

MULTIMODAL EMOTION DETECTION FOR EDUCATION AND WORK ENVIRONMENT BY USING IMPROVED ARTIFICIAL INTELLIGENCE MACHINE VISION SYSTEM

Submitted: July 2024; accepted 6th September 2024

Wan Mohd Bukhari Wan Daud, Adnan Kiral, Mohamed Osman Tokhi, Lee Chung Yee, Muhammad Muzhafar
Mohammad Zawawi

DOI: 10.14313/jamris-2026-019

Abstract:

The use of artificial intelligence (AI) has significantly advanced emotion recognition within human-computer interaction (HCI). This paper aims to develop a multimodal emotion detection system for educational and work environments using an enhanced AI machine vision system. The primary focus is on training and testing a multimodal AI model in Python using convolutional neural networks (CNN). The results from the trained facial emotion AI model demonstrated substantial improvements. Training accuracy increased from 30.49% to 72.21%, while validation accuracy improved from 37.6% to 60.58%. Simultaneously, training loss decreased from 180.69% to 73.65%, and validation loss reduced from 172.97% to 107.53%. This CNN-based model can use OpenCV to detect seven emotions: happy, sad, neutral, angry, afraid, disgusted, and surprised. The ECG emotion AI model, also trained with CNN, also successfully recognized patterns for the same seven emotions. When these two models are combined into a multimodal AI system, they can detect facial and ECG-based emotions simultaneously. This comprehensive approach allows for the detection of both visible and hidden emotions, such as stress or anxiety, which may not be easily discernible through facial expressions alone. The integration of these models into a multimodal AI system provides a more accurate and holistic understanding of human emotions, enhancing applications in educational and work settings. The improved detection capabilities can lead to better user experiences and more effective responses to emotional states, ultimately contributing to advancements in HCI.

Keywords: Artificial Intelligence, Multimodal Emotion Detection, Machine Learning, Convolutional Neural Network, Machine Vision System

1. Introduction

Emotion detection integrates artificial intelligence, machine learning, and psychology in attempting to identify, analyze, and interpret human emotions through computational methods [8]. Its foundation lies in analyzing various data streams that encompass visual cues, like facial expressions captured through image or video data; linguistic patterns from textual information and tonal inflections in speech; physiological signals such as heart rate or electrodermal

activity. They have even been able to analyze behavioral cues. Employing complicated algorithms and techniques like machine learning models, deep neural networks, natural language processing, and computer vision, these systems extract, analyze, and interpret features embedded within these data modalities to differentiate and categorize emotional states [19].

The complexity of this field is amplified by the multifaceted nature of emotions, including slight variations and cultural differences in their expression, as well as the complex interaction between different emotional dimensions. Emotion detection systems continuously evolve, striving for greater accuracy and depth in understanding emotions beyond basic classifications (like happiness, sadness, anger) to address the nuances within each emotional state. These advancements smooth out the way for applications across various sectors, including the workplace, education, healthcare and customer service. Emotion detection, for example, can identify signs of stress or burnout among employees. This allows employers to implement strategies for stress reduction, provide support, and create a healthier work environment at their workplace. In the education field, emotion detection is also core to adaptive learning systems, in which educators to adjust teaching styles and content based on students' emotional states.

Ethical considerations surrounding privacy, biases, and fairness are integral in the development and deployment of these systems. We must ensure their responsible and equitable use while harnessing the potential to build more empathetic and responsive AI technologies that enhance human-machine interactions and enrich our understanding of human emotions and behaviors.

Recently, mental health issues in Malaysia became a major public health concern. The number of individuals experiencing serious long-term mental health issues such as depression, anxiety, stress and so on had increased drastically, and suicides had become more widespread. In Malaysia, there are about 5.5 million adolescents and the statistics have shown that 1 in 5 are depressed, 2 in 5 are anxious and 1 in 10 are stressed. Between 2012 and 2017, the rate of adolescent suicidal behavior—such as suicidal ideation, planning and attempts—had increased. Suicidal ideation increased from 7.9% to 10.0%, suicidal planning increased from 6.4% to 7.3%; and attempts increased

6.8% to 6.9%. As a result, mental health issues at the workplace or education have become a wake-up call to society and need to be solved urgently.

Most people tend to suppress their feelings rather than express their feelings to others. This has made it more difficult for people to recognize emotions in others. Emotion detection using artificial intelligence (AI) has played an important role in order to addressing this problem, as it can detect people's emotion without asking them. Mental health issues in education and the workplace can thus be identified early, and treatment can be started as soon as an abnormal emotional pattern is detected by the AI. The research of X. Li et al. [7] found that emotion detection has garnered significant interest in the field of education during the past few decades, since students' emotions are directly related to their learning effectiveness and academic performance. Based on the timely monitoring of students' emotional states with considerable assistance from AI, instructors can assess the emotions and engagement levels of their students in a way that is difficult to determine from watching them in a big classroom.

2. Literature Survey

To design the AI model, the focus of the literature review was to find:

1. Types of emotional expression.
2. Types of emotion detection systems.
3. Types of artificial intelligence (AI) technologies.
4. Types of machine learning systems.

In 2022, T. Kusunose et al. [3] highlighted that facial expressions are among the most efficient and effective ways to signal emotions, functioning as a near-universal language. This is because the facial cranial nerves that control the muscles involved in facial expressions are highly versatile, surpassing even the vocal cords in conveying prosody and inflection. Specific combinations of facial movements, such as a smile or displaying teeth, can indicate happiness, anger, or concern. Recent advancements in facial recognition technology can detect these expressions in real time, offering valuable applications in healthcare, education, and work environments for assessing emotions and task engagement.

In 2018, L. Shu et al. [29] found that physiological signals like electroencephalogram (EEG), electrocardiogram (ECG), and electromyography (EMG) are crucial for emotion detection due to their direct link to emotional states. These signals provide objective measures of physiological responses influenced by emotional arousal and valence, offering insights into emotions that may be difficult to verbalize. The process involves extracting features from these signals, such as EEG frequency bands or ECG heart rate variability, which are then analyzed and associated with emotions using machine learning algorithms. This method enables real-time monitoring and improves the accuracy and reliability of emotion recognition systems in various applications.

In 2021, X. Li et al. [7] wrote that AI-driven emotion detection focuses on automatically analyzing students' emotional states during educational lessons. AI systems can process large amounts of data from text, speech, facial expressions, and body language, recognizing emotions based on eye and head movements in online learning environments. Machine learning algorithms enable these systems to continuously improve their accuracy by learning from new data, making them effective for monitoring user status and mitigating risks. Additionally, as noted by G. Assunção et al. [8] in 2022, AI systems provide consistent and objective analysis, reduce human bias, and can be scaled for large-scale applications such as customer sentiment analysis and mental health monitoring.

In 2023, G. Shi [10] proposed using machine learning technologies to detect emotions by training AI models on extensive data. These algorithms continuously learn and improve their accuracy in recognizing patterns in facial expressions, speech, and body language. Machine learning is adept at identifying subtle emotional cues and can be customized for specific contexts, languages, or cultures. It enables real-time, scalable emotion detection across various applications—such as computer vision, natural language processing, and audio recognition—without requiring constant human intervention. This approach has significantly enhanced the efficiency and effectiveness of emotion detection.

M. U. Khan et al. [19] recommended the use of using deep learning methods, particularly transfer learning with models like MobileNetV2, to train emotion detection models. These methods automatically learn hierarchical representations from raw data, such as images, enhancing their ability to extract complex features crucial for recognizing emotional cues from facial expressions. Pre-training on large datasets like ImageNet allows these models to leverage learned representations to improve performance on emotion-specific datasets, such as Kaggle's emotion dataset. Adjusting learning rates further optimize accuracy, achieving as much as 98.7% accuracy at a learning rate of 0.0001. Deep learning architectures, including CNNs for image data and multimodal networks for integrating multiple data sources, excel in capturing intricate patterns across various modalities, advancing the accuracy and generalization of emotion detection systems without manual feature engineering.

3. Proposed Methodology

The machine learning method, which uses a convolutional neural network (CNN), is implemented for training and testing the AI model in Figures 1 and 2.

Figure 1 shows a facial emotion recognition process. The input frame, containing images of faces, undergoes pre-processing to enhance and prepare the data. Subsequently, feature extraction identifies crucial characteristics from the processed images. These features are then classified into various categories, leading to the final predicted result—in this example, identification of the "Happy" emotion.

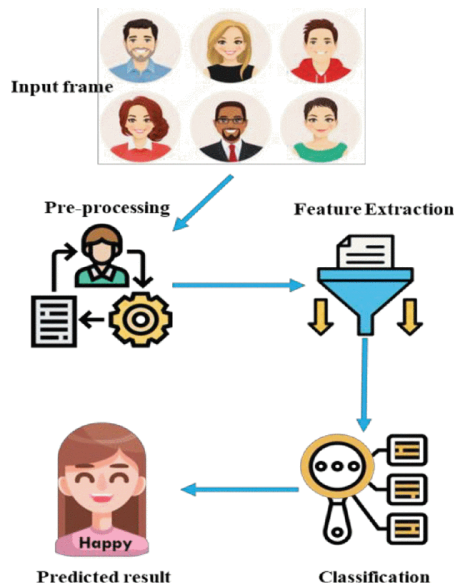


Figure 1. Flow process of training and testing the face emotion detection AI model using machine learning methods

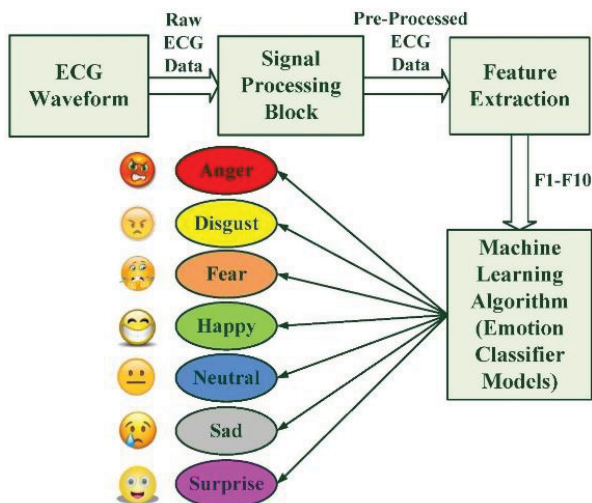


Figure 2. Block diagram of training and testing the ECG emotion detection AI model with machine learning methods

Figure 2 shows an emotion recognition process using ECG data. It begins by capturing the ECG waveform, which is then converted into raw ECG data. This data passes through a signal processing block, resulting in pre-processed ECG data. Feature extraction is performed on this data, generating features, labeled as “F1 score.” The features are then fed into a machine learning algorithm that classifies the data into one of seven emotions: happy, sad, neutral, angry, afraid, disgusted, and surprised.

3.1. Training and Testing the Face Emotion Detection AI Model Using Machine Learning Methods

3.1.1. Training the Face Emotion Detection AI Model

Step 1: Importing Libraries

Import various libraries for data manipulation, numerical operations, plotting, and interaction with

the OS, such as Matplotlib (plotting), NumPy (numerical operations), Pandas (data manipulation), Seaborn (statistical visualization), and ‘os’ (OS interaction). Import necessary deep learning libraries from Keras, including tools for image loading and conversion (load_img, img_to_array), image augmentation (Image Data Generator), building neural network layers (Dense, Input, Dropout), defining network architecture (Model, Sequential), and optimization algorithms (Adam, SGD, RMSprop).

Step 2: Displaying Images

Download the face expression recognition dataset from Kaggle. Define picture_size as 48 for input images, and set folder_path to the dataset directory. This prepares the code to work with 48x48 pixel images from the specified folder for facial expression recognition. Set expression to ‘happy’ and create a 12x12-inch plot for visualizing images related to the ‘happy’ expression. Load and display a 3x3 grid of images from the dataset using load_img and plt.imshow.

Step 3: Making Training and Validation Data

Set batch_size to 128. Create instances of Image Data Generator for both training and validation datasets, configuring data augmentation for training. Generate batches of training and validation data using flow_from_directory, specifying image directories, target size, color mode, batch size, class mode, and data shuffling.

Step 4: Model Building

Import optimizers from Keras and set no_of_classes to 7. Initialize a sequential model, adding convolutional layers with various filters and sizes, followed by batch normalization, ReLU activation, max-pooling, and dropout. Flatten the output before fully connecting layers. Add fully connected layers with 256 and 512 neurons, including batch normalization, ReLU activation, and dropout. Add output layers corresponding to the number of classes with softmax activation. Compile the model using the Adam optimizer with accuracy as the metric and categorical_crossentropy as the loss function. Print the model summary.

Step 5: Fitting the Model with Training and Validation Data

Import Keras optimizers and callbacks. Define the checkpoint for saving the best model, EarlyStopping for stopping training with no improvement, and ReduceLROnPlateau for lowering the learning rate when needed. Set the number of epochs to 48. Assemble the model and start training using fit_generator.

Step 6: Plotting Accuracy and Loss

Set the plot style and create a new figure for accuracy and loss plots. Create subplots to display both loss and accuracy for training and validation.

3.1.2. Training the Facial Emotion Detection AI Model

Step 1: Importing Libraries

Import various libraries for a deep learning model using Keras and OpenCV. These include load_model for loading pre-trained models, sleep for delays, img_to_array for converting images to arrays, image for image preprocessing, cv2 for computer vision tasks, and numpy for numerical operations.

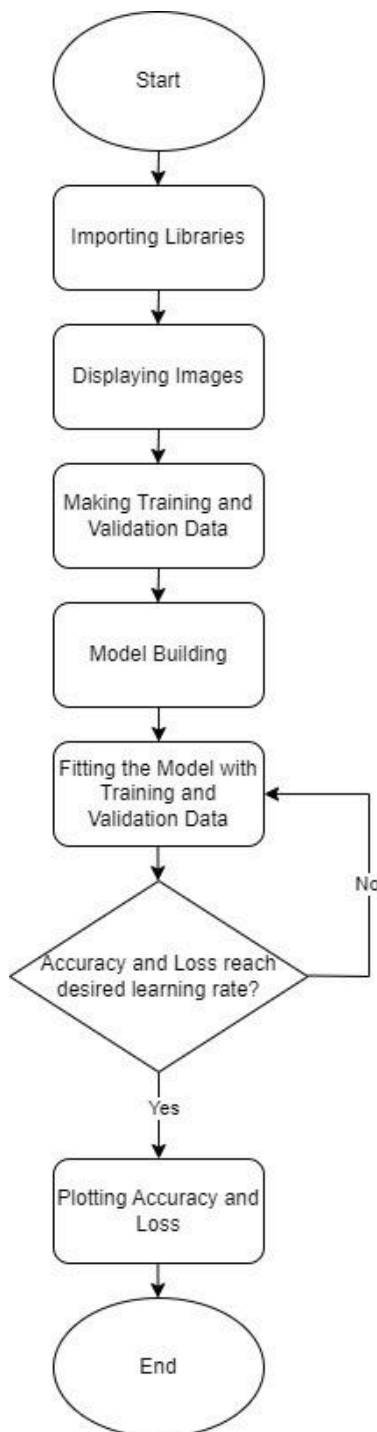


Figure 3. Overall flowchart of facial emotion AI model training process

Step 2: Initialization of AI Model System

Initialize a face cascade classifier using a Haar Cascade XML file, which loads a pre-trained emotion classification model in HDF5 format, defines a list of emotion labels, and initializes a video capture object to start capturing video frames from the default camera.

Step 3: Setting up a Graphical Window for Displaying the Probabilities

Set up a graphical window to display different emotions' probabilities by defining the window dimensions, creating a named window called 'Probabilities' with cv2.namedWindow, and resizing it to the specified dimensions using cv2.resizeWindow.

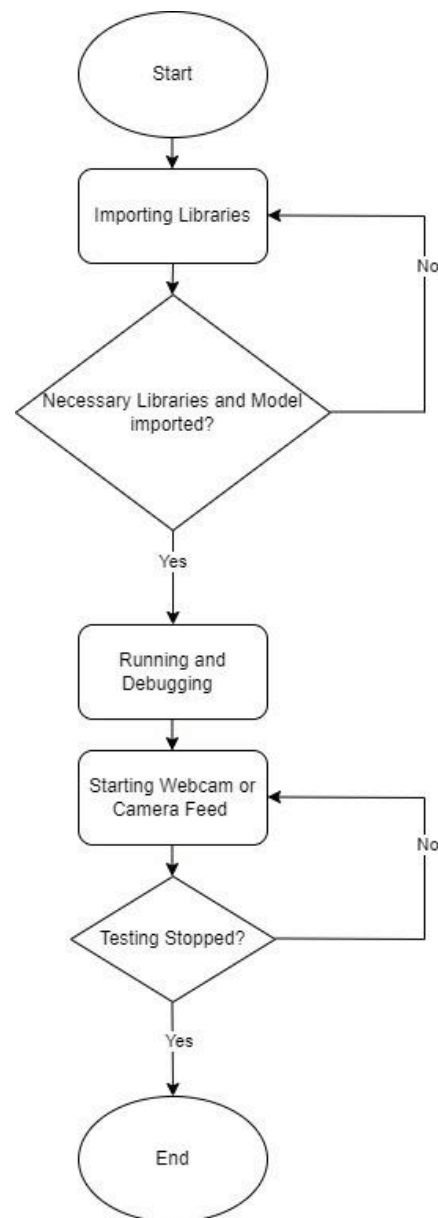


Figure 4. Overall flowchart of facial emotion AI model testing process

Step 4: Running and Debugging the AI Model System

The AI model system runs a real-time video processing loop where it detects faces using the pre-trained face cascade classifier.

3.2. Testing the ECG Emotion Detection AI Model Using Machine Learning Methods

3.2.1. Training the ECG Emotion Detection AI Model

Step 1: Importing Libraries

Import various libraries, including glob, pandas, numpy, os, wget, github, tensorflow.keras, scikit-learn, matplotlib.pyplot, and scipy modules (signal, ndimage, stats, interpolate, and integrate). These libraries provide a wide range of functions for data management, as well as machine learning model creation, evaluation, and signal analysis.

Step 2: Data Processing

Process ECG data files from specified locations by selecting relevant segments and combining metadata

from external dataframes. The code cycles through files, retrieves data up to a predefined length, and parses filenames for session, participant, and video IDs. It filters annotations by comparing IDs with meta-data rows, creating database entries with combined ECG data, participant characteristics, and emotional ratings. The processed data is then transformed into a pandas DataFrame, with missing values filled and columns renamed for clarity.

Step 3: Data Visualization

The data will be categorized into one of the seven emotions—happy, sad, neutral, angry, afraid, disgusted, and surprised— using both self-reported and target emotion labels. A custom plotting function iterates through the filtered data, plotting ECG signals for each category and providing a visual representation of ECG signal variations associated with different emotional states.

Step 4:

Create and train a neural network model for emotion identification based on ECG data using TensorFlow and Keras. The data is preprocessed, organized into features and target labels, and standardized. A sequential neural network model is defined with several dense layers, and the model is compiled with categorical cross-entropy loss and the Adam optimizer. Training is performed over 100 epochs with ModelCheckpoint and EarlyStopping callbacks to save the best model and prevent overfitting. The best model is then saved in the HDF5 format.

Step 5: Model Evaluation

Evaluate the pre-trained neural network ECG emotion detection model using a specific test dataset. The `model.evaluate()` function is used to compute performance metrics, such as accuracy and loss values. The results are stored in a variable and printed, providing a brief overview of the model's predictive performance on unseen test data.

3.2.2. Testing the ECG Emotion Detection AI Model

Step 1: Monitoring Heart Rate in Arduino

Interface with a MAX30100 pulse oximeter sensor to monitor heart rate by establishing serial communication, initializing the sensor, updating readings, and printing the heart rate to the serial monitor every second in Arduino.

Step 2: Real-Time Visualization of Data from Arduino to PyCharm

Capture real-time heart rate data from an Arduino via a serial port, visualize it using Matplotlib, and save the data to a CSV file. The Arduino used to sets up a serial connection (COM4, baud rate 115200) and initializes a Matplotlib plot for dynamic heart rate display. The `read_and_process_data()` function continuously reads serial port lines, extracts valid heart rate values (60-100 bpm), and appends them to `heart_rate_data`. The `update_plot()` function, called by Matplotlib's animation framework every second, updates the plot with the latest readings. At the end of the heart rate measurement, the data is saved to 'heart_rate_data.csv', and this enables real-time monitoring and logging of heart rate data.

Step 3: ECG Testing

Test the AI model for ECG emotion recognition in Pycharm, importing necessary libraries (NumPy, pandas, TensorFlow's Keras, scikit-learn's StandardScaler, and Matplotlib). The `preprocess_ecg` function normalizes ECG data using StandardScaler. Load a pre-trained Keras model (`ecg_emotion_recognizer.h5`) for emotion recognition. ECG data is read from 'heart_rate_data.csv' to simulate real-time data, and the ECG data is normalized using `preprocess_ecg`. The `extract_features` function, which returns the normalized ECG data, is defined for feature extraction. The ECG signal is plotted using Matplotlib, and predictions are made by using the loaded model on extracted features. The predicted emotion label is then decoded and printed.

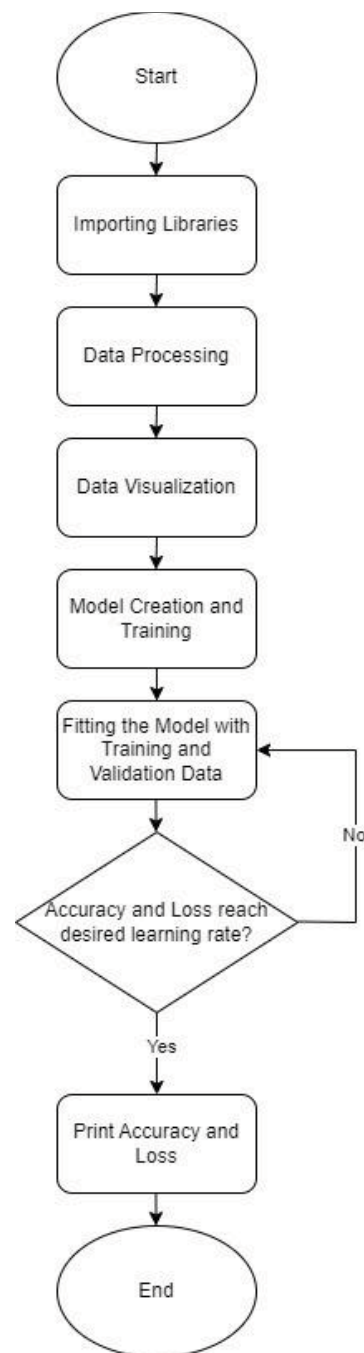


Figure 5. Overall flowchart of ECG emotion AI model training process

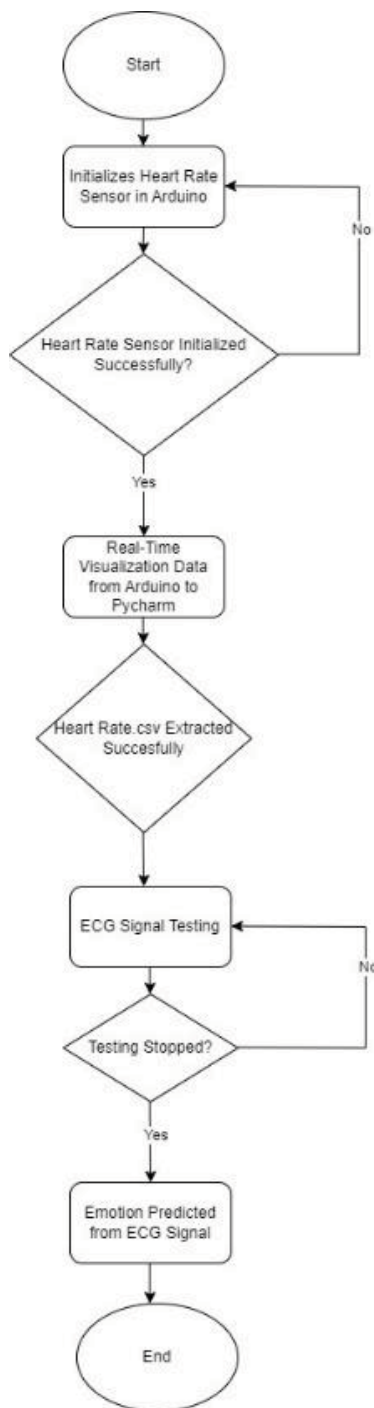


Figure 6. Overall flowchart of ECG emotion AI model testing process

3.3. Testing the Multimodal Emotion Detection AI Model

3.3.1. Testing Multimodal Emotion Detection AI Model Using PyCharm

Step 1: Monitoring Heart Rate in Arduino

Build a multimodal emotion detection AI model using OpenCV for face detection integrating with ECG emotion detection. Facial emotions are predicted using a Haar Cascade classifier, while ECG emotions are predicted from the preprocessed 'heart_rate_data.csv' using a deep learning model. Combined emotions are determined by matching ECG and face predic-

tions, or by defaulting to the ECG emotion if no match occurs. The final combined emotion is then printed for each prediction.

Step 2: Display Combined Emotion in Web Browser

Display a textual description and an emoji in the web browser by the link given, based on the combined emotion from the multimodal emotion detection model.

4. Result with Discussion

In this section, the initial findings from the process of training and testing the multimodal AI model in PyCharm will be discussed.

4.1. Preliminary Results of the Facial Emotion AI Model

Accuracy is a key metric for evaluating convolutional neural networks (CNNs) during both training and validation stages. Training accuracy measures how well the model classifies samples from the training dataset, while validation accuracy assesses the model's ability to generalize to new, unseen data. Both training and validation accuracy increased across epochs, indicating the model's improving performance. The training accuracy reached 72.21%, slightly higher than the validation accuracy, which was 60.58%. A significant gap between these values could indicate overfitting, where the model performs well on training data but fails to generalize to new data. The training process, however, employs an early stopping technique to avoid overfitting.

Loss is a crucial metric for evaluating the performance of convolutional neural networks (CNNs) during the training and validation stages. Training loss measures the discrepancy between the model's predictions and the target values in the training dataset, while validation loss assesses this discrepancy using a separate validation dataset.

The training loss decreased across epochs, indicating that the model's adjustments to its parameters were improving its predictions. The validation loss also decreased, suggesting the model is learning to make accurate predictions on new, unseen data. However, the training loss (73.65%) was significantly lower than the validation loss (107.53%), which could indicate overfitting. This was mitigated, however, by an early stopping technique.

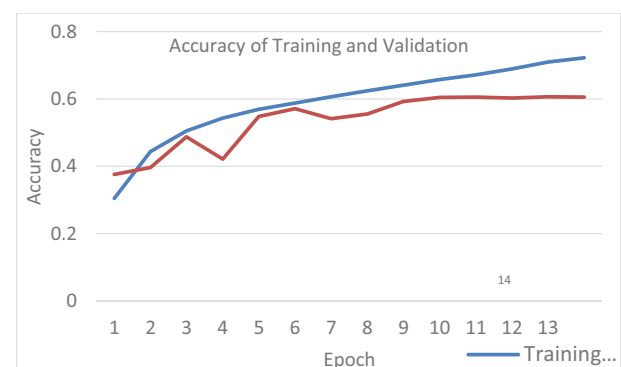


Figure 7. Accuracy of training and validation of facial emotion AI model

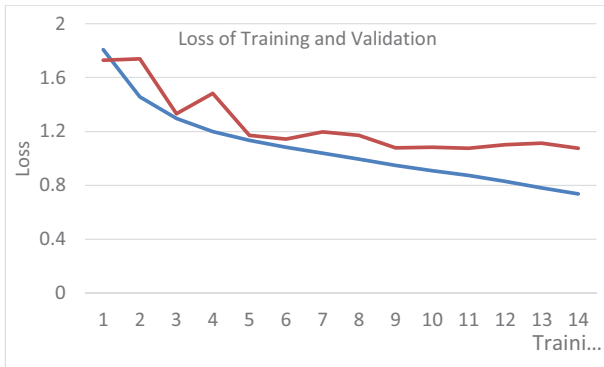


Figure 8. Loss of training and validation of facial emotion AI model

Table 1. Results of experiment.

Emotion	Probability (%)
Happy	98.63
Neutral	56.60
Sad	39.58
Surprise	77.86
Fear	26.36
Angry	41.34
Digust	27.19

Based on Table 1, the highest probability is that the face expression detected by the AI model is “Happy” (98.63%), while the lowest probability is that the face expression detected by the AI model is “Fear” (26.36%). This is because fear often involves subtle facial expressions that can be difficult to capture accurately, especially in real-time or non-controlled environments. Hence, the probability of fear is the lowest, since fear is only expressed in subtle changes like widened eyes or tensed lips, which might be less noticeable, while the “happy” emotion often involves broad smiles and visible changes in facial muscles that are much more noticeable.

4.2. Preliminary Results of the ECG Emotion AI Model

In Figure 9, the model’s training accuracy is 26.47%, slightly lower than the validation accuracy of 28.75%, which suggests no overfitting, but indicates poor performance because both accuracies are only around 20%. This low accuracy can be attributed to insufficient and low-quality data, leading to underfitting and ineffective recognition of ECG signals. To enhance the model’s performance, it is crucial to collect more and higher-quality data.

In Figure 10, the training loss is 197.367%, which is slightly higher than the validation accuracy of 193.61%. As with the accuracy of training and validation of ECG emotion AI model, the high loss percentage is seen as indicating a poor model. Hence, it is necessary to collect more and better-quality data to improve the performance of the ECG emotion AI model, since the quantity of data used to train the AI model is smaller (only 155 ECG signals for 7 classes of emotions).

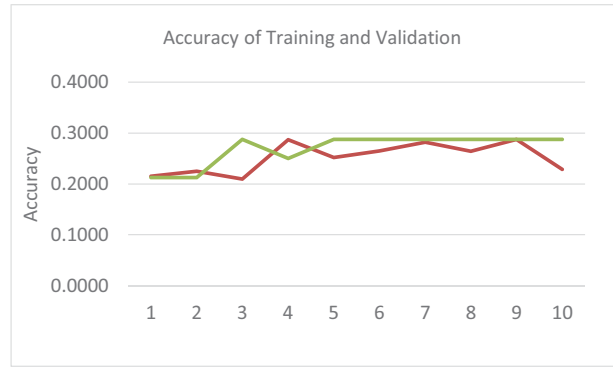


Figure 9. Accuracy of training and validation of facial emotion AI model



Figure 10. Accuracy of training and validation of facial emotion AI model

```

Decision Tree
print(classification_report(test_y, dt_pred_y))
              precision    recall  f1-score   support

 0.0         0.12         0.25         0.17         4
 1.0         0.33         0.30         0.32        10
 2.0         0.48         0.57         0.47         7
 3.0         0.55         0.48         0.51        23
 4.0         0.46         0.35         0.40        17
 5.0         0.18         0.50         0.27         4
 6.0         0.00         0.00         0.00         8
 7.0         0.00         0.00         0.00         7

 accuracy          0.34         80
 macro avg         0.26         0.31         0.27         80
 weighted avg      0.35         0.34         0.33         80
    
```

Figure 11. Accuracy of training and validation of facial emotion AI model

4.3. Performance Measures of the ECG Emotion AI Model

4.3.1. F1-Score

The F1-scores can be calculated by using equation (1) below:

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1}$$

4.3.2. Decision Tree

The decision tree model for the ECG signal emotion detection has an overall accuracy of 34%, with precision ranging from 0.00 to 0.55, recall from 0.00 to 0.57, and F1-scores from 0.00 to 0.51, indicating inconsistent performance across different emotions. Classes 0.0 and 5.0 show low metrics due to fewer instances, while classes 6.0 and 7.0 have zero precision, recall, and F1-scores. The macro average precision, recall, and F1-scores are 0.26, 0.31, and 0.27, respectively, with slightly higher weighted averages of 0.35, 0.34, and 0.33, highlighting the impact of unbalanced data.

4.3.2. Random Forest

The random forest model for ECG signal emotion detection has an overall accuracy of 39%, with precision levels ranging from 0.00 to 0.59; recall ranging from 0.00 to 0.57; and F1-scores ranging from 0.00 to 0.58. This indicates inconsistent performance across different emotions. Classes 0.0 and 5.0 show low metrics due to fewer instances, while classes 1.0 and 6.0 have zero precision, recall, and F1-scores. The macro average precision, recall, and F1-score are 0.26, 0.35, and 0.29, respectively, with slightly higher weighted averages of 0.34, 0.39, and 0.35, underscoring the impact of unbalanced data.

Based on the comparison between decision tree and random forest for an ECG emotion AI model, referring to Figures 10 and 11, the random forest outperforms the decision tree, achieving a higher total accuracy of 39% compared to 34% for the decision tree. The random forest also exhibits superior weighted averages for precision (0.35 vs. 0.34), recall (0.39 vs. 0.34), and F1-score (0.35 vs. 0.33), indicating better performance across various emotion classes and in managing imbalanced data. Therefore, the random forest model proves more effective and consistent in classification tasks for ECG-based emotion recognition. Further tuning, additional data, and/or alternative modeling methods will be required to improve the AI model.

4.4. Preliminary Results of the Multimodal Emotion AI Model

Figure 13 shows the output in a web browser that includes the combined emotion detected from the multimodal emotion AI model, which is "Happy". It also provides some advice for each emotion. "Happy," for example, has the following text: "Good job! Maintain your positive outlook and keep up the great work."

Random Forest

```
print(classification_report(test_y,fpred_y))
```

	precision	recall	f1-score	support
0.0	0.22	0.50	0.31	4
1.0	0.00	0.00	0.00	10
2.0	0.40	0.57	0.47	7
3.0	0.59	0.57	0.58	23
4.0	0.43	0.53	0.47	17
5.0	0.22	0.50	0.31	4
6.0	0.00	0.00	0.00	8
7.0	0.25	0.14	0.18	7
accuracy			0.39	80
macro avg	0.26	0.35	0.29	80
weighted avg	0.34	0.39	0.35	80

Figure 12. Classification report using random forest



Figure 13. Display combining emotion and advice in the web browser

5. Conclusion

In this study, a multimodal emotion detection system was successfully designed for educational and work environments using convolutional neural networks (CNN). The first objective, involving training and testing the AI model with Python and OpenCV, showed that the model accurately recognized facial emotions from seven classes (happy, sad, neutral, angry, afraid, disgusted, and surprised). The second objective focused on developing the system by integrating OpenCV and ECG signals. The AI model also can detect emotions from seven classes by recognizing the ECG signals. The final objective is validating the multimodal AI system, which accurately combines and displays the results, based on facial and ECG predictions, on a web application.

AUTHORS

Wan Mohd Bukhari Daud* – Faculty of Artificial Intelligence & CyberSecurity at Universiti Teknikal, Malaysia Melaka, Durian Tunggal, 76100, Malaysia. He can be contacted at bukhari@utem.edu.my.

Adnan Kiral – Assistant Professor in the Department of Civil Engineering at Recep Tayyip Erdogan University, Fener, Rize, TR53100, Türkiye. He can be contacted at email adnan.kiral@erdogan.edu.tr

Mohamed Osman Tokhi – A Professor of Engineering at University of London South Bank, London, United Kingdom. He can be contacted at tokhim@lsbu.ac.uk

Lee Chung Yee – Fakulti Teknologi & Kejuruteraan, Universiti Teknikal Malaysia Melaka, Durian Tunggal, 76100, Malaysia, lcyee@student.utem.edu.my

Muhammad Muzhafar Mohammad Zawawi – Fakulti Teknologi & Kejuruteraan, Universiti Teknikal Malaysia Melaka, Durian Tunggal, 76100, Malaysia, m112420023@student.utem.edu.my

ACKNOWLEDGEMENT

Gratitude for Progress Catalysts. We extend our heartfelt appreciation to the Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia Melaka (UTeM) for their invaluable financial support. Together, we drive transformative research endeavors.

References

- [1] M. Cui, J. Fang, and Y. Zhao, "Emotion Recognition Of Human Body's Posture In Open Environment," *Proc. 2020. Chinese Control and Decision Conference (CCDC)*, 2020, pp. 3294–3299.
- [2] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova, "Emotion Recognition From Body Movement," *IEEE Access*, vol. 8, 20210 pp. 11761–11781, 2020.
- [3] T. Kusunose, X. Kang, K. Kiuchi, R. Nishimura, M. Sasayama, and K. Matsumoto, "Facial Expression Emotion Recognition Based on Transfer Learning and Generative Model," *Proc. 8th International*

- Conference on Systems and Informatics (ICSAI)*, 2022, pp. 1–6.
- [4] D. Shehada, A. Turkey, W. Khan, B. Khan, and A. Hussain, "A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People," *IEEE Access*, vol. 11, 2023, pp. 36961–36969.
- [5] A. S. Utane and S. L. Nalbalwar, "Emotion Recognition Through Speech," *Proc. 2nd National Conference on Innovative Paradigms in Engineering and Technology (NCIPET)*, 2013.
- [6] J. Chen, "Speech Emotion Recognition Based On Convolutional Neural Network," *Proc. International Conference on Networking, Communications and Information Technology (NetCIT)*, 2021, pp. 106–109.
- [7] X. Li, R. Yue, W. Jia, H. Wang, and Y. Zheng, "Recognizing Students' Emotions based on Facial Expression Analysis," *Proc. 11th International Conference on Information Technology in Medicine and Education (ITME)*, 2021, pp. 96–100.
- [8] G. Assunção, B. Patrão, M. Castelo-Branco, and P. Menezes, "An Overview of Emotion in Artificial Intelligence," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, 2022.
- [9] J. M. Foo, S. S. Anhum, S. Islam, and K. H. Keoy, "Facial Emotion Recognition System for Mental Stress Detection among University Students," *Proc. 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2023, pp. 1–6.
- [10] G. Shi, "Technical Analysis of Machine Learning Algorithms in Artificial Intelligence Image Recognition," *Proc. 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 2023, pp. 1–5.
- [11] S. Ashrita and T. Padmass Machine Learning for Automation Software Testing: Challenges, Use Cases Advantages & Disadvantages," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 9, 2020, pp. 1301–1307.
- [12] K. Vinutha, M. K. Niranjani, J. Makhijani, B. Natarajan, V. Nirmala, and T. R. Vijaya Lakshmi, "A Machine Learning based Facial Expression and Emotion Recognition for Human Computer Interaction through Fuzzy Logic System," *Proc. International Conference on Inventive Computation Technologies (ICICT)*, 2023, pp. 166–173.
- [13] I. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, "Challenges in Representation Learning: A Report on Three Machine Learning Contests," *Neural Networks*, vol. 64, 2013, doi: 10.1016/j.neunet.2014.09.005.
- [14] T. Ergin, M. A. Ozdemir, and A. Akan, "Emotion Recognition with Multi-Channel EEG Signals Using Visual Stimulus," *Proc. Medical Technologies Congress (TIPTEKNO)*, 2019, pp. 1–4.
- [15] L. Huang, K. Yu, T. Song, H. Wang, F. Yuan, Y. Zhang, and H. Yang, "Research on Emotion Recognition Based on Multisource Signals," *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, 2023, pp. 280–284.
- [16] Itai Gat, Hagai Aronowitz, Weizhong Zhu, Edmilson Morais, Ron Hoory, "Speaker Normalization for Self-Supervised Speech Emotion Recognition," *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 7342–7346.
- [17] R. Sato, R. Sasaki, N. Suga, and T. Furukawa, "Creation and Analysis of Emotional Speech Database for Multiple Emotions Recognition," *Proc. 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, 2020, pp. 33–37.
- [18] P. Yue, L. Qu, S. Zheng, and T. Li, "Multi-task Learning for Speech Emotion and Emotion Intensity Recognition," *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2022, pp. 1232–1237.
- [19] M. U. Khan, M. A. Abbasi, Z. Saeed, M. Asif, A. Raza, and U. Urooj, "Deep Learning Based Intelligent Emotion Recognition and Classification System," *Proc. International Conference on Frontiers of Information Technology (FIT)*, 2021, pp. 25–30.
- [20] R. Zhan, J. Lu, Y. Jiang, T. Li, G. Zhong, and W. Lv, "A Multimodal Deep Learning Approach for Typhoon Track Forecast by Fusing CNN and Transformer Structures," *Proc. 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, pp. 391–394.
- [21] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review Of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions," *Journal of Big Data*, vol. 8, 53.
- [22] H. Liu, M. Cocea, and W. Ding, "Decision Tree Learning Based Feature Evaluation and Selection for Image Classification," *Proc. International Conference on Machine Learning and Cybernetics (ICMLC)*, 2017, pp. 569–574.
- [23] F. Aaboub, H. Chamlal, and T. Ouaderhman, "Analysis of the Prediction Performance of Decision Tree-Based Algorithms," *Proc. International Conference on Decision Aid Sciences and Applications (DASA)*, 2023, pp. 7–11.
- [24] E. Alpaydin, *Introduction to Machine Learning*, MIT Press, 2014.

- [25] T. Ghirmai, "Generating Laplace process with desired autocorrelation from Gaussian AR processes," *Proc. IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)*, 2015, pp. 113–117.
- [26] L. Yang, C. Heiselman, J. G. Quirk, and P. M. Djurić, "Class-Imbalanced Classifiers Using Ensembles of Gaussian Processes and Gaussian Process Latent Variable Models," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 3775–3779.
- [27] Y. Li, S. Rao, A. Hassaïne, R. Ramakrishnan et al., "Deep Bayesian Gaussian Processes For Uncertainty Estimation in Electronic Health Records," *Scientific Reports*, vol. 11, no. 1, 20685, 2021.
- [28] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A Review of Emotion Recognition Using Physiological Signals," *Sensors*, vol. 18, no. 2018, 2074, doi: 10.3390/s18072074.