YOU ARE IN MY HEART: Enhancing Human-Robot Collaboration-Safety through Human Emotion-Based Perception in Smart Manufacturing Contexts

HOPE DIAMANTOPOULOS, ZOFIA PRZEDWORSKA, OMAR OBIDAT, JESSE PARRON, AND WEITIAN WANG

All authors are with the School of Computing, Montclair State University, Montclair, NJ 07043, USA

Corresponding author: Weitian Wang (e-mail: wangw@montclair.edu).

ABSTRACT The development of safe human-robot collaborative systems is crucial to the future advancement of the manufacturing industry. This is particularly relevant to companies that already utilize robots alongside human workers in collaborative tasks such as product co-assembly. However, as industry standards require workers to continually be in closer proximity to these robots with powerful capabilities, it is vital that workers are kept safe, especially when entering the robot's envelope or workspace. The proposed approach in this paper allows both the human and robot to make the best use of their respective capabilities, while constantly monitoring the human worker's emotions in order to prevent accidents. This additional layer of understanding allows the robot to respond appropriately to its human counterpart's feelings. Having robots take over tedious, repetitive tasks enables human workers to focus on controlling the more complicated decision-based operations. Allowing robots to understand the nuances within subjective human emotions empowers them to respond to the expression of a given feeling appropriately. This subjectivity of human behavior in human-robot collaborative contexts has not been extensively studied in the field. We explore the potential results of robots appropriately responding to human emotions in a manufacturing setting. To allow for safe human-robot interaction and collaboration, we employ a transfer learning-based method of teaching the robot to discern various human emotional states. This method allows the robot to be able to understand and properly respond to its human partner's emotional states, while collaborating with a human on a job. Within our approach, the speed of the robot's actions changes in response to the human's perceived feeling. For example, detected sadness or discomfort in humans causes the robot to slow down, while perceived happiness and satisfaction result in the robot increasing the pace of its work to match its partner's attitude. Our method allows for greater understanding between a collaborative human-robot pair, leading to both robot-leveraged efficiency and a reduction in safety risks in the workspace. The results of our experiments show how a robot can correctly detect real-time changes in human expressions and apply this understanding towards safely co-assembling a product with a human. The future work of our study is also discussed.

INDEX TERMS Robotics, smart manufacturing, human-robot collaboration, comfort, human factors, computer vision.

I. INTRODUCTION

DESPITE major advancements in large-scale manufacturing technology, there are still many physically taxing tasks carried out only by humans [1], [2]. As industry sectors increasingly adopt smart manufacturing techniques where robots are employed and trained to perform certain kinds of work, tasks will be optimized using collaborative robots in order to alleviate humans' strains on the body and enhance manufacturing efficiency [3]–[5]. Constant change and perpetual evolution are to be expected in today's fast-paced world. For industries to thrive and become more cost-effective and efficient, they must closely track current and future market needs [6]. This cannot be accomplished without delving into futuristic, boundary-pushing technology that is more than likely unfamiliar to current industry professionals. What is considered cutting-edge today can become industry standard tomorrow and, in due time, entirely obsolete [7]. One way to increase workflow efficiency is by utilizing robots to assist humans with tedious, mistake-prone tasks. For example, autonomous mobile robots assist humans by independently carrying heavy payloads through warehouses [8], [9]. Additionally, robots with computer vision can automatically determine whether additional work is required to assure optimal safety and quality [10], [11]. When robots assist humans with tasks, they work with a high level of accuracy that surpasses that of humans working independently [12]. Businesses that employ collaborative robots oversee research on them prior to production implementation, so the robots can be seamlessly added to the formerly human-only job. This ensures that necessary safety regulations are followed, human workers acclimate to their mechanical coworkers, and operations are completed as expected in the industry. The constant changing of the industrial landscape motivates companies to continually adjust their operations towards technological progress; the sectors that succeed in integrating smart technology into their workforce reap the benefits. They harness technology's power to complete intensive work, and humans' safety and comfort are enhanced as a result.

Rather than jeopardizing occupational health and safety by pushing human employees to their physical limits for the sake of keeping up with production, industries can reduce risks while making production more time-efficient with the use of robots. Human-robot collaboration (HRC) enhances the level of human's engaging tasks in the workplace [13]–[15]. It lets people focus on jobs involving leadership, planning, and problem-solving, while eliminating the risk of physical injuries caused by intensive and dangerous work that humans otherwise conduct in the manufacturing setting.

The involvement of collaborative robots and smart technologies is a humane way of alleviating difficulties for industrial employees and increasing the quality of production work. Relative to the limited decision-making capabilities of today's robots, humans hold these competencies naturally and thus are able to direct robots in completing given tasks. As robots' scope of manufacturing skills consists of objective, direct instructions, they are less flexible in going outside the bounds of their abilities unless trained to do so [16]. When testing the use of a robot in the industry, human safety protocols need to consider the static way in which robots perform tasks [17]–[19]. As robots are programmed to perform tasks in a vacuum, outside changes to their environment are not accounted for. Robots continue to run as they are programmed to in an altered setting, unless a change of setting is accounted for in their code. For example, a robot may not slow down its movements when its human operator walks up too close. This is a potential source of human injuries and equipment damage in the workplace. Robots' level of ability in adjusting to unexpected real-time changes should be considered during the research stage of implementing a robot in the work environment.

In anticipation of a technologically driven future of manufacturing, we propose the use of a standard web camera for robots in learning the differentiation between different human emotions. This method is more accessible and direct in human-robot collaboration than alternative safety protocols (e.g., the use of proximity sensors). Our method enables robots to understand human responses to their movements more accurately, improving the human experience in the workplace. Additionally, utilizing robots that understand how people are feeling keeps the decision-making in workers' hands, while removing certain draining activities that drag down the work process. Several studies have been conducted on emotion recognition in robot-assisted healthcare through EEG sensors [20]. However, developing a more intuitive and effective approach (e.g., using a general web camera) to detect and respond to human emotional expression in human-robot collaborative manufacturing contexts remains understudied and underutilized. The contributions derived from our study can be summarized as:

(1) We put forth a novel solution in improving human-robot collaboration safety by enabling robots to proactively empathize and understand human emotions in shared tasks.

(2) We develop a transfer learning-based approach to empower robots to accurately accommodate and assist human partners in collaborative tasks using only a small emotion dataset.

(3) The proposed approach is experimentally implemented in real-world human-robot collaborative manufacturing contexts. The results and evaluations demonstrate that the robot can precisely understand human emotions in real-time and effectively help its human partner in the co-assembly task.

II. RELATED WORK

In recent years, several efforts have been made in consideration of collaboration-safety in human-robot teams. A wide variety of different strategies and safety procedures have been proposed to make working comfortably alongside robots feasible. Researchers across this field are looking into many different techniques to make human-robot collaboration safer. The workplace is moving away from robots being tucked away from humans and is progressing towards enabling a symbiotic relationship between humans and robots [21]–[23].

Some researchers are looking into the concept of creating a non-physical barrier for the robot to remain inside in order to detect whether the robot is following safety procedures. The work [24] introduces a light projection system that helps identify when the robot has gone out of its proper bounds. Sensing when the robot has breached its light-ray boundary allows for other safety measures to be taken. This is contrary to the archaic procedure of keeping humans and robots completely separate. Multiple safety boundaries with unique shapes can be created using this technique.

Although a light boundary can be useful when defining safety parameters, it is not very efficient when humans and robots are enveloped in the same workspace. Assembly lines require parts to be passed along from robot to human and vice versa. There needs to be areas that both the human and the robot are able to occupy simultaneously. Because of this need, tracking where the human is working and moving is vital. Scientists have utilized jackets with different colored sections to keep a robot's eye on where the person's shoulders, elbows, and wrists are located throughout collaboration. This input provided by a triple-camerathe stereo vision system, allows image coordinates to be generated. When the human wears the jacket, the system tracks where t Another study in this field focuses on ways to use sensors to detect unexpected changes in the environment. These sensors acquire the distance from new entities within the workspace. For example, if a human were to enter the workspace, the robot would take notice and select its actions based on a positive desired effect for the person's safety. This is done by utilizing distance sensors that surround the cylindrical part of the robotic arm, which reaches out when completing tasks. These nerve-like sensors covering the robot allow it to sense motion from every direction. It would not be able to do this if the sensors were just arranged in one direction [27]. The more directions the robot can sense, the more accurate safety protocols can be. Calculations of velocity and inertia are used when determining if an action is safe enough to be taken when a person is in the proximity of the robot. If the action is deemed too dangerous, then an alternative, safer action is taken.

Measuring the possible danger level throughout all robotic actions could establish better safety for the human participant in human-robot interaction. This way, if the danger level is too high, the robot can back off or take another action. A perceived danger level can be generated by giving the robot the ability to view and monitor the human's behavior. This is done to make the given human-robot interactions safer. Physiological indicators such as skin conductance response, heart rate, and corrugator muscle activity are monitored. These metrics track anxiety in the human counterpart throughout every interaction [28]. Detecting uneasiness in participants can help prevent this course of action from recurring.

Methods to keep track of scenarios that have the potential to play out the wrong way are also being developed. These methods allow for the creation of safety plans that are initiated if potential danger is sensed, and they help take the safest path possible. This ensures that the robot carefully considers how its next steps affect the potential danger level based on the velocity and distance of the human [29]. As an additional safety measure, this study develops measures that determine the signals for when it is appropriate for the robot to withdraw from a collaborative task (e.g., humans suddenly raising their hands).

However, simply monitoring a variety of physical changes and putting safety procedures in place is not sufficient. Since humans have a much wider perception of their surroundings, they might perceive actions quite differently from the way robots perceive them. Humans can sense danger outside of the scope of a robot's cognition. Therefore, human discomfort in a variety of situations must be acknowledged by having the robot respond in a way that makes the person feel comfortable and listened to.

Emotion recognition-driven human-robot collaboration strategies have been studied to alleviate this gap in recent years. Different approaches and solutions have been developed aiming to facilitate human-robot interaction and collaboration [30], [31]. The authors in [32] presented a facial emotion recognition method based on 2D-Gabor and uniform local binary pattern operator, and tested it in simulation scenarios. A multimodal emotion recognition solution with evolutionary computation was developed for human-robot interaction in service robotics contexts [33]. An approach for adapting the cobot parameters to the emotional state of the human worker was presented in [34], which employed the Electroencephalography technology to characterize and understand the human emotional state. However, few studies have implemented emotion recognition-enabled solutions in real-time physical human-robot collaborative tasks, particularly with collaboration-safety considered.

To solve the issues mentioned above, we present a new approach that teaches a collaborative robot to recognize how its human counterpart individually displays seven different emotions. Allowing the robot to shape its own understanding of each human's individual emotion is important because not every human outwardly displays emotions in the same way. Our technique ensures that the robot can recognize subtle changes in expression, which means that if the data fed into training is accurate, even slight discomfort can be identified. This way, the robot only takes swifter actions when the human is displaying positive emotions. The robot should be able to sense how its human co-worker is feeling in order to respond in a way that lets the person feel that he/she is being catered to.

III. MODELING METHODOLOGY

A. TRANSFER LEARNING

Traditional machine learning models require separate training for each separate class when two or more classes exist. This process is time-inefficient due to the entire training process being restarted and completed for each class. Transfer learning bypasses this issue by drawing upon a wide library of public data; the learning system, or pre-trained model, can be used to transfer new data and the newly assigned task into our model, as shown in Figure 1, in order to generate the desired outcome [35], [36]. By sourcing knowledge from an already trained model, we are able to shorten the process of teaching new operations to a robot based on an existing learning system. New information is not saved in traditional machine learning; when working with multiple individual tasks, differentiated networks should be used for each [37]. Every stage of this procedure should lead to a specified outcome in order to make the model applicable to the studied problems. The model's input must be precise and relevant to the collaborative problem it is designed to solve. In the transfer learning system, a network trained on a different domain can be used as a source task. The learned knowledge will be transferred to the domain and employed in the target task. The previously trained model utilizes millions of labeled images. This large amount of data helps initiate the development process quicker than if we were starting the process from scratch.

The additional component to preparing the desired task with the transfer learning approach is a small set of locally stored data. A model that uses this dataset can be built in a variety of ways, whether it be a single person's emotional expressions or those of multiple people. The model is constructed in a time-efficient manner due to the utilization of the pre-trained weights. Additionally, the learning process in transfer learning is faster and more effective than that in traditional machine learning.

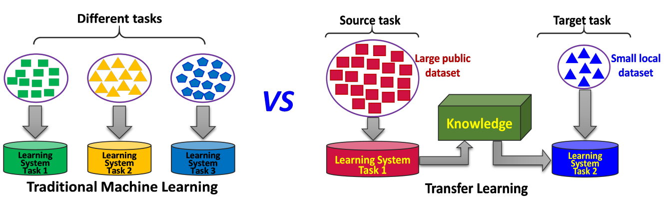


FIGURE 1. Transfer learning VS traditional machine learning.

B. VGG16

We teach the robot to differentiate between human emotional reactions using VGG16 in our transfer learning approach. As a Convolutional Neural Network, VGG16 consists of 3 defense layers and 13 convolutional layers to add a 3 x 3 size filter to the captured image [38], [39]. When new input is processed through the VGG16 network layering, a feature map becomes more accurate. The 13 convolutional layers are organized into two two-layer partitions and three three-layer ones. The partitions include a pooling layer that flattens the image, halves its file size, and highlights whichever characteristics of the image are most prominent. Rectified Linear Units (ReLU) prevent the decay of accumulated information with time. A library of more than 14 million image files goes into training the VGG16 network [40]. Complex Convolutional Neural Network models require a Graphics Processing Unit (GPU) relative to the number of mathematical operations needed to complete the training. This is because the GPU allows calculations to occur in parallel or at the same time. Although a GPU is still very important for this emotion understanding study, utilizing transfer learning with a VGG16 pre-trained model to teach the robot in human-robot collaborative tasks is less strenuous on the hardware requirements.

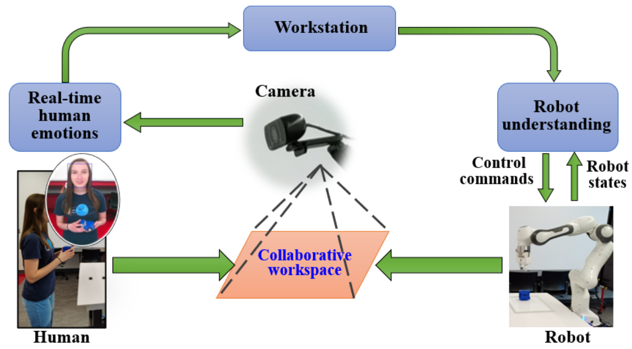


FIGURE 2. Human emotion data collection process.

C. HUMAN EMOTION DATA COLLECTION

In order for the robot to work collaboratively with the human and adjust its actions accordingly, the robot must be able to identify different human emotions. For this level of understanding to be reached, pictures of the human portraying each of the emotions that are to be understood must be acquired to train the robot's cognitive capacity. Figure 2 demonstrates the human emotion data collection process in the experimental space. The participant looks at the camera and forms the seven distinct facial expressions (happiness, anger, fear, sadness, disgust, surprise, and neutral) as the system takes and saves images. Unintentional disruptions and shifts in expression, such as a blink, are accounted for as the camera rapidly takes photos. The human should not only look directly into the camera while these frames are being captured because people's movements in daily life are more unpredictable. Therefore, for the model to work effectively at different angles of view, capturing multi-faceted pictures is essential. This approach is comparable to a person putting their finger on a fingerprint scanner for the first time and having to move their finger on the scanner, so that the machine captures many different angles of the fingerprint. During the adoption of an application, the data collected from its users is also as detailed as possible. The robot absorbs new information in the collected dataset based on the VGG16 pre-trained model. After the data is gathered, we create a system for detecting the photographed emotion in real-time, to be employed by the robot. The emotion recognition system then lets the robot interact with its human partner by sending the appropriate command to the robot's local controller.

D. SMALL LOCAL EMOTION DATASET

A general web camera is utilized as the robot's "eye" to capture human emotion information. We develop a script that collects, names, and archives the photos of the expressions; this is done as the human is modeling emotions in front of the camera. As the field of human-robot interaction encompasses a wide variety of potential situations, as shown in Figure 3, we chose to represent seven kinds of expressions commonly used by humans on a daily basis [41], [42], with 100 pictures taken of each (as contrasted with the millions of photos in the VGG16 library). When collecting human emotion images, the HAAR cascade algorithm [43] is employed to accurately detect and capture the areas of a human face. We found that the human facial expressions were able to be correctly acquired even when the background of the working environment was complex or dynamic. In addition, different angles and varying levels of each emotion are performed by the participant in the image collection process. Human emotions are not always extreme. It is important that our approach can pick up on subtle changes in emotions and still understand how the human is feeling. The completed set of images is saved in the given proportions and grayscale. The bottleneck features are selected within each set of emotion photos. This method is rooted in the VGG16 learning methods, and the robot applies this with transfer learning.

FIGURE 3. Human emotions adopted in this study.

E. HUMAN EMOTION-BASED PERCEPTION IN HRC

In this study, a robot is taught to respond to a human's emotional stimuli in a collaborative setting, with the goal of having a robot learn how to process such stimuli. Using VGG16 and a local library of emotional expression photographs as a framework, the robot deepens its own knowledge base with transfer learning. Its newfound competencies allow the robot to tell which emotion its human coworker displays; the robot can then apply this knowledge to the collaborative process with the human. The newly introduced data and policy are processed using the robot learning, which is shown in Figure 4 with the online emotion understanding method.

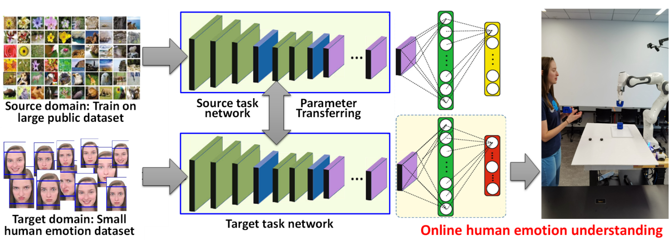


FIGURE 4. The framework of the robot learning approach.

We define the following variables in the context of robot learning: the VGG16 source domain D_S, a target domain for comprehending facial expressions D_E, the VGG16 learning task T_S, and a task for emotion learning T_E. With these variables, we build an emotion prediction function $f_e^*(X^{'})$. The possible results of $f_e^*(X^{'})$ can be produced in D_E by the transfer learning using the knowledge in D_S and T_S, where $D_S \neq D_E$ or $T_S \neq T_E$. The collected emotional expression data is represented by X, and the online emotion data is denoted by $X^{'}$. For every set $X^{'}$, the perceived human emotion E_* can be derived from the predicted outcomes:

$$E^* = \underset{e=1,2,\ldots,M}{\arg\max}\ f_e^*(X^{'}) \tag{1}$$

where M represents the quantity of emotion categories; M is 7 to represent the seven emotions tracked in our study. Eq. (1) will enable the robot to qualify what emotion the human is displaying. Further, the emotion understanding results will allow for humans to be more confident in collaborative tasks with robots; the humans can expect that their emotional states will be analyzed and responded to accordingly by robots through the proposed human emotion-based perception framework in human-robot collaboration.

IV. EXPERIMENTAL PLATFORM

A. EXPERIMENTAL PLATFORM

Figure 5 presents the components of our experimental workspace: a co-assembly robot, a camera, a workstation, an object to be assembled, and a work area to be used by both the human and robot. A vehicle model is used as the target object in the human-robot co-assembly task. This study uses a Franka Emika Panda, a 7-DoF collaborative robot that is equipped with a two-finger parallel gripper, a pilot-user interface, and a Franka Control Interface (FCI) controller [44], [45]. The Franka Emika Panda can safely collaborate with humans in a way that simulates the interaction between humans. The study also uses a ThinkPad P15 with an Intel Core I9-10885H processor and 64GB of memory; this workstation processes emotional images and is the center of our transfer learning system's development. It also assists with detecting emotion in real-time and mapping the robot's actions. The Robot Operating System (ROS) is utilized in managing our robot system [46], [47]. ROS is an open-source framework for inter-platform maneuvering and communication on a large scale. In addition, the study uses MoveIt! and runs it with the ROS operating system. MoveIt! is a library of ROS-compatible packages, tools, and software that enables the user to modify the actions of a robot. ROS simulation software allows for the development of robot path planning algorithms [48]. To plan the robot's movements in human-robot collaborative tasks, the control commands are sent to the libfranka interface, which is a ROS package that allows for the Panda robot to communicate with the FCI controller. The FCI will provide the current robot states and enable the robot to be directly controlled by the commands derived from the robot's real-time perception of human emotions.

B. TASK DESCRIPTION

We validated the time efficiency and correctness of our developed approach by carrying out real-world human-robot collaborative experiments. In these experiments, the human

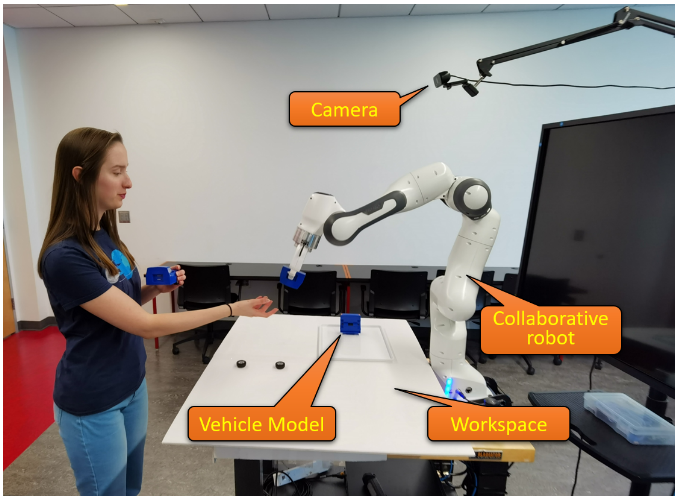


FIGURE 5. The experimental platform.

and robot workers partake in a collaborative manufacturing job. The experiment uses a buildable four-part vehicle (including a cargo bed, backseat, front seat, and front of the car) with four additional detachable wheels as an assembly prompt that the human and robot are to follow. Our co-assembly task consists of the robot handing parts of the car to the human, starting with the backseat and cargo bed. Once handed all the parts, the human can make sophisticated decisions in putting together the pieces; this exemplifies how human-robot collaboration can leave humans in charge of administrative tasks while leaving repetitive labor to machines. Employing our proposed methodology, the robot will be able to understand and analyze the human's real-time emotion expressions. In response to the emotion interpretation data received from the online recognition system, the robot changes its movement and speed so that the human emotional state is accommodated. By way of example, if the human seems upset or sad, the robot will slow down its actions. If the human seems to be happy or in a good mood, the robot will pick up the pace, matching the human worker's energy with a faster speed. If the human feels disgusted, the robot would actively move away from the human and only start to collaborate when the human feels better or looks upbeat. This makes collaborative tasks safer and better the ergonomics of human-robot relationships. The proposed approach is generalized and not limited to tasks in which humans co-assemble with robots, or the adopted emotion types. These validation experiments will serve as stepping-stones to more complex human-centered collaborative tasks in multiple application areas such as healthcare and aerospace exploration.

V. RESULTS AND EVALUATIONS

A. EMOTION TRAINING AND CROSS-VALIDATION ACCURACY

Prior to training the robot, the data received from the images regarding human emotion is categorized into three main categories: emotion training data, emotion cross-validation data, and emotion testing data. The numbers of emotion images in each category are listed in Table 1.

As shown in Figure 6, at around epoch 8.5, the training accuracy is 100%. On the other hand, the cross-validation accuracy reaches 97.82%. Identifying the correct emotions from images presented beforehand at the fastest speed is a priority; however, the system also needs to deduce the correct emotion of new image data that it has not been shown before. Tracking accuracy during the training process assures that our system is working properly and connects each individual emotion with the proper features. It is necessary to confirm that the robot learning system is outputting consistent results, regardless of changes in external circumstances. Figure 6 shows that the accuracy levels of both the emotion training and cross-validation are high once a certain threshold of epochs is crossed. This indicates that the robot has been trained well enough to achieve a high level of cognition of human emotions.

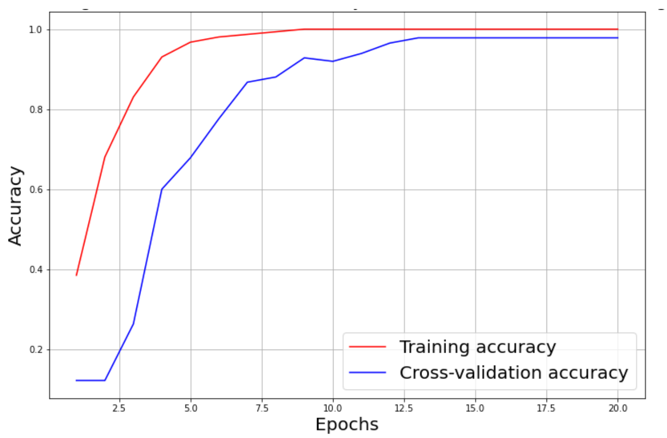


FIGURE 6. Training and cross-validation accuracy of human emotion understanding.

B. EMOTION TRAINING AND CROSS-VALIDATION LOSS

The robot's model must be strong in order to properly work with the seven different emotions. The dataset must be diverse enough so that the model does not overanalyze it. A result of data over-analyzation is the model becoming overly specific to a small number of users or outputting incorrect results due to different conditions, such as the lighting, the background, or the proximity of the human subject. Understanding and appropriately responding to the detected

TABLE 1. The Numbers of Training, Cross-Validation, and Testing Images in Each Type

Emotion Type	Training Images	Cross-validation Images	Testing Images
Anger	67	13	20
Disgust	67	13	20
Fear	67	14	19
Happiness	66	14	20
Neutral	67	13	20
Sadness	66	14	20
Surprise	67	13	20

human emotions is the main priority of the model. In order to tune the weights during the training process, the model was evaluated using the cross-entropy loss function [49]. The predicted probability of each emotion type is compared to the desired emotion type. A loss value is then computed to penalize the probability based on how far it is from the expected emotion type. The penalty is logarithmic in nature, which generates a larger score for major differences tending to 1, and a smaller score for minor differences closing to 0. As the loss value becomes lower, the model becomes more reliable. As shown in Figure 7, the training loss can reach 0 and the cross-validation loss can reach 0.2. These results suggest that the robot's learned model is accurate and reliable.

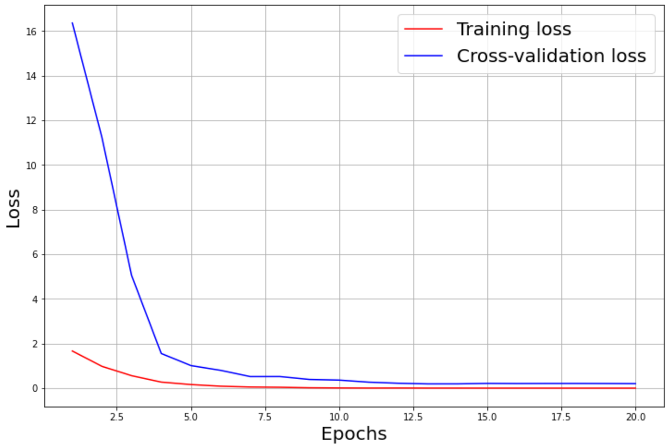


FIGURE 7. Training and cross-validation loss of human emotion understanding.

C. REAL-WORLD HUMAN-ROBOT COLLABORATION

Utilizing the developed approach, human-robot collaborative manufacturing tasks are implemented in a real-world context. During the human-robot collaboration procedure, based on the new comprehension acquired from the robot's learning, the robot acknowledges how its human partner is feeling and responds accordingly. As shown in Figure 8, the robot recognizes when the human is in a good mood and responds accurately by successfully completing the collaborative task. In Figure 8(1), the robot starts the shared job by picking up the backseat of the car that was placed in the workspace in front of it. The human simultaneously collects the other pieces directly in front of them and assembles them. Afterward, the robot examines the human's face to assess the state of emotion she feels. This assessment is based on prior acquired knowledge and is conducted before continuing with the collaborative task. As presented in Figure 8(2), the robot accurately recognizes that the human is expressing a happy emotion, so the robot quickly hands over the backseat to the human. In Figure 8(3) and Figure 8(4), the human interacts with and receives the backseat from the robot. Figure 8(5) and Figure 8(6) display the human joining the pre-assembled parts to the newly received segment. The robot system works with the human at a rapid pace due to the identification of the human's positive emotional state. The consideration of how human feels allows for an efficient and safe human-robot collaborative task. The real-world experimental video is available at: https://youtu.be/v488i96LPE8.

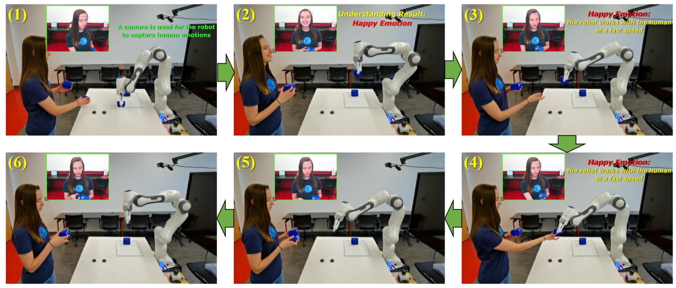


FIGURE 8. "Happiness" emotion understanding and human-robot collaboration in the shared task.

Human emotions are unpredictable, as there are many factors that lead to their changing during the human-robot collaboration process. It is important to consider this information and ensure that the robot is able to recognize when its human partner is in a negative mood, so that it can take the necessary precautions. As shown in Figure 9, the robot identifies when the human collaborator is in a negative mood and adjusts accordingly. As exhibited in the prior positive emotion example, Figure 9(1) presents the robot beginning its task of helping its coworker by picking up the backseat from the workspace in front of it. The robot system then analyzes what emotions the collaborator is portraying before it conducts the collaborative work. As shown in Figure 9(2), the robot is cognizant of the individual's disgusted emotion, and takes precautions by moving away, in order to make the human feel better (Figure 9(3) and Figure 9(4)). The perceived negative emotion in the human causes the robot to stay distant until it recognizes a positive emotional change, for example, happiness. As soon as the robot perceives a positive response, as presented in Figure 9(5) and Figure 9(6), the joint work will continue. With the validation of the proposed approach in the performed tasks, the robot displayed that it can accurately analyze human emotional states and dynamically change its actions while performing collaborative tasks.

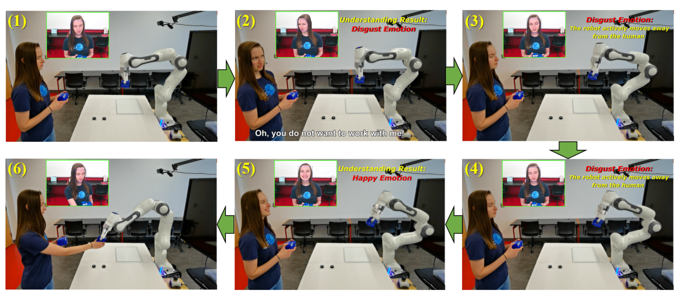


FIGURE 9. "Disgust" emotion understanding and human-robot collaboration in the shared task.

It is crucial that the robot be able to differentiate between positive and negative emotions so that human-robot collaborations function smoothly and safely. The speed of the robot's movement varies based on what emotion it detects. As shown in Figure 10(1) - Figure 10(3), when the negative emotion of sadness is identified, the robot moves to the human at a slower pace. As recorded by the time stamps, the robot takes nine seconds to deliver the backseat to the human. When a positive emotion, such as happiness, is detected (Figure 10(4)), the robot delivers the cargo bed to its human partner at a faster pace (Figure 10(5) and Figure 10(6)). As recorded by the time stamps, the robot moves at a speed of five seconds when happiness is recognized in the human. This implementation demonstrates that, when utilizing the proposed approach, the robot is able to accurately distinguish different human emotions. Based on the understood emotions, the robot can adjust how quickly it moves; this is done to make the human worker feel more comfortable. Setting the robot's assisting actions to an appropriate speed, given the emotion experienced by the human, ensures that the collaborative task is completed safely and at the speed that best accommodates the human's emotion states.

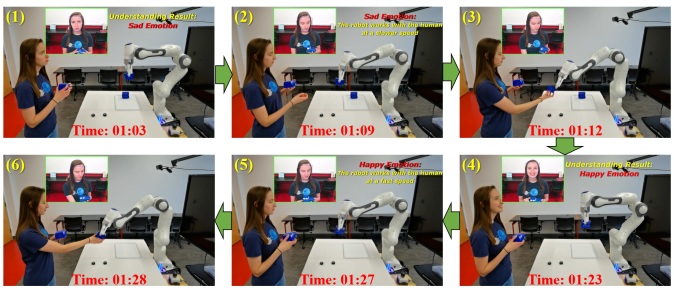


FIGURE 10. Comparison of "Happiness" and "Sadness" emotion understanding in the human-robot collaborative task.

D. CONFUSION MATRIX

To confirm that the robot's learning model correctly predicts emotional states, we evaluate the produced predictions through a confusion matrix. The matrix helps produce a visualization of the number of images that are correctly categorized by the model. There are four possible options for how the model predicted the emotion types within the confusion matrix. The four parameters are true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [50]. A TP occurs when the model correctly predicts the emotion shown in an image. Whereas a TN result shows that the model is correctly predicting that the image does not portray a particular emotion. An FP exists when the model correctly identifies that the emotion shown in an image is incorrect. An FN occurs when the model's prediction states that the image does not show an emotion while the image contains that emotion. The percentage measures along the blue diagonal in Figure 11 represent correctly predicted emotions using the test images, and the percentages in white display the misidentified emotions. With the occurrence of only two misidentified emotions (disgust and surprise), these results exemplify the high accuracy, efficiency, and robustness of the robot's learning model.

FIGURE 11. The confusion matrix of human emotion understanding (the blue diagonal presents the correctly recognized emotions).

Quick summarization of instance reliability is not the only data that can be gathered from Figure 11. Calculating TP, TN, FP, and FN is important because these tests provide valuable insight into where the model is performing best and where it has potential for improvement. As shown below in Table II, most of the possible TPs and TNs are desirable. However, in future model versions with FNs and FPs, more specific classification rules can be developed. Identifying sources of confusion between specific emotions can prevent the same type of misidentification in future applications of the model.

Further, based on each class's TP, FP, TN, and FN, we can obtain the following five parameters for each class: Accuracy, Error Rate, Macro-precision, Macro-recall, and Macro-F1-score [51] (as shown in Table III). Each of these parameters has favorable results. When the emotion metrics are viewed collectively, the model's total average accuracy is 99.59% its average error rate is 0.41%, its average macro-precision is 98.46%, its average macro-recall is 98.57% and the average macro-F1-score is 98.48%. The individualized breakdown shown in Table 3 demonstrates the calculated metrics for each emotion.

E. RECALL MATRIX

Calculating recall produces the proportion of correct desired emotions over the total number of emotions that should be output from the images provided. This is done through dividing the total number of true positives by the sum of

TABLE 2. TP, TN, FP, and FN for Each Emotion Understanding

Emotion Type	TP	TN	FP	FN
Anger	20	119	0	0
Disgust	19	119	0	1
Fear	19	119	1	0
Happiness	20	119	0	0
Neutral	20	118	1	0
Sadness	20	119	0	0
Surprise	19	119	0	1

TABLE 3. The Average Accuracy, Error rate, Macro-Precision, Macro-Recall, and Macro-F1-Score of the Proposed Approach

Emotion Type	Accuracy	Error Rate	Macro-precision	Macro-recall	Macro-F1-score
Anger	100%	0	100%	100%	100%
Disgust	99.28%	0.72%	100%	95%	97.44%
Fear	99.28%	0.72%	95%	100%	97.44%
Happiness	100%	0	100%	100%	100%
Neutral	99.28%	0.72%	95%	100%	97.44%
Sadness	100%	0	100%	100%	100%
Surprise	99.28%	0.72%	100%	95%	97.44%
Average Metric	99.59%	0.41%	98.57%	98.57%	98.54%

the total true positives and false negatives (TP/TP+FN). This calculation shows where the model understands the correct emotion. Conversely, the expression reveals places where positive predictions are not properly identified. As shown in Figure 12, the overall recall level is high, with most of the seven emotions (anger, fear, happiness, neutral, and sadness) being 100% accurate and two (disgust and surprise) being 95%, slightly less than complete accuracy. These calculated outcomes can be seen in the Macro-recall section of Table III. Each result is also displayed along the blue diagonal of the recall matrix in Figure 12. The two instances of inaccuracies are displayed outside of the blue diagonal, where the 0.05 inaccuracy scores are located. In one of these instances, the desired emotion was disgust, but instead neutral was incorrectly understood. In the second case, surprise was the desired emotion, but fear was incorrectly understood. This recall matrix indicates that the overall performance of our model reaches high accuracy.

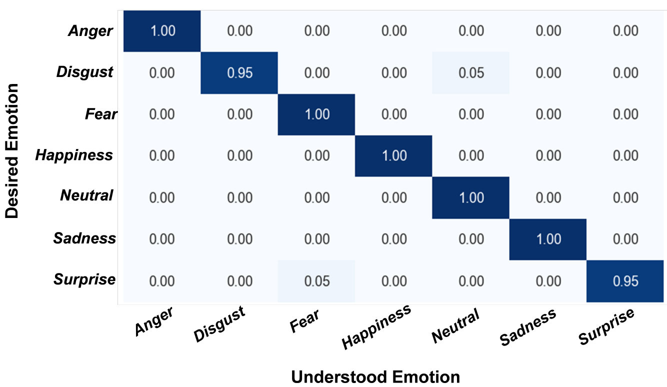


FIGURE 12. The recall matrix of human emotion understanding (the blue diagonal presents the recall value TP/TP+FN of each emotion recognition).

F. PRECISION MATRIX

Calculating precision yields the ratio of correctly matched emotions to the total number of positive predicted emotions. This includes both correct and incorrect positive predictions that the model outputs from the images provided. In other words, the total number of true positives is divided by the sum of the total number of true positives and the total number of false positives: TP/(TP+FP). Calculating precision helps display how reliable a positive prediction is from the individual emotion classifier. According to this formula, the more false positives exist, the worse the predictive ability of the model. As presented in Figure 13, the overall precision of the proposed approach was high, with most emotions (anger, disgust, happiness, sadness, and surprise) being 100% accurate and two (fear and neutral) being 95%, slightly less than completely accurate. These calculated outcomes can be seen in the Macro-precision section of Table III. Each of these results can also be seen along the blue diagonal of the precision matrix in Figure 13, while the inaccuracies are displayed outside of the blue diagonal. There are also two 0.05s in the precision matrix, representing inaccuracies, one where the desired emotion was disgust, but neutral was incorrectly understood instead. The other occurrence is where surprise was the desired emotion, but fear was incorrectly understood.

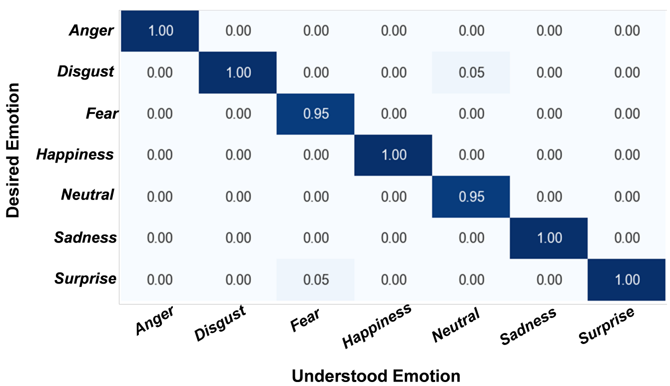


FIGURE 13. The precision matrix of human emotion understanding (the blue diagonal presents the precision of each emotion recognition).

VI. CONCLUSIONS AND FUTURE WORK

This study proposes an effective and novel approach that enables human-robot collaboration in smart manufacturing contexts. The developed transfer learning-based method relies on the robot "understanding" human emotions and it does so based on a relatively small dataset of human emotions. Our method allows for greater understanding between a collaborative human-robot pair, leading to heightened efficiency and a reduction in safety risks in the manufacturing space. This implementation is shown to be efficient in a real-world application of human-robot collaboration. The results verify the approach's validity, in that the robot is able to comprehend its human partner's emotions. This enables the robot to better aid the human in collaborative assembly tasks.

While the developed solution was successfully validated in human-robot collaborative tasks, there are still some misclassifications of human emotions that may adversely affect the collaboration safety of human-robot partnerships. To mitigate potential risks, further development and improvement are needed in order to enable human-robot collaboration in more complex tasks. The development effort includes the compilation of a more generalized dataset that describes human emotions. The dataset is comprised of images of humans conveying seven different emotions, and includes a diverse range of human emotional imagery, involving different ages, genders, races, etc. The inclusion of a wider range of emotions is also a future improvement to our dataset. We will define more human emotions to fortify the robot's perception capacity. In addition, we will explore new methods such as domain adaptation or domain randomization techniques and incorporate them into our approach to improve the model's robustness and generalization. During the human-robot collaboration, environmental conditions, such as lighting and occlusion, may affect the robustness of the developed model. We will verify the performance of the developed solution under varied conditions and compare it with baseline models of emotion recognition. Finally, we plan to recruit more participants to validate and improve our approach in more complex human-robot collaborative tasks for accelerating the transition-to-scale of the proposed approach in future human-centered workplaces and contexts.

REFERENCES

[1] A. J. Scott, "The changing global geography of low-technology, labor-intensive industry: clothing, footwear, and furniture," World Development, vol. 34, no. 9, pp. 1517-1536, 2006.

[2] S. Islam and S. S. Shazali, "Determinants of manufacturing productivity: pilot study on labor-intensive industries," International Journal of Productivity and Performance Management, 2011.

[3] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating Human–Robot Collaborative Tasks by Teaching-Learning-Collaboration From Human Demonstrations," IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 640-653, 2018.

[4] E. Matheson, R. Minto, E. G. Zampieri, M. Faccio, and G. Rosati, "Human–Robot Collaboration in Manufacturing Applications: A Review," Robotics, vol. 8, no. 4, p. 100, 2019.

[5] W. Wang, Z. Przedworska, J. Parron, M. Lyons, M. Zhu, and A. Tuininga, "MCROS: A Multimodal Collaborative Robot System for Human-Centered Tasks," in 2024 International Conference on Networking, Sensing and Control (ICNSC), 2024: IEEE, pp. 1-6.

[6] M. Rüßmann et al., "Industry 4.0: The future of productivity and growth in manufacturing industries," Boston Consulting Group, vol. 9, no. 1, pp. 54-89, 2015.

[7] K. Thoben, S. Wiesner, and T. Wuest, "Industrie 4.0" and smart manufacturing–a review of research issues and application examples," Int. J. Autom. Technol, vol. 11, no. 1, pp. 4-16, 2017.

[8] Y. Chen et al., "A Robotic Lift Assister: A Smart Companion for Heavy Payload Transport and Manipulation in Automotive Assembly," IEEE Robotics and Automation Magazine, vol. 25, no. 2, pp. 107-119, 2018.

[9] B. Hichri, J.-C. Fauroux, L. Adouane, I. Doroftei, and Y. Mezouar, "Lifting mechanism for payload transport by collaborative mobile robots," New Trends in Mechanism and Machine Science, pp. 157-165, 2015.

[10] H. Diamantopoulos and W. Wang, "Accommodating and Assisting Human Partners in Human-Robot Collaborative Tasks through Emotion Understanding," in 2021 International Conference on Mechanical and Aerospace Engineering (ICMAE), 2021: IEEE, pp. 523-528.

[11] R. Halme, M. Lanz, J. Kämäräinen, R. Pieters, J. Latokartano, and A. Hietanen, "Review of vision-based safety systems for human-robot collaboration," Procedia CIRP, vol. 72, pp. 111-116, 2018.

[12] E. Herrera, M. Lyons, J. Parron, R. Li, M. Zhu, and W. Wang, "Learning-Finding-Giving: A Natural Vision-Speech-based Approach for Robots to Assist Humans in Human-Robot Collaborative Manufacturing Contexts," in 2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS), 2024: IEEE, pp. 1-6.

[13] A. Vysocky and P. Novak, "Human-Robot collaboration in industry," MM Science Journal, vol. 9, no. 2, pp. 903-906, 2016.

[14] W. Wang, Y. Chen, R. Li, and Y. Jia, "Learning and comfort in human–robot interaction: a review," Applied Sciences, vol. 9, no. 23, p. 5152, 2019.

[15] O. Obidat, J. Parron, R. Li, J. Rodano, and W. Wang, "Development of a Teaching-Learning-Prediction-Collaboration Model for Human-Robot Collaborative Tasks," in The 13th IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, 2023: IEEE, pp. 1-5.

[16] Z. Pan, J. Polden, N. Larkin, S. Van Duin, and J. Norrish, "Recent progress on programming methods for industrial robots," Robotics and Computer-Integrated Manufacturing, vol. 28, no. 2, pp. 87-94, 2012.

[17] O. Obidat and W. Wang, "TELL ME YOUR FEELINGS: Characterization and Analysis of Human Comfort in Human-Robot Collaborative Manufacturing Contexts," in 2021 6th International Conference on Control, Robotics and Cybernetics, 2021, pp. 1-5.

[18] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," Mechatronics, vol. 55, pp. 248-266, 2018.

[19] J. Parron, T. T. Nguyen, and W. Wang, "Development of A Multimodal Trust Database in Human-Robot Collaborative Contexts," in 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2023: IEEE, pp. 0601-0605.

[20] E. Gümüslü, D. Erol Barkana, and H. Köse, "Emotion Recognition using EEG and Physiological Data for Robot-Assisted Rehabilitation Systems," in Companion Publication of the 2020 International Conference on Multimodal Interaction, 2020, pp. 379-387.

[21] J. Arents, V. Abolins, J. Judvaitis, O. Vismanis, A. Oraby, and K. Ozols, "Human–robot collaboration trends and safety aspects: A systematic review," Journal of Sensor and Actuator Networks, vol. 10, no. 3, p. 48, 2021.

[22] A. Zacharaki, I. Kostavelis, A. Gasteratos, and I. Dokas, "Safety bounds in human robot interaction: A survey," Safety science, vol. 127, p. 104667, 2020.

[23] C. Hannum, R. Li, and W. Wang, "A Trust-Assist Framework for Human-Robot Co-Carry Tasks," Robotics, vol. 12, no. 2, pp. 1-19, 2023.

[24] C. Vogel, M. Poggendorf, C. Walter, and N. Elkmann, "Towards safe physical human-robot collaboration: A projection-based safety system," in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011: IEEE, pp. 3355-3360.

[25] J. T. C. Tan and T. Arai, "Triple stereo vision system for safety monitoring of human-robot collaboration in cellular manufacturing," in 2011 IEEE International Symposium on Assembly and Manufacturing (ISAM), 2011: IEEE, pp. 1-6.

[26] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward safe human robot collaboration by using multiple kinects based real-time human tracking," Journal of Computing and Information Science in Engineering, vol. 14, no. 1, 2014.

[27] G. B. Avanzini, N. M. Ceriani, A. M. Zanchettin, P. Rocco, and L. Bascetta, "Safety control of industrial robots based on a distributed distance sensor," IEEE Transactions on Control Systems Technology, vol. 22, no. 6, pp. 2127-2140, 2014.

[28] D. Kulić and E. Croft, "Pre-collision safety strategies for human-robot interaction," Autonomous Robots, vol. 22, no. 2, pp. 149-164, 2007.

[29] D. Kulić and E. A. Croft, "Real-time safety for human–robot interaction," Robotics and Autonomous Systems, vol. 54, no. 1, pp. 1-12, 2006.

[30] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human-robot interaction: Recent advances and future perspectives," Frontiers in Robotics and AI, vol. 7, p. 532279, 2020.

[31] R. Stock-Homburg, "Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research," International Journal of Social Robotics, vol. 14, no. 2, pp. 389-411, 2022.

[32] Z. Liu et al., "A facial expression emotion recognition based human-robot interaction system," IEEE CAA J. Autom. Sinica, vol. 4, no. 4, pp. 668-676, 2017.

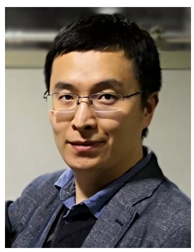
[33] L. Perez-Gaspar, S. Caballero-Morales, and F. Trujillo-Romero, "Multimodal emotion recognition with evolutionary computation for human-robot interaction," Expert Systems with Applications, vol. 66, pp. 42-61, 2016.
[34] A. Toichoa Eyam, W. M. Mohammed, and J. L. Martinez Lastra, "Emotion-driven analysis and control of human-robot interactions in collaborative applications," Sensors, vol. 21, no. 14, p. 4626, 2021.
[35] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques: IGI global, 2010, pp. 242-264.
[36] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," Journal of Big Data, vol. 4, no. 1, pp. 1-42, 2017.
[37] K. R. Weiss and T. M. Khoshgoftaar, "An investigation of transfer learning and traditional machine learning algorithms," in 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), 2016: IEEE, pp. 283-290.
[38] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in 2018 IEEE 8th annual computing and communication workshop and conference (CCWC), 2018: IEEE, pp. 169-175.
[39] D. Theckedath and R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," SN Computer Science, vol. 1, no. 2, pp. 1-7, 2020.
[40] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211-252, 2015.
[41] J. M. Zelenski and R. J. Larsen, "The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data," J Res Pers, vol. 34, no. 2, pp. 178-197, 2000.
[42] D. Trampe, J. Quoidbach, and M. Taquet, "Emotions in everyday life," PLoS One, vol. 10, no. 12, p. e0145450, 2015.
[43] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," in 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 2017: IEEE, pp. 483-487.
[44] P. Tilloo, J. Parron, O. Obidat, and W. Wang, "A POMDP-based Robot-Human Trust Model for Human-Robot Collaboration," in 2022 IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, 2022: IEEE, pp. 1-6.
[45] W. Wang, C. Coutras, and M. Zhu, "Empowering computing students with proficiency in robotics via situated learning," Smart Learning Environments, vol. 8, no. 1, pp. 1-18, 2021, doi: 10.1186/s40561-021-00167-6.
[46] M. Quigley et al., "ROS: an open-source Robot Operating System," in ICRA workshop on open source software, 2009, vol. 3, no. 3.2: Kobe, Japan.
[47] S. Bier, R. Li, and W. Wang, "A Full-Dimensional Robot Teleoperation Platform," in 2020 IEEE International Conference on Mechanical and Aerospace Engineering, 2020: IEEE, pp. 186-191.
[48] S. Chitta, I. Sucan, and S. Cousins, "Moveit!," IEEE Robotics & Automation Magazine, vol. 19, no. 1, pp. 18-19, 2012.
[49] P. Perera and V. M. Patel, "Deep transfer learning for multiple class novelty detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11544-11552.
[50] Y. Altuntaş, Z. Cömert, and A. F. Kocamaz, "Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach," Computers and Electronics in Agriculture, vol. 163, p. 104874, 2019.
[51] Y. Yadav, S. Chand, R. C. Sahoo, B. M. Sahoo, and S. Kumar, "Comparative analysis of detection and classification of diabetic retinopathy by using transfer learning of CNN based models," Journal of Intelligent and Fuzzy Systems, no. Preprint, pp. 1-15, 2022.

HOPE DIAMANTOPOULOS was a research assistant in the School of Computing at Montclair State University. She received her B.S. in Computer Science from Montclair State University and an M.S. in Machine Learning from Stevens Institute of Technology. Her research interests include robotics, artificial intelligence, machine learning, and neurotechnology.

ZOFIA PRZEDWORSKA was a research assistant in the School of Computing at Montclair State University. She received her B.S. in Computer Science from Montclair State University. Her research interests include robotics, smart glove technologies, machine learning, and artificial intelligence.

OMAR OBIDAT was a research assistant in the School of Computing at Montclair State University. He received his M.S. in Computer Science from Montclair State University. His research interests include robotics, human-robot interaction, and artificial intelligence.

JESSE PARRON (Member, IEEE) is currently a research associate in the Collaborative Robotics and Smart Systems Laboratory and instructor in the School of Computing, Montclair State University. His research interests include collaborative robotics, machine learning, Artificial Intelligence, and sensor technologies.



WEITIAN WANG (Senior Member, IEEE) is currently an Associate Professor in the School of Computing and the Founder Director of the Collaborative Robotics and Smart Systems Laboratory (CRoSS Lab) at Montclair State University. His research focuses on collaborative robotics, artificial intelligence, smart systems, and their synergistic CRoSS-disciplinary applications in smart manufacturing and remanufacturing, healthcare, smart ecosystem, intelligent transportation, smart agriculture, daily assistance, data science, cybersecurity, and interactive learning. He is the recipient of IEEE R1 Technological Innovation Award.