Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2024, № 1 (113) https://doi.org/10.33108/visnyk_tntu2024.01
ISSN 2522-4433. Web: ***visnyk.tntu.edu.ua***

UDC 004.852

COMPARATIVE ANALYSIS OF DATA AUGMENTATION METHODS FOR IMAGE MODALITY

Andrii Bokhonko; Nataliia Melnykova; Yurii Patereha

Lviv Polytechnic National University, Lviv, Ukraine

Summary. *The object of research is forecasting processes in the case of short sets of tabular data. The subject of research is the data augmentation method for images. Achieving the goal occurs primarily from the study of existing machine learning tools and data augmentation methods for images. Further software development to implement various data augmentation methods and machine learning models for images. Approbation of the work was carried out by analyzing the effectiveness of various methods of data augmentation for images using quality metrics and statistical methods. Due to the results of the research, an analysis of the influence of various methods of data augmentation on the effectiveness of classifiers in images was carried out.*

Key words: *data modality, classification, data augmentation, sample expansion, machine learning.*

https://doi.org/10.33108/visnyk_tntu2024.01.016 *Received 22.12.2023*

Problem statement and justification. In the modern world, the volumes of available data are growing each year, but in many cases, the quality and quantity of data may be insufficient for effective training of various machine learning and artificial intelligence models. Therefore, to enhance the productivity and accuracy of these models, there is a need to develop effective methods of data augmentation.

Data augmentation is the process of creating new data based on existing data, which helps to improve the representation of various aspects and characteristics of the data. The importance of data augmentation is high in various fields such as computer vision, speech recognition, and data analytics, and it can play a crucial role in building accurate and efficient models.

The data challenges are making the achievement of high accuracy and productivity in machine learning and artificial intelligence models impossible. Without adequate understanding and application of effective data augmentation methods for different modalities, we are very likely to face the following problems [1]:

1. Overfitting: models may become too specific to the training dataset and inadequately generalize their predictions to new data.
2. Insufficient amount of data: the lack of sufficient quantity and quality of data can lead to poor training results and render models unfit for solving practical tasks.
3. Time and computational costs: without using optimal data augmentation methods, model training can become costly in terms of both time and computational resources.

All these problems can lead to low-quality results, wasted efforts in model development, and inefficient utilization of computational resources. This underlines the importance of conducting research and developing data augmentation methods for various modalities that will facilitate solving the above-mentioned problems and achieving better results in the construction and application of machine learning and artificial intelligence models.

Analysis of existing research results. Based on the literature search on this topic, several articles directly related to the research theme have been analyzed. In the work [2], the

 Corresponding author: Andrii Bokhonko; e-mail: andrewboxonko@gmail.com

application of hyperparameter adjustment methodologies for data augmentation to improve the performance of convolutional neural networks in image classification was clearly described. Additionally, the substantial practical use of data augmentation to increase the volume of training data for deep artificial neural networks, especially convolutional neural networks, is discussed in the article [3]. The comparison and analysis of different data augmentation methods for image classification, from classical image transformations (rotation, cropping, zooming, histogram-based methods) to style transfer and generative adversarial networks (GANs), are also presented in the work [4].

Objective of the study. Comparative analysis of data augmentation methods for images. Within the scope of this research, various data augmentation methods for images will be analyzed. The research will be conducted using different algorithms and machine-learning models. Comparing the results of the research will help to determine the effectiveness of different data augmentation methods for images.

Formulation of research tasks. The research tasks can be formulated as follows:

1. Analyze scientific literature and identify key data augmentation methods for images.
2. Develop some criteria for evaluating the effectiveness of data augmentation methods, taking into account the specific features of model tasks and image modalities.
3. Conduct experiments with different data augmentation methods, using the specified criteria to assess their effectiveness on various datasets and tasks.
4. Analyze the results of experiments and identify the most effective data augmentation methods for images, considering their impact on the accuracy and generalization ability of models.

Justification of the chosen modality and data augmentation methods. Data augmentation (enlargement) is a technique used in machine learning, especially in the context of deep learning, to expand and enrich the training dataset by creating new data instances through various transformations applied to existing data. These transformations are designed to preserve the original class labels while introducing variations and diversity in the data, simulating real-world scenarios.

More in detail, data augmentation performs several tasks to help address the data scarcity problem [5]:

1. Increase the size of the dataset.
2. Enhance data diversity.
3. Solving class imbalance problems.
4. Implicit regularization.

Image data augmentation plays a crucial role in improving the performance of machine learning models, particularly in computer vision tasks such as image classification, object detection, and segmentation. By applying various transformations to original images, new training samples are created, making the dataset larger and more diverse. This helps the model learn more generalized functions and become more resistant to variations in real-world data. Some examples of image augmentation techniques are as follows [6]:

1. Random rotation: This technique involves randomly rotating images by a specified angle within a defined range.
2. Width and height shift: These methods involve random horizontal or vertical shifts of images by a certain fraction of the total width or height of the image. These techniques help the model recognize objects despite slight displacements.
3. Brightness adjustment: This method alters the brightness of images by randomly adjusting illumination levels within a specified range. This allows the model to learn features less dependent on lighting conditions.
4. Shear transformation: This operation distorts images by applying a random shear angle within a defined range. Shear transformation helps models recognize objects that may appear deformed or stretched.

5. Horizontal and vertical flipping: This technique involves randomly flipping images horizontally or vertically, creating new images with different perspectives and helping the model generalize various viewpoints.

Algorithm for image data classification and results evaluation metrics. Convolutional Neural Network (CNN) is a type of deep learning model specialized in processing grid-like data, such as images, using convolutional layers to capture local spatial patterns and abstractions. CNNs have become the basic model for numerous image classification tasks due to their ability to learn hierarchical functions, resilience to noise, and variations in input data.

The key layers of CNN [7] for image classification are as follows:

1. Input Layer. The input layer is responsible for receiving the raw image data. Each input image is typically represented as a three-dimensional tensor (width, height, and number of color channels - usually 3 for RGB images).

2. Convolutional Layers. These layers perform the convolution operation, applying a set of learnable filters to the input image or the output of previous layers. During convolution, the filter moves across the image, element-wise multiplying the filter with the local neighborhood and then summing the results to create a feature map. Each filter detects specific features or patterns in the input data, such as edges, lines, textures, or more complex structures at deeper levels of the network.

3. Non-linearity (Activation Functions): Activation functions introduce non-linearity to the network, allowing it to learn and approximate complex, non-linear functions. The most common activation function is the Rectified Linear Unit (ReLU), which retains positive values and sets negative values to zero.

4. Pooling Layers. Pooling layers reduce spatial dimensions and computational complexity, making the model more computationally efficient and invariant to changes. The most commonly used pooling technique is Max-Pooling, which takes the maximum value from a certain local neighborhood (usually 2x2) and moves across the image.

5. Fully Connected Layers. Fully connected layers are used to map the results from the convolutional and pooling layers to the final class score or probability distribution. In a typical architecture, there may be one or more fully connected layers, followed by the softmax activation function to output probabilities for each class.

6. Output Layer. The output layer produces the final class probabilities for image classification. For multi-class tasks, softmax activation is often used, while binary classification may use a sigmoid activation function. The class with the highest probability is selected as the final prediction of the network.

To assess the quality of predictions, five metrics have been chosen [8]:

1. Accuracy. Measures the proportion of correct predictions made by the model compared to the total number of samples. It provides an evaluation of how well the model can correctly classify instances from a given dataset.

2. Precision: evaluates how precise the model is in predicting positive values.

3. Recall: measures how many actual positive cases the model correctly predicted as positive.

4. F1-score. This evaluation metric is particularly effective when dealing with unbalanced datasets or when both false positives and false negatives are of concern.

5. Confusion matrix. This table reflects the classification model's performance by comparing its predicted labels with the actual labels (ground truth) in a cross-tabular format.

Research Results. Image Augmentation. The software implementation was carried out using the Google Colaboratory cloud environment and the Python programming language. Several auxiliary libraries were utilized for the task, including Numpy, Pandas, OpenCV, Scikit-learn, Keras, and Matplotlib.

For image augmentation, a dataset of chest X-ray images with pneumonia and without pneumonia was chosen. It consists of 5863 X-ray images (JPEG) and two categories (pneumonia/normal). Chest X-ray images were selected from a medical center in Guangzhou. All chest X-ray examinations were conducted as part of routine clinical care for patients. To analyze chest X-ray images, all chest radiographs were initially reviewed for quality control by removing any low-quality scans or unreadable images.

In Figures 1 to 2, examples of images from this dataset are presented with pneumonia and without it, respectively.

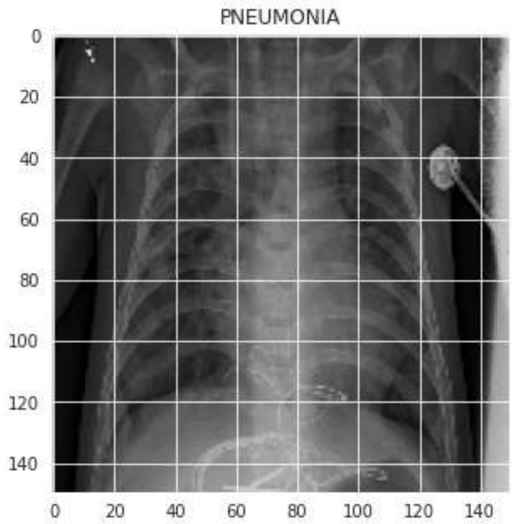

Figure 1. An example of a dataset with pneumonia

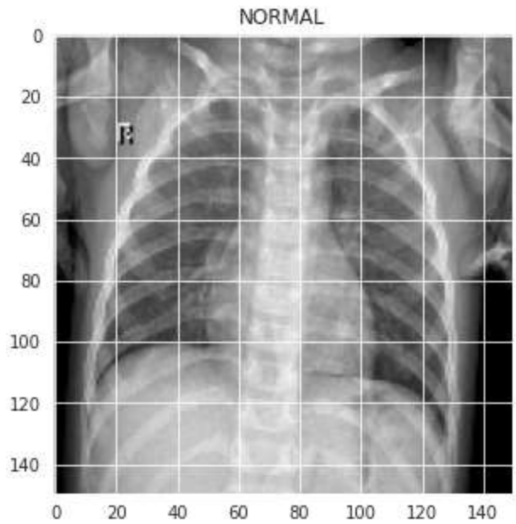

Figure 2. An example of a dataset without pneumonia

In Figure 3, the class balance in the image dataset is depicted. The class of images with the disease predominates over the healthy lungs. With the use of augmentation, it is necessary to balance the classes.

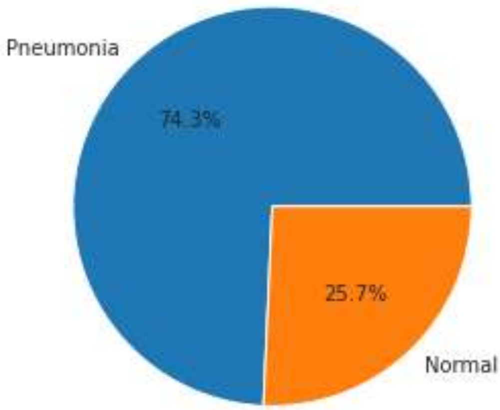

Figure 3. Class distribution in the image dataset

To balance the classes, various image augmentation techniques need to be employed. The ImageDataGenerator from the Keras library works by applying diverse transformations to images in the dataset, such as scaling, rotation, zooming, and flipping. ImageDataGenerator dynamically augments data during each epoch, providing batches to the neural network during training. This ensures that the model sees different variations of the same image in different epochs.

Let's examine how the data will look after augmentation using different combinations of parameters to increase the number of images:

- Combination of image rotations: In Figure 4, images augmented with rotations are presented. As shown, in this case, the images will be randomly rotated horizontally or vertically by a certain angle.

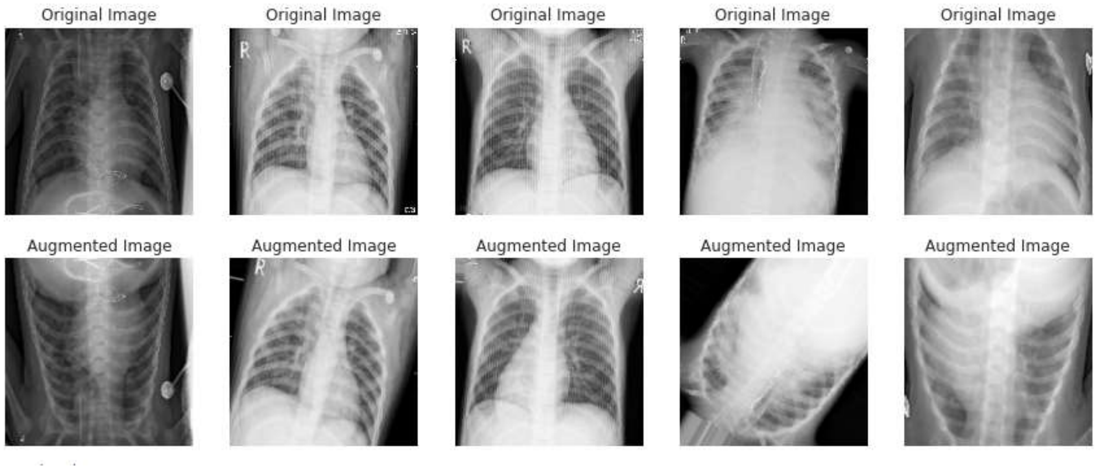

Figure 4. Images are augmented with rotations

- Combination of zooming and height or width shift of images: In Figure 5, images augmented with zooming and shifts are presented. In this case, the images will be randomly zoomed in or out. Additionally, they may be arbitrarily shifted horizontally or vertically by a certain fraction.

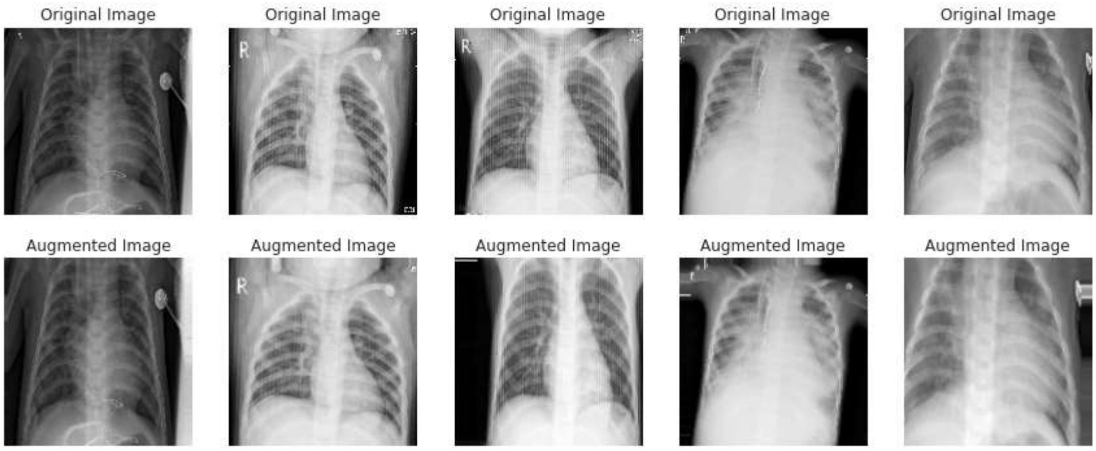

Figure 5. Images are augmented by zooming and shifting

- Combination of brightness adjustment: In Figure 6, images augmented with brightness adjustment are presented. In this case, the brightness of pixels in the images will be randomly altered.

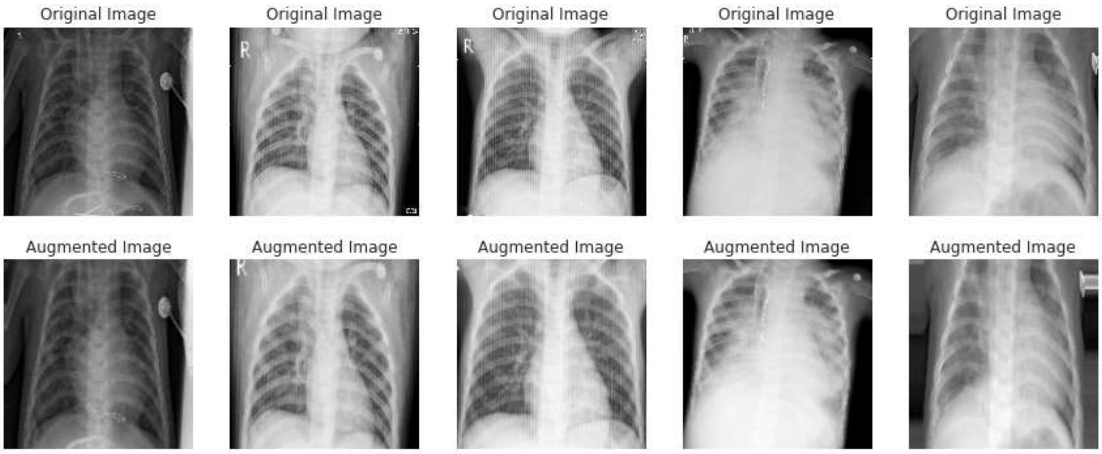

Figure 6. Images are augmented by changing the brightness

Quality Evaluation of Augmented Data. To detect pneumonia, a CNN network was employed. Initially, training and evaluation were conducted on the original set of images. For this training, 15 epochs were utilized with a callback function to reduce the learning rate.

In Figure 7, a classification report of predictions on the original image dataset is presented. Figure 8 displays the Confusion Matrix of predictions on the original image dataset.

	precision	recall	f1-score	support
Pneumonia (Class 0)	0.70	1.00	0.82	390
Normal (Class 1)	0.99	0.29	0.45	234
accuracy			0.73	624
macro avg	0.84	0.65	0.64	624
weighted avg	0.81	0.73	0.69	624

Figure 7. Classification report of predictions on the original image dataset

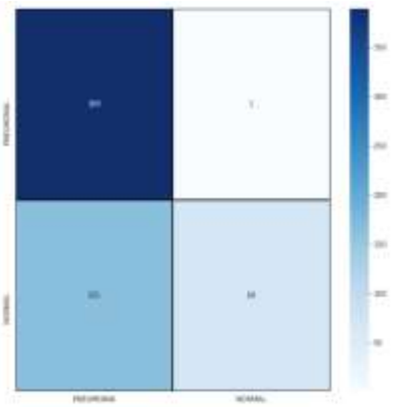

Figure 8. Confusion matrix of predictions on the original image dataset

Using this model, pneumonia prediction proved to be reasonably accurate (Precision = 0.70) and demonstrated high sensitivity in correctly identifying this condition (Recall = 1.00), resulting in a generalized F1-score of 0.82.

However, the model was less effective in identifying cases without pneumonia (normal condition – class 1), with precision at 99% but low sensitivity at only 29%, leading to an F1-score of 0.45.

The overall accuracy of the model is 73%, which, while higher than random chance, still leaves room for improvement.

Now, let's analyze the results of pneumonia prediction using the augmentation methods described above. The ImageDataGenerator from the Keras library was used as an augmentor. It dynamically identifies minority classes and balances the dataset during training. Additionally, it generates new synthetic samples for each class. The results obtained with augmentation are as follows.

Random scaling and shifts in height or width of images. On Fig. 9, the classification report of predictions on the augmented image dataset using shifts and scaling is presented. Fig. 10 shows the Confusion matrix of predictions on the augmented image dataset using shifts and scaling

	precision	recall	f1-score	support
Pneumonia (Class 0)	0.93	0.94	0.94	390
Normal (Class 1)	0.90	0.89	0.89	234
accuracy			0.92	624
macro avg	0.92	0.91	0.91	624
weighted avg	0.92	0.92	0.92	624

Figure 9. A classification report of predictions on an augmented image dataset using shifts and approximations

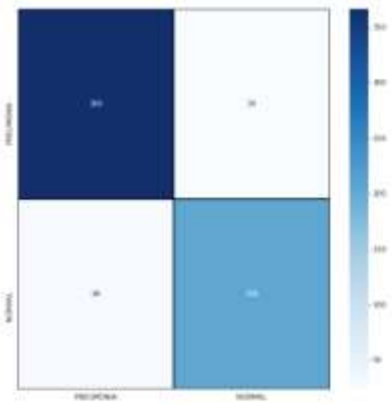

Figure 10. Confusion matrix predictions on an augmented image dataset using shifts and approximations

In comparison to the original dataset, the results of applying CNN to the dataset augmented with random shifts and scaling showed even better performance.

For the pneumonia class, precision increased to 93%, and recall reached 94%. This indicates that the model identified almost all pneumonia cases, and most predictions were correct. The overall f1-score, which is the harmonic mean between precision and recall, is 94%.

In the case of determining the normal state, accuracy increased to 90%, and sensitivity reached 89%. This means that the model more accurately recognizes the normal state of patients and makes fewer mistakes in diagnosing pneumonia. The f1-score for these results reached 89%.

The overall accuracy of the model increased to 92%, demonstrating a significant improvement in its effectiveness in determining the patients' condition.

Random Changes in Brightness. Figure 11 shows the classification report of predictions on the augmented dataset using changes in brightness. Figure 12 presents the Confusion Matrix of predictions on the augmented dataset using changes in brightness.

	precision	recall	f1-score	support
Pneumonia (Class 0)	0.62	0.99	0.77	390
Normal (Class 1)	0.00	0.00	0.00	234
accuracy			0.62	624
macro avg	0.31	0.50	0.38	624
weighted avg	0.39	0.62	0.48	624

Figure 11. A classification report of predictions on an augmented image dataset using luminance variation

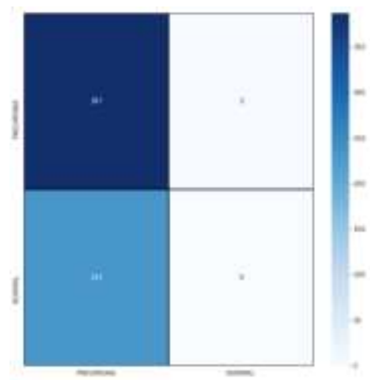

Figure 12. Confusion matrix predictions on an augmented dataset of images using changes in brightness

In comparison to the original dataset, the results of using CNN on the augmented dataset with changes in brightness are significantly worse.

The precision of pneumonia detection decreased to 62%, but recall increased and reached 99%. This means that the model has leaned towards extremes, predicting pneumonia in a large number of cases, including healthy patients. As a result, the f1-score for this class is 77%.

For the determination of the normal state, the model achieved zero values for precision, recall, and f1-score. This indicates that the model completely failed to recognize healthy patients, even though there were 234 of them in the dataset.

The overall accuracy of the model is 62%, and the overall f1-score is 38%.
Such a poor result of the model may be attributed to the fact that changes in brightness affect the contrast of images, which is a key factor in pneumonia recognition on chest X-rays. Thus, the use of brightness changes in this specific task likely distorted the features the model should have focused on, resulting in very weak performance.

Random Image Rotations. Figure 13 presents the classification report of predictions on the augmented dataset using random rotations. Figure 14 shows the Confusion Matrix of predictions on the augmented dataset using random rotations.

	precision	recall	f1-score	support
Pneumonia (Class 0)	0.88	0.93	0.90	390
Normal (Class 1)	0.87	0.78	0.82	234
accuracy			0.87	624
macro avg	0.87	0.86	0.86	624
weighted avg	0.87	0.87	0.87	624

Figure 13. A classification report of predictions on an augmented image dataset using rotations

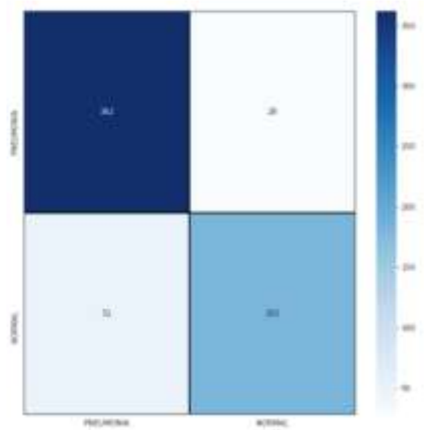

Figure 14. Confusion matrix predictions on an augmented image dataset using rotations

In comparison to the original dataset, CNN trained on the dataset augmented with rotations showed better results.

The precision in pneumonia detection significantly increased, reaching 88%. Sensitivity also increased to 93%, ensuring almost complete detection of all pneumonia cases. The f1-score compared to the previous results increased to 90%.

Regarding the determination of the normal state, the model also improved. Precision increased to 87%, and recall reached 78%. This means that the model became much less «pessimistic» and recognized a larger number of healthy patients. The harmonic mean of these two metrics, f1-score, now stands at 82%.

Overall, the accuracy of the model has significantly increased, reaching 87%. This indicates that the model was much more effective in recognizing the patients' condition, both in the case of pneumonia and in the normal state.

Comparison of results. The results obtained using augmentation were compared with the results obtained on the original dataset. The table reflects the impact of augmentation and the percentage improvement in metrics.

The calculation of the change in results will be carried out according to the formula:

*Variation = (m – n) * 100%*

where m – is the original dataset, n –is the augmented dataset.

The table visually demonstrates the percentage improvement (green color) or deterioration (red color) of the results:

Table 1

A metric variable on a sample of images

Metrics \ Augmentation	Random chance	Random changes in brightness	Random turns
Accuracy	19%	-11%	14%
Precision	8%	-53%	3%
Recall	26%	-15%	21%
F1-score	27%	-26%	22%

Using brightness changes results in significant deterioration. This is because X-ray images are highly sensitive to such alterations, and in some cases, we lose pneumonia data, or conversely, we might erroneously assume pneumonia is present in healthy images. However, augmentation using rotations and shifts proved to be very well-suited for this dataset, resulting in a significant improvement in prediction outcomes.

Conclusions. The conducted research involved an analysis of scientific literature on data augmentation methods for image data modality. The outcome identified the main challenges of data augmentation and outlined the necessary research directions to enhance the accuracy and generalization capabilities of machine learning and artificial intelligence models. Further analysis, and the development of criteria for evaluating augmentation methods, and experiments, the results of which will be used to determine the most optimal data augmentation approaches, laid the groundwork for future research.

The aspects of classification applied to this data modality were described. The classification challenges in the context of working with imbalanced data were highlighted, and the consequences of which can be mitigated by employing artificial data augmentation methods for underrepresented classes. Various classifiers, ranging from logistic regression to convolutional neural networks, can be used in classification depending on several factors. The importance of evaluating the quality of models using various performance metrics to determine their effectiveness in relevant scenarios was emphasized.

Experiments were conducted for the image data modality using three different augmentations: random shifts, random brightness changes, and random rotations. The experiments were carried out on the original dataset and the augmented dataset. A comparative analysis of prediction improvements helped understand both deteriorations and enhancements in the results.

References

1. A. Soofi i A. Awan Classification Techniques in Machine Learning: Applications and Issues. Journal of Basic & Applied Sciences. Vol. 13. 2017. P. 459–465. https://doi.org/10.6000/1927-5129.2017.13.76

2. A. L. C. Ottoni, R. M. de Amorim, M. S. Novo, і D. B. Costa Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. Intl. J. Mach. Learn. Cybern. Vol. 14. 2023. P. 171–186. https://doi.org/10.1007/s13042-022-01555-1
3. L. Taylor і G. Nitschke Improving deep learning with generic data augmentation, in 2018 IEEE symposium series on computational intelligence (SSCI). 2018. P. 1542–1547. https://doi.org/10.1109/SSCI.2018.8628742
4. A. Mikołajczyk і M. Grochowski Data augmentation for improving deep learning in image classification problem, in 2018 international interdisciplinary PhD workshop (IIPhDW). 2018. P. 117–122. https://doi.org/10.1109/IIPHDW.2018.8388338
5. D. Lewy і J. Mańdziuk An overview of mixing augmentation methods and augmentation strategies. Artif Intell Rev. 56 (3). 2023. P. 2111–2169. https://doi.org/10.1007/s10462-022-10227-z
6. C. Shorten і T. M. Khoshgoftaar A survey on Image Data Augmentation for Deep Learning. J. Big Data. 6 (1). 2019. https://doi.org/10.1186/s40537-019-0197-0
7. K. Dunphy, M. N. Fekri, K. Grolinger, і A. Sadhu Data Augmentation for Deep-Learning-Based Multiclass Structural Damage Detection Using Limited Information. Sensors. 22. (16), 2022. https://doi.org/10.3390/s22166193
8. M. Hossin і S. M.N A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process. Vol. 5. 2015. P. 1–11. https://doi.org/10.5121/ijdkp.2015.5201

Список використаних джерел

1. A. Soofi і A. Awan Classification Techniques in Machine Learning: Applications and Issues. Journal of Basic & Applied Sciences. https://doi.org/10.6000/1927-5129.2017.13.76
2. A. L. C. Ottoni, R. M. de Amorim, M. S. Novo, і D. B. Costa Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. Intl. J. Mach. Learn. Cybern. Vol. 14. 2023. P. 171–186. https://doi.org/10.1007/s13042-022-01555-1
3. L. Taylor і G. Nitschke Improving deep learning with generic data augmentation, in 2018 IEEE symposium series on computational intelligence (SSCI). 2018. P. 1542–1547. https://doi.org/10.1109/SSCI.2018.8628742
4. A. Mikołajczyk і M. Grochowski Data augmentation for improving deep learning in image classification problem, in 2018 international interdisciplinary PhD workshop (IIPhDW). 2018. P. 117–122. https://doi.org/10.1109/IIPHDW.2018.8388338
5. D. Lewy і J. Mańdziuk An overview of mixing augmentation methods and augmentation strategies. Artif Intell Rev. 56 (3). 2023. P. 2111–2169. https://doi.org/10.1007/s10462-022-10227-z
6. C. Shorten і T. M. Khoshgoftaar A survey on Image Data Augmentation for Deep Learning. J. Big Data. 6 (1). 2019. https://doi.org/10.1186/s40537-019-0197-0
7. K. Dunphy, M. N. Fekri, K. Grolinger, і A. Sadhu Data Augmentation for Deep-Learning-Based Multiclass Structural Damage Detection Using Limited Information. Sensors. 22. (16), 2022. https://doi.org/10.3390/s22166193
8. M. Hossin і S. M.N A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process. Vol. 5. 2015. P. 1–11. https://doi.org/10.5121/ijdkp.2015.5201

УДК 004.852

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ АУГМЕНТАЦІЇ ДАНИХ ДЛЯ МОДАЛЬНОСТІ ЗОБРАЖЕНЬ

Андрій Бохонко; Наталія Мельникова; Юрій Патерега

Національний університет «Львівська політехніка», Львів, Україна

Резюме. *В сучасному світі машинне навчання стало ключовим інструментом для аналізу та опрацювання даних у різних сферах – від медицини до транспорту. Однак, щоб досягти високої точності та надійності в прогнозуванні, машинне навчання потребує великої кількості даних. Проте у багатьох випадках наявність достатньої кількості даних є неможливою, що ускладнює розроблення ефективних моделей машинного навчання. Однією з важливих проблем у машинному навчанні є аналіз коротких наборів даних. У таких випадках, розроблення ефективної моделі машинного навчання може бути складним завданням, оскільки точність* *таких моделей зазвичай залежить від обсягу даних. Проте застосування методів аугментації даних може збільшити обсяг даних, необхідних для ефективного навчання моделей. Метою роботи є порівняльний аналіз методів аугментації даних для модальності зображень. У рамках роботи досліджено різні методи аугментації даних для зображень. Дослідження виконано за допомогою різних алгоритмів та моделей машинного навчання. Порівняння результатів досліджень дозволить визначити ефективність різних методів аугментації даних для модальності зображень. Об'єктом дослідження є процеси прогнозування у випадку коротких наборів табличних даних. Методами дослідження є алгоритми машинного навчання, які використовуються для роботи з модальностю зображення. Предметом дослідження є методи штучного доповнення даних для зображень. Досягнення мети відбувається, в першу чергу, за рахунок вивчення існуючих засобів машинного навчання та методів доповнення даних для зображень. Подальше розроблення програмного забезпечення для впровадження різних методів доповнення даних і моделей машинного навчання для зображень. Апробацію роботи проведено шляхом аналізу ефективності різних методів доповнення даних для зображень за допомогою метрик якості та статистичних методів. У результаті дослідження отримано порівняльний аналіз впливу різних методів доповнення даних на ефективність класифікаторів зображень із використанням різних метрик, таких, як accuracy, recall, precision, f1-score, confusion matrix.*

Ключові слова: *модальність даних, класифікація, аугментація даних, розширення вибірки, машинне навчання.*

https://doi.org/10.33108/visnyk_tntu2024.01.016 *Отримано 22.12.2023*