Image Downsizing and Compression Impacts on AI-based Medical Image Classification

Edson M. Hung, Renam C. da Silva, Andrey O. O. dos Reis, Darlington Akogo

***Abstract*—In recent years, machine learning has been used for automatic medical image diagnosis. Solutions leveraging machine learning can help physicians in the diagnostic process and also reduce the time spent by medical experts analyzing images and video frames to conclude their assessments. In the studied scenario, images are captured at resource-constrained remote basic health facility and sent to a cloud-based inference model, which returns the results to physicians. We investigate the impact of straightforward low-bit-rate image coding solutions on the classification performance of neural network models targeted at cloud-based image diagnosis solutions. Our experiments show that it is possible to lower the bit rate needs without significant harm to the prediction accuracy of the models using both downsizing and compression. These findings provide evidence of the viability of deploying automated diagnostic systems as a Service over constrained communication infrastructures to assist remote areas.**

***Index Terms*—Machine learning, image classification, compression, JPEG, JPEG2000, downsizing.**

I. Introduction

RECENT years have seen a revolution in the application of machine learning techniques, particularly those based on deep learning, to different problems such as understanding human speech, competing at a high level in strategic game systems such as chess and Go, autonomous vehicle navigation, intelligent routing in content delivery networks, and interpreting complex data, including visual data such as images and videos. We are specifically interested in the transmission of medical images to a cloud infrastructure that implements automatic diagnosis Software as a Service (SaaS), leveraging artificial intelligence techniques. This scenario is illustrated in Fig. 1. With SaaS, end users at remote medical facilities do not need to spend resources to manage any software or maintain dedicated hardware infrastructure. Instead, they can log in with a web browser and connect to the service. SaaS is of great utility in remote locations lacking medical specialists and having severe communication and computational constraints, a condition commonly encountered in several countries and with devastating effects, as witnessed in pandemic times. This type of solution can also contribute to the development of more generalized models. Data-driven learning systems usually benefit from exposure to data diversity resulting from different geographical locations, gender, ethnicity, and age groups. The gathered data may be used to foster further development, produce better models, and lower mistrust in the use of artificial intelligence solutions in medical applications. From the SaaS provider's perspective, system maintenance would be quite easy and practical. For instance, an improved model could be put to run after retraining or using adaptive learning techniques.

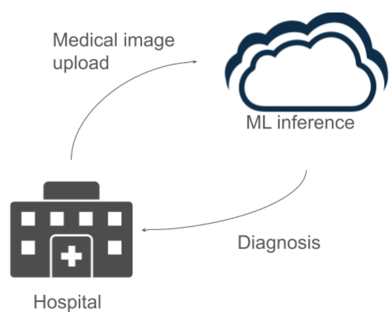


Fig. 1: Application scenario of a software as a service cloud-based automatic medical image diagnosis.

We aim to study the impact of compression and downsizing of medical images on the classification performance of models based on convolution neural networks (CNN) employed as a tool to aid in the diagnosis of two diseases. Loosely speaking, both downsizing and compression are used to selectively ditch information in order to achieve a smaller representation (file size) and fit into a communication channel or save storage space, at the cost of quality degradation. This degradation is most commonly studied from the perspective of consumption by the human visual system. It turns out, in the considered application scenario, the images are meant to feed an artificial intelligence (AI) algorithm for classification. And thus, the rate-distortion trade-off may be different from that for human consumption. For instance, artifacts introduced due to compression may be annoying for human visualization but unimportant for an AI model. On the other hand, compression may get rid of discriminating image features, leading to misclassification and poor performance.

We use off-the-shelf image coding solutions adopted in the DICOM[1] container standard, namely JPEG and JPEG2000, which are lossy compression algorithms. These image codecs,

Edson M. Hung is with the Electrical Engineering Department, Universidade de Brasília, Darcy Ribeiro Campus. ORCID: 0000-0003-1457-1747, e-mail: mintsu@unb.br

Renam C. da Silva is with the Electrical Engineering Department, Universidade de Brasília, Darcy Ribeiro Campus. ORCID: 0000-0002-6147-4475, e-mail: renam.silva@unb.br

Andrey O. O. dos Reis is a doctoral student at the Electrical Engineering Program, Universidade de Brasília. ORCID: 0009-0009-3001-727X, e-mail: andreyotacilio@gmail.com

Darlington Akogo is with minoHealth AI Labs. ORCID: 0000-0001-5652-1815, e-mail: darlington@gudra-studio.com

Digital Object Identifier: 10.14209/jcis.2026.2

Submission: 2025-05-19, First decision: 2025-06-23, Acceptance: 2026-01-15, Publication: 2026-01-28

[1]Digital Imaging and Communications in Medicine (DICOM)

respectively, introduce compression artifacts such as blocking and ringing [1]. These artifacts may reduce the AI model performance and should also be taken into consideration in the system design. Image resizing is also a common operation in many computational solutions for reducing bandwidth.

This work results from the activities led by the Focus Group on Artificial Intelligence for Health (FG-AI4H), a joint effort between the International Telecommunication Union (ITU) and the World Health Organization (WHO) to establish a standardized assessment framework for evaluation of AI-based methods for health, diagnosis, triage, or treatment decisions. In this context, we have developed a library to test the impact on the performance of medical image classifiers in a scenario where the image is compressed and transmitted to SaaS solution. This library will be made publicly available shortly.

II. Related work

Several studies have investigated to what extent compression algorithms such as JPEG and JPEG2000 affect predictive models based on neural networks in visual analysis tasks such as natural image classification and forensic tasks. In [2] the authors discuss how image quality affects deep neural network models employed for the classification of natural images. The authors carried out an evaluation of four state-of-the-art deep neural network models for image classification under five quality distortions, including blur, noise, contrast, JPEG, and JPEG2000 compression. They pointed out that deep neural models are commonly trained using high-quality images, but in many practical applications, such an assumption may not hold as the images may suffer degradation at several steps of the processing pipeline. The studied deep neural models are found to be quite susceptible to blur and noise distortions.

Mandelli *et al.* [3] focused on the effect that JPEG compression has on the training of convolutional neural networks applied to multimedia forensic tasks and image classification. Specifically, they consider the issues that arise from JPEG compression and misalignment of the JPEG grid for the problems of camera model identification, detection of synthetic images generated by generative adversarial networks, and object recognition. They show that one needs to consider these effects when generating a training dataset in order to properly train a forensic detector not losing generalization capability, whereas the model for the object recognition task is much more robust to JPEG compression.

An increasing interest in the interplay of image/video compression solutions and the performance of learning-based models has led to the rise of compression for machines. Besides filling a screen for human consumption, visual information is now more than ever meant for machine consumption, to feed into classification, detection, and segmentation algorithms in several application domains. A further step in this direction is to design end-to-end image and video compression solutions directly optimized for the target visual analysis task [4–8]. Here, however, we focus on the use of off-the-shelf standard image codecs which are more mature, pervasive, and computationally affordable for medical images in the considered application scenario.

Recently, a few research works have turned attention to the impact of compression on the performance of classification models designed for medical image modalities, which may exhibit different textural and structural characteristics from that of natural images. In [9], Zanjani *et al.* investigate the impact of lossy JPEG2000 compression on their deep convolutional neural network model for metastatic cancer detection in histopathological images from breast lymph nodes. They brought attention to the controversial use of lossy compression in medical images, but advocated for the use as long as compression does not interfere with diagnosis, and to cope with storage and transmission challenges to enable cloud-based computer-aided diagnosis (CAD) systems. Their solution, based on a convolutional neural network, is compared to a pathologist's diagnosis in different experimental setups. They found that their model trained on uncompressed high-quality images is robust up to a certain level of compression of test images. The authors highlighted that their findings are specific to the considered application scenario and further studies should be done.

Jo *et al.* [10] studied the impact of lossy JPEG2000 compression on the performance of a predictive model based on deep learning applied to the classification of mammograms into three classes associated with the diagnosis labels normal, benign, and malignant. They carried out a set of experiments to investigate the performance of different models resulting from the training over several versions of the dataset, compressed at different compression ratios. The authors stated that moderate compression ratios do not produce a substantial impact on the classification performance of models.

We further investigate the impact of compression on the classification performance of deep learning models devised for medical image classification. To get a better picture, we have chosen models devised for two different medical image modalities: X-ray images and magnetic resonance images (MRIs). In addition to JPEG2000 compression investigated in previous works, we also study the impact of JPEG compression along with a low-cost, yet effective variant comprising downsizing before compression and interpolation after decoding, which is quite appealing to the application scenario considered in this work.

The remainder of this work is organized as follows. Section III briefly reviews the standard off-the-shelf image coding solutions used. Section IV presents the predictive models whose performances are evaluated before and after coding the test dataset at various compression ratios. Section V discusses the experiments carried out and the obtained results. Finally, the conclusions are presented in Section V.

III. Image coding

To assess the impact of coding on the performance of predictive models devised to classify radiology images, we have chosen two of the most popular image coding algorithms, namely JPEG and JPEG2000. In the sequel, they are briefly described. The reader is referred to the cited works for a thorough treatment of the image codecs discussed in this section.

A. JPEG

The JPEG [11, 12] image coding standard was introduced in the early 1990s and still is widely used in a range of applications, its long-standing success results mainly from back-compatibility needs and computational affordability. Among the four coding modes provided, the most common sequential encoding proceeds as follows. The input image $f(x, y)$ is partitioned in blocks of 8×8 samples. The 8×8 blocks are scanned from left to right and top to bottom order, each 8×8 block of samples is then transformed to a suitable domain, the resulting coefficients at the transformed domain are scalar-quantized and entropy encoded. Fig. 2 depicts schematically the steps carried out by an encoder to produce a JPEG compressed bitstream, as well as the steps to get back an approximated version of the input image. The key coding tools are:

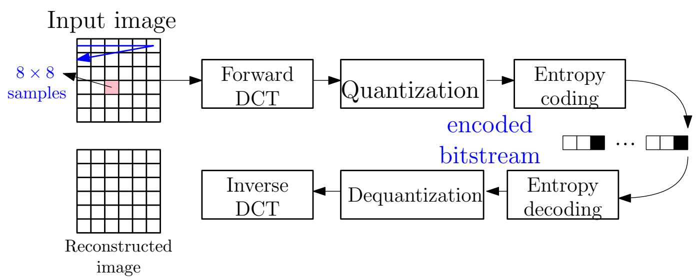


Fig. 2: JPEG processing chain to produce a JPEG-compliant bitstream.

- *Forward DCT*: taking advantage of the energy compaction property, the discrete cosine transform (DCT) is applied in an attempt to produce a representation with only a few non zero coefficients. The resulting coefficients account for the contribution of orthogonal basis functions with varying spatial frequency. At the decoder side, the inverse DCT step does the opposite, that is, weights the basis functions by the transform coefficients to get back the input image block.
- *Quantization*: scalar uniform quantization takes place to map transform coefficients to a set of just a few discrete amplitudes, which turns out zeroing low amplitude transform coefficients and reducing bit rate. The resulting quantized coefficients are arranged in an intermediate format before entropy coding. Quantization is a many-to-one mapping; therefore information is lost, resulting in reconstruction artifacts. An encoder can trade off rate reduction and quality degradation by varying the quantization step size. Dequantization takes place at the decoder side to recover the transform coefficients from quantization bins.
- *Entropy coding*: aiming at compression efficiency, the quantized transform coefficients are scanned in a zigzag fashion and arranged in an intermediate format to exploit long chains of zero coefficients. A special treatment is given to the first transform coefficient, the so-called DC coefficient [11]. After that, entropy coding assigns binary codes based on statistical characteristics, giving shorter codes to frequent symbols and longer codes to less frequent symbols. At the decoder side, the statistical model shared between encoder and decoder is used to decode the compressed bitstream.

Interpolative JPEG

In [13], a JPEG-based interpolative image coding scheme was proposed. In it, image downsizing is performed before JPEG encoding, whereas reconstruction is obtained by carrying out image interpolation after JPEG decoding. In comparison to standard JPEG, the interpolative scheme has shown competitive coding efficiency and significantly reduced blocking artifacts, overall yielding a much more pleasant visual result [13]. Although important for visual consumption, the visual aspect of image coding is not of primary concern, but rather the classification performance of machine learning models. This approach is quite appealing to the considered application scenario, downsizing is an affordable operation to be performed on the client side, and it reduces the computational burden for the compression step.

B. JPEG2000

In addition to JPEG, we have experimented with JPEG2000 as the compression solution prior to image classification. The JPEG2000 standard was defined in early 2000 and brought features amenable to internet image applications, such as quality scalability and embedded bitstream. Considering our use of JPEG2000, we briefly discuss its main tools.

After a pre-processing step that takes the input image to a compression-suited color space, and next splits the color-transformed image into tiles, JPEG2000 starts by decomposing each component of the input image into a set of subsampled spatial frequency subband images [14] using an analysis Discrete Wavelet Transform (DWT). Afterward, another step of DWT analysis is carried out on the low-frequency subband image resulting from the first DWT analysis. In fact, the DWT is repeated in such a way to produce a hierarchical multi-resolution representation of the input image. After image decomposition, samples of the hierarchical representation are grouped into the so-called code-blocks, typically blocks of 32 $\times$ 32 or 64 $\times$ 64 samples. Each code block is encoded in a finely embedded bitstream providing quality scalability. In the sequel, a selection of JPEG2000 modules is briefly discussed; the reader is referred to [14–16] for a detailed treatment of the JPEG2000 coding solution.

- *DWT analysis*: aims to exploit the spatial redundancy as well as to produce a resolution-scalable representation of the input image. The standard supports two transform options, the irreversible and reversible, implemented by two different filters [16]. Image reconstruction may be obtained at a given level of the hierarchical structure by composing the responses of the synthesis DWT to the immediate lower subbands images.
- *Quantization*: the DWT analysis responses within the code-blocks are scalar-quantized before entropy coding, JPEG2000 provides for different quantization schemes, including a uniform scalar dead-zone quantizer that directly maps DWT analysis responses to quantization bins.

An embedded scalar quantization scheme is built upon the uniform scalar dead-zone quantizer to allow progressive decoding. It exploits the equivalence of dropping bits from quantization bins resulting from a fine quantizer and modulating the quantization step size. This allows for progressive coding of the quantization bins bit planes.

- *Entropy coding*: JPEG2000 employs context-adaptive modeling and arithmetic encoding to create a compressed bit stream. The quantized DWT responses within a code block are decomposed into bit planes to facilitate constructing an embedded bit stream. For better probability estimation and efficient coding of the binary symbols, JPEG2000 manages several contexts derived from the immediate neighboring symbols; each context maintains a separate probability model. In short, JPEG2000 employs several thoughtful schemas to reduce the number of bits assigned while providing a scalable bitstream.

In addition to the standard JPEG2000, we also experimented with an interpolative scheme similar to the interpolative JPEG described previously.

IV. Predictive neural models

Considering the application scenario of a SaaS-based cloud system for automatic medical image diagnosis, we have selected two predictive models to study the effect of image coding. Specifically, we employed two convolutional neural network models designed to perform disease detection through image classification. Since this work is part of an initiative to establish a standardized assessment framework for evaluating AI methods in radiology, the chosen models were selected to align with that objective. The selected architectures provide a balanced trade-off between classification performance and computational cost, making them well-suited to the intended scenario. Both can be executed on commodity GPUs, thus avoiding the need for costly infrastructure. Moreover, these models remain influential and continue to serve as the foundation for several recent solutions [17]. We have used models pre-trained on the ImageNet[18] dataset, then we carried out fine-tuning over a large dataset of input-label pairs comprised of medical images and associated annotations gathered from patients assessed by medical experts. The goal here is to reproduce the results of the referenced works and produce the predictive models to be assessed under different compression strategies of the input image. Next, we briefly present the most general specifications for the predictive models used.

A. Covid model

The first predictive model considered, the so-called COVID-Net [19], was devised as a tool to fight the Coronavirus pandemic by helping screen patients using chest radiography. The adopted implementation COVID-Next is an open-source [20] Pytorch implementation[2] inspired by the COVID-Net[19]. The model was designed to classify a chest X-ray (CRX) into three different classes: normal, pneumonia, or COVID-19.

The model has the ResNeXt-50 (34×4d) as its backbone [21], a network with aggregated residual transformations based on VGG [22] and Resnets [23]. The ResNeXt-50 design is highly modularized and features a stacking topology of the residual blocks.

On top of the backbone, a final block with three convolution layers was added. The architectural details of the COVID model are given in Tab. I. As for the training dataset, we used the same dataset as [19], namely COVID chest x-ray [24] and the one from RSNA Pneumonia Detection Challenge [25]. The COVID chest x-ray is an open-access collection designed to support research on COVID-19 and related respiratory diseases. It includes chest x-ray and CT images from patients who tested positive or were suspected of having COVID-19, as well as other viral and bacterial pneumonia. The data are gathered from public repositories and through indirect contributions from hospitals and physicians. The RSNA Pneumonia Detection Challenge dataset was created to support the development of machine learning algorithms capable of detecting pneumonia in chest X-rays automatically.

B. Brain tumor model

The second predictive model considered is a model devised to carry out the classification of 3D volumetric magnetic resonance images (MRIs) from brains. The spatiotemporal neural network model [26] treats slices of 3D MRI as a sequence of images over time.

The model is trained in a large dataset of images obtained from healthy brains and brains stricken by two types of tumors: Low-grade glioma (LGG) and High-grade glioma (HGG). The healthy brain samples images were collected from IXI Dataset [27], whereas pathological brain images were acquired from Brain Tumor Segmentation (BraTS) 2019 dataset [28], T1 contrast-enhanced MRIs were used in the model. The IXI dataset is a publicly available collection of nearly 600 MRI scans from healthy subjects, aimed at supporting research in medical image analysis and brain modeling. All images are provided with corresponding demographic information, making it a valuable benchmark for brain imaging studies, MRI synthesis, and machine learning applications in neuroimaging.

The neural network model is built on top of the ResNet Mixed Convolution [29], a deep residual neural network comprised of a stem layer with a 3D convolution and a layered structure of one 3D and three 2D residual structures. There is a pair of residual units at each residual (3D or 2D) structure. The output of each residual unit consists of the addition of its input (shortcut connection) with the result of the transformations carried out within the block. Between each residual unit within the 2D residual structure, there is a 3D layer with kernel size 1 and stride 2, to halve the input. Each convolution layer is followed by a 3D batch normalization layer and a ReLU activation function. The architectural details of the Brain Tumor model are given in Tab. I, the reader is referred to [29] for further details.

C. Implementation details

In the training stage, aimed at producing the predictive models to be assessed, both models take as input 256×256-

[2]The Pytorch implementation has 5x fewer parameters compared to COVID-Net and achieves equivalent performance.

TABLE I: Architecture of the evaluated convolutional neural networks. (**Left**) COVID-Next was built upon ResNext50 architecture. (**Right**) Brain tumor model with architecture based on ResNet mixed convolution. Layer blocks are inside the brackets, with the quantity of blocks stacked outside the brackets. C=32 in COVID-Next indicates the amount of convolutions grouped with 32 groups.

Stage	Output	COVID-Next	Output	Brain Tumor
conv1	128x128	7x7, 64, stride 2	Lx56x56	3×7×7, 64, stride 1×2×2
conv2	64x64	3x3 pool, stride 2 $\begin{bmatrix} \text{1x1}, 128 \\ \text{3x3}, 128, C=32 \\ \text{1x1}, 256 \end{bmatrix} \times 3$	Lx56x56	$\begin{bmatrix} \text{3x3x3}, 64 \\ \text{3x3x3}, 64 \end{bmatrix} \times 2$
conv3	32x32	$\begin{bmatrix} \text{1x1}, 256 \\ \text{3x3}, 256, C=32 \\ \text{1x1}, 512 \end{bmatrix} \times 4$	$\frac{L}{2}$x28x28	$\begin{bmatrix} \text{1x3x3}, 128 \\ \text{1x3x3}, 128 \end{bmatrix} \times 2$
conv4	16x16	$\begin{bmatrix} \text{1x1}, 512 \\ \text{3x3}, 512, C=32 \\ \text{1x1}, 1024 \end{bmatrix} \times 6$	$\frac{L}{4}$x14x14	$\begin{bmatrix} \text{1x3x3}, 256 \\ \text{1x3x3}, 256 \end{bmatrix} \times 2$
conv5	8x8	$\begin{bmatrix} \text{1x1}, 1024 \\ \text{3x3}, 1024, C=32 \\ \text{1x1}, 2048 \end{bmatrix} \times 3$	$\frac{L}{8}$x7x7	$\begin{bmatrix} \text{1x3x3}, 512 \\ \text{1x3x3}, 512 \end{bmatrix} \times 2$
end	8x8	$\begin{bmatrix} \text{3x3}, 512 \\ \text{1x1}, 1024 \\ \text{3x3}, 512 \end{bmatrix} \times 1$	...	...
	1x1	average pooling, 3-d fully connected softmax	1x1x1	spatiotemporal pooling, 3-d fully connected softmax

pixel images. The input images fed to the models result from a pre-processing step. For COVID-Next, the images are resized to 256×256 pixels, then sequentially subjected to random horizontal flip, vertical flip, and color jitter with even probability. The input to the Brain tumor model is subjected to intensity rescaling and left-right random flip.

In the case of COVID-Next, the model was trained for 40 epochs using 64 image batches using the Adam optimizer with a learning rate of 10^{-4}. The learning rate is reduced on plateaus with a factor of 0.7 and patience of 5 epochs. The brain tumor model had 10 epochs of training using Adam optimizer, batch size 1, and a learning rate of 10^{-5}.

V. Experimental Results

Once we have the predictive models ready to perform inference, we aim to assess their tolerance to degradation in the input image due to compression and downsizing. In other words, we aim to determine a setup able to reduce the bandwidth required to send medical images over constrained channels without severely degrading the performance of prediction models, and whether a SaaS solution can meet such stringent requirements. To this end, we have performed a set of experiments as described next.

Regarding the predictive model for chest X-ray image classification, our baseline model is the COVID-Next [20], a COVID-19 classifier based on the COVID-Net proposed by Wang *et al.* [19]. This model was trained using chest radiography with different resolutions, qualities, and artifacts. The dataset has 15,572 images from which 13,811 are used as training samples and 1,761 for testing. The test accuracy of this model is 93.63% on the pristine test dataset. It is worth stressing that the COVID-Next test dataset is not raw (uncompressed data) but rather slightly compressed (perceptually near-lossless). Conveying pristine images over bandwidth-constrained channels would imply high latency and inefficiency, causing delays in diagnosis and congestion in the communication channel. To cope with such issues, the images may be compressed to reduce bandwidth before transmission to a remote classifier in the cloud, where the analysis takes place using a neural network model trained over a proper dataset.

In order to study such a scenario, the test dataset was compressed with different quality parameters. Experimental results reveal it is possible to achieve significant bit rate savings with a negligible accuracy reduction to enable automatic diagnosis SaaS over bandwidth-constrained communication channels.

Before discussing any experimental results, we should first make clear that the data ratio compression used in our charts is defined as the ratio between the uncompressed data size/volume and the compressed data size/volume, in practice computed as a ratio between files. Thus, for the same level of classification performance, the higher the compression ratio, the better.

Fig. 3-a shows a chest X-ray image classified by a radiologist as normal. By compressing such an image at a low bit rate, blocking artifacts and loss of image details due to JPEG compression become an issue as shown in Fig. 3-b. Similarly, a reduction in image details can be seen in the JPEG2000 compressed image shown in Fig. 3-c. In such cases, compression artifacts [1] as blurring, blocking in the case of JPEG, and ringing in the case of JPEG2000, induced the model to misclassify this image as Covid-19. These effects are well known in lossy image compression and are due to the quantization of transformed coefficients as briefly discussed in Section III. The quantization process gets rid of image information to lower the bit rate, but adversely reduces the image reconstruction quality as well.

On the other hand, by reducing the amount of image information downsizing before lossy compression, it is pos-

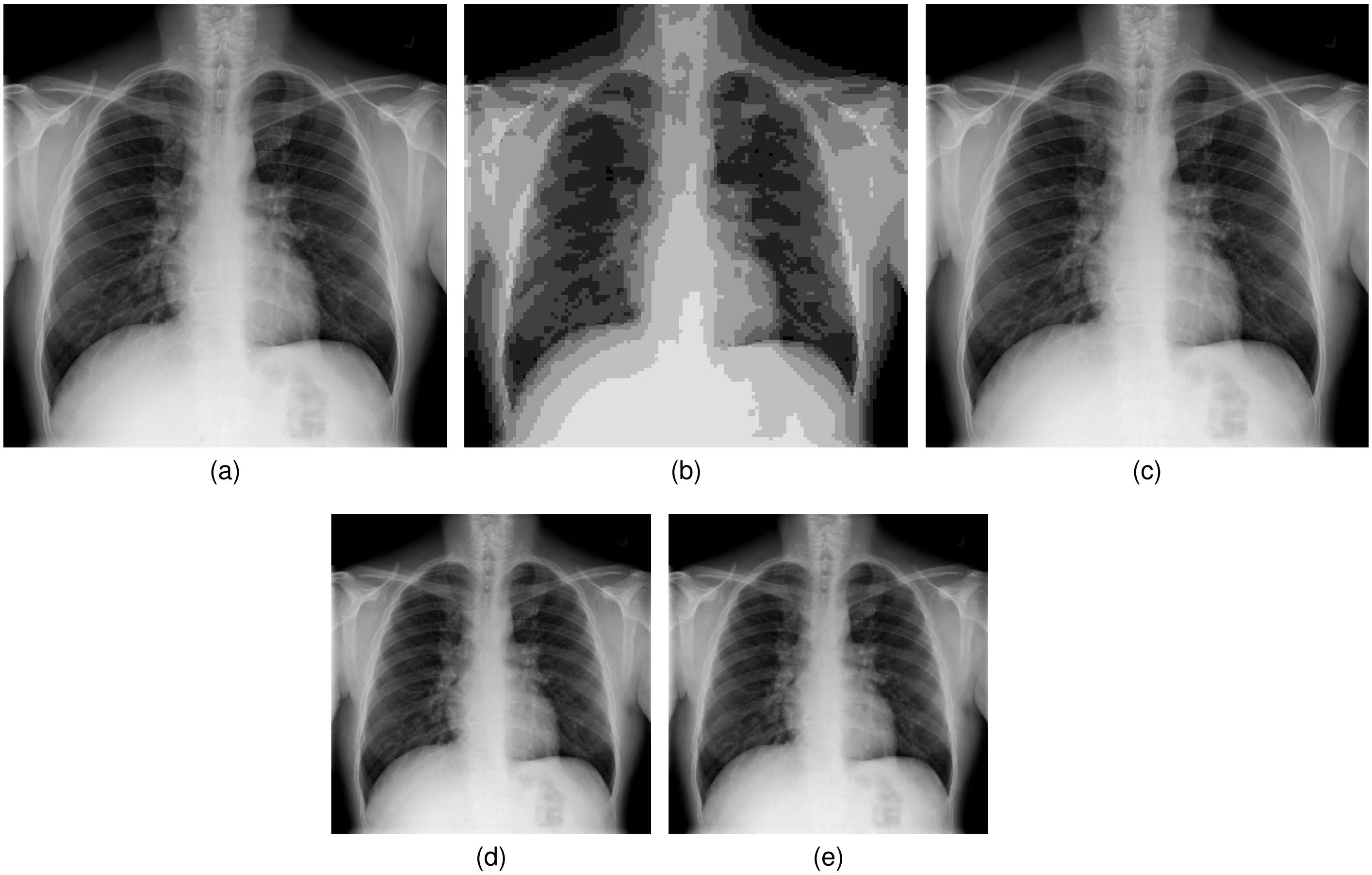


Fig. 3: Subjective comparison of the images: (a) original with 1024×1024 pixels, (b) JPEG-compressed at 29.155 compression ratio, and (c) JPEG2000-compressed at 31.949 compression ratio. Subjective comparison of the downsized 256×256-pixel images: (d) JPEG-compressed at 43.29 compression ratio, and (e) JPEG2000-compressed at 53.191 compression ratio.

sible to achieve a higher compression ratio while maintaining the characteristics necessary for the correct classification. Despite the visual quality degradation due to compression, the compression artifacts are quite reduced as a result of the downsizing before compression. In Fig. 3-d and Fig. 3-e we can subjectively compare the downsized compressed images with the original one shown in Fig. 3-a, the downsized compressed images have better subjective quality than the ones compressed at the original resolution despite requiring fewer bits per pixel on average. We hypothesize that this behavior is linked to the image content covered by the underlying transform. For JPEG, DCT block spans a larger area in the downsampled image relative to the original, introducing less discontinuity, preserving useful features.

Fig. 4 shows the confusion matrices for different scenarios. In Fig. 4-a, the original test set was used to compute the confusion matrix; this result is used as a benchmark for our comparative analysis. Fig. 4-b shows the confusion matrix for the case where the test set is compressed with JPEG with an average of 28.490 compression ratio, no downsizing in this case. The model seems to become biased to misclassify the inputs as Covid-19. However, an improved result can be achieved using JPEG2000 with a quite higher compression ratio 46.296 as shown in Fig. 4-c. In the case where the test set is downsized before compression, either with JPEG or JPEG2000, we can achieve a better trade-off between compression ratio and classification performance. At a compression ratio of 38.911 for JPEG and 45.872 for JPEG2000, the classification model shows a much better performance. As shown in Fig. 4-d and 4-e, the accuracy of the classification model is kept comparable to that resulting from the original images.

Fig. 5 shows a range of operational trade-off points. Examining the curves, we can see that the accuracy is significantly reduced due to compression. The accuracy drops sharply as the average compression ratio increases. In an extreme scenario, we resize the images in the dataset to 256×256 pixels using a Lanczos-4 filter before performing compression; the image size was chosen due to the COVID-Next input dimension. The results for the downsized images compressed with JPEG and JPEG2000 are shown in Fig. 5. In this scenario, the size of the bitstream is significantly reduced, but the accuracy is also significantly reduced, showing that severe compression is detrimental to the COVID-Next as the image quality degrades. However, it is possible to obtain a configuration where the accuracy is kept comparable with that resulting from classify-

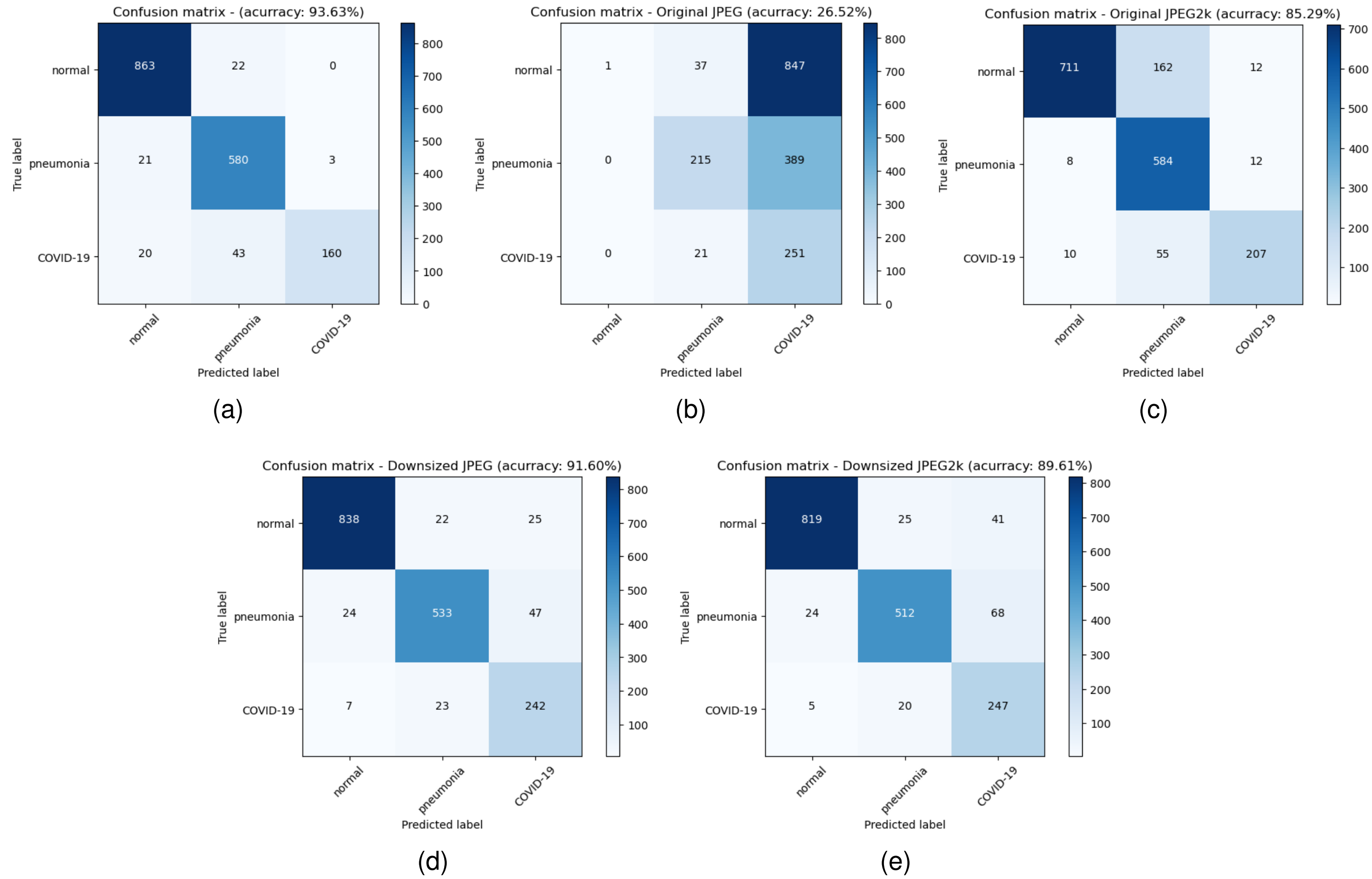


Fig. 4: COVID-Next – Confusion matrix comparison for different compression ratios: (a) original without compression, (b) JPEG compression at 28.490 average compression ratio, (c) JPEG2000 compression at 46.296 average compression ratio. (d) downsized JPEG compression at 38.911 average compression ratio and (e) downsized JPEG2000 compression at 45.872 average compression ratio.

ing pristine images, though at the expense of a significantly smaller bitstream.

We have carried out the same set of experiments for the Brain Tumor model. This model was trained using images of slices of the 3D MRI acquired from brains. The dataset has 594 images from which 415 are used as training samples and 179 for testing. The test accuracy of this model is 95.53% on the pristine test dataset. As in the case of COVID-Next, the Brain Tumor's test dataset is not raw, uncompressed data, but rather slightly compressed, perceptually near-lossless.

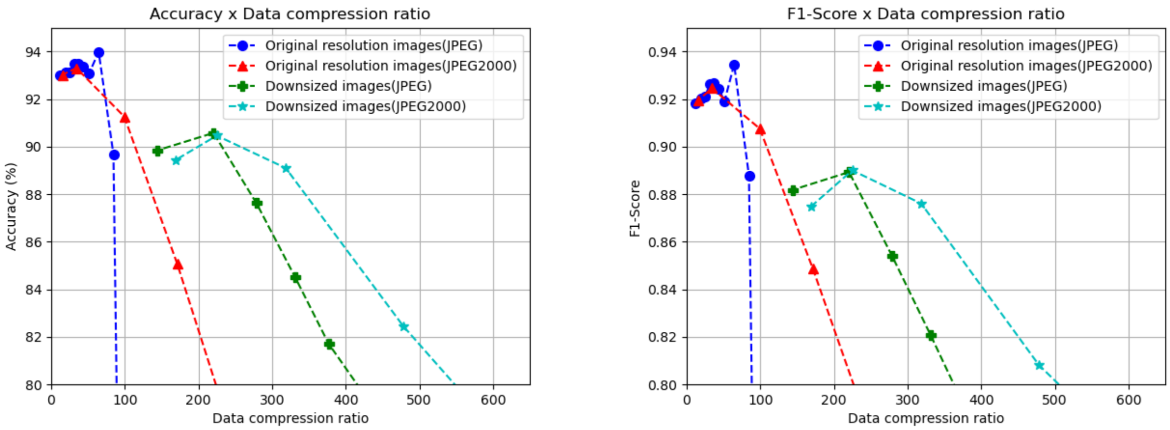
Fig. 6 shows a range of operational trade-off points for the Brain Tumor model. The different curves resulted from the same coding setups as for the COVID-Next model. The difference observed in the Brain Tumor model's behavior regarding the COVID-Next can be associated with the degree of prior compression of the dataset images. Different from the COVID-Next, downsizing had no benefit in terms of better compression ratio and prediction trade-off. However, up to a certain degree, compression still brings benefits as it reduces the latency and bandwidth without significant prediction performance degradation.

Overall, the set of experiments shows the influence of compression artifacts in medical image classification. The results show that, in both cases, there is a combination of downsizing and compression quality that leads to significant bit rate savings without severely impairing the accuracy of the classification model. Additionally, this combination is better than using only compression, both in terms of bitstream reduction and accuracy impact.

To evaluate the effect of image compression in the scenario considered in this work, we have developed a library that calculates a set of metrics such as accuracy, sensitivity, specificity, and F1-score. The library will be publicly available to support further research on the influence of compression on medical image classification.

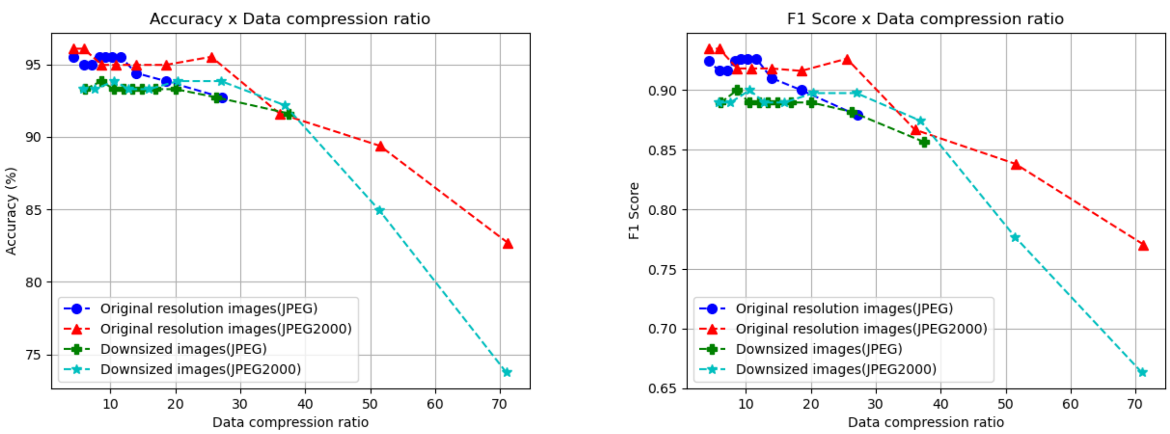
VI. Conclusion

We have investigated the impact lossy compression has on the performance of predictive models devised for the classification of medical images. Two popular off-the-shelf standard lossy compression solutions, JPEG and JPEG2000, were selected due to their maturity, pervasiveness, and computational affordability. The two solutions were used with and without a prior downsizing step to compress the test dataset.



(a) (b)

Fig. 5: Impact of compression on: (a) accuracy of the COVID-Next classifier; (b) F1 score of the COVID-Next classifier. The dataset images were compressed with different compression ratios using JPEG or JPEG2000. The downsized images were scaled to 256×256 pixels before compression. The model's accuracy on the raw images is 93.63% and the F1-score is 0.9384.



(a) (b)

Fig. 6: Impact of compression on: (a) accuracy of the Brain Tumor classifier; (b) F1 score of the Brain Tumor classifier. The dataset images were compressed with different compression ratios using JPEG or JPEG2000. The downsized images were scaled to 256×256 pixels before compression. The model's accuracy on the raw images is 95.53% and the F1-score is 0.9245.

In turn, the compressed dataset is subjected to the analysis of the prediction models, and the results are compared to the results obtained for the pristine test dataset, thus enabling us to assess the impact of standardized compression solutions on the performance of the models. We have chosen two distinctive models devised for different problem domains and medical image modalities. The experimental results have shown that it is possible to reduce the bit rate without significant harm to the prediction accuracy of the models, thus allowing image transmission under severe bandwidth constraints. Such results suggest the viability of implementing automatic diagnosis SaaS to cope with poor communication infrastructure in remote areas or developing countries. The algorithms used to obtain the results will be published as a Python library and adopted by the AI4H (Artificial Intelligence for Health), a focus group of the ITU in partnership with the World Health Organization (WHO), aimed at establishing a standardized assessment framework for the evaluation of AI-based methods for health, diagnosis, triage, or treatment decisions. This tool will define the best strategy to obtain a good transmission rate reduction without a significant performance drop. This is an ongoing initiative of the AI4H focus group on AI for radiology [30].

VII. Limitations and Future Work

Lossy compression in the domain of medical imaging still raises skepticism. Nonetheless, in remote regions facing shortages of medical specialists and stringent communication and computational constraints, it can represent a practical solution. We have provided evidence that one can operate a standard image codec to lower the bitrate while introducing negligible performance impairment to CNN-based classification models targeted at disease detection. However, further investigations are needed to determine whether these findings apply to different model architectures, imaging modalities, and data resolutions. It is also important to assess whether CNN models attend to similar regions in both the compressed and the raw image. Additionally, one can explore the interplay of compression and classification by devising a CNN-based classification model to operate directly in a partially compressed domain, leveraging JPEG coefficients as input data.

References

[1] E. Allen, S. Triantaphillidou, and R. Jacobson, "Image Quality Comparison Between JPEG and JPEG2000. I. Psychophysical Investigation," *Journal of Imaging Science and Technology*, vol. 51, no. 3, pp. 248–258, 2007, doi: 10.2352/J.ImagingSci.Technol.(2007)51:3(248).

[2] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," 2016, doi: 10.48550/arXiv.1604.04004.

[3] S. Mandelli, N. Bonettini, P. Bestagini, and S. Tubaro, "Training CNNs in Presence of JPEG Compression: Multimedia Forensics vs Computer Vision," 2020, doi: 10.48550/arXiv.2009.12088.

[4] S. Wang, Z. Wang, S. Wang, and Y. Ye, "End-to-End Compression Towards Machine Vision: Network Architecture Design and Optimization," *IEEE Open Journal of Circuits and Systems*, vol. 2, 2021, doi: 10.1109/OJ-CAS.2021.3126061.

[5] S. Yang, Y. Hu, W. Yang, L.-Y. Duan, and J. Liu, "Towards Coding for Human and Machine Vision: Scalable Face Image Coding," *IEEE Transactions on Multimedia*, vol. 23, 2021, doi: 10.1109/TMM.2021.3068580.

[6] L. D. Chamain, F. Racapé, J. Bégaint, A. Pushparaja, and S. Feltman, "End-to-end optimized image compression for machines, a study," 2020, doi: 10.1109/DCC50243.2021.00024.

[7] ——, "End-to-end optimized image compression for multiple machine tasks," 2021, doi: 10.1109/DCC50243.2021.00024.

[8] L. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, "Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics," *IEEE Transactions on Image Processing*, vol. 29, 2021, doi: 10.1109/TIP.2020.3016485.

[9] F. G. Zanjani, S. Zinger, B. Pieper, S. Mahmoudpour, P. Schelkens, and P. H. N. de With, "Impact of JPEG 2000 Compression on Deep Convolutional Neural Networks for Metastatic Cancer Detection in Histopathological Images," *Journal of Medical Imaging*, vol. 6, no. 2, 2019, doi:10.1117/1.JMI.6.2.027501.

[10] Y.-Y. Jo, Y. S. Choi, H. W. Park, J. H. Lee, H. Jung, H.-E. Kim, K. Ko, C. W. Lee, H. S. Cha, , and Y. Hwangbo, "Impact of Image Compression on Deep Learning-based Mammogram Classification," *Scientific Reports*, vol. 11, no. 7924, 2021, doi: 10.1038/s41598-021-86726-w.

[11] G. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. 18–34, 1992, doi: 10.1109/30.125072.

[12] G. Hudson, A. Léger, B. Niss, and I. Sebestyén, "JPEG at 25: Still Going Strong," *IEEE MultiMedia*, vol. 24, no. 2, pp. 96–103, 2017, doi: 10.1109/MMUL.2017.38.

[13] B. Zeng and A. Venetsanopoulos, "A jpeg-based interpolative image coding scheme," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 1993, pp. 393–396 vol.5, doi: 10.1109/ICASSP.1993.319830.

[14] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for Interactive Imaging," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1336–1357, 2002, doi: 10.1109/JPROC.2002.800725.

[15] M. W. Marcellin, M. A. Lepley, A. Bilgin, T. J. Flohr, T. T. Chinen, and J. H. Kasner, "An overview of quantization in JPEG 2000," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 73–84, 2002, doi: 10.1016/S0923-5965(01)00027-3.

[16] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2000, doi: 10.1109/30.920468.

[17] Tae-hoon Kim and Asadi Srinivasul and Ravikumar Chinthaginjala and Dhakshayani J and Xin Zhao and Safia Obaidur Rab and Sivarama Prasad Tera, "Improving CNN predictive accuracy in COVID-19 health analytics," *Scientific reports*, vol. 15, no. 29864, 2025, doi: 10.1038/s41598-025-15218-y.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[19] 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," *Scientific reports*, vol. 10, no. 19549, 2020, doi: 10.1038/s41598-020-76550-z.

[20] COVID-Next, "COVID-Next," https://github.com/velebit-ai/COVID-Next-Pytorch, May Accessed on December 2021.

[21] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," 2017, doi: 10.1109/CVPR.2017.634.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, San Diego, May 2015, doi: 10.48550/arXiv.1409.1556.

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

[24] "Covid chest x-ray dataset," Accessed on 11th October 2025. [Online]. Available: https://github.com/ieee8023/

covid-chestxray-dataset

[25] "Rsna pneumonia detection challenge," Accessed on 11th October 2025. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

[26] S. Chatterjee, F. A. Nizamani, A. Nürnberger, and O. Speck, "Classification of brain tumours in MR images using deep spatiospatial models," *Sci Rep 12, 1505*, 2022, doi: 10.1038/s41598-022-05572-6.

[27] "Ixi dataset," Accessed on 21th March 2022. [Online]. Available: https://brain-development.org/ixi-dataset

[28] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," 2019, doi: 10.48550/arXiv.1811.02629.

[29] Du Tran and Heng Wang and Lorenzo Torresani and Jamie Ray and Yann LeCun and Manohar Paluri, "A closer look at spatiotemporal convolutions for action recognition," 2018, doi: 10.1109/CVPR.2018.00675.

[30] FG-AI4H, "Topic description document for the topic group on ai for radiology," Accessed on 15th October 2025. [Online]. Available: https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AI4H-2023-32-PDF-E.pdf

VIII. Biography

Edson M. Hung received the Eng., M.Sc., and D.Sc. degrees from the Dept. of Electrical Engineering at Universidade de Brasília (UnB), Brazil, in 2004, 2007, and 2012, respectively. He is currently an Associate Professor at Universidade de Brasília. His work at UnB has been or is being funded by agencies such as CNPq, CAPES, FAP-DF and companies such as Hewlett-Packard, Eletronorte, Leucotron, Toledo do Brasil, Google, ID-Scan, Saab, Petrobrás, and Samsung. He has experience in project development focusing on Signal Processing, Machine Learning, and Electronics. Since 2018, he has been member of the Joint Picture Experts Group (JPEG) and the Motion Picture Experts Group (MPEG) from the International Organization for Standardization (ISO) and the International Telecommunication Union (ITU). He has also been a member of the Focus Group on Artificial Intelligence for Health (AI4H) of the World Health Organization (WHO) since February 2020.

Renam C. da Silva (Member, IEEE) received the degree in mechatronics engineering from the Universidade do Estado do Amazonas, Brazil, in 2010, and the M.Sc. and D.Sc. degrees in electrical engineering from Universidade Federal do Rio de Janeiro, in 2013 and 2018, respectively. He is currently with the Electrical Engineering Department, Universidade de Brasília. His research interests include multimedia signal processing, image and video compression, and artificial intelligence and deep learning.

Andrey O. O. dos Reis Andrey Otacilio Oliveira dos Reis received the Eng. and M.Sc. degrees from the Dept. of Electrical Engineering at Universidade de Brasília (UnB), Brazil, in 2022 and 2024, respectively. He is currently a Doctoral Student at Universidade de Brasília. He has experience in the field of digital signal processing and machine learning.

Darlington Akogo is a global leader in Artificial Intelligence. He is the Founder and Chief Executive Officer at minoHealth AI Labs, an AI Healthtech company; karaAgro AI, an AI-powered Plant and Pest Disease Detection and Precision Agriculture platform; Runmila AI Institute, an AI and Data Science training institute; and Gudra AI Studio, an organization broadly exploring AI and Exponential Technologies applied various to domains including Transportation, Sanitation and Energy. Akogo has been named one of Forbes 30 under 30, and one of the Global Top 100 Most Influential People of African Descent (MIPAD) in Healthcare. He is the Chair of the Topic Group on AI for Radiology under the United Nations International Telecommunications Union (ITU) and World Health Organization (WHO) Focus Group on Artificial Intelligence for Health (FG-AI4H). At UN FG-AI4H, Akogo leads the development of global regulations and standards for AI in radiology. He also works with African Union, where he serves as the Chair for the working group on Artificial Intelligence Economy, working towards the development of a Continental African Union (AU) Artificial Intelligence (AI) Continental Strategy that includes legislative, regulatory, ethical, policy, and infrastructural frameworks in consultation with stakeholders such as AU Member States' governments, private sector, academia, innovators, and consumers.