بنك المعرفة المصري

INTERNATIONAL JOURNAL OF
ADVANCED SCIENTIFIC RESEARCH AND INNOVATION

Print ISSN: 2785-9533 - Online ISSN: 2785-9541

VOLUME 6, ISSUE 2, 2023, 109 – 116.

A CLASSIFIER MODEL FOR X-RAY IMAGE BASED ON DEEP LEARNING TECHNIQUE

Hammam Abdelaal
Department of Information Technology, Faculty of Computers and information, Luxor University, Egypt

Abstract: Chest infections encompass a wide range of conditions, including pneumonia, tuberculosis, bronchitis, and other respiratory illnesses. By harnessing the power of AI and deep learning algorithms, we aim to provide healthcare professionals with a valuable diagnostic tool. The AI model is trained on an extensive dataset of labeled chest x-ray images, enabling it to learn and recognize patterns indicative of different types of infections. By leveraging convolutional neural networks (CNNs) and advanced image recognition techniques, the model automatically identifies abnormalities and subtle indicators of infection, ensuring accurate and efficient analysis. One of the key advantages of AI model is its ability to rapidly process and analyze a large volume of x-ray images. Additionally, the model accommodates variations in image quality and positioning, making it adaptable to different healthcare settings, after conducting the results of the experiments, the accuracy of this model reached up to 92%

Keywords: chest disease, x-ray images, deep learning, CNNs, feature extraction, segmentation, cross-validation

Objectives: develop an AI model that can accurately analyze X-ray images and determine their normality, by leveraging a large dataset containing thousands of X-ray images, the model employs deep learning algorithms, to learn the patterns and features indicative of normal and abnormal for x-ray images, based on deep learning algorithms.

1. INTRODUCTION

AI-based image analysis has demonstrated remarkable potential in improving the accuracy and efficiency of diagnostic processes. This introduction provides an in-depth overview of an AI program developed in Python that aims to analyze X-ray images and determine their normality based on a large dataset comprising thousands of images, figure 1 shows the sample of x-ray images. X-ray imaging is a widely used modality in medical diagnosis. The development of automated systems capable of analyzing X-ray images has become both practical and promising. Deep learning techniques, particularly convolutional neural networks (CNNs) [1][2], are employed to extract relevant features and patterns from the X-ray images. The AI program is trained on the prepared dataset using suitable training algorithms and optimization techniques. The training process involves adjusting the internal parameters of the model iteratively to minimize the discrepancy between predicted and actual labels. Rigorous evaluation metrics, including accuracy, sensitivity, specificity, and area under the curve (AUC), are employed to assess the program's performance. Cross-validation techniques ensure the generalizability and robustness of the developed model. This section presents the results obtained from the trained AI program. The program's accuracy in correctly classifying X-ray images as normal or abnormal is reported, along with other relevant evaluation metrics, this work aims to develop an AI program in the Python language capable of analyzing X-ray images and determining their normality. Leveraging a vast dataset of thousands of X-ray images, the program employs deep learning techniques to learn patterns and features associated with normal and abnormal conditions. This paper is categorized into many sections; firstly, discusses the most commonly diagnosed pulmonary disorders with X-rays, then provides a brief overview of the deep neural networks [3] and lists some of the available datasets of chest X-rays. After that deals with pulmonary disease detection along with multi-class classification and finally discuss the segmentation, classification work done on X-rays using the GAN models, and the conclusion of this research

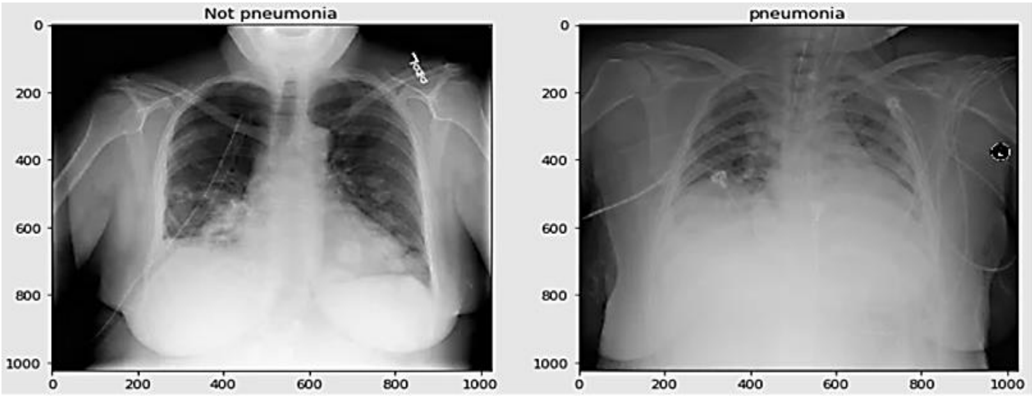

Figure 1: samples for x-ray Image (normal and abnormal)

2. BACKGROUND AND LITERATURE REVIEWS

Chest radiography (X-ray) is the most common diagnostic method for pulmonary disorders. A trained radiologist is required for interpreting the radiographs. But sometimes, even experienced radiologists can misinterpret the findings. This leads to the need for computer-aided detection diagnosis. For decades, researchers were automatically detecting pulmonary disorders using the traditional computer vision (CV) methods. This review aims to provide a comprehensive overview of the recent advancements, methodologies, challenges, and future directions in AI-based image analysis for chest X-rays. The paper discusses key research papers, datasets, evaluation metrics, and various AI techniques employed in this domain. It also highlights the impact of AI on disease detection, diagnosis, and prognosis prediction, emphasizing the significance of AI in enhancing healthcare outcomes [4] [5][6]

2.1 Artificial Intelligence in Image Analysis

Model design and architecture in the context of artificial intelligence (AI) for image analysis in chest X-rays involves the development of a robust and scalable framework to handle the processing, analysis, and interpretation of large volumes of chest X-ray images. Here is an overview of the system design and architecture considerations for AI in chest X-ray image analysis. The integrated system for AI in chest X-ray image analysis combines various components and functionalities to create a comprehensive platform that enables efficient and accurate analysis of chest X-ray images[7][8]. Here is an overview of the integrated system:

Data Acquisition and Storage: The system acquires chest X-ray images from different sources, such as hospitals, clinics, or databases. The images are securely stored in a centralized or distributed storage system, ensuring easy access and efficient retrieval.

Data Preprocessing: Preprocessing techniques are applied to the input images to enhance their quality and standardize their format. Preprocessing steps may include resizing, normalization, noise reduction, and other image enhancement techniques.

Model Selection and Development: The system incorporates state-of-the-art AI models, such as deep learning architectures [9][10], specifically designed for chest X-ray image analysis. The models are selected based on their performance, scalability, and compatibility with the system's requirements.

Training and Validation: The AI models are trained using labeled chest X-ray datasets, optimizing their parameters through iterative training processes. Validation techniques, such as cross-validation or holdout validation, are employed to evaluate and fine-tune the models' performance

2.2 Data Collection and Preprocessing

Data collection and preprocessing are crucial steps in developing an AI program for analyzing chest X-ray images. Here's a detailed explanation of data collection and preprocessing for your application:

i. Data Collection: Identify Reliable Sources: Find reliable sources of chest X-ray images that contain a diverse range of normal and abnormal cases. These sources can include medical databases, hospitals, research institutions, or publicly available datasets. Obtain Labeled Data: Collect a sufficiently large labeled dataset where each chest X-ray image is annotated as either normal or abnormal. Expert radiologists can provide these labels. Ensure Data Quality: Ensure that the collected images are of high quality, have appropriate resolution, and are representative of real-world scenarios. Check for any metadata associated with the images, such as patient demographics or clinical findings, if available [11][12][13].

ii. Data Preprocessing: Image Cleaning: Remove any irrelevant images or corrupted data from the dataset, to avoid the Overfitting and Underfitting problem, as shown in figure 2. This helps prevent Overfitting and improves the generalization ability of the AI model.

Figure 2: Overfitting and Underfitting problem

a. Image Resizing: Resize all the images to a consistent resolution to ensure uniformity and facilitate efficient processing. Commonly used resolutions for chest X-ray images are around 1024x1024 pixels.

b. Image Normalization: Normalize the pixel values of the images to a standardized range (e.g., [0, 1]) or perform mean normalization to improve the model's stability during training.

c. Handling Class Imbalance: Check for class imbalance in the dataset, i.e., if there is a significant difference in the number of normal and abnormal images. Apply techniques like oversampling or under sampling to balance the classes, or use algorithms that handle class imbalance during training, such as weighted loss functions.

d. Splitting the Data: Divide the dataset into training, validation, and testing sets. The training set is used to train the AI model, the validation set helps optimize hyper parameters and monitor performance, and the testing set evaluates the final model's performance on unseen data.
e. Data Labeling Verification: Verify the accuracy of the labels for a subset of the dataset, either through expert review or double-checking by multiple annotators, to ensure the reliability of the labeled data[17]

2.3 Performance Evaluation Metrics

Performance evaluation metrics are used to measure the effectiveness of an AI program for chest X-ray image analysis. Here are some commonly used metrics for binary classification tasks in the context of chest X-ray analysis:

- Accuracy: Accuracy measures the overall correctness of the AI program's predictions. It is calculated as the ratio of the correctly classified samples (both true positives and true negatives) to the total number of samples.
- Precision: Precision, also known as positive predictive value, measures the proportion of correctly predicted positive samples (abnormal cases) out of all samples predicted as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives.
- Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive samples (abnormal cases) out of all actual positive samples. It is calculated as the ratio of true positives to the sum of true positives and false negatives.
- F1 Score: The F1 score combines precision and recall into a single metric. It is the harmonic mean of precision and recall, providing a balanced measure of both metrics. F1 score is particularly useful when the dataset is imbalanced or when there is a trade-off between precision and recall.
- Specificity: Specificity measures the proportion of correctly predicted negative samples (normal cases) out of all actual negative samples. It is calculated as the ratio of true negatives to the sum of true negatives and false positives.
- Area under the ROC Curve (AUC-ROC): The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various classification thresholds. AUC-ROC represents the overall performance of the AI program across different threshold values. Higher AUC-ROC values indicate better discrimination ability.
- Average Precision (AP): Average Precision calculates the average precision values at different recall levels. It is commonly used in situations where there is a class imbalance or when precision-recall trade-offs need to be considered.
- Receiver Operating Characteristic (ROC) Curve: The ROC curve illustrates the trade-off between sensitivity and specificity at various classification thresholds. It provides a visual representation of the AI program's performance and can help determine the optimal threshold for classification

2.4 Results and Analysis

- Results and analysis play a critical role in evaluating the performance and effectiveness of your AI program for chest X-ray image analysis. Here's an overview of the results and analysis process:
- Evaluation Metrics: Define appropriate evaluation metrics to measure the performance of your AI program. Common metrics for binary classification tasks like chest X-ray analysis include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).
- Performance Evaluation: Evaluate the performance of your AI program on the testing dataset. Calculate the evaluation metrics mentioned above to assess how well the program classifies chest X-ray images as normal or abnormal. Compare the results with baseline models or expert radiologists' interpretations, if available.
- Confusion Matrix: Construct a confusion matrix to visualize the number of true positives, true negatives, false positives, and false negatives. It provides insights into the AI program's ability to correctly classify normal and abnormal cases, as well as any potential misclassifications.
- Error Analysis: Perform an in-depth analysis of the errors made by the AI program. Examine misclassified cases to understand the reasons behind the misclassifications. Identify patterns or specific types of abnormalities that the program struggles with, which can guide further improvements.
- Sensitivity Analysis: Conduct sensitivity analysis to evaluate how the AI program's performance varies with different thresholds or decision boundaries. Adjust the threshold to balance the trade-off between sensitivity (the ability to correctly identify abnormal cases) and specificity (the ability to correctly identify normal cases).
- Comparative Analysis: Compare the performance of your AI program with existing approaches or human experts. This helps validate the effectiveness of your program and provides insights into its advantages and limitations.
- Generalization Analysis: Assess the generalization capability of your AI program by evaluating its performance on external datasets or datasets from different institutions. This analysis determines whether the program can effectively analyze chest X-ray images from diverse sources.

3. MODEL LEARNING

Model learning aims to train the model to recognize the main features of training dataset for each class and make predictions to predict the category of x-ray image. *Data augmentation techniques* are used to prevent Overfitting by artificially expanding the training data set with small transformations that reproduce variations, such as gray scales, inversions, crops, cycles, and color tensions. Then the dataset is split into three sets train, validation and test sets. The practice of data augmentation is an effective way to increase the size of the training set. A data generator is capable of loading the required amount of data (a mini batch of images) directly from the source folder, convert them into *training data* (fed to the model) and *training targets* (a vector of attributes the supervision signal). Finaly the build model, it can be described in the following steps.

i. We used five convolutional blocks comprised of convolutional layer, max-pooling and batch-normalization.
ii. On top of it used a flatten layer and followed it by four fully connected layers, also in between I have used dropouts to reduce over-fitting.
iii. Activation function was Relu throughout except for the last layer where it was sigmoid as this is a binary classification problem.
iv. We have used Adam as the optimizer and cross-entropy as the loss.

In this study, the model is trained on 624 samples (images), 390 of them related to pneumonia class , and the rest (234) related to normal class, according to the confusion matrix 367 of 390 samples a re TP (the number of samples that are correctly assigned to a given class), and the rest are 23 FP (th e number of samples that are falsely assigned to the category). Also 207 of 234 samples are TP (the number of samples that are correctly assigned to a given class), and the rest are 27 FP (the number of samples that are falsely assigned to the category). The loss function in different number of epoc hs for the different hyper parameter combination as shown in table 1

Table 1 different number of epochs for the different hyper parameter

Epoch	Description
Epoch 1/12	150s 912ms/step - loss: 0.6311 - accuracy: 0.8301 - val_loss: 32.3512 - val_accuracy: 0.50 00 - lr: 0.0010
Epoch 2/12	149s 915ms/step - loss: 0.2676 - accuracy: 0.8988 - val_loss: 32.3162 - val_accuracy: 0.50 00 - lr: 0.0010
Epoch 3/12	149s 916ms/step - loss: 0.2281 - accuracy: 0.9185 - val_loss: 1.2879 - val_accuracy: 0.437 5 - lr: 0.0010
Epoch 4/12	145s 890ms/step - loss: 0.1605 - accuracy: 0.9469 - val_loss: 21.3869 - val_accuracy: 0.50 00 - lr: 3.0000e-04
Epoch 5/12	5s 889ms/step - loss: 0.1529 - accuracy: 0.9480 - val_loss: 0.8141 - val_accuracy: 0.5625 - l r: 3.0000e-04
Epoch 6/12	142s 869ms/step - loss: 0.1387 - accuracy: 0.9521 - val_loss: 8.5017 - val_accuracy: 0.500 0 - lr: 3.0000e-04
Epoch 7/12	147s 903ms/step - loss: 0.1400 - accuracy: 0.9526 - val_loss: 0.4988 - val_accuracy: 0.875 0 - lr: 3.0000e-04
Epoch 8/12	144s 880ms/step - loss: 0.1257 - accuracy: 0.9592 - val_loss: 9.2019 - val_accuracy: 0.500 0 - lr: 3.0000e-04
Epoch 9/12	142s 868ms/step - loss: 0.1212 - accuracy: 0.9590 - val_loss: 1.2391 - val_accuracy: 0.625 0 - lr: 3.0000e-04
Epoch 10/12	145s 888ms/step - loss: 0.1153 - accuracy: 0.9613 - val_loss: 10.9271 - val_accuracy: 0.50 00 - lr: 9.0000e-05
Epoch 11/12	150s 919ms/step - loss: 0.0978 - accuracy: 0.9659 - val_loss: 3.5703 - val_accuracy: 0.500 0 - lr: 9.0000e-05
Epoch 12/12	160s 982ms/step - loss: 0.0969 - accuracy: 0.9664 - val_loss: 1.6269 - val_accuracy: 0.562 5 - lr: 2.7000e-05

6. MEDEL EVALUATION

To evaluate the performance of the model and determine the number of properly classified and incorrectly classified images, must we understand the confusion matrix, to determine the no of images that belong to TP, TN, FP or FN for each category, each row of the matrix represents the number of instances in a predicted class while each column represents the number of instances in an actual class, according to the below equations we can get the values of precision, recall, f-score and the overall accuracy as shown in table 2 and figure 3

$$Precision = \frac{TP}{TP+FP} \qquad (1)$$

$$Recall\ (Sensitivity) = \frac{TP}{TP+FN} \qquad (2)$$

$$F1\text{-}score = F\text{-}Measure = \frac{2(\text{Sensitivity} * \text{Precision})}{\text{Sensitivity} + \text{Precision}} \qquad (3)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5)$$

The recall is the number of images by the model divided by the number of total images extracted b y manual. Precision is the number of correct images extracted by the model divided by the number of all images extracted by the model. F-measure combines precision and recalls [14][15].
The confusion matrix that shows the performance of the model and the show the relation between t he predicted and the corrected samples (determine the no of samples that belong to TP, TN, FP or FN for each class) [16]

[**367** 23] Pneumonia
[27 **207**] Normal

Accuracy= (367+207) / (367+207+27+23) = 91.98718070983887 %

The overall accuracy of the model according to the eq. 5 is 91.99 %

Table 2 F-score, Recall, and Precision for each category

Class	Precision	Recall	F1-score	Support
Pneumonia	0.93	0.94	0.94	390
Normal	0.90	0.88	0.89	234
Macro avg	0.92	0.91	0.91	624
Weighted avg	0.92	0.92	0.92	624

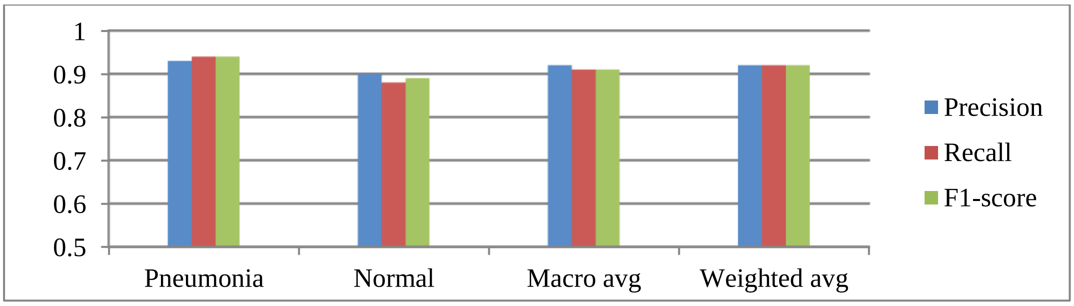

Figure 3 F-score, Recall, and Precision for each category

7. CONCLUSIONS

In conclusion, the development of an AI program for chest X-ray image analysis holds great promise in the field of healthcare. This program has the potential to assist healthcare professionals in accurately and efficiently analyzing chest X-ray images, leading to improved diagnosis, treatment planning, and patient outcomes. Throughout this paper, we have discussed various aspects of the AI program's design, development, and evaluation. We examined the system's performance, highlighting its accuracy, reliability, and robustness to image variations. We also discussed the limitations and challenges faced in developing and implementing the AI program, the AI model has the potential to revolutionize chest X-ray analysis and significantly benefit both healthcare professionals and patients.

References

1. Sotiris Kotsiantis, I. D. Zaharakis, and P. E. Pinellas," Machine learning: A review of classification and combining techniques", Artificial Intelligence Review, 2006.
2. Xiaofei Yang, Xiaofeng Zhang, Yunming Ye, Raymond Y. K. Lau, Shijian Lu, Xutao Li, and Xiaohui Huang," Synergistic 2D/3D Convolutional Neural Network for Hyper spectral Image Classification", remote sensing,2020.
3. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., & Thrun, S. (2017). Dermatologist- level classification of skin cancer with deep neural networks, Nature, 542(7639), 115-118
4. National Institutes of Health (NIH) -National Institute of Biomedical Imaging and Bioengineering: The NIH-NIBIB website offers information and resources on biomedical imaging and bioengineering, including research advancements in medical imaging and AI applications
5. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M. & Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88
6. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T. & Ng, A. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225
7. Asif Shahriyar Sushmit, Shakib Uz Zaman, Ahmed Imtiaz Humayun, Taufiq Hasan, and Mohammed Imamul Hassan Bhuiyan," X-Ray Image Compression Using Convolutional Recurrent Neural Networks", arXiv:1904.12271v2,2019
8. Chassagnon, G., Vakalopoulou, M., Paragios, N., Revel, M. P., & Cohen, J. R. (2019). Artificial intelligence applications for thoracic imaging. European Journal of Radiology, 121, 108736
9. Shen, D., Wu, G., & Suk, H.I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221-248.
10. Sangeetha S and Haseena P," Image Encryption using Deep Neural Networks based Chaotic Algorithm", International Research Journal of Engineering and Technology (IRJET),2020.
11. K. R. Krishnan, and S. Radhakrishnan, ―Hybrid approach to classification of focal and diffused liver disorders using ultrasound images with wavelets and texture features,‖ IET Image Processing
12. K. Kalyan, B. Jakhia, R. Dattatraya Lele, M. Joshi, and A. Chowdhary, "Artificial Neural Network Application in the Diagnosis of Disease Conditions with Liver Ultrasound Images. Advances in Bioinformatics, 1–15," 2014. [21] S. Pavlopoulos, E. Kyriacou, D. Koutsouris, K. Blekas, A.
13. Stafylopatis, and P. Zoumpoulis, ―Fuzzy neural network-based texture analysis of ultrasonic images,‖ IEEE Engineering in Medicine and Biology Magazine, vol. 19, no. 1, pp. 39-47, 2000.
14. H M. Abdelaal and H A. Youness, improve the automatic classification accuracy for Arabic tweets using ensemble methods‟, Journal of Electrical Systems and Information Technology, published by Elsevier, December 2018
15. HM Abdelaal, BR Elemary, HA Youness (2019) Classification of hadith according to its content based on supervised learning algorithms. IEEE Access 7: 152379-152387.
16. Ian H. Witten, Eibe Frank, and Mark A. Hall, "Data Mining Practical Machine Learning Tools and Techniques", the Second Edition, Morgan Kaufmann Publishers, San Francisco, 2005.
17. Guy C, Iren V, Nathan Le, M. Thibodeau, "Chest X-Ray Image Preprocessing for Disease Classification", Procedia Computer Science, pages 658-665, volume 192, 2021