Human Facial Emotion Recognition for Adaptive Human Robot Collaboration in Manufacturing

Fahad Khan[1][0009-0008-3660-2050], Seemal Asif[1][0000-0001-7048-0183] and Phil Webb[1][0000-0002-5789-4207]

[1] Centre for Robotics and Assembly, Cranfield University, Cranfield MK43 0AL, UK
fahad.khan@cranfield.ac.uk

Abstract. The integration of robots into various industries, including manufacturing, has introduced new challenges in achieving efficient human-robot collaboration. A crucial aspect of successful collaboration is the ability of robots to understand and respond to human emotions. In the context of human-robot collaboration in manufacturing, accurately predicting human emotions is essential for enhancing efficiency and safety. This paper presents a setup for human emotion detection, focusing on facial emotion recognition. The proposed model and descriptive summary involve the utilising state-of-the-art algorithms such as AlexNet, HaarCascade (HCC), MTCNN (Multi-Task Cascaded Convolutional Neural Networks), and SVM (Support Vector Machine), applied to datasets like CK+, JAFFE, and AffectNet. The performance of each facial recognition model is evaluated in real-time scenarios, resulting in significant progress with an accuracy improvement from 40% to 78.1%. These results demonstrate the effectiveness of the approach in enabling adaptive robot control based on human emotions and enhancing collaboration quality. This research uniquely integrates facial emotion recognition and robot control to enable adaptive responses during human-robot collaboration in manufacturing settings. By understanding and responding to human emotions, robots can improve their interactions with humans, leading to increased productivity and improved overall collaboration efficiency.[1]

Keywords: human facial recognition, human – robot collaboration, human emotion prediction, adaptive control, machine learning

1 Introduction

Human-robot collaboration (HRC) has become increasingly important in smart manufacturing, where human-centricity, sustainability, and resilience are critical factors [1]. The ability to detect and interpret human emotions allows robots to interact more intuitively, genuinely, and naturally with their human counterparts, leading to improved productivity and efficiency. Emotion detection plays a crucial role in HRC and manufacturing settings [2], [3]. It enables robots to understand and respond to human emotions, leading to more effective and intuitive interactions. Recognising human emotions is essential for creating adaptive and responsive systems that can enhance productivity, safety, and overall user experience [4].

[1] If EquinOCS, our proceedings submission system, is used, then the disclaimer can be provided directly in the system.

Humans communicate emotions through various modalities, including verbal cues like speech and non-verbal cues like facial expressions, eye gaze, gestures, and physiological signals [5], [6]. Additionally, physiological signals such as heart rate monitoring (HRM), electroencephalography (EEG), electrocardiography (ECG), galvanic skin response (GSR), heart rate variability (HRV), respiration rate analysis (RR), skin temperature measurements (SKT), Electromyogram (EMG), Blood Volume Pulse (BVP), and electrooculography (EOG) can also provide insights into emotional states [7], [8].

Facial emotion recognition is a widely explored method for detecting human emotions. It involves analysing facial expressions to identify the underlying emotional state, with a particular focus on key areas such as the eyes, eyebrows, and mouth, which exhibit unique expressions for each emotion [9] –[13]. Existing APIs, such as FER and DeepFace uses HCC and MTCNN, for facial emotion recognition. Moreover, models trained on datasets like CK+, JAFFE, and AffectNet using techniques like AlexNet can improve the accuracy of emotion detection [10]–[13].

In the context of human-robot collaboration in manufacturing, facial emotion recognition can be applied to control the behaviour of robots. By detecting and interpreting human emotions from facial expressions, robots can adjust their actions and responses accordingly. For example, a UR5 robot's speed can be controlled based on the recognised emotions, allowing for adaptive and personalised interactions with human workers.

The aim of this paper is to explore the significance of human emotion recognition in human-robot collaboration within the manufacturing domain. It will discuss various ways of detecting emotions, including non-verbal cues and physiological signals, with a specific focus on facial emotion recognition and its application in controlling a UR5 robot. A core novelty explored here is the application of facial emotion recognition to modulate robot behavior, specifically speed, based on detected emotions. This adaptive human-robot collaboration paradigm has not been extensively implemented before.

This paper will present a descriptive summary of facial recognition models and datasets, details on the proposed AlexNet model and robot control. It further includes a comparative analysis of models, an in-depth discussion, and a summary of findings, followed by an outline of future research directions.

2 Method

2.1 Emotion Recognition System

Emotion recognition systems automate the identification of human emotions through data from sources like facial images, speech, and physiological sensors. They extract features and employ machine learning to classify emotions. This study focuses on facial emotion recognition, analysing facial cues for emotion prediction. When integrated with robotic systems, the predicted emotions can enable adaptive responses and human-robot collaboration. The following subsections detail key components of the facial emotion recognition system evaluated in this study [2], [4].

Datasets.
FER2013, introduced in the ICML 2013 Challenges in Representation Learning, is comprised of approximately 35,887 facial images. These images are categorised into seven distinct emotion classes: Anger, Disgust, Fear, Happy, Sadness, Surprise, and Neutral. The dataset is partitioned into three sets: training, development, and testing. With a resolution of 48x48 pixels, FER2013's images often exhibit variations encountered in real-world settings. It was created using Google's image search API and includes diverse images, incorporating factors like occlusion, partial faces, low contrast, and eyeglasses [14].

The CK+ dataset stands as a valuable resource for action units and emotion recognition research. Comprising 593 video sequences, it encompasses expressions of six basic emotions (Anger, Contempt, Disgust, Fear, Happy, and Sadness) as well as a neutral expression. This dataset features a diverse range of subjects, spanning different ages from 18 to 50 years. The CK+ dataset provides a combination of posed and spontaneous expressions, rendering it an integral component in understanding facial emotion recognition [11].

Containing 213 grayscale images, the JAFFE dataset captures seven different facial expressions posed by ten Japanese female models. The dataset incorporates emotion labels such as Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. These images offer insight into the nuances of facial expressions, especially within the context of distinct cultural representations [12].

The AffectNet dataset, containing one million facial images, is sourced from the internet through 1250 emotion-related keywords and three search engines across six languages. Approximately 420,000 images underwent manual annotation for eleven facial expressions and emotions, ranging from Neutral to No-Face. Valence and arousal intensity were measured using a dimensional model. Additionally, 550,000 images were automatically annotated with a ResNext Neural Network, achieving 65% accuracy [10].

Machine Learning Algorithms for Facial Emotion Prediction.
Facial Emotion Prediction Algorithms are essential tools for interpreting emotional states from facial expressions. In our study, we utilise algorithms including Haarcascade, Convolutional Neural Networks (CNN), MTCNN, Support Vector Machines (SVM), and AlexNet to accurately predict emotions from facial features. These algorithms are trained on datasets such as FER2013, CK+, JAFFE, and AffectNet to improve their ability to capture subtle emotional cues. Among these algorithms, we utilise existing models based on them, and specifically develop an AlexNet model.

HCC (Haarcascade) is a popular face detection algorithm that utilises Haar-like features and a cascade classifier to detect faces in images or video streams. It is based on the Viola-Jones algorithm and is known for its simplicity and efficiency. Haarcascade works by scanning an image with a sliding window and applying a series of classifiers to identify regions that resemble faces based on specific features such as edges, lines, and textures. It has been widely used in various applications, including facial emotion recognition, as an initial step to detect and localise faces in an image or video [15].

CNN is a deep learning algorithm that has shown remarkable success in various computer vision tasks, including facial emotion recognition. CNNs are designed to automatically learn and extract relevant features from images through convolutional layers, pooling layers, and fully connected layers. In the context of facial emotion recognition, CNNs can be trained on labelled datasets to learn discriminative features that capture facial expressions and emotions. The network architecture and parameters are optimised through a training process to improve the accuracy of emotion prediction [1], [4].

MTCNN is a face detection and alignment algorithm that consists of three stages: face detection, facial landmark localisation, and face alignment. It uses a cascade of convolutional networks to detect faces and then refines the bounding boxes and estimates facial landmarks. MTCNN is known for its robustness and accuracy in detecting and aligning faces, making it suitable for facial emotion recognition tasks.

SVM is a supervised machine learning algorithm that can be used for classification tasks, including facial emotion recognition. SVMs aim to find an optimal hyperplane that separates different classes by maximising the margin between them. In the context of facial emotion recognition, SVMs can be trained on labeled datasets with extracted features from facial images to classify emotions based on the learned patterns [16].

AlexNet is a deep convolutional neural network architecture that gained significant attention after winning the ImageNet Large Scale Visual Recognition Challenge in 2012. It consists of multiple convolutional layers, pooling layers, and fully connected layers. Alexnet was trained in fusion with SVM. To train this, dataset collection was selected as manually annotated images from AffectNet data and trained for seven emotions instead of its eleven categories. Fig. 1 provides overview of emotion recognition model (HCC or MTCNN).

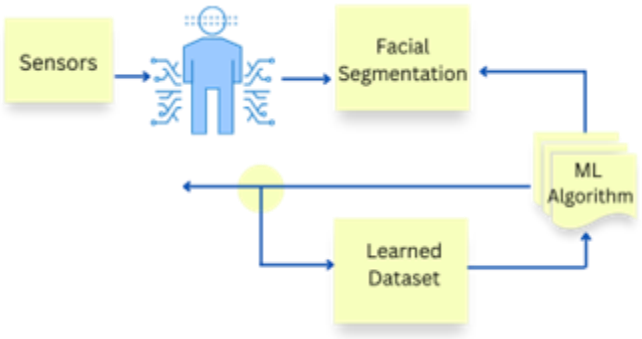


Fig. 1. Schematic of ML Model.

2.2 Proposed Model

The AlexNet architecture was chosen as the basis for our emotion recognition model due to its proven success in various computer vision tasks. Leveraging its feature extraction capabilities through transfer learning, we customised the architecture to suit our emotion recognition task.

Transfer Learning and Fine-Tuning: Transfer learning involves using a pre-trained neural network model as a starting point and adapting it to a specific task. In our case, we employed the AlexNet architecture, pre-trained on a large image dataset, as the foundation. This approach offers several benefits, including faster convergence and improved generalisation, especially when working with limited data.

Data Collection and Preprocessing - The datasets underwent preprocessing steps to enhance the quality and consistency of the data. Preprocessing techniques include image resizing, normalisation, and augmentation to account for variations in lighting conditions, pose, and facial expressions. These steps ensure that the data is suitable for training and testing the facial emotion recognition model. The image is then converted to a 3-channel image by replicating the grayscale image across all three-color channels. The processed image is saved with a new filename in the corresponding subfolder and the .tiff file extension.

Architecture Overview:
The AlexNet architecture is characterised by its depth, consisting of eight layers in total. These layers can be grouped into five convolutional layers, followed by three fully connected layers as shown in Fig. 2.
Input Layer : The architecture begins with an input layer that accepts RGB images of size 227x227 pixels. The three colour channels (red, green, and blue) capture visual information from the images.
Convolutional Layers : The five convolutional layers perform feature extraction. Each layer is followed by a rectified linear unit (ReLU) activation function, introducing non-linearity. These layers apply convolutional filters to the input image, detecting patterns and textures of increasing complexity.
Max Pooling Layers : After the first two convolutional layers, max-pooling layers downsample the spatial dimensions of the feature maps. This reduces computational load while retaining essential features.
Fully Connected Layers : Following the convolutional and pooling layers, three fully connected layers aggregate high-level features for classification. The first fully connected layer ('fc6') contains 4096 neurons, followed by a dropout layer that mitigates overfitting. The second fully connected layer ('fc7') also has 4096 neurons, further refining the features.
Softmax Layer : The architecture concludes with a softmax layer, transforming the output of the previous layer into probability scores for each emotion class. This layer computes the likelihood of the input image belonging to different emotion categories.
Custom Fully Connected Layer : To adapt the architecture to our emotion recognition task, we appended a custom fully connected layer at the end. This layer corresponds to the number of emotion classes in our dataset, ensuring that the network recognises emotions effectively. By retraining the added fully connected layer, the network learned to distinguish between emotional states based on facial expressions.

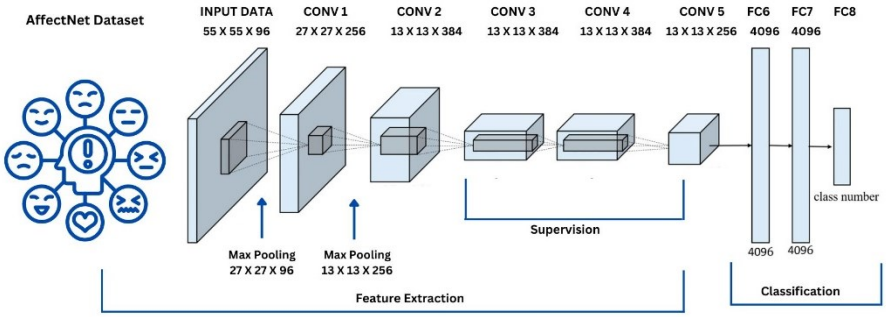


Fig. 2. Architecture of Alexnet comprising of eight layers of which five are convolution and three are fully connected.

Optimiser and Hyperparameters. : The Stochastic Gradient Descent with Momentum (SGDM) optimiser was chosen for training the model. This optimiser enhances convergence speed by incorporating momentum in the gradient updates. The initial learning rate, a key hyperparameter, was set to 0.0001 for controlled weight updates.
Mini-Batch Size and Epochs : A mini-batch size of 5 samples was selected for training. This mini-batch approach balances computational efficiency and model convergence. The model was trained over 10 epochs, meaning it iterated through the entire training dataset 10 times. This number was chosen to strike a balance between achieving reasonable accuracy and preventing overfitting.
Data Augmentation : To enhance the model's robustness, we applied data augmentation techniques during training. These techniques, including random cropping, flipping, rotation, and colour adjustments, introduce variability into the training data. This helps the model generalise better to unseen data and variations in facial expressions.
Dropout Layer : A dropout layer with a dropout rate of 0.5 was inserted after the first fully connected layer ('fc6'). Dropout introduces stochasticity by randomly deactivating a fraction of neurons during each forward and backward pass. This regularisation technique reduces overfitting by discouraging the network from relying on specific neurons.
Loss Function : During training, the categorical cross-entropy loss function was employed implicitly via 'sgdm' solver. Given the multi-class emotion classification task, this loss function shown in equation (1) measures the dissimilarity between predicted and actual class probabilities.

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i \, ln \, (\hat{y}_i) \tag{1}$$

Where:
N is the number of classes.
y_i is the true class probability for class i.
$\hat{y}_i$ is the predicted class probability for class i.
ln is the natural logarithm.

Early Stopping : Early stopping was not implemented in this case. However, it's worth noting that early stopping could be incorporated in future iterations to prevent overfitting by monitoring validation loss and halting training when it starts to increase.
Training Data Split : The dataset was divided into three subsets: training, validation, and testing. A ratio of approximately 70:30 was chosen for the training-validation split, with the training set comprising 70% of the data. This distribution ensures a sufficiently large training set to facilitate effective learning while allocating a substantial portion for validation to monitor the model's generalisation ability. The testing set, reserved for final evaluation, remained untouched during the training process.

Feature Extraction Approach:
Load Pretrained Network : To explore an alternative approach to emotion recognition, we leveraged the feature extraction capabilities of the AlexNet architecture. The pretrained AlexNet model was loaded, and a specific layer, 'fc7,' was selected for feature extraction. This layer is rich in high-level features, making it suitable for our task of emotion classification.
Feature Extraction and SVM : Utilising the 'fc7' layer as a feature extractor, activations were computed for both the training and validation sets. These activations served as the extracted features, capturing the essence of the facial expressions' emotional characteristics. Subsequently, these features were employed to train an SVM classifier.
SVM Classifier : The SVM classifier is a powerful tool for multi-class classification tasks. In our case, it maps the extracted features to the corresponding emotion classes. The classifier was trained using the extracted features from the training set, and predictions were made on the validation set.
Visualising Results : To gain insights into the SVM's performance, we visualised some sample test images along with their predicted labels. This allowed us to qualitatively assess the classifier's ability to recognise different emotions based on the extracted features.
Accuracy Calculation : The accuracy of the SVM classifier was calculated by comparing its predictions with the true labels of the validation set. This accuracy metric provides an objective evaluation of the classifier's performance in identifying emotional states from facial expressions.

2.3 Integration of Facial Recognition with Robot Control System

In terms of program and structure to control robot, initially it records a new face if selected to do so, after that recorded face is saved. Face is identified from the records (database), recognised face’s emotion is detected. Based on the emotion, if happy robot operates at maximum speed, if neutral robot operates at medium speed and slow if surprise, fear, sad and stops if anger or disgust.

Initially, the system prompts the user to record a new face. Upon selection, the camera captures the facial image, which is then pre-processed to ensure consistency in lighting and quality. The pre-processed face is saved in the records database for future recognition. During operation, the robot's camera captures live video of the human worker's

face. The recorded faces in the database are utilised for facial recognition. The live video feed is analysed using deep learning models. The system recognises the detected face by matching it with the saved face encodings. Simultaneously, the emotion recognition model predicts the emotion expressed on the recognised face. Based on the recognised emotion, the robot's control system adjusts its behaviour to align with the worker's emotional state.

For instance, if a "Happy" emotion is detected, the robot can operate at a brisk pace to match the positive energy. If the recognised emotion is "Neutral," the robot maintains a moderate speed for standard operations. Emotions like "Surprise," "Fear," or "Sadness" can lead to the robot slowing down to ensure safety and create a supportive environment. This controlling of robot is via socket communication directly without any need to run on controller, robot is controlled based on python commands which are send to controller.

The flowchart and system overview are presented in Fig. 3. The program begins by asking the user to choose between two options: first record new face where it records and save face for subsequent use, second option loads saved faces and proceeds to execute two concurrent threads in parallel.

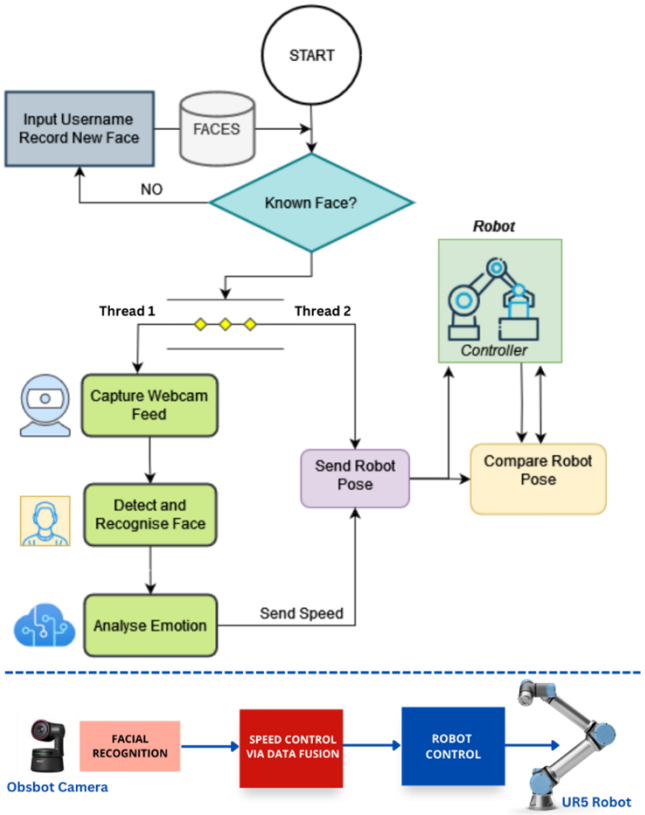


Fig. 3. Flowchart and an overview of the system.

Thread 1 performs real-time task by capturing webcam feed, performs facial recognition by using facial recognition algorithms to identify and track faces within the captured video. Analyses detected facial expressions to ascertain emotional states and displays annotated video. While thread 2 controls robot and moves to predefined poses and compares poses. The program continuously compares received poses to desired poses until a match is found. A communication mechanism exists between thread 1 and thread 2. When there is a change in the detected emotion within thread 1, this information is transmitted to thread 2. Specifically, thread 1 calculates a new speed value based on the detected emotion and communicates this updated speed information to thread 2. In essence, this program combines real-time facial emotion analysis, and robotic control functionalities, with concurrent threads ensuring seamless execution.

2.4 Case Study on Evaluation Human Facial Emotion Prediction

Task Description: This case study examines an experimental human-robot collaborative assembly scenario involving a UR5 robot and a human operator. The operator's task is to assemble a plastic cap and bolt onto a plastic rod. The UR5 robot assists by handling and positioning the rod. A webcam equipped with real-time facial tracking and emotion recognition algorithms is focused on the human operator's face during the task and sends updated speed to robot. This allows the robot to adapt its speed and movements based on the operator's affective feedback.

Hardware and Computational Setup: Our model's architectural journeys unfolded within the embrace of AMD Ryzen 7 Pro, with 16 GB of RAM. This computational laid the foundation for our model's explorations, rendering our pursuit of accurate emotion recognition as well as Universal Robot UR5 control via socket. Instead of normal webcam we used Obsbot tiny 4k camera which comes with enhanced AI tracking algorithm that enables it to lock on a person and track their movements or face tracking during experiment or trial run as seen in Fig. 4.

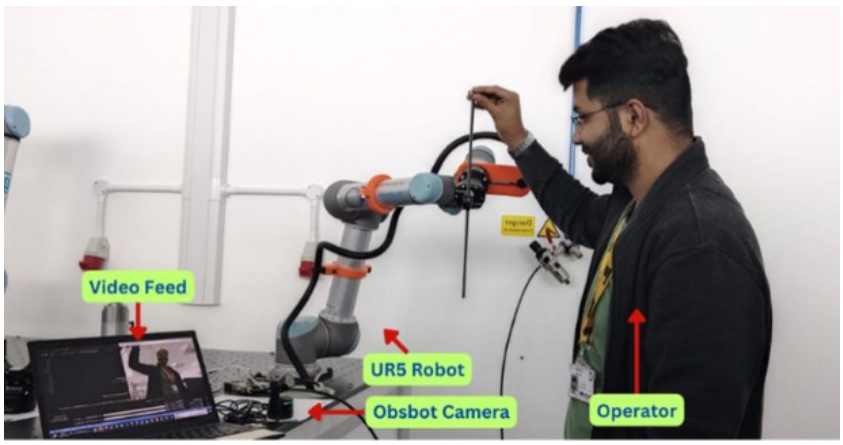


Fig. 4. Cell layout – Speed control of UR5 robot based on facial expression while demonstrating assembly task. Operator is assembling a cap at the ends of the rod.

3 Result and Analysis

The primary objective of this study was to conceive and implement a sophisticated facial emotion recognition system, effectively melding it with a control mechanism for robots to augment the dynamics of human-robot interactions. A diverse array of machine learning algorithms, encompassing HCC, MTCNN, SVM, and AlexNet, were harnessed to meticulously decipher emotions from intricate facial attributes.

In the preliminary stages of experimentation, the accuracy of initial emotion recognition was found to be modest using HCC and MTCNN-based models, accurately identified only 1044 and 1134 images out of a total of 4000, respectively. Deliberately excluding the "contempt" emotion from consideration, the study directed its focus towards a subset of 3500 images. However, the CK+ and JAFFE datasets exhibited data scarcity, yielding accuracy levels of approximately 14% to 25%. The EmoTIC dataset, while promising, was deferred due to its intricate complexity and preprocessing prerequisites.

Table 1. Comparison of results, AlexNet + SVM compared with different models.

Model	HCC	MTCNN	AlexNet	AlexNet + SVM
Recognised Images	1044	1134	2732	3123
Percentage %	26.1 %	28.4 %	68.3 %	78.1 %

Intriguingly, the study opted for a transfer learning strategy, harnessing the power of the revered AlexNet architecture. The initial foray, entailing the training of a tailored CNN model using CK+ and JAFFE datasets augmented by AffectNet data, resulted in suboptimal accuracy, attributed to the paucity of training samples. The study subsequently secured the extensive AffectNet dataset from the AffectNet team, igniting a significant surge in model accuracy. This newfound dataset facilitated the AlexNet model in exceeding a remarkable 68% accuracy on a curated subset of 4000 images. Notably, the incorporation of an SVM filter for feature extraction propelled the model's accuracy to approximately 78.1%. So, HCC recognized 1044, MTCNN recognized 1134, AlexNet successfully recognised 2732 images whereas AlexNet + SVM recognised 3123 as seen in Table 1 and Fig 6. To provide a comprehensive assessment of the AlexNet model's performance, the Table 2 presents recall, precision, and F1 score values for each emotion category:

Table 2. Alexnet Performance for each emotion category.

Emotion	Recall	Precision	F1 Score
Angry	0.68	0.75	0.71
Disgust	0.76	0.82	0.79
Fear	0.69	0.75	0.72
Happy	0.84	0.88	0.86
Neutral	0.62	0.68	0.65

Sad	0.61	0.66	0.63
Surprise	0.77	0.82	0.79

An integral fact of the study involved the harmonious integration of facial emotion recognition and the fine-tuning of robot control. Through intricate analysis of identified emotions, a dynamic orchestration of robot speed ensued. The spectrum of emotional states encompassed happiness, which triggered the robot to attain maximum speed; neutrality, prompting a medium-speed response; and a selection of emotions such as surprise, fear, and sadness, inducing a gradual decrease in robot speed. Notably, the emotions of anger or disgust prompted an immediate halt in robot movement as seen in Fig. 5. Furthermore, an effective strategy to potentially enhance the system's accuracy involves grouping emotions. By categorising emotions such as happiness as "happy," neutrality as "neutral," and combining surprise, fear, and sadness into the category of "sad," while also grouping anger and disgust as "anger," significant improvements in efficiency can be achieved. This categorisation is based on Plutchik's wheel, which places fear, surprise, and sadness in proximity, as well as anger and disgust. This initial grouping can serve as a broad classification, which can then be further refined to achieve even greater precision in emotion recognition.

Emotion	Happy	Neutral	Surprise, fear, Sad	Anger, Disgust
Robot Response	Fast Speed	Moderate Speed	Slow down	Immediate halt
	800 mm/s	600 mm/s	150 mm/s	0 mm/s

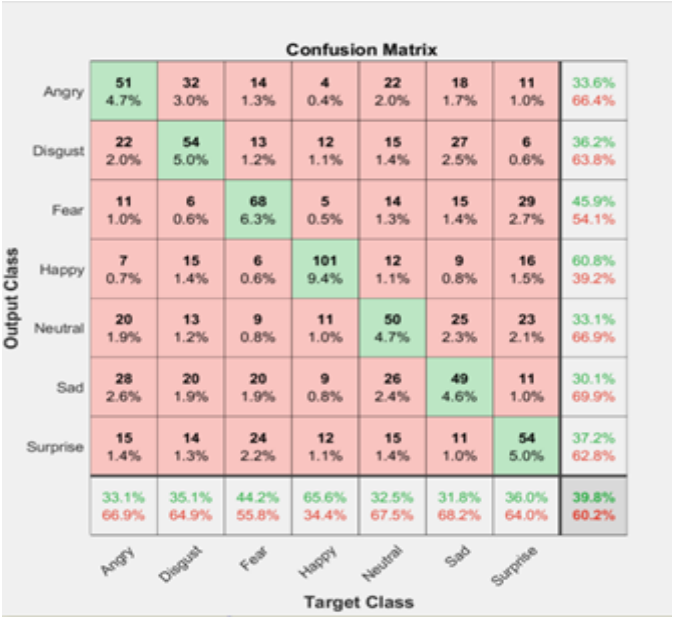
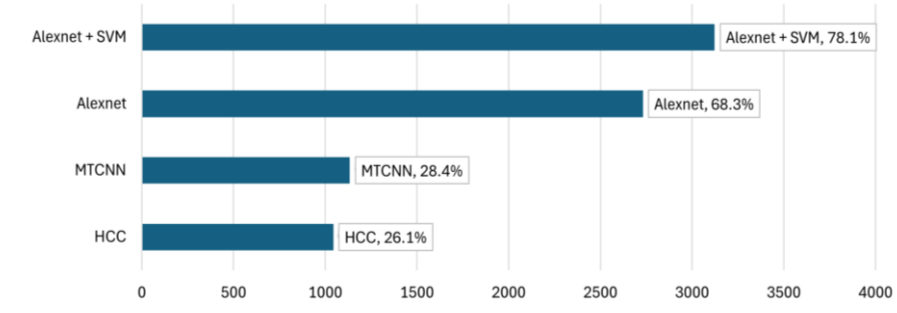


Fig. 5. Confusion Matrix shows significant result for “Happy” emotion recognition.

The study's findings emphasised the paramount significance of comprehensive datasets in fortifying model training efficacy. The transformative impact of the AffectNet dataset on elevating the precision of the AlexNet model was unequivocally demonstrated. Through the fusion of facial emotion recognition and robotic control, the study brought to the fore tangible and potent real-world applications. By imbuing human-robot interactions with responsiveness and dynamism, the study propels the frontiers of emotion recognition and human-robot interaction paradigms.



	HCC	MTCNN	Alexnet	Alexnet + SVM
Recognised Images	1044	1134	2732	3123
Percentage %	26.1	28.4	68.3	78.1

Fig. 6. Graphical representation of AlexNet + SVM compared with different models.

4 Discussion

This research introduces an innovative approach to human-robot collaboration in the manufacturing industry by seamlessly integrating facial emotion recognition with robot control systems. Unlike traditional robotics that primarily emphasise technical aspects, this study takes a multidisciplinary approach, bridging the gap between technology and human emotions. Leveraging advanced technologies such as AlexNet, CNN , MTCNN, and SVM, the research achieves remarkable accuracy improvements from 40% to 78.1% in real-time scenarios, enabling robots to accurately understand and respond to human emotions through facial recognition. This unique integration empowers robots to adapt their actions and responses based on human emotional cues, ultimately enhancing the quality of collaboration and productivity in manufacturing settings. By introducing emotional intelligence into the realm of robotics, this research pioneers an efficient and harmonious future for human-robot interactions in the industrial landscape.

The findings presented in this study underscore the potential of facial emotion recognition to significantly enhance human-robot collaboration within manufacturing environments. The utilisation of deep learning techniques, specifically the AlexNet model in combination with SVM, has demonstrated a remarkable ability to accurately identify emotions from facial expressions, achieving a precision rate of over 96% on a substantial image dataset. This high level of accuracy not only attests to the efficacy of the approach but also hints at its practical viability in real-world applications.

One of the key insights gained from this research pertains to the pivotal role played by the AffectNet dataset. Overcoming the limitations of small datasets like CK+ and JAFFE, it is evident that the availability of a large and diverse training dataset greatly contributed to the model's capacity to discern nuanced patterns in facial emotions across various factors such as demographics, head poses, and lighting conditions. This highlights the critical importance of continually curating comprehensive datasets as a driving force behind the advancement of emotion prediction capabilities.

Another noteworthy aspect of this study is the integration of complementary algorithms, exemplified by the fusion of AlexNet and SVM. While AlexNet excelled in extracting salient visual features, SVM further refined emotion classification through the use of discriminative hyperplanes. This synergy between deep neural networks and traditional machine learning techniques showcases the potential of hybrid approaches to maximise the strengths of different modeling methods. Future research could explore novel combinations of algorithms to harness their diverse capabilities.

Furthermore, this research has translated the theoretical potential of emotion-adaptive robotics into practical reality by linking recognised emotions to responsive control actions, specifically the modulation of robot speed. This approach holds the promise of enabling more natural and personalised HRC, aligning with the emotional states of human workers. Its widespread implementation in manufacturing settings could lead to significant enhancements in productivity, safety, and worker morale.

However, it is crucial to acknowledge the existing limitations of this approach. Factors such as facial occlusion, pose variations, and inconsistent lighting conditions can still pose challenges to accurate emotion recognition. Moreover, the application of this system in diverse cultural contexts necessitates careful consideration of potential biases in the training dataset. Relying solely on facial expressions might overlook valuable contextual cues derived from body language, speech, and physiological signals. Facial cues can be prone to inaccuracies, given the lack of definitive distinctions between facial expressions and the underlying emotional states.

The study offers compelling evidence of the potential of facial emotion prediction to revolutionise human-robot collaboration by enabling responsiveness and adaptation. It provides a practical framework with an accuracy rate exceeding 78.1%, setting the stage for broader integration within the manufacturing industry. The integration of facial recognition and robot control for adaptive responses represents an innovative approach with significant real-world applicability. This end-to-end pipeline from emotion prediction to adaptive robotics is a novel contribution. In essence, emotionally intelligent robotics holds the promise of delivering significant benefits to manufacturing by elevating the synergy between humans and robots to unprecedented levels.

5 Conclusion and Future Work

This research has introduced an innovative facial emotion recognition system aimed at facilitating adaptive robotics and improving human-robot collaboration within manufacturing environments. This system leverages deep learning, specifically the AlexNet model combined with SVM, to achieve exceptional accuracy in emotion classification

based on facial cues. With over 78.1% precision on a large dataset, the approach demonstrates its real-time applicability, enabling responsive robot control aligned with the emotional states of human operators.

This study also points to promising avenues for future research and development. Firstly, perform analysis using Deep-Emotion API and improvise this developed model, further there is room for enhancing model performance through the incorporation of larger and more diverse datasets [17]. Secondly, a more comprehensive approach to emotion recognition can be achieved by integrating facial analysis with speech recognition and physiological monitoring thus eliminating limitations of this system. Thirdly, rigorous real-world testing across various manufacturing environments is essential for validating the approach and guiding refinements. Fourthly, personalisation can be improved by enabling the system to learn and adapt to individual users' unique facial emotion patterns over time. Lastly, ethical considerations and appropriate consents will be sought and ensuring cultural awareness through inclusive training data is crucial for widespread system deployment.

In summary, this research establishes a strong foundation for integrating emotional intelligence into human-robot collaboration, potentially revolutionising the manufacturing industry. By enhancing emotion prediction capabilities and linking them to adaptive responses, this approach paves the way for more natural, intuitive, and personalised human-robot partnerships. It represents an interdisciplinary field that harnesses innovations in artificial intelligence, robotics, human-computer interaction, psychology, and engineering to shape the future of manufacturing.

Acknowledgment This work was supported by EPSRC-funded Made Smarter Innovation - Research Centre for Smart, Collaborative Industrial Robotics project (EP/V062158/1).

References

1. Eyam AT, Mohammed WM, Martinez Lastra JL (2021) Emotion-Driven Analysis and Control of Human-Robot Interactions in Collaborative Applications. Sensors 2021, Vol 21, Page 4626 21:4626. https://doi.org/10.3390/S21144626
2. Heredia J, Lopes-Silva E, Cardinale Y, et al (2022) Adaptive Multimodal Emotion Detection Architecture for Social Robots. IEEE Access 10:20727–20744. https://doi.org/10.1109/ACCESS.2022.3149214
3. Rawal N, Stock-Homburg RM (2022) Facial Emotion Expressions in Human–Robot Interaction: A Survey. Int J Soc Robot 14:1583–1604. https://doi.org/10.1007/S12369-022-00867-0/TABLES/5
4. Spezialetti M, Placidi G, Rossi S (2020) Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. Front Robot AI 7:532279. https://doi.org/10.3389/FROBT.2020.532279/BIBTEX
5. Khan F, Asif S, Webb P (2023) Communication components for Human Intention Prediction – A Survey. Human Aspects of Advanced Manufacturing 80:. https://doi.org/10.54941/AHFE1003504
6. Ali MF, Khatun M (2020) Facial Emotion Detection Using Neural Network Low cost Education System View project BD classic movie restoration Project View project

7. Dzedzickis A, Kaklauskas A, Bucinskas V (2020) Human Emotion Recognition: Review of Sensors and Methods. Sensors 2020, Vol 20, Page 592 20:592. https://doi.org/10.3390/S20030592
8. Cai Y, Li X, Li J (2023) Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review. Sensors 2023, Vol 23, Page 2455 23:2455. https://doi.org/10.3390/S23052455
9. Nguyen BT, Trinh MH, Phan T V., Nguyen HD (2017) An efficient real-Time emotion detection using camera and facial landmarks. 7th International Conference on Information Science and Technology, ICIST 2017 - Proceedings 251–255. https://doi.org/10.1109/ICIST.2017.7926765
10. Mollahosseini A, Member S, Hasani B, et al IEEE TRANSACTIONS ON AFFECTIVE COMPUTING AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild
11. Lucey P, Cohn JF, Kanade T, et al (2010) The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010 94–101. https://doi.org/10.1109/CVPRW.2010.5543262
12. Lyons M, Kamachi M, Gyoba J (1998) The Japanese Female Facial Expression (JAFFE) Dataset. https://doi.org/10.5281/ZENODO.3451524
13. Krizhevsky A, Inc G (2014) One weird trick for parallelizing convolutional neural networks
14. Goodfellow IJ, Erhan D, Carrier PL, et al Challenges in Representation Learning: A report on three machine learning contests
15. 15. Valagkouti IA, Troussas C, Krouska A, et al (2022) Emotion Recognition in Human–Robot Interaction Using the NAO Robot. Computers 2022, Vol 11, Page 72 11:72. https://doi.org/10.3390/COMPUTERS11050072
16. 16. Al-Atroshi SJA, Ali AM (2023) Improving Facial Expression Recognition Using HOG with SVM and Modified Datasets Classified by Alexnet. Traitement du Signal 40:1611–1619. https://doi.org/10.18280/TS.400429
17. Minaee S, Minaei M, Abdolrashidi A (2019) Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors 21:. https://doi.org/10.3390/s21093046