COVID-19 Diagnosis: Edge Impulse-Driven CNN Architecture for High-Accuracy Classification of Lung X-ray Images

1st Christopher de Souza Lima Francisco
Inatel - National Institute of Telecommunications
Santa Rita do Sapucaí, Brazil
christopher@inatel.br

2nd Gabriel Pivoto Ambrósio
Inatel - National Institute of Telecommunications
Santa Rita do Sapucaí, Brazil
gabriel.pivoto@inatel.br

***Abstract*—This study explores developing and training an Edge Impulse model to classify COVID-19 in lung X-ray images. The model architecture encompasses two fundamental blocks: an image block configured for grayscale depth to extract intricate features and a classifier block employing a 2D convolutional neural network (CNN). Through meticulous training involving 55[1] cycles with a learning rate of 0.0005, the CNN architecture leveraged 32 and 16 filters across its layers, alongside a flattening layer and a dropout rate of 0.25. The integration of this CNN architecture enabled the discernment of nuanced patterns within the dataset, resulting in an exemplary accuracy of 92.2% in diagnosing COVID-19. This study showcases the efficacy of Edge Impulse in harnessing advanced AI techniques and CNN architectures for efficient and accurate COVID-19 diagnosis, promising transformative implications for medical diagnostics in resource-constrained environments.**

***Index Terms*—Artificial intelligence, Machine Learning, Edge Impulse, Classification**

I. Introduction

Amid the global COVID-19 pandemic, the rapid and accurate diagnosis of infected individuals has become a critical imperative for public health. Medical imaging, particularly chest X-rays, has emerged as a valuable diagnostic tool for detecting patterns indicative of respiratory infections, including COVID-19. With the advent of advanced technologies, artificial intelligence (AI) and machine learning (ML) have demonstrated unparalleled potential in revolutionizing medical diagnostics [1].

Traditional diagnostic methods for respiratory infections involve scrutiny of radiological images by trained professionals. However, the sheer volume of cases and the urgency for timely diagnosis in a pandemic have underscored the need for automated and efficient screening processes. AI and ML algorithms, when trained on diverse datasets encompassing COVID-19-positive and negative cases, can learn intricate patterns and subtle nuances within chest X-ray images that may elude the human eye [2].

This article also delves into a novel approach, employing Edge Impulse, a cutting-edge platform, to simulate and implement AI-powered classification models to rapidly and accurately identify COVID-19 in lung X-ray images. The convergence of AI and Edge Impulse technology represents a pioneering effort to bring computational intelligence to the forefront of medical diagnostics, particularly in resource-constrained environments [3].

Edge Impulse, renowned for its prowess in facilitating the deployment of machine learning models on edge devices, provides a unique framework for real-time, on-device inference. By leveraging this platform, we address the imperative need for swift diagnostics and embrace the practicality of deploying AI solutions directly at the edge, minimizing the reliance on centralized computing resources [4].

Our exploration encompasses the intricacies of model training, validation, and testing, focusing on the unique challenges posed by the edge environment. Through rigorous experimentation, we elucidate the efficiency and reliability of the Edge Impulse platform in enabling AI-based COVID-19 diagnosis, shedding light on its potential to revolutionize medical diagnostics by making intelligent decisions directly at the source of data acquisition.

We endeavor to advance state-of-the-art diagnostic technology and advocate for the practical implementation of these solutions in diverse healthcare settings. This article contributes valuable insights into the promising landscape of AI and Edge Impulse synergy, paving the way for accessible, real-time, and accurate COVID-19 diagnosis by analyzing lung X-ray images. This article explores the innovative application of AI and ML techniques to classify chest X-ray images and diagnose the presence of COVID-19 with a high degree of accuracy. All work related to this article can be accessed from GitHub repository[2].

II. Related Works

This section presents other articles and studies concerning the use of machine learning for COVID-19 detection:

A. Machine Learning for medical imaging-based COVID-19 detection and diagnosis

Rehouma, Buchert, and Chen [5] review recent progress in machine learning (ML) for COVID-19 detection, emphasizing

[1]This is the max number of epochs possible to stay under 20 min training time, which is the max time allowed on Edgle Impulse's free plan.

[2]https://github.com/chrislima/tp557-covid-classifier

the potential for early and accurate diagnosis to control epidemics and reduce mortality. While nose and nasopharyngeal swabs are vital for clinical detection, medical imaging demonstrates a high positive rate of disease detection, particularly CT scans, X-rays, and ultrasound.

The review compiles publicly available datasets for training ML models, focusing on analyzing models for COVID-19 diagnosis using various image modalities. The U-Net convolutional network stands out for efficiently extracting lung regions and lesions in applications.

The discussion includes similar projects, highlighting their reliance on deep learning. The authors explore chest CT, X-rays, and lung ultrasonography. Supervised ML algorithms and deep learning, primarily using Convolutional Neural Networks (CNN), emerge as common methods for COVID-19 diagnosis.

Limitations addressed include a lack of labeled data, imbalanced datasets, data quality issues, and model evaluation and interpretability challenges. Despite reviewing 62 studies introducing ML-based solutions, the authors note a limitation in comparing ML model performance between X-ray and CT scan images due to a scarcity of studies across multiple image modalities.

Despite ML's effectiveness in COVID-19 research, the shortage of large datasets poses a significant challenge. The authors suggest integrating medical images with clinical and laboratory data will enhance ML model quality and performance.

B. *Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: A machine learning based approach*

Kassania et al. [6] highlight the potential of computer-aided diagnosis systems for early COVID-19 detection, aiding in monitoring disease progression and potentially reducing mortality. The study compares various deep learning-based feature extraction frameworks for automatic COVID-19 classification, including MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, and NASNet, chosen from a pool of deep convolutional neural networks.

The primary motivation is to introduce a generic feature extraction method using convolutional neural networks, eliminating the need for handcrafted or complex features and facilitating application to different modalities such as X-ray and CT images. The study aims to minimize generalization errors for a more accurate diagnosis.

In the initial step, public chest X-ray and CT images are collected, followed by dataset preprocessing using standard image normalization techniques to enhance visual information quality. Prepared input images undergo feature extraction with state-of-the-art CNN descriptors. For the training phase, the generated features feed into machine learning classifiers like Decision Tree, Random Forest, XGBoost, AdaBoost, Bagging classifier, and LightGBM. The authors adopt a 10-fold cross-validation technique to assess classifier generalization performance on test images.

In conclusion, the authors discuss the frameworks and present experimental results. They emphasize that DenseNet121 architecture, trained by a Bagging tree classifier, achieves a remarkable 99.00% classification accuracy on available chest X-ray and CT datasets. The authors suggest future investigations into designing models with transferred and fine-tuned weights within similar domains, as opposed to weights from distinct domains like the ImageNet dataset.

C. *A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods*

Saygılı [7] underscores the significance of technological tools that swiftly and accurately detect COVID-19 infections, aiding medical professionals in overcoming challenges posed by similarities between COVID-19 and other lung infections in imaging modalities such as computed tomography (CT) and X-rays. To address this, computer-aided diagnostic solutions are being developed to enhance the identification of positive COVID-19 cases.

This study investigates the role of machine learning and image processing in rapidly and accurately detecting COVID-19 from two prevalent medical imaging modes: chest X-ray and CT images. The proposed approach is applied to three distinct public COVID-19 datasets, involving five key stages: dataset acquisition, preprocessing, feature extraction, dimension reduction, and classification, each comprising specific sub-operations.

Three datasets are utilized: the first, composed of 349 COVID-19-positive CT images; the second, consisting of 125 positive COVID-19 X-ray images; and the third, a CT dataset with a substantial 1252 CT scans.

In conclusion, the study's application to three datasets with different characteristics underscores its generalizability. The findings reveal that COVID-19 can be detected with high success rates in less than a minute using image processing and classical learning methods. Notably, classical learning methods like k-NN and SVM exhibit exceptional success in detecting COVID-19 positivity, achieving accuracies of 89.41%, 99.02%, and 98.11% for datasets 1 (CT), 2 (X-ray), and 3 (CT), respectively.

III. Metodology

The project revolves around a COVID-19 disease classifier using chest X-rays. To accomplish this, two datasets, (i) Covid-19 Image Dataset[3] and (ii) COVID CXR Image Dataset (Research)[4] were collected from the Kaggle website's Computer Vision section, containing images of chest X-rays. These images were then uploaded to the Edge Impulse platform, where the model was trained.

A. *Dataset*

Training the model required the utilization of two datasets containing chest X-rays. Initially, only one dataset was planned to abide by the limitations of the free usage of the Edge

[3] https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset

[4] https://www.kaggle.com/datasets/sid321axn/covid-cxr-image-dataset-research

Impulse platform. Two different datasets were tested. The first one comprises three classes: Viral Pneumonia, Normal, and Covid, containing 90, 90, and 137 images, respectively. The second dataset, featuring three classes, includes Covid, Normal, and Virus, with 536, 668, and 619 images, respectively. Both datasets were used to enhance the model's training, assuming that the Viral Pneumonia class from the first dataset and the Virus class from the second dataset represent the same category, named Virus. Figures 1, 2, and 3 represent an example of Covid, Normal, and Virus, respectively.

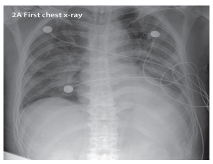

Fig. 1. Covid

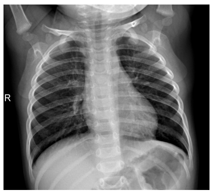

Fig. 2. Normal

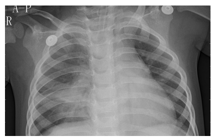

Fig. 3. Virus

B. Edge Impulse

Training the model on the Edge Impulse platform is straightforward. Initially, the data to train the model is uploaded. For the first dataset, images were loaded into separate training and testing sets as it was already provided that way. For the second dataset, a feature of the platform was used to split the images into training and testing sets since the initial division wasn't available. The final data acquisition process resulted in 1919 items divided 80% as train and 20% as test. It's important to note that the total count of items may be lower than anticipated because the data upload process on Edge Impulse verifies whether the images in the dataset are identical. The result is shown in Figure 4.

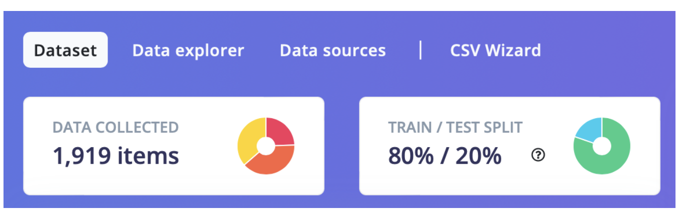

Fig. 4. Model accuracy and Confusion Matrix.

To create the impulse and train the model, two blocks were added: one for the image and another for the classifier. The grayscale color depth was selected in the image block, and the features were subsequently generated. In the classifier block, 55 training cycles and a learning rate of 0.0005 were employed, using a 2D convolution architecture consisting of 32 filters, one with 16, one flattened layer, and a dropout of 0.25.

This study employed a convolutional neural network (CNN) architecture for the task at hand [8]. The architecture is designed to process high-dimensional input data and extract hierarchical features through the utilization of convolutional and pooling layers, followed by flattening and dropout layers for regularization. The detailed configuration of the neural network is outlined below:

C. Input Layer

The input layer comprises 9,216 features representing the raw input data. This layer serves as the neural network's entry point, capturing the dataset's initial information.

D. 2D Convolutional / Pooling Layer (Layer 1)

The first convolutional layer consists of 32 filters with a kernel size of 3x3. This layer is responsible for capturing local patterns and features within the input data. Following the convolution operation, a max-pooling operation with a pool size of 2x2 is applied to down-sample the spatial dimensions, retaining essential information.

E. 2D Convolutional / Pooling Layer (Layer 2)

Subsequently, a second convolutional layer is employed with 16 filters and a 3x3 kernel size. Similar to the first layer, this convolutional layer is followed by a max-pooling operation (2x2) to abstract the learned features further.

F. Flatten Layer

The flattening layer is introduced after the convolutional layers to transform the 2D feature maps into a 1D vector, enabling the transition from spatial hierarchies to a format suitable for fully connected layers.

G. Dropout Layer

A dropout layer with a dropout rate of 0.25 is applied after the flattening layer to mitigate overfitting and enhance the model's generalization capability. This layer randomly deactivates a fraction of the neural units during training, preventing the network from relying too heavily on specific nodes and improving its robustness.

The architecture's configuration was determined through a combination of domain expertise and iterative experimentation to optimize performance on the specific task under investigation. Furthermore, the Python code generated from the model in Edge Impulse was exported, along with supplementary files including the application's Dockerfile. Using these resources, an application container was constructed to facilitate local model training upon initialization. Upon completion of the training process, a compiled model file is produced for utilization. Following this, a dedicated Python script was crafted to evaluate the effectiveness of this model through the classification of lung images.

IV. Results

The model underwent various training methodologies until achieving the optimal outcome. Initially, solely the first dataset was employed. Utilizing this dataset, 20 epochs were executed with a learning rate of 0.0005. None of the convolutional layers were altered throughout any of the tests, maintaining the initial layer with 32 filters, the subsequent one with 16 filters, followed by a flattening layer, and finally, a dropout of 0.25.

By utilizing only this dataset and these specified parameters, the model achieved an accuracy of 84%.The second evaluation comprised maintaining the same parameters while exclusively using the second dataset. This model demonstrated an accuracy of 87.5%.

Ultimately, the decision was made to merge the two datasets to augment the number of samples for both training and testing, thereby enhancing the model's accuracy. Employing the same parameters, an accuracy of 92% was attained. By extending the number of training epochs to 55, which was the maximum allowable within the free version to avoid surpassing the time limit, an accuracy of 92.2% was achieved. This level of accuracy was deemed excellent and deemed sufficient for the model's requirements. Figure 5 demonstrates the model's accuracy and its confusion matrix.

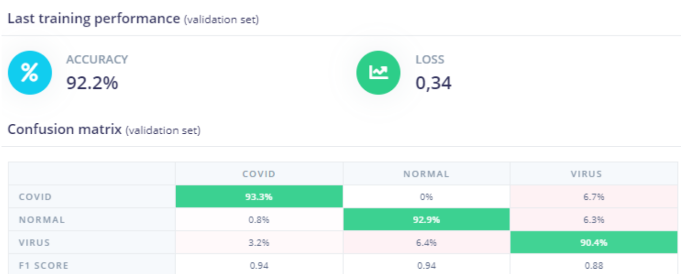

Fig. 5. Model accuracy and Confusion Matrix.

To conduct the model test, it was deployed on a smartphone to assess its functionality. Once deployed on the smartphone, X-ray images of lungs were searched with captions indicating Covid, Normal, or Virus. It is worth noting that none of the images contained any text during the model's training. It was observed that the model correctly predicts the vast majority of classifications but makes errors in some instances. Figure 6 illustrates a correct classification by the model for a Covid X-ray.

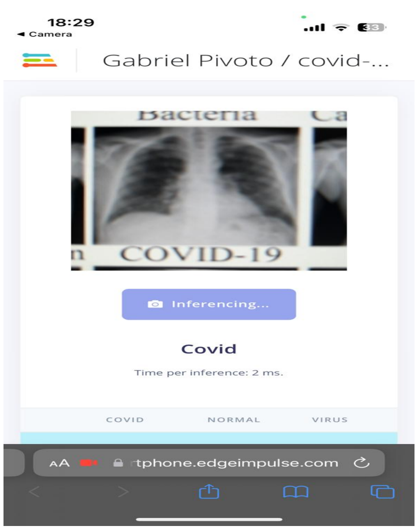

Fig. 6. Model classification of Covid

Furthermore, the model was evaluated using the exported code from Edge Impulse. Using the application container, a 91.26% accuracy was achieved after model training. It was observed that the classifications obtained with this compiled model are superior to those of the application deployed on the smartphone.

V. Conclusion

In conclusion, using Edge Impulse, this study employed a systematic approach to train and evaluate an AI-powered classification model for rapidly and accurately identifying COVID-19 in chest X-ray images. The results demonstrated the model's evolution through different training methodologies, initially utilizing separate datasets and subsequently merging them to enhance sample diversity. Notably, the model achieved an accuracy of 92.2% after 55 training epochs, surpassing the deemed sufficient threshold for its intended diagnostic requirements within the study scope.

In developing and training the Edge Impulse model, two pivotal blocks were incorporated to effectively capture the intricacies of the data: an image block and a classifier block. The image block, tailored for grayscale color depth, served as the initial stage where features were meticulously extracted from the input data. Subsequently, in the classifier block, the model underwent 55 training cycles with a learning rate of 0.0005. This critical phase embraced a 2D convolutional neural network (CNN) architecture, characterized by 32 filters in the initial layer, followed by another layer with 16 filters. The architecture also featured a flattening layer, facilitating the transition from convolutional to fully connected layers, and a dropout of 0.25, imparting a regularizing effect to enhance the model's generalization capability. This CNN architecture played a pivotal role in discerning complex patterns within the dataset, ultimately contributing to the model's robust performance and high accuracy in COVID-19 diagnosis from lung X-ray images.

The fusion of Edge Impulse technology with AI represents a groundbreaking effort in the realm of medical diagnostics, especially in the context of resource-constrained environments. Our exploration of the convergence of AI, Edge Impulse, and medical imaging addressed the pressing need for swift and accurate diagnostics, showcasing the platform's efficacy in facilitating on-device inference and minimizing dependence on centralized computing resources.

Looking ahead, future work in this domain should extend the current study's findings by exploring additional avenues to enhance further the effectiveness and generalizability of the AI-powered classification model developed on Edge Impulse. Specifically, a promising direction involves the replication and implementation of the convolutional neural network (CNN) architecture used in Edge Impulse within a Python-based framework. This comparative analysis would enable a comprehensive evaluation of the model's performance, shedding light on potential variations in results arising from different development environments. Furthermore, exploring alternative datasets or expanding the current dataset to encompass a more diverse range of chest X-ray images would contribute to the model's robustness and ensure its applicability across varied demographic groups. Additionally, investigating the impact of hyperparameter tuning and architecture modifications on the model's performance could provide valuable insights for further optimization.

Our findings contribute valuable insights into the synergistic application of AI and Edge Impulse, offering a practical and accessible solution for real-time COVID-19 diagnosis.

References

[1] S. Huang et al. "Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives". In: *International journal of biological sciences* 17.6 (2021), p. 1581.

[2] I. Ozsahin et al. "Review on diagnosis of COVID-19 from chest CT images using artificial intelligence". In: *Computational and Mathematical Methods in Medicine* 2020 (2020).

[3] S. Hymel et al. "Edge Impulse: An MLOps Platform for Tiny Machine Learning". In: *arXiv preprint arXiv:2212.03332* (2022).

[4] S. H. Hong. "Edge impulse machine learning for embedded system design". In: *Journal of Korea Society of Digital Industry and Information Management* 17.3 (2021), pp. 9–15.

[5] R. Rehouma, M. Buchert, and Y.-P. P. Chen. "Machine learning for medical imaging-based COVID-19 detection and diagnosis". In: *International Journal of Intelligent Systems* 36.9 (May 2021), pp. 5085–5115. DOI: 10.1002/int.22504. URL: https://doi.org/10.1002/int.22504.

[6] S. H. Kassania et al. "Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach". In: *Biocybernetics and Biomedical Engineering* 41.3 (July 2021), pp. 867–879. DOI: 10.1016/j.bbe.2021.05.013. URL: https://doi.org/10.1016/j.bbe.2021.05.013.

[7] A. Saygılı. "A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods". In: *Applied Soft Computing* 105 (July 2021), p. 107323. DOI: 10.1016/j.asoc.2021.107323. URL: https://doi.org/10.1016/j.asoc.2021.107323.

[8] R. Chauhan, K. K. Ghanshala, and R. Joshi. "Convolutional neural network (CNN) for image detection and recognition". In: *2018 first international conference on secure cyber computing and communication (ICSCCC)*. IEEE. 2018, pp. 278–282.