Research Article

AI-Based Call Center Management

Nevra Kazanci[1*], Erçin Tevfik Öztuncel[2*], Metin Akuş[3*]

[,1] Affiliatio Sakarya University, Department of Industrial Engineering, Faculty of Engineering 1, Orcid ID: https://orcid.org/0000-0002-9525-1755, E-mail: nakbilek@sakarya.edu.tr

[2] CMC İletişim ve Çağrı Merkezi Hizmetleri Anonim Şirketi 2, Orcid ID: https://orcid.org/0009-0006-3233-6416, E-mail: Ercin.Oztuncel@cmcturkey.com

[3] CMC İletişim ve Çağrı Merkezi Hizmetleri Anonim Şirketi 3, Orcid ID: https://orcid.org/0009-0005-4510-2531, E-mail: Metin.Akus@cmcturkey.com

[*] Correspondence: nakbilek@sakarya.edu.tr; Tel.: (optional; include country code)

Reference: Kazancı, N., Öztuncel, E. T., & Akuş, M. (2024). AI-based call center management. The European Journal of Research and Development, 4(4), 338-351.

Abstract

Call centers today operate within complex ecosystems where surveillance technology, digitalization, and process automation are pivotal. These advancements enable multi-channel communication, personalized service, and proactive customer support. Unlike traditional models centered solely on phone interactions, modern call centers leverage digital tools to enhance operational efficiency. A significant innovation lies in the application of image processing techniques, including face recognition algorithms. These technologies automate tasks, minimizing human intervention and optimizing workflow. In this context, a proposed artificial intelligence-driven call center management system aims to replicate office environments remotely. It focuses on ensuring high service quality and security through real-time monitoring of representatives. Key features include facial recognition accuracy rates of 99% for detection and 96.88% for recognition. This system distinguishes live faces from photographs using cascade location detection, a novel approach that enhances fraud prevention compared to current methods. Integrating such advanced technologies into call centers marks a transformative step towards efficient, secure, and personalized customer service experiences in the digital age. Only the video call recordings are utilized for all analyses without additional equipment or data sources. Therefore, this easily implementable management system is introduced at a minimal cost.

Keywords: *Call center management; image processing; artificial intelligence; remote working; digitalization*

1. Introduction

Call centers have become essential in customer relationship management within the last 30 years [1]. Also, they contribute significantly to employment in many countries [2]. The personnel costs constitute 60-80 percent of all operating costs in the call center [3]. The call center has two most important success criteria: Accessibility and Quality. Accessibility is a factor that affects all other parameters that may come with it, and it is the beginning of everything. The second most important point that your customers expect from you, the other most crucial factor distinguishing you from other companies, is how much "quality" service you provide.

Operating remotely, assessing the performance of customer representatives during calls, evaluating call efficiency, and ensuring representative presence are challenging tasks using manual methods. Typically, such evaluations are conducted sporadically on a small sample of personnel. Companies, especially those in sectors like legal services, require stringent oversight of employee work hours. Given the scale of the call center industry with its thousands of employees, there is an urgent and critical demand for an automated system that is both highly accurate and professional to manage this oversight efficiently. Existing products partially fulfill these needs but come with prohibitively high costs. Therefore, an automated monitoring system is not just a luxury but a necessity to track employee activities and gauge productivity effectively.

Working from home has been increasing recently, especially with the pandemic effect. In a press release from 2022, Tele Performance announced that more than 70% of its workforce was engaged in remote work [4]. Sitel, a prominent operator in the contact center industry, has prominently mentioned on its website that 65% of its extensive staff of 160,000 individuals are presently engaged in remote work as of 2023 [5]. Due to the widespread adoption of telecommuting accelerated by the COVID-19 pandemic, particularly in service industries, ensuring the effectiveness and safety of organizational operations has become paramount. For instance, banking regulations mandate that customer representatives in call centers must be official employees working under appropriate conditions and meeting performance standards.

In essence, remote work necessitates using communication technologies and digital platforms within business processes. Consequently, a solution has been proposed where webcam footage of home-based customer representatives undergoes analysis to detect any potential irregularities. This involves employing a model that utilizes image processing, face recognition, and verification of reality and vitality. The study scrutinizes factors that could impact customer satisfaction, such as whether a representative is present in front of the camera during remote sessions, the individual's identity, orientation displayed during customer interactions.

This paper proposes an artificial intelligence-based call center management system that monitors call center representatives and creates alerts for the managers in case of any unexpected situation. In particular, the primary contributions of this study include the following:

The proposed system aims to replicate the challenging oversight processes call center companies face in managing remote representatives. Our research initiative is specifically designed within

the framework of legal obligations set forth by the BRSA, particularly in banking transactions. According to BRSA regulations, one out of every ten interactions with each customer representative must be video recorded daily, with corresponding security checks enforced. Remarkably, there is currently no existing AI-supported framework in literature that fulfills this specific legal requirement.

Additionally, the proposed system serves specialized service requests from companies, contributing to industry practices. For instance, VIP customers are directed to connect directly with a customer representative immediately after the IVR (Interactive Voice Response) system. Options such as reporting lost/stolen cards and handling suspicious transactions are prioritized in the main menu, ensuring customers are swiftly connected to a representative. Furthermore, bank call centers must prioritize connecting users requesting card cancellations directly to a human agent, bypassing robot agents or chatbots. Elderly customers aged 60 and above, identified through their registered number or ID, are also provided the option to connect directly to a call center representative, regardless of the registered calling number.

Face identification in video frames is employed for verification purposes, achieving an accuracy rate of 99% in identification and 96.88% in recognition. The system also automatically monitors the duration customer representatives spend with customers and their break times. Additionally, it verifies whether the detected face belongs to a live representative or is a photograph, using location change detection of facial features. This approach effectively prevents potential deceptive behaviors by customer representatives.

Exclusively utilizing video call recordings for all analyses without additional equipment or data beyond what is found in existing literature, the system's interface oversees various key aspects: detection of multiple persons and detection of still images, as depicted in Figure 9. This approach effectively automates call center management, offering a cost-effective solution. The system is designed to be user-friendly and widely accessible, aiming to stand out as an easily deployable solution with broad appeal.

The research has seven sections: first is an introduction, second is a literature review, third is call center management, fourth is machine-learning-based management, fifth is experimental results, sixth is the discussion about future work and limitations, and seventh is the conclusion.

2. Related Literature

The proposed framework begins by detecting and verifying the presence of a live face, followed by facial recognition to authenticate the representative—a crucial initial security measure. Subsequently, the system scrutinizes the work environment, automatically identifying new persons beyond authorized employees from recorded images. This aspect of the system aims to maintain productivity to meet service quality standards, triggering alerts when identification problems are detected in agents.

Moreover, representatives exhibiting positive moods are strategically assigned to handle calls involving challenging or VIP customers and repeated issues. Additionally, the analysis of agent head direction supports compliance with business quality standards. However, the primary focus remains on security measures, as any behavioral anomalies detected by the system prompt immediate warnings to prevent potential fraud. The proposed system's infrastructure is visually represented in Figure 9, while its comprehensive monitoring capabilities.

2.1. Face Recognition

Computer vision and deep learning are currently the focus of extensive research efforts, particularly in the realm of facial detection and identification. These technologies find application in various settings such as stores, banks, schools, public venues, and private properties for managing access control and ensuring security.

Facial recognition is a biometric method that involves accurately analyzing and comparing facial patterns based on distinctive facial features to verify an individual's identity [6]. The core challenge in automatic facial recognition lies not only in recognizing faces but also in reliably detecting them across diverse environments [7]. Precise facial detection in any given scene is crucial as it simplifies the subsequent recognition process, highlighting the importance of initial face detection as a prerequisite for effective face recognition [8], [9], [10].

Paul Viola and Michael Jones made a significant breakthrough in computer face detection in 2001, introducing the Viola-Jones face recognition algorithm. This algorithm revolutionized the field by enabling rapid framing of faces within images following years of less successful attempts [11]. Their approach, known as the detector cascade, has garnered widespread academic interest and consists of progressively more complex face classifiers. The Viola-Jones detector cascade has also found practical applications in various commercial products, including digital cameras and mobile phones [7].

While cascade detectors excel in detecting noticeable faces, they encounter challenges with partially visible faces or faces captured from different angles, prompting researchers to seek solutions. Using the Viola-Jones method, Wu et al. [12] developed specialized detector cascades for different face views. Mathias et al. [13] introduced integral channel features and a soft cascade approach to enhance multi-view face detection performance, incorporating face orientation annotations. Viola and Jones [14] proposed a tree classifier for face pose estimation and validated detections through cascades tailored to specific face poses to address the computational demands of multi-view face detection.

In addition to cascade-based methods, deformable part models [15] offer an alternative approach. They define faces as assemblies of facial parts trained through supervised and unsupervised methods. This technique, employing Latent SVM, proves robust against detecting partially visible faces, as detection relies on recognizing specific parts rather than the entire face.

Hangaragi et al. [16] pioneered the detection and recognition of faces using Face Mesh technology. Initially, a machine learning framework identifies faces and extracts multiple landmarks, followed by 3D face reconstruction via Face Mesh. Deep learning techniques then match the reconstructed face with a database for identification, utilizing 468 extracted facial landmarks. Despite its capability to handle diverse conditions such as varying illumination, backgrounds, gender, age, and race, the proposed model is noted for its high computational complexity.

While some researchers, including those referenced in [17], continue to leverage machine learning for face recognition, recent advancements have been significantly influenced by deep learning methodologies. Convolutional Neural Network (CNN)--based object detectors can now effectively distinguish faces from other objects within images.

In this context, the framework presented in this article demonstrates its proficiency in face detection and recognition using the ResNet model. Trained with the ResNet50 architecture in Python, this neural network leverages the "residual layer" structure introduced by He et al. [18] in 2016, which has proven instrumental in advancing deep learning applications, particularly in the field of image processing.

The proposed system utilized Haar Cascades from the OpenCV library in Python for the purpose of face detection. Haar feature-based cascade classifiers, originally introduced
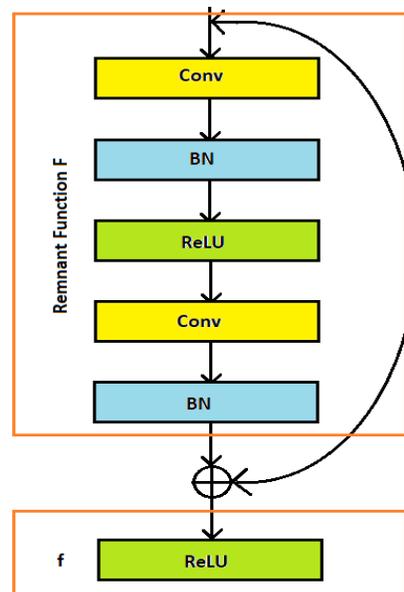

Figure 1: Residual unit [23].

by Viola and Jones [19], were employed to establish the fundamentals of face detection. This approach, deeply rooted in machine learning, involves training a cascading function with a vast number of positive and negative image samples. Haar features, such as edges, lines, and four-

rectangle patterns, akin to convolutional kernels, are extracted from these images. The integral image, an intermediate representation of an image, plays a crucial role in efficiently computing these Haar features across different regions of an image.

3. Materials and Methods

3.1. Face Recognition

Residual neural networks (ResNet) were employed for face identification, leveraging recent advancements in deep learning techniques. ResNet has demonstrated remarkable success in extracting powerful features from images, leading to efficient and accurate classification. The core element of ResNet is the residual unit, which consists of multiple convolutional layers (Conv), batch normalization (BN) stages, and rectified linear unit (ReLU) layers. Figure 1 illustrates a schematic representation of a residual unit.

Let be the input for the first residual unit. The algorithm's result is computed as given in equation 1:

$$x_{l+1} = f(x_l + F(x_l, W_l)) \qquad (1)$$

where F(x) represents the residual mapping learned by the convolutional layers, BN stages, and ReLU layers within the residual unit. This additive approach facilitates the training of deeper networks by mitigating the vanishing gradient problem, thereby enhancing the network's ability to learn intricate patterns and features essential for face.

$$x_{l+1} = f(W, x_l + F(x_l, W_l)) \qquad (2)$$

The residual function incorporates a weight parameter and is characterized by its adaptable structure. ReLU (Rectified Linear Unit) serves as a non-linear activation function within this context. It's crucial that the dimensions of the input xxx and the output are identical. In scenarios where input and output channels differ, a linear mapping is employed to align dimensions, as detailed in equation 2 [20]. This ensures compatibility and smooth transition between layers, facilitating effective training and optimization of ResNet architectures for diverse tasks, including face identification and classification.

Additional convolutional layers can be incorporated to construct a deep residual network by assembling numerous residual units. Input data passes through sequences of Convolutional (Conv), Rectified Linear Unit (ReLU), and Batch Normalization (BN) layers. Outputs are generated after processing through residual units, BN layers, and several fully connected layers. The architectural framework employed for face recognition is depicted in Figure 1, illustrating the configuration of residual units essential for this task. The model was trained using images from nearly 1000 individuals, with the entire implementation written in Python, encompassing CNN and other necessary components. Figure 2 provides a comprehensive overview of the network architecture designed for face detection and recognition. Identity blocks consist of three convolutional layers, while convolutional blocks feature a similar structure, which is crucially employed in identity verification modules.
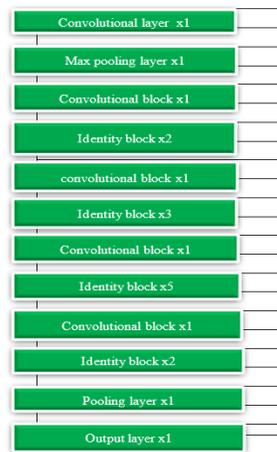

Figure 2: The network architecture utilized for facial detection(a) and recognition(b)

4. Results

We outline the evaluation criteria employed to assess the effectiveness of our AI-driven call center management system across its key functionalities: face detection and recognition. Each subsection within this segment elaborates on specific metrics tailored to our classification problem, influenced by related studies such as [14], [15], [21].

In calculating these metrics, understanding True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) is crucial. For face detection, these terms are defined as follows:

- TP: Correctly detecting a face when present in the scene.
- TN: Correctly not detecting a face when absent from the scene.
- FP: Incorrectly detecting a face when none exist in the scene.
- FN: Failing to detect a face when they are actually present.

These metrics collectively provide comprehensive insights into the system's performance, guiding further enhancements and optimizations in call center

3.1. Face detection and recognition task

For this task, first, we were required to detect faces in video frames and then recognize the identities for the automated monitoring of customer representatives. Haar cascades were used to detect face regions, and residual neural networks (ResNet) were used to recognize the identities of detected faces. Table. **1** represents the experimented parameter values for the ResNet-based face recognition model. Bold values refer to the best parameters that produce the highest accuracies.

Table 1: Experimented hyper parameter values for face recognition

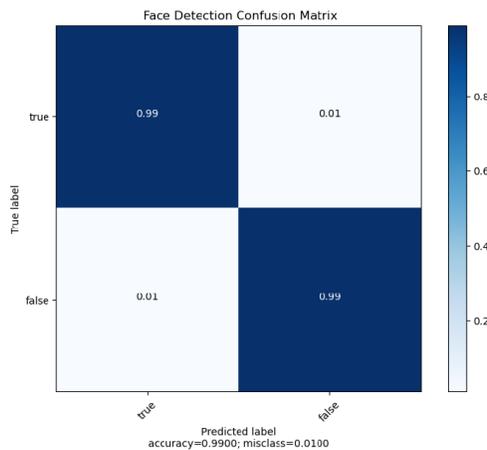| Hyper parameter	Experimented values
Activation function	**ReLU**, sigmoid, softmax, tanh
Loss function	**Binarycross-entropy**, categorical_crossentropy
Optimization algorithm	**Adam**, SGD, Nadam, Rmsprop
Learning rate	0.0001, **0.001**, 0.01
Batch size	**32**, 16
Dropout rate	**0.2**, 0.5

Figure 3: Confusion matrix of Face detection

As the confusion matrix in Figure 3 shows, Haar feature-based cascade classifiers with OpenCV module in Python detect almost all faces. 99% is a successful detection accuracy for the proposed agent monitoring system. Details of the Haar feature-based classifier are given in the section Face Detection and Recognition.

More deterministic and discriminative features are required for the identification of faces. Hence, more features were extracted through Resnet-based models that are mostly used for face recognition tasks. The proposed ResNet-based face recognition model is given in Figure 5. ResNetV2 model [22] was customized by adjusting the number of residual and fully connected layers and fine-tuning them. Our best model has four residual layers; its properties are given in Figure 5. The vanishing gradient problem encountered in deep neural networks disappears in ResNet by activating and adding to the current layer before activation. That's why the contribution of some neurons in deeper layers is guaranteed, and the problems are solved successfully.
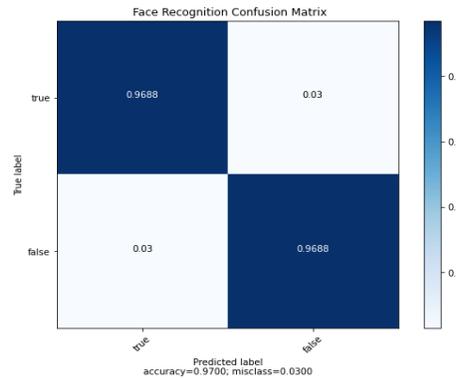

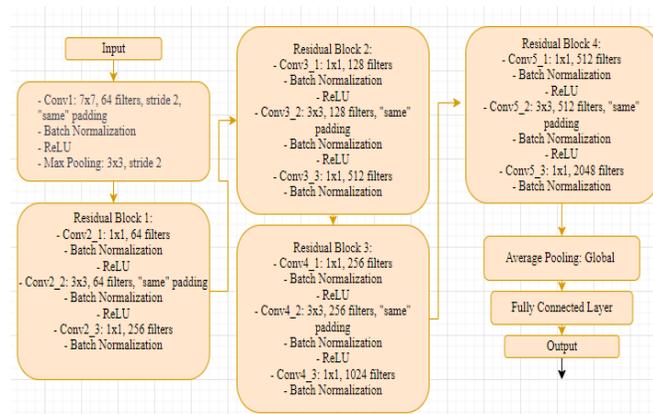*Figure 4: Confusion matrix of Face recognition task*

*Figure 5: Proposed ResNet-based model*

For the model's training, we used frames in the video streams of each customer representative. Six frames were used for one agent and as an input image for that agent. 978 images belonging to 163 agents were used in total for the experiments on the face recognition task. Figure 4 shows the face recognition task's confusion matrix, implying a satisfactory success rate with about 97% accuracy—3% of the images needing correct recognition.

Table 2 shows the performance metrics for face recognition on the CMC Turkey company and the Olivetti datasets. Additional experiments on the public Olivetti dataset [23] were also conducted to prove the success of the model we created for face recognition. There are 400 face images from 40 distinct samples in the dataset. Face images were taken between April 1992 and April 1994 by AT&T Laboratories in Cambridge. They were taken at different times, with varying lighting, facial expressions, and facial detail conditions. The size of each image is 64x64, and image pixel values were scaled to [0, 1] intervals.

Table 2: The performance metrics values of the face recognition application

Performance metrics for the face recognition task on datasets (Average-%)				
Dataset	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CMC Turkey	96.88	94.50	95.00	95.50
Olivetti	86.00	80.00	85.00	86.00

Furthermore, the sitting position of the employees, which complicates the decision about the camera field of view and the data quality of input images, may impair the effect of extracted features and cause losses. Training and validation accuracy and loss graphs of the face recognition model for 60 epochs are given in Figure 6. The graph indicates that the model can successfully recognize samples with no overfitting/under fitting. Because at a certain point (around the 40th epoch), training and validation curves become consistent, which means the model learned well.
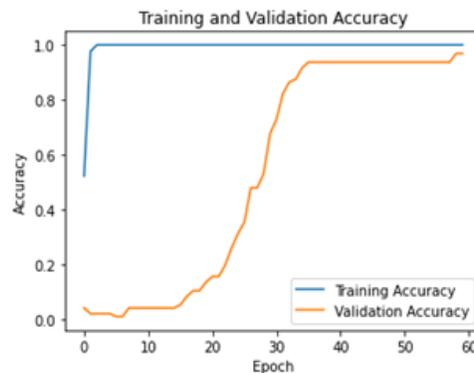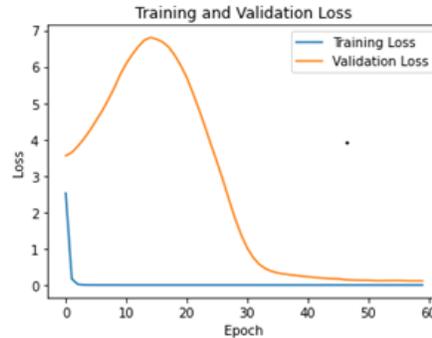
Figure 6. Training and validation accuracy and loss graphs of ResNet-based face recognition model on CMC Turkey dataset
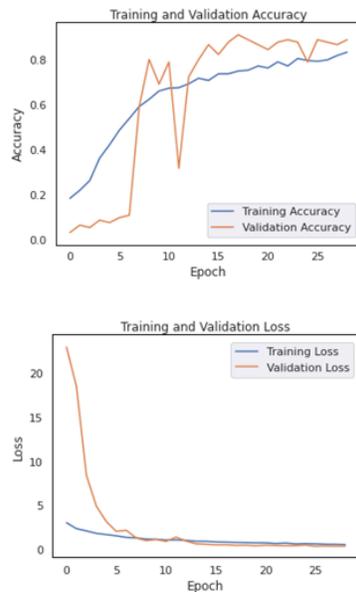

Figure 7. Training and validation accuracy and loss graphs of ResNet-based face recognition model on the Olivetti dataset

Figure 7 shows training and validation accuracy-loss graphs for the Olivetti dataset. The metrics and validation graphs indicate acceptable rates for the created ResNet-based face recognition model on a public Olivetti dataset. For better performance, agents must comply with their sitting patterns and see themselves on the screen during work to enable better input images for the analysis. An ergonomic design for the physical environment of the agents should be developed, and employees should be able to follow their pictures on the screen. Increasing the camera quality will also improve the decision process by reducing the need for preprocessing. In addition, you can see the success values obtained from all transactions in Table 3 below.

Table 3. Face detection and recognition results.

Table 2: The performance metrics values of the face recognition application

Task	Accuracy	Precision	Recall	F1-Score
Face Detection	0.99	0.99	0.99	0.99
Face Recognition	96.88	94.50	95.00	95.50

Also, the DenseNet face recognition model is applied to the CMC Turkey dataset, as shown in Figure 8. DenseNet, short for "Densely Connected Convolutional Networks," is an architecture introduced in a 2017 paper by Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger [24]. It is a convolutional neural network (CNN) known for its unique connectivity pattern; it connects each layer to every other layer in a feed-forward model.
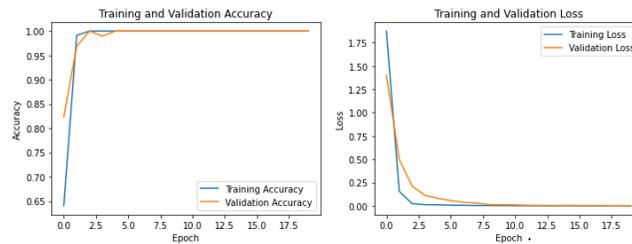

Figure 8. Training and validation accuracy and loss graphs of the DenseNet face recognition model on the CMC Turkey dataset.

5. Discussion and Conclusion

Remote work leverages digital tools, communication platforms, and automation technologies to streamline workflows, facilitate remote collaboration, ensure security, and boost productivity. These tools foster close cooperation and interaction between humans and machines. The goal is establishing a symbiotic relationship where machines handle repetitive tasks and data processing. Decision-making authority is shared between humans and machines, supported by real-time data from intelligent systems. To this end, we propose a novel framework for fully automated call center management that is cost-effective, easily deployable across any organization and addresses common challenges in call center operations. Tasks such as monitoring break times of customer representatives during interactions for effective communication and customer satisfaction, and identifying representatives are efficiently managed through an enhanced artificial intelligence-driven management system. The inefficiencies of manual inspection of customer representatives are effectively eliminated with our proposed framework. In the future, we anticipate widespread adoption of this mechanism by

nearly all call center enterprises to enhance customer relationships and optimize operational costs.

Looking ahead, a potential evolution of our remote management system could involve real-time monitoring. Unlike our current offline approach, future iterations could instantly alert representatives during interactions, making the transition to real-time applications more purposeful for companies. This exciting prospect opens up new possibilities for call center management, promising even greater efficiency and customer satisfaction.

References

[1]   Örmeci, E. L., Salman, F. S., & Yücel, E. (2014). Staff rostering in call centers providing employee transportation. *Omega*, *43*, 41-53.

[2]   Phung-Duc, T., & Kawanishi, K. I. (2014). Performance analysis of call centers with abandonment, retrial and after-call work. *Performance Evaluation*, *80*, 43-62.

[3]   Xia, C. H., & Dube, P. (2007). Dynamic pricing in e-services under demand uncertainty. *Production and Operations Management*, *16*(6), 701-712.

[4]   Christl, W. (2023). Surveillance and algorithmic control in the call centre. *A case study on contact and service center software, automated management and outsourced work. Cracked Labs*.

[5]   https://foundever.com/

[6]   Yaswanthram, P., & Sabarish, B. A. (2022, January). Face recognition using machine learning models-comparative analysis and impact of dimensionality reduction. In *2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC)* (pp. 1-4). IEEE.

[7]   Farfade, S. S., Saberian, M. J., & Li, L. J. (2015, June). Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* (pp. 643-650).

[8]   Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4690-4699).

[9]   Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 212-220).

[10]   Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

[11]   Viola, P., & Jones, M. (2001, December). 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. CVPR 2001* (Vol. 1, pp. I-I). Ieee.

[12]   Wu, B., Ai, H., Huang, C., & Lao, S. (2004, May). Fast rotation invariant multi-view face detection based on real adaboost. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* (pp. 79-84). IEEE.

[13] Mathias, M., Benenson, R., Pedersoli, M., & Van Gool, L. (2014). Face detection without bells and whistles. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13* (pp. 720-735). Springer International Publishing.

[14] Jones, M., & Viola, P. (2003). Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, *3*(14), 2.

[15] Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008, June). A discriminatively trained, multiscale, deformable part model. In *2008 IEEE conference on computer vision and pattern recognition* (pp. 1-8). Ieee.

[16] Hangaragi, S., Singh, T., & Neelima, N. (2023). Face detection and Recognition using Face Mesh and deep neural network. *Procedia Computer Science*, *218*, 741-749.

[17] Hu, K., Allon, G., & Bassamboo, A. (2022). Understanding customer retrials in call centers: Preferences for service quality and service speed. *Manufacturing & service operations management*, *24*(2), 1002-1020.

[18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[19] Viola, P., & Jones, M. (2001, December). 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. CVPR 2001* (Vol. 1, pp. I-I). Ieee.

[20] Xu, W., Fu, Y. L., & Zhu, D. (2023). ResNet and its application to medical image processing: Research progress and challenges. *Computer Methods and Programs in Biomedicine*, *240*, 107660.

[21] Putra, Y. C., & Wijayanto, A. W. (2023). Automatic detection and counting of oil palm trees using remote sensing and object-based deep learning. *Remote Sensing Applications: Society and Environment*, *29*, 100914.

[22] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[23] AT&T Laboratories Cambridge, "The Olivetti faces dataset," *scikit learn*. [Online]. Available: https://scikit-learn.org/0.19/datasets/olivetti_faces.html. [Accessed: Feb. 05.2024]

[24] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).