УДК 004.93
DOI: 10.25559/SITITO.019.202303.575-580

Оригинальная статья

# Применение моделей трансформеров для классификации рентгеновских снимков грудной клетки

**Э. Аррохо Эрнандес**
ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы», г. Москва, Российская Федерация
Адрес: 117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6
1142221454@pfur.ru

## Аннотация

В данной работе оценивается качество моделей визуальных трансформеров для решения задачи классификации рентгеновских снимков грудной клетки. Рентгеновские снимки грудной клетки являются наиболее известным и распространенным клиническим методом диагностики пневмонии. Однако диагностика пневмонии по рентгеновским снимкам грудной клетки является сложной задачей даже для опытных радиологов. Были проведены компьютерные эксперименты по применению переноса обучения для распознавания пневмонии на рентгеновских снимках грудной клетки. Для этого в качестве базовых моделей обучения были выбраны глубокие нейронные сети – трансформеры ViT, Swin и глубокие сверточные сети ResNet и VGG-16, предобученные на датасете ImageNet. Обучение моделей проводилось с функцией потерь CrossEntropyLoss и показателями точности accuracy, precision, recall, f1-score и AUC (Area Under Curve). После обучения выбиралась лучшая предварительно обученная модель на основе вышеуказанных метрик точности, полученных на тестовом наборе. В результате экспериментов наилучшую точность классификации показала модель Swin (Tiny) с показателями точности accuracy, precision и recall, равными 88, 89, 94% соответственно. После тонкой настройки показатели достигли значений 90, 94, 90, 92 и 90% соответственно.

**Ключевые слова:** рентгеновские снимки грудной клетки, пневмония, перенос обучения, искусственные нейронные сети, трансформеры

**Конфликт интересов:** автор заявляет об отсутствии конфликта интересов.

**Для цитирования:** Аррохо Эрнандес Э. Применение моделей трансформеров для классификации рентгеновских снимков грудной клетки // Современные информационные технологии и ИТ-образование. 2023. Т. 19, № 3. С. 575-580. https://doi.org/10.25559/SITITO.019.202303.575-580

© Э. Аррохо Эрнандес, 2023

Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

Original article

# Application of Transformer Models for Classification of Chest X-rays

**E. Arrokho Ernandes**
Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow, Russian Federation
Address: 6 Miklukho-Maklaya St., Moscow 117198, Russian Federation
aenoela@gmail.com

## Abstract

This article evaluates the quality of models of visual transformers for solving the problem of classification of chest X-rays. Chest X-raying is the most well-known and widespread clinical method of diagnosing pneumonia. However, the diagnosis of pneumonia by chest X-rays is a difficult task even for experienced radiologists. Computer experiments were conducted on the application of learning transfer for the recognition of pneumonia on chest X-rays. For this purpose, deep neural networks transformers ViT, Swin and deep convolutional networks ResNet and VGG-16, pre-trained on the ImageNet dataset, were selected as basic training models. The models were trained with the CrossEntropyLoss loss function and accuracy, precision, recall, f1-score and AUC (Area Under Curve) accuracy metrics. After training, the best pre-trained model was selected based on the above accuracy metrics obtained on the test set. As a result of the experiments, the best classification accuracy was shown by the Swin (Tiny) model with accuracy, precision and recall accuracy indicators equal to 88%, 89%, 94%, respectively. After fine-tuning, the metrics reached the values 90%, 94%, 90% respectively.

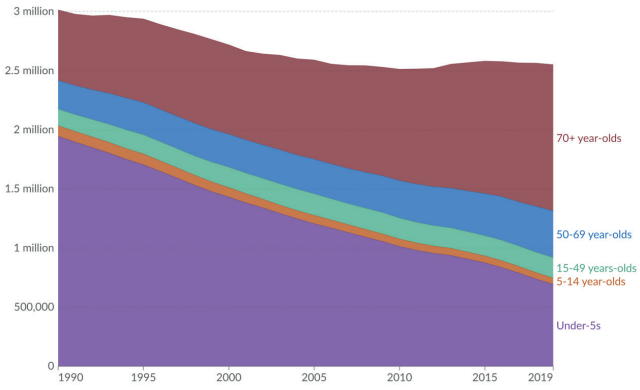
**Keywords:** chest x-rays; pneumonia; transfer learning; artificial neural networks, transformers

**Conflict of interests:** The author declares no conflict of interests.

**For citation:** Arrokho Ernandes E. Application of Transformer Models for Classification of Chest X-rays. *Modern Information Technologies and IT-Education.* 2023;19(3):575-580. https://doi.org/10.25559/SITITO.019.202303.575-580

## Введение

Пневмония – одна из наиболее распространённых острых респираторных инфекций, воздействующая на легкие. В 2019 году 2.5 миллиона человек умерли от пневмонии. Почти треть всех жертв составили дети младше 5 лет, это основная причина смертности детей в возрасте до 5 лет. Также высок уровень смертности среди пожилых людей в возрасте 70 лет и старше, 1.13 миллиона человек умерло от пневмонии в 2019 в этой возрастной группе[1] (рис. 1).




Р и с. 1. Смертность от пневмонии в мире с 1990 по 2019[2]
F i g. 1. Mortality from pneumonia in the world from 1990 to 2019[2]

Пневмонию можно лечить с помощью антибиотиков и противовирусных препаратов. Однако крайне важна ранняя диагностика для начала своевременного лечения для предотвращения осложнений, которые приводят к смерти. Рентгеновские снимки грудной клетки являются наиболее известным и распространенным клиническим методом диагностики пневмонии. Однако диагностика пневмонии по рентгеновским снимкам грудной клетки является сложной задачей даже для опытных радиологов. Внешний вид пневмонии на рентгеновских снимках часто неясен, ее можно спутать с другими заболеваниями, и она может вести себя как многие другие доброкачественные аномалии [1]. Таким образом, существует потребность в компьютерных системах поддержки, которые помогли бы рентгенологам диагностировать пневмонию по рентгеновским снимкам грудной клетки. Последние разработки в области глубокого обучения, основанные на визуальных трансформерах (vision transformers), показали большой успех в классификации изображений [2-5]. Целью данной работы является оценка качества моделей визуальных трансформеров с применением трансферного обучения в задаче классификации рентгеновских снимков грудной клетки [6].

## Визуальные трансформеры

Сверточные нейронные сети (CNN) уже много лет являются неотъемлемой частью исследований в области анализа медицинских изображений [7-8]. Несмотря на их выдающуюся производительность, CNN страдают концептуальными ограничениями и изначально неспособны моделировать явные зависимости на большом расстоянии из-за ограниченного поля восприятия ядер свертки [9, 10]. На основе достижений моделей трансформеров в задачах обработки естественного языка Досовицкий с коллегами предложили модель визуального трансформера (ViT) [11]. Она не содержит сверток и основана исключительно на механизме внимания (attention). Это позволяет избавиться от проблем сверхточных нейронных сетей. Однако для обучения трансформеров требуется очень большое количество данных.

## Перенос обучения в распознавании изображений

Трансфертное обучение (transfer learning) — это метод машинного обучения, при котором знания, полученные для решения одной задачи, повторно используются для решения смежных. Обучение глубоких нейронных сетей с нуля требует большого количества данных, больших вычислительных мощностей и много времени [12]. Поэтому зачастую используется предварительно обученные глубокие нейронные сети на очень большом наборе данных, таком как ImageNet, который содержит 1,2 миллиона изображений и 1000 классов [13-15]. Для того чтобы использовать предобученную нейронную сеть для решения требуемой задачи, требуется сначала зафиксировать веса всех ее слоев и заменить последний полносвязанный слой с функцией активации softmax, указав требуемое число классов [16]. Затем модель обучают на требуемом наборе данных, после чего модель готова для решения конкретно поставленной задачи.

Для повышения качества модели возможно проведение тонкой настройки модели (Fine tuning Model). В данном подходе после первого обучения все слои нейронной сети «размораживаются», после чего проводится обучение с уменьшенным значением шага обучения (learning rate).

## Описание набора исследуемых данных

В качестве базы данных для сравнительного анализа производительности исследуемых моделей глубокого обучения в работе использовались наборы рентгеновских снимков грудной клетки [17]. В наборе имеются 5856 изображений, 4273 из которых принадлежат классу Pneumonia (пневмония), а 1583 изображения — классу Normal (нет пневмонии) (рис. 2.1-2.2). Для использования трансферного обучения требуется привести изображения к тому формату, на котором нейронная сеть обучалась. Для сетей, предобученных на ImageNet, требуется:

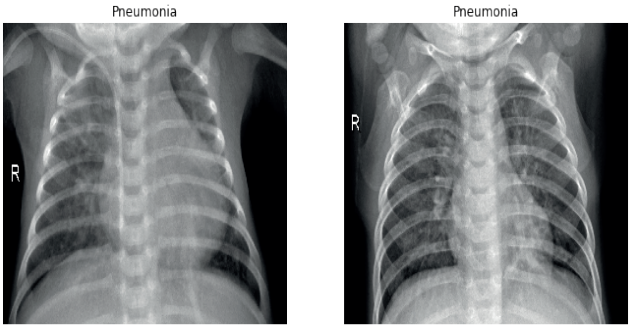
- размер изображений – 224x224x3 пикселя;
- нормализация всех пикселей.

Для повышения точности предсказания нейронных сетей используют метод искусственного увеличения данных (data augmentation). Благодаря ему возможно увеличить обучаю-

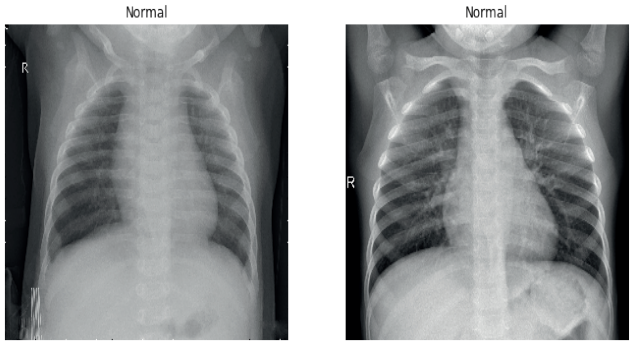
[1] Dadonaite B., Roser M. Pneumonia [Электронный ресурс] // Our World in Data, 2019. URL: https://ourworldindata.org/pneumonia (дата обращения: 11.08.2023).

[2] IHME, Global Burden of Disease (2023) – with minor processing by Our World in Data [Электронный ресурс]. URL: https://www.healthdata.org/research-analysis/gbd (дата обращения: 11.08.2023).

щую выборку в несколько раз. Для этого используют изображения, например, их случайным образом поворачивают или растягивают. Далее датасет разделяется на обучающую (train), валидационную и тестовую выборки.




Р и с. 2.1. Примеры изображений, принадлежащих классу Pneumonia [17]
F i g. 2.1. Examples of images belonging to the class "Pneumonia" [17]




Р и с. 2.2. Примеры изображений, принадлежащих классу Normal [17]
F i g. 2.2. Examples of images belonging to the class "Normal" [17]

## Компьютерные эксперименты по применению методов трансферного обучения в распознавании

Были проведены компьютерные эксперименты по применению переноса обучения для распознавания пневмонии на рентгеновских снимках грудной клетки. Для этого в качестве базовых моделей обучения были выбраны глубокие нейронные сети — трансформеры ViT, Swin [18] и глубокие сверточные сети ResNet [19] и VGG-16 [20], предобученные на датасете ImageNet [21-23]. Выбранные сети реализованы с использованием фреймворка PyTorch на языке программирования Python в операционной системе Ubuntu с графическом процессором Nvidia P100[3].

Обучение моделей проводилось с функцией потерь CrossEntropyLoss и показателями точности accuracy [24], precision, recall, f1-score и AUC (Area Under Curve). Формулы метрик приведены ниже:

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

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

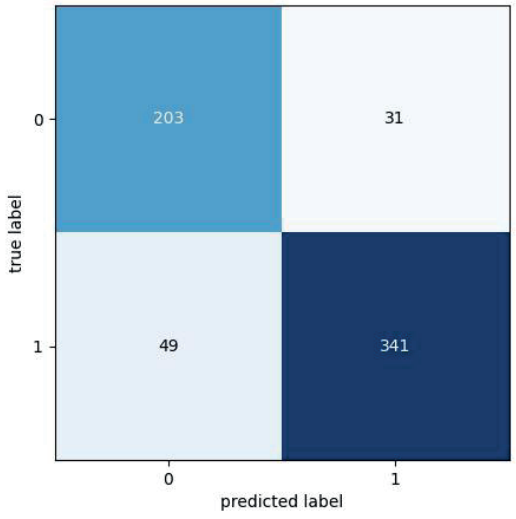
$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$F1\ \text{score} = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

где TP, TN, FN, FP — истинно положительные, истинно отрицательные, ложноотрицательные и ложноположительные результаты соответственно.

Модель обучалась в течение 15 эпох с начальной скоростью обучения 1e-3 и batch_size = 32. В качестве оптимизатора использовался Adam. После обучения производился fine tuning модели.

После обучения выбиралась лучшая предварительно обученная модель на основе вышеуказанных метрик точности, полученных на тестовом наборе [25]. В результате экспериментов наилучшую точность классификации показала модель Swin (Tiny) с показателями точности accuracy, precision и recall, равными 88, 89, 94 % соответственно (таблица 1). После тонкой настройки показатели достигли значений 90, 94, 90 % (таблица 2). На рисунке 3 приведена confusion matrix обученный модели.




Р и с. 3. Confusion matrix обученной модели Swin
F i g. 3. Confusion matrix of the trained Swin model

*Источник:* составлено автором.
*Source:* Compiled by the author.

Т а б л и ц а 1. **Показатели моделей**
T a b l e 1. **Indicators of models**

| | Swin_small | ViT_b_16 | ResNet18 | VGG-16 |
|---|---|---|---|---|
| **Accuracy** | **0.88** | 0.87 | 0.85 | 0.81 |
| **Precision** | 0.89 | **0.92** | 0.91 | 0.89 |
| **Recall** | **0.94** | 0.87 | 0.86 | 0.79 |
| **f1-score** | **0.91** | 0.90 | 0.88 | 0.84 |
| **AUC** | **0.87** | **0.87** | 0.86 | 0.82 |

*Источник:* здесь и далее в статье все таблицы составлены автором.
*Source:* Hereinafter in this article all tables were drawn up by the author.

[3] Chollet F. Deep Learning with Python. New York, NY : Manning, 2017. 384 p. URL: https://www.manning.com/books/deep-learning-with-python (дата обращения: 11.08.2023).

Таблица 2. **Показатели моделей с fine-tuning**
Table 2. **Indicators of models with fine-tuning**

| | Swin_small | Vit_b_16 | ResNet18 |
|---|---|---|---|
| **Accuracy** | **0,90** | 0,90 | 0,86 |
| **Precision** | **0,94** | **0,94** | 0,90 |
| **Recall** | **0,90** | 0,88 | 0,88 |
| **f1-score** | **0,92** | 0,91 | 0,89 |
| **AUC** | **0,90** | 0,89 | 0,85 |

## Заключение

В работе исследована эффективность методов переноса обучения глубоких нейронных сетей трансформеров, их тонкая настройка в контексте анализа медицинских изображений в качестве альтернативы сверточным нейронным сетям. Анализ проводился на наборе данных рентгеновских снимков грудной клетки, содержащих снимки больных пневмонией и изображения здорового человека. В качестве базовых моделей использовались такие модели, как Swin, ViT и ResNet.

Как показали результаты компьютерных экспериментов по классификации рентгеновских снимков грудной клетки, модель Swin показала себя лучше сверточных нейронных сетей ResNet и VGG-16. Таким образом, можно заключить, что модели визуальных трансформеров эффективны для решения задачи обнаружения пневмонии на рентгеновских снимках грудной клетки, осуществляя его лучше сверточных нейронных сетей.

## References

[1] Jain R., Gupta M., Taneja S., Hemanth D.J. Deep learning based detection and analysis of COVID-19 on chest X-ray images. *Applied Intelligence*. 2021;51(3):1690-1700. https://doi.org/10.1007/s10489-020-01902-1

[2] Kwon T., Lee S.P., Kim D., Jang J., Lee M., Kang S.U., et al. Diagnostic performance of artificial intelligence model for pneumonia from chest radiography. *PLoS ONE*. 2021;16(4):e0249399. https://doi.org/10.1371/journal.pone.0249399

[3] Krishnan Ko.S., Krishnan Ka.S. Vision Transformer based COVID-19 Detection using Chest X-rays. In: 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC). Solan, India: IEEE Computer Society; 2021. p. 644-648. https://doi.org/10.1109/ISPCC53510.2021.9609375

[4] Ma Y., Lv W. Identification of Pneumonia in Chest X-Ray Image Based on Transformer. *International Journal of Antennas and Propagation*. 2022;2022:5072666. https://doi.org/10.1155/2022/5072666

[5] Parveen Z., Adinarayana S., Aamani R., Santoshi S., Tulasi S. Efficient pneumonia detection in chest X-ray images using convolution neural network. *International Journal of All Research Education and Scientific Methods.* 2021;9(7):2900-2905. https://doi.org/10.1371/journal.pone.0256630

[6] Liang H., Fu W., Yi F. A Survey of Recent Advances in Transfer Learning. In: 2019 IEEE 19th International Conference on Communication Technology (ICCT). Xi'an, China: IEEE Computer Society; 2019. p. 1516-1523. https://doi.org/10.1109/ICCT46805.2019.8947072

[7] Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning. *Journal of Big Data*. 2019;6(1):1-48. https://doi.org/10.1186/s40537-019-0197-0

[8] Shin H.-C., Roth H.R., Gao M., Lu L., Xu Z., Nogues I., Yao J., Mollura D., Summers R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*. 2016;35(5):1285-1298. https://doi.org/10.1109/TMI.2016.2528162

[9] Shamshad F., Khan S., Zamir S.W., Khan M.H., Hayat M., Khan F.S., Fu H. Transformers in medical imaging: A survey. *Med Image Analysis*. 2023;88:102802. https://doi.org/10.1016/j.media.2023.102802

[10] Khoiriyah S.A., Basofi A., Fariza A. Convolutional Neural Network for Automatic Pneumonia Detection in Chest Radiography. In: 2020 International Electronics Symposium (IES). Surabaya, Indonesia: IEEE Computer Society; 2020. p. 476-480. https://doi.org/10.1109/IES50839.2020.9231540

[11] Dosovitskiy A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv:*2010.11929v2. https://doi.org/10.48550/arXiv.2010.11929

[12] Rajpurkar P., Irvin J., Zhu K., et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv*:1711.05225v3. https://doi.org/10.48550/arXiv.1711.05225

[13] Ayan E., Ünver H.M. Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning. In: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). Istanbul, Turkey: IEEE Computer Society; 2019. p. 1-5. https://doi.org/10.1109/EBBT.2019.8741582

[14] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*:1409.1556v6. https://doi.org/10.48550/arXiv.1409.1556

[15] Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017;60(6):84-90. https://doi.org/10.1145/3065386

[16] Shchetinin E.Yu., Sevastianov L.A. On transfer learning methods in biomedical images classification tasks. *Informatics and Applications*. 2021;15(4):59-64. (In Russ., abstract in Eng.) https://doi.org/10.14357/19922264210408

[17] Kermany D., Zhang K., Goldbaum M. Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data*. 2018;2. https://doi.org/10.17632/rscbjbr9sj.2
[18] Liu Z., et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE Computer Society; 2021. p. 10012-10022. https://doi.org/10.48550/arXiv.2103.14030
[19] Giełczyk A., et al. Pre-processing methods in chest X-ray image classification. *PLoS ONE*. 2022;17(4):e0265949. https://doi.org/10.1371/journal.pone.0265949
[20] Shelke A., et al. Chest X-ray classification using deep learning for automated COVID-19 screening. *SN computer science.* 2021;2(4):300. https://doi.org/10.1007/s42979-021-00695-5
[21] Showkat S., Qureshi S. Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia. *Chemometrics and Intelligent Laboratory Systems*. 2022;224:104534. https://doi.org/10.1016/j.chemolab.2022.104534
[22] Xia K., Wang J. Recent advances of transformers in medical image analysis: a comprehensive review. *MedComm – Future Medicine*. 2023;2(1):e38. https://doi.org/10.1002/mef2.38
[23] Zhou H.-Y., et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nature Biomedical Engineering*. 2023;7:743-755. https://doi.org/10.1038/s41551-023-01045-x
[24] He K., Zhang X., Ren Sh., Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE Computer Society; 2016. p. 770-778. https://doi.org/10.1109/CVPR.2016.90
[25] Thrun S., Pratt L. Learning to Learn. Berlin: Springer Science and Business Media Publ.; 2012. 354 p. https://doi.org/10.1007/978-1-4615-5529-2

*Поступила 11.08.2023; одобрена после рецензирования 28.09.2023; принята к публикации 02.10.2023.*
*Submitted 11.08.2023; approved after reviewing 28.09.2023; accepted for publication 02.10.2023.*

Об авторе:

**Аррохо Эрнандес Эноэль**, аспирант кафедры математического моделирования и искусственного интеллекта факультета физико-математических и естественных наук, ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6), **ORCID: https://orcid.org/0009-0006-0301-264X**, 1142221454@pfur.ru

*Автор прочитал и одобрил окончательный вариант рукописи.*

About the author:

**Enoel Arrokho Ernandes**, Postgraduate Student of the Department of Mathematical Modeling and Artificial Intelligence, Faculty of Science, Peoples' Friendship University of Russia named after Patrice Lumumba (6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation), **ORCID: https://orcid.org/0009-0006-0301-264X**, 1142221454@pfur.ru

*The author has read and approved the final manuscript.*