ITM Web of Conferences **68**, 01009 (2024) https://doi.org/10.1051/itmconf/20246801009
ICAETM 2024

Cutting-Edge AI for Healthcare: Optimized Pneumonia Diagnosis on MAX78000 Microcontroller

Rahul Gowtham Poola [1*], *Tejaswi* Chodavarapu[2], and *Siva Sankar* Yellampalli[3]

[1]Department of Electronics and Communication Engineering, SRM University, Andhra Pradesh, India
[2]Department of Electronics and Communication Engineering, SRM University, Andhra Pradesh, India
[3]Department of Electronics and Communication Engineering, SRM University, Andhra Pradesh, India

Abstract. The methodology discussed in the research herein results in the design of a pneumonia diagnosis system that can be executed in real time on hardware edge board. The methodology includes the design and optimization of a neural network classification model for a given pneumonia classifier within the scope of the study. The specifics of the neural network architecture are based upon the designed architecture, which has been modified to fit the computation requirements of the MAX78000 microcontroller. Once trained, the model is exported to file formats supported by the MAX78000 processor, which allows it to be used in the Eclipse programming environment. The final point is the deployment of the trained neural network model with a parallel camera and display which eliminates the need for external software and enables real-time pneumonia diagnosis. This integration emphasizes the flexibility and robustness of the system, and it demonstrates how a low-power system as the MAX78000 allows rethinking the approaches towards the medical diagnostics. After the system is deployed, it is able to deliver accurate predictions where it achieved 97% normal class prediction and 88% pneumonic class prediction. These results reaffirm the system's efficacy in real-time classification tasks, highlighting the diagnostic potential.

1 Introduction

Pneumonia is one of the most commonly reported respiratory ailments across the world and is a causative agent of immense concern as it can be fatal along with severe complications. The word pneumonia is derived from the Greek word pneumonitis, and refers to the infection of lung tissue. Pneumonia can be triggered by a variety of microorganisms including bacteria, viruses, and fungi. It is also a common disease for certain predisposed populations; for example, children, the elderly, and patients with immune deficiency. Internal investigations have shown that pneumonia still poses challenges within the healthcare system. The standard of care focuses on multi-step diagnosis which requires patients to have imaging mainly x-rays of the chest and have them evaluated by a physician. In many countries this procedure is highly complex and may cause important delays in treatment that are often necessary in urgent situations or when help is not readily available. Clinists use non-invasive methods with chest x-ray imaging techniques and radiological changes in these patients to assist in diagnosing pneumonia before it reaches advanced stages. X-ray images show lung pneumonia: infiltrates, abnormal shadows or patches, and increased lung volume are changes that many doctors have long relied upon. The acknowledgement and understanding of these radiographic patterns is of crucial importance in ensuring that the diagnosis of pneumonia is achieved, that treatment options are selected, and that the prognosis is well monitored. The ability to precisely and quickly interpret these imaging findings plays a great role in timely interventions and enhanced patient outcomes.

In this regard, the use of advanced technology like the MAX78000 Microcontroller makes it possible to detect and diagnose pneumonia quickly and accurately, even in areas with limited resources. The technology attached to this microcontroller allows HCPs to analyze pneumonia from X-ray films, in real time and without the need for guesswork. Using X-ray images, pneumonia is usually diagnosed only with the help of healthcare experts. The MAX78000 microcontroller does, however, change the situation because it is designed to carry out intelligent analysis of X-ray images with the help of a pre-trained model without human intervention. The consumption of power is lower and the inference has high-performance because of the particular hardware accelerators for neural networks which the MAX78000 has. This is appropriate for Ai in healthcare because it is used widely in such applications. With this technology, the MAX78000 microcontroller can quickly take and analyze X-ray pictures, search for signs of pneumonia, and arrive at a diagnosis in a matter of seconds. This integration has some possible benefits such as diagnosis in real time, lesser reliance on expertise, and availability in areas far from centers, timely intervention and efficiency of resources being used. Successful implementation is dependent on the development of a model, optimization, and model

* Corresponding author: rahulgowtham_poola@srmap.edu.in

embedding into the MAX78000's structure in adherence to medical standards and regulation criteria. Development of a reliable and useful diagnostic instrument to support pneumonia diagnosis by X-ray images using the MAX78000 microcontroller requires synergy between medical professionals, AI specialists, and hardware engineers.

2 Related Work

Interventional radiology employs Biomedical imaging modalities, X-ray and CT scans, to diagnose lung complications of pneumonia [1-4]. Recent studies [5-8] have employed deep learning algorithms such as VGG19, U-Net and convolutional neural networks (CNNs) to study X-Ray images to detect whether a COVID-19 positive or negative patient is in the picture. These studies adopted techniques such like segmentation preprocessing, transfer learning, data augmentation and achieved very high detection accuracies from about 97% to more than 98%. In one noteworthy study [9], researchers designed a new CNN architecture that differentiates between COVID-19 patients and healthy individuals with perfect accuracy. While differentiating healthy volunteers with COVID-19 patients and pneumonia patients, the model performed classification with an accuracy of 93.75%. In another research [10], hand-crafted cues, radiological cues with deep cues computed from pre-trained networks were differentiated for improving COVID-19 diagnoses. It was found that combining several different types of features produced higher classification accuracy than using any single feature set alone. Moreover, in [11], Inception-V3, MobileNet, Xception, and DenseNet based models were implemented for the detection of COVID-19 from X-ray images. From these deep learning algorithms, MobileNet emerged the best in performance based on accuracy, recall, F1-score metrics.

Studies [12–14] have involved CT and radiograph imaging patterns and deep learning algorithms based on VGG19, ResNet-50 and MobileNet-V2 for the detection of COVID-19. It was noted during these studies that deep learning techniques were successful in determining the Covid-19 presence from imaging datasets. Similar research [15] combined the use of deep learning and machine learning algorithms for COVID-19 detection using chest x-ray images which provided high sensitivity, specificity, accuracy, and AUC values. The research utilizing a novel variation of multi-objective adaptive differential evolution (MADE) within a CNN framework was also very accurate and outperformed other machine learning techniques.

At the same time, more recent works try to implement machine learning models into edge devices such as microcontroller units (MCUs) or edge processors with low power [16]. This type of trend allows the feature detectors to be embedded directly onto these devices without the available features which are resource-heavy. For adjusting deep learning methods, model quantization has been proven to be successful in lowering resources for deep learning models on low-power platforms while enhancing the efficiency of computations [17]. Another major and revolutionary system is the one developed by Analog Devices which is the MAX78000 [18]. It provides high computational performance with very low power consumption. The main features of MAX78000 include sufficient RAM for model storage and a dedicated CNN accelerator. Thus, it is regarded as a promising solution for learning models based on deep learning for applications in edge computing.

3 Proposed Methodology

The methodology developed for a successful pneumonia detection and diagnosis system begins by obtaining a set of X-Ray images of patients with normal conditions as well as in cases where the pneumonia afflicts the patient. After the dataset is obtained, the images are then preprocessed to enhance the image quality as well as grab relevant information required for the shape analysis. A target pneumonia classification neural network model is then developed and optimized for the memory constrained MAX78000 CUT microcontroller. The parameters of the model are trained, then saved in a model compatible with the MAX78000 MC and the model is set up on the epos software. The system is also developed such that it is efficient in the usage of memory and processing time while ensuring that the required performance of the system is met, with the relevant optimization employed where necessary. The integration of parallel camera interface with a touch display increases the functionality of the system in real time diagnosis of pneumonia, showing the effectiveness and flexibility of the system as a solution to diagnostic problems. It illustrates the need for computational efficiency as well as accuracy in classification when designing an image classification system suitable for use in a hardware edge board. Fig. 1 shows the procedure used to identify pneumonia.

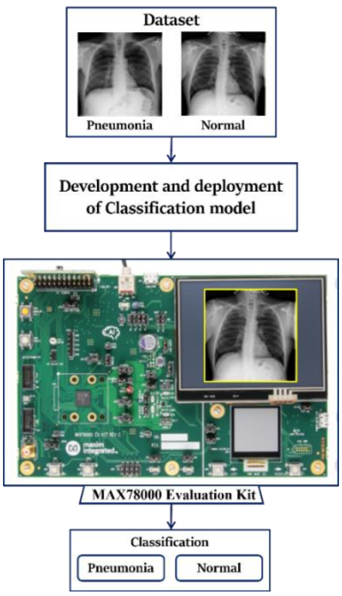

Fig. 1. Proposed Methodology

3.1 Aggregating Radiographic Data

The dataset contains two groups of subjects, patients suffering from pneumonia and completely healthy individuals composed of 3,616 pneumonia patients and 10,192 able-bodied people's chest radiographs. This data was obtained through Kaggle [19]. Furthermore, Figshare [20] managed to add 900 radiographs of healthy individuals and 900 radiographs of pneumonia cases. All images are PNG in a width, height, and depth ranging from 229 x 229 x 3 up to 1024 x 1024 x 3 pixels. As a result of the data augmentation process, all images were resized to 128 x 128 x 3 pixels for the purpose of having standard dimensions and to help in eliciting pseudo-X-ray pictures. Furthermore, for model performance and resource consumption efficiency the X-ray images were also compressed.

3.2 Pre-processing

For the purposes of pneumonia diagnosis, the term pre-processing refers to the set of elementary actions that are performed to improve the quality and integrity of the medical images. This would include the elimination of some irrelevant structures, the normalization of the volume of images by reducing their size, and satellite imaging which provides details of the features of the interest area. Chest x-rays are very important in the diagnosis of pneumonia as they allow the chest structure to be seen including the lungs. Nevertheless, these images also have distortions, interferences and defects that may compromise the effectiveness of image diagnostics.

3.3 MAX78000 microcontroller Implementation

Pneumonia diagnosis system deployment on the MAX78000 microcontroller requires a number of steps which may involve the identification of a suitable neural network, saving the model, exporting it to Tensorflow, making files for the MAX78000 microcontroller, and implementing the workflow in Eclipse software. The procedure unfolds as follows:

3.3.1 Model Development:

- Create and train a neural network model specialized in pneumonia classification.
- The model should accept X-ray images as input and produce binary classification outcomes (Normal or Pneumonia).

3.3.2 Saving Trained Model as mat File:

Following training, save the neural network model as a .mat file through MATLAB.

3.3.3 Exporting Model to TensorFlow File:

Convert the .mat file into a TensorFlow-compatible format using the command "exportNetworkToTensorFlow(net, 'folder name')".

3.3.4 Conversion to MAX78000 Compatible Files:

Transforming a TensorFlow model into MAX78000 embedded C code entails multiple phases: Training and Synthesis.

- Training:

- Optimize and train the TensorFlow model.
- Fine-tune the model for desired accuracy and performance.
- Convert the trained TensorFlow model into frozen weights and biases.
- Extract and format the model parameter values suitable for the MAX78000 microcontroller.
- Utilize sparse embedding layers in the TensorFlow graph for efficient resource utilization.

- Synthesis:

- Extract and store the model parameter values in YAML format, outlining model specifications.
- Translate the TensorFlow model into equivalent C code.
- Employ the MAX78000 TensorFlow Lite Converter to produce MAX78000-compatible files (e.g., `.c`, `.h`).
- Load the converted C code onto the MAX78000 microcontroller for deployment.

3.3.5 Eclipse IDE Workflow Implementation:

- Configure the Eclipse IDE Workspace for MAX78000 development.
- Establish a new project and incorporate MAX78000 libraries.
- Import the converted model files into the project.
- Load and execute the model, handling image classification.
- Develop code to interface with the parallel camera interface for X-ray image capture.
- Showcase classification outcomes and corresponding input X-ray images on the 3.5-inch TFT touch-enabled display.

3.3.6 Performance Assessment:

- Evaluate the performance of the pneumonia diagnosis algorithm on the MAX78000 microcontroller using the X-ray images.

Pneumonia diagnosis based systems particular one based on neural networks is dependent on its performance metrics such the architecture design and models. The structure of the classification model is composed of seven layers, each having a different role to play. The network architecture consists of several types of layers such as the convolutional layers which are used for feature extraction, the Rectified Linear Units (ReLU) as an activation function to introduce non – linearlity, pooling layers functional for the highest values, a softmax layer for the distribution of probability, an input layer with the corresponding image data and the output layer for the classification. The designs are expected as the results given in the third objective in which the model is described to an extent to know how the design of each model layer contributes to the overall diagnostic accuracy of the system being targeted in this case pneumonia diagnosis. The workflow stress on the key strategies for construction, importing and applying the model for pneumonia diagnosis in consideration of the MAX78000 microcontroller. This model is implemented with the parallel camera interface and touch display to provide an all-in-one solution for diagnosis. Fig. 2 shows the overall setup of the hardware implementation target for real-time pneumonia diagnosis using the MAX78000 microcontroller. Through the following Fig. 3, the classification model and the various layer aspects used in the diagnosis of pneumonia has been presented.

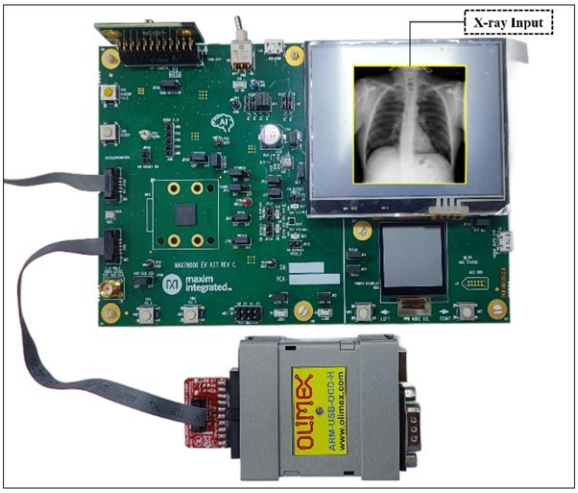

Fig. 2. Hardware implementation setup

	Layer Name	Type	Activations	Learnable Properties
1	input 128x128x3 images with 'zerocenter' normalization	Image Input	128(S) × 128(S) × 3(C) × 1(B)	-
2	conv1 16 3x3 convolutions with stride [11] and padding 'same'	2-D Convolution	128(S) x 128(S) × 16(C) x 1(B)	Weights 3 x 3 x 3 x 16 Bias 1 × 1 × 16
3	relu1 ReLU	ReLU	128(S) × 128(S) × 16(C) x 1(B)	-
4	maxpool1 2x2 max pooling with stride [22] and padding [0 0 0 0]	2-D Max Pooling	64(S) x 64(S) × 16(C) x 1(B)	-
5	conv2 32 3x3 convolutions with stride [11] and padding 'same'	2-D Convolution	64(S) × 64(S) × 32(C) × 1(B)	Weights 3 × 3 × 16 × 32 Bias 1 x 1 x 32
6	relu2 ReLU	ReLU	64(S) × 64(S) × 32(C) × 1(B)	-
7	maxpool2 2x2 max pooling with stride [22] and padding [0 0 0 0]	2-D Max Pooling	32(S) x 32(S) x 32(C) x 1(B)	-
8	conv3 64 3x3 convolutions with stride [11] and padding 'same'	2-D Convolution	32(S) × 32(S) × 64(C) × 1(B)	Weights 3 × 3 × 32 × 64 Bias 1 x 1 x 64
9	relu3 ReLU	ReLU	32(S) x 32(S) x 64(C) x 1(B)	-
10	maxpool3 2x2 max pooling with stride [22] and padding [0 0 0 0]	2-D Max Pooling	16(S) × 16(S) × 64(C) × 1(B)	-
11	conv4 32 3x3 convolutions with stride [11] and padding 'same'	2-D Convolution	16(S) x 16(S) x 32(C) × 1(B)	Weights 3 × 3 × 64 × 32 Bias 1 x 1 x 32
12	relu4 ReLU	ReLU	16(S) x 16(S) x 32(C) × 1(B)	-
13	maxpool4 2x2 max pooling with stride [22] and padding [0 0 0 0]	2-D Max Pooling	8(S) × 8(S) × 32(C) × 1(B)	-
14	conv5 32 3x3 convolutions with stride [1 1] and padding 'same'	2-D Convolution	8(S) × 8(S) × 32(C) × 1(B)	Weights 3 × 3 × 32 × 32 Bias 1 x 1 x 32
15	relu5 ReLU	ReLU	8(S) × 8(S) × 32(C) × 1(B)	-
16	conv6 16 3x3 convolutions with stride [11] and padding 'same'	2-D Convolution	8(S) × 8(S) × 16(C) × 1(B)	Weights 3 x 3 x 32 x 16 Bias 1 × 1 × 16
17	relu6 ReLU	ReLU	8(S) × 8(S) × 16(C) × 1(B)	-
18	fc 2 fully connected layer	Fully Connected	1(S) x 1(S) × 2(C) x 1(B)	Weights 2 x 1024 Bias 2 x 1
19	softmax softmax	Softmax	1(S) × 1(S) × 2(C) x 1(B)	-
20	output	Classification Output	1(S) × 1(S) × 2(C) × 1(B)	-

Fig. 3. Model Description

4 Results and Discussion

Completing the MAX78000 microcontroller image classification algorithm is an important milestone on the way of combining the theoretical progress with its practical deployment. This initiative illustrates the multifaceted problems and important issues related to the use of deep neural networks for embedded systems with limited resources. Operating under such conditions that include memory, computational capabilities, and energy consumption - poses a challenge that must be met if real-time efficiency is to be achieved. Theoretical considerations include understanding these constraints and the need to devise specific approaches for the algorithm in order for it to execute seamlessly on the microcontroller. Performance evaluation presented in Table 1 demonstrates the advantages of real-time implementation of the MAX78000 microcontroller as it was designed and targeted to operate within certain constraints. Furthermore, Fig. 4 present results of the prediction accuracy of the proposed device system based on the MAX78000 microcontroller and pneumonia diagnosis system which measures the efficiency of the developed system relative to the performance of existing ones. This type of analysis reinforces the claims and claims useful for the application of the algorithm in practical medical situations. Taken altogether, the precise adaptation and implementation of the image classification algorithm on the MAX78000 microcontroller is a novel way of addressing the issues related to the implementation of deep learning in low-powered devices, which has a potential for improving diagnostic capabilities in the healthcare industry. In addition, Fig. 5 illustrates a Diagnostic prediction for the test X-ray images through the graphical user interface, demonstrating its clinical responsiveness and accuracy on actual cases.

Table 1. Performance Evaluation of real time MAX 78000 Microcontroller implementation

Test image	Algorithm	Classification Result	Prediction Accuracy
	Neural Network Classification model	Normal	97%
	Neural Network Classification model	Normal	94%
	Neural Network Classification model	Pneumonia	88%
	Neural Network Classification model	Pneumonia	80%

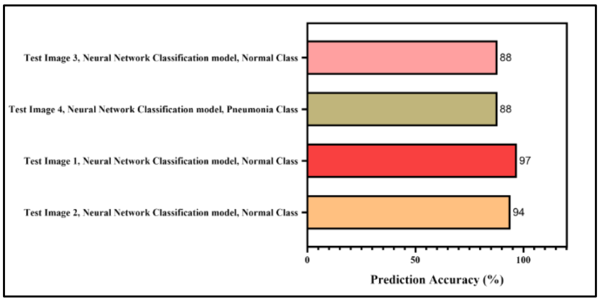

Fig. 4. MAX78000 microcontroller accuracy results

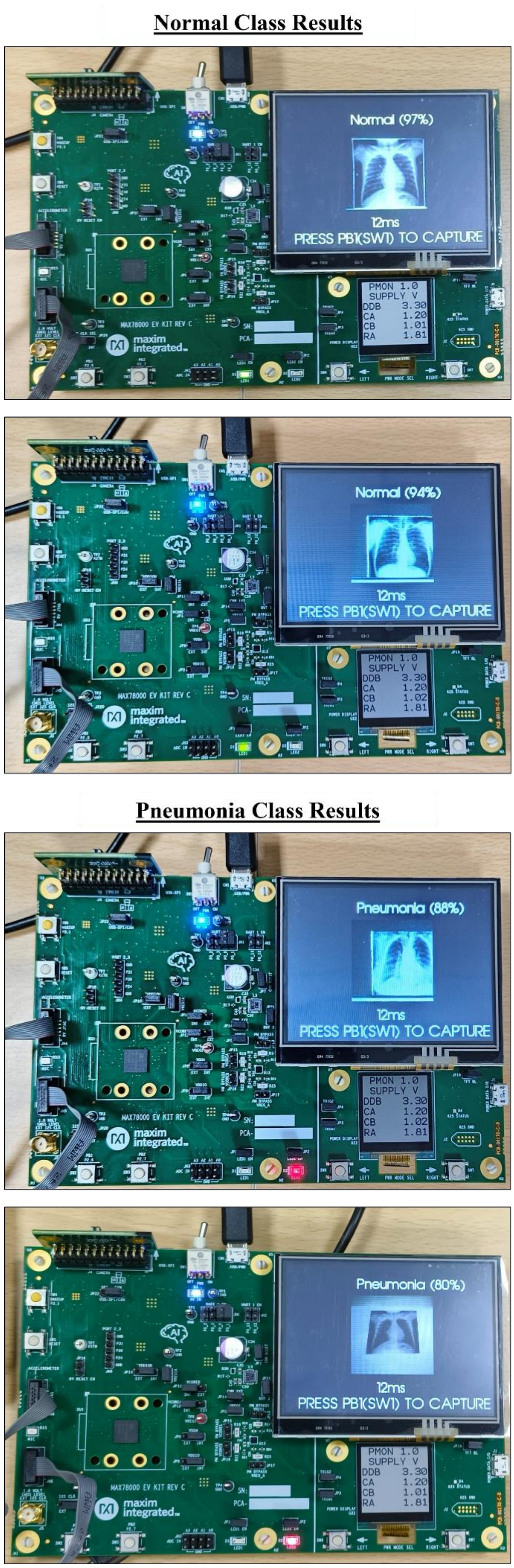

Fig. 5. Real-time Diagnostic predictions for test X-ray

5 Conclusion

Finally, the implementation of the neural network-based pneumonia diagnosis system using the MAX 78000 microcontroller is an important improvement in the area of medical diagnosis performed in real-time. With the deployment of this thoroughly optimized neural network model on this domain specific resource-constrained hardware, the research obtains stunning prediction accuracies, going relatively further in the field. The accuracy levels of 97% and 88% for Normal and Pneumonia classes respectively are in particular quite remarkable. The achieved high accuracies strong show the strength and efficiency of the implemented system in making pneumonia diagnosis cases, even when bound by the limitations of memory and the processing capabilities of the microcontroller. The implications of these results are significant for clinical practice. The capacity to perform the diagnosis process on edge hardware in real time allows doctors to make quick decisions when it is required, which is likely to reduce the time taken to start treatment and thus improve patient results. In addition, the fact that the system can also be used with a parallel camera and a display makes the system more user friendly and practical in the clinical environment so that it meets the needs of the health practitioners. In conclusion, the effective implementation and deployment of the neural network model on the MAX 78000 microcontroller brings to attention the potential that edge computing platforms have in medical diagnostics. This achievement therefore unlocks the development of low cost and effective diagnostic tools which can operate most efficiently in third world countries and enhance the experience of patients as well as healthcare professionals.

References

1. Kundu, R., Das, R., Geem, Z. W., Han, G. T., & Sarkar, R. (2021). Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PloS one*, *16*(9), e0256630. https://doi.org/10.1371/journal.pone.0256630
2. Manickam, A., Jiang, J., Zhou, Y., Sagar, A., Soundrapandiyan, R., & Samuel, R. D. J. (2021). Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures. *Measurement*, *184*, 109953. https://doi.org/10.1016/j.measurement.2021.109953
3. Hashmi, M. F., Katiyar, S., Hashmi, A. W., & Keskar, A. G. (2021). Pneumonia detection in chest X-ray images using compound scaled deep learning model. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *62*(3-4), 397-406. https://doi.org/10.1080/00051144.2021.1973297
4. Yao, S., Chen, Y., Tian, X., & Jiang, R. (2021). Pneumonia detection using an improved algorithm based on faster r-cnn. *Computational and*

Mathematical Methods in Medicine, *2021*, 1-13. https://doi.org/10.1155/2021/8854892

5. Arias-Garzón, D., Alzate-Grisales, J. A., Orozco-Arias, S., Arteaga-Arteaga, H. B., Bravo-Ortiz, M. A., Mora-Rubio, A., ... & Tabares-Soto, R. (2021). COVID-19 detection in X-ray images using convolutional neural networks. *Machine Learning with Applications*, *6*, 100138. https://doi.org/10.1016/j.mlwa.2021.100138
6. Bushra, K. F., Ahamed, M. A., & Ahmad, M. (2021). Automated detection of COVID-19 from X-ray images using CNN and Android mobile. *Research on Biomedical Engineering*, *37*(3), 545-552. https://doi.org/10.1007/s42600-021-00163-2
7. VJ, S. (2021). Deep Learning Algorithm for COVID-19 Classification Using Chest X-Ray Images. *Computational and Mathematical Methods in Medicine*, *2021*(1), 9269173. https://doi.org/10.1155/2021/9269173
8. Brunese, L., Martinelli, F., Mercaldo, F., & Santone, A. (2020). Machine learning for coronavirus covid-19 detection from chest x-rays. *Procedia computer science*, *176*, 2212-2221. https://doi.org/10.1016/j.procs.2020.09.258
9. Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2022). Automated detection of COVID-19 through convolutional neural network using chest x-ray images. *Plos one*, *17*(1), e0262052. https://doi.org/10.1371/journal.pone.0262052
10. Ho, T. K. K., & Gwak, J. (2022). Feature-level ensemble approach for COVID-19 detection using chest X-ray images. *Plos one*, *17*(7), e0268430. https://doi.org/10.1371/journal.pone.0268430
11. Rawat, R. M., Garg, S., Jain, N., & Gupta, G. (2021, May). COVID-19 detection using convolutional neural network architectures based upon chest X-rays images. In *2021 5th international conference on intelligent computing and control systems (ICICCS)* (pp. 1070-1074). IEEE. doi: 10.1109/ICICCS51141.2021.9432134
12. Zouch, W., Sagga, D., Echtioui, A., Khemakhem, R., Ghorbel, M., Mhiri, C., & Hamida, A. B. (2022). Detection of COVID-19 from CT and chest X-ray images using deep learning models. *Annals of Biomedical Engineering*, *50*(7), 825-835. https://doi.org/10.1007/s10439-022-02958-5
13. Aggarwal, S., Gupta, S., Alhudhaif, A., Koundal, D., Gupta, R., & Polat, K. (2022). Automated COVID-19 detection in chest X-ray images using fine-tuned deep learning architectures. *Expert Systems*, *39*(3), e12749. https://doi.org/10.1111/exsy.12749
14. Dileep, P., Jayasri, N., & Raghavender, G. (2023). DETECTION OF COVID-19 FROM CHEST X-RAY IMAGES USING CONVOLUTIONAL NEURAL NETWORKS. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *14*(03), 604-615. https://doi.org/10.17762/turcomat.v14i03.14097
15. Singh, D., Kumar, V., Yadav, V., & Kaur, M. (2021). Deep neural network-based screening model for COVID-19-infected patients using chest X-ray images. *International Journal of Pattern Recognition and Artificial Intelligence*, *35*(03), 2151004. https://doi.org/10.1142/S0218001421510046
16. Giordano, M., Piccinelli, L., & Magno, M. (2022, June). Survey and comparison of milliwatts micro controllers for tiny machine learning at the edge. In *2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)* (pp. 94-97). IEEE. https://doi.org/10.1109/AICAS54282.2022.9870017
17. Wang, X., Magno, M., Cavigelli, L., & Benini, L. (2020). FANN-on-MCU: An open-source toolkit for energy-efficient neural network inference at the edge of the Internet of Things. *IEEE Internet of Things Journal*, *7*(5), 4403-4417.doi:10.1109/JIOT.2020.2976702
18. A. Devices, “Max78000,” 2022. [Online]: https://www.analog.com/en/products/MAX78000.html
19. X-ray Dataset-1: https://www.kaggle.com/datasets
20. X-ray Dataset-2: Haghanifar, Arman; Molahasani Majdabadi, Mahdiyar; Ko, Seokbum (2020). Chest X-Ray Image Repository. figshare. https://figshare.com/articles/dataset/COVID-19_Chest_X-Ray_Image_Repository/12580328