas & Chovil, 1997); with trust and anticipation regarded as cognitive or evaluative processes related to expectations and beliefs.

Gratch and Marsella (2013) claimed that individuals do not have to be consciously aware of the stimuli in order to form a response to surrounding objects and/or people. The human amygdala, which is critical to human threat or reward stimuli response, receives visual input from the eyes, via a rapid subcortical pathway, which conveys 'coarse' information (using low-spatial frequencies). If participants are, for a few milliseconds, shown a picture of emotion filled faces – particularly fearful faces – a measurable physiological response occurs (Vlastos et al., 2020). Although participants are not consciously aware of the existence of these fearful images, the existence of the sub-cortical visual pathways helps humans to react unconsciously to pre-attentive stimuli (Lundqvist et al., 2014). This unconscious response can impact visual saccades (McSorley and Van Reekum, 2013), can prompt attention towards threat-relevant stimuli (Öhman et al., 2001); and even causes activation of the amygdala and cortical areas of the brain; supporting threat stimuli responses (Maratos et al., 2009).

The human visual system can detect faces and emotions seemingly effortlessly in a wide range of conditions; either considering the face as a whole and/or focusing on the presence and geometric relationship of specific features (Devue et al., 2019). This ability to focus on contextual features means that human pre-attentive processing tends to "fill in" the missing pieces of a face - even in cases of occlusion (e.g., a mask) - and can continuously process the information despite movement of the object; e.g., differences in ethnicity, gender, and/or changes to facial and head characteristics - such as the presence of a moustache, or beard etc. (Chellappa et al., 2002). This implies that identifying the best resolution containing the relevant emotion-based facial features (Ekman et al., 1980; Kohler et al., 2004; Rosenberg & Ekman, 2020) will aid in the human participants to successfully classify the emotion type. This can further inform the least amount of information required by the machine learning algorithms for best classification performance.

In summary, the human visual system identifies facial expressions very quickly and, in the case of threat-relevant facial stimuli, humans pre-attentively react to the existence of negative or unpleasant emotion states even when the stimuli have undergone considerable feature information and/or intensity loss. The question is whether machine learning algorithms, and automated emotion classification, can match human performance.

2.2. Machine learning algorithms (MLA)

Even though humans can identify facial expressions with ease, the process of using automated machine learning algorithms includes a series of actions: (i) detection of the face within the image itself; (ii) extraction of the information relevant to facial expressions; and (iii) classification of the facial expression into the relevant emotional category (Mohana & Subashini, 2024).

Multiple studies have explored the use of computer vision and/or real-time identification, and classification of basic emotional states from visual stimuli, and multiple standard methods have been used including

Principal Component Analysis (PCA) (Aggarwal et al., 2021; Thuseethan & Kuhanesan, 2016), Singular Value Decomposition (SVD) (Siam et al., 2022), and Convolutional Neural Networks (CNNs) (Ayyalasomayajula et al., 2021; Belaiche et al., 2020; Jaiswal & Nandi, 2020). CNNs, which are biomimetically inspired deep-learning models, have been found to support the highest accuracies in FER (Jaiswal & Nandi, 2020); with deep learning emerging as the state-of-the-art computer vision approach. CNNs offered performance with 99 % accuracy (topmost), which compared to an average accuracy of 96 % when using traditional methods (topmost by Support Vector Machines (SVM) (Mohana & Subashini, 2024); e.g. vector machines (SVM), K-Nearest Neighbours (KNNs), Adaboost, and Random forests. Although accurate, there are multiple issues linked to the use of deep-learning based Facial Emotion Recognition approaches; such as a requirement of large data training sets, massive computational processing power, large memory demands, and data-related computational complexities (Mohana & Subashini, 2024). Many of these issues are critically tested within the following research design, which aims to understand the threshold or minimum level of information needed by humans to determine the emotion correctly. Although there are other generalisation issues concerning pose variation, aging, illumination, partial occlusion such as masks, glasses, beards and cosmetics (Mohammed & Al-Tuwaijari, 2022; Mohana & Subashini, 2024), this study focuses on issues associated with low resolution and subsequent impacts on MLA's performance.

2.3. Hypothesis framing

The accuracy of FER systems varies significantly depending on a range of real-world factors. Previous research shows that the type of emotion being expressed impacts the accuracy of facial emotion recognition classifications (Barros et al., 2023). Barros et al., using only happy, angry and neutral images, showed that participants identify angry images faster and more accurate than happy images. Rushu et al. also showed that FER systems perform better with angry, fearful, and happy images. Accordingly, in this study, hypothesis 1 states that there is a significant effect between emotion type and successful emotional recognition. Wang et al. (2017) measured participant confidence level, i.e., when judging emotions, and showed that confidence had a significant impact on the successful recognition of emotions. Thus, our hypothesis 2 states that there is a significant relationship between participant confidence level and successful emotion recognition. Calvo and Lundqvist (2008) showed that participant reaction time differs based on the emotion type during FER experiments. Accordingly, our hypothesis 3 states that there is a significant relationship between reaction time (RT) and successful emotional recognition. The Reaction Time (RT) is the time taken by a participant to categorise the emotion displayed. Use of high-resolution media can help mitigate many confounding factors for better neural network performance, yet high resolution media is not always available. Accordingly, our hypothesis 4 states that there is a significant effect of image granularity (level of information loss) on successful facial emotion recognition. Finally, we compare human and MLA performance, when categorising facial emotions from pre-processed lossy images, to determine the level of granularity that humans and Machine Learning Algorithms (MLAs) require to successfully categorise specific emotion types.

3. Research design

Due to the need to compare human performance against Machine Learning Algorithm (MLA) performance, and ensure controlled causal inference, replicability, and researcher objectivity, a quantitative experimental research approach was selected for use in this study.

3.1. Image data Sources

Images (53 % females and 47 % male) were selected from the Warsaw Set of Emotional Facial Expression Pictures (WSEFEP), because (i) of its balanced gender representation and (ii) the fact that it contains previously validated images - having been extensively tested by multiple researchers (Hartling et al., 2021; Olszanowski et al., 2015; Vlastos et al., 2020). This set of emotional face images used in this study covered Ekman's set of basic emotional expressions - i.e., happy, anger, sad, surprise, disgust, and fear. An additional neutral (non-emotional) expression was added as a control image, specifically as it had been clearly defined using action units (AU) and validated, by the dataset providers (Olszanowski et al., 2015). This dataset is composed of emotion-filled faces and contains facial expressions of people from a white ethnic background which could be a limitation for this study.

3.2. Image pre-processing and resolution reduction

Bryt and Elad (2008) tested two forms of dimension reduction, i.e., (i) resolution reduction and (ii) Singular Value Decomposition. In the first set of images, Bryt and Elad used 8-bit grayscale ID images (358*441) to present a coarse version of images. In the second set of images, the researcher used Singular Value Decomposition (SVD) to compress the image, allowing reconstruction of low feature images with defined error thresholds. Results showed that SVD performed better than resolution reduction; i.e., offering much better space savings than a reduction of resolution (Bryt & Elad, 2008).

The images used in this study were pre-processed using the following steps: (i) Select a range of WSEFEP Images (ten of each emotion), (ii) convert the images from.jpg to.pgm format; (iii) convert the images to greyscale using the pixmap library (Bivand et al., 2025); (iv) images of all six emotion types, and neutral faces, were selected and cropped to show only the face (see Fig. 1), and the images were aligned to have the same pixels of 491 (width) * 718 (height) with 8-bit depth; (v) luminance was adjusted to ensure consistency across all images; (vi) Singular Value Decomposition (SVD) processing of the faces was then performed to generate 20 copies of the same image (using the SVD function in R studio) with stepped levels of granularity variance (i.e. from 2 (least number of singular vectors) to 20 singular image vectors - see Fig. 1 (with the most amount of retained visual information)); and (vii) the images were converted back into.png file format to facilitate use. The same set of pre-processed images (70 in total) were then used for both the online (human experiment) and machine learning algorithm (MLA) tests.

A simple dedicated online tool was developed to support the remote online experiment; using PHP, JavaScript, HTML (for the user interface), and SQL (for the database); viewable at https://stephenrgulliver.online/Faces_deploy/expLAB.php.

3.3. Experimental process

All procedures performed in the study were checked, and undertaken, in accordance with the ethical rules defined within Henley Business School (University of Reading). An online (web) experiment asked participants to classify images containing displayed emotions. All participants took part in the study voluntarily. Compensation was not offered to participants for their involvement in the experiment. Participants were sent the experimental online tool URL. As part of the experiment introduction, however, all participants were (i) asked to watch a small video demo (https://youtu.be/dmKQEU2rMN0) introducing the research, (ii) given the researchers' details, to facilitate contact if they have any questions or concerns, and (iii) informed that they could leave the experiment at any point - resulting in an incomplete data sample, which would be removed before analysis. When participants had completed the demographic questions, and had agreed to proceed with the experiment, each participant was presented with a random high loss image containing two singular image vectors (see Fig. 2a). The participant was asked to click the 'Increase Granularity' button, thus adding to the dimensional value of n, until the participant

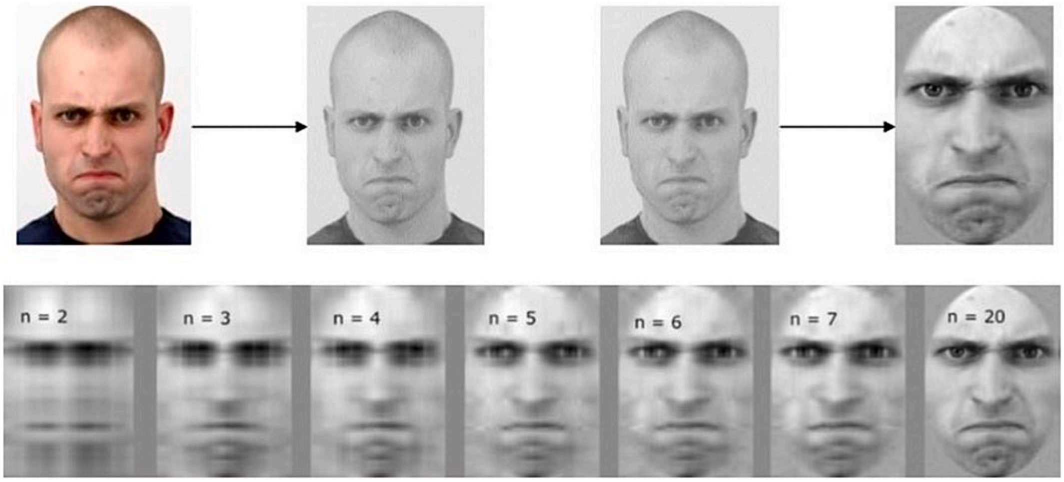

Fig. 1. Steps (iii) (iv) and (vi), i.e., Conversion to greyscale, cropping of image, and creation of images with varying granularity (n = 2 to 20) using SVD.

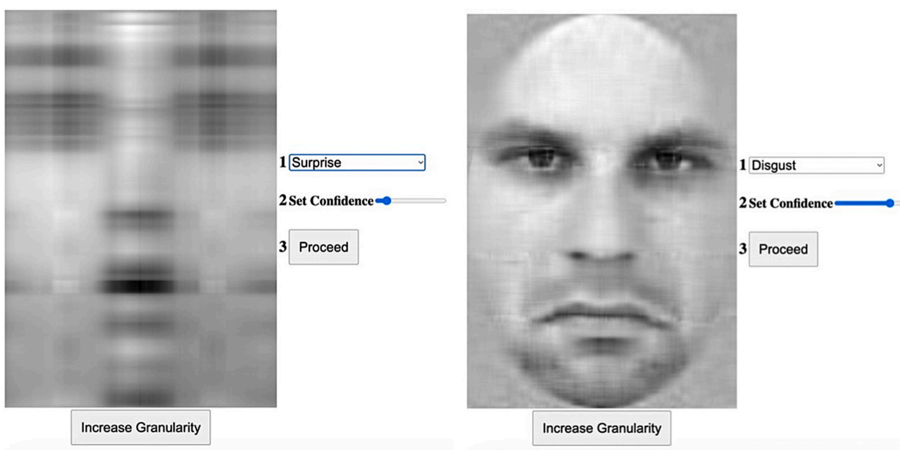

Fig. 2. Online Experimental layout a) n = 2, and b) n = 12.

believed they could effectively categorise the image - in 'data entry point one; i.e., disgust, fear, joy, surprise, sadness, anger, or neutral (see Fig. 2).

In data entry point two (see Fig. 2), the participant was asked to enter how confident they felt about the categorisation decision. The person was encouraged in the video to only increase the granularity if they were not confident that the accuracy of the categorisation was good enough; thus, allowing the researchers to better appreciate when humans use visual categorisation as the basis of decision making. Once the participant was happy with their choice, they clicked data entry point three, which allowed them to move on to the next trial. In total each experiment contained five images for each of the six emotions, and five images for neutral faces, i.e., 35 images in total. The five images chosen for each emotion category were selected randomly, and the order in which the 35 images were presented to participants was also randomized to counter any order effects. Once all 35 images had been presented, and processed by the participant, a thank you message was presented on screen, with a request to close the internet browser.

3.4. Experimental sample

In total 125 volunteers (67 female, 56 male, and 2 undisclosed; age above 18 years) participated in the web experiment. All participants used their own devices and online setup. The participants were recruited using email-based advertisements. 15 (10 female, 5 male) participants were identified as having either (i) a neuropsychiatric disorder, or (ii) had participated previously in an initial pilot experiment. Their records were omitted from the final dataset, leaving a final sample size of 110 (57 female, 51 male, 2 undisclosed). All participants stated that they had normal or corrected-to-normal vision. All participants were older than 18, undertook the experiment remotely, and took part in the study voluntarily. Electronic consent was obtained from all the participants, and participants were briefed about the experiment before proceeding to the actual experiment via an online information sheet. A confidentiality agreement was included in the briefing, which provided an overview of how data will be used and stored for a limited time. No identifiable information was collected. Due to the use of repeated measures, i.e., each participant viewing 5 different faces for each emotion type, 550 samples were collected for each emotion.

Key variables include: 1) granularity level (numeric - level 2 to 20); 2) Confidence (percentage - 0 %–100 %); 3) Emotional expressions shown on the image (neutral, happy, anger, sad, surprise, disgust, or fear); 5) outcome (Success/Failure); 6) other demographic details (Sex, Neuropsychiatric disorder, and information identifying whether participant had previously participated in similar studies).

3.5. Machine-based data capture

CNNs (Convolutional Neural Networks) use convolutional layers to learn spatial hierarchies of features (textures and edges) from the input image. After convolution, an activation function helps the network link lower order level features into complex patterns. Next, a pooling layer is used to remove non-critical features. This output is then flattened as a vector and processed by fully connected layers, identifying global patterns and relationships between the features extracted in earlier layers. CNNs are the preferred machine learning algorithm for use with FER image processing (Jaiswal & Nandi, 2020). Convolutional Neural Networks (CNN) have been used widely in FER (Ayyalasomayajula et al., 2021; Belaiche et al., 2020; Jaiswal & Nandi, 2020). Interestingly, however, the accuracy of CNN models appears to vary significantly depending on the application area, the number and type of layers, and the efficacy of the training set. For example, Jaiswal and Nandi (2020) used CNNs to detect real-time facial expressions, achieving an accuracy of 74 %. Belaiche et al. (2020) used a CNN, consisting of three convolution layers, to identify micro-expression: achieving an accuracy of 60.2 %.

Python (Using Google Collab) was used to create a Convolutional Neural Network (CNN), which was used to test and categorise all emotional images of the Warsaw set (at all levels of granularity). The input images were resized to 48x48 pixels to reduce computational cost and to simultaneously preserve facial structure as we can assume linear independence of the input features. Convolution layer one had 64 (3*3) feature maps as the input. The output was linked to a 2 × 2 max pooling layer, which extracted the maximum number of neighbouring pixels/features. A second convolution layer (using 125 5*5 feature maps as the input) was then used, followed by a second 2 × 2 max pooling layer. Third and fourth convolution layers (both with 512 3*3 feature maps as the input) were then followed by a final 2 × 2 max pooling layer. The output was then flattened using two fully connected layers (layer 1–265 neurons, and layer 2–512 neurons). The two fully connected layers were separated by a drop out layer, which nullified the contribution of some neurons in layer one moving to layer two. The CNN algorithm was trained with full resolution images (training set – contains all images in the original image dataset) and validated with images processed to different levels of granularity (test set); in addition, the algorithm used random drops to avoid overfitting.

3.6. Data validation and analysis

A python script was developed to transform the data captured from experiments into long format; as required by R-studio for analysis. Metadata for each trial sample was captured (emotion type presented, emotion type selected, confidence, granularity, demographic factors) and mapped to the unique participant ID (auto-generated by the online tool).

Normality tests for the reaction time were performed qualitatively. Both parameters had non-normal distribution, as indicated by Q-Q plots, however, after applying log transformation, the distributions became fairly normal (the datapoints followed the 45-degree reference line except for the top right corner) and thus the results are presented with caution. Transformed values were used in human experimental data analysis. Homogeneity of variances were met, according to qualitative checks using boxplots (the boxes for the different emotion types were fairly of the same size). Analysis of raw data indicated that female participants performed better than males and had 8.5 % higher accuracy or success (overall) in the classification of emotions (Female participants: Anger: 75 %, Disgust: 73 %, Fear: 62 %, Happy: 95 %, Neutral: 86 %, Sad: 71 %, Surprise: 81 %; Male participants: Anger: 63 %, Disgust: 56 %, Fear: 52 %, Happy: 92 %, Neutral: 86 %, Sad: 58 %, Surprise: 76 %.). The misclassifications followed a similar trend as that of the consolidated classification matrix presented in Table 1.

The collected data was analysed using Linear Mixed Effect Models (LMER) provided by the lme4 library (Bates et al., 2015). Code has been shared in this GitHub link. LMER models were created to support fixed and random effects, and the sjPlot package (Lüdecke, 2025) was used to get marginal R-squared (variance explained by fixed variables) and conditional R-squared (variance explained by fixed and random variables); facilitated using the tab_model() function. The plots were produced with the plot_model() function, contained in the sjPlot library (Lüdecke, 2025), and merTools function, contained in plotREsim, plotFEsim and REsim libraries (Knowles & Frederick, 2025). Post-hoc tests were produced using the glht() function from the multcomp library to identify the actual differences (Hothorn et al., 2008).

4. Results

4.1. Effect of the type of emotion on Emotion type classification

LMER modelling, using participant ID as the random variable, with success (i.e., correct classification) as the dependent variable (DV), and emotion type as the independent variable (IV) (Field et al., 2012), showed that emotion type has a significant effect on the participant's ability to successfully categorise specific emotions ($\chi2$ (6) = 317.1, $p < 0.05$). This result suggests that, with the exception of disgust and sadness emotion types, the accuracy of human emotion categorisation is significantly linked to the emotion type (see Fig. 3).

The results suggest that the type of emotion impacts the accuracy of categorisation; with happy emotion images scoring highest (see Fig. 3). The variance explained by the fixed effect, however, was small (~7 %), with the random effect explaining a larger amount of variance (total

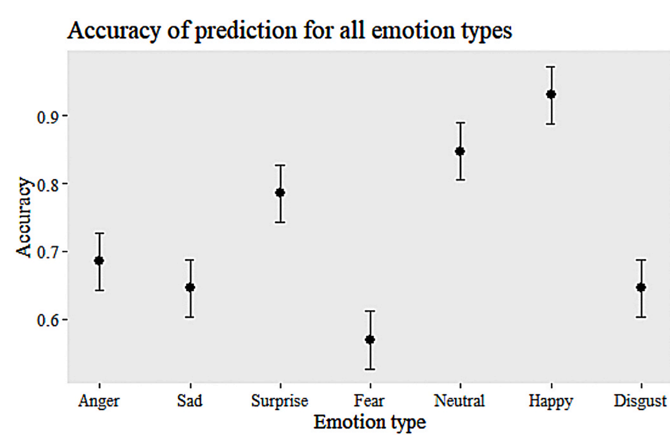

Fig. 3. Accuracy of prediction for all emotion types.

Table 1
Confusion matrix for human experimentation (online). This table depicts the classification matrix of reference vs. predictions of the online experiment.

		Predictions						
		Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
Reference	**Anger**	**69 %**	10 %	4 %	0 %	6 %	11 %	1 %
	Disgust	23 %	**65 %**	2 %	1 %	3 %	5 %	1 %
	Fear	1 %	2 %	**57 %**	1 %	1 %	4 %	35 %
	Happy	1 %	1 %	0 %	**93 %**	1 %	1 %	2 %
	Neutral	3 %	2 %	1 %	2 %	**85 %**	5 %	1 %
	Sad	3 %	5 %	4 %	1 %	22 %	**65 %**	1 %
	Surprise	1 %	2 %	13 %	2 %	2 %	2 %	**79 %**

variance of the model being 16 %). Thus, the null hypothesis has been rejected and the alternative hypothesis that emotion type has an effect on successful emotion type classification has been accepted.

A confusion matrix has been provided (Table 1) to enable further understanding and analysis into misclassifications. The confusion matrix confirms that the emotion type with the highest accuracy is happiness, which has been misclassification predominately as surprise. The emotion type with the lowest accuracy (fear) has been mostly misclassified as surprise (35 %) and sadness (5 %).

4.2. *Effect of confidence on Emotion type classifications*

Wang et al. (2017) measured confidence level as the degree of variability in judging the emotion of a presented image. LMER modelling was used with participant ID as the random variable, with confidence as the dependent variable (DV), and types of emotions as the independent variables (IV). Results showed that the type of emotion had a significant effect on classification confidence, $\chi 2\ (6) = 83.83$, $p < 0.05$. Our results suggests that participant confidence does impact categorisation accuracy; with people most confident when categorising happy faces, see Fig. 4. Thus, the null hypothesis has been rejected and the alternative hypothesis that confidence of the participants has an effect on successful emotion type classification has been accepted.

4.3. *Effect of Emotion type on reaction time (RT)*

The Reaction Time (RT) is the time taken by a participant to categorise the emotion displayed. LMER modelling was used with participant ID as the random variable, with log of reaction time (RT) as the dependent variable (DV), and type of emotion as the independent variable (IV). Results showed that type of emotion has a significant overall effect on RT, $\chi 2\ (6) = 200.59$, $p < 0.05$. Results suggest that reaction time has an effect on emotion categorisation; and that categorisation of happy faces is significantly faster than other emotion types (see Fig. 5). Thus, the null hypothesis has been rejected and the alternative hypothesis that reaction time has an effect on successful emotion type classification has been accepted.

4.4. *Effect of Emotion type on granularity*

LMER modelling was used with participant ID as the random variable, with granularity as the dependent variable (DV), and emotion type as the independent variable (IV). Results showed that emotion type has a significant effect on granularity, $\chi 2\ (6) = 430.38$, $p < 0.05$. Results suggest that classification of happy emotions required the lowest amount of granularity, followed by surprise (see Fig. 6). Thus, the null hypothesis has been rejected and the alternative hypothesis that emotion type has an effect on granularity level chosen, has been accepted.

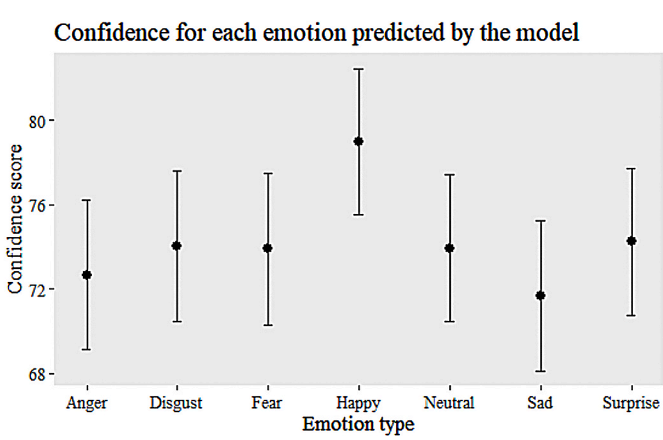

Fig. 4. Confidence for each emotion predicted by the model.

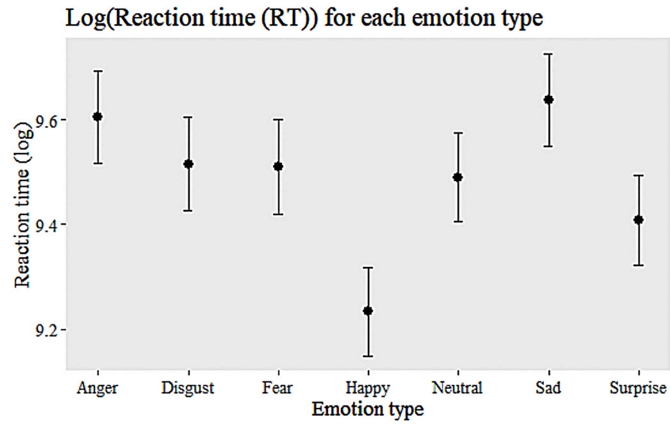

Fig. 5. Reaction time (RT) for each emotion type predicted by the model.

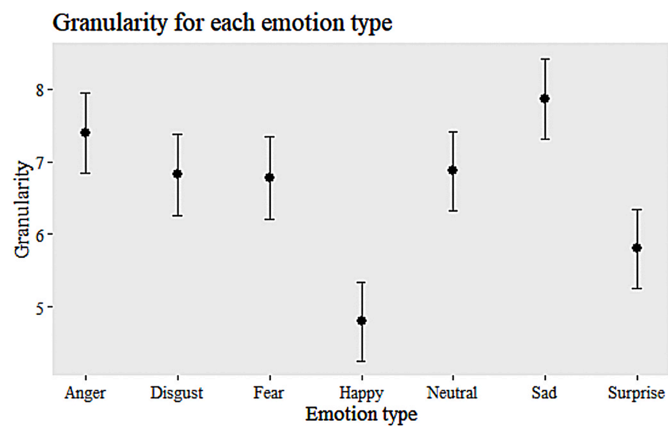

Fig. 6. Granularity level estimates for each Emotion Type.

4.5. *CNN model tests*

The CNN model was trained using all high resolution images and validated with the granularities from 2 to 20 singular image vectors (i.e., creating confusion matrices for each granularity level) (as mentioned in section 3.6). The performance of each granularity model was measured with 100 epochs, and validation accuracy was analysed with the LMER model (i.e., 100 trials for each granularity measuring validation accuracy for each epoch). The data was not split within granularities to avoid overfitting, yet a dropout rate of 25 % was included in each CNN layer to prevent overfitting. To facilitate comparison, human data for different emotion types was combined, i.e., to get the participant level accuracy for each granularity irrespective of the emotion type.

LMER modelling was used with model ID as the random variable, validation accuracy as the dependent variable (DV), and granularity as the independent variable (IV). Granularity was found to have a significant effect on the validation accuracy ($b0 = 0.46$, $t(1900) = 20.30$, $p < 0.05$); with the model performing best at granularity 20 (no information loss) ($b1 = 0.39$, $t(1900) = 12.22$, $p < 0.05$ and second best was at granularity 12 ($b2 = 0.34$, $t(1900) = 10.84$, $p < 0.05$, $R2m = 0.15$, $R2c = 0.15$) – see Fig. 8. All granularities had a significant effect on validation accuracy. The machine unsurprisingly performs best without loss (85 % accuracy), however, the second-best was at granularity 12 (with 80 % accuracy). Normalized confusion matrix for granularity 12 – i.e., the highest performing lossy CNN model - showed that surprise (at this level) achieved the highest accuracy followed by happiness, neutral, disgust, and anger (see Fig. 7). CNN models struggled consistently to

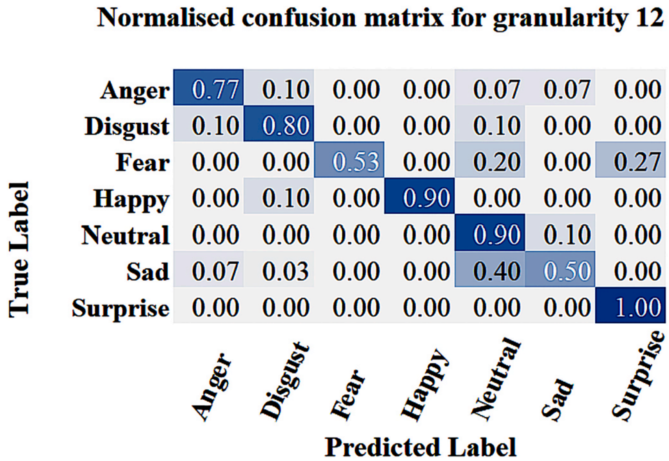

Fig. 7. Normalized confusion matrix for granularity 12.

categorise fear and sadness emotion types (see Fig. 7).

4.6. Human vs CNN trials

Our results show that with 12 singular image vectors the CNN was able to achieve an accuracy of 80 % (including fear and sadness); outperforming human subjects at the same resolution and achieving accuracies similar to those possible when provided full resolution images (see Fig. 8). However, this difference was not statistically significant, as can be seen from the overlapping intervals in the confidence interval plot (Fig. 8).

Humans can cope with information loss down to 3 singular image vectors and still correctly identify facial emotional expressions with approximately 85 % accuracy (see Fig. 8). For the machine learning algorithm to perform at 80 % accuracy at least 12 singular image vectors are required. In addition, the accuracy achieved by our CNN, for media above this resolution, was higher than that of the study by Jaiswal and Nandi (2020) in which a validation accuracy of 74 % was achieved.

5. Discussion

5.1. Considering the research hypothesis

The online experiment aimed to examine four hypothesis, i.e., whether: (i) the emotion type plays a role in the accuracy of emotion classification; (ii) the confidence of the participant determines the chances of successful classification; (iii) the reaction time impacts successful emotion classification; and (iv) humans can classify emotions accurately despite information loss. Moreover, the research aimed to determine whether current machine learning approaches outperform or underperform facial emotion classification tasks, i.e. compared to humans when presented with varying degrees of image information loss.

Our findings show that the emotion type being observed by participants (i.e., happy, sad, surprise, neutral, anger, fear, or disgust) does have a significant effect on the emotion categorisation accuracy; thus, supporting our hypothesis 1. Post-hoc tests, and qualitative analysis, from the human experiment findings, indicate that overall happy images scored with the highest accuracy and response-rate; followed by neutral, surprise, anger, disgust, and sadness. Fear had the lowest accuracy in human experiments. Our hypothesis 1 results align well with previous studies, which also identified that humans find happiness easiest to classify, and conclude that people often get confused between (i) sadness and neutral, (ii) anger and sadness, and (iii) anger and disgust (Du and Martinez, 2014). Our analysis showed that participant confidence had a significant effect on emotional face classification success; thus, supporting hypothesis 2. Results indicate a significant effect, with happiness classified with the highest confidence, and sadness classified with the lowest confidence. There were no significant differences between other emotion types. With regards to our hypothesis 3, concerning the effect of reaction time on categorisation success, our study showed that participants were much quicker at classifying happy faces (affirming hypothesis 1), and took the longest time to classify sad faces. There were no substantial differences between other emotion types. Research suggests that humans can recognise and react quickly to fearful situations, despite resolution/information reduction in peripheral vision (Smith & Rossit, 2018). Although no evidence occurred to suggest that humans can classify fearful faces quickly in the presence of image information loss, we did show that granularity has an effect of emotional face recognition overall, thus supporting our hypothesis 4. Moreover, even though granularity plays a critical role in successfully classifying emotions, we note that the degree of granularity needed by humans for

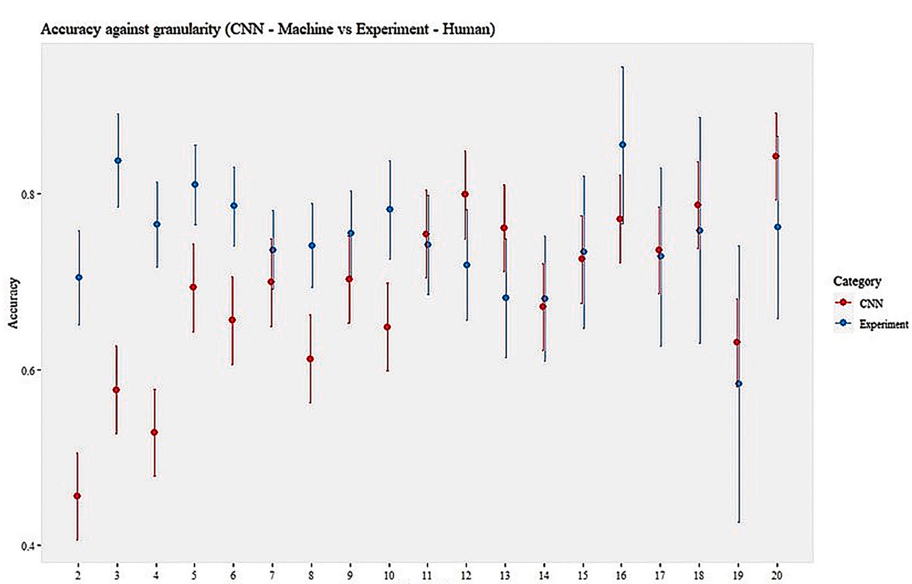

Fig. 8. Comparison between human and machine-based experiments.

decision-making differs significantly between emotion types. Participants were able to successfully categorise happy and surprised faces at low levels of granularity (see figure Fig. 6) yet required much higher granularity to effectively categorise sad faces. Our findings align with the findings of Smith and Rossit (2018), who stated that happy and surprised facial expressions are most recognized in peripheral vision.

5.2. Comparison (Human vs machine)

To the best of our knowledge no previous research has compared human and MLA FER performance across different granularity levels. Our results revealed that humans out-perform machines at lower granularity levels (see Fig. 8). CNN model validation, with 12 singular image vectors (see Fig. 7), however identified surprised faces with the highest accuracy; followed by happy, neutral, disgusted, and angry (see Fig. 8). The CNN model inaccurately categorized fear and sadness emotional faces, achieving respectively only 53 % and 50 % accuracy.

Humans achieve high levels of emotion categorisation accuracy at granularity level 3 and above. CNNs are able to achieve similar outcomes at mid granularity levels (between 11 and 17 singular image vectors), yet outperformed our participants at higher resolutions (i.e., >17 singular image vectors), although this result is not significant and cannot be generalized. This finding correlates with previous research which indicates that machine learning algorithms can predict emotions accurately when provided with full resolution information (Martínez et al., 2020).

These findings are highly relevant to practical contexts such as emotion detection in crowd scenes (particularly in cases of fear or anger) (Nan et al., 2022), patient monitoring systems using mobile and edge devices (Bisogni et al., 2022; Shaik et al., 2023), student engagement levels using mobile devices (Savchenko et al., 2022), and other applications fast decision making in a wide range of mobile intelligent models requiring lightweight neural network models (Savchenko, 2021).

5.3. Limitations

This paper does not address the issue of noisy data, such as images with background or blurry images, or faces shown in a certain angle. Lighting conditions, body posture, and overall image quality can influence accuracy and reliability of emotion recognition models (Hossain et al., 2021; Kaur & Kumar, 2024; Mahfoudi et al., 2022; Prasad & Chandana, 2023). Research shows that changes in illumination, such as variations in brightness or sharpness, and the presence of shadows, affect the system's ability to accurately identify individuals, often leading to elevated false positive or negative rates, particularly in unconstrained or real-world settings (Li et al., 2021). Similarly, differences in posture such as head tilts, non-frontal poses, or occluded facial features in real life settings, can degrade recognition performance by altering the biometric features used for identification (Jin et al., 2023). Furthermore, limited ethnic diversity within training datasets continues to produce racial bias, with algorithms demonstrating reduced accuracy and higher error rates for individuals from underrepresented groups, a trend confirmed by contemporary systematic reviews and empirical analysis on facial emotion recognition technologies (Halberstadt et al., 2022; Xu et al., 2020). As discussed earlier, this study also uses image datasets that are of actors with the same ethnicity (White European), and this might impact the generalizability of the findings. We recommend further studies to include denoising mechanisms, frontalisation and larger datasets including different ethnicities and different head posture variations, to overcome these limitations. We also suggest that the interpretability of the model is prioritised to enhance explainability of the model outcomes.

6. Conclusion and future work

The importance of Facial Emotion Recognition (FER) for society is multifaceted. The use of FER has profound implications across a range of fields, and has the potential to transform our interaction with technology; facilitating better mental health support, increased consumer engagement, enhanced public safety, etc. The automatic identification of human emotion, from low-resolution web or CCTV cameras, is therefore critical to public affective computing technologies becoming more dynamic, and more widely adopted.

This study concluded that the impact of granularity varies between different emotion types. The reaction time of the participants, confidence of the judgment, and granularity level plays a vital role in categorisation accuracy; with happy faces having the lowest reaction time and highest confidence levels at low granularity level, i.e., three singular image vectors. Provision of effective information is critical to emotion categorisation (by both humans and machines), and although humans are still better with low information media, CNNs can match human performance when presented with controlled mid and high-resolution media.

While recognising the importance of facial emotion recognition, it's essential to address associated challenges, such as privacy concerns, ethical considerations, and potential biases, to ensure responsible and ethical deployment of machine learning algorithms. The pilot study model shows good validity, yet repeated measures had to be used to ensure sample adequacy. In addition, since this is an online study, further standardizing of settings, via use of a lab-based controlled environment, is warranted. Furthermore, although the experimental sample was diverse, the CNN result implications and applications of our results are limited by the use of Warsaw Set of Emotional Facial Expression Pictures (WSEFEP), which contain only a very limited ethnic sample. Since physical and cultural diversity influences emotion expression, our current CNN models are likely to be more accurate when assessing a white European face. This is a limitation, but it is hard to avoid this when more diverse image libraries are limited. The domain demands additional capture and validation of a more comprehensive library of emotions (test images), covering a wider range of faces, colours, and ethnic backgrounds. Only when researchers have access to a full range of globally validated samples could appropriate global FER MLA tests be achieved. Making sure that FER works for all races, across all skin colour, and for a range of ethnic groups would (i) support the increased effectiveness of future FER solutions, and (ii) reduce the inherent bias in FER solution results caused by unrepresentative sample media. Furthermore, we encourage further research on the different machine learning approaches that can be used to artificially classify emotions with better accuracy and performance in this particular task.

Achieving Facial emotion recognition is important to the future of human computer interaction, security and surveillance, marketing and advertising, and personalise solutions; yet the benefit of automated machine-based FER will only occur, and lead to the development of effective artificially intelligent heuristics if we fully understand the limitations of machine based solutions. Our results showed that accuracy of CNN is widely dependant on the input dataset, the CNN layer settings, the training sample, image quality, lighting, and parameter configuration (Du et al., 2014). However, our recommendation would be to implement the model with at least 12 singular image vectors in order to maximise accuracy whilst optimising storage and computing use.

The authors believe that this comparison paper, which considers the differences and limitations of human and MLAs, presents critical findings that support (i) the effective use of CNN with lower-resolution images, and (ii) the long-term development of useable facial recognition heuristics.

CRediT authorship contribution statement

Diwakar Y. Dube: Resources, Investigation, Formal analysis, Data curation. **Mathy Vandhana Sannasi:** Writing – original draft, Validation, Project administration, Formal analysis. **Markos Kyritsis:** Writing

– review & editing, Supervision, Methodology, Conceptualization. **Stephen R. Gulliver:** Writing – review & editing, Visualization, Supervision.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NA.

Data availability

The data that has been used is confidential.

References

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current Opinion in Neurobiology, 18*(2), 166–172.
Aggarwal, A., Alshehri, M., Kumar, M., Sharma, P., Alfarraj, O., & Deep, V. (2021). Principal component analysis, hidden Markov model, and artificial neural network inspired techniques to recognize faces. *Concurrency and Computation: Practice and Experience, 33*(9), Article e6157.
AlEisa, H. N., Alrowais, F., Negm, N., Almalki, N., Khalid, M., Marzouk, R., Alnfiai, M. M., Mohammed, G. P., & Alneil, A. A. (2023). Henry gas solubility optimization with deep learning based facial emotion recognition for human computer interface. *IEEE Access, 11*, 62233–62241.
Andalibi, N., & Buss, J. (2020). The human in emotion recognition on social media: Attitudes, outcomes, risks. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.
Ayyalasomayajula, S. C., Ionescu, B., & Ionescu, D. (2021). A CNN Approach to Micro-Expressions Detection. In *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*.
Banskota, N., Alsadoon, A., Prasad, P. W. C., Dawoud, A., Rashid, T. A., & Alsadoon, O. H. (2023). A novel enhanced convolution neural network with extreme learning machine: facial emotional recognition in psychology practices. *Multimedia Tools and Applications, 82*(5), 6479–6503. https://doi.org/10.1007/s11042-022-13567-8
Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68.
Barros, F., Soares, S. C., Rocha, M., Bem-Haja, P., Silva, S., & Lundqvist, D. (2023). The angry versus happy recognition advantage: the role of emotional and physical properties. *Psychological Research, 87*(1), 108–123. https://doi.org/10.1007/s00426-022-01648-0
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01
Bavelas, J. B., & Chovil, N. (1997). 15. Faces in dialogue. *The psychology of facial expression*, 334.
Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. *PLoS One, 6*(6), Article e21584.
Belaiche, R., Liu, Y., Migniot, C., Ginhac, D., & Yang, F. (2020). Cost-Effective CNNs for Real-Time Micro-Expression Recognition. *Applied Sciences, 10*(14), 4959. https://www.mdpi.com/2076-3417/10/14/4959.
Bisogni, C., Castiglione, A., Hossain, S., Narducci, F., & Umer, S. (2022). Impact of deep learning approaches on facial expression recognition in healthcare industries. *IEEE Transactions on Industrial Informatics, 18*(8), 5619–5627.
Bivand, R., Leisch, F., Maechler, M., & Zeileis, A. (2025). pixmap: Bitmap Images/Pixel Maps. https://doi.org/10.32614/CRAN.package.pixmap.
Bryt, O., & Elad, M. (2008). Compression of facial images using the K-SVD algorithm. *Journal of Visual Communication and Image Representation, 19*(4), 270–282. https://doi.org/10.1016/j.jvcir.2008.03.001
Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40*(1), 109–115.
Chellappa, R., Wilson, C. L., & Sirohey, S. (2002). Human and machine recognition of faces: A survey. *Proceedings of the IEEE, 83*(5), 705–741.
Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33*(2), 218–236.
Dantas, P. V., Sabino da Silva Jr, W., Cordeiro, L. C., & Carvalho, C. B. (2024). A comprehensive review of model compression techniques in machine learning. *Applied Intelligence, 54*(22), 11804–11844.
Derntl, B., Seidel, E.-M., Kainz, E., & Carbon, C.-C. (2009). Recognition of emotional expressions is affected by inversion and presentation time. *Perception, 38*(12), 1849–1862.
Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General, 148*(6), 994.
Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences, 111*(15), E1454–E1462. https://doi.org/10.1073/pnas.1322355111
Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion, 98*(45–60), 16.
Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*(6), 1125.
Field, A., Field, Z., & Miles, J. (2012). *Discovering statistics using R*.
Gratch, J., & Marsella, S. (2013). *Social emotions in nature and artifact*. Oxford University Press.
Grauman, K., & Leibe, B. (2011). *Visual object recognition*. Morgan & Claypool Publishers.
Halberstadt, A. G., Cooke, A. N., Garner, P. W., Hughes, S. A., Oertwig, D., & Neupert, S. D. (2022). Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion, 22*(3), 403.
Hartling, C., Metz, S., Pehrs, C., Scheidegger, M., Gruzman, R., Keicher, C., Wunder, A., Weigand, A., & Grimm, S. (2021). Comparison of Four fMRI Paradigms Probing Emotion Processing. *Brain Sciences, 11*(5), 525. https://www.mdpi.com/2076-3425/11/5/525.
Holland, C. A. C., Ebner, N. C., Lin, T., & Samanez-Larkin, G. R. (2019). Emotion identification across adulthood using the Dynamic FACES database of emotional expressions in younger, middle aged, and older adults. *Cognition & Emotion, 33*(2), 245–257. https://doi.org/10.1080/02699931.2018.1445981
Hossain, S., Umer, S., Asari, V., & Rout, R. K. (2021). A Unified Framework of Deep Learning-Based Facial Expression Recognition System for Diversified Applications. *Applied Sciences, 11*(19), 9174. https://www.mdpi.com/2076-3417/11/19/9174.
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal, 50*(3), 346–363.
Hull, R. H. (2016). The art of nonverbal communication in practice. *The Hearing Journal, 69*(5), 22–24.
Jacintha, V., Simon, J., Tamilarasu, S., Thamizhmani, R., yogesh, K. T., & Nagarajan, J. (2019). A Review on Facial Emotion Recognition Techniques. In *2019 International Conference on Communication and Signal Processing (ICCSP)*.
Jaiswal, S., & Nandi, G. C. (2020). Robust real-time emotion detection system using CNN architecture. *Neural Computing & Applications, 32*(15), 11253–11262. https://doi.org/10.1007/s00521-019-04564-4
Jin, L., Zhou, Y., Ma, G., & Song, E. (2023). Quaternion deformable local binary pattern and pose-correction facial decomposition for color facial expression recognition in the wild. *IEEE Transactions on Computational Social Systems, 11*(2), 2464–2478.
Kaur, M., & Kumar, M. (2024). Facial emotion recognition: A comprehensive review. *Expert Systems, 41*(10), Article e13670.
Kim, H.y., & McGill, A. L. (2025). AI-induced dehumanization. *Journal of Consumer Psychology, 35*(3), 363–381.
Knowles, J. E., & Frederick, C. (2025). merTools: Tools for Analyzing Mixed Effect Regression Models. https://github.com/jknowles/mertools.
Kohler, C. G., Turner, T., Stolar, N. M., Bilker, W. B., Brensinger, C. M., Gur, R. E., & Gur, R. C. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research, 128*(3), 235–244. https://doi.org/10.1016/j.psychres.2004.07.003
Lau, W. K., Chalupny, J., Grote, K., & Huckauf, A. (2022). How sign language expertise can influence the effects of face masks on non-linguistic characteristics. *Cognitive research: Principles and Implications, 7*(1), 53.
Li, K., Chen, H., Huang, F., Ling, S., & You, Z. (2021). Sharpness and brightness quality assessment of face images for recognition. *Scientific Programming, 2021*(1), Article 4606828.
Li, S., Ham, J., & Eastin, M. S. (2024). Social media users' affective, attitudinal, and behavioral responses to virtual human emotions. *Telematics and Informatics, 87*, Article 102084. https://doi.org/10.1016/j.tele.2023.102084
Lo, L., Ruan, B.-K., Shuai, H.-H., & Cheng, W.-H. (2023). Modeling uncertainty for low-resolution facial expression recognition. *IEEE Transactions on Affective Computing, 15*(1), 198–209.
Lüdecke, D. (2025). sjPlot: Data Visualization for Statistics in Social Science. https://CRAN.R-project.org/package=sjPlot.
Mahfoudi, M.-A., Meyer, A., Gaudin, T., Buendia, A., & Bouakaz, S. (2022). Emotion expression in human body posture and movement: A survey on intelligible motion factors, quantification and validation. *IEEE Transactions on Affective Computing, 14*(4), 2697–2721.
Maratos, F. A., Mogg, K., Bradley, B. P., Rippon, G., & Senior, C. (2009). Coarse threat images reveal theta oscillations in the amygdala: A magnetoencephalography study. *Cognitive, Affective, & Behavioral Neuroscience, 9*(2), 133–143.
Martínez, F., Hernández, C., & Rendón, A. (2020). Identifier of human emotions based on convolutional neural network for assistant robot. *TELKOMNIKA (Telecommunication Computing Electronics and Control), 18*(3), 1499–1504.
McKone, E., & Robbins, R. (2011). Are faces special. *Oxford handbook of face perception*.
Melinte, D. O., & Vladareanu, L. (2020). Facial expressions recognition for human–robot interaction using deep convolutional neural networks with rectified adam optimizer. *Sensors, 20*(8), 2393.

Menaka, K., & Yogameena, B. (2021). Face detection in blurred surveillance videos for crime investigation. *Journal of Physics: Conference Series, 1917*(1), Article 012024. https://doi.org/10.1088/1742-6596/1917/1/012024

Mohammed, O. A., & Al-Tuwaijari, J. M. (2022). Analysis of challenges and methods for face detection systems: A survey. *International Journal of Nonlinear Analysis and Applications, 13*(1), 3997–4015.

Mohana, M., & Subashini, P. (2024). Facial Expression Recognition Using Machine Learning and Deep Learning Techniques: A Systematic Review. *SN Computer Science, 5*(4), 432. https://doi.org/10.1007/s42979-024-02792-7

Molas, G., & Nowak, E. (2021). Advances in Emerging Memory Technologies: From Data Storage to Artificial Intelligence. *Applied Sciences, 11*(23), Article 11254. https://www.mdpi.com/2076-3417/11/23/11254.

Nan, F., Jing, W., Tian, F., Zhang, J., Chao, K.-M., Hong, Z., & Zheng, Q. (2022). Feature super-resolution based facial expression recognition for multi-scale low-resolution images. *Knowledge-Based Systems, 236*, Article 107678.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466.

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01516, 1516-1516.

Park, H., Shin, Y., Song, K., Yun, C., & Jang, D. (2022). Facial Emotion Recognition Analysis Based on Age-Biased Data. *Applied Sciences, 12*(16), 7992. https://www.mdpi.com/2076-3417/12/16/7992.

Pascual, A. M., Valverde, E. C., Kim, J.-i., Jeong, J.-W., Jung, Y., Kim, S.-H., & Lim, W. (2022). Light-FER: A Lightweight Facial Emotion Recognition System on Edge Devices. *Sensors, 22*(23), 9524. https://www.mdpi.com/1424-8220/22/23/9524.

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Science Information, 21*(4–5), 529–553. https://doi.org/10.1177/053901882021004003

Prasad, S. B. R., & Chandana, B. S. (2023). Mobilenetv3: a deep learning technique for human face expressions identification. *International Journal of Information Technology, 15*(6), 3229–3243.

Rosenberg, E. L., & Ekman, P. (2020). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press.

Russell, J. A., & Fernandez-Dols, J. M. (1997). *The psychology of facial expression*. Cambridge university press.

Sabry, F., Eltaras, T., Labda, W., Alzoubi, K., & Malluhi, Q. (2022). Machine learning for healthcare wearable devices: the big picture. *Journal of Healthcare Engineering, 2022*(1), Article 4653923.

Savchenko, A. V. (2021). Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. In *2021 IEEE 19th international symposium on intelligent systems and informatics (SISY)*.

Savchenko, A. V., Savchenko, L. V., & Makarov, I. (2022). Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing, 13*(4), 2132–2143.

Shaik, T., Tao, X., Higgins, N., Li, L., Gururajan, R., Zhou, X., & Acharya, U. R. (2023). Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13*(2), Article e1485.

Siam, A. I., Soliman, N. F., Algarni, A. D., Abd El-Samie, F. E., & Sedik, A. (2022). Deploying machine learning techniques for human emotion detection. *Computational Intelligence and Neuroscience, 2022*(1), Article 8032673.

Smith, F. W., & Rossit, S. (2018). Identifying and detecting facial expressions of emotion in peripheral vision. *PLoS One, 13*(5), Article e0197160.

Stoychev, S., & Gunes, H. (2022). The effect of model compression on fairness in facial expression recognition. *International Conference on Pattern Recognition*.

Tekkök, S.Ç., Söyünmez, M. E., Bostancı, B., & Ekim, P. O. (2021). Face detection, tracking and recognition with artificial intelligence. In *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*.

Thuseethan, S., & Kuhanesan, S. (2016). *Eigenface based recognition of emotion variant faces*.

Vlastos, D. D., Kyritsis, M., Varela, V. A., Gulliver, S. R., & Spiroulia, A. P. (2020). Can a Low-Cost Eye Tracker Assess the Impact of a Valent Stimulus? A Study Replicating the Visual Backward Masking Paradigm. *Interacting with Computers, 32*(1), 132–141. https://doi.org/10.1093/iwc/iwaa010

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion, 83*, 19–52.

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I. B., & Chung, J. M. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications, 8*(1), Article 14821.

Wright, J. (2023). Suspect AI: Vibraimage, emotion recognition technology and algorithmic opacity. *Science Technology & Society, 28*(3), 468–487.

Xu, T., White, J., Kalkan, S., & Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. In A. Bartoli, & A. Fusiello (Eds.), *Computer Vision – ECCV 2020 Workshops Cham*.

Ye, K., & Kovashka, A. (2021). A case study of the shortcut effects in visual commonsense reasoning. *Proceedings of the AAAI conference on artificial intelligence*.