Techniques for facial affective computing: A review

Bashir Eseyin Abdullahi, Emeka Ogbuju, Taiwo Abiodun, Francisca Oladipo

Abstract
Facial affective computing has gained popularity and become a progressive research area, as it plays a key role in human-computer interaction. However, many researchers lack the right technique to carry out a reliable facial affective computing effectively. To address this issue, we presented a review of the state-of-the-art artificial intelligence techniques that are being used for facial affective computing. Three research questions were answered by studying and analysing related papers collected from some well-established scientific databases based on some exclusion and inclusion criteria. The result presented the common artificial intelligence approaches for face detection, face recognition and emotion detection. The paper finds out that the haar-cascade algorithm has outperformed all the algorithms that have been used for face detection, the Convolutional Neural Network (CNN) based algorithms have performed best in face recognition, and the neural network algorithm with multiple layers has the best performance in emotion detection. A limitation of this research is the access to some research papers, as some documents require a high subscription cost.

Practice implication: The paper provides a comprehensive and unbiased analysis of existing literature, identifying knowledge gaps and future research direction and supports evidence-based decision-making. We considered articles and conference papers from well-established databases. The method presents a novel scope for facial affective computing and provides decision support for researchers when selecting plans for facial affective computing.

Keywords: facial affective computing; face recognition; face detection; emotion detection; artificial intelligence.

INTRODUCTION

The term affective computing was introduced by Rosalind Picard in 1995 as a type of computing that deliberately influences emotion" (Calvo et al., 2015). Affective computing is a branch of artificial intelligence that deals with the interaction between humans and computers and allows the computer to study and understand signs from humans, like gestures, facial expressions, body language and voice tones, to assess emotional state. Thus, affective computing is analysing facial looks in images and text videos using computer vision technology to analyse human emotional status (Gaudenz, 2022). Also, Affective computing can be explained as the interaction between humans and computers in which a computer can detect and respond to human emotions and other stimuli. Facial expressions, gestures, posture, speech and changes in the hand's temperature on the mouse can all indicate changes in the user's emotional state. The computer can identify and interpret all these (Flynn et al., 2020). According to Gu et al. (2019), we have seven ways emotional states can be classified: anger, fear, sadness,
happiness, disgust, contempt and surprise. These emotional states can be examined by
an affective computing system though there are variations in classification. There are
many application areas of affective computing, such as in the health sector, to monitor
the entry and exit of patience, staff records, patient diagnoses and premises security.
Also, it can be applied in the sporting arena to check the activities of spectators and
other areas.

Facial Emotion Recognition (FER) is a technology researchers use to analyse
sentiments from different sources, like videos and pictures. FER belongs to the class of
technologies, which is frequently called "affective computing", a multidisciplinary field
of research on the ability of the computer to identify and interpret human emotions
based on Artificial Intelligence technologies. The technology analyses facial expressions
from static images and videos to reveal information on human emotional states. Facial
expressions play an important role when recognising emotions and can be used in the
non-verbal communication process and the identification of people (EDPS, 2021). The
computer can recognise emotional states from people's faces the same way as
humans; consequently, facial expression signs are often used in automatic systems
(Samadiani et al., 2019). FER analysis has three steps which include (i) face detection,
(ii) detection of expression, and (iii) classification of expression to emotional state.
Emotion detection is based on studying facial landmark positions such as eyebrows,
nose shape, lips position etc. Additionally, changes in video positions are also analysed
in videos to identify shrinkages in a cluster of facial muscles (EDPS, 2021).

The remaining part of this paper consists of "the methodology", which has the
method that was used to achieve the aim and objectives of this review, "The search
process", which shows the process by which research papers were searched and
retrieved for this research, "result and discussion" which shows and discuss the state-
of-the-art artificial intelligence techniques that are being used for face detection, face
recognition, and emotion detection as well as the state-of-the-art dataset that is being
used for emotion detection, and finally "the conclusion" which shows the summary of
the work that has been done in this research.

**Methodology**

The method used in this research is a systematic literature review method, which is
known to be in-depth and reliable (Kitchenham et al., 2009). This systematic literature
review guideline was proposed by (Jokonowo et al., 2018). The method involves
formulating some research questions and then finding the solutions to the prepared
questions by reviewing and analysing relevant studies based on the inclusion and
exclusion criteria.

**Inclusion criteria**

Papers written in the English language and peer/blind-reviewed journals, workshops
and conferences between the year 2018 and 2023 were included. Additionally, these
papers must be relevant to artificial intelligence, face recognition, face detection and
emotion detection to qualify for inclusion.
Exclusion criteria

Papers not written in English and unrelated to face recognition, face detection or emotion detection were excluded. Reports whose contributions to knowledge were not stated explicitly in the abstract were all excluded from the papers reviewed in this work.

Search string

We constructed a search phrase using keywords based on the stated research questions in section 4 to create a good search string. The databases presented in Table 1 were used to carry out our searches. The search strings are "Facial affective systems", "face recognition", "face detection", "emotion detection", "machine learning algorithms for face recognition", "machine learning algorithms for face detection", "deep learning algorithms for face recognition", "deep learning algorithms for face detection", and "review on facial affective systems".

Table 1. Academic database

<table>
<thead>
<tr>
<th>S/N</th>
<th>Database</th>
<th>Publisher</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Springer Link</td>
<td>Springer</td>
<td><a href="https://link.springer.com/">https://link.springer.com/</a></td>
</tr>
<tr>
<td>2</td>
<td>IEEE Explore</td>
<td>IEEE</td>
<td><a href="https://ieeexplore.ieee.org/Xplore/home/">https://ieeexplore.ieee.org/Xplore/home/</a></td>
</tr>
<tr>
<td>3</td>
<td>Web of Science</td>
<td>Clarivate</td>
<td><a href="https://www.sciencedirect.com/">https://www.sciencedirect.com/</a></td>
</tr>
<tr>
<td>4</td>
<td>Scopus</td>
<td>Elsevier</td>
<td><a href="https://www.scopus.com/home/">https://www.scopus.com/home/</a></td>
</tr>
</tbody>
</table>

Search process

1,267 papers were obtained from the database using the search strings; nevertheless, many were unrelated to the research, so the documents were first separated based on abstract and titles of papers, and 683 articles were removed, leaving 584 pieces. Furthermore, 38 relevant reports were extracted based on the inclusion and exclusion criteria.

Research questions

Three research questions were formed, and the answers to these questions were provided in this paper's result and discussion section. The research questions are:

RQ1: What AI techniques are being used for face detection, and how well have these techniques performed?

RQ2: What AI techniques are used for face recognition, and how well have these techniques performed?

RQ3: What AI techniques are being used for emotion detection, and how well have these techniques performed?

RQ4: What are the state-of-the-art dataset for emotion detection?

This literature review will answer these questions in the result and discussion part.

RESULT AND DISCUSSION

Answers to RQ1

AI techniques used for Face Detection
Gretchen et al. (2018) developed a system to track humans by detecting the head and face in real time. The researchers used computer vision techniques with the OpenCV library; the Python programming language was used for the coding. The archers’ dataset contains images from webcam and Closed-Circuit Television (CCTV). The algorithms used for the implementation are the Haar-like classifier to detect the head and Consensus-based Matching and Tracking of keypoint for the object (CMT) object-tracking algorithm to track the head. During the system evaluation, the CMT that follows the charge was 68% accurate, while the Haar-like classifier that detects the human head was 83% accurate.

Lee et al. (2018) researched to solve face detection issues in different lighting states. The study creates an intelligent and effective face detection system in another lightning form. OpenCV and Visual Studio2015 were used for their implementation. The model was able to achieve an accuracy of 80%.

A framework that uses the Viola and Jones algorithm on employee facial database to identify personal employee statements was developed by Sriratana et al. (2018); the researchers used the OpenCV library, and the face detection model that was created was able to achieve an accuracy of 90%.

Gorbatevich et al. (2019) worked on a system that detects human faces in 2019. The researchers adopted the Convolutional Neural Network (CNN) based algorithm for face detection purposes, and a dataset called the WIDER FACE dataset was used to train the face detection model. The CNN algorithm was able to achieve an accuracy of 62.9%.

Seunghyun et al. (2019) proposed an approach to detect faces in real-time video. The study makes use of the Boundary-Saliency-Multi-Branch Network (BSMNet), the Multi-Task-Cascaded Convolutional Neural Network (MTCNN) algorithms to see smaller fronts and the You-Only-Look-Once (YOLO) algorithm to detect larger faces. From the study, it was observed that optimisation of speed and accuracy is possible if some of the hidden layers in the neural network are removed.

A model that uses Red-Green-Blue (RGB) approach for facial and fundamental object recognition was developed by Ferdousi et al. (2019). The researchers used the YOLO and MTCNN algorithms for object detection and facial recognition. The OpenCV library and the researchers used datasets locally curated from the environment. An accuracy of 63-80% and 80-100% was achieved by the YOLO and the MTCNN algorithms, respectively.

Mehariya et al. (2020) developed a system that can take attendance of students attending a class or workers at work by detecting their faces. The researchers trained the model using images of students collected from a course as a dataset. The OpenCV, occupancy ratio, and see Multiscale techniques were adopted for the research. The researchers used the Haar-cascade algorithm to train the face detection model, which achieved an accuracy ranging between 90% and 100%.

Gupta et al. (2020) proposed a system based on image processing that can help to improve attendance systems in higher institutions. The system can also reduce the wastage rate when taking attendance. The attendance system aims to manage attendance students by estimating the times a student is present and the number of times the student is absent, then automatically producing a document that shows the report. The OpenCV library was used for the implementation. The Haar-cascade
algorithm was used for face detection, and the LBPH was used for face recognition. From the evaluation, the model achieved an accuracy of 97%.

A system that can identify faces from video streams and images was developed by Hoque et al. (2020). The ATMega328p Microcontroller and the OpenCV were used for image processing. The researcher trained the model with an open-source dataset from Kaggle. The researchers used AdaBoost, Haar-cascade Camshift, Viola Jonas and Housedorff Gap to identify human faces. And the Haar-cascade was used to classify faces. The face classification model was 83% accurate.

A system that detects a human face before extracting and using features from the detected face to recognise the beginning was developed by Singh & Jasmine (2019). The viola-jones detection algorithm, which uses the AdaBoost learning algorithm and integral image as classifiers, was used alongside the haar cascade and KLT algorithm for face detection. The viola-jones performed best in different lighting conditions compared to other algorithms used for face detection.

Table 2. **Summary of techniques used for face detection**

<table>
<thead>
<tr>
<th>Papers</th>
<th>Techniques</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gretchen et al. (2018)</td>
<td>Haar-like classifier to detect the head and CMT object-tracking algorithm to track the head</td>
<td>CMT that tracks the head was 68% accurate, while the Haar-like classifier that detects the human's lead was 83% accurate.</td>
</tr>
<tr>
<td>Lee et al. (2018)</td>
<td>Haar cascade algorithm</td>
<td>The model was able to achieve an accuracy of 80%.</td>
</tr>
<tr>
<td>Sriratana et al. (2018)</td>
<td>Viola and Jones's algorithm</td>
<td>The face detection model that was developed achieved an accuracy of 90%.</td>
</tr>
<tr>
<td>Gorbatsevich et al. (2019)</td>
<td>CNN algorithm</td>
<td>The CNN algorithm was able to achieve an accuracy of 62.9%</td>
</tr>
<tr>
<td>Lee et al. (2019)</td>
<td>BSMNet and MTCNN algorithms to detect smaller faces and the Yolo algorithm to detect larger faces</td>
<td>Not stated</td>
</tr>
<tr>
<td>Ferdousi et al. (2019)</td>
<td>YOLO and MTCNN algorithms for face detection and facial recognition, respectively</td>
<td>An accuracy of 63-80% and 80-100% was achieved by the YOLO and the MTCNN algorithms, respectively.</td>
</tr>
<tr>
<td>Mehariya et al. (2020)</td>
<td>Haar-cascade algorithm</td>
<td>The algorithm achieved an accuracy that ranges between 90% and 100%.</td>
</tr>
<tr>
<td>Gupta et al. (2020)</td>
<td>The Haar-cascade algorithm was used for the face detection process</td>
<td>The model achieved an accuracy of 97%.</td>
</tr>
<tr>
<td>Hoque et al. (2020)</td>
<td>Haar-cascade algorithm</td>
<td>The model was 83% accurate</td>
</tr>
<tr>
<td>Singh &amp; Jasmine (2019)</td>
<td>Haar-cascade and KLT</td>
<td>Not stated</td>
</tr>
</tbody>
</table>

Based on our reviewed literature, we observed that the major techniques used for face detection are machine learning and deep learning techniques, which includes Haar-cascade, KLT, YOLO, BSMNet, Viola and Jones, CMT object-tracking, and MTCNN algorithms. The haar-cascade algorithm, which is a machine learning algorithm used to identify objects in images and video, is mostly used for face detection and has achieved high accuracy that ranges from 83% to 97% as used by researchers like
The BSMNet, MTCNN, and YOLO algorithms have also been used for face detection and performed well by researchers like Gorbatsevich et al. (2019) and Lee et al. (2019). Other researchers like Ferdousi et al. (2019) used nonparametric algorithms for face detection.

Answer to RQ2

AI techniques used for face recognition

An attendance record system based on face recognition for Nigerian universities was proposed by Ogbuju et al. (2020). The method uses the Faster Region Convolutional Neural Netowrk (FRCNN) algorithm for implantation. Google Colab was used to design and implement the model. The dataset used to train the model was collected locally by capturing the faces of students using the camera. The developed system, called "DLFacer," has 97-99% accuracy.

A system that recognises human faces based on a computational-intelligent algorithm was developed by AbdELminaam et al. (2020). The researchers made use of the model created with the help of Deep Neural Networks (DNN). The result from the proposed method was compared with three different machine learning algorithms, which are Decision Tree (DT), K-Nearest Neighbor (KNN) and the Support Vector Machine (SVM). The DNN performed better than all the compared algorithms with the highest accuracy (99.06%), the highest recall (99.07), the highest precision (99.12%) and the highest specificity (99.10%).

A study on the application of facial recognition-based algorithms on examination officers' security systems at Omar Al-Mukhtar University was carried out by Bazama et al. (2021). The Independent Component Analysis (ICA) was utilised for face recognition. The data used to train the model consists of 100 images of people's faces at the school's College of Science. The ICA algorithm was able to achieve 86.7%.

A deep learning method was used for face identification and face recognition by Teoh et al. (2020). The CNN algorithm was used to train the face detection model. The researchers also utilised the use of the OpenCV library in Python. The model was able to achieve 91.7% in face recognition.

Wahana et al. (2020) designed a face recognition system that can be used for indoor applications. The face recognition process used the Local-None-Negative-Matrix-Factorization (LNMF) and the none-negative-matrix-factorisation-suppressed-carrier (NMFsc) methods. Videos of classroom students with 640X480 pixels were captured and used to train the face recognition model. The LNMF algorithm was 71.61% accurate with a computation time 152,634. In comparison, the NMFsc algorithm achieved an accuracy of 86.76% with a computation time of 152,785.

Singh & Jasmine (2019) introduced a system that automatically recognises people's faces. The system works by first detecting an individual's face, extracting the future from the face that has been seen and recognising the face. The Principal Component Analysis (PCA) algorithm was used for face recognition purposes, and the model outperformed other compared algorithms used in face recognition. The PCA algorithm achieved 77% accuracy.

A paper that proposes a robust and efficient method for face recognition in real-time was submitted by Vinay et al. (2018). The proposed method was based on Multilayer Perception (MLP), which uses the gradient-descent algorithm. The MLP was
fed in with binary patterns extracted from the images. The method was tested on four different datasets, including the Grimace, FACES96, FACES95 and FACES94 datasets. The technique was able to achieve an accuracy of 91%.

Onyema et al. (2021) presented a study on a deep learning-based method for facial recognition. The deep learning method presented for the research implementation is the Convolutional Neural Network (ConvNet). The dataset used for the implementation is the FER2013 dataset, which was gathered automatically using image search API from Google. The proposed model was able to achieve an accuracy of 70%.

An attendance system based on face recognition was developed by Bhagat et al. (2021). The system works by discovering and recognizing faces from the classroom’s live stream footage and then sending it to the appropriate faculty. The researchers generated a dataset from photos of kids in a class. The skin-splitting approach was used. The system was able to reach an accuracy of 82.3%

The research by Gupta et al. (2020) to improve the attendance system in higher institutions also used the Local Binary Patterns Histogram (LBPH) for the face recognition process. The research used the OpenCV library for the implementation. The LBPH achieved an accuracy of 97%.

Table 3. Summary of techniques used for face recognition

<table>
<thead>
<tr>
<th>Papers</th>
<th>Techniques</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogbujo et al. (2020)</td>
<td>Faster Region Convolutional Neural Network (FRCNN)</td>
<td>The model has an accuracy of 97-99%</td>
</tr>
<tr>
<td>AbdElminaam et al. (2020)</td>
<td>deep-convolutional-neural-networks (DCNN)</td>
<td>The DCNN has an accuracy of 99.06%</td>
</tr>
<tr>
<td>Bazama et al. (2021)</td>
<td>The independent component analysis (ICA)</td>
<td>The ICA algorithm was able to achieve 86.7%.</td>
</tr>
<tr>
<td>Teoh et al. (2020)</td>
<td>Convolutional Neural Network (CNN)</td>
<td>The model was able to achieve 91.7% accuracy</td>
</tr>
<tr>
<td>Wahana et al. (2020)</td>
<td>The local-none-negative-matrix-factorization (LNMF) and the none-negative-matrix-factorization-suppressed-carrier (NMFsc)</td>
<td>The LNMF algorithm was 71.61% accurate, while the NMFsc algorithm achieved an accuracy of 86.76%.</td>
</tr>
<tr>
<td>Singh &amp; Jasmine (2019)</td>
<td>PCA algorithm</td>
<td>The PCA algorithm achieved 77% accuracy</td>
</tr>
<tr>
<td>Vinay et al. (2018)</td>
<td>gradient-descent algorithm</td>
<td>The method was able to achieve an accuracy of 91%</td>
</tr>
<tr>
<td>Onyema et al. (2021)</td>
<td>convolutional-neural-network (ConvNet)</td>
<td>The model was able to achieve an accuracy of 70%.</td>
</tr>
<tr>
<td>Bhagat et al. (2021)</td>
<td>The skin-splitting approach</td>
<td>The model achieved 82.3%</td>
</tr>
<tr>
<td>Gupta et al. (2020)</td>
<td>LBPH algorithm</td>
<td>The model achieved 97% accuracy</td>
</tr>
</tbody>
</table>

From our reviewed paper on the commonly used techniques for face detection, we notice that CNN, RCNN, ICA, LNMF, NMFsc, PCA, ConvNet, gradient-descent, LBPH, and skin splitting algorithms have recently been used by researchers like (Bazama et al., 2021; Teoh et al., 2020; Wahana et al., 2020; Singh & Jasmine 2019; Vinay et al., 2018; Onyema et al., 2021; Bhagat et al., 2021; Gupta et al., 2020). It is also observed that
most of these algorithms used in the reviewed papers are related to the deep learning CNN algorithm and have performed excellently when used for face detection purposes. The CNN-based algorithms have achieved high accuracies of 91.7% and 99.06%, except for the ConvNet, which earned a lesser accuracy of 70% due to the limited dataset the researchers used.

**Answers to RQ3**

AI techniques used for emotion detection

Ogbuju et al. (2021) presented research that detects emotions from online news/text. The authors used twelve (12) different online news channels in Nigeria as datasets to train the proposed model and a hybrid method of corpus-based lexicon and dictionary approaches to achieve the research aim.

Gupta (2018) proposed a method used to detect facial emotion. The detect emotions on faces in real-time; after the feeling has been seen, it will now be categorised according to the emotions category using the Support Vector Machine (SVM) algorithm. During the evaluation that was done on the model, it was observed that the SVM was able to achieve an accuracy of 93.7%. The researchers also state in their report that the facial landmarks may be tweaked to increase the model’s accuracy.

Patel et al. (2018) developed a model that can detect a driver sleeping when driving based on facial expression, the research aimed to prevent road accidents. The researchers used the OpenCV library. Also, a vision device that monitors the face to detect the driver's facial emotion and eye blinking was used for the implementation. Artificial Neural Network (ANN) algorithm was used for the modelling, and at the evaluation stage, it was discovered that the algorithm was 90% accurate.

Jain et al. (2018) used a hybrid of CNN, and Recurrent Neural Networks (RNN) were used for face emotion detection since Deep Neural Networks (DNN) always perform better than traditional models when compared by researchers. Two publicly available datasets were used to train the hybrid of deep learning algorithms. The hybrid model was 90% accurate.

Benamara et al. (2019) developed a facial emotion detection system; the researchers used an ensemble of YOLO and CNN algorithms to train the emotion detection model. the model was evaluated in real-time, and the ensemble model attained an accuracy of 72.47%.

A real-time system that monitors elders' living conditions in case of emergency was developed by Punidha et al. (2020). The system monitors the elders and sends messages to the relatives to alert them during an emergency. The researcher used a locally curated dataset. The CNN algorithm was used for the implementation purpose, although the research's accuracy level was not stated in the report.

A study based on emotion recognition of speech and graphic visualisation of expression in an intelligent learning environment was carried out by Lu (2021). The model was trained using the Convolutional Neural Network Bi-Directional Long Short-Term Memory (CNN-BiLSTM) algorithm. The researcher conducted a simulation experiment to check the algorithm’s performance. The result shows that the CNN-BiLSTM was 98.75% accurate.

Farhoumandi et al. (2021) tried to predict alexithymia using a facial emotion recognition model. Several students from the University of Tabriz were selected based
on their scores on the Toronto Alexithymia scale, which was used as a dataset. The facial expression model was developed using SVM and Feed Forward Neural Network (FFNN). An accuracy ranging between 72.7 to 81.8 was obtained after evaluating the model’s performance using the area under curve (AUC).

Research presented a real-time method for implementing a model that detects emission and deploying it in a robot conducted by Siam et al. (2022). Three datasets were used to develop the model: the Cohn-Kande dataset, which contains 593 video sequences; the JAFFE dataset, which contains facial expressions of female Japanese; and a dataset from the real-world face effective database. Machine learning algorithms such as SVM, KNN, Naïve Bayes (NB), Linear Regression (LR), Random Forest (RF) and MLP were used to train the model. At the end of the training, only MLP was deployed because it outperformed all the algorithms used. The MLP has the highest accuracy of 97%.

A facial emotion recognition model that can recognise emotions such as happiness, anger, neutrality, surprise, sadness, fear and disgust was proposed by Debnath et al. (2022). A deep learning algorithm called ConvNet was used to develop the model. The ConvNet algorithm used four layers for facial emotion detection. A dataset known as the JAFFE dataset was used to train the model. The ConvNet was able to achieve 98.13% accuracy.

Singh et al. (2022) Developed deep learning-based facial emotion detection using the CNN algorithm. The paper aims to detect emotions in humans and extract human behaviour. This model successfully saw emotions such as anger, happiness, fear, calmness, sadness, surprisingness and disgust. The model has an accuracy of 90-100.

Table 4. Summary of techniques used for emotion detection

<table>
<thead>
<tr>
<th>Authors</th>
<th>Techniques</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogbuju et al. (2021)</td>
<td>corpus-based lexicon and dictionary approaches</td>
<td>Accuracies from 90-99%</td>
</tr>
<tr>
<td>Gupta (2018)</td>
<td>SVM algorithm</td>
<td>The algorithm achieved an accuracy of 93.7%</td>
</tr>
<tr>
<td>Patel et al. (2018)</td>
<td>ANN algorithm</td>
<td>The algorithm was 90% accurate</td>
</tr>
<tr>
<td>Jain et al. (2018)</td>
<td>A hybrid of CNN and RNN</td>
<td>The hybrid model was 90% accurate</td>
</tr>
<tr>
<td>Benamara et al. (2019)</td>
<td>An ensemble of YOLO and CNN algorithm</td>
<td>The ensemble model was 72.47% accurate</td>
</tr>
<tr>
<td>Punidha et al. (2020)</td>
<td>CNN algorithm</td>
<td>The model achieved 92% accuracy</td>
</tr>
<tr>
<td>Lu (2021)</td>
<td>The CNN-BiLSTM algorithm</td>
<td>The CNN-BiLSTM was 98.75% accurate</td>
</tr>
<tr>
<td>Farhoumandi et al. (2021)</td>
<td>SVM and FFNN algorithms</td>
<td>An accuracy ranging from 72.7 to 81.8 was obtained after evaluating the model’s performance using AUC.</td>
</tr>
<tr>
<td>Siam et al. (2022)</td>
<td>Machine learning algorithms such as SVM, KNN, NB, LR, RF and MLP were used to train the model.</td>
<td>MLP has performed best, with the highest accuracy of 97%.</td>
</tr>
<tr>
<td>Deb Nath et al. (2022)</td>
<td>Four-layer ConvNet deep learning algorithm</td>
<td>The ConvNet was able to achieve 98.13% accuracy</td>
</tr>
<tr>
<td>Singh et al. (2022)</td>
<td>CNN algorithm</td>
<td>The model has an accuracy of 90-97%</td>
</tr>
</tbody>
</table>
The techniques being used for emotion detection are ANN, CNN, FFNN, MLP, CNN-BiLSTM, YOLO, assemble of YOLO and CNN, Hybrid of CNN and RNN, SVM, KNN, NB, LR, and RF. We have also observed that many researchers have employed machine learning and deep learning techniques for emotion detection purposes. As reviewed in this research, one of the algorithms that were being used is the SVM which was used by researchers like (Gupta 2018; Farhoumandi et al., 2021; Siam et al., 2022) went ahead and compared the SVM with other algorithms which include KNN, NB, LR, RF and a multilayer perceptron (MLP) algorithm, eventually the MLP outperform the compared machine learning algorithm. Other algorithms used are deep learning algorithms which include the CNN algorithm, which was used by (Punidha et al., 2020; Singh et al. 2022). The ANN algorithms were used by Patel et al. (2018), a four-layer ConvNet deep learning algorithm used by Debnath et al. (2022), CNN-BiLSTM used by Lu (2021), a hybrid of CNN and RNN used by Jain et al. (2018) and an ensemble of YOLO and CNN algorithm which was used by Benamara et al. (2019). From the reviewed literature, we observed that the CNN-BiLSTM performed the best compared to other studied algorithms in this paper and achieved an accuracy of 98.75%. The deep learning four layers ConvNet also achieved excellently with an accuracy of 98.13% which is very close to the performance of the CNN-BiLSTM. Other algorithms, like CNN, RNN, a hybrid of CNN and RNN, and the YOLO and CNN algorithm ensemble, have performed well in the reviewed papers.

Answers to RQ4

State-of-the-art emotion detection datasets.
- **Kaggle emotion detection dataset**: This is an emotion detection dataset from Kaggle, which contains about 35,685 images of faces in greyscale. The dataset is already divided into train data and test data. The images in the dataset are categorised into seven different emotions: happiness, anger, sadness, surprise, fear, disgust and neutral (Ares, 2021).
- **AffectNet**: This dataset is one of the largest facial emotion detection datasets. The dataset contains over 1 million images of different faces, which are collected from other sources from the Internet using about 1250 keywords related to emotion detection in six languages. The languages include English, Arabic, Portuguese, Farsi, German and Spanish (Mahoor, 2023).
- **Ascertain**: This dataset was introduced by Subramanian et al. (2018). The Ascertain database is a multimodal database for affect recognition and implicit personality using commercial sensors. Ascertain dataset has 58 users' emotional self-ratings with synchronously recorded Electrocardiogram (ECG), Electroencephalogram (EEG), facial activity data and Galvanic Skin Response, which were recorded using an off-the-shelf sensor.
- **Dreamer**: This is a multimodal dataset containing ECG and EEG signals documented during effect elicitation through audio-visual stimuli. The dataset includes calls from 23 participants and their emotional state after each stimulation (Katsigiannis & Ramzan, 2018).
- **EMOTIC Dataset**: This dataset contains real-time images of people with different emotions. The dataset images are annotated with 26 categories of emotions with three continuous dimensions: Arousal, Valence and Dominance. The dataset contains about 23,571 photos, and some ideas are collected manually from the Internet using the Google search engine (Kosti et al., 2019).
- **Google Facial Expression Comparison Dataset**: This dataset contains face images of people with different facial expressions; the dataset set was gathered to help researchers on topics that are related to facial expression analysis like expression-based photo album summarisation, expression-based picture retrieval, classification of emotion, expression synthesis etc. (Vemulapalli & Agarwala, 2018).

- **K-EmoCon**: This dataset is a multimodal dataset with comprehensive annotation of continuous emotions during a naturalistic conversation. The dataset contains multimodal measurements, including ECG, audio visual recordings and peripheral physiological signals, acquired during about 16 different 10-minute debates based on social issues (Park et al., 2020).

- **CLASS dataset**: This is a multimodal dataset that was majorly developed to support researchers consisting of synchronous recording of ECG, Plethysmography (PPG), and ElectroDermal Activities (EDA) of 62 volunteers (Markova et al., 2020).

- **JAFFE dataset**: The JAFFE dataset, which stands for the Japanese Female Facial Expression dataset, contains 213 images that show the facial expressions of ten different volunteer female Japanese, each of the female Japanese leading seven other facial expressions (Lyons et al., 2020).

Table 5. **Summary of state of the art emotion detection datasets**

<table>
<thead>
<tr>
<th>S/N</th>
<th>Authors</th>
<th>Dataset</th>
<th>Size</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ares (2021)</td>
<td>Kaggle emotion detection dataset</td>
<td>35,685</td>
<td>Seven categories of emotions</td>
<td>images of faces in greyscale</td>
</tr>
<tr>
<td>2</td>
<td>Mahoor (2023)</td>
<td>AffectNet dataset</td>
<td>1,000,000</td>
<td>Seven categories of emotions</td>
<td>images of different faces are collected from the Internet</td>
</tr>
<tr>
<td>3</td>
<td>Subramannian et al. (2018)</td>
<td>Ascertain dataset</td>
<td>58</td>
<td>Five categories of emotions</td>
<td>synchronously recorded ECG, EEG, facial activity data and Galvanic skin response</td>
</tr>
<tr>
<td>4</td>
<td>Katsigiannis &amp; Ramzan (2018)</td>
<td>Dreamer dataset</td>
<td>23</td>
<td>Nine categories of emotions</td>
<td>ECG and EEG signals that were documented during effect elicitation through audio-visual-stimuli</td>
</tr>
<tr>
<td>5</td>
<td>Kosti et al. (2019)</td>
<td>EMOTIC dataset</td>
<td>23,571</td>
<td>26 categories of emotions</td>
<td>images of people collected manually from the Internet. recorded ECG, EEG, facial activity</td>
</tr>
<tr>
<td>6</td>
<td>Vemulapalli &amp; Agarwala (2018)</td>
<td>Google Facial Expression Comparison dataset</td>
<td>51,060</td>
<td>Three categories of emotions</td>
<td>Measurement of ECG, audio-visual recordings and peripheral physiological signals</td>
</tr>
<tr>
<td>7</td>
<td>Park et al. (2020)</td>
<td>K-EmoCon dataset</td>
<td>20</td>
<td>Nine categories of emotions</td>
<td>Measurement of ECG, audio-visual recordings and peripheral physiological signals</td>
</tr>
<tr>
<td>8</td>
<td>Markova, et al. (2020)</td>
<td>CLAS dataset</td>
<td>62</td>
<td>Five categories of emotions</td>
<td>Synchronously recorded ECG, PPG, and EDA.</td>
</tr>
<tr>
<td>9</td>
<td>Lyons et al. (2020)</td>
<td>JAFFE dataset</td>
<td>213</td>
<td>Seven categories of emotions</td>
<td>Collection of images of female Japanese facial expressions</td>
</tr>
</tbody>
</table>
The state-of-the-art emotion dataset is Kaggle emotion detection, AffectNet, Ascertain, Dreamer, EMOTIC, Google Facial Expression Comparison, K-EmoCon, CLAS, and JAFFE datasets. Most of these dataset contains synchronously recorded ECG, EEG, EDA, facial activity data and Galvanic skin Response with different emotion categories. In contrast, others consist of facial images of people with different emotional states.

**CONCLUSION**

This paper has presented the AI techniques for facial affective computing by answering some research questions. Firstly, the report showed the methods of face detection based on the reviewed paper to be Haar-cascade, KLT, YOLO, BSMNet, Viola and Jones, CMT object-tracking, and MTCNN algorithms. Secondly, papers on face recognition were reviewed, and the recently used techniques were found to be CNN, RCNN, ICA, LNMF, NMFsc, PCA, ConvNet, gradient-descent, LBPH and the skin splitting algorithms. Thirdly, we reviewed a paper on emotion detection and found out that ANN CNN, CovNet, FFNN, MLP, CNN-BiLSTM, YOLO, assemble of YOLO and CNN, Hybrid of CNN and RNN, SVM, KNN, NB, LR, and RF have been recently utilised for emotion detection. Finally, state of the art emotion detection dataset has been reviewed in this paper. The reviewed emotion detection datasets are Kaggle, AffectNet, Ascertain, Dreamer, EMOTIC, Google Facial Expression Comparison, K-EmoCon, CLAS, and JAFFE datasets. We observed from the checked state-of-the-art dataset that most datasets are gathered from different online platforms, and others contain images from foreign countries. For these reasons, we recommend a localised dataset that contains images of emotions from people from the locality.

**RECOMMENDATION**

It is observed that the haar-cascade algorithm performed all the algorithms used for face detection in the reviewed papers in terms of accuracy. Nevertheless, other deep learning algorithms like CNN will perform well for face detection. Hence, to improve accuracy and precision we recommend a hybrid of haar-cascade and CNN algorithms for face detection. Secondly, out of the recently used algorithms for face recognition, we observed that the CNN-based algorithms have also performed excellently as used by the researchers. However, other algorithms like the LBPH and gradient-descent algorithms have proven effective in recognising faces. Thus we recommend a hybrid of the CNN algorithm with either the LBPH algorithm or the gradient-descent algorithm. Also, based on our reviewed paper on emotion detection, we recommend neural network algorithms with multiple layers for emotion detection.

**REFERENCES**


About the authors:
Bashir Eseyin Abdullahi, Federal University Lokoja, Nigeria. bashir.abdullahi-pg@fulokoja.edu.ng
Emeka Ogbuju, Federal University Lokoja, Nigeria. emeka.ogbuju@fulokoja.edu.ng
Taiwo Abiodun, Federal University Lokoja, Nigeria. taiwo.abiodun-pg@fulokoja.edu.ng
Francisca Oladipo, Federal University Lokoja, Nigeria. francisca.oladipo@fulokoja.edu.ng