Artificial Intelligence and Social Computing, Vol. 72, 2023, 100–110
https://doi.org/10.54941/ahfe1003278

Emotional Analysis of Candidates During Online Interviews

Alperen Sayar[1], Tunahan Bozkan[1], Seyit Ertuğrul[1], Mert Güvençli[2], and Tuna Çakar[3]

[1]Tam Finans Faktoring A.Ş, Sisli, Istanbul 34360, Turkey
[2]Apsiyon Bilişim Sistemleri San. Tic. A.Ş., Maltepe, İstanbul, Turkey
[3]MEF University, Sarıyer, Istanbul 34240, Turkey

ABSTRACT

The recent empirical findings from the related fields including psychology, behavioral sciences, and neuroscience indicate that both emotion and cognition are influential during the decision-making processes and so on the final behavioral outcome. On the other hand, emotions are mostly reflected by facial expressions that could be accepted as a vital means of communication and critical for social cognition, known as the facial activation coding in the related academic literature. There have been several different AI-based systems that produce analysis of facial expressions with respect to 7 basic emotions including happy, sad, angry, disgust, fear, surprise, and neutral through the photos captured by camera-based systems. The system we have designed is composed of the following stages: (1) face verification, (2) facial emotion analysis and reporting, (3) emotion recognition from speech. In this study, several classification methods were applied for model development processes, and the candidates' emotional analysis in online interviews was focused on, and inferences about the situation were attempted using the related face images and sounds. In terms of the face verification system obtained as a result of the model used, 98% success was achieved. The main target of this paper is related to the analysis of facial expressions. The distances between facial landmarks are made up of the starting and ending points of these points. 'Face frames' were obtained while the study was being conducted by extracting human faces from the video using the VideoCapture and Haar Cascade functions in the OpenCV library in the Python programming language with the image taken in the recorded video. The videos consist of 24 frames for 1000 milliseconds. During the whole video, the participant's emotion analysis with respect to facial expressions is provided for the durations of 500 milliseconds. Since there are more than one face in the video, face verification was done with the help of different algorithms: VGG-Face, Facenet, OpenFace, DeepFace, DeepID, Dlib and ArcFace. Emotion analysis via facial landmarks was performed on all photographs of the participant during the interview. DeepFace algorithm was used to analyze face frames through study that recognizes faces using convolutional neural networks, then analyzes age, gender, race, and emotions. The study classified emotions as basic emotions. Emotion analysis was performed on all of the photographs obtained as a result of the verification, and the average mood analysis was carried out throughout the interview, and the data with the highest values on the basis of emotion were also recorded and the probability values have been extracted for further analyses. Besides the local analyses, there have also been global outputs with respect to the whole video session. The main target has been to introduce different potential features to the feature matrix that could be correlated with the other variables and labels tagged by the HR expert.

Keywords: Face verification, Facial emotion analysis, Emotional recognition, Artificial intelligence in recruitment

INTRODUCTION

One of the most significant focus areas is the "workforce" management process, which is one of the most essential management techniques utilized and used by corporate policies to guarantee that all forms of production-to-service firms may continue to exist and grow their activities (Kathuria and Partovi, 1999). Businesses have been searching for effective and efficient methods to use their staff for decades. To manage and sustain this process with a maximally optimum knowledge and flow, it is evident that the process of delivering human resources must be addressed with care and diligence from the outset (Barney and Wright, 1998). In every recruiting process, it is still a subject of discussion that selecting the most qualified and well-equipped candidate is not the optimal strategy. It is vital to concentrate on individuals who will bring long-term added value to enterprises, who will own their position in company operations with their own will and desire, i.e., who have the drive and discipline to fulfill the established job description and strive to enhance it (Hamza, Othman, Gardi, Sorguli, Aziz, Ahmed, Sabir, Ismael, Ali and Anwr, 2021). Directly and indirectly, the company will benefit from the ability to identify the ideal applicant for this position at the earliest stage of the recruiting process. Since the emphasis is on people, making such a judgement is as challenging as possible. Artificial Intelligence is one of the most widely used and deployed methods for overcoming this difficulty for human-centered, predictable, and interpretable goals (van den Broek, Sergeeva and Huysman, 2021). "Biometric System" is considered as the artificial intelligence topic that might provide a solution to this challenge (Winston and Hemanth, 2019).

These systems, which are mostly utilized for security reasons in organizations nowadays, might vary based on their use requirements, and each firm can examine it independently. The most prevalent biometric systems include recognition, hand geometry detection, fingerprinting, DNA chain analysis, and so on (Dargan and Kumar, 2020).

DEFINING THE PROBLEM

A traditional approach remains in the current recruitment process. In this traditional approach, job interviews are not recorded for later evaluation, evaluations are made simultaneously by the relevant expert at the end of the interview, and after the meeting with the relevant position manager, an instant decision is made and the process is continued or stopped. At this point, question marks arise about determining the right person according to the mood of the decision makers in the candidate to be taken for the position. The fact that it is only based on business knowledge in choosing the right candidate, and the human opinion and emotional state vary, increases the anxiety about choosing the right candidate (Behroozi, Shirolkar, Barik and Parnin, 2020). Recent empirical findings from related fields such as psychology, behavioral sciences, and neuroscience show that both emotion and cognition are influential on decision-making processes and thus on the final behavioral outcome (Mobbs, Trimmer, Blumstein and Dayan, 2018).

METHODS

Within the scope of the project, at the beginning of the recruiting procedures, a technique titled "Emotional Analysis of Candidates Throughout Online Interviews" was used in order to begin with the facial expressions of the candidates during the interviewing process. As a consequence of this project, the ability to determine if an applicant is qualified for the job description for which they have applied has increased efficiency and decreased resource consumption throughout the recruiting process.

Within the scope of the project, an architecture consisting of 12 different steps was created (Figure 1). These are given as follows:

1. Application Reception
2. Schedule a Live Call
3. Environment Preparation
4. Transfer to Server
5. Processing with Python
6. Face Recognition
7. Face Verification
8. Emotional Analysis
9. Recording and Reporting of Analysis Results
10. Action Based on Results
11. 6–12 Monthly Performance Evaluation Notification
12. Supervised Learning (Classification).

Application Reception

An announcement was issued for the appropriate job within the scope of business requirements, and the project was begun after passing the "must have" filters defined by the Human Resources Department and the manager of the relevant post. In the experiment that began collecting data thirteen months ago, 634 adverts were opened. There were around 25,000 applications for a total of 634 advertising. The human resources department and manager of the job unit processed twenty-five thousand candidates via the

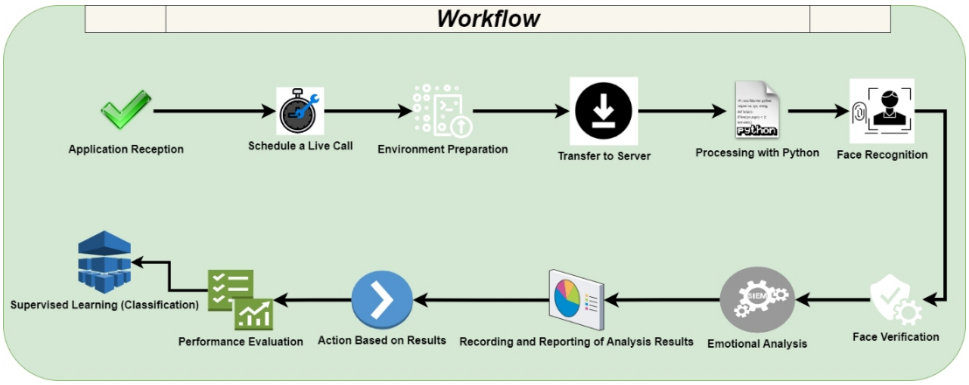

Figure 1: Emotional analysis workflow.

"must have" filters for the relevant roles. Following the screening procedure, 3,574 applicants were phone interviewed. With 2203 applicants, the live interview process began after the telephone interview.

Schedule a Live Call

Day, hour, ambient lighting, camera angle, internet connection, etc., are of the utmost importance for these sessions, which are the source of future project phases. The relevant candidates were provided with information on such concerns. For the average applicants who cleared the previous step and will be included in the live interview, a timetable for 5–7 days later has been set.

Environment Preparation

A server capable of storing the data of ten thousand persons was assigned at the outset of the project to guarantee the proper functioning of the project architecture. In order to be prepared for unanticipated and potentially dangerous server scenarios, a backup server with identical functionality was assigned.

Transfer to Server

Each interview is connected with a dedicated server, and photos of the persons' faces to be sampled from various angles via video recording are captured at certain frame intervals and stored on company-allocated local servers. So far, 546 composed of 64,639 minutes of recorded conversation have been uploaded to the server. These videos contain the records of online interviews stored in mp4 or similar format.

MODEL DEVELOPMENT

Processing with Python

Preparation is the transformation of data into a usable format for further analysis and processing. The correctness of unprocessed data must be verified prior to processing. Input is the process of encoding or converting validated data into machine-readable format so that it may be processed by an application. Utilizing a keyboard, scanner, or data input from an existing source, data is entered. This time-consuming operation needs speed and precision. At this point, a large amount of processing power is necessary to decompose complicated data, hence the majority of data must adhere to a rigorous and rigid grammar. Processing is when data is exposed to different strong technological transformations employing Machine Learning and Artificial Intelligence algorithms to provide an output or interpretation. The process may be made up of many threads of execution that concurrently execute instructions, depending on the kind of data (Arooj, Farooq, Akram, Iqbal, Sharma and Dhiman, 2022). (Arooj, Farooq, Akram, Iqbal, Sharma and Dhiman, 2022). In this perspective, data processing consisted of validating the original data's precision.

Face Recognition

Face recognition is a biometric application that employs image processing methods and pattern matching to identify an individual based on an image analysis and comparison. Face recognition systems use numeric codes derived from eighty unique spots on the human facial known as "faceprints" (Khan, Akram and Usman, 2020). It consists of all individual characteristics, such as cheekbones, the depth and structure of the eye sockets, and the length, breadth, and depth of the nose. The face recognition system was created by comparing these numerical codes or values with previously inputted numerical values in the database.

For facial recognition, the open-source packages OpenCV and Deepface were employed. A contemporary face recognition pipeline has five stages: detect, align, normalize, represent, and verify (Xu, Wang, Shou, Ngo, Sadick and Wang, 2021). While Deepface handles these frequent phases in the background, you do not require in-depth understanding of its underlying procedures (Du, Shi, Zeng, Zhang and Mei, 2022).

Deepface is a hybrid model for facial recognition. As of now, it encompasses several cutting-edge Face Recognition models (Figure 2). Facial recognition models essentially encode face pictures as vectors with several dimensions. Sometimes these vectors are required directly. Deepface includes a unique representation function. Deepface is essentially a Tensorflow Framework-based open source module (Wang and Deng, 2021).

Deepface feeds facial photographs into a model of an evolving neural network, but that is not the point. It employs CNN models to discover embeddings comparable to autoencoders (Fig. 3). (Abdulnabi, Wang, Lu and Jia, 2015).

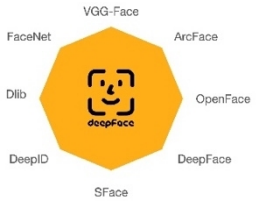

Figure 2: State-of-the-art face recognition models.

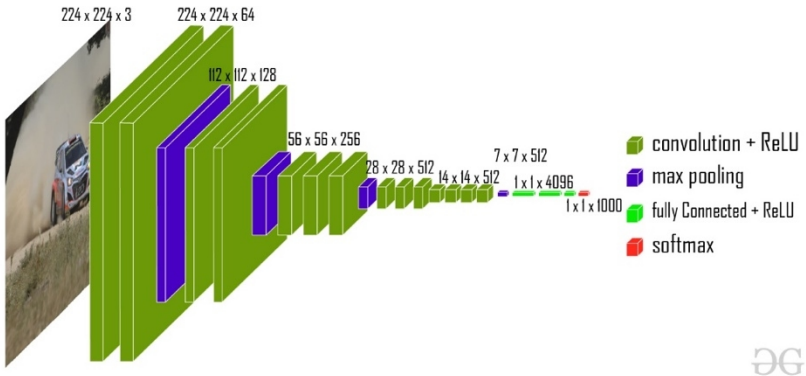

Figure 3: Convolutional neural network (CNN) model.

The candidate's facial recognition was done using Deepface and taught to the system. Thus, the process of facial recognition is concluded.

Face Verification

During the step of face verification, it was established if the face pairings belonged to the same or distinct individuals. The verify function of Deepface demands complete image paths as input. It also handles Numpy and base64-encoded picture transmissions. Then it returns a dictionary, and only after checking its confirmed key are we done validating. In cases where the interviewee for the relevant post could not be confirmed, the Face recognition step was retried and the relevant video clip was analyzed using various embedding techniques. The facial verification process was then repeated.

Emotional Analysis

In picture analysis, it was possible to identify a person's emotional state based on his facial expressions, providing an intuitive depiction of his mental state. Intuitive reflection was examined in relation to seven primary emotional states.

These are;

- Happy
- Sad
- Mad
- Disgust
- Fear
- Surprise
- Neutral

With Deepface, the score values were extracted for these 7 basic emotions.

RESULTS

Recording and Reporting of Analysis Results

The suitable candidate's emotional analysis was discovered and logged in the oracle database. The data was then examined using a number of free source visualization packages (AutoViz, Matplotlib, ggplot, seaborn etc.). In addition, QlikSense and QlikView, which are used by Tam Finans Faktoring Inc., were analyzed. The findings of these analyses were then immediately communicated to the Human Resources officer and the manager in charge of the relevant role.

Action Based on Results

After reporting the findings to the appropriate position's manager and the human resources officer, the procedure was either continued or terminated. Managers stressed that it is of more importance to them that the 'pleasant' and 'neutral' traits of the applicant predominate. Some employers want the applicant to be inquisitive and receptive to innovation, with "Surprise" dominating and "Fear" being as inconspicuous

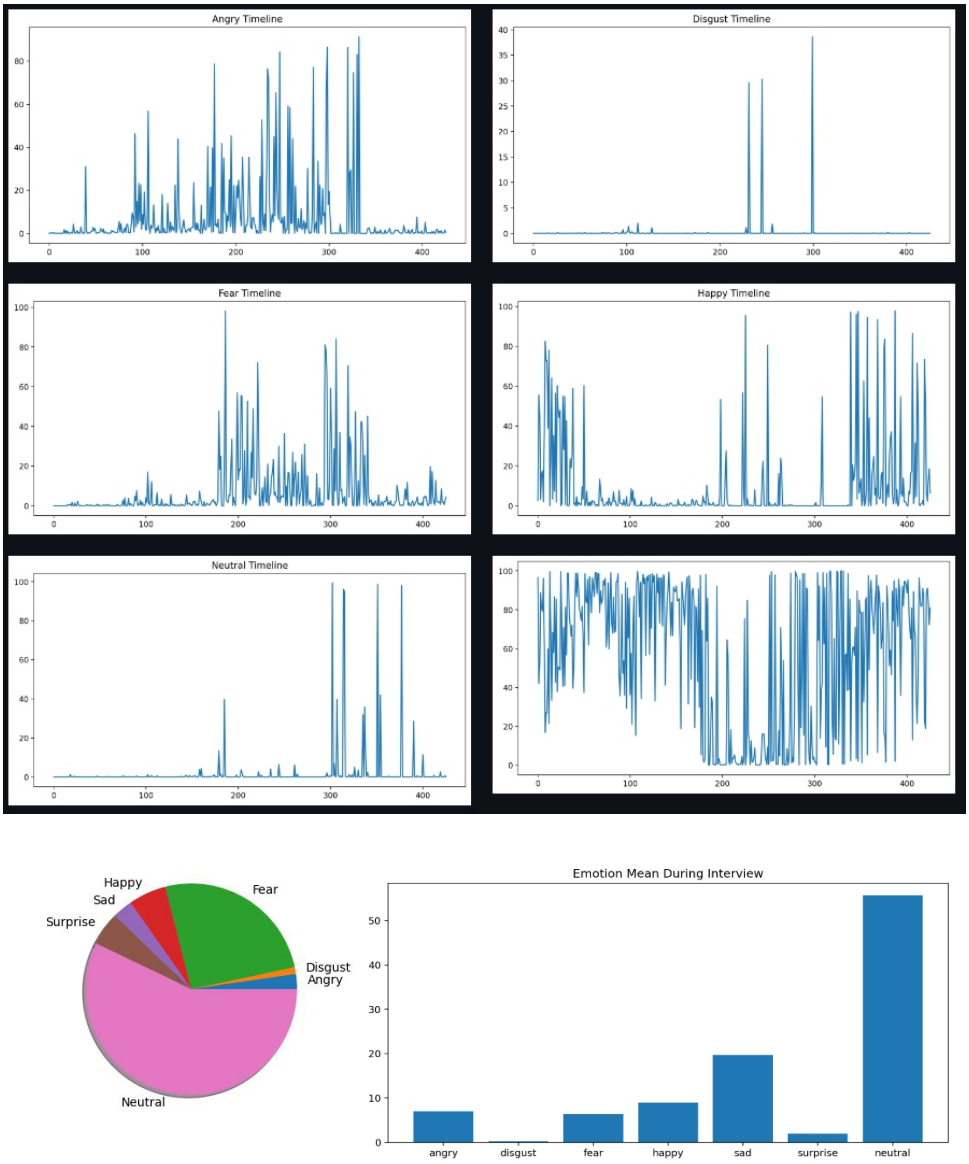

Figure 4: Emotion analysis reports during interview.

as possible. Managers modified these scenarios, and the process proceeded by examining the candidate's emotional analyses from a variety of angles.

6-12 Monthly Performance Evaluation Notification

455 individuals were selected from a pool of 2203 hopefuls. These recruited applicants' information was stored in a database. After their start date, their boss evaluated each employee's performance. Similarly, performance assessment results were transmitted to the database. At the conclusion of the sixth, 18 of the 455 people recruited for this position

voluntarily resigned. In terms of managers, the remaining 437 contenders' results were much better than in prior years. A 43% increase was seen compared to the 6-month performance review of the staff employed the prior year. In the 12th month, the same analysis was conducted for evaluation purposes. 13 of 437 applicants left the position willingly. The performance of the remaining 424 individuals was evaluated. Again, a 46.4% rise was seen compared to the previous year's recruiting performance evaluations.

Supervised Learning (Classification)

At the conclusion of a year, the examination of 424 candidates for categorization using each characteristic began. Our aim variable at this time was performance assessment outcomes. The performance assessment score was out of 100 points. For these 424 contestants, the lowest score was 76 and the best was 99. Their mean score was 89. The distributions were examined using visualizations, and the threshold value of 91 was selected based on the views of experts. Those with a performance score of less than 91 were assigned 0, while those with a performance score of more than 1 were assigned. There were a total of 149 candidates from 1 and 275 candidates from 0.

A 23-column data set with a total of 22 characteristics and one target variable was constructed. People have picture information on many lines, thus the average has been deduplicated. A dataframe with 424 rows and 23 columns was generated after deduplication. According to the distribution, corresponding assignments were performed for values of null. It was pulled from the third and first quarters for outliers. One-Hot encoding, Label Encoding, and Target Encoding conversions were used for string-type variables storing the applicants' demographic information.

The train test split function of the sklearn package was used to divide the dataset into training and testing by dividing the training set by 80% and the test set by 20%. In order to evaluate the precision of the training set, the cross val score metric from the sklearn framework was used, with cv set to 5.

In model assessment, prediction, recall, precision, and f1-score were considered. RandomForest, XGBoost, kNN, Decision Tree, Stochastic Gradient Descent, Support Vector Machine, Naive Bayes, and Logical Regression were the models used. XGBoost has been the most popular algorithm in recent years, and its success rate is even greater than that of competing models (Chen and Guestrin, 2016). The XGBoost approach yielded the greatest f1-score of 96%, and the HalvingGridSearchCV module was subsequently used to optimize the model. The optimization led to a success rate of 98 percent. Likewise, the average CV score was 98.

DISCUSSION AND CONCLUSION

For time-based features CNN has the lowest prediction rate and for pretrained representations the highest rate is the same as MLPC. This shows that CNN and time and frequency-based features have less ability to classify speech emotions with conventional audio features. In addition to these, to avoid potential overfitting problems, networks regularised with applying

dropout in each layer. Neurons designed for information processing and their learning process, interfere with parameters. Audio signals and time series gained from emotional utterances are not linear and they have complicated forms. Results of experiments in this study demonstrate that SVM has less ability to classify emotions among the other neural network techniques. It appears that certain patterns for emotions occur in the neural network's input space more obviously than SVM's plane and neural networks outperform the classical classifier. Despite the close results between MLPC and CNN, MLPC performs remarkably faster than CNN. Computational efficiency and consuming energy are crucial points of a speech emotion recognition system. From this point of view, MLP and CNN models are more robust than SVM models whereas MLP Classifier is the most advantageous technique for speech emotion classification.

Speech conveys an explicit message which is the linguistic part and implicit message which is the paralinguistic non-verbal part. Linguistic features not used in this work and the non-linguistic part seems to provide satisfactory results. Audio propagates by wave motions and at the course of wave motions energy is emitted. Furthermore, it changes over time and this change causes difficulties. Owing to these and many other anatomic reasons there are a myriad of derivable features. It is still unknown which audio features are more relevant with emotions. In this study for examining which features represent the emotion better, time and frequency dimensions of audio are taken into consideration. Independent of these dimensions, pre-trained representations are practised instead of choosing features manually. Results point out that the input obtained from the pre-trained representations provides superior predictions. Pre-trained model's main advantage over hand-crafted features is that it takes into consideration all of the audio features and attends prominent features. The pre-trained model is an unsupervised model and in this model the aim is finding the regularities in the input, irrelevant or little related emotional attributes are identified and removed. Both time and frequency features are suitable for emotion recognition. However, in the spectrum shape-based feature there are more discriminative features than time-based features. It seems that the time domain features are supplemental features for a SER system.

Schneider et al. (2019) implemented wav2vec which is an unsupervised pre-training model for speech recognition and achieve superior results than the next best-known character-based speech recognition model. Similarly, in this study better results have been obtained with the wav2vec large model. From these results it can be claimed that pre-training models make more accurate predictions for speech processing. Another study on raw spectrograms is made by Satt et al. They calculate spectrograms from audios then apply deep learning to spectrograms directly without extracting features. In the first stage, each sentence in the database was split for 3 seconds. These new sentences are used for labelling throughout the system. Whole sentences are used in the testing phase. They tried limiting the prediction latency even though it was losing accuracy. For each sentence, a spectrogram is calculated and then normalised. Convolutional Networks and Recurrent (LSTM) Networks are used for classifying. As a pre-process, non-speech noise sounds are removed

from log spectrograms based on harmonic filtering. Clean spectrograms test data has the highest recognition rate of 68.8.

Due to the project's main motivation, massive volumes of data have been accumulated. During the course of thirteen months, a total of 634 job posts were published and around 25,000 applications were received. 3,574 applicants were interviewed by phone after 25,000 applications were screened to ensure they met the minimum requirements. Video interviews were done with 2203 individuals out of 3574 applicants for this project. A total of 64,639 minutes of interviews with 2203 persons were recorded. In accordance with Emotional Analysis, the Human Resources officer and the position manager progressed or halted the procedure. Following the selection process, a total of 455 applicants were hired. Eighteen of the 455 applicants willingly quit their positions in the first six months, and thirteen in the final six months.

The managers of 424 employees who were still employed conducted a six-month and subsequently a twelve-month performance review. Comparing the 6-month performance review to the prior year, a 43% rise in performance scores was noted. As a consequence of the 12-month review of performance, a 46.4% rise was noted.

Using this data, classification was conducted at the end of 12 months. The performance assessment cutoff for 424 applicants (between 0 and 100) was judged to be 91. 149 individuals with grades over 91 and 275 individuals with grades below 91 were labeled as 1 and 0, respectively. The system was subsequently moved to Python for data processing and modeling. With an f1-score of 98%, the XGBoost algorithm obtained the highest score.

REFERENCES

Abdulnabi, A. H., Wang, G., Lu, J. and Jia, K., 2015. Multi-task CNN model for attribute prediction. IEEE Transactions on Multimedia, 17(11), pp. 1949–1959.

Arooj, A., Farooq, M. S., Akram, A., Iqbal, R., Sharma, A. and Dhiman, G., 2022. Big data processing and analysis in internet of vehicles: architecture, taxonomy, and open research challenges. Archives of Computational Methods in Engineering, 29(2), pp. 793–829.

Barney, J. B. and Wright, P. M., 1998. On becoming a strategic partner: The role of human resources in gaining competitive advantage. Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, 37(1), pp. 31–46.

Behroozi, M., Shirolkar, S., Barik, T. and Parnin, C., 2020, June. Debugging hiring: What went right and what went wrong in the technical interview process. In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society (pp. 71–80).

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794).

Dargan, S. and Kumar, M., 2020. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. Expert Systems with Applications, 143, p. 113114.

Du, H., Shi, H., Zeng, D., Zhang, X. P. and Mei, T., 2022. The elements of end-to-end deep face recognition: A survey of recent advances. ACM Computing Surveys (CSUR), 54(10s), pp. 1–42.

Hamza, P. A., Othman, B. J., Gardi, B., Sorguli, S., Aziz, H. M., Ahmed, S. A., Sabir, B. Y., Ismael, N. B., Ali, B. J. and Anwr, G., 2021. Recruitment and selection: The relationship between recruitment and selection with organizational performance. International Journal of Engineering, Business and Management, 5(3).

Kathuria, R. and Partovi, F. Y., 1999. Work force management practices for manufacturing flexibility. Journal of Operations Management, 18(1), pp. 21–39.

Khan, S., Akram, A. and Usman, N., 2020. Real time automatic attendance system for face recognition using face API and OpenCV. Wireless Personal Communications, 113(1), pp. 469–480.

Mobbs, D., Trimmer, P. C., Blumstein, D. T. and Dayan, P., 2018. Foraging for foundations in decision neuroscience: insights from ethology. Nature Reviews Neuroscience, 19(7), pp. 419–427.

Schneider, S., Baevski, A., Collobert, R., & Auli, M., wav2vec: Unsupervised pre-training for speech recognition, ArXiv, abs/1904.05862, 2019.

van den Broek, E., Sergeeva, A. and Huysman, M., 2021. When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. MIS Quarterly, 45(3).

Wang, M. and Deng, W., 2021. Deep face recognition: A survey. Neurocomputing, 429, pp. 215–244.

Winston, J. J. and Hemanth, D. J., 2019. A comprehensive review on iris image-based biometric system. Soft Computing, 23(19), pp. 9361–9384.

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A. M. and Wang, X., 2021. Computer vision techniques in construction: a critical review. Archives of Computational Methods in Engineering, 28(5), pp. 3383–3397.