Su, Y. S., Suen, H. Y., & Hung, K. E. (2021). Predicting behavioral competencies automatically from facial expressions in real-time video recorded interviews. *Journal of Real-Time Image Processing, 18*(4), 1011–1021. https://doi.org/10.1007/s11554-021-01071-5

Predicting Behavioral Competencies Automatically from Facial Expressions in Real-Time Video-Recorded Interviews

Yu-Sheng Su, Hung-Yue Suen, Ku-En Hung

Abstract

This work aims to develop a real-time image and video processor enabled with an artificial intelligence (AI) agent that can predict a job candidate's behavioral competencies according to his or her facial expressions. This is accomplished using a real-time video-recorded interview with a histogram of oriented gradients and support vector machine (HOG-SVM) plus convolutional neural network (CNN) recognition. Different from the classical view of recognizing emotional states, this prototype system was developed to automatically decode a job candidate's behaviors by their microexpressions based on the behavioral ecology view of facial displays (BECV) in the context of employment interviews using a real-time video-recorded interview. An experiment was conducted at a Fortune 500 company, and the video records and competency scores were collected from the company's employees and hiring managers. The results indicated that our proposed system can provide better predictive power than can human-structured interviews, personality inventories, occupation interest testing, and assessment centers. As such, our proposed approach can be utilized as an effective screening method using a personal-value-based competency model.

Keywords

Behavioral Ecology View of Facial Displays (BECV), Convolutional Neural Network (CNN), Employment Selection, Histogram of Oriented Gradients (HOG), Real-time Image and Video Processing, Support Vector Machine (SVM)

1. **Introduction**

Competency refers to "a set of behavior patterns that the incumbent needs to bring to a position in order to perform its tasks and functions" [1]. This can help to predict how a job candidate will perform or behave at a specific job for which he or she is applying [2]. Therefore, competency is also called behavioral competencies that reflect the attributes underpinning a behavior, including knowledge, skills, abilities, and other characteristics (KSAOs) associated with successful performance in an area of work [3]. Simply put, behavioral competencies are any observable characteristics that can distinguish higher performers from others in an organization [4]. Some underlying characteristics, such as personality traits and values, play more important roles in success than technical knowledge and skills [5].

There are different approaches for developing competency models in human resource management [6]: the job-based approach, the future-based approach, the person-based approach, and the value-based approach.

Job-based competency defines what should be done for a specific job and role and is commonly adopted with a static context with specific job duties and requirements. Future-based competency defines what will be done in an organization in the future and requires a clear organizational vision, roadmap, and expected behaviors based on existing culture. Person-based competency defines which generic personal attributes are translated into behavioral patterns that can support the human capital advantage in an organization (not for a specific job), such as individual intelligence and growth mindset at Microsoft. This is suitable for organizations that focus on innovation and flexibility in the context of dynamic change. Value-based competency defines how things should be done by everyone in an organization and is useful when an organization wants to promote core values that may not gain a short-term advantage but can achieve sustainability. In a rapid change environment, an increasing number of organizations adopt person-based plus value-based competency models to lead organizational transformation.

When generic (person-based) and core (value-based) competencies can be identified successfully, the competency model can be applied to many organizational activities, especially personnel selection [7]. To assess whether a job candidate's generic person-based or value-based competency is valid, human experts need to assign ratings [8], such as during employment interviews that are commonly used in personnel selection [9]. The interview questions can focus specifically on experiences (e.g., what have you done?) or situations (e.g., what would you do?), or can focus more broadly on job-related KSAOs in both behavioral (e.g., tell me about your experience with…) or situational formats (e.g., tell me about your knowledge of…). The logic behind the employment interview is that a candidate's interview performance can predict his or her future behavior associated with prospective competencies [9] according to the candidates' verbal and nonverbal responses [10].

However, it is not cost-effective to interview every job candidate [11], and predictive validity is concerned with the relationship between interview performance and actual competencies due to interviewer-respondent interactions [12] and interviewees' fake behaviors [13]. Some interdisciplinary scholars in the computer science and human resources fields have suggested that video-recorded interviews, also called asynchronous video interviews (AVIs), can be used as an alternative to synchronous interviews, including face-to-face interviews, phone interviews, and conferencing interviews that require both the interviewer and interviewee to be present at the same time. In an AVI, the interviewee can answer predefined text questions at his or her convenience, and then the recordings of these answers are evaluated by a human [14]. Moreover, a job candidate's nonverbal responses, including facial expressions, are not easy to fake compared with verbal responses during an employment interview [10]. Recordings of facial movement can be used as more reliable cues to predict a job candidate's emotional state [15], personality traits [16], and communication skills [17] by computer vision enabled with deep learning. Hirevue.com is an example that combines artificial intelligence (AI, a deep-learning agent) and the Internet of Things (IoT: the AVI on mobile devices) to automatically predict a job interviewee's future performance according to his or her facial expressions [18].

A neuroscience study found that primates' (as well as humans') facial expressions are disclosed as a part of an embodied multimodal system (communication, emotional experience, and cognitive aspects that work together) that can predict the actor's likely future behavior [19]. This is not limited to emotional states [20] or social perception by observers [21]. Therefore, a job candidate's facial expressions recorded by an AVI platform is a good indicator of his or her competencies.

The study aims to develop a deep-learning agent that can detect job candidates' facial expressions according to the interview records from an AVI interface, while the agent can utilize the cues to predict their generic and core competencies in real-time based on a person plus value-based competency model of a Fortune 500 computer company in 2020.

The rest of the paper is organized as follows: related works are reviewed in section 2; the experimental procedure, image processing, and modeling are proposed in section 3; and the results are presented in section 4. Finally, the discussion and conclusions are drawn in section 5.

2. **Related works**

2.1 **Decoding facial expressions**

Research on human facial expressions has two different approaches [22]: basic emotion theory (BET) and the behavioral ecology view of facial displays (BECV). BET is also known as the "common view" or "classic view" [23] and assumes that human internal feelings and emotions are externalized by facial expressions at both the macro and micro levels [24]. Macroexpressions can be subjectively identified by human observation and categorized into six universal basic emotions: happy, fear, sad, surprise, anger, and disgust. By comparison, microexpressions are provoked involuntarily in brief and short durations, and they are not easy to recognize by the human eye [25]. Technically, the conceptual difference between macro- and microexpressions is their duration, and the threshold is 1/5 of a second [65][66]. From the perspective of BET, a person's emotional state can be recognized by his or her facial muscle movement [23]. For example, smiling reveals happiness [22].

However, several studies have discovered that a person's facial expressions do not reflect his or her true feelings [26]. Instead, a facial expression is a signal that reflects a person's behavioral intention toward any social interactants, including human or nonhuman objects, which is the other view of BECV [22]. From the perspective of BECV, facial expressions are social tools that signal and predict what the sender wants to happen next but are not a mirror about himself or herself [27]. For example, smiling is a tool used to influence interactants to play or affiliate and is not about happiness. A "disgusted" face may not indicate that you are disgusted with the interactant (e.g., an AVI interface or an AI agent), but it may signal that you want it to convey a different interaction [22]. This is not to say that people always attempt to manipulate others with their facial expressions, but this may well be instinctive when an interactant is not physically near [15][28] because people are nerve along psychologically [22][22].

In line with a competency model based on person-based plus value-based approaches [6], employers are more interested in predicting a job candidate's future behavior rather than his or her feelings during job interviews. Therefore, the BECV is more appropriate for recognizing facial expressions to predict the sender's behaviors in the context of an employment interview.

In the field of computer science, most related studies have been conducted based on BET to recognize a person's facial expressions and match or label them with a person's different emotional states from video datasets [15]. In contrast, this study was conducted based on BECV to recognize job candidates' facial expressions and match and label them with their observable behaviors as consistent with the target company's core values on an individual basis. Thus, facial expressions are conceptualized as indicators of the actors' (such as the job candidates in a job interview) likely future behaviors [19] in this study.

2.2 **Recognizing facial expressions**

Facial expression recognition can be divided into four major processes: face detection, preprocessing, feature extraction, and classification [15]. Facial detection refers to identifying the region(s) of interest (ROIs) of the human face in the frames of large image sequences [28]. Traditionally, the Viola-Jones object detection framework [29] combined with Haar-like features, integral images, the AdaBoost algorithm, and Cascade classifier stages is commonly used to classify a human face and discard nonfaces quickly on mobile devices [30]. This framework is also implemented to autodetect facial expressions in OpenCV, an open-source computer vision library [31].

Preprocessing aims to suppress unwanted distortions and enhances some features of facial expressions. Then, the facial images can be cropped using a bounding box and resized to allow for faster processing to obtain the feature descriptors [32]. Moreover, using a grayscale model can reduce the effect of shooting conditions, cosmetics, and background noise [33].

Feature extraction is used to identify important features or data attributes, which is a vital process in facial expression recognition [34]. Feature detection and feature extraction are two primary activities under facial feature extraction [15]. Discriminative response map fitting (DRMF) is an oft-performed Holistic texture-based model for detecting facial features from each facial expression in a video clip. A total of 68 landmark points are annotated on the face [35] and can serve as the input for the patterns of facial expression recognition [36].

To extract the features, each image is divided into small spatial regions, and then the gradients and orientation are calculated for every pixel in the image. The method of the feature extraction process is called a histogram of oriented

Su, Y. S., Suen, H. Y., & Hung, K. E. (2021). Predicting behavioral competencies automatically from facial expressions in real-time video recorded interviews. *Journal of Real-Time Image Processing, 18*(4), 1011–1021. https://doi.org/10.1007/s11554-021-01071-5

gradients (HOG) [37]. These features can be used to train machine learning algorithms such as linear support vector machines (SVMs) that can perform classification using a nonlinear decision boundary with a polynomial kernel [38]. Dlib is a pretrained facial landmark detector and library based on HOG and SVM for automatically extracting facial features, and users can customize the model at their own discretion [39, 40]. Given recent advances in deep learning, robust facial landmark detectors have been developed based on convolutional neural networks (CNNs). These detectors include "Dlib CNN" and outperform handcrafted feature methods [41], such as "Dlib HOG-SVM." However, CNN-based methods require more memory power and are more time consuming in real-time feature extraction, whereas HOG-SVM-based methods can run faster with fewer memory resources and produce more accurate results for extracting nonocclusions and frontal facial landmarks [42-44] in the context of AVI.

Classification refers to analyzing facial expression features and organizing the data into predefined categories [45]. The CNN is a widely used, high-performance, and simple method for classifying facial movement features into different categories [46], primarily for emotional states based on either macro- [47] or microexpressions [48]. Other than emotional states, facial expressions have been used as clues for understanding human behavioral intentions by using a CNN from open-source image datasets [49] or video frames of television series [50].

3. **Methodology**

3.1 **Participants**

To collect facial expression data in the context of AVI, we solicited a sponsoring organization that provided AVI records of new employees who passed their probationary period when provided with appropriate training. Additionally, included were their competency ratings (personal value-based) from their immediate supervisors. The organization is a global original design manufacturer providing services for information and communication technology products, including desktop systems, information appliances, handheld devices, storage and servers, networking and communication products, and notebook PCs. Its headquarters is located in Taiwan, and it does business in Asia, Europe, and North America. Its total revenue in 2020 was ranked on the list of global Fortune 500 companies [51]. A total of 298 newly hired employees from different functions were asked by the company to participate as AVI interviewees in this study, and the interviewees' competency ratings were evaluated during their probation (training for six months) by their immediate supervisors (239 hiring managers). The newly hired employees' profiles are listed in Table 1.

Table 1 Data Profile

Function		Gender		Average age
Research & Development	158	Male	210	35.16
Engineering	40	Female	88	32.73
Sales	29			
Marketing	33			
Information System	29			
Finance	3			
Administration	6			
Overall	298		298	34.44

3.2. **Procedure**

The participants were required to answer six core competency interview questions on mobile applications through an AVI platform developed by the authors before being hired as a permanent employee. The six behavioral-based structured interview questions were developed and used by the company to interview job candidates of different functions. These questions were asked when screening every full-time job candidate by the hiring managers to ensure that new hires had a certain level of generic (personal-based) core (value-based) competencies to fit the company's digital transformation. Each interviewee answered the same six questions in order, as listed in Table 2.

Table 2 Competency Interview Questions and Rating Scale

Competency	**Interview Questions**
Perseverance	Q1: What was the greatest obstacle you had to overcome in the past? Please describe the situation and how you handled it?
Openness	Q2: When you joined a new team, how did you engage it?
Initiation	Q3: Tell me about a time when you took initiative on a team, considering the previous question.
Innovation	Q4: Explain a new digital tool (e.g., Microsoft Planner, Teams, Slack, RPA, or Python) that you have tested. Describe how you pitched it and what the results were.
Critical thinking	Q5: Give me an example when you determined that your manager or a coworker made an incorrect decision. What did you do?
Risk taking	Q6: How did you decide to proceed when all conditions were variable in work outcomes?

The interviewees were invited by their HR department to sign up for the AVI on any Android or iOS mobile device, and the interviewees could decide when they were ready to start the interview within a week. The software guided them through the interview step by step, and the interviewees were informed that their answers would be used as a reference for the hiring manager's decision and that their facial expressions would be analyzed by an AI algorithm to predict the six competencies. The questions were displayed on the screen, and one minute was allowed to think after each question was announced. Audiovisual function was automatically started upon entering the answer screen. Three minutes were provided to answer each question. If an interviewee completed the question within three minutes, they could choose to skip to the next question, or the system would automatically move on to the next question after three minutes. The entire AVI process for each interviewee lasted approximately 15-18 minutes. In this study, we only collected records for the interviewees' facial expressions to predict their competencies. The interview ratings by the hiring manager were used only as a reference for the company's internal staffing decision. The ratings were not disclosed to the authors and were not used in this study.

After a three-month probationary period for the interviewees who were formally hired by the company, we collected the competency ratings evaluated by each interviewee's immediate supervisor regarding his or her observations toward the interviewee's six competencies based on the six competency behavioral incidents and rating scale without watching the video. The results are listed in Table 3.

Table 3 Competency Rating Scale

Competency	No	Behavioral incidents	α
Risk taking	1	Take actions when the probability of success is unclear.	0.805
	2	Take decisions that involve risk.	
	3	Try new but unproven approaches to solving problems.	
Perseverance	4	Achieve the goal when the objectives are made.	0.772
	5	Try new approaches for changed situations	
	6	Adjust effectively to frequently changing assignments.	
	7	Keep trying to achieve goals despite obstacles.	
Openness	8	Collect different ideas and explore their opportunity to implement.	0.782
	9	Generate new ideas and make them happen.	
	10	Be willing to implement new ideas from others.	
Initiation	11	Gather extra data that might be useful for work.	0.868
	12	Explore alternatives and positions to reach outcomes in difficult situations.	
	13	Conduct risk analysis and initiate action despite outcome uncertainty	
	14	Analyze the question beyond the question.	
Innovation	15	Try new approaches in an uncertain situation.	0.736
	16	Identify different approaches to achieve the same goals.	
	17	Explore new methods to solve problems with colleagues.	
Critical thinking	18	Question the legend points.	0.664
	19	Clarify issues and answers through debate.	
	20	Think outside the box.	
	21	Question established norms or rules.	

Rating scale (Five-points):
5: Demonstrates relevant behaviors accurately, consistently, and independently as a good example.
4: Demonstrates relevant behaviors accurately and consistently in most situations with minimal guidance.
3: Demonstrates relevant behaviors accurately and consistently on familiar procedures and needs supervisor guidance.
2: Demonstrates relevant behaviors inconsistently, even with repeated instruction or guidance.
1: Fails to demonstrate relevant behaviors regardless of the guidance provided.

Every interviewee who passed the probationary period had their six competency ratings averaged by three to four items, as listed in Table 3. All the Cronbach's alpha (α) as Eq. (1) of the six competency dimensions were acceptable (>0.6), as shown in Table 3. This indicates that the items can be grouped within a specific dimension for the same construct [52].

$$\frac{N*\bar{c}}{\bar{v}+(N-1)*\bar{c}} \quad (1)$$

N = the number of items
$\bar{c}$ = average covariance between item pairs
$\bar{v}$ = average variance

3.3 **Detection and preprocessing**

In the stage of facial expression detection, we utilized OpenCV to detect each participant's facial expressions frame-by-frame from the AVI records at a frame rate of 20 images per second (FPS) [53]. We resized all frame images at a fixed 640-pixel width to eliminate the variability in different image frames caused by rotation and shifting [53]. Moreover, we located facial landmark points on each image, as shown in Fig. 1, to capture the face ROI and facial deformations.

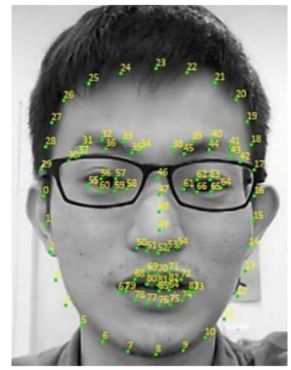

Fig. 1 Facial landmark points

With regard to the rotation, we set the top landmark point at 8, the forehead and the bottom landmark at point 23, and the chin to calculate the angle to align the face for all images as Eq. (2)-(7):

$$\text{Vertical Line: } (x1, y1) - (x1, y2) \quad (2)$$
$$\text{Landmark Point 8 to 23: } (x1, y1) - (x2, y2) \quad (3)$$
$$L1 = (Xa, Ya). Xa = 0, Ya = y2 - y1 \quad (4)$$
$$L2 = (Xb, Yb). Xb = x2 - x1, Yb = y2 - y1 \quad (5)$$
$$\cos\theta = \frac{L1 \cdot L2}{|L1||L2|} = \frac{XaXb+YaYb}{\sqrt{Xa^2+Ya^2}\cdot\sqrt{Xb^2+Yb^2}} \quad (6)$$
$$\text{Angle} = \text{acos}\left(\frac{XaXb+YaYb}{\sqrt{Xa^2+Ya^2}\cdot\sqrt{Xb^2+Yb^2}}\right)\cdot\frac{180}{\pi} \quad (7)$$

Afterward, the images were rotated and cropped to cover the entire face and transformed to grayscale with the purpose of reducing the noise (e.g., cosmetics, hair, and illumination) and normalizing the large-scale images [49] for the preprocessing stage. This is shown in Fig. 1.

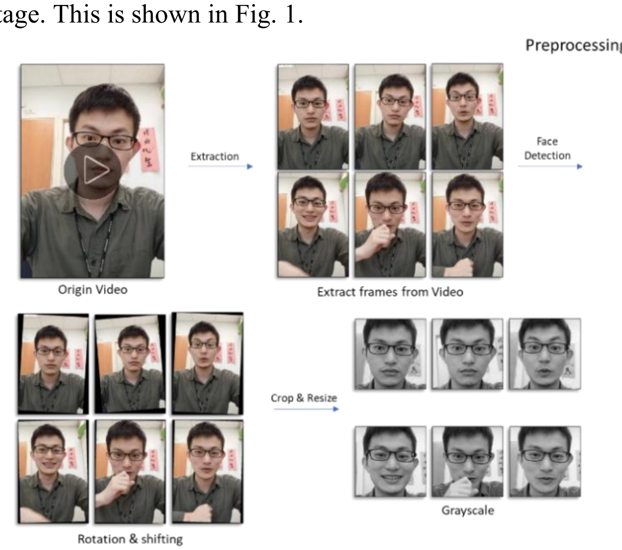

Fig. 2 Detection and preprocessing of facial expression

3.4 **Feature extraction**

In the feature extraction stage, for better results, we located 86 landmark points, including the forehead region (more than 68 points were used in DRMF without the forehead region), in each image [54], as shown in Fig. 1, by using Dlib HOG-SVM to capture the interviewees' microexpressions with a vector size of 4,096 values. This is shown in Fig. 2.

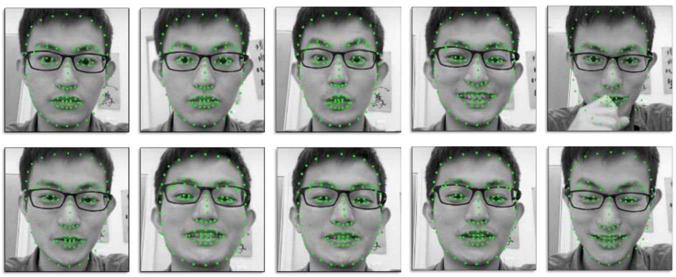

Fig. 2 Feature extraction of facial expressions

3.5 **Classification**

After collecting data as mentioned in section 3.2, we utilized a CNN in our proposed framework of different facial expressions to classify the six competency scores, including each mean score (mean), standard deviation (SD), maximum score (max), and minimum score (min), as shown in Table 4.

Table 4 Descriptive statistics of competency scores

Competency	Mean	SD	Max	Min
Risk taking	3.41	0.78	5.00	1.33
Perseverance	3.98	0.63	5.00	1.50
Openness	3.38	0.77	5.00	1.00
Initiation	3.60	0.78	5.00	1.25
Innovation	3.73	0.67	5.00	1.67
Critical thinking	3.33	0.65	5.00	1.50

N=298

Every interviewee was assigned six competency scores, and we found that the interviewees' job function, gender, and age did not influence their six competency scores based on an analysis of variance (ANOVA) and Pearson correlation analysis. Afterward, we trained and tested the six competency models separately. We randomly split the 298 data points into a training set (148), testing set (75), and validation set (75) at a 50-25-25 ratio [55]. The process aimed to build a model that can discriminate the six competency scores represented by the extracted facial expression features.

Our proposed models for the six competencies were conducted by TensorFlow CNN (version 1.15.2), as illustrated in Fig. 3. The network consists of three convolutional layers followed by a max-pooling layer. A rectified nonlinear unit (ReLU) was used with convolution in the CNN models to decrease the vanishing gradient problem in a sigmoid function [56].

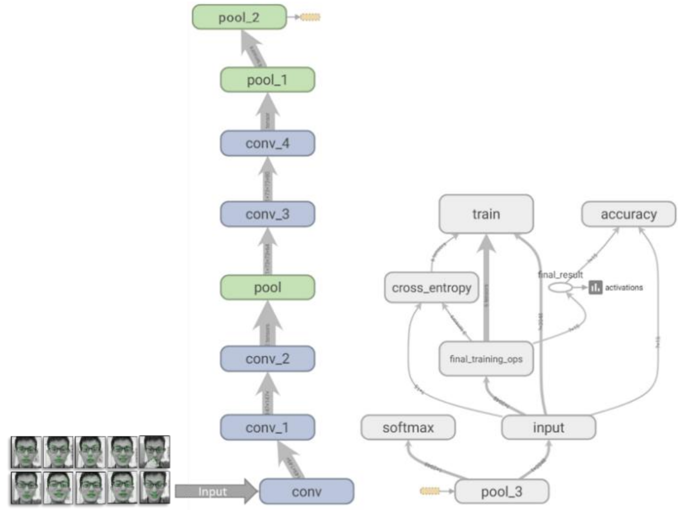

Fig. 3 Proposed CNN models for classification

Conv layer filters the 640×640 input images. The conv layer reduced the image to 320×320×32. Conv_1 and conv_2 took the output of conv as input and scanned the images to create a feature map and reduce the image to 160×160×64. The pooling layer reduces the feature tensor obtained in conv_2 and reduces the image to 80×80×64. The same flow is output in conv_3 and conv_4 to pool_1 and pool_2 with a feature map and image size of 8×8×2048. The final convolution with a special 1×1 filter followed by average polling collected the facial information in a 1×1×4,096 feature map. The output of the final pool was passed to the softmax classifier for the classification of competency scores. The input block was used to save the weight for the next training step and the statistics. We split each output into 100 classes for the six competencies; therefore, the final layer of the model was a softmax layer with 600 (6 × 100) possible outputs.

A learning rate dropout (LRD) of 0.01 was applied to reduce overfitting. This was followed by a sequence of four convolutional layers. The first two had a filter size of 640 each and then downsized to 320, 160, 80, and 40 each. A single max-pooling layer followed these four layers with a dropout rate of 0.25. To convert the output into a single-dimensional vector, the output of the previous layers was flattened. A fully connected layer was subsequently used along with an additional dropout rate of 0.5. Eventually, a fully connected layer with a softmax activation function served as the output layer [57]. Then, 4,000 training iterations were conducted with a 0.01 learning rate and 10 evaluation frequencies in this process. Moreover, we inserted batch normalization between the convolutions to normalize the inputs to layers within the network to prevent overfitting of the model [67].

4. Results

The proposed CNN models for the six competencies were trained for 256 epochs independently and reached 83%-85% validation accuracy based on our predictive scores (X_i) and the supervisors' rated scores (Y_i) for the interviewees' six competencies in this field study. The training and validation loss and the validation accuracies for the six CNN-based models are shown in Fig. 4.

Su, Y. S., Suen, H. Y., & Hung, K. E. (2021). Predicting behavioral competencies automatically from facial expressions in real-time video recorded interviews. *Journal of Real-Time Image Processing, 18*(4), 1011–1021. https://doi.org/10.1007/s11554-021-01071-5

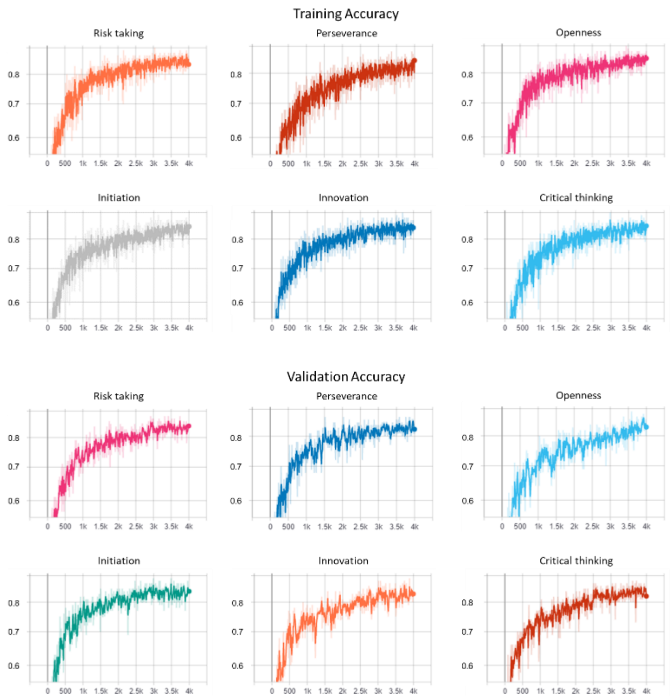

Fig. 4 Accuracies of training and validation for the six CNN-based model

Competency prediction by facial expressions was performed using AVI plus CNN, and we followed [16, 17] to present the accuracy, F1 score, Pearson correlation coefficient (R), explained variation (R^2), and mean square error (MSE) to measure the predictive validity of our proposed models. This is shown in Table 5.

The accuracy indicates the percentage of correct predictions as Eq. (8):

$$\left(\frac{\sum diag(n)}{n}\right) \tag{8}$$

The F1 score measures the overall model's accuracy and combines precision and recall as Eq. (9):

$$\left(\frac{\sum_{i=1}^{n} f_1(X_i)}{\mathrm{n}}\right) \tag{9}$$

R measures the strength of the association between the predictive scores of our modeling and the supervisors' rated scores as Eq. (10):

$$\left(\frac{\sum_i (x_i-\bar{x})}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(x_i-\bar{x})^2}}\right) \tag{10}$$

R^2 represents the proportion of the variance in the supervisors' rated scores that can be explained by the predictive scores from the facial expressions as Eq. (11):

$$\left(\frac{\sum_i (x_i-\bar{x})}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(x_i-\bar{x})^2}}\right)^2 \tag{11}$$

MSE calculates the sum of squared distances between the supervisors' rated scores and predicted scores. In contrast to the previous metrics, MSE is a risk factor. Thus, the smaller the MSE is, the better the model fits the supervisors' ratings, as shown in Eq. (12):

$$\left(\frac{\sum(\hat{Y}-Y)^2}{\mathrm{n}}\right) \tag{12}$$

The results in Table 5 show that the facial expressions extracted by our model can predict an interviewee's six competencies as required by the Fortune 500 company. The measures of accuracy and F1 outperform the state-of-the-art methods of facial expression recognition and were between 60% and 80% [49]. In terms of predictive power, our models can explain 65% to 74% of the supervisors' ratings toward their employees' competency scores. With regard to personnel selection, the correlations of our model achieved 0.81 to 0.87, which was far greater than the operational validities of structured employment interviews at 0.41, personality inventory at 0.14 to 0.25, occupational interest testing at 0.34, and assessment centers at 0.37 for training performance [58], similar to the probationary evaluation in this study. Moreover, the MSEs are all less than one SD for the six competency scores, which indicates that the risk of error is acceptable. Therefore, our proposed models can be considered alternatives to traditional selection tools for screening candidates and predicting their training performance after the probationary period. In addition, the cost of AVI plus an AI function is less than that of human-conducted selection tools, and human biases can be decreased by the AI function [14, 16, 17].

Table 5 Metrics of Modeling Results

Competency	Accuracy	F1	R	R^2	MSE
Risk taking	85%	78%	82%	67%	0.11
Perseverance	83%	82%	81%	65%	0.13
Openness	83%	79%	87%	76%	0.04
Initiation	84%	78%	84%	71%	0.11
Innovation	83%	79%	81%	65%	0.12
Critical thinking	83%	83%	86%	74%	0.12

5. Conclusions

In this study, we developed a prototype system that implements a HOG-SVM detection and feature extraction method of facial expressions plus a CNN classification architecture with fewer data [59] to significantly predict the personal- and value-based competencies in a field environment. The interviewees' facial expressions were extracted by an AVI platform, and then they were analyzed by deep learning to successfully converge the patterns of facial expressions associated with the six competency scores evaluated by the employees' immediate supervisors.

The study results support and highlight the theory of BECV, which assumes that human facial expressions can reflect future behaviors [22]. In addition, the novel system can automatically mark, detect, and extract an interviewee's microexpressions and output the predictive scores of different value-based personal competencies. This can replace human interviewers and traditional competency assessments in the stage of screening job candidates with higher predictive validity and lower selection cost.

To the best of our knowledge, there are no previous studies that used facial expressions (neither macro nor micro) as predictors to foresee personal-value-based competencies in the context of employment interviews. Therefore, our study proposed a state-of-the-art human resource selection and assessment method that can be used to predict behavioral competencies automatically from facial expressions in asynchronous video interviews.

Despite fewer data from a single company, the results achieved were promising. These results indicate that novel image processing combined with a deep-learning approach of such a framework can be a better method for producing a discriminative model for competency assessment based on facial expressions. This can be an indicator of future behavior within social interaction as opposed to the current emotional state alone, which is one of the current challenges in the field of AI applications of facial expressions [22]. Simply put, AI can estimate how you will behave by reading your microexpressions [60, 61, 68-75].

The present study leaves a few intriguing questions open for future exploration. First, we focused on specific categories of a personal-value-based competency model that varied across organizations and industries. Future research can explore whether the effects described here also converge for other competency models, such as job-based or future-based competency approaches for knowledge, skills, or cognitive abilities.

Second, although our proposed system achieves better performance, the results are explored for a specific company and culture. Future work will be extended to other organizations in different sectors and countries.

Third, this study carried out a probationary evaluation to label competencies as criteria and achieved higher validity compared to the other selection tools. In the future, we will collect task performance records as criteria to examine criterion-related validity.

Finally, we utilized OpenCV and Dlib HOG-SVM to detect and extract frontal facial expressions. However, more challenging aspects of detecting and extracting facial expressions should be considered, including orientation variance and pose changes. Moreover, to achieve better performance, other detection and extraction methods can be implemented and evaluated, including the main directional mean optical flow (MDMO) [61], facial dynamics mapping (FDM) [62], local binary pattern-three orthogonal planes (LBP-TOP) [63], and the Gabor wavelet filter [64].

References

1. Woodruffe, C.: What is meant by competency? In: Boam, R., Sparrow, P. (eds.) Designing and Achieving Competency, pp.1–29. McGraw-Hill, New York (1992)
2. Hofrichter, D.A., Spencer, L.M.: Competencies: the right foundation the right foundation for effective human resources management. Compens. Benefits Rev. 28, 21–26 (1996)
3. Moore, D., Cheng, M.I., Dainty, A.: Competence, competency and competencies: performance assessment in organisations. Work Study 51, 314–319 (2002)
4. Kochanski, J.T.: Introduction to special issue on human resource competencies. Hum. Resour. Manag. 35, 3–6 (1996)
5. Spencer, L.M., Spencer, S.M.: Competence at Work: Models for Superior Performance. John Wiley & Sons, New York (1993)
6. Cardy, R.L., Selvarajan, T.T.: Competencies: alternative frameworks for competitive advantage. Bus. Horiz. 49, 235–245 (2006)
7. Feltham, R.: Using competencies in selection and recruitment. In: Boam, R., Sparrow, P. (eds.) Designing and Achieving Competency. A Competency-based Approach to Developing People and Organizations. pp. 89–103. McGraw-Hill, London (1992)
8. Nikolaou, I.: The development and validation of a measure of generic work competencies. Int. J. Test. 3, 309–319 (2003)
9. Hartwell, C.J., Johnson, C.D., Posthuma, R.A.: Are we asking the right questions? Predictive validity comparison of four structured interview question types. J. Bus. Res. 100, 122–129 (2019)
10. DeGroot, T., Gooty, J.: Can nonverbal cues be used to make meaningful personality attributions in employment interviews? J. Bus. Psychol. 24, 179–192 (2009)
11. Nikolaou, I., Foti, K.: Personnel selection and personality. In: Zeigler-Hill, V., Shackelford, T. (eds.) The SAGE Handbook of Personality and Individual Differences. pp. 659–677. Sage, London (2018)
12. Huffcutt, A.I., Van Iddekinge, C.H., Roth, P.L.: Understanding applicant behavior in employment interviews: a theoretical model of interviewee performance. Hum. Resour. Manag. Rev. 21, 353–367 (2011)
13. Melchers, K.G., Roulin, N., Buehl, A.K.: A review of applicant faking in selection interviews. Int. J. Sel. Assess. 28, 123–142 (2020)
14. Suen, H.Y., Chen, M.Y.C., Lu, S.H.: Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? Comput. Hum. Behav. 98, 93–101 (2019)
15. Takalkar, M., Xu, M., Wu, Q., Chaczko, Z.: A survey: facial micro-expression recognition. Multimed. Tools Appl. 77, 19301–19325 (2018)
16. Suen, H., Hung, K., Lin, C.: TensorFlow-based automatic personality recognition used in asynchronous video interviews. IEEE Access 7, 61018–61023 (2019)
17. Suen, H.Y., Hung, K.E., Lin, C.L.: Intelligent video interview agent used to predict communication skill and perceived personality traits. Hum.-Centric Comput. Inf. Sci. 10, 3 (2020)
18. Hilke, S., Bellini, J.: Artificial intelligence: the robots are now hiring. The Wall Street Journal. https://www.wsj.com/articles/artificial-intelligence-the-robots-are-now-hiring-moving-upstream-1537435820 (2018). Accessed 20 September 2018
19. Waller, B.M., Whitehouse, J., Micheletta, J.: Rethinking primate facial expression: a predictive framework. Neurosci. Biobehav. Rev. 82, 13–21 (2017)
20. Fridlund, A.J.: Human Facial Expression: An Evolutionary View. Academic Press, San Diego, CA (1994)
21. Chanes, L., Wormwood, J.B., Betz, N., Barrett, L.F.: Facial expression predictions as drivers of social perception. J. Pers. Soc. Psychol. 114, 380–396 (2018)
22. Crivelli, C., Fridlund, A.J.: Facial displays are tools for social influence. Trends Cogn. Sci. 22, 388–399 (2018)
23. Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A.M., Pollak, S.D.: Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. Psychol. Sci. Public Interest 20, 1–68 (2019)
24. Ekman, P.: What scientists who study emotion agree about. Perspect. Psychol. Sci. 11, 31–34 (2016)
25. Ekman, P., Friesen, W.V.: Nonverbal leakage and clues to deception. Psych. 32, 88–106 (1969)
26. Crivelli, C., Carrera, P., Fernández-Dols, J.M.: Are smiles a sign of happiness? Spontaneous expressions of judo winners. Evol. Hum. Behav. 36, 52–58 (2015)
27. Fridlund, A.J.: The behavioral ecology view of facial displays, 25 years later. In: Fernández-Dols, J.M., Russell, J.A. (eds.) Oxford Series in Social Cognition and Social Neuroscience. The Science of Facial Expression. pp. 77–92. Oxford University Press, Oxford, UK (2017)
28. Rehman, B., Ong, W.H., Tan, A.C.H., Ngo, T.D.: Face detection and tracking using hybrid margin-based ROI techniques. Vis. Comput. 36, 633–647 (2020)
29. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. I–511–I–518. IEEE, Kauai, HI, USA, (2001)
30. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. 57, 137–154 (2004)
31. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro- and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pp. 51–56. IEEE, Santa Barbara, CA, USA, (2011)
32. Pitaloka, D.A., Wulandari, A., Basaruddin, T., Liliana, D.Y.: Enhancing CNN with preprocessing stage in automatic emotion recognition. Procedia Comput. Sci. 116, 523–529 (2017)
33. Yudin, D.A., Dolzhenko, A.V., Kapustina, E.O.: The usage of grayscale or color images for facial expression recognition with deep neural networks. In: Kryzhanovsky, B., Dunin-Barkowski, W., Redko, V., Tiumentsev, Y. (eds.) Advances in Neural Computation, Machine Learning, and Cognitive Research III, pp. 271–281. Springer International Publishing, Cham, (2020)

34. Sadeghi, H., Raie, A.A.: Human vision inspired feature extraction for facial expression recognition. Multimed. Tools Appl. 78, 30335–30353 (2019)
35. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Robust discriminative response map fitting with constrained local models. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3444–3451. IEEE, Portland, OR, USA, (2013)
36. Merget, D., Rock, M., Rigoll, G.: Robust facial landmark detection via a fully-convolutional local-global context network. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 781–790. IEEE, Salt Lake City, UT, USA, (2018)
37. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 886–893. IEEE, San Diego, CA, USA, (2005)
38. Carcagnì, P., Del Coco, M., Leo, M., Distante, C.: Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus 4, 645 (2015)
39. King, D.: Dlib-ml: A Machine Learning Toolkit. J. Mach. Learn. Res. 10, 1755–1758 (2009)
40. Csaba, B., Tamás, H., Horváth, A., Oláh, A., Reguly, I.Z.: PPCU sam: open-source face recognition framework. Procedia Comput. Sci. 159, 1947–1956 (2019)
41. Pursche, T., Clauß, R., Tibken, B., Möller, R.: Using neural networks to enhance the quality of ROIs for video based remote heart rate measurement from human faces. In: IEEE International Conference on Consumer Electronics (ICCE), pp. 1–5. IEEE, Las Vegas, NV, USA, (2019)
42. Johnston, B., Chazal, P.D.: A review of image-based automatic facial landmark identification techniques. EURASIP J. Image Video Process. 2018, 86 (2018)
43. Aslan, M.F., Durdu, A., Sabanci, K., Mutluer, M.A.: CNN and HOG based comparison study for complete occlusion handling in human tracking. Measurement 158, 107704 (2020)
44. Adouani, A., Henia, W.M.B., Lachiri, Z.: Comparison of Haar-like, HOG and LBP approaches for face detection in video sequences. In: 16th International Multi-Conference on Systems, Signals & Devices (SSD), pp. 266–271. Istanbul, Turkey, (2019)
45. Hammal, Z., Couvreur, L., Caplier, A., Rombaut, M.: Facial expression classification: an approach based on the fusion of facial deformations using the transferable belief model. Int. J. Approx. Reason. 46, 542–567 (2007)
46. Liu, Y., Zhang, J., Yan, W., Wang, S., Zhao, G., Fu, X.: A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Trans. Affect. Comput. 7, 299–310 (2016)
47. Mehendale, N.: Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020)
48. Zhao, Y., Xu, J.: A convolutional neural network for compound micro-expression recognition. Sensors (Basel, Switzerland) 19, 5553 (2019)
49. González-Lozoya, S.M., de la Calleja, J., Pellegrin, L., Escalante, H.J., Medina, M.A., Benitez-Ruiz, A.: Recognition of facial expressions based on CNN features. Multimed. Tools Appl. 79, 13987–14007 (2020)
50. Sajjad, M., Zahir, S., Ullah, A., Akhtar, Z., Muhammad, K.: Human behavior understanding in big multimedia data using CNN based facial expression recognition. Mob. Netw. Appl. 25, 1611–1621 (2020)
51. Fortune: Fortune 500 in 2020. https://fortune.com/fortune500/2020/ (2020). Accessed 11 August 2020
52. Taber, K.S.: The use of Cronbach's Alpha when developing and reporting research instruments in science education. Res. Sci. Educ. 48, 1273–1296 (2018)
53. Oh, Y.H., See, J., Le Ngo, A.C., Phan, R.C.W., Baskaran, V.M.: A survey of automatic facial micro-expression analysis: databases, methods, and challenges. Front. Psychol. 9, 1128 (2018)
54. Deng, J., Trigeorgis, G., Zhou, Y., Zafeiriou, S.: Joint multi-view face alignment in the wild. IEEE Trans. Image Process. 1 (2019).
55. Cıbuk, M., Budak, U., Guo, Y., Cevdet Ince, M., Sengur, A.: Efficient deep features selections and classification for flower species recognition. Measurement 137, 7–13 (2019)
56. Krishnaraj, N., Elhoseny, M., Thenmozhi, M., Selim, M.M., Shankar, K.: Deep learning model for real-time image compression in Internet of Underwater Things (IoUT). J. Real-Time Image Process 17, 2097–2111 (2019).
57. Saravanan, A., Perichetla, G., Gayathri, D.K.S.: Facial emotion recognition using convolutional neural networks. arXiv:1910.05602 (2019)
58. Schmidt, F.L., Oh, S., Shaffer, J.A.: The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings (Fox School of Business Research Paper). Temple University, Philadelphia, PA (2016)
59. Lopes, A.T., de Aguiar, E., De Souza, A.F., Oliveira-Santos, T.: Facial expression recognition with Convolutional Neural Networks: coping with few data and the training sample order. Pattern Recognit. 61, 610–628 (2017)
60. Smith, J.: You are what you will: Kant, schopenhauer, facial expression of emotion, and affective computing. Ger. Life Lett. 70, 466–477 (2017)
61. Poiesi, F., Cavallaro, A.: Predicting and recognizing human interactions in public spaces. J Real-Time Image Proc 10, 785–803 (2015)
62. Hannuna, S., Camplani, M., Hall, J. et al.: DS-KCF: a real-time tracker for RGB-D data. J Real-Time Image Proc 16, 1439–1458 (2019)
63. Bordallo López, M., Nieto, A., Boutellier, J. et al.: Evaluation of real-time LBP computing in multiple architectures. J Real-Time Image Proc 13, 375–396 (2017)
64. Pang, W., Choi, K., Qin, J.: Fast Gabor texture feature extraction with separable filters using GPU. J Real-Time Image Proc 12, 5–13 (2016)
65. Shen, X.B., Wu, Q., Fu, X.I.: Effects of the duration of expressions on the recognition of microexpressions. Journal of Zhejiang University SCIENCE B. 13, 221–230 (2012)
66. Queiroz, R.B., Musse, S.R., Badler, N. I.: Investigating Macroexpressions and Microexpressions in Computer Graphics Animated Faces. MIT Press. 23, 191–208 (2014)
67. Garbin, C., Zhu, X., Marques, O.: Dropout vs. batch normalization: an empirical study of their impact to deep learning. Multimedia Tools and Applications. 79, 12777–12815 (2020)
68. Dai, C., Liu, X., Lai, J., Li, P..: Human Behavior Deep Recognition Architecture for Smart City Applications in the 5G Environment. IEEE Network. 33, 206–211 (2019)
69. Su, Y.S., Chou, C.H., Chu, Y.L., Yang, Z.F.: A finger-worn device for exploring Chinese printed text with using CNN algorithm on a micro IoT processor. IEEE Access. 7, 116529–116541 (2019)
70. Dai, C., Liu, X., Lai, J.: Human action recognition using two-stream attention based LSTM networks. Applied Soft Computing. 86, 105820 (2020)
71. Su, Y.S., Lin, C.L., Chen, S.Y., Lai, C.F.: Bibliometric study of social network analysis literature. Libr. Hi Tech.. 38, 420–433 (2019)
72. Dai, C., Liu, X., Chen, W., Lai, C.F.: A low-latency object detection algorithm for the edge devices of IoV systems. IEEE Transactions on Vehicular Technology. 69, 11169–11178 (2020)
73. Su, Y.S., Chen, H.R.: Social Facebook with Big Six approaches for improved students' learning performance and behavior: A case study of a project innovation and implementation course. Front. Psychol.. 11, 1166 (2020)
74. Dai, C., Liu, X., Yang, L.T., Ni, M., Ma, Z., Zhang, Q. Deen, M.J.: Video scene segmentation using tensor-train faster-RCNN for multimedia IoT systems. IEEE Internet of Things Journal. (2020)
75. Su, Y.S., Ni, C.F., Li, W.C., Lee, I.H., Lin, C.P.: Applying deep learning algorithms to enhance simulations of large-scale groundwater flow in IoTs. Applied Soft Computing. 92, 106298 (2020)