

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ARTIFICIAL INTELLIGENCE AND COGNITIVE INFORMATION TECHNOLOGIES

doi: 10.17586/2226-1494-2023-23-5-1009-1020

System for customers' routing based on their emotional state and age in public services systems

Guedes M. Soma¹✉, Georgy D. Kopanitsa²

^{1,2} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ guedes.soma@mail.ru✉, <https://orcid.org/0000-0001-9004-3554>

² kopanitsa@itmo.ru, <https://orcid.org/0000-0002-6231-8036>

Abstract

In this paper, we have developed a system for assigning customers to the routes based on their emotional state and age in Public Service Systems (PSSs). The Squeeze-and-Excitation (SE) method was used to develop the models, it improves the efficiency of the Deep Convolutional Neural Networks (DCNN) architecture by increasing the information flow between layers and enhancing important features. The method is based on compressing and exciting information at each convolution stage, which allows obtaining a vector of channel importance scores and using it to reweight the channels of the feature map. The study showed that this method allowed improving the quality of classification and reducing the model training time. The model of emotional target routing was developed based on the Newton interpolation polynomial to route customers based on their emotional state and age. The interpolation function in this model calculates the waiting time for customers according to their emotional state. Three models of binary classification of emotions and ages were developed, namely, two models for recognizing the emotional state of the customer, and one model for recognizing their age. The first and third models utilize DCNN from scratch using the new SE approach based on the attention mechanism. The second model uses the Support Vector Machine (SVM) method. The evaluate method was used to test the model after training, which allows evaluating the quality of the model on new data that was not used during training. This is done to check how accurately the model can predict the values of the target variable on new data. The evaluate method utilizes quality evaluation metrics such as accuracy, recall, and F1-score to assess the performance of the model. According to the experimental data obtained, the first and the second developed models achieved the validation accuracy of 72 % and 66 %, respectively, on the FER-2013 and Adience datasets. Their sizes were 0.69 MB and 369 MB, respectively. At the same time, the age recognition model achieved the accuracy of 88 % with the size of 1.68 MB. The mathematical model of emotional target routing (TERSS) was developed to minimize conflicts in public service systems. The developed system can automatically route customers based on their emotional state (presence of anger) and age to the appropriate operator. Thus, customers over 60 years old or with the anger level of 60–80 % are directed to a senior operator who knows how to communicate with elderly or emotionally excited customers, while customers with the anger level of 80–100 % are directed to a psychologist. This research can be applied in PSSs to detect the features of customers' age and anger. Moreover, it can be applied in various areas where there is a contact with a large number of people, such as banks, supermarkets, airports access control systems, police stations, subways, and call centers.

Keywords

facial expression, emotion, age, classification, DCNN, PSS

For citation: Soma G.M., Kopanitsa G.D. System for customers' routing based on their emotional state and age in public services systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 5, pp. 1009–1020. doi: 10.17586/2226-1494-2023-23-5-1009-1020

УДК 004.85

Система маршрутизации клиентов на основе их эмоционального состояния и возраста в системах обслуживания населения

Гедеш Мануэл Сом¹✉, Георгий Дмитриевич Копаница²^{1,2} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация¹ guedes.soma@mail.ru✉, <https://orcid.org/0000-0001-9004-3554>² kopanitsa@itmo.ru, <https://orcid.org/0000-0002-6231-8036>

Аннотация

Введение. Разработана система маршрутизации клиентов на основе их эмоционального состояния и возраста в системах обслуживания населения. Для разработки моделей предложен метод Squeeze-and-Excitation (SE), который отличается от существующих методов тем, что позволяет улучшить эффективность глубокой сверточной нейронной сети архитектуры, повышая информационный поток между слоями и усиливая важные признаки. SE метод основан на сжатии и возбуждении информации на каждом этапе свертки, что позволяет получить вектор оценок важности по каналам и использовать его для перевзвешивания каналов карты признаков. В работе показано, что SE метод позволил улучшить качество классификации и сократить время обучения модели. **Метод.** Для проведения маршрутизации клиентов в зависимости от их эмоционального состояния и возраста разработана математическая модель эмоциональной целевой маршрутизации на основе интерполяционного многочлена Ньютона. Интерполяционная функция в предложенной модели выполняет расчет времени ожидания клиентов в соответствии с их эмоциональным состоянием. Разработаны три модели бинарной классификации эмоций и возрастов: две модели для распознавания эмоционального состояния клиента и одна — для распознавания возраста. В первой и третьей моделях представлена глубокая сверточная нейронная сеть с использованием нового подхода SE на основе механизма внимания. Во второй модели применен метод опорных векторов. **Основные результаты.** Для тестирования модели после ее обучения применен метод evaluate, позволяющий оценить качество модели на новых данных, которые не были использованы при ее обучении. Это осуществлено для проверки, насколько точно модель может предсказывать значения целевой переменной на новых данных. Метод evaluate использует метрики оценки качества модели, такие как точность (accuracy), полнота (recall), F1-мера (F1-score). Выполнено сравнение эффективности методов на наборах данных FER-2013 и Adience, а также результатах, полученных в настоящей работе, с результатами, описанными в научных работах, опубликованных в базе данных Scopus. Согласно полученным экспериментальным данным, первая и вторая разработанные модели достигли точности валидации 72 % и 66 % соответственно, а их размеры составили 0,69 МБ и 369 МБ. В то же время модель распознавания возраста достигла точности 88 % при размере 1,68 МБ. Математическая модель эмоциональной целевой маршрутизации разработана для минимизирования конфликтов в системах обслуживания населения. Разработанная система может автоматически направлять клиентов по маршрутам в зависимости от их эмоционального состояния (наличия злости) и возраста к соответствующему оператору. Таким образом, клиенты старше 60 лет или с уровнем злости 60–80 % направляются к старшему оператору, который имеет опыт общения, с пожилыми клиентами или клиентами в возбужденном эмоциональном состоянии, а клиенты с уровнем злости 80–100 % направляются к психологу. **Обсуждение.** Результаты исследования могут быть применены в системах обслуживания населения для распознавания возраста и признаков злости клиентов. Разработанная система может быть применена, в различных областях, где есть контакт с большим количеством людей (например, банки, супермаркеты, системы контроля доступа в аэропортах, полицейские участки, метро и колл-центры).

Ключевые слова

выражение лица, эмоция, возраст, классификация, DCNN, PSS

Ссылка для цитирования: Сом Г.М., Копаница Г.Д. Система маршрутизации клиентов на основе их эмоционального состояния и возраста в системах обслуживания населения // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 5. С. 1009–1020 (на англ. яз.). doi: 10.17586/2226-1494-2023-23-5-1009-1020

Introduction

The stability and prosperity of a company largely depend on its ability to respond in a timely manner to changes in the market situation. A modern company should be manageable, and the degree of manageability in any organization is largely determined by how well it works with information, namely, with the collection, processing, and analysis for decision-making. Such a process is performed daily in Public Service Systems (PSSs) which is characterized by increased stress and extreme responsibility for decision-making as well as the severity of the consequences in case of conflicts due to the customer's emotional state. PSSs is a service provided

by the government to people living within its jurisdiction, either directly (through the public sector) or by funding the private provision of services [1]. In professions involving constant communication with people, communicative competence acts as an integral part of professional competence. Researchers in [2, 3] note that the activity in PSSs is characterized by a high level of stress and extremes, greater responsibility for decision-making, and the severity of consequences in case of mistakes. In [4] the authors developed a deep learning model for predicting conflict attacks in Nigeria. Despite the fact that the data set was incomplete, noisy, and unbalanced by classes when predicting the type of conflict attacks, after cleaning and normalizing the multiclass dataset, a qualitative dataset

was obtained. The deep neural network model showed good prediction with 98 % accuracy on the training dataset and 89 % on the test dataset with losses of 2 % and 11 %, respectively.

In [5], the author highlights the causes of conflicts in organizations, which according to the research, include a lack of coordination and communication as well as a lack of proper definition of responsibilities. The author emphasizes that during the data collection in PSSs in United Arab Emirates organizations, many respondents agreed that it is necessary to define an appropriate set of rules and regulations so that people know what they should do and how they should behave. According to the author, managers and leaders should deal with and solve the problems their organization faces. Each of them has their own inclinations and ways of resolving conflicts. It all depends on how and when problems are recognized and what is done to solve them.

After analyzing several works, we came to the conclusion that none of them solves the problem of automatic distribution of conflicted customers along the route to a proper operator in PSSs depending on their emotional state and age. These facts allow us to consider the topic of the paper as a relevant research area for solving the above-mentioned tasks.

Goal and Objectives

The aim of the work is to develop the Target Emotional Routing System Software (TERSS) to improve the quality of customer service in the PSSs. To achieve this goal, the following objectives need to be addressed:

- Develop a Deep Convolutional Neural Networks (DCNN) model for recognizing customer's emotions and age using the Squeeze-and-Excitation (SE) method;
- Develop a mathematical model of TERSS based on the Newton interpolation polynomial;
- Compare the effectiveness of methods on FER-2013 and Adience datasets as well as the results obtained in this study with the results described in scientific papers previously published in the Scopus database.

Subject of Research

The subject of this research is creating or adapting models for recognizing customer's emotional state and age through visual channels as well as a mathematical model for distributing customers along various routes to the appropriate operator based on the Newton interpolation polynomial.

Analytical Review

In recent years, we may have seen a rapid development in parallel computing technologies, particularly with the development of graphics processors which are now being used for more than just computer graphics. This allowed training even the most complex neural networks in terms of their architecture and opened a whole horizon of previously unsolvable tasks [6]. These days, intelligent systems are not only focused on recognizing images from the input image,

but also learn to extract metadata from the recognized objects, such as emotions, mood, gender, and age of a person [7–9]. In particular, emotions play an important role in human life and interpersonal dialogue [10]. As it is well-known, emotion is a psychological state of the human mind. Research in different fields has different opinions about the process of emotional development [11]. Some philosophers believe that emotions are the result of changes (positive or negative) in personal situations or the environment. However, some biologists believe that the nervous and hormonal systems are primarily responsible for the development of emotions [12]. Although there is no consensus on what causes emotion; it is a fact that its arousal is usually accompanied by some manifestation in our appearance, such as changes in facial expression, voice, gesture, posture, and other physiological conditions. Nevertheless, it is the facial expression that most accurately and clearly reflects person's emotions and age. Facial expressions provide relevant information about an individual's emotional state and play a key role in human interaction and as a form of non-verbal communication. They can complement verbal communication or even convey a complete message on their own [13]. Research demonstrates that the verbal part or spoken words of a message contribute only 7 % to the impact of the message as a whole; the vocal part (intonation) contributes 38 %, while the speaker's facial expression contributes 55 % to the impact of the spoken message [14]. Facial expressions consist of one or more movements or positions of face muscles. In a specific visual structure on the face, these muscles are related to a specific emotion. Facial expressions are related to the non-verbal form of communication [15]. A person recognizes an emotion by identifying the structure of the facial expression. To learn how to recognize and classify emotions, machines need to mimic the structure of the neurons in the human brain in a similar way. In this work, we focus on recognizing the emotional state of anger and the age of the customers; moreover, anger is the most important emotional state in our study. By its semantic structure, the word 'anger' is also understood as 'resentment, a feeling of strong indignation' [16]. This condition can lead to serious conflicts. We chose the age of 60+ years because, according to [17], most customers over the age of 60 are disabled in the second group for various reasons. Basically, they suffer from loss of vision, hearing, and memory, which makes communication difficult and creates barriers to communication. In addition, elderly customers can experience long waiting times in queues, and conflicts may arise, especially if the operator is a trainee. Therefore, in order to prevent possible conflicts during maintenance, we have proposed the following solutions:

- Recognition of the emotional state and age of the customer by facial expression;
- Route distribution of a conflicted or elderly customer over 60 years to a senior operator or a psychologist using the mathematical model of emotional target routing developed by us based on the Newton interpolation polynomial;
- The use of a hardware system (camera, computer, and coupon issuing machine) to make an automatic decision based on the emotional state and age of the customer.

Convolutional Neural Networks (CNN) is a technology actively used in Artificial Intelligence (AI) to classify images by creating a classification model of human emotions based on a dataset of facial expression images. Currently, there is a significant number of models that can automatically recognize human emotions and age from facial expressions [18–21]. However, these existing models do not make an automatic decision about people's emotions or age, and the quality of these emotion and age recognition systems is also reduced due to the following number of problems [22, 23]:

- 1) Poorly constructed neural network architecture;
- 2) Small amount of data for training and testing;
- 3) Intraclass differences and interclass similarity;
- 4) Presence of fictitious emotions;
- 5) Low light levels;
- 6) Different head rotation angles;
- 7) Differences in facial proportions.

In accordance with the above-mentioned problems, we have created hypotheses to solve them, such as:

- 1) Applying the SE method to DCNN architecture;
- 2) Using a dataset with large amounts of data for training and testing;
- 3) Adjusting the images in the dataset to be balanced;
- 4) Adjusting the parameters of the classifier under development (weights and bias);
- 5) Adjusting the loss function to avoid overfitting.

The scientific novelty of this work lies in the use of machine learning and emotion recognition algorithms allowing the determination of the customer's emotional state and age, as well as in the development of a mathematical model that determines the customer direction to a proper operator. This improves the quality of customer service and prevents potential conflicts, which is an important factor in increasing customer satisfaction and improving the company's image. Additionally, it can improve organizational efficiency by reducing the time spent on the resolving conflicts which can lead to the increased customer loyalty and profitability.

Methods

The first and third models utilize the SE method which includes Attention mechanisms to dynamically adjust the contribution of each feature channel depending on the image context. The models are trained to determine which filter channels correspond to a specific context and to manage the contribution of these channels during the processing at the next layer. In binary classification, the SE method is employed to improve the performance of the model by increasing its ability to distinguish between two classes. This can be especially useful when classes have complex dependencies that are difficult to detect using conventional convolutional neural networks. The SE method improves the accuracy of the model without increasing its complexity. This method effectively leverages information from both classes, leading to higher accuracy. In the second model, we used the Support Vector Machine (SVM) method which represents a machine learning algorithm employed for binary classification. It is based on finding a hyperplane that separates the data into two

classes and searches the distance (margin) between them. The larger the margin, the more confidently SVM classifies the data. Despite SVM is a powerful tool for binary classification, it can still be difficult to use and requires parameter tuning to achieve optimal results.

Proposed TERSS algorithm

The working principle of our algorithm utilizes the hierarchical structure of the deep neural network and the SE method to improve the accuracy of facial expression recognition and decision-making in routing customers to a senior operator or a psychologist at the expense of emotional state — 'anger' over 60 % and age over 60. It is known that conflict management in public services is performed by clarifying work requirements and setting clear professional and organizational goals [24]. The developed system classifies the image (frame) by searching the characteristics (signs) of the basic level; in case the customer's emotion of anger is over 60 % or the age is over 60, the system automatically directs the customer to a senior operator or a psychologist to prevent possible conflicts in PSSs. The main parts of the TERSS system are:

- A camera — an electronic device that is used to capture images and/or videos of customers in real-time;
- The trained model — an algorithm with optimized parameters (weights) obtained after training the classifier;
- The image pre-processing which performed by the MediaPipe face detection method. MediaPipe Face Detection is an ultra-fast face detection solution with six landmarks and possibility for multiple face recognition that allows finding the area of the face in the image, then cropping and resizing that area and adjusting the brightness of the image (Fig. 1);
- Emotion and age recognition which aims to detect the emotional states and age of customers in PSSs to prevent possible conflicts;
- The distribution of conflicted customers among routes to the appropriate operators is performed using the mathematical model developed in our work;
- Process efficiency evaluation — at this stage, data is collected on the state of customers before and after receiving the service. This data can be used to improve the process and increase the level of customer service.

Mathematical model of TERSS

The task of constructing the interpolating function $F(x)$ has several solutions, because an infinite number of curves can be drawn through the points X_i , Y_i , each of which is a graph of a function; and all the interpolation conditions are satisfied for this function. For our task, we have created a function to calculate the customer's waiting time as a function of their emotional state. This function is necessary because the initial waiting time is three minutes (i.e., 180 seconds). We want the waiting time to vary depending on the current average value of the customer's anger level. The average anger level is the sum of this level in all the frames divided by the total number of frames for a given period of time (in our case, the program detects 20 frames per second and sets 6 seconds to determine the average anger level).

To create such a function, experimental interpolated nodes are defined. Interpolated nodes are the link between the correspondence of time and anger level in the graph

which is used to determine the formula of the function. The function formula is derived using Newton's interpolation polynomial rule written as:

$$P_n(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)(x - x_1) + \dots + A_n(x - x_0) \dots (x - x_{n-1}).$$

When A_n — divided differences are defined as:

$$A_0 = f(x_0), A_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, A_2 = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}, \dots, \\ A_n = \frac{f(x_1, \dots, x_n) - f(x_0, \dots, x_{n-1})}{x_n - x_0}.$$

It is possible to write the Newton polynomial in recursive form as:

$$P_k(x) = P_{k-1}(x) + A_k(x - x_0) \dots (x - x_{k-1}), \\ P_{k-1}(x) = P_{k-2}(x) + A_{k-1}(x - x_0) \dots (x - x_{k-2}).$$

Setup experiments

To evaluate the quality of the developed classifiers, we used the classification report function from the Sklearn library and the confusion matrix. There are three main metrics in classification report that are used to evaluate the quality of our models [25]:

- 1) Precision: percentage of correct positive predictions relative to the total number of positive predictions;
- 2) Recall: percentage of correct positive predictions relative to the total actual positive result;
- 3) F1-score: weighted harmonic means of accuracy and recall; the closer to 1, the better of the model.

$$F1 \text{ Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}).$$

Support is the relevant number of occurrences of the class in the specified dataset [25].

According to the main parameters (see Table 1: model size, execution time which is measured on the NVIDIA "Tesla P100-PCIE" GPU), losses are the value of the loss function used to evaluate the quality of the decisions made in the frame of model training. The model error is minimized with each successive training epoch. Training is the process of teaching a neural network by changing the weights of the network using the gradient descent method, minimizing the loss function at each iteration [26].

Practical implementation of the proposed TERSS algorithm

As mentioned earlier, image pre-processing is performed using the MediaPipe face detection method (Fig. 1). i_1, i_2, i_3 and o_1, o_2 refer to the input and output data in the neural network. i_1, i_2, i_3 represent the input image that will be processed by the neural network. The image is represented as a matrix of pixels. o_1, o_2 represent the results of the neural network which can vary depending on the learning objective. In our case, it contains information about recognized emotions and age based on the processing of the input image. Once the image is pre-processed, the emotion and age classifications are performed through the developed DCNN using the SE method.

Fig. 1 illustrates the pre-processing of images in three steps: face detection, feature extraction, and image classification. In step (Fig. 1, a), the MediaPipe face detection method is used to identify the area of the face in the image. In step (Fig. 1, b), the system extracts the features from the detected face, such as facial landmarks and a texture, to be used in the subsequent classification step. Finally, in step (Fig. 1, c), the developed DCNN using the SE method performs the image classification based on the extracted features.

Fig. 2 demonstrates the developed emotion and age recognition algorithm. The algorithm works in PSSs in real time by using a visual channel (webcam). In the beginning, the system analyzes the customer for a short amount of time (six seconds currently) to record their initial emotional state values, estimated age, storing their face and giving it an ID so that the algorithm can track them later. The system receives a frame (image) as an input, and then processes the image to detect the faces appearing on it. After that, it applies a face recognition to match the current frame faces with the faces stored previously to update their emotional state values based on the new received image and the previous emotional state values. If the values are in the higher ranges of anger [60–80], [80–100] the waiting time for the customers experiencing these emotions will be updated accordingly to the developed Polynomial Newton function which, in its turn, hastens the queue waiting time for the particular customers. The speed of timer expiration depends strongly on the intensity of the emotions that the customer is experiencing during his/her visit. Since the estimated age of the customers has an important role in speeding them up to receive the service, the customers

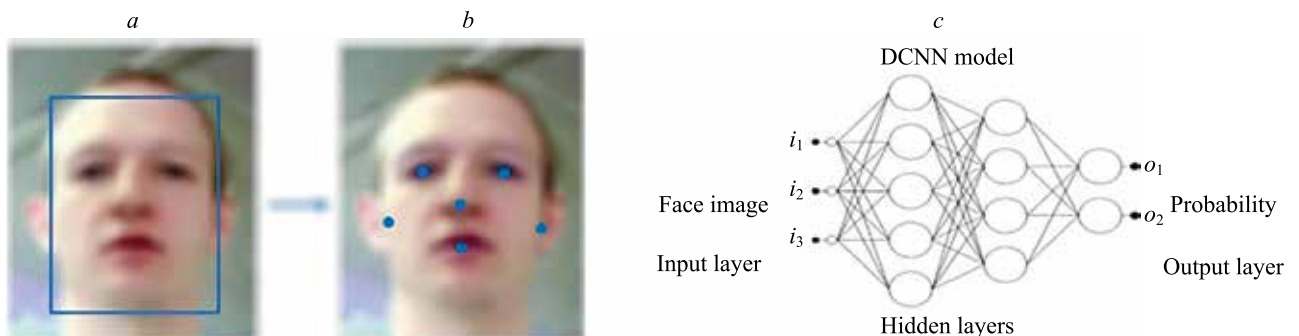


Fig. 1. Pre-processing image: face detection (a); feature extraction (b); image classification (c)

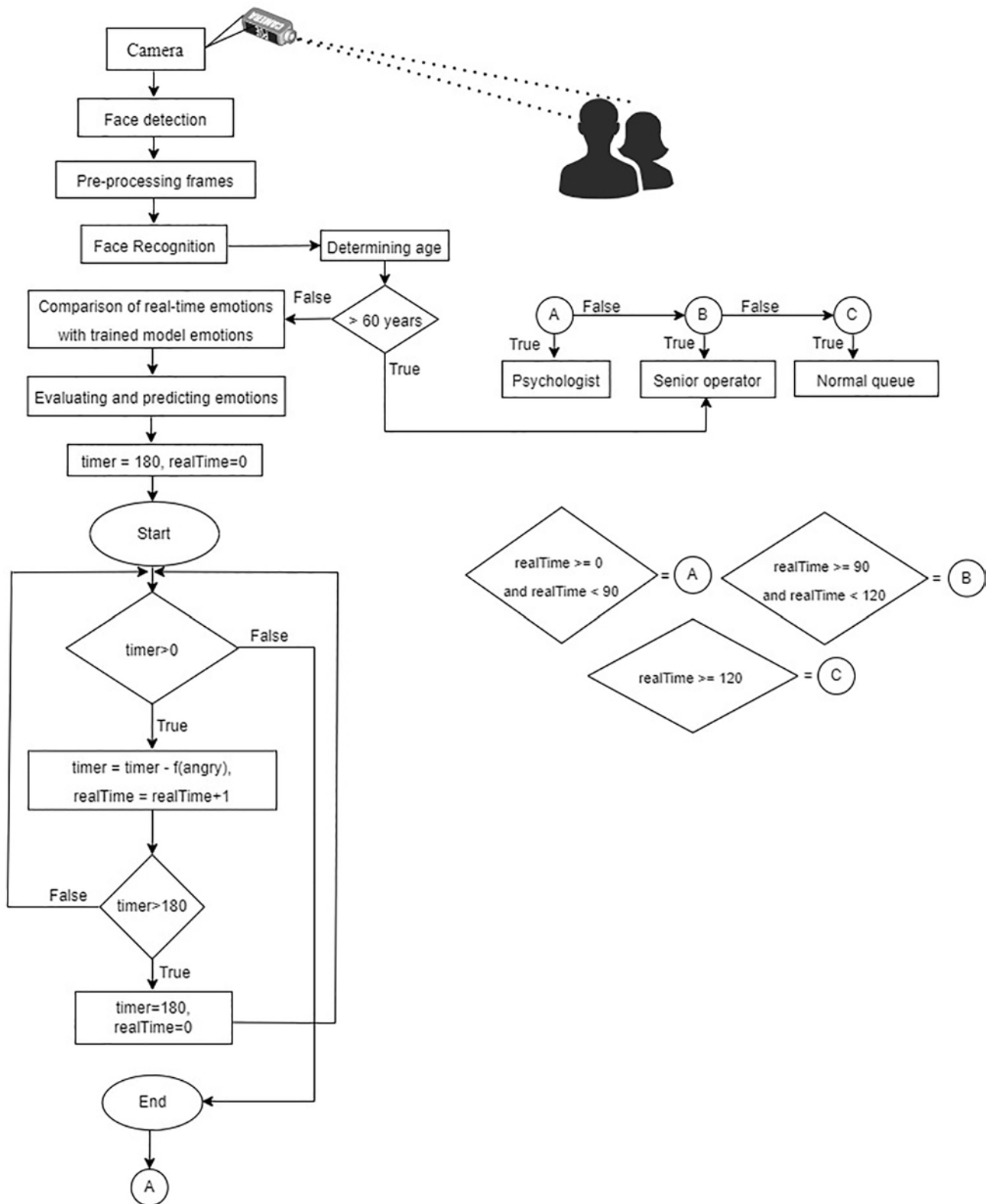


Fig. 2. Overall procedure of the developed TERSS

whose age is above 60 will be sent directly to a senior operator or a psychologist without waiting in the queue.

Fig. 3 depicts the DCNN architecture using the SE method of the emotion and age recognition model. The parameters are an input layer ($48 \times 48 \times 1$) for emotions and ($80 \times 80 \times 3$) for age, receiving a greyscale image

(48×48) and RGB image (80×80). These architectures have 21 hidden layers, one of which is an output layer with two outputs determining the probability that the detected emotion and age belongs to each of the given classes (angry/not angry) and (60+/60-). Conv2D is a convolution layer with parameters such as F (filters) —

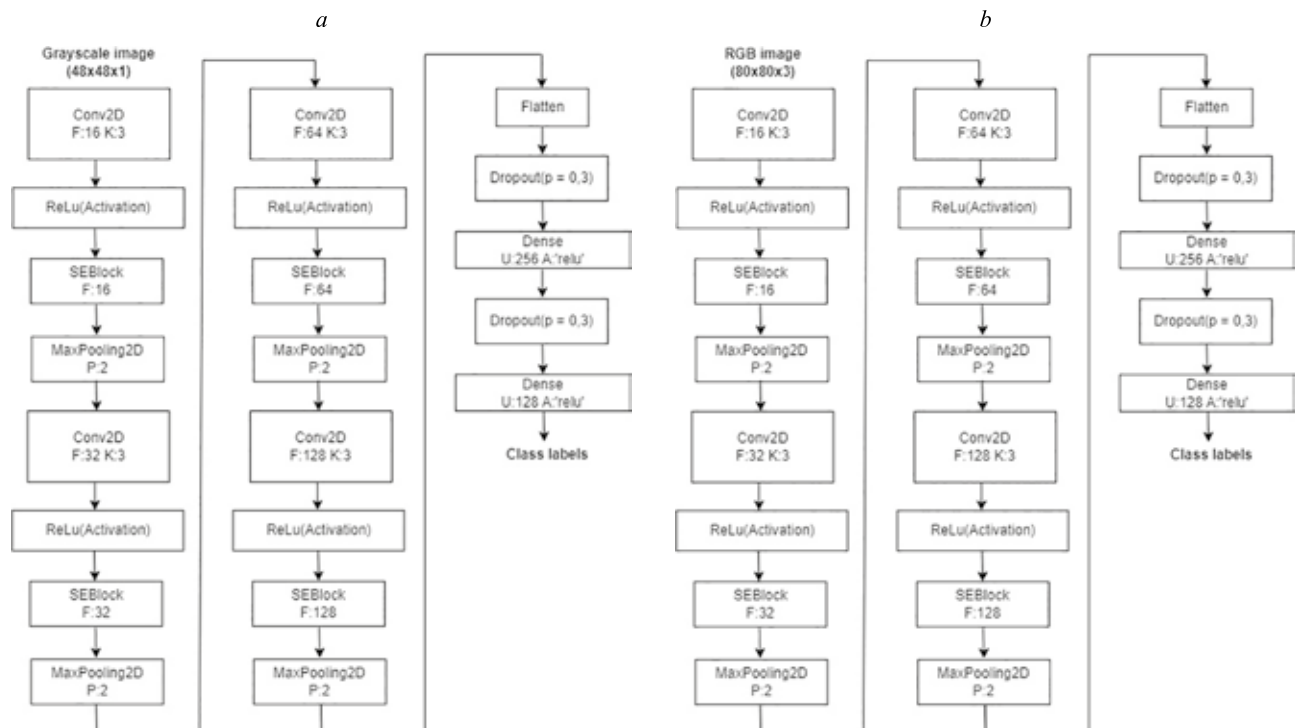


Fig. 3. DCNN architecture of the emotion (a) and age (b) recognition models

number of output filters in the convolution; K (kernel size) — convolution window length. ReLu is an activation function. SEBlock (Squeeze-and-Excitation Block) is a block in the neural network architecture that is used to improve its performance by dynamically adjusting the weights of features. It consists of two stages, namely, squeezing and excitation. In this block, features from the previous layer are aggregated, and then compressed into a single vector which is fed to a fully connected layer and activated by the sigmoid function. This allows the model to highlight the most significant features and reduce the amount of noise in the data which, in turn, improves the quality of the neural network predictions. MaxPooling2D is a subsampling layer where P (pool size) allows the input data to be reduced by a factor of n; Dropout provides a random dropout of neurons to avoid oversaturation of the neural network. Flatten transforms the input matrix into a single array. Dense is a fully connected layer, where U (units) is the dimensionality of the output space; a 'relu' is

ReLU activation function; and p is a layer that cancels some of the weights in the neural network model.

As shown in Fig. 4, in order to recognize the emotional state of the customer, it is necessary to consider the customer characteristics and the information sources about the customer which represent the general information provided by the customer through non-verbal communication.

Fig. 5 presents a scheme of the operator's initial data consisting of the operator's sources and characteristics. For operators in PSSs, information is required on personal characteristics, such as level of training (trainee, experienced operator, operator-psychologist), high communicative competence, competence by credit and salary, behavioral flexibility, and ability to communicate without conflicts [27].

Fig. 6 presents the functional diagram of the TERSS. The diagram shows four different algorithms for customer

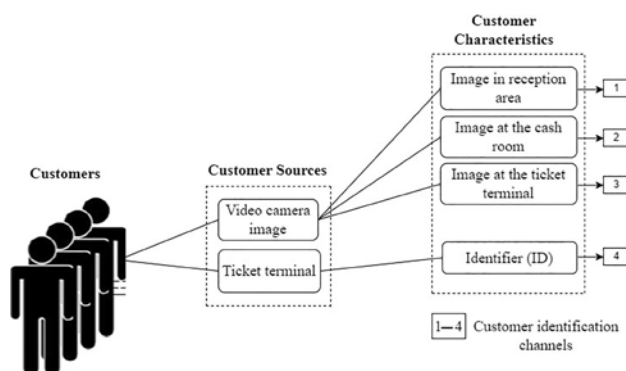


Fig. 4. Customer input data scheme

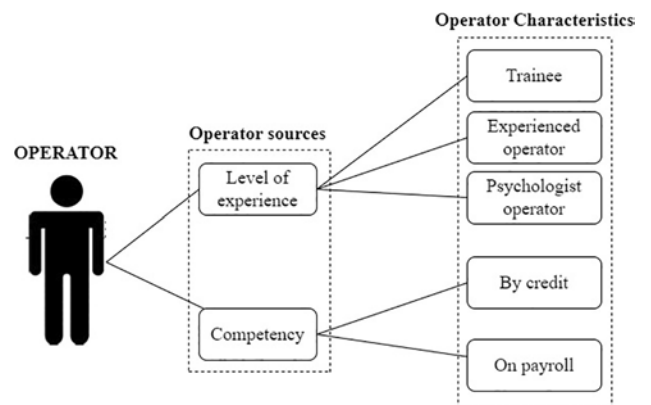


Fig. 5. Operator input data scheme

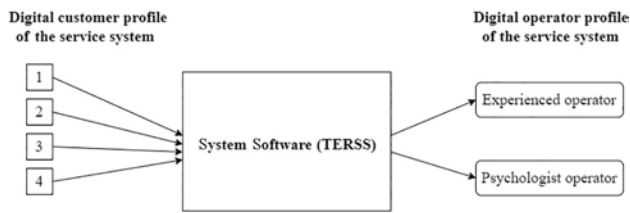


Fig. 6. TERSS operating scheme

identification, which is the input of TERSS, while the output is the routing of customers to senior operators or psychologist-operators based on their (customers') emotional state and age.

TERSS mathematical model

The resulting mathematical model of TERSS based on Newton's interpolation polynomial formula is presented below:

$$f(x) = \begin{cases} -1.00; & [0, 0.2[\\ -1.00 + 3.33(x - 0.2) + & [0.2, 0.6[\\ + 6.67(x - 0.2)(x - 0.35); & \\ +1.00 + 6.50(x - 0.6) + & [0.6, 0.9[\\ + 1.67(x - 0.6)(x - 0.8); & \\ +3.00; & [0.9, 1[\end{cases}$$

This function determines the number of time units elapsing in one second depending on the emotional state of the incoming customer. The definition ranges of this function are percentages that determine the level of customer anger. For example, if this level is in the range of $[0.2, 0.6[$ according to the definition of the visual channel, the waiting time (the distance from the customer to the operator) will be longer than for the customer whose anger level is defined in the range of $[0.6, 0.9[$. For customers in the interval of $[0, 0.2[$ the waiting time increases, because in this emotional state such customers are served in the normal queue order.

The function changes according to the graph below:

$$\begin{aligned} \text{Time_initial} &= 180 \text{ seconds} \\ \text{Time_current} &= \text{Time_initial} - F(x + y). \end{aligned}$$

Discussion and experimental results

As the accuracy of facial expression recognition algorithms is dependent on the dataset used for training and

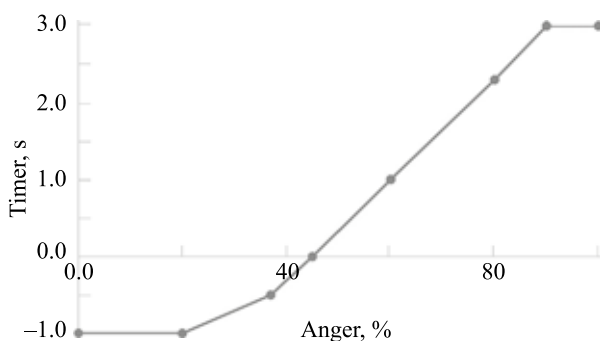


Fig. 7. TERSS Emotion Tracking Graph [28]

testing, it is appropriate to compare different methods on the same dataset. Therefore, to compare facial expression recognition methods on still images, we used the FER-2013 and Adience datasets to examine the results. In [29], an intelligent system was proposed that recognizes and classifies emotions through facial expressions using a multilayer convolutional neural network model. Another learning method was also investigated using transfer learning technology on a pre-trained ResNet50 model trained on the FER-2013 dataset. To demonstrate the effectiveness of the trained model, a basic live video application was developed to track and record facial emotions in real-time during a live video broadcast and summarize the overall responses at the end. Following the training and testing, the model achieved a validation accuracy of 61.47 %. In [30], the authors proposed a new multimodal method for age classification using a Seg-Net based architecture and a machine learning based on SVM classifier. Seg-Net is a neural network architecture used for semantic image segmentation, while SVM is a machine learning method used for data classification. In this case, the method is used for classifying age groups using the Adience dataset. The authors in [30] emphasize that the proposed method showed better results with an accuracy of 74.5 % on the Adience dataset.

According to the graphs presented in Fig. 8, we have observed good results during the training of the developed emotion classifier using SE method on DCNN architecture. The SE method is a technique used in DCNN architecture to improve its efficiency. It involves squeezing information and then exciting it at each convolutional stage. During the squeezing process, spatial dimensions of the feature map are averaged to obtain a vector that contains channel-wise importance scores. This vector is then used during the excitation process to reweight the feature map channels. Using the SE method improves the efficiency of the DCNN by enhancing information flow between layers and amplifying important features. The loss function value decreases by 0.5606, and the validation accuracy increases up to 72 %. In the developed age classifier using SE method on DCNN, the value of the loss function decreases by 0.2890, and the validation accuracy increases up to 88 %. Our classifiers achieved the highest validation accuracy in comparison with two previously published scientific results on emotion classification in the FER-2013 dataset and age classification in the Adience dataset, as shown in Table 1. The equally important result of our trained models is a comparison of weights and real-time recognition speed for emotions and ages. We aimed to make the model lighter and faster. Table 1 demonstrates the execution time of the model which efficiently and quickly recognizes emotions within seconds. We reduced the size of the model by decreasing the number of hidden layers and neurons. To calculate the execution time, we used two separate photos from the Test_set and calculated the recognition time on each photo, dividing this time by the number of photos in the test sample.

Tables 2, 3, and 4 demonstrate the predictions of the models after training. Using the three metrics from the classification report function, we can understand how well our developed classification models have been trained and are able to predict further results.

Table 1. Classification model test results

Model	Method	Model size, MB	Dataset	Execution time, s	Training accuracy	Validation accuracy
Emotion classifier developed	Squeeze-and-Excitation	0.69	FER-2013	0.00087	0.8	0.7200
Age classifier developed	Squeeze-and-Excitation	1.68	FER-2013	0.00280	0.9	0.8800
Emotion classifier developed	SVM	369	Adience	0.05400	0.8	0.6600
ResNet50 by transfer-learning [29]	—	—	FER-2013	—	—	0.6147
Age classifier [30]	Multimodal method, SVM	—	Adience	—	—	0.7450

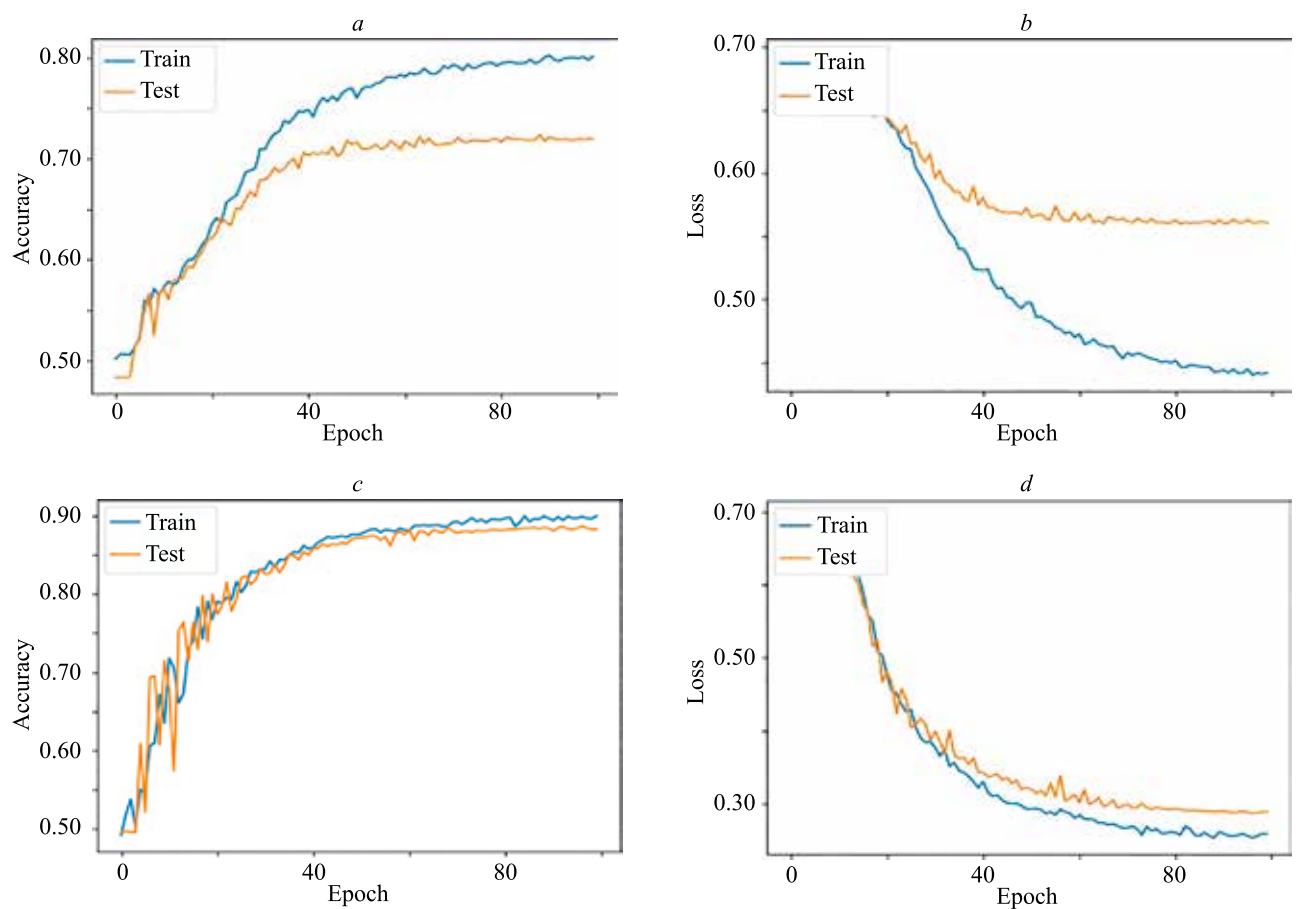


Fig. 8. Training and testing of accuracy and loss: developed emotion classifier (a, b), and age classifier (c, d)

Table 2. Classification report of the developed emotion model using SE method

	Precision	Recall	F1-score	Support
Angry	0.75	0.68	0.71	1279
Not Angry	0.69	0.76	0.72	1198
Accuracy	—	—	0.72	2477
Macro avg	0.72	0.72	0.72	2477
Weighted avg	0.72	0.72	0.72	2477

Table 3. Classification report of the developed emotion model using SVM method

	Precision	Recall	F1-score	Support
Angry	0.66	0.69	0.67	1015
Not Angry	0.66	0.62	0.64	967
Accuracy	—	—	0.66	1982
Macro avg	0.66	0.66	0.66	1982
Weighted avg	0.66	0.66	0.66	1982

Table 4. Classification report of the developed age model using SE method

	Precision	Recall	F1-score	Support
60+	0.89	0.86	0.88	668
60–	0.87	0.89	0.88	657
Accuracy	—	—	0.88	1325
Macro avg	0.88	0.88	0.88	1325
Weighted avg	0.88	0.88	0.88	1325

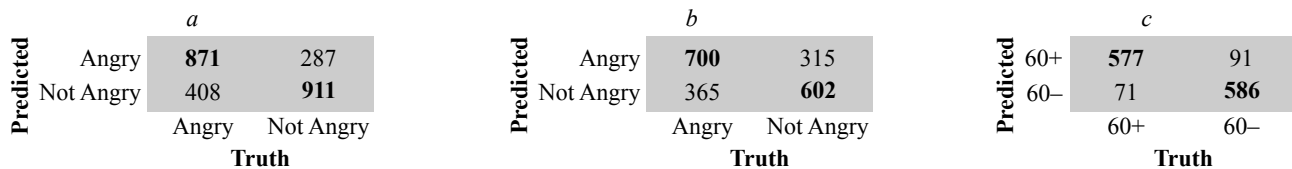


Fig. 9. Confusion matrix of the developed model using SE and SVM method for emotions (a, b) and age model using SE method (c)

Fig. 9 demonstrates that the classifiers developed for emotions and age performed well on the Unweighted Average Recall, UAR, according to the resulting confusion matrices. The confusion matrix allows us to understand the performance of a classifier (on the base of the data presented in the Fig. 9). In the confusion matrix, the rows and the columns correspond to the actual labels and to the predicted labels, respectively. The diagonals (the bold numbers) show the correctly classified samples and the standard text numbers show the incorrectly classified samples.

Limitations of the Study

In this article, we analyzed the emotional state of customers, specifically 'angry' and 'not angry', to prevent conflicts in PSSs. However, customer behavior in PSSs was not analyzed. We recognize that there are two types of customers in PSSs, namely, those who may 'get angry' and then leave the PSSs without causing a scene due to queueing for a long time, and those who may cause a scene in the queue when they get angry and their facial expression changes. Our task is to recognize facial expressions and track customers using the identifier (ID) received upon entering PSSs. If a customer leaves the queue and exits the PSSs, all information is stored in the database, which allows assessing their satisfaction levels in PSSs. We limited ourselves to using the FER2013 dataset instead of the AffectNet dataset in training our model due to the difficulty of obtaining it from the official website. The versions we found on the Internet did not meet all of our requirements and were cluttered.

Conclusions

To summarize, we compared our results to the similar studies where the authors trained emotion classifiers using pre-prepared models from ResNet50 on the same FER2013 dataset [29], and age classifiers in the Adience dataset [30]. The validation accuracies obtained in [29] and [30] were 61.47 % and 74.5 %, respectively.

Based on our results, we can conclude that our developed age model and emotion model using SE method give better results than the compared studies. In the future, we plan to improve our models to achieve over 90 % validation accuracy. Additionally, we would like to apply this work to various types of PSSs in Angola, such as banks, supermarkets, and other customer service settings.

Conclusion

In this study, we developed a mathematical model of Target Emotional Routing System Software based on Newton polynomial interpolation for Public Service Systems (PSSs). We also developed an emotion and age classifier using Deep Convolutional Neural Network with the Squeeze-and-Excitation (SE) method as well as an emotion classifier using Support Vector Machine (SVM). For training the emotion classifier using SE method, we use the Adam optimizer and trained three times for 100 epochs. The accuracy gradually increased, and the loss gradually decreased for a few epochs, after which no further improvements in training were observed. Based on the results obtained after training, we observed that the

emotion classifier using SE method achieved a validation accuracy of 72 %, a size of 0.69 MB and an execution time of 0.00087 s. The Developed age classifier using SE method achieved a validation accuracy of 88 %, a size of 1.68 MB

and execution time of 0.0028 s. The emotion classifier using SVM method achieved a validation accuracy of 66 %, a size of 369 MB, and execution time of 0.054 s.

References

1. Adigwe P., Okoro E. Human communication and effective interpersonal relationships: an analysis of client counseling and emotional stability. *International Journal of Economics & Management Sciences*, 2016, vol. 5, no. 3. <https://doi.org/10.4172/2162-6359.1000336>
2. Giangreco A., Carugati A., Sebastiano A., Al Tamimi H. War outside, ceasefire inside: An analysis of the performance appraisal system of a public hospital in a zone of conflict. *Evaluation and Program Planning*, 2012, vol. 35, no. 1, pp. 161–170. <https://doi.org/10.1016/j.evalprogplan.2010.11.004>
3. Knutson B. Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 1996, vol. 20, no. 3, pp. 165–182. <https://doi.org/10.1007/bf02281954>
4. Compas B.E. Psychobiological processes of stress and coping: implications for resilience in children and adolescents—comments on the papers of Romeo & McEwen and Fisher et al. *Annals of the New York Academy of Sciences*, 2006, vol. 1094, no. 1, pp. 226–234. <https://doi.org/10.1196/annals.1376.024>
5. Zapf D. Emotion work and psychological well-being: A review of the literature and some conceptual considerations. *Human Resource Management Review*, 2002, vol. 12, no. 2, pp. 237–268. [https://doi.org/10.1016/s1053-4822\(02\)00048-7](https://doi.org/10.1016/s1053-4822(02)00048-7)
6. Dong Y., Liu Y., Lian S. Automatic age estimation based on deep learning algorithm. *Neurocomputing*, 2016, vol. 187, pp. 4–10. <https://doi.org/10.1016/j.neucom.2015.09.115>
7. Giannakakis G., Koujan M.R., Roussos A., Marias K. Automatic stress analysis from facial videos based on deep facial action units recognition. *Pattern Analysis and Applications*, 2022, vol. 25, no. 3, pp. 521–535. <https://doi.org/10.1007/s10044-021-01012-9>
8. Gwyn T., Roy K., Atay M. Face recognition using popular deep net architectures: A brief comparative study. *Future Internet*, 2021, vol. 13, no. 7, pp. 164. <https://doi.org/10.3390/fi13070164>
9. Kumar S., Singh S., Kumar J., Prasad K.M.V.V. Age and gender classification using Seg-Net based architecture and machine learning. *Multimedia Tools and Applications*, 2022, vol. 8, no. 29, pp. 42285–42308. <https://doi.org/10.1007/s11042-021-11499-3>
10. Leist A., Playne D.P., Hawick K.A. Exploiting graphical processing units for data-parallel scientific applications. *Concurrency and Computation: Practice and Experience*, 2009, vol. 21, no. 18, pp. 2400–2437. <https://doi.org/10.1002/cpe.1462>
11. Madhavi M., Gujar I., Jadhao V., Gulwani R. Facial emotion classifier using convolutional neural networks for reaction review. *ITM Web of Conferences*, 2022, vol. 44, pp. 03055. <https://doi.org/10.1051/itmconf/20224403055>
12. Vdovina M.V. The regulation of conflict between social worker and recipient of social services. *Wschodnioeuropejskie Czasopismo Naukowe (East European Scientific Journal)*, 2015, vol. 3, no. 3, pp. 124–131. (in Russian)
13. Mouatasim A.E. Fast gradient descent algorithm for image classification with neural networks. *Signal, Image and Video Processing*, 2020, vol. 14, no. 8, pp. 1565–1572. <https://doi.org/10.1007/s11760-020-01696-2>
14. Nguyen H.-D., Kim S.-H., Lee G.-S., Yang H.-J., Na I.-S., Kim S.-H. Facial expression recognition using a temporal ensemble of multi-level convolutional neural networks. *IEEE Transactions on Affective Computing*, 2022, vol. 13, no. 1, pp. 226–237. <https://doi.org/10.1109/TAFFC.2019.2946540>
15. Nikishina I.Y. *The expression of «anger» concept in the modern English and American fiction*. Extended abstract of Candidate of Philological Sciences Dissertation. Moscow, 2008. 23 p. (in Russian).
16. Olaide O.B., Ojo A.K. A model for conflicts' prediction using deep neural network. *International Journal of Computer Applications*, 2021, vol. 183, no. 29, pp. 8–13. <https://doi.org/10.5120/ijca2021921667>

Литература

1. Adigwe P., Okoro E. Human communication and effective interpersonal relationships: an analysis of client counseling and emotional stability // *International Journal of Economics & Management Sciences*. 2016. V. 5. N 3. <https://doi.org/10.4172/2162-6359.1000336>
2. Giangreco A., Carugati A., Sebastiano A., Al Tamimi H. War outside, ceasefire inside: An analysis of the performance appraisal system of a public hospital in a zone of conflict // *Evaluation and Program Planning*. 2012. V. 35. N 1. P. 161–170. <https://doi.org/10.1016/j.evalprogplan.2010.11.004>
3. Knutson B. Facial expressions of emotion influence interpersonal trait inferences // *Journal of Nonverbal Behavior*. 1996. V. 20. N 3. P. 165–182. <https://doi.org/10.1007/bf02281954>
4. Compas B.E. Psychobiological processes of stress and coping: implications for resilience in children and adolescents—comments on the papers of Romeo & McEwen and Fisher et al. // *Annals of the New York Academy of Sciences*. 2006. V. 1094. N 1. P. 226–234. <https://doi.org/10.1196/annals.1376.024>
5. Zapf D. Emotion work and psychological well-being: A review of the literature and some conceptual considerations // *Human Resource Management Review*. 2002. V. 12. N 2. P. 237–268. [https://doi.org/10.1016/s1053-4822\(02\)00048-7](https://doi.org/10.1016/s1053-4822(02)00048-7)
6. Dong Y., Liu Y., Lian S. Automatic age estimation based on deep learning algorithm // *Neurocomputing*. 2016. V. 187. P. 4–10. <https://doi.org/10.1016/j.neucom.2015.09.115>
7. Giannakakis G., Koujan M.R., Roussos A., Marias K. Automatic stress analysis from facial videos based on deep facial action units recognition // *Pattern Analysis and Applications*. 2022. V. 25. N 3. P. 521–535. <https://doi.org/10.1007/s10044-021-01012-9>
8. Gwyn T., Roy K., Atay M. Face recognition using popular deep net architectures: A brief comparative study // *Future Internet*. 2021. V. 13. N 7. P. 164. <https://doi.org/10.3390/fi13070164>
9. Kumar S., Singh S., Kumar J., Prasad K.M.V.V. Age and gender classification using Seg-Net based architecture and machine learning // *Multimedia Tools and Applications*. 2022. V. 81. N 29. P. 42285–42308. <https://doi.org/10.1007/s11042-021-11499-3>
10. Leist A., Playne D.P., Hawick K.A. Exploiting graphical processing units for data-parallel scientific applications // *Concurrency and Computation: Practice and Experience*. 2009. V. 21. N 18. P. 2400–2437. <https://doi.org/10.1002/cpe.1462>
11. Madhavi M., Gujar I., Jadhao V., Gulwani R. Facial emotion classifier using convolutional neural networks for reaction review // *ITM Web of Conferences*. 2022. V. 44. P. 03055. <https://doi.org/10.1051/itmconf/20224403055>
12. Вдовина М.В. Регулирование конфликта между социальным работником и получателем социальных услуг // *Wschodnioeuropejskie Czasopismo Naukowe (East European Scientific Journal)*. 2015. Т. 3. № 3. С. 124–131.
13. Mouatasim A.E. Fast gradient descent algorithm for image classification with neural networks // *Signal, Image and Video Processing*. 2020. V. 14. N 8. P. 1565–1572. <https://doi.org/10.1007/s11760-020-01696-2>
14. Nguyen H.-D., Kim S.-H., Lee G.-S., Yang H.-J., Na I.-S., Kim S.-H. Facial expression recognition using a temporal ensemble of multi-level convolutional neural networks // *IEEE Transactions on Affective Computing*. 2022. V. 13. N 1. P. 226–237. <https://doi.org/10.1109/TAFFC.2019.2946540>
15. Никишина И.Ю. Выражение концепта anger («гнев») в современной английской и американской художественной литературе: автореферат диссертации на соискание ученой степени кандидата филологических наук / МГУ им. М.В.Ломоносова. М., 2008. 23 с.
16. Olaide O.B., Ojo A.K. A model for conflicts' prediction using deep neural network // *International Journal of Computer Applications*. 2021. V. 183. N 29. P. 8–13. <https://doi.org/10.5120/ijca2021921667>

17. Paper D. Classification from complex training sets. *Hands-on Scikit-Learn for Machine Learning Applications*. Apress, Berkeley, CA, 2020, pp. 71–104. https://doi.org/10.1007/978-1-4842-5373-1_3
18. Piatak J., Romzek B., LeRoux K., Johnston J. Managing goal conflict in public service delivery networks: Does accountability move up and down, or side to side? *Public Performance & Management Review*, 2018, vol. 41, no. 1, pp. 152–176. <https://doi.org/10.1080/15309576.2017.1400993>
19. Pollak S.D., Camras L.A., Cole P.M. Progress in understanding the emergence of human emotion. *Developmental Psychology*, 2019, vol. 55, no. 9, pp. 1801–1811. <https://doi.org/10.1037/dev0000789>
20. Reichel L. Newton interpolation at Leja points. *BIT*, 1990, vol. 30, no. 2, pp. 332–346. <https://doi.org/10.1007/BF02017352>
21. Rodrigues M.V. *Perspectives of Communication and Communicative Competence*. Concept Publishing Company, 2000, 390 p.
22. Ryumina E.V., Karpov A.A. Analytical review of methods for emotion recognition by human face expressions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 2, pp. 163–176 (in Russian). <https://doi.org/10.17586/2226-1494-2020-20-2-163-176> (in Russian)
23. El-Glaly Y.N., Quek F. Digital reading support for the blind by multimodal interaction. *ICMI '14: Proc. of the 16th International Conference on Multimodal Interaction*, 2014, pp. 439–446. <https://doi.org/10.1145/2663204.2663266>
24. Hameduddin T., Engbers T. Leadership and public service motivation: a systematic synthesis. *International Public Management Journal*, 2022, vol. 25, no. 1, pp. 86–119. <https://doi.org/10.1080/10967494.2021.1884150>
25. Thrassou A., Santoro G., Leonidou E., Vrontis D., Christofi M. Emotional intelligence and perceived negative emotions in intercultural service encounters: Building and utilizing knowledge in the banking sector. *European Business Review*, 2020, vol. 32, no. 3, pp. 359–381. <https://doi.org/10.1108/eb-04-2019-0059>
26. Varma S., Shinde M., Chavan S.S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers. *Techno-Societal 2018: Proc. of the 2nd International Conference on Advanced Technologies for Societal Applications. V. 1*, 2020, pp. 109–119. https://doi.org/10.1007/978-3-030-16848-3_11
27. Petri V., Jari K. Public service systems and emerging systemic governance challenges. *International Journal of Public Leadership*, 2015, vol. 11, no. 2, pp. 77–91. <https://doi.org/10.1108/IJPL-02-2015-0007>
28. Wu Z., Shen C., van den Hengel A. Wider or deeper: Revisiting the ResNet model for visual recognition. *Pattern Recognition*, 2019, vol. 90, pp. 119–133. <https://doi.org/10.1016/j.patcog.2019.01.006>
29. Tian Y. Evaluation of face resolution for expression analysis. *Proc. of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 82. <https://doi.org/10.1109/CVPR.2004.334>
30. Zaghibani S., Bouhleb M.S. Multi-task CNN for multi-cue affects recognition using upper-body gestures and facial expressions. *International Journal of Information Technology*, 2022, vol. 14, no. 1, pp. 531–538. <https://doi.org/10.1007/s41870-021-00820-w>
17. Paper D. Classification from complex training sets // *Hands-on Scikit-Learn for Machine Learning Applications*. Apress, Berkeley, CA, 2020. P. 71–104. https://doi.org/10.1007/978-1-4842-5373-1_3
18. Piatak J., Romzek B., LeRoux K., Johnston J. Managing goal conflict in public service delivery networks: Does accountability move up and down, or side to side? // *Public Performance & Management Review*. 2018. V. 41. N 1. P. 152–176. <https://doi.org/10.1080/15309576.2017.1400993>
19. Pollak S.D., Camras L.A., Cole P.M. Progress in understanding the emergence of human emotion // *Developmental Psychology*. 2019. V. 55. N 9. P. 1801–1811. <https://doi.org/10.1037/dev0000789>
20. Reichel L. Newton interpolation at Leja points // *BIT*. 1990. V. 30. N 2. 332–346. <https://doi.org/10.1007/BF02017352>
21. Rodrigues M.V. *Perspectives of Communication and Communicative Competence*. Concept Publishing Company, 2000. 390 p.
22. Рюмина Е.В., Карпов А.А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // *Научно-технический вестник информационных технологий, механики и оптики*. 2020. V. 20. N 2. P. 163–176. <https://doi.org/10.17586/2226-1494-2020-20-2-163-176>
23. El-Glaly Y.N., Quek F. Digital reading support for the blind by multimodal interaction // *ICMI '14: Proc. of the 16th International Conference on Multimodal Interaction*. 2014. P. 439–446. <https://doi.org/10.1145/2663204.2663266>
24. Hameduddin T., Engbers T. Leadership and public service motivation: a systematic synthesis // *International Public Management Journal*. 2022. V. 25. N 1. P. 86–119. <https://doi.org/10.1080/10967494.2021.1884150>
25. Thrassou A., Santoro G., Leonidou E., Vrontis D., Christofi M. Emotional intelligence and perceived negative emotions in intercultural service encounters: Building and utilizing knowledge in the banking sector // *European Business Review*. 2020. V. 32. N 3. P. 359–381. <https://doi.org/10.1108/eb-04-2019-0059>
26. Varma S., Shinde M., Chavan S.S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers // *Techno-Societal 2018: Proc. of the 2nd International Conference on Advanced Technologies for Societal Applications. V. 1*. 2020. P. 109–119. https://doi.org/10.1007/978-3-030-16848-3_11
27. Petri V., Jari K. Public service systems and emerging systemic governance challenges // *International Journal of Public Leadership*. 2015. V. 11. N 2. P. 77–91. <https://doi.org/10.1108/IJPL-02-2015-0007>
28. Wu Z., Shen C., van den Hengel A. Wider or deeper: Revisiting the ResNet model for visual recognition // *Pattern Recognition*. 2019. V. 90. P. 119–133. <https://doi.org/10.1016/j.patcog.2019.01.006>
29. Tian Y. Evaluation of face resolution for expression analysis // *Proc. of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*. 2004. P. 82. <https://doi.org/10.1109/CVPR.2004.334>
30. Zaghibani S., Bouhleb M.S. Multi-task CNN for multi-cue affects recognition using upper-body gestures and facial expressions // *International Journal of Information Technology*. 2022. V. 14. N 1. P. 531–538. <https://doi.org/10.1007/s41870-021-00820-w>

Authors

Guedes M. Soma — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57479369200](mailto:guedes.soma@mail.ru), <https://orcid.org/0000-0001-9004-3554>, guedes.soma@mail.ru

Georgy D. Kopanitsa — PhD (Medical & Biology Sciences), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 55326019500, <https://orcid.org/0000-0002-6231-8036>, kopanitsa@itmo.ru

Авторы

Сом Гедеш Мануэл — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57479369200, <https://orcid.org/0000-0001-9004-3554>, guedes.soma@mail.ru

Копаница Георгий Дмитриевич — PhD медико-биологические науки, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 55326019500, <https://orcid.org/0000-0002-6231-8036>, kopanitsa@itmo.ru

Received 18.01.2023

Approved after reviewing 20.06.2023

Accepted 21.09.2023

Статья поступила в редакцию 18.01.2023

Одобрена после рецензирования 20.06.2023

Принята к печати 21.09.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»