UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU

Collegium Medicum
im. Ludwika Rydygiera w Bydgoszczy

Bydgoszcz 2023

UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU
Wydział Lekarski
Collegium Medicum w Bydgoszczy

Anna Kloska

Use of Artificial Intelligence Methods for Classification of X-Ray Images of Patients with Lung Diseases

Rozprawa na stopień doktora nauk medycznych

Promotor:

Dr hab. n. med. Marcin Woźniak, prof. UMK

Drugi promotor:

Prof. dr hab. Zbigniew Serafin

Promotor pomocniczy:

Dr inż. Agata Giełczyk

Bydgoszcz 2023

„Nikt prawie nie wie,

dokąd go zaprowadzi droga,

póki nie stanie u celu.”

J.R.R. Tolkien

Table of contents

1. List of scientific papers included in the dissertation 6

  1.1. Pre-processing methods in chest X-ray image classification 6

  1.2. A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images 6

  1.3. Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification 6

2. List of abbreviations 7

3. Introduction 8

  3.1. AI as a diagnostic decision support tool in radiography 9

  3.2. Strengths and weaknesses of the application of AI methods in diagnostic imaging 11

  3.3. Description of the methods and algorithms used 12

    3.3.1. Data pre-processing 15

    3.3.2. Data augmentation 15

    3.3.3. Bypassing the limitations of the most commonly used algorithms 16

  3.4. Evaluation metrics 18

  3.5. Ethical issues of using AI for diagnostic purposes 21

    3.5.1. Medical data security 23

    3.5.2 Interpretability and explainability of AI models 24

4. Study aims 26

5. Summary of works included in the series of publications 27

  5.1. Original paper I – Pre-processing methods in chest X-ray image classification 27

  5.2. Original paper II – A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images 29

  5.3. Original paper III – Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification 32

6. Publications that are the subject of the dissertation 34

  6.1. Original paper I – content of the publication “Pre-processing methods in chest X-ray image classification” 34

  6.2. Original paper II – content of the publication “A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images” 45

  6.3. Original paper III – content of the publication “Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification” 55

7. Conclusions 66

8. References 68

9. Statements of co-authors of publications included in the series 75

9.1. Attachment No. 1 75
9.2. Attachment No. 2 77
9.3. Attachment No. 3 79
9.4. Attachment No. 4 81
9.5. Attachment No. 5 83
9.6. Attachment No. 6 85
9.7. Attachment No. 7 87
9.8. Attachment No. 8 89
10. Consent of the bioethics committee 91
Streszczenie 93
Summary 94

1. List of scientific papers included in the dissertation

This dissertation includes three original papers published in peer-reviewed journals included in the ministerial list of scientific journals. The total Impact Factor of the publications constituting the dissertation is 9.046 and 310 Ministry of Education and Science (Ministerstwo Edukacji i Nauki, MEiN) points.

1.1. Pre-processing methods in chest X-ray image classification

Agata Giełczyk, Anna Marciniak*, Martyna Tarczewska, Zbigniew Lutowski. PLOS One, 17(4): e0265949. Published: 5 April, 2022, Pages 1-11, https://doi.org/10.1371/journal.pone.0265949

IF = 3.752

MEiN = 100

1.2. A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images

Agata Giełczyk, Anna Marciniak*, Martyna Tarczewska, Sylwester Michal Kloska, Alicja Harmoza, Zbigniew Serafin, Marcin Woźniak. Journal of Clinical Medicine 2022 Sep 20;11(19):5501. https://doi.org/10.3390/jcm11195501

IF = 4.964

MEiN = 140

1.3. Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification

Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester M. Kloska, Adrian Michalski, Zbigniew Serafin, Marcin Woźniak. Polish Journal of Radiology. https://doi.org/10.5114/pjr.2023.126717

MEiN = 70

2. List of abbreviations

AI – artificial intelligence

CNN – convolutional neural network

CT – computed tomography

DBHL – Deep Boosted Hybrid Learning

DHL – Deep Hybrid Learning

FN – false negative

FP – false positive

ML – machine learning

MRI – magnetic resonance imaging

PCR – polymerase chain reaction

ROC – receiver operating characteristic

RNA – ribonucleic acid

RT-PCR – reverse transcriptase-polymerase chain reaction

SGD – stochastic gradient descent

TN – true negative

TP – true positive

xAI – explainable AI

3. Introduction

X-ray imaging has been widely used in clinical practice as a non-invasive and cost-effective tool for detecting various lung diseases, such as pneumonia (1–3), tuberculosis (4), lung cancer (5), and due to recent pandemic, COVID-19 (6). However, the interpretation of chest X-ray images can be challenging due to various factors, like workload (7), lowered satisfaction among professionals (8,9), but also hidden defects, which can go unnoticed even by experienced specialists. Another factor that affects and limits the accuracy of image analysis by specialists is fatigue (10,11).

One potential solution to these issues is the use of artificial intelligence (AI) methods and machine learning (ML) algorithms to assist in the analysis of chest X-ray images. AI-powered algorithms can provide several benefits, including the ability to operate continuously without fatigue, detect subtle differences that may not be visible to the human eye, and automate the analysis process (12,13).

There are several ways in which AI can help to improve the diagnostic capabilities of chest X-ray image analysis. AI algorithms can be trained on large datasets of chest X-ray images to learn the patterns and characteristics associated with specific lung diseases. Afterwards the training, these algorithms can be used to automatically analyze new, previously unseen, X-ray images and provide a diagnosis suggestion or a risk factor score that indicates probability of a particular disease or excludes it. Additionally, AI can help to reduce the workload of human radiologists by performing automated image triage, prioritizing suspicious cases for further review (14,15).

However, there are some challenges that need to be addressed to make AI-powered chest X-ray image analysis a reliable tool for clinical practice. One significant challenge is the lack of high-quality labeled datasets that accurately represent the diverse range of lung diseases and imaging variations. There are also important ethical issues that should be addressed, such as who is responsible for a misdiagnosis based on an ML model score. Moreover, there is a need to ensure the transparency and interpretability of AI algorithms to build trust among healthcare professionals and patients (16).

In this dissertation the use of AI-powered algorithms in chest X-ray image analysis for COVID-19 diagnostic purposes will be described. AI holds great promise in improving the accuracy and efficiency of lung disease diagnosis. However, it is crucial to address the technical and ethical challenges to ensure that these methods are safe, reliable, and beneficial to patients.

3.1. AI as a diagnostic decision support tool in radiography

Throughout the history of radiography, the profession has consistently adapted to new technologies and advancements, leading to changes in radiographic practice. While these changes have typically been driven by the adoption of new imaging technologies, recent technological advances have instead focused on the integration of complex ML algorithms and AI systems into equipment operation and image review processes Diagnostic imaging techniques such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound have revolutionized modern medicine by providing non-invasive means for visualizing internal organs and tissues. However, the interpretation of medical images can be complex and challenging, requiring highly trained experts with a deep understanding of the human anatomy and pathology. As a result, the impact of these technologies on radiography practice has yet to be fully examined. (14). AI algorithms can be used as powerful tools to support diagnostic imaging.

AI is a comprehensive term that encompasses the study and creation of computer systems capable of executing tasks that typically necessitate human intelligence, such as decision-making, prediction, visual perception, and speech recognition. It is a data-dependent approach that aligns with the technology-focused practice of current medical imaging, especially in computer vision tasks. Recently, there has been a substantial increase in AI-based proposals for diagnostic imaging (17) in both academic and industrial settings, with most emphasis on enhancing and supporting the analysis of radiological tests results. Accordingly, there is a developing area of AI applications that can be directly implemented in radiography practice (18,19). AI algorithms have emerged as a promising solution to the challenges of medical image analysis (20). Especially, the use of AI in the analysis of X-ray images for the diagnosis of lung diseases has gained significant attention in recent years (1,4,6). These algorithms are trained on large datasets of labeled medical images to learn the patterns and features associated with specific diseases or conditions. Once trained, AI algorithms can be used to analyze new images, providing automated diagnosis, or assisting human experts in the interpretation of complex images.

There are several advantages to using AI algorithms in medical imaging analysis. Firstly, AI algorithms can process large volumes of images quickly and accurately, potentially reducing the time and costs associated with manual analysis. Secondly, AI algorithms can detect subtle differences and patterns that may be difficult to discern by human experts, improving the accuracy and consistency of diagnoses. Thirdly, AI algorithms can reduce the workload of human experts, enabling them to focus on more complex cases or other tasks. In addition, it has been reported in a number of cases that AI algorithms can have efficiency comparable to or even better than experienced professionals (21–26).

However, the use of AI algorithms in medical imaging analysis also presents some challenges. AI has the potential to greatly aid radiographers in their work, but its full implementation could also significantly diminish their roles and responsibilities. As a result, healthcare organizations must be careful in exploring the use of AI technologies to improve radiology departments throughput and efficiency, and must fully understand and manage associated risks and liabilities. Current regulations require strict human oversight and auditing of AI solutions that are deployed clinically, which means that vendors must develop systems that require human oversight (27). Earlier research has demonstrated that individuals have a tendency to heavily rely on advice they receive and may even struggle to disregard inaccurate advice. This phenomenon has been observed in both clinical settings among medical professionals (28,29) and in other contexts involving decision-making. Considering that physicians can be influenced by advice, it is important to explore the most effective ways of presenting clinical advice to enhance its impact.

The introduction of ML systems presents a new challenge for the radiography profession, as they must now become adept at interacting with and supervising semi-automated processes driven by AI. Another significant challenge is the lack of high-quality labeled datasets that accurately represent the diversity of diseases and imaging variations (13). Additionally, there is a need to ensure the transparency and interpretability of AI algorithms to build trust among healthcare professionals and patients. Furthermore, there are concerns regarding the potential biases and ethical implications of using AI algorithms in medical decision-making (30).

To summarize, AI algorithms have the potential to revolutionize medical imaging analysis by providing fast, accurate, and consistent analysis of medical images. However, it is crucial to address the technical and ethical challenges associated with the use of AI algorithms in medical decision-making to ensure that these methods are safe, reliable, and beneficial to patients. Scientists and medical practitioners must continue to work towards developing AI algorithms that can be used as tools to support diagnostic imaging while also ensuring that they are transparent, ethical, and trustworthy.

3.2. Strengths and weaknesses of the application of AI methods in diagnostic imaging

It is important to consider both the advantages and disadvantages of using AI to analyze X-ray images in the diagnosis of lung diseases.

The use of AI in X-ray image analysis offers several potential benefits. One of the most significant advantages is the ability to analyze large volumes of data quickly and accurately. This is particularly important in the field of medical imaging, where there is often a need to process large amounts of data in a timely manner. AI algorithms can process this data much faster than human experts, thus reducing the time and costs associated with manual analysis. Furthermore, the use of AI can provide consistent and unbiased analysis. Human experts are subject to error and bias, which can affect the accuracy of their diagnoses. AI algorithms can help eliminate errors and reduce the risk of misdiagnosis, which can have significant consequences for patients' health outcomes.

Another benefit of using AI in X-ray image analysis is the potential to improve diagnostic accuracy. With the ability to process vast amounts of data, AI algorithms can detect subtle differences in images that may be missed by human experts. This can help improve the accuracy of diagnoses, potentially leading to earlier detection and treatment of lung diseases.

However, the use of AI in X-ray image analysis also presents significant ethical challenges (30) and technical challenge related to the creation of a good quality dataset. One concern is the potential for algorithmic bias. If the dataset used to train the algorithm is not representative, the algorithm may produce biased results. AI algorithms are only as unbiased as the data they are trained on. Bias may appear on the stage of training dataset preparation and labeling. Labeling the image data used in training datasets is extremely crucial to obtain a reliable model. This is a very difficult, time-consuming step because expert opinion on a given image can vary (31).

One important concern is that the data utilized for training algorithms may not be an accurate representation of the population for which the algorithm will be used. This is due to the likelihood that the available data used for training may have an over-representation of positive findings. Clinical data that yields positive research findings may be more readily available, whereas data from negative studies are often under-reported. As a result, AI tools may over-interpret the incidence of disease, leading to over-fitting. Additionally, selection bias can arise if training data are derived from a sub-population that does not accurately represent the entire population. This is likely to happen if the data is obtained only from a patient cohort attending a specialized center, which can result in bias against specific groups based on age, gender, ethnic origin, height, or other characteristics (32). Such bias can lead to the under-reporting or over-reporting of disease in patients on whose data AI used (33). This could lead to misdiagnosis and therefore, lead to delayed treatment, which could harm patients.

To date, there is no general consensus on who should bear responsibility for medical and diagnostic decision-making when AI is involved. While doctors are ultimately responsible for making medical decisions, the use of AI algorithms complicates the picture, as the authors of the algorithms are usually not medically trained and may not fully understand the implications of their algorithms on medical decision-making. This raises important ethical issues related to accountability and transparency in medical decision-making. Medical decisions made using AI algorithms must be transparent, and the process by which the algorithm makes its recommendations must be understandable to both medical professionals and patients. Furthermore, there must be clear guidelines for who bears responsibility in cases where medical decisions are made using AI algorithms.

In conclusion, while the use of AI in diagnostic imaging holds significant promise in improving the accuracy and efficiency of medical diagnoses, it also raises important ethical issues related to responsibility and accountability.

3.3. Description of the methods and algorithms used

Chest X-rays have been widely used in the diagnosis and management of respiratory diseases, including COVID-19 (34–36). The diagnosis of COVID-19 through chest X-ray images is a crucial task for timely treatment and isolation of patients (26). With the advent of AI and ML algorithms, computer-aided diagnosis has become a promising approach to improve the accuracy and efficiency of chest X-ray analysis (Fig. 1).

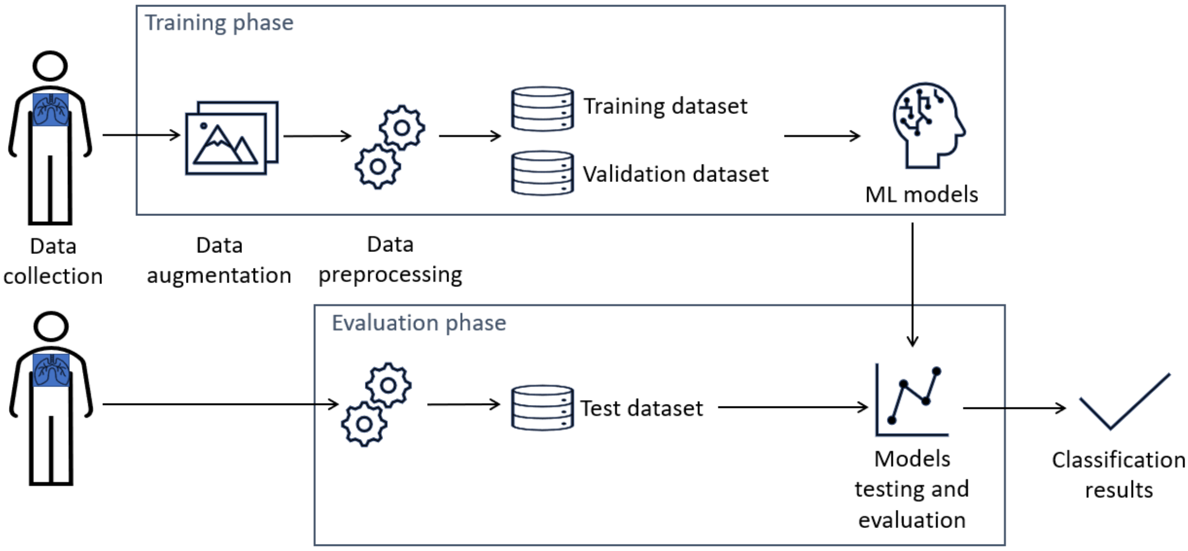

Figure 1 – Block diagram of the overall system.

Various ML techniques have been utilized to classify chest X-ray images and detect the presence of COVID-19. The performance of several popular ML algorithms in detecting COVID-19 from chest X-ray images was compared and analyzed in my research (37). These algorithms were convolutional neural networks (CNNs), ResNet-18, ResNet-34, XGBoost, and LightGBM. Each of these methods has its advantages and disadvantages.

CNN is a deep learning algorithm that is commonly used for image classification and recognition tasks due to its high accuracy and efficiency (38). However, CNNs require a large amount of data for training, which can be a challenge in medical diagnosis due to the limited availability of annotated medical images. CNNs have multiple layers, each of which performs a specific function. The first layer detects edges and curves in the image, while subsequent layers detect more complex patterns by combining features learned in the previous layers. In the case of COVID-19 diagnosis, CNN can be trained on a dataset of chest X-ray images, learning to distinguish between normal and abnormal images (39,40).

ResNet-18 and ResNet-34 are variants of CNNs that use residual connections to improve the model's training efficiency and accuracy and reduce the risk of model's overfitting (matching of the model to the training data resulting in too week generalization ability). These models are effective in medical image classification tasks due to their ability to extract relevant features from the images. However, they can be computationally expensive, which may limit their applicability in real-time diagnosis. Both, ResNet-18, and ResNet-34 are a CNN architecture that consists of 18 and 34 layers, respectively. This type of CNN is a modified version of the original ResNet, which stands for residual network. ResNet uses skip connections, which allow the network to learn residual functions. This helps to overcome the problem of vanishing gradients, which can occur when training deep neural networks. Both ResNets were proven to be used with high accuracy rates (41,42).

XGBoost and LightGBM are gradient boosting algorithms that are commonly used in tabular data analysis, classification and regression tasks, and have been successfully applied in medical image analysis (43–45). Gradient boosting is an ensemble learning method that combines multiple weak learners to create a strong learner. XGBoost and LightGBM both use decision trees as the base learner and perform gradient boosting to improve the accuracy of the model. They have shown promising results in COVID-19 diagnosis based on medical image analysis (46–49). These models can handle missing data and outliers, making them suitable for noisy medical datasets. However, their interpretability can be limited, which may be a concern in medical diagnosis where the decision-making process needs to be transparent.

These methods rely on the extraction of relevant features from the chest X-ray images, which are then used to train the AI/ML models. The models learn to recognize patterns in the images that are associated with COVID-19 infection, such as the presence of ground-glass opacities and consolidation in the lungs. All of these methods have been used successfully for COVID-19 diagnosis using chest X-ray images.

Overall, each of the methods described above has its strengths and weaknesses. CNNs are accurate but require large amounts of training data, ResNet models are effective in feature extraction but can be computationally expensive, and XGBoost and LightGBM are suitable for noisy datasets but may lack interpretability. Therefore, the choice of the method should depend on the specific requirements and characteristics of the medical dataset.

In conclusion, there is no single best method for COVID-19 diagnosis using chest X-ray images. The choice of the method should depend on the specific characteristics and requirements of the dataset. However, the combination of multiple methods, such as CNNs and ResNet models, may improve the accuracy and efficiency of the diagnosis process. In conducted research techniques such as data pre-processing and data were also used. The aim of their use was to check whether these methods can improve model performance.

3.3.1. Data pre-processing

Data pre-processing is a critical step in the ML pipeline that involves transforming raw data into a format suitable for use in ML algorithms. It is uncommon to have an image segmentation model that is trained on a dataset that is perfectly sized and labeled, particularly in the case of medical imaging applications, where acquiring both data and annotations can be expensive. Medical image segmentation datasets are often limited by a shortage of annotations, where there is only a small amount of annotated data available for training, or by weak annotations, where the training data has only sparse, noisy, or image-level annotations (50). In such cases, even the most advanced segmentation models may fail to generalize to datasets from real-world clinical settings. Therefore, data pre-processing typically involves a series of operations that clean, transform, and normalize the data, in order to make it more useful for training ML models.

Data pre-processing is a crucial step in image classification as it can greatly impact the performance of machine learning models. In order to improve the quality of data, pre-processing techniques such as normalization, resizing, and data augmentation are applied. Normalization involves transforming the pixel values of the image to have a mean of zero and unit variance, which helps to reduce the effects of varying lighting conditions and makes the training process more stable. Resizing involves scaling the images to a fixed size, which can help to reduce computational costs and also improve the accuracy of the model. Data augmentation involves generating additional training data by applying transformations such as rotations, flips, and zooms to the original images. This not only increases the size of the training dataset but also helps the model to generalize better to unseen data. Overall, proper pre-processing of image data can lead to more accurate and efficient machine learning models for image classification (51). Proper data pre-processing can help to improve the accuracy and robustness of the ML models and can help to avoid issues such as overfitting and poor generalization to new data (52).

3.3.2. Data augmentation

Data augmentation is a widely used technique in ML that involves generating additional training data from existing data. This is typically achieved by applying a variety of transformations to the original data, such as rotation (53), flipping, scaling, cropping (54), and adding noise (55). By augmenting the data in this way, the ML model is exposed to a greater variety of data, which can improve its ability to generalize to new and unseen data (56).

Data augmentation is particularly useful when working with limited or small datasets (26), where the model may be prone to overfitting due to the limited diversity of training examples. By artificially increasing the size and diversity of the training dataset, data augmentation can help to mitigate overfitting and improve the accuracy and robustness of the ML model (48). The effectiveness of data augmentation depends on the specific problem domain and the types of transformations applied to the data. The selection of appropriate data augmentation techniques should be based on the characteristics of the input data and the desired performance of the ML model.

3.3.3. Bypassing the limitations of the most commonly used algorithms

Bypassing the limitations of commonly used algorithms is a topic of great interest in the field of AI and ML. While algorithms such as CNN, ResNet-18, ResNet-34, XGBoost, and LightGBM have been widely used for various applications, they have certain limitations that may hinder their performance in certain scenarios.

For instance, CNNs are known to have a high number of parameters, which can lead to overfitting and require large amounts of data for training. ResNet-18 and ResNet-34 have shown better results in terms of accuracy and speed, but they still suffer from the same limitations as CNNs. XGBoost and LightGBM, on the other hand, are popular for their ability to handle structured data, but they may not perform as well on unstructured data such as images. To bypass these limitations, researchers have proposed various methods such as transfer learning, meta-learning, and ensembling.

Transfer learning involves using pre-trained models and fine-tuning them for specific tasks. The utilization of transfer learning is an effective approach for image classification, particularly when the dataset is limited in size. This technique can lead to the creation of highly sophisticated models that would otherwise require excessive computational resources. Transfer learning leverages pre-trained networks, which significantly shortens the learning process. It can significantly contribute to the improvement of an overall performance of the model (34).

Meta-learning, also known as "learning to learn," is a subfield of ML that focuses on training algorithms to improve their learning and generalization abilities across different tasks and datasets (57). Meta-learning algorithms are designed to learn and adapt quickly to new tasks with minimal data, by leveraging the knowledge and experience gained from previous tasks. In the context of AI-based X-ray image analysis, meta-learning can be used to enhance the performance and robustness of ML models by enabling them to quickly adapt to new datasets and clinical scenarios. For instance, a meta-learning algorithm can be trained on multiple X-ray image datasets with different imaging protocols, pathologies, and patient demographics, to learn the common patterns and features that are relevant to the classification and segmentation tasks. Then, the trained algorithm can be used to quickly adapt to new X-ray image datasets, with minimal or no additional training, by fine-tuning the model's parameters and architectures based on the new dataset's characteristics. Meta-learning can also be used to address some of the challenges of X-ray image analysis, such as limited data, class imbalance, and domain shift, by enabling models to learn from related tasks and domains and transfer the learned knowledge to the target task and domain. Moreover, meta-learning can help to automate and optimize the hyperparameter tuning process, such as the selection of the learning rate, optimizer, and regularization, by learning the optimal combinations of hyperparameters across different tasks and datasets.

Ensembling is a ML technique that involves combining the predictions of multiple models to improve their overall performance (58). In the context of AI-based X-ray image analysis, ensembling can be used to enhance the accuracy, robustness, and generalization abilities of ML models by reducing the effects of individual model biases, errors, and uncertainties. Ensembling can be achieved using various methods, such as bagging, boosting, and stacking. Bagging involves training multiple models on different subsets of the data and combining their predictions by averaging or voting. Boosting involves sequentially training multiple models, where each model focuses on the data points that were misclassified by the previous models. Stacking involves training a meta-model on the predictions of multiple base models, where the meta-model learns to combine the predictions based on their relative performance and reliability. Ensembling can provide several benefits in AI-based X-ray image analysis, such as improving the accuracy and robustness of the classification and segmentation tasks, reducing the risk of overfitting and underfitting, and capturing the diversity and complementarity of the models. Moreover, ensembling can help to address some of the challenges of X-ray image analysis, such as the limited availability and quality of annotated data, the heterogeneity and complexity of the clinical scenarios, and the variability and uncertainty of the imaging modalities and protocols.

In conclusion, while CNNs, ResNet-18, ResNet-34, XGBoost, and LightGBM are commonly used algorithms in AI and ML, they have certain limitations that can be overcome using methods such as transfer learning, meta-learning, and ensembling. By bypassing these limitations, researchers can improve the performance of these algorithms and expand their applications to various domains.

3.4. Evaluation metrics

As AI and ML continue to advance in the field of medical diagnostics, it becomes increasingly important to evaluate their effectiveness using appropriate metrics. Metrics provide quantifiable measures of how well an AI or ML model is performing, and can help researchers and clinicians understand its strengths and limitations.

One tool commonly used to calculate the metrics and evaluate the performance of AI and ML models is the confusion matrix (59,60). The confusion matrix is a graphical representation (Fig. 2) that summarizes the performance of a binary classification model, such as a model used to diagnose lung disease, such as COVID-19, by comparing its predictions against the true values. Usually, the rows of the matrix correspond to the actual classes of the data (e.g. sick or healthy patients), while the columns correspond to the predicted classes (e.g. predicted as sick or healthy by the model).

	Actually positive (1)	Actually negative (0)
Predicted positive (1)	True Positives (TP)	False Positives (FP)
Predicted negative (0)	False Negatives (FN)	True Negatives (TN)

Figure 2 An exemplary confusion matrix.

In the context of medical diagnostics including lung disease diagnosis, the confusion matrix can be used to evaluate the model's ability to correctly identify patients who are sick (positive cases) and those who are healthy (negative cases). The four possible outcomes are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Thus, a TP corresponds to a patient who is correctly classified as sick by the model, while a TN corresponds to a healthy patient who is correctly classified as such by the model. FP correspond to healthy patients who are erroneously classified as sick by the model, while FN correspond to sick patients who are erroneously classified as healthy by the model.

There are several evaluation metrics that are commonly used in the field of AI and ML, including accuracy, precision, sensitivity, specificity, F1-score, and the area under the receiver operating characteristic (ROC) curve:

- Accuracy is perhaps the most basic evaluation metric and refers to the percentage of correct predictions made by the model (Eq. 1). Accuracy is the most commonly used metric to evaluate the quality of classification (59). It describes what proportion of patients, out of all those classified, were correctly classified. However, accuracy does not always provide a good description of the classification process (see the explanation below). It is important to remember that higher accuracy is better, and an accuracy of 1.00 indicates a perfect match where the algorithm did not make any mistakes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Precision, on the other hand, measures the proportion of TP among all predicted positives, indicating the model's ability to correctly identify positive cases (Eq. 2) (59). Precision, like accuracy and sensitivity, should have a value as close to 1.00 as possible.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

- Sensitivity, also known as recall, measures the proportion of true positives among all actual positives, indicating the model's ability to detect positive cases (Eq. 3) (59). It is important to note that if an algorithm does not misclassify any positive cases (i.e., nothing falls into the FN category), its sensitivity will be 1, even if it misclassifies negative cases as positive (FP). Sensitivity is also a parameter that should have a value approaching 1.00.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- Specificity, on the other hand, measures the proportion of TN among all predicted negatives, indicating the model's ability to correctly identify negative cases (Eq. 4). The specificity is to the *negative* class what sensitivity is to the *positive* class. Specificity measures how many of all negative cases were actually classified into this category (59). High specificity shows that the classifier rarely makes mistakes when it comes to negative cases. Therefore, if a model of high specificity shows that something is positive, it can be expected with high probability that it is indeed positive.

$$Specificity = \frac{TN}{TN + FN} \quad (4)$$

- The F1-score is a metric used to evaluate the performance of classification algorithms, and it is calculated as the harmonic mean between precision and sensitivity (recall) (Eq. 5) (59). A value closer to 1 indicates better performance. The F1-score can reach a maximum value of 1 when there is perfect precision and sensitivity.

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

- The ROC curve is a plot of sensitivity versus 1-specificity, which provides an overall measure of how well the model is able to distinguish between positive and negative cases (Fig. 3).

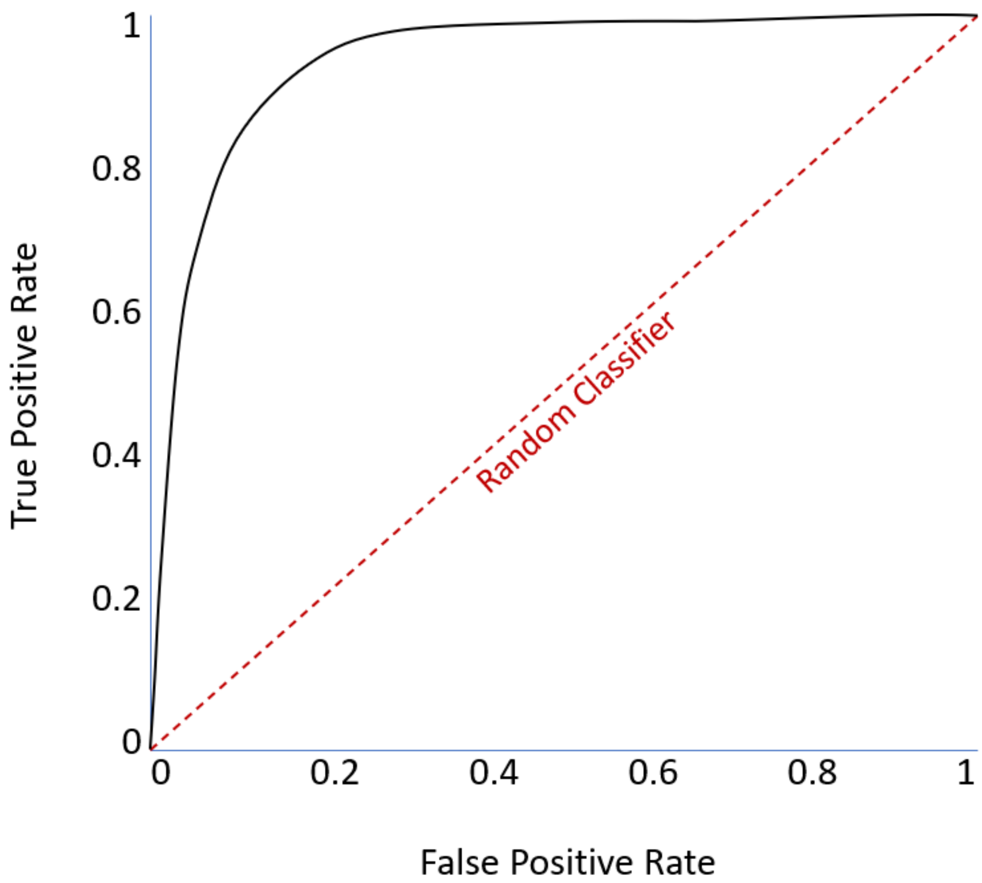

Figure 3 Exemplary ROC curve.

It is important to remember and understand that an unbalanced dataset, which refers to a scenario in which one of the classes is significantly more prevalent than the other, can lead to a skewed understanding of the accuracy of a model. This is particularly relevant in binary classification problems, where the model may appear to have high accuracy due to its ability to correctly classify the majority class while performing poorly on the minority class. However, this can ultimately result in erroneous diagnostic decision-making, as the model may fail to detect important patterns in the minority class. Therefore, addressing class imbalance through techniques such as oversampling, undersampling, or generating synthetic samples is a crucial step in achieving accurate and reliable machine learning models. When using accuracy as a metric to evaluate the performance of a model, an unbalanced dataset can result in a falsely high accuracy score. This is because the model may simply predict the majority class for most of the cases, resulting in a high percentage of correct predictions overall. For example, if 90% of the cases belong to one class, a model that always predicts that class will have a 90% accuracy score, even if it is not actually able to differentiate between the two classes. This problem can lead to overestimating the diagnostic performance of a model, especially in medical contexts where the consequences of a false positive or false negative can be severe. As a result, it is important to use metrics that are more robust to unbalanced datasets, such as precision, recall, and F1-score. Unbalanced datasets can lead to inaccurate assessment of model performance, and it is crucial to use appropriate metrics and take into account the prevalence of each class when evaluating diagnostic decision support models.

Overall, evaluation metrics and the confusion matrix provide important tools for assessing the effectiveness of AI and ML models in medical diagnostics. By carefully selecting appropriate metrics and interpreting the results of a confusion matrix, researchers and clinicians can gain insight into the strengths and limitations of tools used for medical diagnosis.

3.5. Ethical issues of using AI for diagnostic purposes

The application of AI in diagnostics raises several ethical issues (61), especially in terms of responsibility for decisions made by the algorithm. The question of who should bear responsibility for any errors or incorrect diagnoses remains a topic of debate (62,63). It is considered whether the liability should be on the side of the author/developer of the algorithm, the physician who ordered the test or maybe the hospital, which purchased the rights to use the algorithm as for diagnostic purposes? The physician's role is critical in the decision-making process, but the algorithm's output could contradict the physician's diagnosis, creating a moral dilemma. For this reason, it is important to remember that AI-based tools and algorithms are meant to support radiologists, not to make decisions on their own.

Moreover, the lack of transparency and interpretability in AI algorithms used for diagnosis is the main reason of lack of trust between physicians and the AI system. Physicians may be reluctant to use AI systems for diagnosis due to the difficulty in determining how the algorithm arrived at a particular diagnosis or recommendation. For this reason, it is hard for physicians to trust the algorithm's decisions on a matter as important as a patient's health. Additionally, the potential consequences of AI misdiagnosis may pose a threat to patient safety and lead to malpractice claims against physicians or the algorithm's author. In situations where a patient is harmed due to an AI algorithm's incorrect diagnosis, determining the responsible party can be challenging. In certain situations where a patient's health has been compromised, legal action may be taken against the healthcare provider. This can be especially true in cases where a doctor relies solely on the output of an AI algorithm without conducting a thorough analysis of the patient's medical history or physical examination. In such cases, the doctor may be held responsible for any harm caused to the patient. This highlights the ethical dilemma of using AI in healthcare and the need for a clear understanding of the role and responsibility of both the healthcare provider and the AI system in the diagnostic process. The use of AI in healthcare is still relatively new and there is no clear consensus on how to navigate these ethical issues.

AI should serve as a support for medical professionals in their work, rather than a replacement (64,65). The main objective of AI in the field of healthcare is to augment the capabilities of physicians, allowing them to make more accurate and efficient diagnoses and treatments. The use of AI in medical imaging, for instance, can help radiologists to identify early-stage diseases that might not be apparent to the naked eye.

There has been concern about whether AI will ultimately replace doctors in the future. However, it is important to recognize that AI is not intended to replace doctors, but rather to enhance their abilities. AI algorithms can assist in analyzing large amounts of data quickly and accurately, helping doctors to make informed decisions and providing them with new insights. In this way, doctors can focus on more complex tasks, such as communicating with patients and designing tailored treatment plans. Therefore, it is essential that we view AI as a tool that can aid medical professionals in their challenging work, rather than as a replacement for their expertise. With the proper use of AI, the quality of healthcare can be improved, and ultimately improve the patient experience, prognosis, and outcomes.

It is crucial for both doctors and AI developers to work together to establish clear guidelines and protocols for the safe and ethical use of AI in healthcare. To address these issues, regulatory bodies such as the US Food and Drug Administration have issued guidelines for the development and use of AI algorithms in healthcare. These guidelines stress the importance of transparency, interpretability, and accountability in AI-based healthcare systems (66,67).

In conclusion, ethical issues surrounding the use of AI in diagnosis include determining responsibility for decisions made by the algorithm, the lack of trust between physicians and AI systems, and the potential consequences of AI misdiagnosis. These issues must be addressed through increased transparency and accountability to ensure that the benefits of AI in diagnosis outweigh the risks.

3.5.1. Medical data security

The use of medical data is essential for the development and improvement of healthcare systems, as well as for the training and evaluation of AI models. However, the sharing and handling of medical data also poses significant risks in terms of privacy breaches and malicious use. Unauthorized access to medical records can lead to data/identity theft (68), financial fraud, and discrimination (69). Moreover, the leakage of sensitive medical information can have severe consequences on patients' well-being, as well as on their trust in healthcare systems. The misuse of medical data for targeted advertising or political manipulation can also have ethical implications, compromising the autonomy and dignity of individuals.

To mitigate the risks associated with medical data sharing, it is crucial to establish and enforce strong data protection regulations (70), including informed consent, de-identification (71), and secure storage and transmission. Additionally, the implementation of robust access control and monitoring mechanisms can help prevent unauthorized data access and usage (72,73).

In summary, while the use of medical data is critical for improving healthcare systems and developing AI models, it also presents significant risks that must be carefully managed and mitigated. The development of effective data protection policies and technologies is essential to ensure that medical data is used for its intended purposes and in a secure and ethical manner.

3.5.2 Interpretability and explainability of AI models

Trust from patients regarding the safe and beneficial utilization of AI is a fundamental requirement for its successful implementation in medical applications, particularly radiology. As humans, we have an innate need to comprehend how decisions affecting our health are made. Surrendering decision-making responsibilities to machines or software beyond our understanding may not be acceptable to patients. Hence, patient preferences regarding the use of AI in healthcare must take precedence, even if they do not align with the logic of computer algorithms.

It remains unclear whether individuals would tolerate imperfections in AI-driven healthcare for the potential benefit of the population at large. For example, in situations where medical imaging is solely protocol-driven and interpreted by algorithms, it is unclear whether there will be enough room for common sense and balancing individual and population-based risks against specific patient expectations for radiation exposure (74).

Furthermore, AI-driven radiology is acknowledged to be imperfect and continually evolving. The question remains whether it will be accepted by the public due to its lower cost and reduced labor in comparison to human-provided radiology. In societies with a more litigious culture, the public's tolerance for anything less than perfect delivery of medical care by humans is decreasing. Moreover, people tend to accept decisions coming from other people than from algorithms. However, in line with the idea of algorithm appreciation, people may be more likely to trust algorithmic decisions than those made by other people (75).

The use of artificial neural networks in decision-making introduces a level of opaqueness into AI-based medical practices, commonly known as the *black box* problem (76–78). This can result in doctors being reduced to mere agents of the software, carrying out its decisions without understanding how they were made. The aim of improving the integration of AI-generated advice in decision-making processes can be achieved by increasing the transparency of AI models (79). This can be done by providing additional reasoning for AI recommendations, such as visual annotations on X-rays. This approach has been shown to increase trust in and reliance on the advice, even when it is incorrect (29). However, it is not clear how different levels of task expertise affect the effectiveness of this approach. It is possible that physicians with less specialized training in reviewing medical images may benefit more from diagnostic advice with a visible annotation indicating the region that influenced the advice. To the best of my knowledge, the effect of explanations on the diagnostic decisions of physicians with varying levels of task expertise has not been studied before.

To aid in the interpretation of AI-generated diagnoses from X-ray images, heat maps (80–82) can be used to indicate which areas of the image were most influential in the model's decision-making process. However, while heat maps are commonly used in medical imaging, they have been identified as problematic in other fields of explainability literature (83). Although the hottest parts of the heat map indicate the regions of the image deemed most significant in the diagnosis of pneumonia, this does not necessarily reveal the exact features that the model relied on to make the diagnosis. Consequently, clinicians are unable to determine whether the model considered important factors such as airspace opacity, heart border shapes, or the texture of certain pixels, which may have more to do with image acquisition than the underlying disease. Therefore, heat maps have limitations in providing clear explanations for AI-generated diagnoses in medical imaging.

However, achieving interpretability and explainability in AI models presents a challenge, as too much transparency can leave the algorithm vulnerable to malicious attack or intellectual property theft, and too much explainability can limit the effectiveness of true deep learning. The advantage of AI lies in its ability to perform faster analyses and identify relationships that are beyond human capability, and this should not be compromised.

4. Study aims

The primary objective of this work was to investigate the feasibility and effectiveness of using AI models for diagnostic classification of COVID-19 patients based on their chest X-ray images.

The partial objectives were:

1. Examination of the impact of augmentation methods on the classification performance of the models.
2. Evaluation of the effect of chest X-ray images pre-processing on the classification abilities of the models.
3. Features extraction from chest X-ray images using neural networks and then comparing the classification abilities of tree-based methods (XGBoost and LightGBM).
4. Discussion of the ethical issues related to the use of AI in the clinical imaging diagnostics.

To achieve the main objective, various deep learning models such as CNN, ResNet-18, and ResNet-34 were utilized to classify the chest X-ray images into COVID-19 positive or negative classes. These models were trained and evaluated using different performance metrics such as accuracy, sensitivity, specificity, and AUC.

In addition to the main objective, the research was aimed to explore the impact of data augmentation techniques, such as rotation, flipping, and scaling, on the performance of the classification models. Furthermore, different pre-processing techniques, such as contrast enhancement and histogram equalization, were applied to the chest X-ray images to evaluate their impact on the classification accuracy. The research also focused on feature extraction from the chest X-ray images using deep learning models such as CNN. The extracted features were then used to train and compare the classification abilities of tree-based models, such as XGBoost and LightGBM.

In addition, this study aimed to address the ethical considerations surrounding the use of AI in clinical image diagnosis. The potential benefits and drawbacks of relying on ML models for medical decision-making were carefully examined, along with the question of who bears responsibility for the decisions made by these models. The results of this study may inform future discussions and policies regarding the appropriate use of AI in medical contexts.

5. Summary of works included in the series of publications

5.1. Original paper I – Pre-processing methods in chest X-ray image classification

The COVID-19 pandemic has affected various aspects of society, including social, medical, psychologic, economic, and industrial aspects. The current, gold standard screening method for detecting COVID-19 infections is the reverse transcriptase-polymerase chain reaction (RT-PCR) test, which detects SARS-CoV-2 ribonucleic acid (RNA) extracted from specimens inhabiting respiratory tract of patients. However, abnormalities on chest X-ray images can also be indicative of infection. The use of ML-based methods can improve the efficiency of X-ray analysis, support medics in diagnosis, and lighten the burden on healthcare systems. The paper presents a study that investigates the impact of pre-processing methods on the classification of chest X-rays into three classes: normal (healthy), COVID-19, and pneumonia. The proposed ML-based method was able to classify the images accurately, and the implementation of AI and ML in COVID-19 and other lung diseases is seen as a natural progression.

The paper includes a brief review of the state-of-the-art solutions. The text discusses various deep learning-based methods for automatic lung disease recognition, particularly for COVID-19 detection, using medical imaging. Milestones in pre-processing, feature extraction, and assigning a classification were required to achieve the desired results. Despite some challenges, including imperfect datasets and difficulty in acquiring data and annotations, promising results have been achieved using various methods such as U-NET, ResNet, ReCoNet, and Bayes-SqueezeNet. Several studies have reported high accuracy rates using pre-trained deep CNN models, such as ResNet-50, InceptionV3, Xception, ResNeXt, and SVM classifiers. The importance of having representative and well-balanced datasets in ML experiments is discussed.

The study utilized a public dataset containing posteroanterior (PA) chest X-ray images of COVID-19, pneumonia, and normal cases. The dataset consisted of 6,939 samples with 2,313 samples for each class. To evaluate the effect of pre-processing on the classification results, six different pre-processing approaches were examined. A CNN was implemented to classify the images, and its architecture consisted of 12 layers, including convolution layers, rectified linear unit activation functions, pooling layers, and batch normalization operations. The dataset was divided into training, validating, and testing subsets, and the experiments were conducted using the Kaggle notebook. The output from the neural network showed the probability of an image belonging to one of the three classes, with the highest probability chosen as the final result. The study demonstrated the effectiveness of pre-processing on improving the accuracy of the classification of chest X-ray images.

The text reports on experiments conducted to evaluate the effectiveness of various image pre-processing methods on improving the accuracy, precision, recall, and F1-score of a classification model. Four evaluation parameters were used, and a confusion matrix was employed to report the number of true positives, true negatives, false positives, and false negatives. Results showed that without any pre-processing, the accuracy was 93% and the F1-score for the three classes ranged between 91% to 96%. Applying histogram equalization improved the parameters by 2%, while combining histogram equalization with Gaussian blurring and adaptive masking resulted in all evaluated parameters exceeding 97%, making it the most promising approach.

The paper proposes a fully automated approach for the analysis of COVID-19 chest X-ray images using a neural network. The proposed method successfully distinguished images into three classes: COVID-19, pneumonia, and normal. However, there are issues to keep in mind if the modality is implemented in patient care. One of them is responsibility, as the proposed method is not a tool for replacing the educated specialist but to improve their work and support the diagnostic process. Another issue is the quality of images used in the learning process, and the authors trust that the images provided are labeled correctly and submitted by an expert. We also proposed an improvement in the pre-processing part of the ML-based system, which increases the efficiency of the system as the F1-score raised from 93% to over 97%. The results are comparable to other similar ML-based approaches in the literature, but there are plenty of pre-processing methods that can improve the efficiency of the system and be implemented in future work.

My contribution to the research reported in this paper has been conceptualization and realization of the study. I performed a thorough state-of-the-art approaches analysis. I collected the necessary datasets for the research and algorithm preparation. I performed the investigation and compared obtained results to the literature data. I wrote the original draft of this paper.

5.2. Original paper II – A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images

The article discusses the challenges faced during the COVID-19 pandemic with regards to the diagnosis of the disease caused by the SARS-CoV-2 virus. The PCR test has been the gold standard for confirming SARS-CoV-2 infection, but it is not error-free and false results are possible. Additionally, some infected individuals may not display symptoms and therefore not be referred for testing. There have been reports of successful use of ML methods to detect COVID-19 infections on X-ray images of the lungs. We proposed a novel approach to chest X-ray image analysis using a CNN-based features extraction method to diagnose COVID-19. The authors obtained a new dataset containing samples from confirmed COVID-19 cases and uninfected patients and performed augmentation to increase the dataset's size. The proposed features extraction method was implemented for different classifiers, and promising results were obtained. The article concludes with a discussion of the results and comparisons with other state-of-the-art approaches, as well as future perspectives for work in this area.

The review of state-of-the-art approaches indicated several deep-learning-based methods proposed by researchers for accurate and quick diagnosis of COVID-19 using X-ray images. These methods use various deep learning frameworks such as Deep Hybrid Learning (DHL), Deep Boosted Hybrid Learning (DBHL), and CNN for lung segmentation and localization of specific changes caused by COVID-19. Transfer learning is also used to create a model capable of detecting COVID-19 changes in X-ray images of the lungs. The authors of papers included in this review have shown that these models can achieve high accuracy and sensitivity, making them effective tools for radiologists to quickly diagnose COVID-19. Additionally, the paper mentions that ML-based methods can support vaccine discovery, and there are seven key requirements identified by the European Commission for the implementation of AI, including human agencies and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination, and fairness, societal and environmental well-being, and accountability.

The proposed method for COVID-19 diagnosis using chest X-ray images involves data augmentation, pre-processing, feature extraction, and classification using ML algorithms. Real data from 60 chest X-ray images were used, 30 from healthy individuals and 30 COVID-19 positive confirmed by PCR test. The dataset was divided into three subsets for training, validation, and testing. Augmentation was performed by rotations, noise addition, and zooming out to increase the size of the training dataset. Data pre-processing involved normalization and mask application to select the region of interest. ML-based methods using CNN were used for feature extraction and classification. The CNN architecture included three pairs of convolutional layers and max pooling layers, and the flattened layer and dense layer were used with the most promising feature extraction that had 57 features. Tree-based classifiers including XGBoost, Random Forest, LightGBM, and CatBoost were examined, and CNN with softmax activation function was used for binary classification. All experiments were carried out using Python 3.7 and the TensorFlow platform with scikit-learn, Xgboost, Lightgbm and Catboost libraries.

The study treated the problem of disease diagnosis as a binary classification task and evaluated and compared ML-based methods using confusion matrices and four metrics TP, FP, FN, and TN. The study found that using XGBoost and LightGBM classifiers provided the highest accuracy, precision, recall, and F1-score, with LightGBM having a faster training and prediction time, making it the optimal classifier. The text also explains that Tesla systems with GPU support are a powerful and cost-effective alternative to traditional high-power computing systems for image processing and medical diagnostics. Finally, the text provides an overview of how LightGBM works, using leaf-sage techniques to achieve the optimal number of leaves in trees and using the minimum amount of data in the tree.

Summarizing, the paper discusses the potential of ML methods as a valuable tool for COVID-19 diagnosis, specifically in screening with X-ray images, which are less expensive and faster than PCR testing. We noted that while ML-based methods cannot replace an experienced medical doctor in the final diagnosis, they can significantly assist in the process, relieving the burden on health care and improving the diagnostic process. We also emphasized the importance of explainable AI (xAI) in clinical applications and propose a model using pre-trained networks (ResNet-18 and DenseNet-121) to perform image classification with an AUC score of 0.81. The presented model is fast, efficient, and does not require high computing power, making it suitable for hospital laboratories. We acknowledged the limitations of the study due to the small number of original images that formed the basis of the used dataset and called for further cooperation with hospitals to provide more learning data. We also noted that there are potential future improvements required for the presented model, such as validation on a larger, different dataset, and implementing xAI to provide explanations for ML-based decisions.

My contribution to the research reported in this paper has been conceptualization and realization of the study. I performed a thorough state-of-the-art approaches analysis. I collected the necessary datasets for the research and algorithm preparation. I performed the investigation and compared obtained results to the literature data. I was involved in the funding acquisition, to publish the results of conducted research. I wrote the original draft of this paper.

5.3. Original paper III – Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification

The COVID-19 pandemic caused widespread health, mental, and social issues, as well as overwhelming healthcare systems around the world. AI has been suggested as a tool to aid in various aspects of the pandemic crisis, including medical diagnosis, drug development, patient treatment, epidemiology, and socioeconomics. This paper focuses on the use of AI to improve the diagnosis of COVID-19 patients using chest X-ray scans. The researchers implemented a baseline transfer learning schema to detect COVID-19 symptoms in X-ray images and tested different scenarios of augmentation to evaluate their impact on evaluation metrics such as accuracy, precision, recall, and F1-score. The proposed system was validated on a dataset of real data obtained from hospitals, and the results were compared to other state-of-the-art analytical algorithms. The paper highlights the advantages of augmentation, including its cost-effectiveness, accuracy, controllability, and ability to overcome data sample limitations and overfitting problems. The use of AI in COVID-19 diagnosis has the potential to significantly accelerate the diagnostic process and keep it cost-effective.

The study utilized a balanced dataset of 30,386 chest X-ray images, obtained from various sources and divided into COVID-19 positive and negative classes. The proposed method used data augmentation, pre-processing, and classification with a CNN. Seven different data augmentation approaches were evaluated, including color manipulation, contrast and brightness adjustment, noise addition, geometric transformation, and rotation. The albumentations library was used for augmentation, and each group was trained and validated independently. Pre-processing included resizing and masking using a ResNet34 segmentation model. The impact of masking on classification metrics was evaluated by running experiments both with and without segmentation. The study aimed to determine the impact of data augmentation methods on the final classification result and to improve the model's ability to generalize.

CNN model, specifically the ResNet18 architecture, was used for the classification task. The dataset used for training and validation comprised 14,191 healthy images and 16,194 COVID-19 positive images, which were shuffled and divided into subsets. The learning parameters for the CNN model were experimentally set, including the optimizer (stochastic gradient descent - SGD), loss function (cross-entropy), number of epochs (200), batch size (16), and early stopping rounds (10). The method was evaluated using a pre-prepared hospital dataset comprising 62 chest X-ray images, with 30 healthy individuals and 32 COVID-19 positive patients confirmed by RT-PCR tests. Four validation metrics, including accuracy, precision, recall, and F1-score, were used to evaluate the performance of the proposed method. These metrics were based on the TP, FP, FN, and TN measures and are considered a golden standard in ML-based studies.

The model evaluation showed that augmentation improved its performance. The most promising augmentation technique was mixed augmentation, and masking also significantly improved the evaluation metrics. The experiments were performed using Nvidia Tesla GPU, which provided high computing power and was an attractive alternative to traditional high-power computing systems.

The paper highlights the importance of augmentation techniques for improving the accuracy of COVID-19 detection on lung X-ray images. The proposed schema of the augmentation technique involves classical image processing methods and ML techniques, including transfer learning and GAN-based augmentation. The results show that a combination of all described groups of augmentations is the most promising approach. The proposed schema was compared with other state-of-the-art solutions previously proposed, and it was found to be competitive. Explainability is also discussed, and visual explanation is proposed for the model's interpretability. The paper concludes by highlighting two possible future improvements: further explainability improvement and complexity reduction of the proposed schema to decrease energy consumption and carbon footprint.

My contribution to the research reported in this paper has been conceptualization and realization of the study. I performed a thorough state of the art analysis. I collected the necessary datasets for the research and algorithm preparation. I performed the investigation and compared obtained results to the literature data. I wrote the original draft of this paper.

6. Publications that are the subject of the dissertation

6.1. Original paper I – content of the publication “Pre-processing methods in chest X-ray image classification”

PLOS ONE

RESEARCH ARTICLE

Pre-processing methods in chest X-ray image classification

Agata Giełczyk[1☯*], **Anna Marciniak**[1,2☯], **Martyna Tarczewska**[1☯], **Zbigniew Lutowski**[1☯]

1 Bydgoszcz University of Science and Technology, Bydgoszcz, Poland, **2** Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Bydgoszcz, Poland

☯ These authors contributed equally to this work.
* agata.gielczyk@pbs.edu.pl

Abstract

Background

The SARS-CoV-2 pandemic began in early 2020, paralyzing human life all over the world and threatening our security. Thus, the need for an effective, novel approach to diagnosing, preventing, and treating COVID-19 infections became paramount.

Methods

This article proposes a machine learning-based method for the classification of chest X-ray images. We also examined some of the pre-processing methods such as thresholding, blurring, and histogram equalization.

Results

We found the F1-score results rose to 97%, 96%, and 99% for the three analyzed classes: healthy, COVID-19, and pneumonia, respectively.

Conclusion

Our research provides proof that machine learning can be used to support medics in chest X-ray classification and improving pre-processing leads to improvements in accuracy, precision, recall, and F1-scores.

OPEN ACCESS

Citation: Giełczyk A, Marciniak A, Tarczewska M, Lutowski Z (2022) Pre-processing methods in chest X-ray image classification. PLoS ONE 17(4): e0265949. https://doi.org/10.1371/journal.pone.0265949

Editor: Zahid Mehmood, University of Engineering and Technology Taxila Pakistan, PAKISTAN

Received: November 25, 2021

Accepted: March 10, 2022

Published: April 5, 2022

Data Availability Statement: Data available for free at the Kaggle repository. www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

The occurrence of the COVID-19 pandemic in 2020 has shaken up the modern world. It has caused societies to close, crowded streets to become deserted, pubs and clubs to be silenced, and popular meeting places to die down. Currently, people all over the world are doing their best to overcome the pandemic’s impact on the social, medical, psychologic, economic, and industrial aspects of society. Currently, the main screening method for detecting COVID-19 infections is reverse transcriptase-polymerase chain reaction (RT-PCR) testing. The RT-PCR test can detect SARS-CoV-2 ribonucleic acid (RNA) from respiratory specimens (collected through nasopharyngeal or oropharyngeal swabs). In addition, patients suffering from COVID-19 can also present with abnormalities on chest X-ray images that are characteristic of infection [1]. This imaging modality is highly available and accessible in many clinical locations, and it is considered standard equipment in most healthcare systems. Moreover, CXR imaging is more widely available than CT imaging, especially in developing countries due to high equipment and maintenance costs. However, X-ray analysis can be time-consuming and requires highly educated specialists to interpret. But, the use of machine learning (ML)-based methods can improve efficiency, support medics in the diagnosis of COVID-19, speed up the time to diagnosis, and lighten the already burdened health care system.

At the same time, modern technologies have gathered more interest. Artificial intelligence (AI) and ML can be used in numerous applications such as cybersecurity [2], pedestrian detection [3], telemedicine [4], biometrics [5] or sports analytics [6]. Thus, the implementation of AI and ML in COVID-19 and other lung diseases seems to be the desired natural progression.

In this article, we present the impact pre-processing can have on the results of a classification system. We tested 5 different pre-processing methods and investigated their effect on the final classification. We conducted the study using a large public dataset. The proposed ML-based method was able to classify chest X-rays into 3 classes: normal (healthy), COVID-19, and pneumonia, which can be similar to images of patients infected by COVID-19.

2 Related work

As the COVID pandemic intensified, more investigators focused on automatic lung disease recognition. Milestones in pre-processing, feature extraction, and assigning a classification were required to achieve the required results. In addition, improvements were made at each step of the workflow.

The authors in [7] used a method based on U-NET and ResNet to perform the segmentation of CT images with an accuracy reaching 95%. The main obstacle in overcoming the segmentation problem is imperfect datasets. As mentioned in [8], medical image segmentation datasets suffer from scarce and weak annotations. In addition, acquiring the medical image's data and annotations can be extremely difficult and expensive. In the article by [9], the authors proposed the use of a multi-level CNN-based preprocessor. The main reason for using this preprocessor was to dynamically enhance the lung regions that are useful in detecting COVID-19. Experiments using ReCoNet for differentiating COVID vs Pneumonia vs Normal were shown to have an accuracy > 97%. Authors in [10] proposed a novel, hybrid, multimodal deep learning system. With the use of Contrast-Limited Adaptive Histogram Equalization (CLAHE) and a Butterworth bandpass filter, the authors were able to enhance the contrast of X-ray images and eliminate the noise leading to an accuracy of 99.93%. The article by [11] highlights that pre-processing can improve a system's accuracy. In this publication, the visibility of the diaphragm on the chest X-ray was mentioned. It was observed as a very light object in the bottom part of the chest. However, experiments using a convolutional neural network (CNN) reported improved results when the diaphragm was removed from the sample. In [12], an interesting and efficient approach based on the Bayes-SqueezeNet method was proposed. What is more, the authors described some details concerning data augmentation. In this specific study, the augmentation was performed offline using Gaussian blur, sheering, and decreasing/increasing brightness. The presented experiments provided promising results, namely an F1 score of 0.983. Mahdy et al. in [13], proposed a method to automatically classify COVID-19 chest X-rays using a multi-level threshold based on the Otsu algorithm and support vector machine (SVM). A SVM was also utilized in the article by [14]. In the presented approach, image enhancement was performed by increasing contrast, the Histogram of

Table 1. The review of available X-ray datasets for COVID-19 classification.

Name	Classes and samples	Source
COVID-19 RADIOGRAPHY DATABASE	COVID—3616, Lung opacity—6012, Normal—10.2k, Viral pneumonia—1345	Kaggle
Covid19 Image Dataset	COVID—137, Normal—90, Viral pneumonia—90	Kaggle
Covid-19 X Ray 10000 Images	COVID—70, Normal—28	Kaggle
Chest X-ray (Covid-19 & Pneumonia)	COVID—576, Normal—1583, Pneumonia—4273	Kaggle
COVID19 Pneumonia Normal Chest Xray PA Dataset	COVID—2313, Normal—2313, Pneumonia—2313	Kaggle
Covid Chestxray Dataset	PA view—481, AP view—173, for over 15 different lung diseases	Github
Covid Patients Chest X-ray	COVID—162, Normal—162	Kaggle

https://doi.org/10.1371/journal.pone.0265949.t001

Oriented Gradients was used for features extraction, and Linear Regression and SVM were implemented for X-Ray classification resulting in an accuracy of 96%.

Models are using neural networks to analyze lung X-rays in cancer [15], pneumonia [16], and other lung diseases [17]. In the wake of the recent pandemic, deep learning methods have been used to analyze X-rays of patients potentially infected with the SARS-CoV-2 virus. A standard state-of-art approach using pre-trained CNNs was presented by authors in [18]. In the evaluation of AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception CNNs, the best results were achieved by Xception and ResNet-101, with an AUC of 0.994 for both networks and an accuracy of 99.02% and 99.51%, respectively. The authors in [19] connected pre-trained deep CNN models with various kernels SVM classifiers. This approach was compared with end-to-end training CNN models that performed worse. The best results (accuracy of 94.7%) were obtained by ResNet50 and SVM with linear kernels. Authors of [20], compared 3 deep-learning based CNN models, InceptionV3, Xception, and ResNeXt. Analysis was performed on 6,432 X-ray scans collected from the Kaggle repository. The highest accuracy was obtained for the Xception model (97.97%). Authors in [21] also tested InceptionV3 and compared the results with 8 more pre-trained CNNs. In that study, the overall accuracy for InceptionV3 was 54.41%, whereas the highest accuracy was achieved by the VGG16 model (95.88%). CoroDet, a novel CNN model, was developed by the authors of [22]. It allows X-ray images and CT scans to be classified into 2, 3, or 4 classes (COVID, Normal, non-COVID viral pneumonia, and non-COVID bacterial pneumonia) with an accuracy of 99.1%, 94.2%, and 91.2%, respectively.

As the COVID-19 pandemic started, researchers focused on providing datasets for performing scientific experiments. Selected datasets are listed in Table 1. As one can see, they contain not only COVID-19 chest X-ray images but also images of other lung diseases like pneumonia and SARS. In ML, the dataset must meet certain requirements. It should be representative of the disease and population being studied, consist of a large sample of data points, and be well balanced. Unfortunately, some of the listed datasets do not meet all of the above-mentioned requirements.

3 Materials and methods

In this research, we used the dataset available for the public at www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset. It consists of images collected from the GitHub repository, Kaggle, Radiopedia, Italian Society of Radiology (SIRM), and Figshare data repository websites. The dataset was organized into 3 classes (COVID-19, pneumonia, and normal) containing posteroanterior (PA) chest X-ray images. A total of 6,939

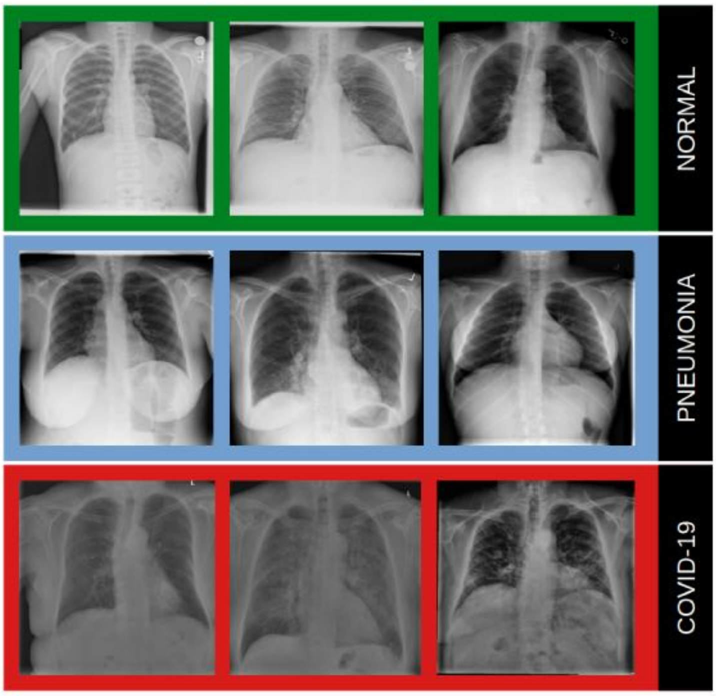

Fig 1. Examples of samples from the dataset: Normal (healthy), pneumonia and COVID-19.

https://doi.org/10.1371/journal.pone.0265949.g001

samples were used in the experiment, and 2,313 samples were used for each class. Some examples of images used are presented in Fig 1.

To verify how the selected pre-processing method affects the final classification result, we proposed a baseline system. The general overview of this system is presented in Fig 2. The black box visible in Fig 2 marks the selected pre-processing method. The pre-processing step is an important element in the image analysis schema. It can enhance the original image and reduce noise or unwanted details. In our research, we examined 6 different approaches to pre-processing:

1. None—the baseline approach is not to use any method apart from size reduction.
2. Histogram equalization—this method extends the pixel's intensity range from the original range to 0 to 255. Thus, the enhanced image has a wider range of intensity and slightly higher contrast.
3. Hist. eq. + Gaussian blur—this filter reduces some noise and unwanted details that can be confusing for the neural network; the filter kernel size was experimentally set to 5×5 size.
4. Hist. eq. + bilateral filter—this filter also reduces some noise and unwanted details that can be confusing for the neural network, but its main feature is to preserve edges; the experimentally set up parameters of the filter: $diameter = 5$, $\sigma_{color} = \sigma_{space} = 75$.

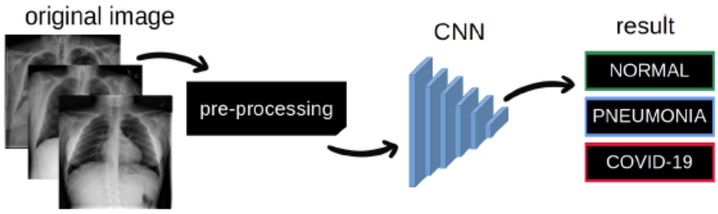

Fig 2. The overview of the proposed method.

https://doi.org/10.1371/journal.pone.0265949.g002

5. Adaptive masking—in [11] the authors proved that by removing the diaphragm from the sample it is possible to improve the classification results. In this proposed pre-processing method, we first found the maximum (max) and minimum (min) intensity of pixels and then applied the binary thresholding using the threshold expressed in Eq 1. The next step used morphologic closing. This creates the adaptive mask that after bitwise operation removes the diaphragm from the source image.
6. Adaptive masking + hist. eq. + Gaussian blur—this method joins adaptive masking with histogram equalization and Gaussian blur.

$$threshold = min + 0.9 \cdot (max - min) \tag{1}$$

At image classification, a CNN was implemented. The CNN model can provide human-like accuracy in classifying various images (Fig 2). A convolution network can be described as a chain of convolution layers, with rectified linear unit activation functions, pooling layers, and batch normalization operations. The architecture of a CNN is analogous to that of the connectivity pattern of neurons in the human brain. In fact, it was inspired by the organization of the visual cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. A collection of these fields overlap to encompass the entire visual area. The hierarchical network provides high-level feature maps, reduced computation complexity, and improved generalization ability. The advantages of CNNs have led to their wide implementation in image processing. For chest X-ray analysis, CNN was implemented in [23–27]. The neural network used in this paper consists of 12 layers:

- Conv2D—it is a convolution layer with 64 filters with dimensions of 3×3 and the activation function ReLU, the dimensions of the input data is also introduced
- MaxPooling with a size of 2×2
- Conv2D—it is a convolution layer with 64 filters with dimensions of 3×3 and the activation function ReLU
- MaxPooling with a size of 2x2
- Conv2D—it is a convolution layer with 128 filters with dimensions of 3×3 and the activation function ReLU
- MaxPooling with a size of 2x2
- Conv2D—it is a convolution layer with 128 filters with dimensions of 3×3 and the activation function ReLU

- MaxPooling with a size of 2x2
- Flatten—it is a data flattening layer, it has no additional parameters
- Dropout layer—randomly sets input units to 0 with a frequency rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1—rate) such that the sum over all inputs are unchanged, in this case, the rate equals 0.2
- Dense layer—in which each neuron is connected to each neuron of the previous layer with the unit parameters (positive integer, dimensionality of the output space) equal to 512, with the activation function ReLU
- Dense—with unit parameters equal to 3, with softmax activation function.

The output from the neural network shows the probability of an image belonging to one of the three classes thanks to the softmax activation function in the last layer. The network selects the classification with the highest probability and identifies it as the final result. All the experiments were executed using the online Kaggle notebook. There were almost 7,000 samples in the dataset. We decided to divide the dataset into three disjoint subsets: training-65%, validating-15%, and testing-20%. All of the experiments were executed 3 times to prove their independence from the learning data. Due to a balanced dataset, we did not need any sample augmentation.

4 Results

The above-mentioned experiments provided some promising results. We used 4 parameters for the evaluation methods—accuracy, precision, recall, and F1-score. The parameters were calculated using a confusion matrix (presented in Fig 3) reporting the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The evaluation parameters were calculated using Eqs 2–5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5}$$

The results without any additional pre-processing resulted in an accuracy of 93% and a F1-score in the range of 91% to 96% for the three evaluated classes. Introducing a pre-processing method improved the parameters, for instance, applying histogram equalization raised the precision, recall, and F1-score by 2%. Nevertheless, the most promising approach was joining histogram equalization with Gaussian blurring and adaptive masking. This approach ensured all evaluated parameters exceeded 97%. The results for each pre-processing method are presented in Table 2.

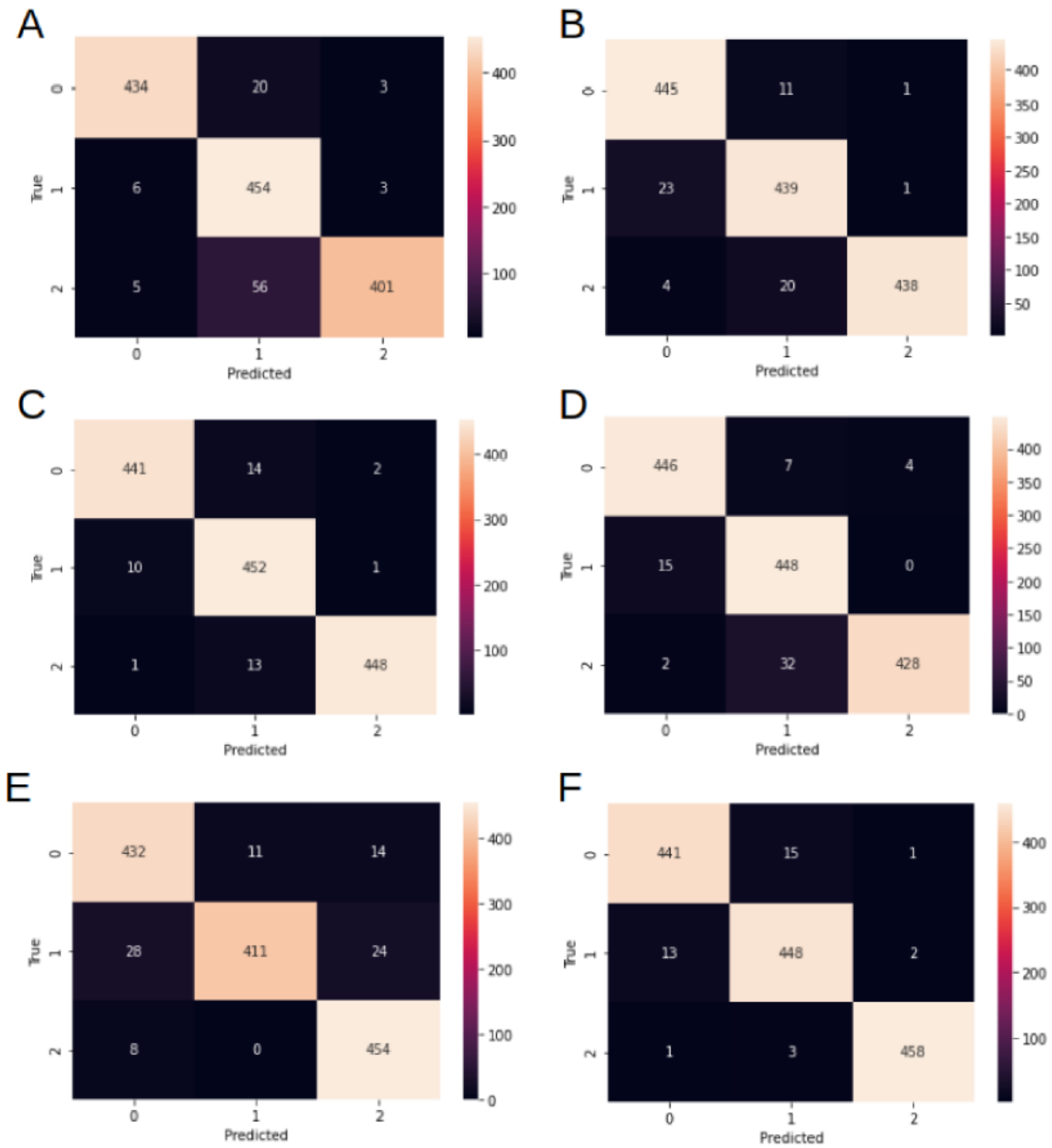

Fig 3. Confusion matrices. Confusion matrices for the experiments varying by the pre-processing method—A: none, B: histogram equalization, C: histogram equalization + Gaussian blur, D: histogram equalization + bilateral filter, E: adaptive mask, F: adaptive mask + histogram equalization + Gausssian blur.

https://doi.org/10.1371/journal.pone.0265949.g003

Table 2. Obtained results for different pre-processing methods: 1—none, 2—histogram equalization, 3—Gaussian blur + hist. equalization, 4—bilateral filter + hist. equalization, 5—adaptive masking, and 6—adaptive masking + Gauss. blur + hist. eq.

Method	Class	Accuracy	Precision	Recall	F1-score
1	Normal	0.9754	0.9753	0.9497	0.9623
	COVID-19	0.9385	0.8566	0.9806	0.9144
	Pneumonia	0.9515	0.9853	0.8680	0.9229
	Average	0.9551	0.9390	0.9327	0.9332
2	Normal	0.9712	0.9428	0.9737	0.9580
	COVID-19	0.9602	0.9340	0.9482	0.9411
	Pneumonia	0.9812	0.9955	0.9481	0.9712
	Average	0.9711	0.9574	0.9567	0.9567
3	Normal	0.9805	0.9757	0.9650	0.9703
	COVID-19	0.9725	0.9436	0.9762	0.9597
	Pneumonia	0.9877	0.9933	0.9697	0.9814
	Average	0.9802	0.9709	0.9703	0.9704
4	Normal	0.9566	0.9633	0.9759	0.9696
	COVID-19	0.9609	0.9199	0.9676	0.9432
	Pneumonia	0.9725	0.9907	0.9264	0.9575
	Average	0.9566	0.9580	0.9566	0.9567
5	Normal	0.9559	0.9231	0.9453	0.9341
	COVID-19	0.9544	0.9739	0.8877	0.9288
	Pneumonia	0.9667	0.9228	0.9827	0.9518
	Average	0.9590	0.9399	0.9386	0.9382
6	**Normal**	**0.9782**	**0.9692**	**0.9650**	**0.9671**
	COVID-19	**0.9761**	**0.9614**	**0.9676**	**0.9645**
	Pneumonia	**0.9949**	**0.9935**	**0.9913**	**0.9924**
	Average	**0.9831**	**0.9747**	**0.9746**	**0.9747**

https://doi.org/10.1371/journal.pone.0265949.t002

5 Discussion

5.1 Threads to validity

The presented method is very powerful in diagnosing COVID-19 and pneumonia. However, there are some issues to keep in mind if this modality is implemented in patient care. The first issue is responsibility—who is going to be responsible for the ML-based decision? Thus, we have to specify that the proposed method is not a tool for replacing the educated specialist but to improve his/her work and support the diagnostic process. Furthermore, to implement the proposed method, the explainability of the module must be added to the described pipeline. Explanation of the decisions would be the main task for such a module. It would give the reasons why the sample was classified to a specific class and would be very helpful in marking the part of the image responsible for the decision. A few explainable ML-based methods have been published recently (see: [28–30]).

The second issue is the quality of images used in the learning process. In the presented approach, the dataset was obtained from numerous sources: Github, Kaggle, Radiopedia, SIRM, and Figshare data repository websites. We deeply trust that the images provided are labeled correctly and submitted by an expert. The labeling process seems to be the most challenging, expensive, and time-consuming part of the chest X-ray image analysis system. Due to the relatively recent identification of COVID-19, the number of samples in the datasets are limited. Some commercial projects are working on gathering COVID-19 data (including chest X-rays), but these datasets are currently not available to the public.

Table 3. Comparison between the proposed method and SOTA methods.

Reference	Dataset	Result
proposed	**Kaggle**	**Acc. = 98.31%, Prec. = 97.47%, Rec. = 97.46%, F1 = 97.47%**
Ahmed et. al. [9]	GitHub	Acc. = 97.48%, Prec. = 97.39%, Spec. = 97.53%, MCC = 92.49%
Al-Waisy et. al. [10]	Github, Kaggle	Acc. = 99.93%, Prec. = 100%, Rec. = 99.90%, F1 = 99.93%
Mahdy et. al. [13]	GitHub	Acc. = 97.48%, Prec. = 95.276%, Spec. = 99.7%
Ucar et. al. [12]	arXiv, Kaggle	Acc. = 98.3%, Prec. = 98.3%, Rec. = 98.3%, F1 = 98.3%

https://doi.org/10.1371/journal.pone.0265949.t003

5.2 Comparison to SOTA

The detection of COVID-19 and other lung diseases using chest X-ray imaging has recently been widely investigated. Table 3 provides detailed results from the current literature. Unfortunately, not all authors evaluate the same set of parameters as in our study, namely: Accuracy, Precision, Recall, and F1-score. Even though, the provided results prove that our solution is comparable to other SOTA methods.

6 Conclusions

COVID-19 is a highly infectious disease caused by the most recently discovered coronavirus and is considered a pandemic according to the World Health Organization. Even though vaccines were introduced at the beginning of 2021, there is a strong need for fast and accurate tools to improve the efficiency of the healthcare system.

In this article, we proposed a novel approach for the fully automated analysis of COVID-19 chest X-ray images using a neural network. Our approach was successful in distinguishing images into three classes: COVID-19, pneumonia, and normal (healthy). We also presented an improvement in the proposed method, namely the pre-processing part of the ML-based system. In this early step of image analysis, a few crucial operations are performed: adaptive masking (the part of the image that is very light is removed), histogram equalization, and Gaussian blur (removes noise and some unwanted details). We proved that the proposed pre-processing method increases the efficiency of the system as the F1-score raised from 93% to over 97%. Our results are comparable to other similar ML-based approaches in the literature, but there are plenty of pre-processing methods that can improve the efficiency of the system and be implemented in future work.

Author Contributions

Conceptualization: Zbigniew Lutowski.

Investigation: Agata Giełczyk, Anna Marciniak, Martyna Tarczewska.

Methodology: Agata Giełczyk, Martyna Tarczewska.

Writing – original draft: Agata Giełczyk, Anna Marciniak.

Writing – review & editing: Anna Marciniak, Martyna Tarczewska, Zbigniew Lutowski.

References

1. Ng Ming-Yen, et al. "Imaging profile of the COVID-19 infection: radiologic findings and literature review." Radiology: Cardiothoracic Imaging 2.1 (2020): e200034. https://doi.org/10.1148/ryct.2020200034 PMID: 33778547

2. Pawlicki Marek, Michał Choraś, and Rafał Kozik. "Defending network intrusion detection systems against adversarial evasion attacks." Future Generation Computer Systems 110 (2020): 148–154. https://doi.org/10.1016/j.future.2020.04.013

3. Nowosielski Adam, et al. "Embedded night-vision system for pedestrian detection." IEEE Sensors Journal 20.16 (2020): 9293–9304.

4. Szymkowski Maciej, et al. "A novelty approach to retina diagnosing using biometric techniques with SVM and clustering algorithms." IEEE Access 8 (2020): 125849–125862. https://doi.org/10.1109/ACCESS.2020.3007656

5. Giełczyk Agata, Michał Choraś, and Rafał Kozik. "Lightweight verification schema for image-based palmprint biometric systems." Mobile Information Systems 2019 (2019).

6. Ćwiklinski Bartosz, Agata Giełczyk, and Michał Choraś. "Who will score? a machine learning approach to supporting football team building and transfers." Entropy 23.1 (2021): 90. https://doi.org/10.3390/e23010090

7. Saood Adnan, and Iyad Hatem. "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet." BMC Medical Imaging 21.1 (2021): 1–10. https://doi.org/10.1186/s12880-020-00529-5

8. Tajbakhsh Nima, et al. "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation." Medical Image Analysis 63 (2020): 101693. https://doi.org/10.1016/j.media.2020.101693 PMID: 32289663

9. Ahmed Sabbir, et al. "Reconet: Multi-level preprocessing of chest x-rays for covid-19 detection using convolutional neural networks." medRxiv (2020).

10. Al-Waisy A. S., et al. "COVID-DeepNet: hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images." Computers, Materials and Continua 67.2 (2021): 2409–2429. https://doi.org/10.32604/cmc.2021.012955

11. Heidari Morteza, et al. "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms." International journal of medical informatics 144 (2020): 104284. https://doi.org/10.1016/j.ijmedinf.2020.104284 PMID: 32992136

12. Ucar Ferhat, and Deniz Korkmaz. "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images." Medical hypotheses 140 (2020): 109761. https://doi.org/10.1016/j.mehy.2020.109761 PMID: 32344309

13. Mahdy Lamia Nabil, et al. "Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine." MedRxiv (2020).

14. Ullah Syed Irfan, et al. "COVID-19 lung image classification based on logistic regression and support vector machine." European, Asian, Middle Eastern, North African Conference on Management & Information Systems. Springer, Cham, 2021.

15. Woźniak Marcin, et al. "Small lung nodules detection based on local variance analysis and probabilistic neural network." Computer methods and programs in biomedicine 161 (2018): 173–180. https://doi.org/10.1016/j.cmpb.2018.04.025 PMID: 29852959

16. Jaiswal Amit Kumar, et al. "Identifying pneumonia in chest X-rays: a deep learning approach." Measurement 145 (2019): 511–518. https://doi.org/10.1016/j.measurement.2019.05.076

17. Bharati Subrato, Prajoy Podder, and Mondal M. Rubaiyat Hossain. "Hybrid deep learning for detecting lung diseases from X-ray images." Informatics in Medicine Unlocked 20 (2020): 100391. https://doi.org/10.1016/j.imu.2020.100391 PMID: 32835077

18. Abbasian Ardakani Ali, et al. "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks." Computers in biology and medicine 121 (2020): 103795. https://doi.org/10.1016/j.compbiomed.2020.103795

19. Ismael Aras M., and Abdulkadir Şengür. "Deep learning approaches for COVID-19 detection based on chest X-ray images." Expert Systems with Applications 164 (2021): 114054. https://doi.org/10.1016/j.eswa.2020.114054 PMID: 33013005

20. Jain Rachna, et al. "Deep learning based detection and analysis of COVID-19 on chest X-ray images." Applied Intelligence 51.3 (2021): 1690–1700. https://doi.org/10.1007/s10489-020-01902-1 PMID: 34764553

21. Makris Antonios, Ioannis Kontopoulos, and Konstantinos Tserpes. "COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks." 11th hellenic conference on artificial intelligence. 2020.

22. Hussain Emtiaz, et al. "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images." Chaos, Solitons & Fractals 142 (2021): 110495. https://doi.org/10.1016/j.chaos.2020.110495 PMID: 33250589

23. Tawsifur Rahman, et al. "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray." Applied Sciences 10.9 (2020): 3233.

24. Jia Guangyu, Hak-Keung Lam, and Yujia Xu. "Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method." Computers in biology and medicine 134 (2021): 104425. https://doi.org/10.1016/j.compbiomed.2021.104425 PMID: 33971427

25. Tobias Rogelio Ruzcko NMI, et al. "CNN-based deep learning model for chest X-ray health classification using tensorflow." 2020 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, 2020.

26. Rahman Mohammad Marufur, et al. "HOG+CNN Net: Diagnosing COVID-19 and pneumonia by deep neural network from chest X-Ray images." Sn Computer Science 2.5 (2021): 1–15. https://doi.org/10.1007/s42979-021-00762-x PMID: 34254055

27. Pathan Sameena, Siddalingaswamy P. C., and Ali Tanweer. "Automated Detection of Covid-19 from Chest X-ray scans using an optimized CNN architecture." Applied Soft Computing 104 (2021): 107238. https://doi.org/10.1016/j.asoc.2021.107238 PMID: 33649705

28. Luca Brunese, et al. "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays." Computer Methods and Programs in Biomedicine 196 (2020): 105608. https://doi.org/10.1016/j.cmpb.2020.105608

29. Singh Rajeev Kumar, Rohan Pandey, and Babu Rishie Nandhan. "COVIDScreen: explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays." Neural Computing and Applications 33.14 (2021): 8871–8892. https://doi.org/10.1007/s00521-020-05636-6 PMID: 33437132

30. Hou Jie, and Terry Gao. "Explainable DCNN based chest X-ray image analysis and classification for COVID-19 pneumonia detection." Scientific Reports 11.1 (2021): 1–15. https://doi.org/10.1038/s41598-021-95680-6

6.2. Original paper II – content of the publication “A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images”

Journal of *Clinical Medicine*

MDPI

Article

A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images

Agata Giełczyk [1,*], Anna Marciniak [1,2], Martyna Tarczewska [1], Sylwester Michal Kloska [2], Alicja Harmoza [2], Zbigniew Serafin [2] and Marcin Woźniak [2]

1 Faculty of Telecommunications, Computer Science and Electrical Engineering, Bydgoszcz University of Science and Technology, 85-796 Bydgoszcz, Poland
2 Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, 85-067 Bydgoszcz, Poland
* Correspondence: agata.gielczyk@pbs.edu.pl

Abstract: Background: This paper presents a novel lightweight approach based on machine learning methods supporting COVID-19 diagnostics based on X-ray images. The presented schema offers effective and quick diagnosis of COVID-19. Methods: Real data (X-ray images) from hospital patients were used in this study. All labels, namely those that were COVID-19 positive and negative, were confirmed by a PCR test. Feature extraction was performed using a convolutional neural network, and the subsequent classification of samples used Random Forest, XGBoost, LightGBM and CatBoost. Results: The LightGBM model was the most effective in classifying patients on the basis of features extracted from X-ray images, with an accuracy of 1.00, a precision of 1.00, a recall of 1.00 and an F1-score of 1.00. Conclusion: The proposed schema can potentially be used as a support for radiologists to improve the diagnostic process. The presented approach is efficient and fast. Moreover, it is not excessively complex computationally.

Keywords: features extraction; X-ray images; COVID-19; machine learning; image processing

check for updates

Citation: Giełczyk, A.; Marciniak, A.; Tarczewska, M.; Kloska, S.M.; Harmoza, A.; Serafin, Z.; Woźniak, M. A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images. *J. Clin. Med.* **2022**, *11*, 5501. https://doi.org/10.3390/jcm11195501

Academic Editors: Adam Piorkowski, Rafal Obuchowicz, Andrzej Urbanik, Michał Strzelecki and Davide Chiumello

Received: 3 August 2022
Accepted: 15 September 2022
Published: 20 September 2022

1. Introduction

COVID-19 is a disease caused by the SARS-CoV-2 virus. It has a wide range of symptoms, most of which affect the respiratory tract. It can lead to serious inflammation of the lungs and, consequently, pneumonia [1]. The COVID-19 pandemic has exposed healthcare problems around the world. The large number of patients to diagnose and the limited number of tests and staff available turned out to be a significant problem. As a result, the number of diagnostic tests performed was very often too low. The diagnostic method that turned out to be the gold standard for the confirmation of SARS-CoV-2 infection was the polymerase chain reaction (PCR) test. However, this method is not error free and sometimes gives false results [2]. Another difficulty comes from the fact that some people, despite being infected with the SARS-CoV-2 virus, do not develop symptoms of the disease [3] and are not referred for PCR testing. This can cause problems with the correct diagnosis of the disease. Despite its latency, the disease can cause serious changes in the lungs. In these cases, the diagnosis is possible on the basis of an X-ray of the lungs. For this reason, machine learning methods have been used to detect COVID-19 infections on X-ray images. Methods based on machine learning (ML) turned out to be effective and useful during the analysis and assessment of the impact of diseases (e.g., COVID-19 or pneumonia) on X-ray images of the lungs [4–6]. The major contributions of this paper are as follows:

- We propose a novel approach to chest X-ray image analysis in order to diagnose COVID-19 using an original CNN-based features extraction method.

- We obtained a new dataset containing samples from confirmed COVID-19 cases as well as from uninfected patients. The infection status of both groups was confirmed by a PCR test. We performed an augmentation in order to increase the dataset's size.
- We implemented the proposed features extraction for different classifiers, obtaining promising results.

Further parts of this paper are constructed in the following manner: (a) a brief review of the state-of-the-art is presented in Section 2; (b) in Section 3, we describe the dataset, augmentation process and the proposed approach to the classification; (c) in Section 4, we present the obtained results; (d) in Section 5 (the Discussion) we compare our results with other state-of-the-art approaches, we pull out some conclusions, and we present perspectives for future work on this topic.

2. Related Work

To combat the challenges posed by the pandemic to the healthcare service, Khan et al. [7] proposed a deep-learning-based method of accurate and quick diagnosis of COVID-19 using X-ray images. They proposed a method consisting of two novel deep learning frameworks: Deep Hybrid Learning (DHL) and Deep Boosted Hybrid Learning (DBHL). The use of both of these frameworks led to an improvement in their COVID-19 diagnostic methods. The result was a model capable of identifying COVID-19 in X-ray images with over 98% accuracy on a previously unseen dataset. This method has been shown to be effective in reducing both false positives and false negatives and has proven to be a useful supportive tool for radiologists.

Tahir et al. [8] proposed a model based on a convolutional neural network (CNN) capable of lung segmentation and localization of specific changes caused by COVID-19. The dataset used consisted of nearly 34,000 X-ray images, including lung images of people with COVID-19 and pneumonia and of healthy people. An important element of the study was the appropriate marking of photos by specialists. This method recognized COVID-19 and its effects on the lung image with sensitivity and specificity values over 99%.

In [9], Brunese et al. used transfer learning to create a model capable of detecting COVID-19 changes in X-ray images of the lungs. This model is applicable (1) to the classification of healthy people and patients with changes in lung X-ray images; (2) to distinguish between COVID-19 and other lung diseases; and (3) to distinguish lung lesions caused by the SARS-CoV-2 infection. This model was based on the VGG-16 (16-layered convolutional neural network) and underwent transfer learning. This study included 6523 X-ray images from healthy individuals, patients with various lung diseases and patients with COVID-19. The model trained on this dataset achieved a sensitivity equal to 0.96 and a specificity of 0.98 (accuracy of 0.96) for distinguishing between healthy individuals and patients with lung diseases, and a sensitivity of 0.87 and a specificity equal to 0.94 (accuracy of 0.98) for distinguishing lung diseases from COVID-19. The image analysis process itself is extremely fast and takes only about 2.5 s. The data presented by the authors indicate that the developed model achieves good and reliable results.

Chakraborty et al. [10] presented a COVID-19 detection method based on a Deep Learning Method (DLM) using X-ray images. The authors used different architectures of deep neural networks in order to achieve optimal results. They combined several pre-trained models such as ResNet18, AlexNet, DenseNet, VGG16, etc. This approach showed to be effective and cost significantly less than standard laboratory diagnostic methods. The dataset consisted of 10,040 chest X-ray images, which included a normal/healthy population, COVID-19 patients and patients with pneumonia. The presented model was highly accurate (96.43%) and sensitive (93.68%). This work showed the high usefulness of ML models for determining changes in X-ray images, which can facilitate the work of radiologists who, as a result of this quick method, can refer patients directly to treatment.

Civit-Masot et al. [11] noted that traditional tests to identify SARS-CoV-2 infection are invasive and time consuming. Imaging, on the other hand, is a useful method for assessing disease symptoms. Due to the limited number of trained medical doctors who can reliably assess X-ray images, it is necessary to invent ways to facilitate this type of assessment. The authors used the VGG-16-based Deep Learning model to identify pneumonia and COVID-19. The presented results indicate high accuracy (close to 100%) and specificity of the model, which qualify it as an effective screening test.

It is also worth mentioning that ML-based methods can support not only radiology specialists. In [12], we can see the transfer learning approach to discovering the impact of the stringency index on the number of deaths caused by the SARS-Cov-2 virus. As presented in [13], ML can be also implemented in order to predict the COVID-19 diagnosis based on symptoms. Statistical analyses revealed that the most frequent and significant predictive symptoms are fevers (41.1%), coughs (30.3%), lung infections (13.1%) and runny noses (8.43%). A total of 54.4% of people examined did not develop any symptoms that could be used for diagnosis. Moreover, ML can also be a useful tool in vaccine discovering, as presented in [14].

Recently, numerous ML-based approaches for rapid diagnostics have been published. In addition, they have gathered increasingly more attention from some government and international agencies. For example, the European Commission published a White Paper entitled 'On Artificial Intelligence—A European approach to excellence and trust [15]. Seven key requirements were identified and are described in the document:

- Human agencies and oversight;
- Technical robustness and safety;
- Privacy and data governance;
- Transparency;
- Diversity, non-discrimination and fairness;
- Societal and environmental well-being;
- Accountability.

The following statement in the EC publication is worth noting: AI can and should itself critically examine resource usage and energy consumption and should be trained to make choices that are positive for the environment. It follows from the above citation that it is extremely important to focus on providing solutions that are not only cost and time effective but also that spare energy used for computations. This kind of approach to AI has become introduced as the 'GreenAI' [16] and has gathered increasingly more attention recently [17]. Research working on the GreenAI have also proposed some novel metrics [18]. As a result of this metric, researchers can compare not only the accuracy and precision of the proposed method, but also its sustainability and eco-friendliness.

3. Materials and Methods

The general overview of the proposed method is presented in Figure 1. The image presents an example of the data obtained from the hospital and some consequential steps: the augmentation process and pre-processing of the sample. In Figure 1, the features extraction and classification steps are presented. Finally, the proposed method gives the answer of "*true*" for the COVID-19-positive sample and "*false*" for the healthy sample. Each step of the process is described in detail in this section. All experiments were carried out with the use of Python 3.7 and the TensorFlow platform. Among others, we used the following libraries: scikit-learn, Xgboost, Lightgbm and Catboost.

3.1. Dataset

In this research, anonymized real data were used. The data were obtained from Antoni Jurasz University Hospital No. 1 in Bydgoszcz, Department of Radiology and Imaging Diagnostics. A total of 60 chest X-ray images were obtained; 30 were from healthy individuals, and 30 had COVID-19 confirmed by a PCR test. The images were provided in the DICOM format. The images were in a raw form, without masks. Some samples from the dataset are presented in Figure 2.

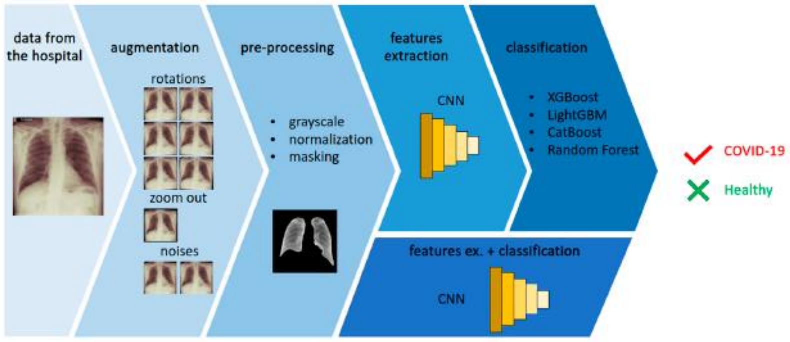

Figure 1. The following steps of processing in the proposed method: data acquisition, data augmentation, sample pre-processing, features extraction and binary classification of COVID-19 as positive or negative (healthy).

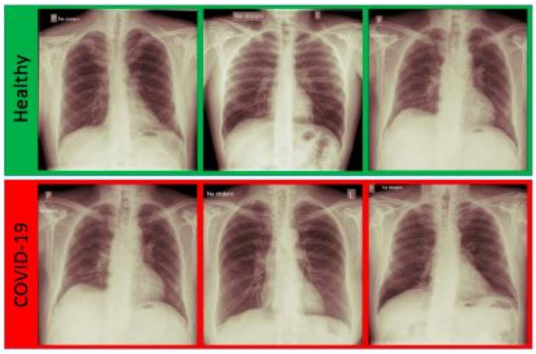

Figure 2. The exemplary images from the dataset divided into two classes: Healthy and COVID-19 confirmed by a PCR test.

3.2. Data Augmentation

In order to perform training, the dataset was divided into 3 disjointed subsets: the training set (80%), validation set (10%) and testing set (10%). Unfortunately, the quantity of samples was not enough to use any ML technique. Thus, we decided to use augmentation for increasing the size of the training dataset. As a result of the augmentation, 10 samples from one single image were obtained. The initial proper balance in the dataset was unchanged, and as a result, the dataset was still well balanced. The following methods for augmentation were implemented:

- rotations—1°, 2° and 3° both clockwise and anti-clockwise;
- noises—a random Gaussian noise and a salt and pepper noise were added;
- zooming out—the image was resized to obtain 95% of its original size.

3.3. Data Pre-Processing

First of all, the samples were moved to grayscale images and were normalized. The goal of normalization was to improve the quality of the images, e.g., by enhancing the contrast, as described in [19]. The data obtained from the hospital were not masked. Thus, the essential step of processing was to provide proper masks to help in selecting the region of interest. The goal of this step was to prevent the ML-based model from learning information that is useless from the point of view of COVID-19 diagnostics, such as images of a collar bone or a stomach. Hand crafting masks would be time consuming, and it would require the involvement of a specialist. On the other hand, proposing a novel method of segmentation can be treated as a separate scientific problem, as presented in [20,21]. Therefore, it was decided to use the pretrained model, which is widely available and very powerful in masking X-ray images [22].

3.4. ML-Based Methods

ML-based methods were used in two steps of processing, namely features extraction and classification. As a baseline, a convolutional neural network (CNN) was used for both steps. Then, almost the same CNN architecture was used solely for extracting features, since it was reported as very promising and efficient [23,24].

The features extraction step can be an essential one for the whole image processing system. It can reduce the complexity of the problem, and consequently, it can make the proposed approach more efficient, require less computing time and, therefore, more eco-friendly. The general schema of the CNN is presented in Figure 3. The input in this architecture was grayscale images with sizes of 512 × 512 pixels. Then, three pairs of convolutional layers and max pooling layers were used. Each of them was responsible for performing operations between the filters and the input of each corresponding layer. The convolutional layers consisted of 64, 128 and 256 filters, respectively. On each layer, an ReLU activation function was used to implement and perform nonlinear transformations. Then, the flattened layer and the dense layer were used. The proposed neural networks against the dense layer neuron quantity were examined. The values, with a range of [5,200], were tested. Involving the validation subset, it was observed that the most promising was using 57 features. This features extraction type was qualified for further research and development. Then, numerous classifiers were examined: XGBoost [25], Random Forest [26], LightGBM [27] and CatBoost [28]. It was decided to use these classifiers due to some reasons. First of all, they are tree-based algorithms, and they perform very well in binary classification problems. Secondly, they have been already used in similar applications, as presented in [29–31], providing promising results.

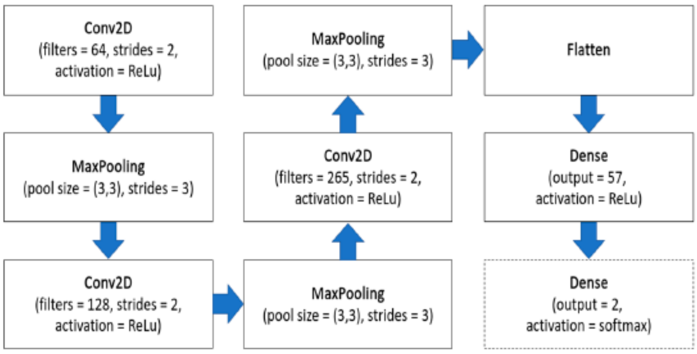

Figure 3. The architecture of the CNN used in the research. In dashed lines, the added Dense network was in solely a CNN-based approach.

The approach based on solely CNN both for features extraction and for classification had one change in the architecture. In this case, the next dense layer with the softmax activation function (presented in Figure 3 in dashed line) was added. It enabled the binary classification of COVID-19 positive or negative. This approach was trained, validated and tested using the above-mentioned training, validation and testing datasets, respectively. The training parameters were: 50 epochs, a learning rate equal to 0.00001 and a loss function set to SparseCategoricalCrossentropy.

4. Results

Since the hospital data represent two classes (COVID-19 positive and healthy), the problem of the disease diagnostics can be treated as a binary classification. In this research, the confusion matrices were used in order to evaluate and compare the ML-based methods. Four measures were defined, as follows:

- TP—true positives—COVID-19-infected patients classified as sick;
- FP—false positives—healthy patient images classified as COVID-19 infected;
- FN—false negatives—COVID-19-infected patients classified as healthy;
- TN—true negatives—healthy patients classified as healthy.

Each model in the research was evaluated using accuracy (Equation (1)), precision (Equation (2)), recall (Equation (3)) and F1-score (Equation (4)), which use the above-mentioned measures of TP, FP, FN and TN.

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{1}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{3}$$

$$\text{F1-score} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \tag{4}$$

All experiments were performed using a Tesla with GPU support. As a result of its enormous computing power, low price, relatively low demand for electricity and the CUDA environment support, Tesla systems have become an attractive alternative to traditional high-power computing systems, such as CPU clusters and supercomputers. This kind of device can be extremely helpful in image processing and also in medicine diagnostics.

The obtained results from all the experiments are provided in Table 1. All the evaluated metrics are given: accuracy, precision, recall and F1-score. The approach using the CNN both for features extraction and for classification provided the less promising results. Two examined classifiers provided the highest results: XGBoost and LightGBM, with accuracy = 1.0, precision = 1.0, recall = 1.0 and F1-score = 1.0. For selecting the optimal classifier for the presented solution, the computational time for both classifiers, namely the training time and prediction time for a single image, was compared. It was decided to use this parameter for optimization because the goal was to provide a light, sustainable and eco-friendly solution. For XGBoost, the average training time was equal to 242 ms, and the prediction time for a single image was above 11ms. For LightGBM, those times were 132 ms and less than 2 ms, respectively. That is why LightGBM is marked in bold in Table 1.

Table 1. Obtained results: accuracy, precision, recall and F1-score for all experiments.

F. Extractor	Classifier	Accuracy	Precision	Recall	F1-Score
CNN	CNN	0.86	0.75	1.00	0.86
CNN	XGBoost	1.00	1.00	1.00	1.00
CNN	Random Forest	0.91	0.86	1.00	0.92
CNN	LightGBM	1.00	1.00	1.00	1.00
CNN	CatBoost	0.91	0.86	1.00	0.92

One could ask 'what makes LightGBM faster than XGBoost?'. The following are the features of LightGBM that affect its effectiveness and the mathematics (Equation (5)) behind LightGBM that allow one to understand the answer to this question [32,33].

$$V_j(d) = \frac{1}{n} \tag{5}$$

where $A_l = \{x_i \in A : x_{ij} \leq d\}$, $A_r = \{x_i \in A : x_{ij} > d\}$, $B_l = \{x_i \in B : x_{ij} \leq d\}$, $B_r = \{x_i \in B : x_{ij} > d\}$, d is the point in the data where the split is calculated to find the optimal gain in variance and the coefficient $\frac{1-a}{b}$ is used to normalize the sum of the gradients over B back to the size of A^C.

LightGBM produces trees and finds the leaves with the greatest variance to perform division with the use of leaf-sage techniques. LightGBM achieves the optimal number of leaves in the trees and uses the minimum amount of data in the tree.

5. Discussion

ML methods can be valuable tool for COVID-19 diagnosis. ML-based methods cannot replace an experienced medical doctor in the final diagnosis, but they help significantly in the process, relieving the burden on health care and improving the diagnostic process. Screening with X-ray images is less expensive and faster than PCR testing [34]. This is one of the reasons why it is worth developing ML-based techniques to assist specialists in diagnostics.

In [35], the authors paid particular attention to the explainability AI (xAI), as it is essential in clinical applications. Explainable approaches increase the confidence and trust of the medical community in AI-based methods. They noted that X-ray imaging was not a method of choice when diagnosing COVID-19. However, the changes visible in the X-ray images of the lungs allow for the detection of pathological changes at an early stage of their development. For this reason, the authors indicated the usefulness of models supporting radiologists in their work and improving the decision-making process. The dataset of X-ray images used by the authors contained nearly 900 X-ray images of both COVID-19 patients and healthy patients, which was a significantly bigger dataset than that presented in this paper. Therefore, it could learn a wider range of differences between the images. In this work, the authors used pre-trained networks (ResNet-18 and DenseNet-121) to perform image classification with the best AUC score of 0.81. The sensitivity and specificity results obtained by authors were significantly lower than ours, however, which may be caused by dataset size differences. The model proposed in this research is fast, efficient and does not require high computing power; thus, it can be used in ordinary computers in hospital laboratories. The presented model obtained satisfactory results of evaluation metrics, which confirm its accuracy. These results are comparable and potentially better than those reported in the state-of-the-art review (Section 2). Some detailed results provided in the literature are presented in Table 2. However, our concern remains on the small number of original images that formed the basis of the database used. We believe that the method of data augmentation used may introduce bias; however, with such a small amount of data, this step was necessary. We believe that this is an aspect that could be improved in the course of further cooperation with hospitals that would provide more learning data.

Table 2. Results compared to other state-of-the-art methods, namely accuracy, precision, recall, F1-score and AUC. The results not provided by the authors are marked with '-'.

Authors	Method	Acc.	Prec.	Rec.	F1	AUC
Rajagopal [27]	CNN + SVM	0.95	0.95	0.95	0.96	-
Júnior et al. [30]	VGG19 + XGBoost	0.99	0.99	0.99	0.99	-
Nasari et al. [29]	DenseNet169 + XGBoost	0.98	0.98	0.92	0.97	-
Ezzoddin et al. [36]	DenseNet169 + LightGBM	0.99	0.99	1.00	0.99	-
Laeli et al. [28]	CNN + RF	0.99	-	-	-	0.99
Proposed	CNN + LightGBM	1.00	1.00	1.00	1.00	1.00

Table 2 presents numerous approaches providing comparable results. It is essential to mention that some of them are based on very complex architectures, use an extremely big number of training parameters, need excessive computational power and require long training. However, we decided to use fewer than 60 features and a light, fast classifier in our approach. Thus, the approach proposed in this paper is lightweight, efficient and fast. In the literature, we can observe very complex and resource-consuming approaches (such as COVID-Net, as proposed in [37], or ResNet, as implemented in [38]). It is worth emphasizing that the proposed solution is lighter but still equally or more efficient. Unfortunately, it is very difficult to compare the eco-friendliness of different computing-based diagnostic approaches, as it is not customary to provide such information in scientific publications. Hopefully, in the near future, the sustainability of the proposed methods will become more significant for researchers and editors.

We are also aware that there are some future improvements required for the presented model. It is necessary to validate the model on a larger, different dataset. X-ray images made with the use of various equipment exhibit different features, which may be an obstacle to the universality of the presented model. It is worth checking how the model performs on a new set of data from the same hospital and, alternatively, on a set from a different source. Likely, the additional pre-processing can help to make all samples uniform. Another potential extension of this work is providing the xAI. Its main aim is not only giving the classification but also providing an explanation of why such a decision was made by a ML-based method. Implementing the xAI can allow radiologist doctors to evaluate the model and verify whether it makes the decisions based on real COVID-19 lesions.

Author Contributions: Conceptualization, A.G., M.T. and A.M.; methodology, A.G.; software, M.T.; validation, A.M., Z.S. and S.M.K.; formal analysis, M.W.; investigation, A.M. and S.M.K.; resources, Z.S. and A.H.; data curation, M.W. and A.H.; writing—original draft preparation, S.M.K.; writing—review and editing, A.G.; visualization, M.T.; supervision, M.W.; project administration, A.G.; funding acquisition, A.G and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Bydgoszcz University of Science and Technology grant number DNM 32/2022. Part of this research was supported by the National Centre for Research and Development of Poland (Narodowe Centrum Badań i rozwoju, NCBR) with project number POWR.03.02.00-00-I019/16-00.

Institutional Review Board Statement: The project was approved by the Ethics Committee of the Nicolaus Copernicus University in Toruń Collegium Medicum in Bydgoszcz (decision no. KB 454/2022.)

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the research in a raw, anonymized form are available at https://github.com/UTP-WTIiE/Xray_data.git (accessed on 2 August 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zheng, J. SARS-CoV-2: An Emerging Coronavirus that Causes a Global Threat. *Int. J. Biol. Sci.* **2020**, *16*, 1678–1685. [CrossRef] [PubMed]
2. Tahamtan, A.; Ardebili, A. Real-time RT-PCR in COVID-19 detection: Issues affecting the results. *Expert Rev. Mol. Diagn.* **2020**, *20*, 453–454. [CrossRef] [PubMed]
3. Pang, L.; Liu, S.; Zhang, X.; Tian, T.; Zhao, Z. Transmission Dynamics and Control Strategies of COVID-19 in Wuhan, China. *J. Biol. Syst.* **2020**, *28*, 543–560. [CrossRef]
4. Momeny, M.; Neshat, A.A.; Hussain, M.A.; Kia, S.; Marhamati, M.; Jahanbakhshi, A.; Hamarneh, G. Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images. *Comput. Biol. Med.* **2021**, *136*, 104704. [CrossRef]
5. Kassania, S.H.; Kassanib, P.H.; Wesolowskic, M.J.; Schneidera, K.A.; Detersa, R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern. Biomed. Eng.* **2021**, *41*, 867–879. [CrossRef]
6. Jain, G.; Mittal, D.; Thakur, D.; Mittal, M.K. A deep learning approach to detect Covid-19 coronavirus with X-ray images. *Biocybern. Biomed. Eng.* **2020**, *40*, 1391–1405. [CrossRef]
7. Khan, S.H.; Sohail, A.; Khan, A.; Hassan, M.; Lee, Y.S.; Alam, J.; Basit, A.; Zubair, S. COVID-19 detection in chest X-ray images using deep boosted hybrid learning. *Comput. Biol. Med.* **2021**, *137*, 104816. [CrossRef]
8. Tahir, A.M.; Chowdhury, M.E.; Khandakar, A.; Rahman, T.; Qiblawey, Y.; Khurshid, U.; Kiranyaz, S.; Ibtehaz, N.; Rahman, M.S.; Al-Maadeed, S.; et al. COVID-19 infection localization and severity grading from chest X-ray images. *Comput. Biol. Med.* **2021**, *139*, 105002. [CrossRef]
9. Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A. Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. *Comput. Methods Programs Biomed.* **2020**, *196*, 105608. [CrossRef]

10. Chakraborty, S.; Murali, B.; Mitra, A.K. An Efficient Deep Learning Model to Detect COVID-19 Using Chest X-ray Images. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2013. [CrossRef]
11. Civit-Masot, J.; Luna-Perejón, F.; Domínguez Morales, M.; Civit, A. Deep Learning System for COVID-19 Diagnosis Aid Using X-ray Pulmonary Images. *Appl. Sci.* **2020**, *10*, 4640. [CrossRef]
12. Ayesha, S.; Yu, Z.; Nutini, A. COVID-19 Variants and Transfer Learning for the Emerging Stringency Indices. *Neural Process. Lett.* **2022**, 1–10. [CrossRef]
13. Ahamad, M.M.; Aktar, S.; Rashed-Al-Mahfuz, M.; Uddin, S.; Liò, P.; Xu, H.; Summers, M.A.; Quinn, J.M.W.; Moni, M.A. A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. *Expert Syst. Appl.* **2020**, *160*, 113661. [CrossRef] [PubMed]
14. Kannan, S.; Subbaram, K.; Ali, S.; Kannan, H. The role of artificial intelligence and machine learning techniques: Race for covid-19 vaccine. *Arch. Clin. Infect. Dis.* **2020**, *15*, e103232. [CrossRef]
15. EC. On Artificial Intelligence—A European Approach to Excellence and Trusty. Available online: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf/ (accessed on 2 July 2022).
16. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green ai. *Commun. ACM* **2020**, *63*, 54–63. [CrossRef]
17. Yigitcanlar, T. Greening the artificial intelligence for a sustainable planet: An editorial commentary. *Sustainability* **2021**, *13*, 13508. [CrossRef]
18. Lenherr, N.; Pawlitzek, R.; Michel, B. New universal sustainability metrics to assess edge intelligence. *Sustain. Comput. Inform. Syst.* **2021**, *31*, 100580. [CrossRef]
19. Giełczyk, A.; Marciniak, A.; Tarczewska, M.; Lutowski, Z. Pre-processing methods in chest X-ray image classification. *PLoS ONE* **2022**, *17*, e0265949. [CrossRef]
20. Hasan, M.J.; Alom, M.S.; Ali, M.S. Deep learning based detection and segmentation of COVID-19 & pneumonia on chest X-ray image. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 27–28 February 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 210–214.
21. Munusamy, H.; Muthukumar, K.J.; Gnanaprakasam, S.; Shanmugakani, T.R.; Sekar, A. FractalCovNet architecture for COVID-19 chest X-ray image classification and CT-scan image segmentation. *Biocybern. Biomed. Eng.* **2021**, *41*, 1025–1038. [CrossRef]
22. Alimbekov, R.; Vassilenko, I.; Turlassov, A. Lung Segmentation Library. Available online: https://github.com/alimbekovKZ/lungs_segmentation/ (accessed on 15 May 2022).
23. Rajagopal, R. Comparative analysis of COVID-19 X-ray images classification using convolutional neural network, transfer learning, and machine learning classifiers using deep features. *Pattern Recognit. Image Anal.* **2021**, *31*, 313–322. [CrossRef]
24. Laeli, A.R.; Rustam, Z.; Pandelaki, J. Tuberculosis Detection based on Chest X-rays using Ensemble Method with CNN Feature Extraction. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 7–8 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 682–686.
25. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
26. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
27. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
28. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Volume 31.
29. Nasiri, H.; Hasani, S. Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost. *Radiography* **2022**, *28*, 732–738. [CrossRef] [PubMed]
30. Júnior, D.A.D.; da Cruz, L.B.; Diniz, J.O.B.; da Silva, G.L.F.; Junior, G.B.; Silva, A.C.; de Paiva, A.C.; Nunes, R.A.; Gattass, M. Automatic method for classifying COVID-19 patients based on chest X-ray images, using deep features and PSO-optimized XGBoost. *Expert Syst. Appl.* **2021**, *183*, 115452. [CrossRef] [PubMed]
31. Jawahar, M.; Prassanna, J.; Ravi, V.; Anbarasi, L.J.; Jasmine, S.G.; Manikandan, R.; Sekaran, R.; Kannan, S. Computer-aided diagnosis of COVID-19 from chest X-ray images using histogram-oriented gradient features and Random Forest classifier. *Multimed. Tools Appl.* **2022**, 1–18. [CrossRef] [PubMed]
32. Machado, M.R.; Karray, S.; Sousa, I.T.D. LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In Proceedings of the 14th International Conference Science and Education, Toronto, ON, Canada, 19–21 September 2019.
33. Misshuari, I.W.; Herdiana, R.; Farikhin, A.; Saputra, J. Factors that Affect Customer Credit Payments During COVID-19 Pandemic: An Application of Light Gradient Boosting Machine (LightGBM) and Classification and Regression Tree (CART). In Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management, Singapore, 7–11 March 2021.
34. Nikolaou, V.; Massaro, S.; Fakhimi, M.; Stergioulas, L.; Garn, W. COVID-19 diagnosis from chest x-rays: Developing a simple, fast, and accurate neural network. *Health Inf. Sci. Syst.* **2021**, *9*, 1–11. [CrossRef]
35. Barbano, C.A.; Tartaglione, E.; Berzovini, C.; Calandri, M.; Grangetto, M. A two-step explainable approach for COVID-19 computer-aided diagnosis from chest x-ray images. *arXiv* **2021**, arXiv:2101.10223.

36. Ezzoddin, M.; Nasiri, H.; Dorrigiv, M. Diagnosis of COVID-19 Cases from Chest X-ray Images Using Deep Neural Network and LightGBM. In Proceedings of the 2022 International Conference on Machine Vision and Image Processing (MVIP), Ahvaz, Iran, 23–24 February 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–7.
37. Wang, L.; Lin, Z.Q.; Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Rep.* **2020**, *10*, 19549. [CrossRef] [PubMed]
38. Rajpal, S.; Lakhyani, N.; Singh, A.K.; Kohli, R.; Kumar, N. Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images. *Chaos Solitons Fractals* **2021**, *145*, 110749. [CrossRef]

6.3. Original paper III – content of the publication "Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification"

© Pol J Radiol 2023
DOI: https://doi.org/10.5114/pjr.2023.126717

Received: 28.03.2023
Accepted: 30.03.2023
Published:

POLISH JOURNAL of RADIOLOGY

http://www.polradiol.com

Original paper

Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification

Anna Kloska[1,2,A,E,F], Martyna Tarczewska[2,E,G], Agata Giełczyk[2,D,F], Sylwester M. Kloska[1,A,C,F], Adrian Michalski[3,B,G], Zbigniew Serafin[1,B,G], Marcin Woźniak[1,A,G]

[1] Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Bydgoszcz, Poland
[2] Bydgoszcz University of Science and Technology, Bydgoszcz, Poland
[3] Department of Analytical Chemistry, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University in Torun, Bydgoszcz, Poland

Abstract

Purpose: A pandemic disease elicited by the virus SARS-CoV2 has become a serious health issue due to infecting millions of people all over the world. Recent publications proof that artificial intelligence (AI) can be used for medical diagnosis purposes including X-ray images interpretation. X-ray scanning is relatively cheap and scan processing is not computationally demanding.

Material and methods: In our experiment a baseline transfer learning schema of processing the lung X-ray images, including augmentation, in order to detect the COVID-19 symptoms was implemented. Seven different scenarios of augmentation were proposed. The model was trained on a dataset consisting of more than 30,000 X-ray images.

Results: The obtained model was evaluated using real images from one of the Polish hospitals, with the use of standard metrics, and achieved accuracy = 0.9839, precision = 0.9697, recall = 1.0000, and F1-score = 0.9846.

Conclusions: Our experiment proved that augmentations and masking could be important steps of data pre-processing and could contribute to improvement of the evaluation metrics. As medical professionals often tend to lack confidence in AI-based tools, we have designed the proposed model so that its results would be explainable and could play a supporting role for radiology specialists in their work.

Key words: image processing, data augmentation, machine learning, COVID-19.

Correspondence address:
Anna Kloska, Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Jagiellońska 13/15, 85-067 Bydgoszcz, e-mail: 503015@stud.umk.pl

Authors' contribution:
A Study design · **B** Data acquisition · **C** Data analysis · **D** Data interpretation · **E** Method creation · **F** Manuscript preparation · **G** Manuscript revision

Introduction

A pandemic disease elicited by the virus SARS-CoV2 has caused serious (health, mental, social etc.) issues by infecting millions of people all over the world. It was reported that more than 200 countries have been affected by the coronavirus pandemic. Apart of causing disease symptoms (like: fever, fatigue, cough, and respiratory distress) the COVID-19 pandemic caused a failure in the health services due to the lack of medical staff or overloading entire healthcare systems. However, recent publications suggest that artificial intelligence (AI) could be used to aid in various aspects of pandemic crisis including: medical diagnosis, novel drug development, patient treatment, epidemiology, and socioeconomics[1].

Even though the 'golden standard' for COVID-19 diagnosis is RT-PCR (reverse transcription-polymerase chain reaction) test, the radiological screening, such as lungs CT scans or lungs X-ray, can help to monitor the disease and quickly isolate infected people. However, the increased number of COVID-19 patients and the need for manual analysis of chest X-ray imaging imposed a significant burden on medical staff. Therefore, intelligent technologies could improve greatly the disease's diagnosis procedures. X-ray analysis may be reported as less accurate than CT scans but at the same time X-ray scanning is less expensive and data processing is less computationally demanding. An automated system of COVID-19 diagnostics based on X-ray scans could work continuously, analyzing input data relatively fast and without breaks. As the result, such a system could significantly accelerate the diagnostic process for COVID-19 patients and keep it cost effective.

The aim of this paper were as follows:

- to implement a baseline transfer learning schema of processing the lung X-ray images in order to detect the COVID-19 symptoms;
- to test different scenarios of augmentation and evaluate them in terms of obtained improvements in evaluation metrics, such as Accuracy, Precision, Recall, and F1-score;
- to use augmentation scenarios in two modes: with and without segmentation and to assess the influence of segmentation on the model effectiveness;
- to validate the proposed system on a dataset containing real data obtained from the hospital (COVID-19 or healthy lung X-ray images confirmed with a RT-PCR test);
- to compare the obtained results to other, state-of-the-art analytical algorithms.

In this research we focused on the influence of augmentation on the performance of lung X-rays classification. Augmentation can help in overcoming the limitation of data samples in particular image datasets. Khalifa et al.[2] describe the following advantages of augmentation: 1) It can be an inexpensive way of gathering more data when compared with regular data collection with its label annotation; 2) It can be very accurate, as it is originally generated from ground-truth data; 3) It can be controllable, so that it is possible to generate well balanced data; 4) It can help in overcoming the overfitting problem; 5) It can provide better testing accuracy.

Materials and methods

In this study, we utilized a dataset that can be found on the website https://www.kaggle.com/datasets/andyczhao/covidx-cxr2. The example of use and detailed description were provided by Wang et al.[3]. The dataset included images obtained from various sources such as the GitHub repositories and Open Radiology Database (RICORD). All images were anonymized. The images were divided into two categories: COVID-19, and normal, and all of them were postero anterior (PA) chest X-rays. A total of 30,386 images were used in the experiment. The dataset can be treated as balanced because there were ~16,000 samples in 'COVID-19 positive' class and ~14,000 samples in 'COVID-19 negative' class.

Some examples of images from the dataset and the general overview of the proposed method are presented in Figure 1. It shows the following steps of the proposed pipeline: augmentation, pre-processing (normalization and masking by ResNet34), and classification using ResNet18's pretrained convolutional neural network. Finally, the proposed method gives the answer of `true' for the COVID-19-positive sample and `false' for the healthy sample. Each step of the process is described in detail in paragraphs below. All analyses were carried out with the use of Python 3.7 and the PyTorch platform. It is worth noting that in pre-processing part masking is marked with dashed line. It is to emphasize that we performed some experiments with and some without masking.

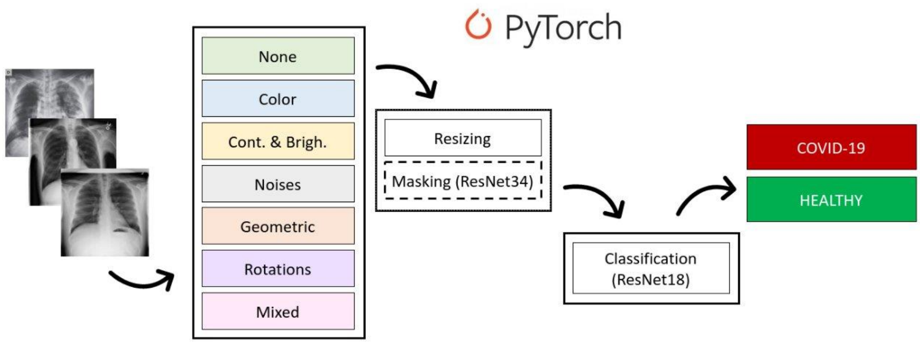

Figure 1 Pipeline of the proposed architecture: images from the dataset, 7 options of augmentation, pre-processing (resizing and masking) and classification performed by ResNet18 and finally the result: positive or negative.

Data augmentation and pre-processing

We proposed a baseline system to determine the impact of the chosen data augmentation method on the final classification result. Data augmentation is an important step in image analysis as it allows for an increase in the size of the dataset, thus, it could possibly contribute to the improvement of the model evaluation metrics. The augmentation can also improve the model's ability to generalize. In our study, we evaluated seven different data augmentation approaches:

- None — no augmentation methods were used in the baseline approach.
- Group 1 — manipulating the color of the image; this group of methods consists of:
  - RandomGamma - applying random gamma correction to the image to change the overall brightness.
  - ColorJitter - applying random changes in brightness, contrast, saturation, and hue to the image.
  - ToGray - converting the image to grayscale.
- Group 2 — manipulating the contrast and brightness of the image; this group of methods consists of:
  - CLAHE (Contrast Limited Adaptive Histogram Equalization) - adjusting the image intensity to improve the contrast and visibility of the lung structures.
  - RandomBrightness - adjusting the brightness of an image by a random amount.

- RandomContrast - adjusting the contrast of an image by a random amount.
- Sharpen - sharpening the image to increase its contrast and highlight details.

- Group 3 — this group of methods adds noises to the image. All used parameters of the noises were set experimentally:
  - MotionBlur - adding blur to the image to simulate motion blur; blur limit was set to 5.
  - MedianBlur - blurring the image by replacing each pixel's value with the median value of the pixels in its neighborhood, blur limit was set to 3.

  - OpticalDistortion - applying distortion to the image to simulate lens distortion.
  - GridDistortion - applying a grid distortion to the image, simulating distortions that can occur in images captured through a grid or mesh.
- Group 5 — rotating an image by a fixed angle to simulate different orientations of the image; images were rotated by angle in range $<-3,3>$ expressed in degrees.
  - Mixed — a mix of all mentioned augmentation methods.

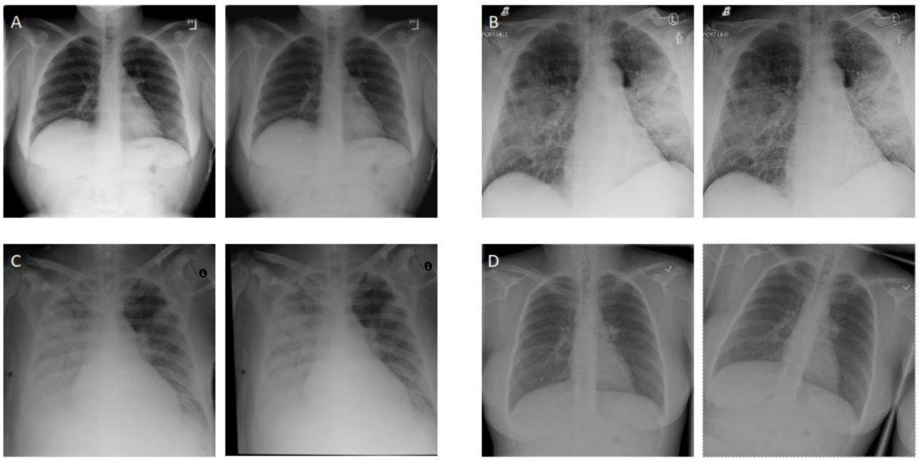

Figure 2 Examples of augmentation techniques. A: baseline and modified images G1, B: baseline and modified images G2, C: baseline and modified images G4, D: baseline and modified images G5.

  - Blur - blurring the image using a box filter, blur limit was set to 4.
  - GaussianBlur - blurring the image using a Gaussian filter with kernel (3,7);
- Group 4 — this group of methods applied geometric transformations to the image:
  - ElasticTransform - applying a non-rigid deformation to the image using displacement fields.

All the used methods come from the albumentations library[4]. Models for each of the seven groups were trained and validated independently. At each epoch, one augmentation was randomly (with equal probability) selected from among those available in the group. Only the training data was augmented. Examples of images created by augmentation are presented in Fig.2. In the forementioned Fig.2 baseline images with their modifications from selected group (G1 - color modification, G2 - contrast and brightness modification, G4 - geometric operations, G5 - rotations) are given. As visible, the differences between the baseline and the modified image are sometimes difficult to be observed with human eye. But, for the computer vision and understanding they are sufficiently different.

In view of studying the impact of augmentation methods, we decided to limit pre-processing methods to resize and apply masks with use of pretrained segmentation model ResNet34[5]. However, in order to evaluate the influence of masking on the classification metrics, we decided to run all experiment twice: with and without segmentation.

ML-based methods

For classification task the CNN was implemented. In order to focus on augmentation point of the research, pretrained CNN was used - ResNet18[5]. ResNet is a type of CNN, which popularity continously increases, also in COVID-19 detection from X-rays[3], [6]–[8]. The whole dataset (14,191 images representing class healthy and 16,194 images representing class COVID-19) were shuffled and divided into training and validation subsets. Some more detailed experimentally set learning parameters were:

- optimizer - SGD (stochastic gradient descent);
- loss function - cross entropy;
- number of epochs - 200;
- batch size - 16;
- early stopping rounds - 10.

Method's evaluation

The method was evaluated on a pre-prepared hospital dataset available at https://github.com/UTP-WTIiE/Xray_data.git and previously described and used in[9]. Images from this dataset are anonymized realistic data. They were obtained from Antoni Jurasz University Hospital No. 1 in Bydgoszcz, Department of Radiology and Diagnostic Imaging. A total of 62 chest X-ray images were obtained; 30 came from healthy individuals, and 32 came from COVID-19 positive patients, confirmed by a RT-PCR test. The images were provided in a raw form, without masks. The dataset was introduced and described in[9]. Each model in this research was evaluated using four validation metrics as follows: Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), and F1-score (Eq. 4), which use the below mentioned measures TP, FP, FN, and TN. These metrics can be treated as a golden standard in ML-based studies. They can help also in comparison of the proposed method to the state-of-the-are results.

- TP - true positives - COVID-19 patient classified as sick;
- FP – false positives - COVID-19 patient classified as healthy;
- FN - false negatives - healthy patient classified as sick;
- TN - true negatives - healthy patient classified as healthy.

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

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

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

$$F1 - score = 2\frac{precision \times recall}{precision + recall} \quad (4)$$

Results

All experiments were performed using the Nvidia Tesla with GPU support. Thanks to its enormous computing power, low price, relatively low demand for electricity, and the CUDA environment support, Tesla systems have become an attractive alternative to traditional high-power computing systems, such as CPU clusters and supercomputers. This kind of devices can be extremely helpful in image processing especially in medicine diagnostics.

The results obtained from all augmentation experiments are provided in Table 1. All the evaluated metrics are given: accuracy, precision, recall, and F1-score. Clearly, the augmentations can improve evaluation metrics. For example F1 metrics value raised from 87.5% (no augmentation, non-masked) to over 95% (mixed augmentations, non-masked). None of the proposed augmentation schema resulted in lowering the evaluation metrics. The most promising scenario both for masked and non-masked images was the last one i.e. with mixed augmentations. It is also noteworthy that masking can significantly improve the evaluation metrics (raising F1 from 95.2% to 98.5%).

Table 1 The results obtained for the selected augmentation methods. The most promising results, achieved on masked data with mixed augmentation, are highlighted in bold.

Augmentation	Masked data				Non-masked data			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
None	0.9032	0. 9063	0.9063	0.9063	0.8710	0.8750	0.8750	0.8750
Rotation	0.9516	0.9677	0.9375	0.9524	0.9194	0.9355	0.9063	0.9026
Color	0.9516	0.9677	0.9375	0.9524	0.9194	0.9355	0.9063	0.9026
Contrast and Brightness	0.9355	0.9667	0.9063	0.9355	0.8871	0.9032	0.8750	0.8889
Geometric	0.9516	0.9677	0.9375	0.9524	0.9194	0.9355	0.9063	0.9026
Noises	0.9516	0.9677	0.9375	0.9524	0.9194	0.9355	0.9063	0.9026
Mixed	**0.9839**	**0.9697**	**1.0000**	**0.9846**	0.9516	0.9677	0.9375	0.9524

Discussion

The augmentation methods can be very important element of the data pre-processing in the image analysis system that improves the obtained results[10]. In this paper we presented the baseline schema of COVID-19 detection on lung X-ray images and improved it by proposing very powerful augmentation technique. However, it should be mentioned that the utility of the augmentations vary, thus making some augmentations more and some less useful. In the proposed schema the most promising approach was to join all described groups and to implement them together.

In general, in the image analysis domain the augmentation can be performed by some image processing methods (classical augmentation) or by machine learning (ML) techniques e.g. GAN (Generative Adversarial Network). First approach is further less complicated computationally but surprisingly effective. GAN-based augmentation was described by Bargshady et al.[11]. In this paper the CycleGAN architecture performed an image-to-image translation. Then, the whole augmented dataset is used for training the finetuned, pretrained Inception V3 network resulting in an accuracy over 94%[11].

Another key AI-based element in our research is transfer learning. This is an approach extremely useful for image classification. It can be very powerful when the dataset is not sufficiently big. Moreover, using transfer learning allow creation of a very complicated model without extreme computations. Transfer learning uses pretrained network, making the learning process far shorter.

As described by Dogan et al.[12], three most excessively used ML-based architectures in researches concerning COVID-19 were: convolutional neural networks (CNN), random forest (RF), and ResNet. Whereas for large-scale image classification problems the most often used architectures were pre-trained networks: ResNet, UNet, VGG, Xception, GoogleNet, and XGBoost classifier.

Khan et al.[13] proposed the Deep Boosted Hybrid Learning (DBHL) architecture for effective COVID-19 detection in X-ray lung images. This approach used transfer learning and augmentation. The proposed framework was evaluated on radiologists' authenticated chest X-ray data with satisfying results (accuracy=98.5%, F1-score=98.0%, and precision=98.0%).

Rahman et al.[14] evaluated the importance of the pre-processing step of the ML-based system. Various transfer learning approaches (e.g. Resnet, DenseNet, InceptionV3) were compared with image enhancement methods, such as: histogram equalization, contrast limited adaptive histogram equalization (CLAHE), image inversion, Gamma correction, balance contrast enhancement technique (BCET). Since the used dataset was not balanced, the augmentation was implemented. For segmentation issue the U-net architecture was used. It was observed that DenseNet201 was the best performing network for the segmented lung CXR images in COVID-19 detection using gamma-corrected samples. The network achieved accuracy=95.11%, precision=94.55%, recall=94.56%, and F1-score=94.53%.

Motamed et al.[15] used GAN (IAGAN and DCGAN) for augmentation and classification, on a dataset from GitHub/IEEE and a second dataset of images of patients with pneumonia. The authors did classification including three classes (healthy, pneumonia, COVID-19). For comparison, the authors performed standard augmentations using random rotations in the range of 20 degrees, width and height shift in the range of 0.2 and zoom in the range of 0.2. In this way, eight new images each were generated, augmenting the dataset. On the COVID-19 dataset, the best ROC score obtained using IAGAN was 0.76, while baseline was 0.74, and typical augmentation was 0.75. The approach presented by the authors therefore allowed a slight improvement in the results.

Nishio et al.[16] presented a classification method which use a pretrained VGG-16. The authors utilised an optimal combination of the three types of data augmentation methods (conventional method, mixup, and RICAP). Similarly to the above mentioned studies, the dataset included X-ray images that were derived from patients representing three classes: healthy patients, COVID-19 patients, and patients with pneumonia. The authors achieved solid results, with accuracy equal to 83.68% on testing data.

Sakib et al.[17] developed a custom CNN model. It was trained on real data and augmented data. The suggested DARI (data augmentation of radiology images) algorithm creates artificial X-ray pictures through the use of a combination of a specialized GAN structure and common data augmentation methods like zooming and rotation, which are chosen adaptively. The proposed solution achieves promising results: accuracy = 94.3%, precision = 95.3%, recall = 97.8%, and F1-score =96.5%.

Narin A. et al.[18] compared some CNN-based models: ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2 for different binary classification issues: COVID-19 vs. Viral Pneumonia, COVID-19 vs. Bacterial Pneumonia, and COVID-19 vs. healthy. The dataset was unbalanced and contained only 341 COVID-19 samples (80% for training and 20% for testing). Authors reported ResNet50 as most promising for COVID-19 vs. normal classification with results: accuracy=96.1% and F1-score=83.5%.

Ozturk et al. proposed a novel deep model, called DarkCovidNet, for early detection of COVID-19 cases using X-ray images[19]. In this approach a Darknet-19 model was used as a baseline. The proposed network had fewer layers and filters as compared to the original DarkNet. Even though the dataset was limited, Authors used neither the augmentation, nor the pre-processing. The obtained results were the following: sensitivity = 95.13%, specificity = 95.30%, and F1-score = 96.51%.

In Table 2 we presented some results obtained from a literature review. The table contains summarized results from several papers from the tine period 2020-2022 and their most promising proposed architecture. Due to the fact that not all authors provided accuracy, precision, recall, and F1-score, there is only the accuracy presented in the table. The comparison proves that the approach proposed in this paper is competitive compared to the other state-of-the-art solutions previously proposed. It is also possible that if the augmentation techniques were used in these approaches, their results could be more remarkable. Furthermore, Table 2 shows that the transfer learning technique is most often used in case of X-ray images classification.

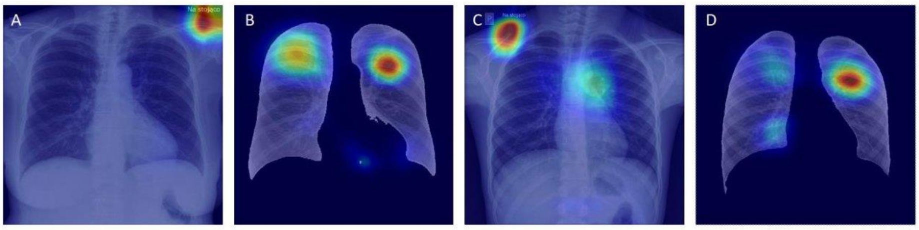

Figure 3 xAI examples. A: a heatmap of non-masked image; B: a heatmap of masked A image; C: a heatmap of non-masked image; D: a heatmap of masked C image.

Table 2 Accuracy comparison between the proposed and SOTA methods.

Reference	Architecture	Augmenta-tion	Result
proposed	double ResNet	✓	Acc.=98.39%
Bargshady G. et al.[11]	Inception V3	✓	Acc.=94.20%
Khan S.H. et al.[13]	DBHL	✓	Acc.=98.53%
Rahman T. et al.[14]	DenseNet201	✓	Acc.=95.11%
Nishio et al.[16]	VGG-16	✓	Acc.=83.68%
Narin A. et al.[18]	ResNet50	×	Acc.=96.10%
Ozturk et al.[19]	DarkCovidNet	×	Acc.=98.08
Wang L. et al.[20]	COVID-Net	×	Acc.=93.30%

Apart from numeric metrics for model evaluation, it is crucial to introduce some explainability to the ML-based system[21], [22]. Although AI models have achieved human-like performance, their use is still limited, partly because they are seen as a black box[23], [24]. As presented by Jia et al.[25] the explainability in an emerging issue particularly in ML-based healthcare systems. The problem with the use of AI-based tools in medicine continues to be the lack of confidence of medical professionals in such solutions and the perception that they lack the 'intuition' that experienced professionals possess[26], [27]. Authors emphasized the role of explainability and its potential implementations: explanation by approximation, explanation by example, feature relevance explanation, and visual explanation. In our research we focused on visual explanation. In Fig. 3 some examples of results obtained for selected samples samples are visually presented as a heatmap. These images show the points which gathered more network's attention. As presented in A section of the figure the main focus points are placed outside the lungs. It is significantly improved in section B of the figure. It should be noticed that the classifier focused on points inside the lung - there are some patterns suggesting lung changes caused by COVID-19. Similar situation is visualized in sections C and D of the figure: in section C the classifier focused more on points outside the lungs, in section D the focus points were improved.

Currently, there are two possible future improvements of the proposed schema. The first one is further explainability improvement so that the medical personnel could increase their trust the AI's predictions. However, this issue is very (technically, mentally, and legally) demanding and probably, it is not possible to achieve in 2 years perspective. Second possible way of improving the proposed schema would be focusing on complexity reduction of the proposed schema. This would allow for reduction of energy consumption and decrease the carbon footprint of the performed calculations, which is still significant even though modern computers are extremely fast.

Data availability

The dataset analysed during the current study are available in the github repository at https://github.com/UTP-WTIiE/Xray_data.git} and Kaggle website https://www.kaggle.com/datasets/andyczhao/covidx-cxr2.

Author contributions statement

A.K. prepared the conception or design of the work, created methods used in the work and prepared the manuscript. M.T. created methods used in the work and revised the manuscript A.G. performed the interpretation of data and prepared the manuscript. S.M.K. prepared the conception or design of the work, analyzed the data and prepared the manuscript. A.M. performed the data acquisition and revised the manuscript. Z.S. performed the data acquisition and revised the manuscript. M.W. prepared the conception or design of the work and revised the manuscript. All authors reviewed the manuscript.

Additional information

Authors do not declare any conflicts of interests.

Bibliography

[1] M. N. Islam, T. T. Inan, S. Rafi, S. S. Akter, I. H. Sarker, and A. K. M. N. Islam, "A Systematic Review on the Use of AI and ML for Fighting the COVID-19 Pandemic," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 258–270, 2020, doi: 10.1109/TAI.2021.3062771.

[2] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2351–2377, 2022, doi: 10.1007/s10462-021-10066-4.

[3] H. Chen, T. Zhang, R. Chen, Z. Zhu, and X. Wang, "A Novel COVID-19 Image Classification Method Based on the Improved Residual Network," *Electronics*, vol. 12, no. 1. 2023, doi: 10.3390/electronics12010080.

[4] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2. 2020, doi: 10.3390/info11020125.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[6] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *arXiv Prepr. arXiv2003.14395*, 2020.

[7] S. Showkat and S. Qureshi, "Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia," *Chemom. Intell. Lab. Syst.*, vol. 224, p. 104534, 2022, doi: https://doi.org/10.1016/j.chemolab.2022.104534.

[8] S. Rajpal, N. Lakhyani, A. K. Singh, R. Kohli, and N. Kumar, "Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images," *Chaos, Solitons & Fractals*, vol. 145, p. 110749, 2021, doi: https://doi.org/10.1016/j.chaos.2021.110749.

[9] A. Giełczyk *et al.*, "A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images," *Journal of Clinical Medicine*, vol. 11, no. 19. 2022, doi: 10.3390/jcm11195501.

[10] A. H. Barshooi and A. Amirkhani, "A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-Ray images.," *Biomed. Signal Process. Control*, vol. 72,

p. 103326, Feb. 2022, doi: 10.1016/j.bspc.2021.103326.

[11] G. Bargshady, X. Zhou, P. D. Barua, R. Gururajan, Y. Li, and U. R. Acharya, "Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-ray images.," *Pattern Recognit. Lett.*, vol. 153, pp. 67–74, Jan. 2022, doi: 10.1016/j.patrec.2021.11.020.

[12] O. Dogan, S. Tiwari, M. A. Jabbar, and S. Guggari, "A systematic review on AI/ML approaches against COVID-19 outbreak," *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2655–2678, 2021, doi: 10.1007/s40747-021-00424-8.

[13] S. H. Khan *et al.*, "COVID-19 detection in chest X-ray images using deep boosted hybrid learning," *Comput. Biol. Med.*, vol. 137, p. 104816, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104816.

[14] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, p. 104319, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104319.

[15] S. Motamed, P. Rogalla, and F. Khalvati, "Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images," *Informatics Med. Unlocked*, vol. 27, p. 100779, 2021, doi: https://doi.org/10.1016/j.imu.2021.100779.

[16] M. Nishio, S. Noguchi, H. Matsuo, and T. Murakami, "Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods," *Sci. Rep.*, vol. 10, no. 1, p. 17532, 2020, doi: 10.1038/s41598-020-74539-2.

[17] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and M. Guizani, "DL-CRC: Deep Learning-Based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach," *IEEE Access*, vol. 8, pp. 171575–171589, 2020, doi: 10.1109/ACCESS.2020.3025010.

[18] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, 2021, doi: 10.1007/s10044-021-00984-y.

[19] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, 2020, doi: https://doi.org/10.1016/j.compbiomed.2020.103792.

[20] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, p. 19549, Nov. 2020, doi: 10.1038/s41598-020-76550-z.

[21] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, and the P. consortium, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 310, 2020, doi: 10.1186/s12911-020-01332-6.

[22] S. Reddy, "Explainability and artificial intelligence in medicine," *Lancet Digit. Heal.*, vol. 4, no. 4, pp. e214–e215, Apr. 2022, doi: 10.1016/S2589-7500(22)00029-2.

[23] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Comput. Methods Programs Biomed.*, vol. 226, p. 107161, 2022, doi:

https://doi.org/10.1016/j.cmpb.2022.107161.

[24] M. Choraś, M. Pawlicki, D. Puchalski, and R. Kozik, “Machine Learning – The Results Are Not the only Thing that Matters! What About Security, Explainability and Fairness? BT - Computational Science – ICCS 2020,” 2020, pp. 615–628.

[25] Y. Jia, J. McDermid, T. Lawton, and I. Habli, “The role of explainability in assuring safety of machine learning in healthcare,” *IEEE Trans. Emerg. Top. Comput.*, vol. 10, no. 4, pp. 1746–1760, 2022.

[26] S. Kundu, “AI in medicine must be explainable,” *Nat. Med.*, vol. 27, no. 8, p. 1328, 2021, doi: 10.1038/s41591-021-01461-z.

[27] T. P. Quinn, M. Senadeera, S. Jacobs, S. Coghlan, and V. Le, “Trust and medical AI: the challenges we face and the expertise needed to overcome them,” *J. Am. Med. Informatics Assoc.*, vol. 28, no. 4, pp. 890–894, Apr. 2021, doi: 10.1093/jamia/ocaa268.

7. Conclusions

The COVID-19 pandemic has posed a significant threat to global public health, with the need for accurate and efficient diagnosis becoming more pressing than ever before. In my research, I have successfully investigated the feasibility and effectiveness of using AI models for diagnostic classification of COVID-19 patients based on their chest X-ray images. The primary objective was to develop deep learning models that could accurately distinguish between COVID-19 positive and negative cases, and I have achieved this goal with high levels of accuracy, sensitivity, specificity, and AUC.

My studies have also addressed several partial objectives related to the impact of augmentation methods and chest X-ray images pre-processing on the classification abilities of the models. I have found that applying data augmentation techniques such as rotation, flipping, and scaling can improve the performance of the models. Furthermore, I have shown that certain pre-processing techniques, such as contrast enhancement and histogram equalization, can also have a positive impact on classification accuracy.

In addition, my studies have explored the use of deep learning models for feature extraction from chest X-ray images, and I have compared the classification abilities of tree-based models, such as XGBoost and LightGBM. I have found that deep learning models can extract informative features from chest X-ray images, which can be used to train tree-based models that can classify COVID-19 patients with high levels of accuracy.

Finally, my studies have also addressed the ethical considerations surrounding the use of AI in clinical image diagnosis. I have discussed the potential benefits and drawbacks of relying on ML models for medical decision-making and raised important questions about the responsibility for the decisions made by these models. The obtained results may inform future discussions and policies regarding the appropriate use of AI in medical contexts, especially in cases where human decision-making may be biased or subject to error.

In my research I have successfully achieved all of its objectives and has demonstrated the potential of using AI models for diagnostic classification of COVID-19 patients based on their chest X-ray images. My findings have significant implications for the development of accurate and efficient diagnostic tools for COVID-19 and other medical conditions, and for the ethical considerations surrounding the use of AI in medical contexts. I hope that my study will contribute to the ongoing efforts to improve public health and healthcare delivery through the responsible and effective use of AI.

The potential future continuation of this study could focus on several aspects to further improve the accuracy and effectiveness of using AI models as medical devices to support physicians in analyzing X-ray images. One important area of focus could be the development of more advanced and sophisticated deep learning models that can better handle the complexities and variations in chest X-ray images. For example, researchers could explore the use of more advanced neural network architectures, such as capsule networks or attention mechanisms, to improve feature extraction and classification performance.

Another area of focus could be the collection and analysis of larger and more diverse datasets to improve the robustness and generalizability of the models. The current study used a relatively small dataset of chest X-ray images, and future research could benefit from using larger datasets that are more representative of the population.

In addition, future research could also focus on addressing some of the ethical concerns related to the use of AI as medical devices. For example, researchers could explore the development of explainable AI models that can provide transparent and interpretable results to support clinical decision-making. They could also investigate ways to ensure that the use of AI models does not lead to bias or discrimination against certain groups of patients.

Despite some of the challenges and ethical considerations surrounding the use of AI in medical contexts, there are many potential benefits to using AI as medical devices that support physicians in analyzing X-ray images. AI models can provide rapid and accurate diagnosis, reduce the workload of physicians, and improve the efficiency of healthcare delivery. They can also assist physicians in making more informed and data-driven decisions, leading to better patient outcomes.

In conclusion, the potential future continuation of this study could focus on improving the accuracy and effectiveness of using AI models as medical devices to support physicians in analyzing X-ray images. As AI technology continues to evolve and improve, there is great potential for AI models to revolutionize the field of medical imaging and improve patient outcomes. However, it is important to carefully consider the ethical implications of using AI in healthcare and to ensure that these technologies are developed and deployed responsibly and ethically.

8. References

1. Jain DK, Singh T, Saurabh P, Bisen D, Sahu N, Mishra J, et al. Deep Learning-Aided Automated Pneumonia Detection and Classification Using CXR Scans. Comput Intell Neurosci. 2022;2022.

2. Lamia A, Fawaz A. Detection of Pneumonia Infection by Using Deep Learning on a Mobile Platform. Comput Intell Neurosci. 2022;2022.

3. Singh S. Pneumonia Detection using Deep Learning. In: 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). 2021. p. 1–6.

4. Nijiati M, Zhang Z, Abulizi A, Miao H, Tuluhong A, Quan S, et al. Deep learning assistance for tuberculosis diagnosis with chest radiography in low-resource settings. J Xray Sci Technol. 2021;29(5):785–96.

5. Yoo H, Lee SH, Arru CD, Doda Khera R, Singh R, Siebert S, et al. AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset. Eur Radiol. 2021;31(12):9664–74.

6. Jain G, Mittal D, Thakur D, Mittal MK. A deep learning approach to detect Covid-19 coronavirus with X-Ray images. Biocybern Biomed Eng [Internet]. 2020;40(4):1391–405. Available from: https://doi.org/10.1016/j.bbe.2020.08.008

7. Lohikoski K, Roos M, Suominen T. Workplace culture assessed by radiographers in Finland. Radiography [Internet]. 2019 Nov 1;25(4):e113–8. Available from: https://doi.org/10.1016/j.radi.2019.05.003

8. Hutton D, Beardmore C, Patel I, Massey J, Wong H, Probst H. Audit of the job satisfaction levels of the UK radiography and physics workforce in UK radiotherapy centres 2012. Br J Radiol [Internet]. 2014;87(1039):20130742. Available from: https://doi.org/10.1259/bjr.20130742

9. Singh N, Wright C, Knight K, Baird M, Akroyd D, Adams RD, et al. Occupational burnout among radiation therapists in Australia: Findings from a mixed methods study. Radiography [Internet]. 2017 Aug 1;23(3):216–21. Available from: https://doi.org/10.1016/j.radi.2017.03.016

10. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long Radiology Workdays Reduce Detection and Accommodation Accuracy. J Am Coll Radiol [Internet]. 2010;7(9):698–704. Available from: https://www.sciencedirect.com/science/article/pii/S1546144010001341

11. Vertinsky T, Forster B. Prevalence of eye strain among radiologists: Influence of viewing variables on symptoms. Am J Roentgenol. 2005;184(2):681–6.

12. Hassantabar S, Ahmadi M, Sharifi A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches. Chaos, Solitons and Fractals [Internet]. 2020;140:110170. Available from: https://doi.org/10.1016/j.chaos.2020.110170

13. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. Interdiscip Sci –

Comput Life Sci [Internet]. 2021;13(1):103–17. Available from: https://doi.org/10.1007/s12539-020-00403-6

14. Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: Impact on the radiography profession. Br J Radiol. 2020;93(1108).

15. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med [Internet]. 2020;121(April):103792. Available from: https://doi.org/10.1016/j.compbiomed.2020.103792

16. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak [Internet]. 2020;20(1):1–9. Available from: https://doi.org/10.1186/s12911-020-01332-6

17. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer [Internet]. 2018;18(8):500–10. Available from: https://doi.org/10.1038/s41568-018-0016-5

18. Ahn SY, Chae KJ, Goo JM. The Potential Role of Grid-Like Software in Bedside Chest Radiography in Improving Image Quality and Dose Reduction: An Observer Preference Study. Korean J Radiol [Internet]. 2018 Jun;19(3):526–33. Available from: https://doi.org/10.3348/kjr.2018.19.3.526

19. Syed Adam C. AB. Z. Artificial Intelligence in Radiology: Current Technology and Future Directions. Semin Musculoskelet Radiol [Internet]. 2018;22(05):540–5. Available from: http://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0038-1673383

20. Ker J, Wang L, Rao J, Lim T. Deep Learning Applications in Medical Image Analysis. IEEE Access. 2017;6:9375–9.

21. Civit-Masot J, Luna-Perejón F, Morales MD, Civit A. Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images. Appl Sci. 2020;10(13).

22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115–8.

23. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol [Internet]. 2018 Aug 1;29(8):1836–42. Available from: https://doi.org/10.1093/annonc/mdy166

24. Killock D. AI outperforms radiologists in mammographic screening. Nat Rev Clin Oncol [Internet]. 2020;17(3):134. Available from: https://doi.org/10.1038/s41571-020-0329-7

25. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Med [Internet]. 2018;15(11):1–17. Available from:

https://doi.org/10.1371/journal.pmed.1002686

26. Tahir AM, Chowdhury MEH, Khandakar A, Rahman T, Qiblawey Y, Khurshid U, et al. COVID-19 infection localization and severity grading from chest X-ray images. Comput Biol Med [Internet]. 2021;139(October):105002. Available from: https://doi.org/10.1016/j.compbiomed.2021.105002

27. Urzędowy D, Polskiej R. UCHWAŁA NR 196 RADY MINISTRÓW z dnia 28 grudnia 2020 r. w sprawie ustanowienia „Polityki dla rozwoju sztucznej inteligencji w Polsce od roku 2020". Monit Pol. 2021;poz 23(196):1–72.

28. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say : susceptibility in deployment of clinical decision-aids. npj Digit Med [Internet]. 2021; Available from: http://dx.doi.org/10.1038/s41746-021-00385-9

29. Jacobs M, Pradier MF, Jr THM, Perlis RH, Doshi-velez F, Gajos KZ. How machine-learning recommendations in fl uence clinician treatment selections : the example of antidepressant selection. Transl Psychiatry [Internet]. 2021; Available from: http://dx.doi.org/10.1038/s41398-021-01224-x

30. Brady AP, Neri E. Artificial Intelligence in Radiology—Ethical Considerations. Diagnostics [Internet]. 2020;10(4). Available from: https://www.mdpi.com/2075-4418/10/4/231

31. Geijer H, Geijer M. Added value of double reading in diagnostic radiology,a systematic review. Insights Imaging. 2018 Jun;9(3):287–301.

32. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med [Internet]. 2022;28(1):31–8. Available from: https://doi.org/10.1038/s41591-021-01614-0

33. Artificial intelligence and medical imaging 2018: French Radiology Community white paper. Diagn Interv Imaging. 2018 Nov;99(11):727–42.

34. Khan SH, Sohail A, Khan A, Hassan M, Lee YS, Alam J, et al. COVID-19 detection in chest X-ray images using deep boosted hybrid learning. Comput Biol Med [Internet]. 2021;137(July):104816. Available from: https://doi.org/10.1016/j.compbiomed.2021.104816

35. Ouchicha C, Ammor O, Meknassi M. CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. Chaos, Solitons and Fractals. 2020;140.

36. Padma T, Usha Kumari C. Deep Learning Based Chest X-Ray Image as a Diagnostic Tool for COVID-19. Proc - Int Conf Smart Electron Commun ICOSEC 2020. 2020;(Icosec 2020):589–92.

37. Dogan O, Tiwari S, Jabbar MA, Guggari S. A systematic review on AI/ML approaches against COVID-19 outbreak. Complex Intell Syst [Internet]. 2021/07/05. 2021;7(5):2655–78. Available from: https://pubmed.ncbi.nlm.nih.gov/34777970

38. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET). 2017. p. 1–6.

39. Bassi PRAS, Attux R. A deep convolutional neural network for COVID-19 detection using chest X-rays. Res Biomed Eng. 2022;38(1):139–48.

40. Khanday NY, Sofi SA. Deep insight: Convolutional neural network and its applications for COVID-19 prognosis. Biomed Signal Process Control [Internet]. 2021;69(February):102814. Available from: https://doi.org/10.1016/j.bspc.2021.102814

41. Kochgaven C, Mishra P, Shitole S. Detecting Presence of COVID-19 with ResNet-18 using PyTorch. In: 2021 International Conference on Communication information and Computing Technology (ICCICT). 2021. p. 1–6.

42. Yadav A, Verma VK, Pal V, Singh S. Automatic Detection of COVID 19 Infection Using Deep Learning Models from X-Ray Images. IOP Conf Ser Mater Sci Eng [Internet]. 2021 Mar;1099(1):12050. Available from: https://dx.doi.org/10.1088/1757-899X/1099/1/012050

43. Liew XY, Hameed N, Clos J. An investigation of XGBoost-based algorithm for breast cancer classification. Mach Learn with Appl [Internet]. 2021;6:100154. Available from: https://www.sciencedirect.com/science/article/pii/S2666827021000773

44. Ogunleye A, Wang Q-G. XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Trans Comput Biol Bioinforma. 2020;17(6):2131–40.

45. Song R, Li T, Wang Y. Mammographic Classification Based on XGBoost and DCNN With Multi Features. IEEE Access. 2020;8:75011–21.

46. Ezzoddin M, Nasiri H, Dorrigiv M. Diagnosis of COVID-19 Cases from Chest X-ray Images Using Deep Neural Network and LightGBM. In: 2022 International Conference on Machine Vision and Image Processing (MVIP). 2022. p. 1–7.

47. Fangoh AM, Selim S. Using CNN-XGBoost Deep Networks for COVID-19 Detection in Chest X-ray Images. In: 2020 15th International Conference on Computer Engineering and Systems (ICCES). 2020. p. 1–7.

48. Giełczyk A, Marciniak A, Tarczewska M, Kloska SM, Harmoza A, Serafin Z, et al. A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images. J Clin Med. 2022;11(19):5501.

49. Nasiri H, Hasani S. Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost. Radiography [Internet]. 2022;28(3):732–8. Available from: https://doi.org/10.1016/j.radi.2022.03.011

50. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med Image Anal [Internet]. 2020;63:101693. Available from: https://www.sciencedirect.com/science/article/pii/S136184152030058X

51. Giełczyk A, Marciniak A, Tarczewska M, Lutowski Z. Pre-processing methods in chest X-ray image classification. PLoS One [Internet]. 2022;17(4 April):1–11. Available from: http://dx.doi.org/10.1371/journal.pone.0265949

52. Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the

performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. Int J Med Inform [Internet]. 2020;144:104284. Available from: https://www.sciencedirect.com/science/article/pii/S138650562030959X

53. Sakib S, Tazrin T, Fouda MM, Fadlullah ZM, Guizani M. DL-CRC: Deep Learning-Based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach. IEEE Access. 2020;8:171575–89.

54. Motamed S, Rogalla P, Khalvati F. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. Informatics Med Unlocked [Internet]. 2021;27:100779. Available from: https://www.sciencedirect.com/science/article/pii/S2352914821002501

55. Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses [Internet]. 2020;140:109761. Available from: https://www.sciencedirect.com/science/article/pii/S0306987720307702

56. Khalifa NE, Loey M, Mirjalili S. A comprehensive survey of recent trends in deep learning for digital images augmentation. Artif Intell Rev [Internet]. 2022;55(3):2351–77. Available from: https://doi.org/10.1007/s10462-021-10066-4

57. Vanschoren J. Meta-learning. Autom Mach Learn methods, Syst challenges. 2019;35–61.

58. Dietterich TG. Ensemble Methods in Machine Learning BT - Multiple Classifier Systems. In Berlin, Heidelberg: Springer Berlin Heidelberg; 2000. p. 1–15.

59. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep [Internet]. 2022;12(1):1–9. Available from: https://doi.org/10.1038/s41598-022-09954-8

60. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. Engineering [Internet]. 2020;6(10):1122–9. Available from: https://www.sciencedirect.com/science/article/pii/S2095809920301636

61. Sand M, Durán JM, Jongsma KR. Responsibility beyond design: Physicians' requirements for ethical medical AI. Bioethics [Internet]. 2022 Feb 1;36(2):162–9. Available from: https://doi.org/10.1111/bioe.12887

62. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics [Internet]. 2021 Dec 1;47(12):e3 LP-e3. Available from: http://jme.bmj.com/content/47/12/e3.abstract

63. Verdicchio M, Perin A. When Doctors and AI Interact: on Human Responsibility for Artificial Risks. Philos Technol [Internet]. 2022;35(1):11. Available from: https://doi.org/10.1007/s13347-022-00506-6

64. What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging. 2019 Apr;10(1):44.

65. Salvatore C, Interlenghi M, Monti CB, Ippolito D, Capra D, Cozzi A, et al. Artificial intelligence applied to chest x-ray for differential diagnosis of covid-19 pneumonia. Diagnostics. 2021;11(3):1–12.

66. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. npj Digit Med [Internet]. 2020;3(1):47. Available from: https://doi.org/10.1038/s41746-020-0254-2

67. Regulation P. Regulation (EU) 2016/679 of the European Parliament and of the Council. Regul. 2016;679:2016.

68. Jiang J (Xuefeng), Bai G. Types of Information Compromised in Breaches of Protected Health Information. Ann Intern Med [Internet]. 2019 Sep 24;172(2):159–60. Available from: https://www.acpjournals.org/doi/abs/10.7326/M19-1759

69. Kaye J. The Tension Between Data Sharing and the Protection of Privacy in Genomics Research. Annu Rev Genomics Hum Genet [Internet]. 2012 Sep 12;13(1):415–31. Available from: https://doi.org/10.1146/annurev-genom-082410-101454

70. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell [Internet]. 2020;2(6):305–11. Available from: http://dx.doi.org/10.1038/s42256-020-0186-1

71. Taylor MJ, Wilson J. Reasonable Expectations of Privacy and Disclosure of Health Data. Med Law Rev [Internet]. 2019 Aug 1;27(3):432–60. Available from: https://doi.org/10.1093/medlaw/fwz009

72. Cheu A, Smith A, Ullman J, Zeber D, Zhilyaev M. Distributed Differential Privacy via Shuffling. 1:1–30.

73. Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. Found Trends® Theor Comput Sci. 2014;9(3–4):211–407.

74. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol Gen. 2015;144(1):114–26.

75. Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. Organ Behav Hum Decis Process [Internet]. 2019;151:90–103. Available from: https://www.sciencedirect.com/science/article/pii/S0749597818303388

76. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Heal [Internet]. 2021 Nov 1;3(11):e745–50. Available from: https://doi.org/10.1016/S2589-7500(21)00208-9

77. Wang F, Kaushal R, Khullar D. Should Health Care Demand Interpretable Artificial Intelligence or Accept "Black Box" Medicine? Ann Intern Med [Internet]. 2019 Dec 17;172(1):59–60. Available from: https://www.acpjournals.org/doi/abs/10.7326/M19-2548

78. Winfield AFT, Jirotka M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos Trans Ser A, Math Phys Eng Sci. 2018

Oct;376(2133).

79. Gaube S, Suresh H, Raue M, Lermer E, Koch TK, Hudecek MFC, et al. Non - task expert physicians benefit from correct explainable AI advice when reviewing X - rays. Sci Rep [Internet]. 2023;1–11. Available from: https://doi.org/10.1038/s41598-023-28633-w

80. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.

81. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). 2017. p. 618–26.

82. Ribeiro MT, Guestrin C. “ Why Should I Trust You ?” Explaining the Predictions of Any Classifier. 2016;

83. Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. Sanity checks for saliency maps. Adv Neural Inf Process Syst. 2018;31.

9. Statements of co-authors of publications included in the series

9.1. Attachment No. 1

Agata Giełczyk, PhD Eng.
Faculty of Telecommunications, Computer Science and Electrical Engineering
Bydgoszcz University of Science and Technology

Statement

I hereby declare that in the paper "Pre-processing methods in chest X-ray image classification" published in the journal *PLOS One* in 2022; https://doi.org/10.1371/journal.pone.0265949, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, and Zbigniew Lutowski, my substantive contribution was to supervise the research and co-edit the paper (senior author).

I hereby declare that in the paper "A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution was to supervise the research and co-edit the paper (senior author).

I hereby declare that in the article „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in *Polish Journal of Radiology* in 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak my substantive contribution was to supervise the research and co-edit the paper (senior author).

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

Agata Giełczyk

dr inż. Agata Giełczyk
Wydział Telekomunikacji, Informatyki i Elektrotechniki
Politechnika Bydgoska im. J. i J. Śniadeckich

Oświadczenie

Oświadczam, że w artykule „Pre-processing methods in chest X-ray image classification" opublikowanym w czasopiśmie *PLOS One* w 2022r.; https://doi.org/10.1371/journal.pone.0265949, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska i Zbigniew Lutowski, mój wkład merytoryczny polegał na opiece nad badaniami i współredagowaniu artykułu (senior author).

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na opiece nad badaniami i współredagowaniu artykułu (senior author).

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak mój wkład merytoryczny polegał na opiece nad badaniami i współredagowaniu artykułu (senior author).

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Agata Giełczyk

9.2. Attachment No. 2

Inż. Martyna Taczewska
Wydział Telekomunikacji, Informatyki i Elektrotechniki
Politechnika Bydgoska im. J. i J. Śniadeckich w Bydgoszczy

Oświadczenie

Oświadczam, że w artykule „Pre-processing methods in chest X-ray image classification" opublikowanym w czasopiśmie *PLOS One* w 2022r.; https://doi.org/10.1371/journal.pone.0265949, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska i Zbigniew Lutowski, mój wkład merytoryczny polegał na przygotowaniu części programistycznej doświadczenia (przygotowanie fragmentów kodu).

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na przygotowaniu części programistycznej doświadczenia (przygotowanie fragmentów kodu).

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak mój wkład merytoryczny polegał na przygotowaniu części programistycznej doświadczenia (przygotowanie fragmentów kodu).

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Martyna Tarczewska
....

(Podpis)

Martyna Tarczewska, BSc
Faculty of Telecommunications, Computer Science and Electrical Engineering
Bydgoszcz University of Science and Technology

Statement

I hereby declare that in the paper "Pre-processing methods in chest X-ray image classification" published in the journal *PLOS One* in 2022; https://doi.org/10.1371/journal.pone.0265949, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, and Zbigniew Lutowski, my substantive contribution was to prepare the programming part of the experience (preparing code snippets).

I hereby declare that in the paper "A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution was to prepare the programming part of the experience (preparing code snippets).

I hereby declare that in the article „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in *Polish Journal of Radiology* in 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak my substantive contribution was to prepare the programming part of the experience (preparing code snippets)

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

Martyna Tarczewska
....

(Signature)

9.3. Attachment No. 3

Sylwester Kloska, M.Sc.
Faculty of Medicine
Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University in Toruń

Statement

I hereby declare that in the paper " A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution was to advise on substantive issues and co-editing the paper.

I hereby declare that in the article „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in *Polish Journal of Radiology* in 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak my substantive contribution was to advise on substantive issues and co-editing the paper.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

Sylwester Kloska

(Signature)

Mgr Sylwester Kloska
Wydział Lekarski
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

Oświadczenie

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na doradztwie w kwestiach merytorycznych i współredagowaniu artykułu.

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak mój wkład merytoryczny polegał na doradztwie w kwestiach merytorycznych i współredagowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloska jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Sylwester Kloska
(Podpis)

9.4. Attachment No. 4

dr hab. Marcin Woźniak, prof. UMK
Wydział Lekarski
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

Oświadczenie

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na konsultacji koncepcji doświadczenia oraz współredagowaniu artykułu.

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak mój wkład merytoryczny polegał na konsultacji koncepcji doświadczenia oraz współredagowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr inż. Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Marcin Woźniak
....
(Podpis)

Prof. Marcin Woźniak
Faculty of Medicine
Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University in Toruń

Statement

I hereby declare that in the paper " A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution was to consult the concept of the research and co-edit the paper.

I hereby declare that in the article „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in *Polish Journal of Radiology* in 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak my substantive contribution was to consult the concept of the research and co-edit the paper.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

....

(Signature)

9.5. Attachment No. 5

prof. dr hab. Zbigniew Serafin
Wydział Lekarski
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

Oświadczenie

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na doradztwie w kwestiach medycznych i współredagowaniu artykułu.

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak mój wkład merytoryczny polegał na doradztwie w kwestiach medycznych i współredagowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Serafin
....
(Podpis)

Kierownik
Katedry Radiologii i Diagnostyki Obrazowej
prof. dr hab. Zbigniew Serafin

Prof. Zbigniew Serafin
Faculty of Medicine
Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University in Toruń

Statement

I hereby declare that in the paper " A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution was counseling on medical issues and co-edit the paper.

I hereby declare that in the article „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in *Polish Journal of Radiology* in 2023r.; doi: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester Michał Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak my substantive contribution was counseling on medical issues and co-edit the paper.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

S.Seraf

....

(Signature)

Kierownik
Katedry Radiologii i Diagnostyki Obrazowej
prof. dr hab. Zbigniew Serafin

9.6. Attachment No. 6

Zbigniew Lutowski, PhD
Faculty of Telecommunications, Computer Science and Electrical Engineering
Bydgoszcz University of Science and Technology

Statement

I hereby declare that in the paper "Pre-processing methods in chest X-ray image classification" published in the journal *PLOS One* in 2022; https://doi.org/10.1371/journal.pone.0265949, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, and Zbigniew Lutowski, my substantive contribution was to supervise the research and co-edit the paper.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

Zbigniew Lutowski
(Signature)

dr inż. Zbigniew Lutowski
Wydział Telekomunikacji, Informatyki i Elektrotechniki
Politechnika Bydgoska im. J. i J. Śniadeckich w Bydgoszczy

Oświadczenie

Oświadczam, że w artykule „Pre-processing methods in chest X-ray image classification" opublikowanym w czasopiśmie *PLOS One* w 2022r.; https://doi.org/10.1371/journal.pone.0265949, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska i Zbigniew Lutowski, mój wkład merytoryczny polegał na opiece nad badaniami i współredagowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Zbigniew Lutowski
(Podpis)

9.7. Attachment No. 7

Adrian Michalski, MSc.
Faculty of Pharmacy
Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University in Toruń

Statement

I hereby declare that in the paper „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" published in the journal *Polish Journal of Radiology* in 2023r., DOI: https://doi.org/10.5114/pjr.2023.126717, whose co-authors are Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester M. Kloska, Adrian Michalski, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution consisted of advising on medical issues and editing the article.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

Adrian Michalski
....

(Signature)

mgr Adrian Michalski
Wydział Farmaceutyczny
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

Oświadczenie

Oświadczam, że w artykule „Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification" opublikowanym w czasopiśmie *Polish Journal of Radiology* w 2023r., DOI: https://doi.org/10.5114/pjr.2023.126717; którego współautorami są Anna Kloska, Martyna Tarczewska, Agata Giełczyk, Sylwester M. Kloska, Adrian Michalski, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na doradztwie w kwestiach medycznych i edytowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloska jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

Adrian Michalski
....
(Podpis)

9.8. Attachment No. 8

Alicja Harmoza, M.D.
Faculty of Medicine
Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University in Toruń

Statement

I hereby declare that in the paper " A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" published in the journal *Journal of Clinical Medicine* in 2022; https://doi.org/10.3390/jcm11195501, whose co-authors are Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin, and Marcin Woźniak, my substantive contribution consisted of advising on medical issues and editing the article.

At the same time, I agree to submit the above-mentioned publications by Anna Kloska, M.Sc., as part of a doctoral dissertation in the form of a thematically coherent collection of scientific papers published in scientific journals. I declare that the sharing of works will not infringe on the copyrights of third parties.

16.05.23r. Harmoza

(Signature)

Lek. Med. Alicja Harmoza
Wydział Lekarski
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

Oświadczenie

Oświadczam, że w artykule „A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images" opublikowanym w czasopiśmie *Journal of Clinical Medicine* w 2022r.; https://doi.org/10.3390/jcm11195501, którego współautorami są Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, Sylwester Michał Kloska, Alicja Harmoza, Zbigniew Serafin i Marcin Woźniak, mój wkład merytoryczny polegał na doradztwie w kwestiach medycznych i edytowaniu artykułu.

Jednocześnie wyrażam zgodę na przedłożenie w/w publikacji przez mgr Annę Kloskę jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych. Oświadczam, że udostępnienie utworów nie będzie naruszało praw autorskich osób trzecich.

15.05.23r Harmoza
(Podpis)

10.Consent of the bioethics committee

Uniwersytet Mikołaja Kopernika w Toruniu

Collegium Medicum im L. Rydygiera w Bydgoszczy

KOMISJA BIOETYCZNA

Ul. M. Skłodowskiej-Curie 9, 85-094 Bydgoszcz, tel.(052) 585-35-63, fax.(052) 585-38-11

KB 454/2022 Bydgoszcz, 23.08.2022 r.

Działając na podstawie art. 29 ustawy z dnia 5.12.1996 r. o zawodach lekarza i lekarza dentysty Dz.U. z 1997 r. Nr 28 poz. 152 (wraz z późniejszymi zmianami), rozporządzenia Ministra Zdrowia i Opieki Społecznej z dnia 11.05.1999 r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych (Dz.U. Nr 47 poz.480) oraz Zarządzenia Nr 21 Rektora UMK z dnia 4.03.2009 r. z późn. zm. w sprawie powołania oraz zasad działania Komisji Bioetycznej Uniwersytetu Mikołaja Kopernika w Toruniu przy Collegium Medicum im Ludwika Rydygiera w Bydgoszczy oraz zgodnie z zasadami zawartymi w DH i GCP

Komisja Bioetyczna przy UMK w Toruniu, Collegium Medicum w Bydgoszczy

(skład podano w załączeniu), na posiedzeniu w dniu **23.08.2022 r.** przeanalizowała wniosek, który złożył kierownik badania:

dr hab. n. med. Marcin Woźniak, prof. UMK
Katedra Medycyny Sądowej
Collegium Medicum w Bydgoszczy UMK w Toruniu

z zespołem w składzie:

- mgr inż. Anna Kloska, dr inż. Agata Giełczyk, dr hab. n. med. Marcin Woźniak, prof. UMK

w sprawie badania:

„Diagnostyka zakażenie wirusem SARS-CoV-2 na podstawie analizy zdjęć rentgenowskich przy użyciu metod sztucznej inteligencji."

Po zapoznaniu się ze złożonym wnioskiem i w wyniku przeprowadzonej dyskusji oraz głosowania Komisja podjęła:

Uchwałę o pozytywnym zaopiniowaniu wniosku

w sprawie przeprowadzenia badań w zakresie określonym we wniosku pod warunkiem uzyskania zgody osób badanych na przetwarzanie danych osobowych w celach naukowych, a w przypadku braku takiej zgody, analizowania jedynie danych zanonimizowanych, pozbawionych danych personalnych (zgodnie z RODO). Zgoda obejmuje tylko dane z dokumentacji uczestników badania z okresu od 01.01.2020 r. do 23.08.2022 r.

Zgoda obowiązuje od daty podjęcia uchwały (23.08.2022 r.) do końca 2023 r.

Wydana opinia dotyczy tylko rozpatrywanego wniosku z uwzględnieniem przedstawionego projektu; każda zmiana i modyfikacja wymaga uzyskania odrębnej opinii

Prof. dr hab. med. Karol Śliwka

Przewodniczący Komisji Bioetycznej

Otrzymuje:
dr hab. n. med. Marcin Woźniak, prof. UMK
Katedra Medycyny Sądowej
Collegium Medicum w Bydgoszczy UMK w Toruniu

Lista obecności

na posiedzeniu Komisji Bioetycznej

w dniu 23.08.2022 r.

Lp.	Imię i nazwisko	Funkcja/ Specjalizacja	Podpis
1.	Prof. dr hab. n. med. Karol Śliwka	*medycyna sądowa* przewodniczący	[illegible]
2.	Mgr prawa Joanna Połetek-Żygas	*prawniczka* zastępca przewodniczącego	[illegible]
3.	Prof. dr hab. n. med. Mieczysława Czerwionka-Szaflarska	*pediatra, alergologia i gastroenterologia dziecięca*	[illegible]
4.	Prof. dr hab. n. med. Marek Grabiec	*położnictwo, ginekologia onkologiczna*	[illegible]
5.	Prof. dr hab. n. med. Maria Kłopocka	*choroby wewnętrzne, gastroenterologia*	[illegible]
6.	Prof. dr hab. n. med. Zbigniew Włodarczyk	*chirurgia ogólna, transplantologia kliniczna*	
7.	Dr hab. n. med. Maciej Słupski, prof. UMK	*chirurgia ogólna, transplantologia kliniczna*	
8.	Dr hab. n. med. Katarzyna Sierakowska, prof. UMK	*anestezjologia i intensywna terapia*	
9.	Ks. dr hab. Wojciech Szukalski, prof. UAM	*duchowny*	[illegible]
10.	Dr n. med. Radosława Staszak-Kowalska	*pediatria, choroby płuc*	[illegible]
11.	Mgr prawa Patrycja Brzezicka	*prawniczka*	
12.	Mgr farm. Aleksandra Adamczyk	*farmaceutka*	Adamczyk
13.	Mgr Lidia Iwińska-Tarczykowska	*pielęgniarka*	[illegible]

Streszczenie

Pandemia COVID-19 uwypukliła potrzebę dokładnej i skutecznej diagnostyki, skłaniając do badań nad wykorzystaniem modeli sztucznej inteligencji (AI) do klasyfikacji diagnostycznej pacjentów COVID-19 na podstawie obrazów RTG klatki piersiowej.

Rozprawa zawiera trzy oryginalne artykuły badawcze dotyczące zastosowania metod opartych na uczeniu głębokim (ang. deep learning) do klasyfikacji osób zdrowych i pacjentów z COVID-19 na podstawie obrazów RTG klatki piersiowej. Opracowane modele precyzyjnie rozróżniają pozytywne i negatywne przypadki pacjentów z COVID-19 z wysokim poziomem dokładności, czułości i specyficzności. W pracy zbadano również wpływ technik rozszerzania danych (ang. data augmentation) oraz metod wstępnego przetwarzania danych (ang. data pre-processing) na zdolności klasyfikacyjne, a także zastosowanie modeli głębokiego uczenia do ekstrakcji cech i porównanie z modelami opartymi na drzewach.

Omówiono rozważania etyczne, w tym potencjalne korzyści i wady polegania na modelach uczenia maszynowego (ML) w procesie podejmowania decyzji medycznych oraz implikacje rutynowego wdrażania AI/ML w praktyce klinicznej. Jednym z kluczowych rozważań etycznych jest potencjalny wpływ tych technologii na jakość podejmowania decyzji medycznych. Chociaż modele AI/ML wykazały obiecujące wyniki w kilku zastosowaniach medycznych, ich przejrzystość, niezawodność i dokładność muszą być starannie ocenione, aby zapewnić, że nie zagrażają one jakości opieki nad pacjentem. Aby rozwiązać te kwestie etyczne, ważne jest ustanowienie wytycznych i przepisów dotyczących rozwoju, wdrażania i stosowania technologii AI/ML w opiece zdrowotnej.

Pomimo wyzwań i rozważań etycznych, modele AI mają ogromny potencjał w zakresie poprawy obrazowania medycznego i wyników pacjentów. Potencjalne korzyści z zastosowania modeli AI w obrazowaniu medycznym są liczne i zostały dobrze udokumentowane w różnych badaniach.

Podsumowując, chociaż wdrożenie modeli AI w obrazowaniu medycznym wiąże się z wyzwaniami i rozważaniami etycznymi, potencjalne korzyści są znaczące i nie można ich zignorować. Ponieważ technologia nadal się rozwija i ulepsza, istotne jest, aby zająć się tymi obawami i zapewnić, że wykorzystanie AI w obrazowaniu medycznym odbywa się w sposób odpowiedzialny i etyczny. Dzięki współpracy specjalistów AI i lekarzy, pełny potencjał AI może zostać uwolniony, poprawić wyniki pacjentów i zrewolucjonizować obrazowanie medyczne.

Summary

The COVID-19 pandemic has highlighted the need for accurate and efficient diagnosis, prompting research into using artificial intelligence (AI) models for diagnostic classification of COVID-19 patients based on their chest X-ray images.

The dissertation includes three original research articles on the application of deep learning-based methods to classify healthy and COVID-19 patients based on chest X-ray images. The successfully developed deep learning models accurately distinguish between COVID-19 positive and negative cases with high levels of accuracy, sensitivity, specificity. The dissertation also explored the impact of data augmentation techniques and pre-processing methods on classification abilities, and the use of deep learning models for feature extraction and comparison with tree-based models.

Ethical considerations were discussed, including the potential benefits and drawbacks of relying on machine learning (ML) models for medical decision-making and the implications of routine AI/ML implementation in clinical practice. One of the key ethical considerations is the potential impact of these technologies on the quality of medical decision-making. Although AI/ML models have shown promising results in several medical applications, their transparency, reliability, and accuracy need to be carefully evaluated to ensure that they do not compromise the quality of patient care. To address these ethical considerations, it is important to establish guidelines and regulations for the development, deployment, and use of AI/ML technologies in healthcare.

Despite challenges and ethical considerations, AI models have great potential to improve medical imaging and patient outcomes. The potential benefits of using AI models in medical imaging are numerous and have been well-documented in various studies.

In conclusion, while the implementation of AI models in medical imaging poses challenges and ethical considerations, the potential benefits are significant and cannot be ignored. As the technology continues to evolve and improve, it is essential that these concerns are addressed and ensured that the use of AI in medical imaging is done in a responsible and ethical manner. With careful consideration and collaboration between researchers, practitioners, and stakeholders, the full potential of AI can improve patient outcomes and revolutionize medical imaging.