http://doi.org/10.35784/iapgos.7156 received: 26.01.2025 | revised: 10.04.2025 | accepted: 05.05.2025 | available online: 27.06.2025

ANALYSIS OF SELECTED METHODS OF PERSON IDENTIFICATION BASED ON BIOMETRIC DATA

Marcin Amadeusz Rudzki, Paweł Powroźnik
Lublin University of Technology, Department of Computer Science, Lublin, Poland

Abstract. *The article explores the challenge of identifying individuals using biometric data through advanced deep learning methods. The research employs three ground-breaking convolutional neural network architectures: ResNet50, EfficientNetB0, and VGG16. The project's objective was to examine the influence of critical factors, such as image quality and data processing techniques, on the performance of face identification systems. A series of experiments were carried out based on predefined test scenarios, allowing for the verification of hypotheses regarding the effects of input image resolution and data transformations on model accuracy. The experimental results highlight the substantial impact of both the chosen architecture and processing parameters on the system's identification accuracy. The article presents valuable conclusions that can inform the further development of biometric systems. Notably, the EfficientNetB0 model achieved the best performance across various metrics, including the confusion matrix and activation heatmaps, demonstrating its superior capability in identifying biometric data from facial images.*

Keywords: biometric identification, convolutional neural networks, image processing

ANALIZA WYBRANYCH METOD IDENTYFIKACJI OSÓB NA PODSTAWIE DANYCH BIOMETRYCZNYCH

Streszczenie. *W artykule porównano metody identyfikacji osób przy użyciu danych biometrycznych za pomocą zaawansowanych metod głębokiego uczenia. W trakcie badań zostały użyte trzy architektury sieci neuronowych konwolucyjnych: ResNet50, EfficientNetB0 i VGG16. Celem badań była ocena wpływu czynników krytycznych, takich jak jakość obrazu i techniki przetwarzania danych, na wydajność systemów identyfikacji twarzy. Przeprowadzono serię eksperymentów w oparciu o wstępnie zdefiniowane scenariusze testowe, co pozwoliło na weryfikację hipotez dotyczących wpływu rozdzielczości obrazu wejściowego i transformacji danych na dokładność modelu. Wyniki eksperymentów podkreślają istotny wpływ zarówno wybranej architektury, jak i parametrów przetwarzania na dokładność całego identyfikacji systemu. W artykule przedstawiono cenne wnioski, które mogą posłużyć do dalszego rozwoju systemów biometrycznych. Co godne uwagi, model EfficientNetB0 osiągnął najlepszą wydajność w różnych metrykach, w tym macierzy pomyłek i mapach cieplnych aktywacji, co dowodzi jego wyższej zdolności do identyfikacji danych biometrycznych z obrazów twarzy.*

Słowa kluczowe: identyfikacja biometryczna, konwolucyjne sieci neuronowe, przetwarzanie obrazów

Introduction

Identification of individuals based on biometric data, such as facial images, is a fundamental component of modern security systems. Biometric features, including the eyes, facial structure, nose, and other unique characteristics like voice samples, provide distinctive and hard-to-forge markers that enable accurate recognition. Advances in technology and the development of artificial intelligence, particularly deep learning algorithms, have significantly enhanced the capabilities of biometric systems [11]. In recent years, convolutional neural networks (CNNs) have revolutionized this field, achieving high accuracy in facial recognition. However, the effectiveness of these systems is influenced by various factors, including image quality and the data processing techniques employed.

The motivation for exploring this topic arises from the need to understand how factors like image quality, data transformations, and model selection influence the performance of biometric identification systems. In an age where security concerns are steadily growing, optimizing these systems is paramount for reducing vulnerabilities and bolstering accuracy. By investigating the interplay of these elements, more efficient methods can be uncovered, leading to better real-world applications.

Recent advances in face recognition emphasize the importance of robust deep learning architectures and feature extraction methods. In particular, convolutional neural networks have shown remarkable performance improvements, highlighting the need for careful consideration of computational efficiency and interpretability [4].

The aim of this study was to analyse the impact of image quality and data transformations on the effectiveness of three selected CNN architectures: ResNet50, EfficientNetB0, and VGG16. The paper proposes two main research hypotheses:
H1: The effectiveness of face identification methods depends on the quality of the camera image.
H2: The appropriate selection of the model improves the performance of the face identification system.

To evaluate these hypotheses, experiments were conducted under two training scenarios, differing in image resolution and the application of data transformations.

The rest of the paper consists as follows: a review of the literature on convolutional neural networks and their applications in biometric identification systems, a description of the methodology and research scenarios employed, an analysis of the results, and a discussion of their significance. The conclusions drawn from the experiments are summarized in the final section, which also includes a presentation of Grad-CAM visualizations for the obtained results.

1. Related works

Over the past decades, convolutional neural networks (CNNs) have become a standard in computer vision tasks. By automatically extracting features from input data, CNNs have revolutionized the field of image analysis, enabling high accuracy in tasks such as classification, object detection, and identification [11]. They are now widely used not only in biometric systems, but also in various areas such as medical diagnostics or broadly understood image or video analysis.

Recent studies highlight the integration of advanced algorithms to enhance facial recognition accuracy. For instance, Kamyab et al. propose a combination of genetic algorithms and neural networks to select facial features, which significantly improves the recognition process [5]. Similarly, Zarei discusses various face recognition methods, emphasizing the effectiveness of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in feature extraction [19]. The work of Paul and Aslan further supports this by demonstrating an improved real-time face recognition system that utilizes Local Binary Pattern Histogram and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance performance under low-resolution conditions [8].

The application of deep learning techniques has significantly advanced face recognition systems. Wang et al. analysis indicates that algorithms such as DeepFace and FaceNet have achieved

remarkable improvements in recognition efficiency and accuracy compared to traditional methods [17]. This is corroborated by the findings of Song and Ji, who developed a Siamese network that maintains high recognition rates while improving computational speed under non-restricted conditions [14]. Moreover, study performed by Wu introduces a masked face recognition algorithm that effectively addresses occlusions, which is a common challenge in real-world applications [18].

The performance of facial recognition systems is often evaluated against human capabilities. Phillips et al. provide a comparative analysis of face recognition algorithms against human examiners, revealing that while machines can perform at par with skilled humans, optimal results are achieved through collaboration [9].

2. Materials and methods

This section presents a detailed description of the materials and methods used to conduct experiments such as, dataset preparation process, the architecture of the network used, the training and evaluation procedure will be also presented. Basic metrics for assessing the quality of the model, like accuracy and precision, allowing for an objective verification of the achieved results, will also be discussed.

2.1. Convolutional Neural Networks

The CNN architecture consists of convolutional, pooling, and fully connected layers. Convolutional layers are responsible for detecting complex patterns such as edges or object structures, and pooling layers reduce the data dimensions, reducing the number of parameters and improving computational efficiency. For the final classification, fully connected layers are used, which integrate the extracted features, allowing for obtaining final results such as classification. Activation functions such as ReLU or softmax introduce nonlinearity to the model, allowing for modelling complex relationships. Regularization and normalization techniques are also used in the image analysis process, which improves the stability and efficiency of the network. CNNs are used in medical diagnostics, autonomous vehicles, and security systems such as biometric identification. Thanks to advanced mechanisms of automatic feature extraction and the ability to work with large data sets without the participation of manual human work, these networks are a key tool in the analysis of images and other forms of visual data [10].

2.2. VGG16 model

The VGG16 model, designed by the Visual Geometry Group at the University of Oxford is characterized by simplicity and efficiency. It uses 13 convolutional layers with fixed-size 3×3 filters, which enables detailed feature capture while reducing the number of parameters [13]. The architecture also includes 3 fully connected layers and 5 MaxPool layers of size 2×2, which gradually reduce the image resolution, allowing the network to focus on the most important high-level features. The model supports input images of sizes from 32×32 to 224×224 pixels, and can process larger images thanks to the possibility of using a global average pooling layer. The final VGG16 output layer consists as many neurons as there are classes in dataset.

2.3. ResNet50 model

ResNet50, presented in 2016 in the article "Deep Residual Learning for Image Recognition" [2], introduces a key innovation in the architecture of deep neural networks – residual learning. Traditional deep neural network architectures have experienced the problem of model accuracy degradation. As the number of layers in traditional neural networks increases, the models start to achieve lower accuracy on training and test sets because they have trouble effectively propagating the back signal (gradient), which leads to its vanishing (vanishing gradient problem). Moreover, parameter optimization becomes more difficult with the addition of subsequent layers due to the appearance of the overfitting effect, which leads to an increase in the amount of unnecessary information that can deteriorate the quality of prediction. ResNet solves the problem of accuracy degradation in deep neural networks by using residual connections. This technique allows the signal to be passed directly through the network layers, bypassing those transformations that do not significantly contribute to improving learning, allowing very deep models to be trained efficiently while maintaining high prediction accuracy.

2.4. EfficientNetB0 model

The EfficientNetB0 model is one of the versions of the EfficientNet family of models. These networks introduce an innovative approach to scaling convolutional network architectures that optimize performance in terms of accuracy and computational efficiency. While traditional networks are often scaled in one of the dimensions (depth, width, or output resolution), EfficientNet uses a Compound Scaling method that uniformly increases all three dimensions by a scaling factor, allowing for better utilization of computational resources. The B0 version is the lightest and most basic version of this family of models, designed to operate on limited computational resources. EfficientNet networks, including B0, are able to achieve high accuracy with significantly fewer parameters and faster inference time compared to earlier convolutional networks such as ResNet or MobileNet. EfficientNet-B7, the most advanced version in the family, achieves 84.3% accuracy on ImageNet while being 8.4 times smaller and 6.1 times faster than the best publicly existing models [15].

2.5. Evaluation metrics

In the process of evaluating the performance of neural network models, various comparative metrics are used. The key tool is the confusion matrix, which analyses the prediction results in detail. The matrix consists of four categories: True Positive (correct assignment to the positive class), True Negative (correct assignment to the negative class), False Positive (incorrect assignment to the positive class), and False Negative (incorrect assignment to the negative class). Commonly used evaluation metrics include accuracy, precision, and F1 score [3]. The confusion matrix provides important information about the types of errors, facilitating the evaluation and interpretation of model performance. Key performance metrics such as accuracy, precision, recall, and F1-score can be derived from the confusion matrix, providing a comprehensive evaluation of classification models [16]:

Accuracy – defines the ratio of correct predictions (TP + TN) to the total number of cases (classes). It is expressed by the following equation

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision – measures what percentage of predicted positive cases are true positives. High precision means fewer false positives. It is expressed by the equation (2):

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Sensitivity (Recall) – determines what percentage of real cases were correctly identified by the model. A high sensitivity score means that the model rarely misses positive cases. It is expressed by the equation (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1 Score is the harmonic mean of the above two values, sensitivity and precision. It is expressed by the equation (4):

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \quad (4)$$

An additional technique that allows to increase the transparency of models based on convolutional neural networks (CNN) is the visualization of input regions,

the Gradient-weighted Class Activation Mapping (Grad-CAM) method [12]. Grad-CAM allows to identify key regions of the input image that have the greatest impact on the model predictions. The process starts with the analysis of gradients calculated with respect to activations in the last layer of the convolutional neural network. The gradients indicate how changing the value (from 0 to 1) at a given location affects the classification result for the selected class. Based on the gradients, weights representing the importance of each feature map in the last convolutional layer are determined, and then these weights are applied to the feature maps, which when combined, give an activation map. The given activation map is interpolated to the size of the input image to allow it to be matched to the original image. The result of these (and other) operations is a heat map, which is overlaid on the input image, visualizing the regions, which are the most important for the model predictions. High intensity areas on the map indicate parts of the image that are crucial to the model's decisions (most often marked with red colours), while low intensity areas are less important (marked with blue).

2.6. Research stand

Training neural networks requires high computing power due to the large number of operations necessary to perform. The most important criterion is an efficient graphics card (or a set of graphics cards), which is mainly responsible for performing the most complex calculations, such as convolutional operations, pooling, due to the parallel calculations it is able to perform compared to the processor unit. The study was conducted in an independent environment, using resources available at home. A series of experiments was conducted on a computer – ASUS TUF Gaming A15 laptop, equipped with AMD Ryzen 7 4800H (2.9GHz), 16GB of RAM, NVIDIA GeForce GTX 1660Ti graphics card with 6GB of built-in memory and a 1TB SSD drive. This experiment was successfully completed, which proves that training small neural networks is possible without the need for specialized equipment, however, it is associated with certain limitations, such as the number of operations performed per second or the time required to train the network.

2.7. Dataset

The dataset is one of the key elements of the training process of selected deep network architectures. In order to provide the necessary data for the research, a modified version of the VGGFace2 [1] set was selected, which contains images in higher resolution than the original set. The set is characterized by a large number of photos at the level of 629 thousand, which make up 4605 classes, which represent different people. Due to limited computational resources, the set was pre-processed so that it was possible to properly conduct the research scenarios. The data preparation process included the following steps:

1. Class selection – from the selected set, which contained over 4605 classes and 629 thousand photos, 20 classes were randomly selected and used in the experiments. The selected classes are visible in the results, such as confusion matrices.
2. Data cleaning – in the selected directories, photos that did not meet the criteria, such as images not showing faces, distorted or damaged, were manually removed. In addition, the number of images in each directory was limited to 150, by removing distorted images.
3. Data division – each of the 20 selected classes was divided in the proportion of 80/10/10% into training, validation and test data. Such division allows for an appropriate assessment of the effectiveness of selected scenarios for CNN models.

The actions taken were aimed at adapting the data set to the requirements of the study, ensuring both the quality and uniformity of samples in individual classes.

2.8. Research scenarios

This section presents research scenarios aimed at analysing the effectiveness of selected convolutional neural network models in the face identification task. The study is based on two hypotheses:

H1: The effectiveness of face identification methods depends on the quality of the camera image.

H2: The appropriate selection of the model improves the quality of the face identification system.

Hypothesis H1 assumes that the effectiveness of face identification methods is strongly dependent on the quality of the image recorded by the camera. Higher resolution and better image quality positively affect the accuracy of face identification systems. Hypothesis H2 suggests that the appropriate selection of the algorithm can significantly improve the quality of the face identification system. Particular emphasis is placed on the ability of the models to operate effectively in conditions of variable lighting and at different angles of face registration, which translates into greater system reliability.

Scenario 1: Impact of image quality on face identification performance

The first research scenario aims to verify hypothesis H1, which assumes that the effectiveness of person identification depends on image quality. The experiment examined the influence of image resolution on the results of model prediction accuracy in the context of face identification. The dataset described in Chapter 3.1 was analysed, containing photos in a resolution of 512×512 pixels. In order to adjust the resolution to the requirements of the scenario, dynamic image cropping was used during the process of loading data into the model. This process was implemented by the data transformation function, which scaled and cropped images in real time to a smaller resolution of 224×224 pixels based on a parameter specified in the script supporting a given scenario for a given model. The advantage of such a solution was the flexible adjustment of the image resolution, without the need to create additional versions of the dataset. The experiment also used the early stop loss mechanism to shorten the training time and prevent overfitting of models. Transfer learning [7], based on pre-trained models available in the PyTorch library, allowed for the efficient use of available resources without the need to train all networks from scratch, which enabled the maximum use of the models' potential and an objective comparison of their effectiveness.

Scenario 2: Data transformation and model selection on face identification performance

The second research scenario was designed to verify hypothesis H2, which assumes that appropriate model selection and the use of input data modification techniques improve the quality of the face identification system. For this purpose, experiments were conducted to analyse the impact of selected neural network model architectures on the identification efficiency. The scenario included input data modification, including methods such as rotation, mirror reflections, and brightness modifications, which aimed to increase the models' ability to generalize in changing conditions. The models were trained and evaluated in two variants: without and with modifications. The experiments used modification techniques that were individually adjusted to each image automatically. The parameters of these modifications were randomly selected from specific value ranges during each image loading into the model, which allowed for the creation of a more diverse training data set, without the need to create additional versions of the data set. Similarly, to scenario 1, transfer learning and an early stopping mechanism were used.

3. Study results

In order to implement the two research scenarios, four experiments were prepared. H1.1 and H1.2 concerned the analysis of the effect of image size (224×224 and 512×512 pixels, respectively), while H2.1 and H2.2 tested the effect of data modification (no change vs. random rotation, mirroring, brightness and contrast adjustment). Such division allowed for the assessment of both the importance of input resolution and the use of augmentation of data during model training.

3.1. Analysis of the impact of image size

The first scenario involved the analysis of two different input image sizes for the models. The analysis shows that the VGG16 model gained the largest increase in accuracy in the H1.2 scenario (Fig. 2) compared to the other models, improving its validation accuracy by 0.0733. The EfficientNetB0 model showed no change in the two experiments, which shows that this model does not care what resolution of the input images the model is supposed to process. The number of epochs and the accuracy remain practically the same, but it should be noted that the EfficientNetB0 model achieved the highest possible accuracy score in the two experiments. The ResNet50 model achieved a higher accuracy score in H1.2, however, the cost of more than twice as many epochs is disproportionate to the obtained increase. For a given model, the optimal approach was to use 224×224 pixel images with H1.1, delivering high accuracy over a much smaller range of epochs (Fig. 1).

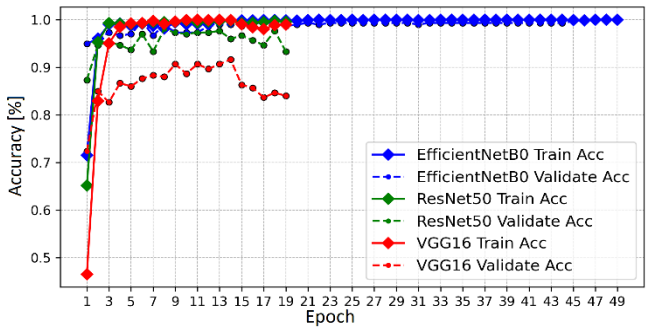

Fig. 1. Accuracy on training and validation sets for models from scenario H1.1 – the size of input images was 224×224 pixels

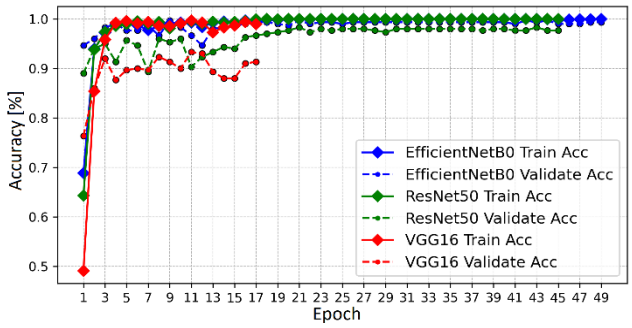

Fig. 2. Accuracy on training and validation sets for models from scenario H1.2 – the size of input images was 512×512 pixels

The summary of metrics obtained in the process of model evaluation for scenario 1 is presented in table 1. The EfficientNetB0 model demonstrates the highest results in the image classification task, achieving results in the range of 98–99%, which makes it the best among the selected models. The superior performance of EfficientNet in this study is consistent with the architecture's efficiency and accuracy demonstrated in its original publication [15], achieving the highest state of the art. In the ranking, immediately after the EfficientNet model, ResNet50 was observed, the effectiveness of which exceeds the threshold of 95%, noting an improvement of metrics in the H1.2 experiment by one percentage point compared to H1.1. Based on the presented data, a particular difference in the effectiveness of the VGG16 model was observed. While EfficientNetB0 and ResNet50 achieve scores in the range of 96–99%, VGG16 achieves the lowest scores for all three metrics. The state-of-the-art approach used in the state-of-the-art architectures clearly outperforms the models in terms of the obtained metrics. Across the two experiments, VGG16 shows the largest metric improvement for H1.2 in the precision metric, by two percentage points.

Table 1. Evaluation metrics from the final epoch of training for the EfficientNetB0, ResNet50 and VGG16 models for scenario 1

Model	Scenario	Precision	Recall	F1-Score
VGG16	H1.1	85.0%	83.0%	83.0%
	H1.2	87.0%	82.0%	83.0%
ResNet50	H1.1	96.0%	96.0%	96.0%
	H1.2	97.0%	97.0%	97.0%
EfficientNetB0	H1.1	99.0%	98.0%	98.0%
	H1.2	98.0%	98.0%	98.0%

The analysis of the confusion matrix (Fig. 3) shows that the main diagonal of the matrix contains high values (dominance of numbers 14 and 15), which indicates very good classification performance for most classes. This means that the model correctly identifies most cases. The number of misclassifications is low, but single mispredictions were observed, mainly recorded for class n000063, amounting to 12 correct predictions from the base number of test images of 15. This is due to the fact that the images are much more diverse (shots taken from different angles, in one of the images the person is wearing glasses) compared to other classes. Classification errors occur in only 3 classes, where the number of mistakes is one (for two classes) and 3 mistakes (for one class n000063). No clear error pattern was detected, which would indicate systematic problems with classification of specific classes. The EfficientNetB0 model in this experiment obtained a result confirming high metrics.

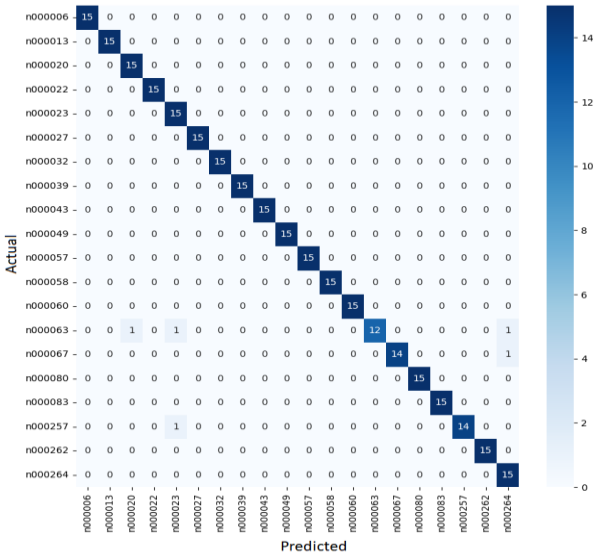

Fig. 3. Confusion matrix for the EfficientNetB0 model for experiment H1.1

Analysis of the results for scenario 1 shows that using larger image sizes improves the accuracy results only slightly (improvement by a few percentage points in a few areas), but on the example of the ResNet50 model it significantly increased the time and computational resources necessary to train the models. The optimal approach was to use a smaller image size (scenario H1.1), which allowed to obtain satisfactory accuracy results (above 95%), excluding the VGG16 model, which achieved a slightly worse result of 84% for the validation set. The benefits of using smaller sizes (H1.1) outweighed those of using larger sizes (H1.2), which contradicts the assumption H1.

3.2. Data transformation impact analysis

The second scenario concerned the analysis of the impact of the input data modification on the learning process of neural network models. Two cases were compared: H2.1, in which no data modification was applied, and H2.2, where modifications were introduced, described in Section 2.8. The accuracy results on the training and validation sets are presented in figures 3 and 4. In the case of the ResNet50 model, both training and validation accuracy remained at the same high level, but the number of epochs is higher in H2.2 by 6 epochs compared to H1.1. The highest accuracy result was achieved among all models in both experiments, which suggests that this model is resistant to data diversity and copes well with generalization without the need for additional modifications, however, it should be mentioned that the learning process of this model was the longest among the others. In the results of the study of scenario 2, the worst impact on the VGG16 model was observed in the case of the application of data modification (H2.2), which caused an increase in the number of epochs (H2.1 needed 14 epochs, while H2.2 – as many as 38), moreover, both training and validation accuracy deteriorated slightly (for H2.1 the validation accuracy is 0.87, for H2.2 – 0.86). A negative impact of data modification was also observed for the EfficientNetB0 model, where the number of epochs was more than twice as large (for H2.1 – 19 epochs, for H2.2 – 46 epochs), where both training and validation accuracy differed in hundredths, which can be considered a marginal impact.

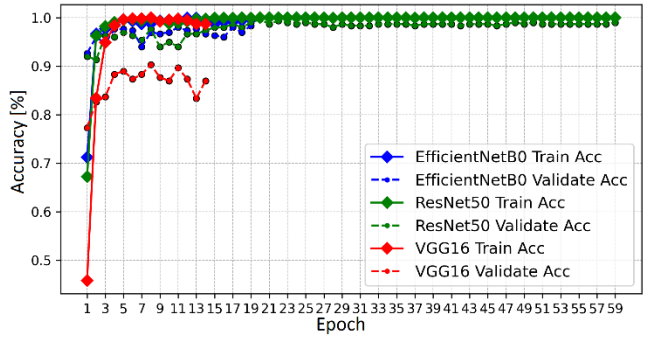

Fig. 4. Accuracy on the training and validation sets for models from scenario H2.1 – no data transformations applied

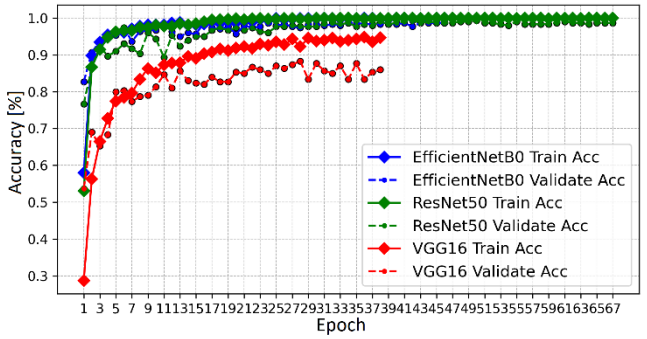

Fig. 5. Accuracy on the training and validation sets for models from scenario H2.2 – applied data transformations

The summary of the metrics obtained in the model evaluation process for scenario 2 is presented in table 2. All models showed an improvement in H2.2 relative to H2.1, with the VGG16 model showing the largest increase in metrics by two percentage points for sensitivity and F1 score, and by one percentage point for precision. The metric results obtained in the two experiments of scenario 2 show that the improvement in metrics is achievable by introducing modifications to the input images, however, the value of the increase depends on the selected model. The highest increase between experiments for the VGG16 model is not reflected in the position of the model ranking, due to the fact that this model obtained the lowest results compared to other selected architectures, achieving only 83% precision in H2.2. The EfficientNetB0 and ResNet50 models achieve identical results in each of the two experiments, achieving 97% for the three metrics in H2.1 and 98% for the three metrics in H2.2. We observe an improvement in H2.2 over H2.1 of one percentage point for all three metrics.

Table 2: Evaluation metrics from the final epoch of training for the EfficientNetB0, ResNet50 and VGG16 models for scenario 2

Model	Scenario	Precision	Recall	F1-Score
VGG16	H1.1	82.0%	79.0%	79.0%
	H1.2	83.0%	81.0%	81.0%
ResNet50	H1.1	97.0%	97.0%	97.0%
	H1.2	98.0%	98.0%	98.0%
EfficientNetB0	H1.1	97.0%	97.0%	97.0%
	H1.2	98.0%	98.0%	98.0%

Analysis of the confusion matrix (Fig. 6) shows that the model mostly correctly predicts the class membership predictions for most classes. There were sporadic, single errors, concerning 7 classes out of 20, where out of a total of 300 images the number of errors is only 7 images, which confirms the metrics in table 2. No patterns were detected in the incorrectly classified classes, and in total errors in individual classes occurred sporadically and are of an incidental nature. The overall picture of the matrix suggests high classification quality in the second scenario.

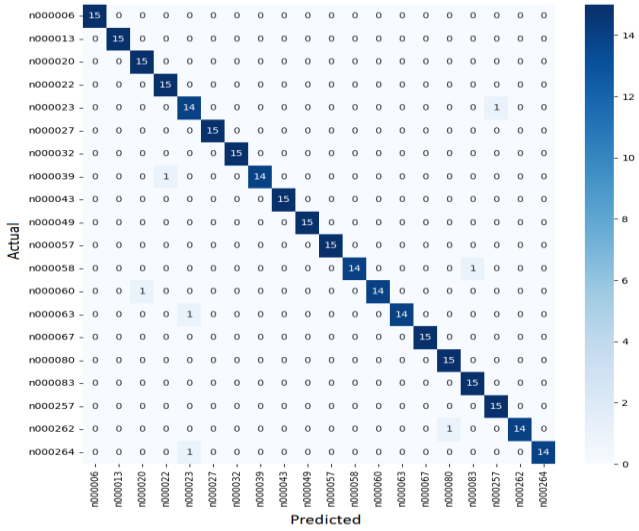

Fig. 6. Confusion matrix for the EfficientNetB0 model for study H2.2

Analysis of the results of scenario 2 suggests that the data modification is disadvantageous for all of the tested models. VGG16 benefited the most, but required a much larger number of epochs, which made the final effect disproportionate. ResNet50 did not show significant changes in the number of epochs or accuracy, although the training process stabilized faster in H2.1 (epoch 5) than in the H2.2 (epoch 17).

3.3. GradCAM evaluation

Figure 7 shows a mosaic of generated heat maps (Grad-CAM) for selected models. The superimposed heat maps indicate key areas during the model's decision-making. As can be seen, the Efficient-NetB0 and ResNet50 models most often focused on the eyes, nose, and forehead. In this case, the VGG16 model significantly differed from the others, focusing its attention on areas outside the face, especially in the hair area, with a certain exception in scenario H1.2, where the heat area was marked around the mouth and eye. The individual models differ in the areas of activation, which suggests that each of them shaped its interpretation of the image and biometric data in a slightly different way. The EfficientNetB0 and ResNet50 models definitely performed best in the task of identifying biometric data, in the areas where they focused their interest, which confirms

the high results obtained in the training process. In the case of the first scenario, a significant deterioration of the areas in the context of biometric data was noted, as a result of which the model's attention was shifted from the eye area (H1.1) to the forehead area (H1.2). This may indicate that the studied effect of higher resolution negatively affects the model's ability to capture key facial biometric features, and consequently worsens the quality of Grad-CAM interpretation in this scenario. In the second scenario, the analysis of heat maps confirms that the EfficientNetB0 and ResNet50 models still focused on key biometric features (eye area, nose). Data modifications in the form of changing brightness, introducing rotation and mirroring were less invasive compared to the first scenario, but their introduction did not negatively affect the interpretation of important image features.

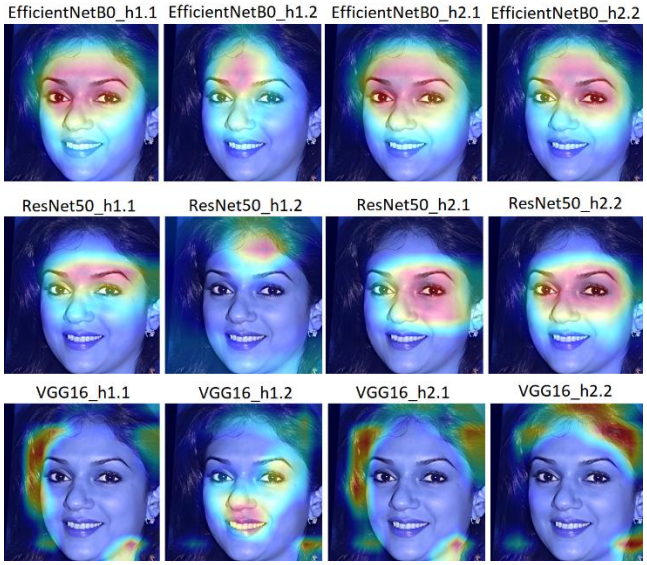

Fig. 7. Grad-CAM summary for selected models. The columns contain the individual research hypotheses described in Section 3, while the rows contain the results for the subsequent models

4. Conclusions and future works

The results presented in this paper indicate that both hypotheses formulated in Section 3.5 were confirmed. In the first scenario, an improvement in the prediction results of the models was observed in H1.2 containing images of 512×512 pixels compared to H1.1 using images of 224×224 pixels. In the second scenario, a minimal improvement in the metrics was also observed after introducing data modifications in H2.2 compared to the unmodified data in H2.1. The key factor in the choice between increasing the image size and/or introducing data modifications should be a decision that takes into account aspects such as network training time, available computational resources or the need to achieve the highest possible model prediction efficiency. The best results in the studied scenarios were obtained by the EfficientNetB0 model, which confirms the high capabilities offered by this modern architecture, followed by ResNet50 in the ranking, and the VGG16 model in the last place. Each model achieved high results exceeding 80% accuracy on both training and validation sets, however, the Grad-CAM analysis showed that the Efficient-NetB0 model performed best in the biometric data identification task (keeping the focus on the face region, especially the eyes), which was the aim of this work.

Future and robustness. Additionally, efforts will be directed toward developing dedicated solutions leveraging the capabilities of vision transformers for advanced analysis and feature extraction [6]. These approaches aim to combine the strengths of multiple methodologies to achieve more accurate and reliable outcomes.

References

[1] Cao Q. et al.: Vggface2: A dataset for recognising faces across pose and age. 13th IEEE international conference on automatic face & gesture recognition, 2018, 67–74.
[2] He K. et al.: Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition, 2016, 770–778 [https://doi.org/10.48550/arXiv.1512.03385].
[3] Hossin M., Sulaiman M. N.: A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process 5(2), 2015, 1 [http://dx.doi.org/10.5121/ijdkp.2015.5201].
[4] Jayaraman U. et al.: Recent development in face recognition. Neurocomputing, 408, 231–245.
[5] Kamyab T. et al.: Combination of Genetic Algorithm and Neural Network to Select Facial Features in Face Recognition Technique. International Journal of Robotics and Control Systems 3(1), 2023, 50–58.
[6] Karczmarek P. et al.: Quadrature-inspired generalized Choquet integral. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–7, 2022.
[7] Nandre J., Rai S., Kanawade B. R.: Comparative Analysis of Transfer Learning CNN for Face Recognition. 2nd International Conference on Intelligent Technologies (CONIT), 1–6, 2022 [https://doi.org/10.1109/CONIT55038.2022.9847946].
[8] Paul K. C., Aslan S.: An improved real-time face recognition system at low resolution based on local binary pattern histogram algorithm and CLAHE. 2021, arXiv preprint arXiv:2104.07234 [https://doi.org/10.48550/arXiv.2104.07234].
[9] Phillips P. J. et al.: Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. Proceedings of the National Academy of Sciences 115(24), 2018, 6171–6176.
[10] Powroźnik P., Wojcicki P., Przylucki S. W.: Scalogram as a representation of emotional speech. IEEE Access 9, 2021, 154044–154057 [https://doi.org/10.1109/ACCESS.2021.3127581].
[11] Schroff F., Kalenichenko D., Philbin J.: Facenet: A unified embedding for face recognition and clustering. IEEE conference on computer vision and pattern recognition, 2015, 815–823.
[12] Selvaraju R. R. et al.: Grad-CAM: Why did you say that?, arXiv preprint arXiv:1611.07450, 2016 [https://doi.org/10.48550/arXiv.1611.07450].
[13] Simonyan K., Zisserman A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 [https://doi.org/10.48550/arXiv.1409.1556].
[14] Song C., Ji S.: Face recognition method based on siamese networks under non-restricted conditions. IEEE Access 10, 2022, 40432–40444.
[15] Tan M., Le Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning, 2019, 6105–6114 [https://doi.org/10.48550/arXiv.1905.11946].
[16] Tharwat A.: Classification assessment methods. Applied computing and informatics 17(1), 2021, 168–192 [https://doi.org/10.1016/j.aci.2018.08.003].
[17] Wang X. et al.: A survey of face recognition. arXiv preprint arXiv:2212.13038, 2022.
[18] Wu H. et al.: Face recognition based on Haar like and Euclidean distance. Journal of Physics: Conference Series, 1813, 1, 012036.
[19] Zarei S., Andi T.: Face recognition methods analysis. International Journal Artificial Intelligent and Informatics 1(1), 2018, 1–12.

B.Eng. Marcin Amadeusz Rudzki
e-mail: s102913@pollub.edu.pl

He graduated an engineering degree at the Kielce University of Technology, majoring in computer science with a specialization in teleinformatics in 2019. The author's main interests areas of interest included computer vision, deep learning and artificial intelligence.

https://orcid.org/0009-0006-2528-008X

Ph.D. Paweł Powroźnik
e-mail: p.powroznik@pollub.pl

He is an assistant professor at the Department of Computer Science at the Lublin University of Technology, since 2016. He received the B.Sc. and M.Sc. degrees in Computer Science at the Maria Curie-Sklodowska University in Lublin, Poland in 2009 and 2011, respectively. In 2018 he defended his Ph.D. thesis at the Lublin University of Technology.
His main research interests are focused on machine learning, data classification, computer vision, analysis of the usability of graphical interfaces, and intangible cultural heritage as well as motion capture, 3D motion analysis. He participates in educational, as well as, scientific national and international grants.
He is a member of IEEE, Polish IT Society, Association for Image Processing, Lublin Scientific Society.

https://orcid.org/0000-0002-5705-4785