Deep Learning applied to Intracranial Hemorrhage Detection

Luis Cortes-Ferre [1,†] , **Miguel Angel Gutiérrez-Naranjo** [1,†] , **Juan José Egea-Guerrero** [2,3,†] , **Soledad Pérez-Sánchez** [4,5,†] **and Marcin Balcerzyk** [6,7,†]*

1. Department of Computer Sciences and Artificial Intelligence, University of Seville, Sevilla, Spain; luiscortesferre@gmail.com; magutier@us.es
2. Hospital Universitario Virgen del Rocio, Avda. Manuel Siurot, 41013 Sevilla, Spain; juanj.egea.sspa@juntadeandalucia.es
3. Instituto de Biomedicina de Sevilla (Universidad de Sevilla – CSIC – Junta de Andalucía), 41013 Sevilla, Spain; juanj.egea.sspa@juntadeandalucia.es
4. Stroke Unit, Neurology Department, Hospital Universitario Virgen Macarena, Sevilla, Spain;
5. Neurovascular Research Laboratory, Instituto de Biomedicina de Sevilla-IBiS, Seville, Spain; soledad.perez.sanchez@gmail.com
6. Department of Medical Physiology and Biophysics, University of Seville, Sevilla, Spain; mbalcerzyk@us.es
7. Centro Nacional Aceleradores (Universidad de Sevilla – CSIC – Junta de Andalucía), 41092 Sevilla, Spain; mbalcerzyk@us.es
* Correspondence: mbalcerzyk@us.es; (M.B.)
† These authors contributed equally to this work.

Abstract: Intracranial hemorrhage is a serious medical problem that requires rapid and often intensive medical care. Identifying the location and type of any hemorrhage present is a critical step in the treatment of the patient. Diagnosis requires an urgent procedure and detection of hemorrhage is a difficult and time-consuming process for human experts. In this paper, we propose methods based on EfficientDet's deep-learning technology that can be applied to the diagnosis of hemorrhages at a patient level and thus become a decision-support system. Our proposal is two-fold. On the one hand, the proposed technique classifies slices of computed tomography scans for the presence hemorrhage or its lack, evaluates if the patient is positive in terms of hemorrhage achieving 92.7% accuracy and 0.978 ROC-AUC. On the other hand, our methodology provides visual explanations of the chosen classification using the Grad-CAM methodology.

Keywords: Image Detection; Intracranial Hemorrhage; Deep Learning; Decision Support System.

1. Introduction

Spontaneous intracranial hemorrhage (ICH) occurs when a diseased blood vessel within the brain bursts, allowing blood to leak inside the brain. Several factors, like disruption of blood-brain barrier and leakage of fluids and proteins, inflammatory cascade or sudden increase in intracranial pressure after bleeding develop into a brain injury in the areas surrounding the blood spill.

Nearly 66 % of all deaths caused by neurological diseases worldwide are related to hemorrhagic stroke [1]. According to [2], the overall incidence of spontaneous ICH worldwide is 24.6 per 100,000 people each year and approximately half of ICH related mortality occurs within the first 24 hours [3].

The severity and outcome of an ICH depend on its cause, location in the brain, size of the bleed, the amount of time that passes between the bleed and treatment, clinical presentation, etc. ICH usually occurs in particular parts of the brain, including the basal ganglia, cerebellum, brain stem, or cortex.

Damage can be severe and result in physical, mental, and task-based disability. Various types of ICH strike people of all ages. Although ICH is usually associated with older adults, it can also occur in the younger population. If not treated correctly and immediately, an ICH can cause disability or death. As a heterogeneous disease, image characteristics help to determine the cause, the prognosis and how to treat the disease.

This paper presents a model based on deep learning techniques that can be useful for classifying ICH in patients from computer-tomography (CT) image slices. We begin

with a short introduction explaining what intracranial bleeding is and present the latest state-of-the-art (SOTA) artificial intelligence (AI) methods applicable to ICH. Then, we present the network architecture that we use in our experiments. On the one hand, it is worth mentioning that our approximation follows the current research guidelines on Green AI [4]. Roughly speaking, Green AI tries to reduce the carbon footprint of the current large computational processes by searching a balance between the consumed resources and the efficiency of the method. As it will be detailed below, our methodology obtains competitive results by using a training dataset and a deep learning network smaller than some competitors found in the literature. On the other hand, our methodology also includes visual explanations to understand model decision-making.

One of the main disadvantages of using deep learning techniques in real-world problems is that it lacks a clear explanation. Although the accuracy achieved in many cases is high, experts find it difficult to extract knowledge from deep learning models and their results that justify decisions. To address this lack of explanation, various approaches have been published in recent years. One of these approaches is the field of visual explanations. Classification models of computational images incorporating such an approach usually provide an accurate classification, and they can also show the region of the image (the set of pixels) which has been important in the classification process.

One of the methods of visual explanations in deep learning is the Gradient-weighted Class Activation Mapping (Grad-CAM) [5] method, where the authors aim to provide visual explanations via gradient-based localization. This approach takes as input an image and uses the gradients of any output neuron in a classification network flowing into the final convolutional layer to produce a heatmap that highlights the relevant regions in the image to make the corresponding prediction. This method has also been used in [6]. Grad-CAM is a form of post-hoc attention, i.e. it is a method applied to a neural network already trained. The basic idea behind this method is to take advantage of the spatial information that is preserved through the neural network to understand which parts of an input image were important for the classification decision. Grad-CAM provides a heat map that focuses on the pixels that were the most influential in the decision.

Beyond the use of Grad-CAM for visual explanations, this paper considers one of the most successful architectures for solving complex technical problems of neuronal networks (as vanishing/exploding gradients): the so-called residual neural networks (or ResNet, for short) [7]. They are inspired by pyramidal cells in the cerebral cortex. ResNet tries to alleviate these problems by introducing the so-called residual blocks. These blocks introduce a direct connection, which skips some layers in between. This connection is called the *skip connection* and is the core of residual blocks. The use of ResNet is integrated in our model with EfficientDet [8], which combines weighted bidirectional feature pyramid network (BiFPN) with a compound scaling method (see below). ResNet and EfficientDet have been well described in literature, but the novelty of this paper is the efficient combination of these technologies. From a technical point of view, to the best of our knowledge, it is the first time that ResNet, EfficientDet and Grad-CAM have been used together. Besides, the proposed model is not a simple juxtaposition of techniques as described in the literature, but an integrated model.

The CT scan slices provide us with 2D matrices with the scan in the Hounsfield unit (HU). The HU scale is a linear transformation of the original linear attenuation coefficient measurement into one in which the radiodensity of distilled water at standard pressure and temperature (STP) is defined as zero HU, while the radiodensity of air at STP is defined as -1000 HU. In a voxel with average linear attenuation coefficient , the corresponding HU value is therefore given by:

$$HU = 1000 \times \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \qquad (1)$$

where μ_{water} and μ_{air} are respectively the measured linear attenuation coefficients of water and air. Thus, a change of one HU represents a change of 0.1% of the attenuation coefficient of water, since the attenuation coefficient of air is nearly zero.

To understand brain bleeds, it is important to have a basic understanding of the different types based on the location where the bleed occurs (for a detailed introduction to ICH, see, e.g., [9] or [10]). After a stroke, the cause (blood or blood clot) must be determined to initiate appropriate treatment. Short-term medical treatment can help reduce brain damage and improve the likelihood of recovery. Survival and recovery from intracranial haemorrhage are related to the site, mass effect, and intracranial pressure of the underlying hematoma, and by subsequent cerebral edema from perihematomal neurotoxicity or inflammation and complications from long-term neurological dysfunction.

Surgery may be necessary since bleeding (hemorrhage) may require immediate decompression of the brain to release pooled blood and relieve pressure. A craniectomy incision (partial removal of the skull to allow the swelling brain to expand), or a craniotomy (opening of the skull cavity) are practiced. Depending on the type, location, and extent of the brain bleed, many patients do not survive the initial bleeding event. Mortality rate is as high as 50% in the first 30 days [11].

Research indicates several therapeutic avenues of benefit in intracerebral hemorrhage modelled around an active management plan, aggressive supportive care, early blood pressure control, and targeted surgery as a life-preserving strategy [12], [13]. The time between symptoms, arrival in hospital, and diagnosis of bleeding is crucial. The earlier the brain hemorrhage is confirmed and evaluated, the sooner a treatment plan can be made.

The paper is organized as follows. In Section 2, some current approaches to the problem of the detection of ICH by deep learning and Grad-CAM techniques are recalled. Section 3, Materials and Methods, is devoted to explaining our proposal. In Section 4, the results obtained are presented. The paper finishes with the discussion of the results (Section 5) and some conclusions (Section 6). Some appendices have also been added with technical details.

2. Deep Learning for the Detection of Intracranial Hemorrhage

Deep Learning [14] is a series of models, technologies, and architectures based on Artificial Neural Networks which have undergone a revolution in AI in the last years. Its doubtless success in real world problems reaches areas such as face recognition [15], music composition [16] o multilingual translation [17]. In the last decade, thousands of researchers have proposed new models that overtake the achievements of previous architectures, so it is impossible to give a general framework that covers all possible approaches in deep learning. Nevertheless, in general, deep learning models can be described as mathematical functions that generate output from a given input. For example, some models provide the names of people when faces are given, or sentences are translated into different languages from a given one.

From a technical point of view, such mathematical function is obtained by composition of typically thousands, or millions of simple mathematical functions called neurons due to their original biological inspiration. These sets of neurons are the basis of the neural network and the links among the neurons determine the architecture. The basic architecture corresponds to the so-called multilayer perceptron (MLP, see, e.g., Figure 1 (top)), where the neurons are arranged in layers. Each neuron is connected to all neurons in the following layer and there is no connection between neurons of the same layer. As we will see below, from this basic architecture more and more efficient ones can be built. The main feature of such a neural network is that its parameters are self-adjusted by trying to minimize a loss function on a training dataset in a process known as learning. After the training process, the neural network with the parameters obtained can be used to solve problems that were unapproachable some years ago.

Deep neural networks extract many interpretive patterns and characteristics from data and learn very complex and significant representations. The extraction and discovery
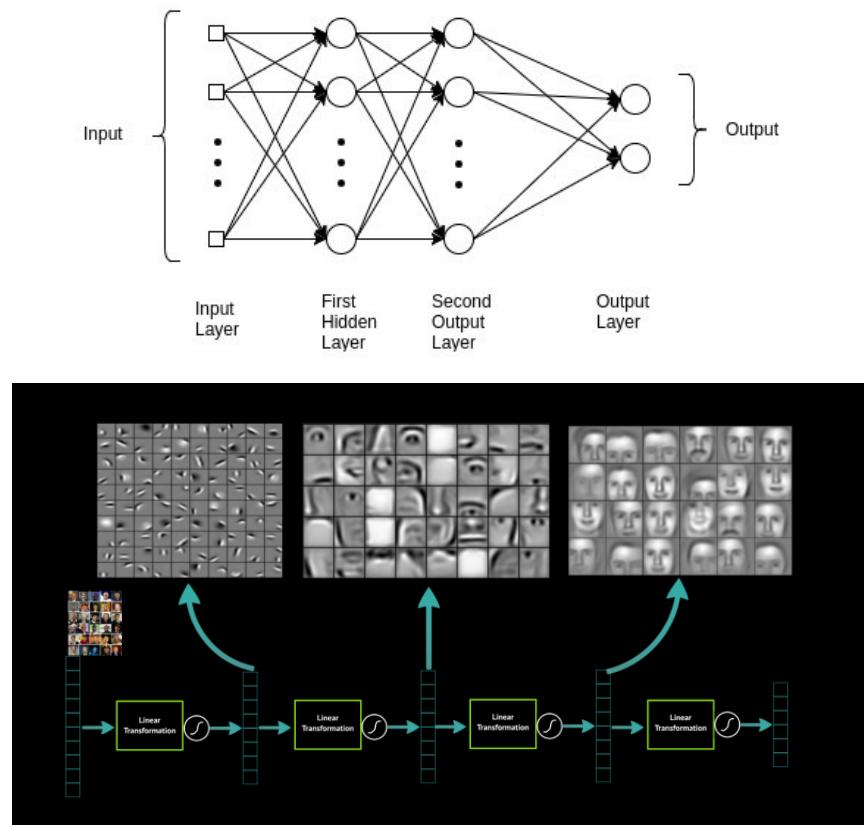
Figure 1. *(Top) Example of an MLP. Source: [18]; (Bottom) Extraction, learning, and integration of features. Source [19].*

of these characteristics or patterns is attributed to the depth of the neuron network, as they are more likely to be found in the later layers of the network.

The intuition behind this is that these layers gradually learn more complex features. For example, in the case of image recognition, the first layer may learn to detect edges, the second layer may learn to identify textures, and similarly the third layer can learn to detect objects, etc. (see Figure 1 (bottom)). As the nature of problems input to the neural network becomes increasingly difficult, researchers begin to use more and more advanced models to achieve better results, as we will see below.

One of the most interesting areas of deep learning research is digital image processing, especially medical image processing. Among many other applications, we can cite disease classification [20], ROI segmentation [21] or medical object detection [22]. In this paper, we apply Deep Learning techniques to the study of ICH detection.

3. Materials and Methods

3.1. Deep learning architectures

The first basic architecture used in our model is ResNet. The second is EfficientDet [8]. The main motivation for this architecture is to optimize the resources of the deep learning model for the detection of digital image objects.

Normally, the big success of deep learning models is obtained only by expensive large computers that can be used by large companies. Currently, many efforts are trying to adapt these models to simpler (and cheap) hardware without losing accuracy. EfficientDet is a clear example of this research line. The main contribution of EfficientDet is two-fold. On the one hand, it uses a novel computing unit as the base of the model, the so-called weighted bidirectional feature pyramid network (BiFPN) and, on the other hand, the authors propose a compound scaling method for scaling up several features of the model (resolution, width, and depth), which leads to a new family of object detectors called EfficientDet family, from
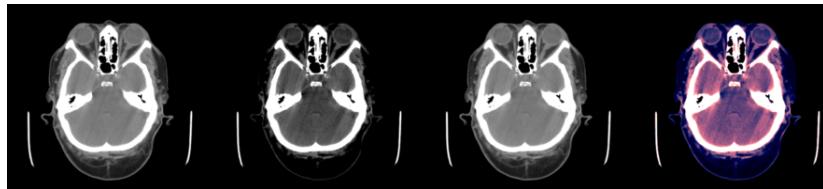
Figure 2. Example of slice windowing as preprocessing of a HU matrix to obtain three different matrices (window used between brackets). From left to right: (1) slice with [0,80], (2) slice with [-20, 180], (3) slice with [-150, 230], (4) the three windowed slices stacked (each channel is associated to one color RGB for a visualization purpose.

D0 (the smallest one) and D7 (the largest one). According to the Green AI guidelines, we have optimized our system in order to have competitive results with a minimum amount of resources, and hence, only the D0 model, the smallest one, has been considered.

In our study, we propose a neural network architecture that combines the EfficientDet [8] and ResNet [7] architectures. Technical details of ResNet and EfficientDet architectures can be found in Appendix A.1.

3.2. Dataset

Kaggle[1] is an online community of data scientists and machine learning practitioners that allows users to find and publish datasets, explore, and build models in a web-based data science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

The RSNA Intracranial Hemorrhage Competition [23] was a competition hosted by Kaggle at the end of 2019. This competition provides a high number of annotated data, indicating if there is hemorrhage in the slice and if so, indicating one of five subtypes. The competition was divided into two stages. In this work, we only use the data from the second stage (more than 400 GB).

The dataset contains 752,799 scan slices in Digital Imaging and Communications in Medicine (DICOM) format, from 18,938 patients.

The set of patients was divided into three randomly divided patients' groups: train (90%, 17044 patients), validation (5%, 947 patients) and testing (5%, 947 patients) groups to avoid bias since there are various slices from the same patient. However, it is important to note that there is a remarkable class imbalance since the groups contain the following number of slices:

- Train: 97525 with ICH, 580934 without ICH.
- Validation: 5401 with ICH, 31174 without ICH.
- Test: 5007 with ICH, 32758 without ICH.

To mitigate this class imbalance, we randomly select the same number of non-ICH slices as the ICH-sliced slices. It turns into 195050 slices from 17044 patients for training, 10802 slices from 947 patients for validation, and 10014 slices from 947 patients for testing.

In addition, since our neural network takes three channels as input, the scans with the HU matrices must be pre-processed to obtain three matrices of size 512×512 pixels, one for each channel. To obtain such three matrices, three different windows are applied to every slice. Applying a window with X as lower bound and Y as upper bound means that all HU values lower than X are converted to X, all HU values greater than Y are converted to Y and the rest remains the same (see Figure 2).

We build a test dataset with all the slices of the test patients, that is, 5007 slices with ICH and 32758 without ICH (37765 slices in total). However, in this case we evaluate the model performance annotating every patient with ICH if any slice has ICH and noICH otherwise. Following this annotation criteria, 371 of 947 patients have ICH.

---

[1] https://www.kaggle.com/

In the Kaggle dataset, the slice thickness was uniform for one patient, but there are different thicknesses between patients.

We added a test group of 55 images of clinical cases of control (eight patients), intracranial hemorrhage (ten patients), and subarachnoid hemorrhage (37 patients) provided by HUVR and HUVM hospitals, where two authors are working. The controls were not healthy volunteers, but patients scanned by CT at admission for who any other type of hemorrhage was excluded. The research project in HUVR was overseen and approved by the Ethics Committee of the Hospital Universitario Virgen del Rocio in Seville, Spain (Cod. CEI2012PI/228). Informed consent was obtained from all patients participating in the study or from their close relatives. The images of the patients at the Hospital Virgen Macarena were obtained after they signed the informed consent form for CT in which they accepted the potential use of the images for the purposes of promoting knowledge and research in relation to the disease from which they suffer. The individual slices were 512×512 pixels.

The images were evaluated for diagnosis per patient by experienced radiologists, neurologists and neurocritical doctors. They were not evaluated on slice-by-slice basis, which is not done in clinical practice. The mean age of patients was 54 years, range [21...84]. There were 25 men and 30 women (55 %).

3.3. Grad-CAM

Grad-CAM uses the feature maps produced by a convolutional layer (one of the types of layer used in deep learning to analyze data structured in grids, as computational images). The authors argue that the last convolutional layers have the best compromise between high-level semantics and detailed spatial information. The output of Grad-CAM is a class-discriminative localization map, i.e., a heatmap where the hot part corresponds to the part of the image that has been relevant for the model to perform the classification. Our model provides two different kinds of information:

- Firstly, our model provides as output whether the patient has ICH or not. Technically, the output is the probability of having ICH. The model outputs an affirmative answer if such a probability is greater than a fixed threshold
- Also, our model provides a color map on the input image where the red area corresponds to the pixels in the image which have been determinant in the decision.

We applied GRAD-CAM and obtained a heat map with focus in the ICH area. For the Grad-CAM implementation, we used the *pytorch-grad-cam* library [24]. In addition, in our Grad-CAM experiments, we used enhancement smoothing, and according to the library documentation, it applies a combination of horizontal flips and image multiplications by $[0.9, 1.0, 1.1]$, having the effect of better centering the CAM around the objects.

Figure 3 (a) illustrates the use of our model. It shows an image where the model predicts bleeding (ICH) with a probability of 0.9897, and in addition Grad-CAM was able to determine the bleeding area to make that decision. Appendix A.2 provides a more technical description of this technique.

Our code is available online[2]. More details of the framework can be found in Appendix A.3. Let us recall that the CAM techniques for obtaining visual explanations of Deep Learning decision have been previously used by other researchers (see, e.g., [25] or [26]), but, to the best of our knowledge, it is the first time when ResNet, EfficientDet and Grad-CAM are used together. The combination of such technologies is a contribution of this paper which can shed light to future research in this area.

3.4. Metrics

The metrics used to evaluate the experiments are accuracy, sensitivity, specificity, positive predictive value and negative predictive value, considering that the model predicts yes (slice with ICH) for a certain image slice if the probability is greater than 0.8 to reduce

---

[2] https://github.com/Keredu/Intracranial-Hemorrhage-Detection

the false positive predictions. Most binary models assume a cut-off value of 0.5, lower than ours.

4. Results

4.1. Kaggle test dataset

Table 1 shows the results of the evaluation of our model, called EffClass, with RSNA Kaggle competition test data and the application of different thresholds for the prediction of our model (i.e., the model predicts ICH if the probability returned by the model is higher than the threshold). The chosen thresholds of probability are 0.5, 0.7, and 0.9. The highest values are marked in **bold**.

Kaggle dataset is a set of CT scan slices, however the DICOM files of slices retain the information of the patient, therefore we could have use these data to train our models at patient level, but we had to develop our own new criteria for evaluation if the *patient* is to be classified as having hemorrhage or not. This is outside of the available data of Kaggle competition and we claim that if the patients CT scan has for example only one slice classified as having hemorrhage, that is not enough to classify and most importantly to diagnose him positively as a whole patient as having brain hemorrhage. Therefore, we evaluate the patients considering the following additional criteria: *The patient is diagnosed with hemorrhage if the model predicts, at least, n slices of the image with ICH being n a percentage (5% or 10%) of the number of slices since the number of slices per patient varies.* We do not check for the continuity of the slices position, just the percentage.

For example, given two patients with 30 and 40 as the number of slices, respectively, and a threshold of 10%, the first patient would be classified as ICH patient with ICH if the model predicts at least three (3) ICH slices, but the second patient would be classified as ICH patient with ICH if the model predicts at least four (4) slices with ICH. The results of the experiments are shown in Table 2.

4.2. Test results obtained with clinical data

Despite the results that we obtained in the previous section being very promising, we needed to test our models with external data to validate that our models could be used not only with Kaggle data but also with external data.

For this purpose, we evaluated the images of 55 patients from the Hospital Universitario Virgen Macarena and Hospital Universitario Virgen del Rocío patients. Both hospitals are in Seville, Spain.

In Figure 3 (a) a slice with ICH is shown. It is quite difficult to see the ICH, but the EfficientClassification neural network predicts ICH with a probability of 0.9897. Figure 3 (b) and Figure 3 (c) show the next consecutive two slices where ICH is clearly seen and the neural network predicts ICH with a probability of 0.9972 and 0.9997, respectively. Figure 3 (d) shows a slice with ICH, where the hemorrhage is not detected according to our threshold (0.8), and the probability was 0.7805. Figure 3 (e) shows a positive slice where the probability was just above the threshold, 0.8101. On the other hand, in Figure 3 (f) a slice without ICH is shown.

The results of the classification of additional test 55 patients are provided in the last row or Table 2. All patients with ICH (47) were identified correctly. All control patients

Table 1: Results per model and threshold (**TH**). Higher metrics values remarked. **EffClass**: EfficientClassification, **ACC:** Accuracy, **TPR:** Sensitivity (Recall), **TNR:** Specificity, **PPV:** Positive Predictive Value (Precision), **NPV:** Negative Predictive Value, **PR:** Precision-Recall AUC, **ROC:** ROC-AUC.

Model (TH)	ACC	TPR	TNR	PPV	NPV	PR	ROC
EffClass (0.5)	**0.927**	**0.914**	0.94	0.938	**0.916**	**0.979**	**0.978**
EffClass (0.7)	0.918	0.872	0.965	0.961	0.883	**0.979**	**0.978**
EffClass (0.9)	0.881	0.774	**0.987**	**0.984**	0.814	**0.979**	**0.978**

Table 2: Results per model, with at least n% of the slices with ICH. Best metric per model and threshold (**TH**) are marked in bold. **EffClass**: EfficientClassification, **ACC:** Accuracy, **TPR:** Sensitivity (Recall), **TNR:** Specificity, **PPV:** Positive Predictive Value (Precision), **NPV:** Negative Predictive Value, **N:** Number of patients, **ICH (%):** Percentage of patients with ICH. (*): No confidence interval provided. (**) Confidence interval impossible to calculate.

Model (TH)	ACC	TPR	TNR	PPV	NPV	N	ICH (%)
Aidoc (FDA/501k) [27]	92.9% (*)	93.6% (86.6-97.6)	92.3% (85.4-96.6)	91.7% (84.9-95.6)	94.1% (88.4-97.2)	198	47.5%
Aidoc [27]	97.2% (*)	92.3% (88.9-94.8)	97.7% (97.2-98.2)	81.3% (77.6-84.5)	99.2% (98.8-99.4)	3605	9.7%
EffClass (5%)	95.5% (93.9-96.7)	94.9% (92.1-96.9)	95.8% (93.9-97.3)	93.6% (90.8-95.6)	96.7% (94.9-97.8)	947	39.2%
EffClass (10%)	93.6% (91.8-95.0)	88.1% (84.4-91.3)	97.0% (95.3-98.3)	95.1% (92.3-96.9)	92.7% (90.6-94.4)	947	39.2%
EffClass test (10%) 55 patients	100.0% (93.5-100.0)	100.0% (92.5-100.0)	100.0% (63.1-100.0)	100.0% (**)	100.0% (**)	55	85.5%

were identified as negative (8) There were no false positives diagnosed in control patients. NPV and PPV are 100% because there was no false negatives nor false positives.

5. Discussion

Recently, some studies have been published on the application of AI techniques to Intracranial Hemorrhage Detection, and many researchers have started to pay attention to it. Among many other studies, we can cite [28], where the authors use a fully convolutional neural network for classification and segmentation with examination of ICH with computed tomographies; [29] where the InceptionV3 and DenseNet Deep Learning models for dealing with CT; or [30], where the authors combine convolutional neural networks with other machine learning techniques to deal with ICH detection. In [31], the authors use a Dense U-net architecture for the detection of ICH. In [32] A novel deep learning technique based on the Monte Carlo algorithm [33] was applied.

The paper by Voter *et al.* [27], published in April 2021, deserves special attention. It is very important to note that we cannot make a fair comparison, since the datasets used in our experiments and the Aidoc experiments [27] are different. In the reference [27] dataset, only the 9.7%, 349 of 3065 patients, were positive (i.e., patients with ICH) while in our test dataset, the 39.2%, 371 of 947 patients were positive. In addition, the results are highly influenced by the types of ICH that comprise the dataset and their proportion.

The study presented in [27] shows high accuracy (ACC) and negative predictive value (NPV). The highest accuracy value obtained from our experiments is 95.5%, with the EfficientClass model with threshold = 5% (overlapping confidence intervals). The study [27] shows the same true negative rate (TNR) as EfficientClass with threshold = 5%. With respect to the true positive rate (TPR), the EfficientClass model with threshold = 5% has a higher value (0.949). Finally, with respect to the predictive positive value (PPV), the highest value is 0.951, and it was obtained with the EfficientClass model with threshold = 10%.

A standard study design is needed to enable a rigorous and reproducible site-to-site comparison of new deep learning technologies. Our goal is to classify patients by their CT images if they have ICH or not. This means that we estimate a binary probability distribution in which each image obtains the probability P that the patient has ICH and the probability $1 - P$ that the patient does not have ICH. To this end, the original Kaggle
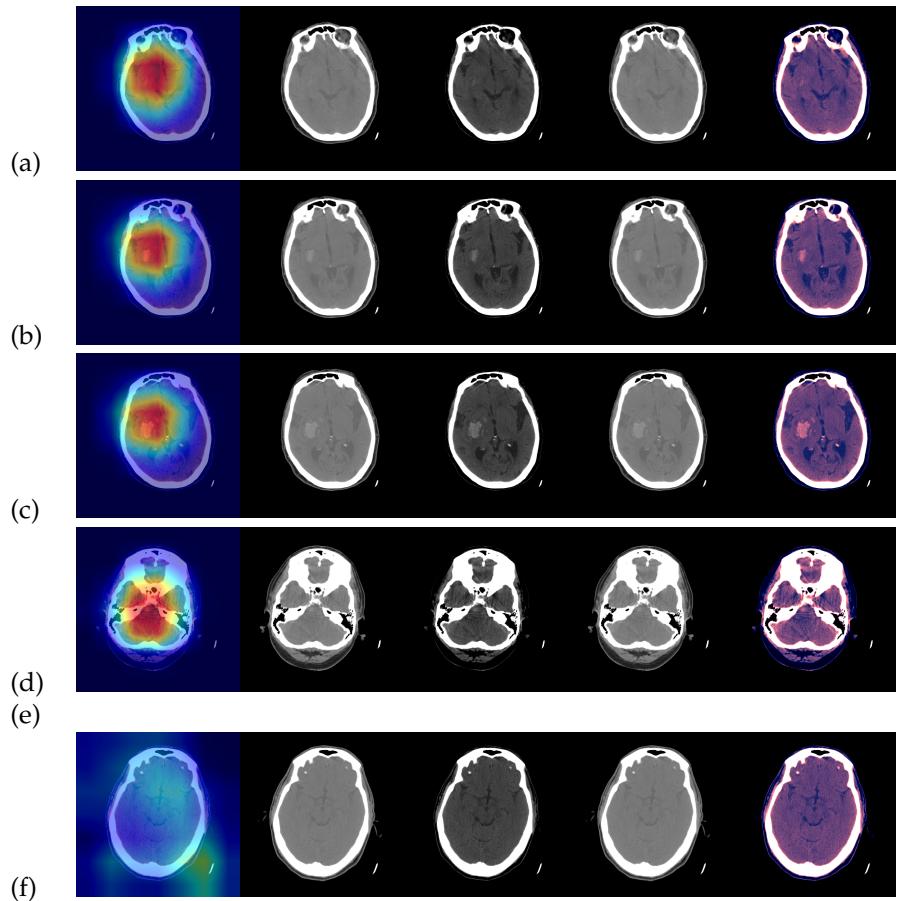
Figure 3. The model EfficientClassification predictions were all correct. In addition Grad-CAM was able to focus on the bleeding area to make that decision. In each image, from left to right: Grad-CAM, [0,80] window, [-20,180] window, [-150,230] window, the three windows stacked. From top to bottom, the probabilities of ICH in the slice are 0.9897, 0.9972, 0.9997, 0.7805, 0.8101 and 0.0097.

dataset was pre-processed to train our model. In addition to many other technical details, preprocessing changes the classification label of all input data from the original label to a binary label marking if the image shows if the patient has ICH or not. Bearing in mind that the target of the models presented to the Kaggle competition is to associate to each CT image a label with the type of ICH, not to classify if the patient has ICH or not, a comparison in the accuracy of our model and the ones presented to the competition is not possible.

A typical detected minimum size (width) by Aidoc software is shown in Fig. 2 A of [27] and is about 2 mm. In clinical practice there are rarely less than 25 slices per brain scan. We did not check or train the model for the size of the hemorrhage. Let us note that the model was trained on slices with yes/no for all the slices, and we did not train the model for the location of the hemorrhage. The Grad-CAM map should not be mistaken for the location of the hemorrhage. It can be interpreted as the usual location of the hemorrhages and the pixels considered in the evaluation of the presence of hemorrhages for this slice, not a precise location.

In paper [27], the authors have the objective of determining the diagnostic precision of an AI decision support system (DSS), developed by Aidoc [34], in diagnosing intracranial hemorrhage (ICH) in non-contrast head CT and to assess the potential generalizability of an AI DSS. This retrospective study included 3605 consecutive, emergent, non-contrast adult head CT scans.

Each scan was assessed for the ICH by a certified neuroradiologist and Aidoc. The authors determined the diagnostic accuracy of the AI model and performed a failure mode analysis with quantitative CT radiomic image characterization. The results were as follows: Of the 3605 scans, 349 cases of ICH (9.7% of studies) were identified. The neuroradiologist and Aidoc interpretations were consistent in 96.9% of the cases and overall sensitivity, specificity, positive predictive value, and negative predictive value were 92.3%, 97.7%, 81.3% and 99.2%, respectively.

Aidoc was wrong in 101 of 3605 studies, which corresponds to an accuracy of 97.2%. This study revealed decreased diagnostic accuracy of an AI DSS at their institution. Despite its extensive evaluation, they were unable to identify the source of this discrepancy, raising concerns about the generalizability of these tools.

In [35], the authors also used Kaggle dataset and compared their results with clinical data available to them keeping the classification at slice level without applying of their model at patient level. These results further highlight the need for a standardized study design to allow a rigorous and reproducible site-to-site comparison of emerging deep learning technologies.

The irruption of deep learning as a new set of AI techniques represents a major challenge in medical applications. This type of technology after CE (*Conformité Européenne*, in French) security marking, Food and Drug Administration (FDA) or respective national agency clearance can be very useful for those centers, especially small community hospitals, where there are no radiologists on call to interpret these images and depend on non specialized physicians.

In many fields, radiologist expertise is the key to obtaining a correct diagnosis. In this case, the use of the latest technology for computer image classification in medical images is a major contribution to human experts.

In fact, specific diagnostic support software is already used in ischemic stroke in a wide range of hospitals. In this way, this paper presents EffClass, a deep learning model that combines two of the most successful models for image classification and we show that our model is competitive with some of the current most widely used commercial, CE certified and FDA approved methods such as Aidoc [34] or e-ASPECTS [36].

The efficiency of manual SAH diagnosis can be 98.9% [37], but at the expense of time. Radiologist usually needs from 15 to 30 minutes to diagnose a patient and inform about it in writing. Diagnosis of ICH has been a subject of several papers and is a field of intensive research using traditional techniques [9] and deep learning [28–31]. The solution [38] has AUC 0.846 (CI 0.837–0.856). Another is BRAINSCAN.AI [39] which has AUC/ROC of 0.976 for hemorrhages. It is a CE-certified solution and is now installed in more than 40 hospitals in Poland [40].

However, the use of neural network methods in medicine as a support in decision processes has an important drawback. They are usually considered as black boxes, where the high accuracy is not considered due to the lack of explanations. Human experts want to know not only the decision of the AI system but also the motivation of such a decision. In such a way, our model not only provides high accuracy on the decision, but it also provides visual explanation with an intuitive color map on the original image. In such a way, the Grad-CAM method has been used to highlight the area of the input image that has been relevant in the decision. This mixture of high accuracy together with the visual explanation makes the proposed model one of the most competitive among the current ones.

The limitation, as of today, is that the method is for research use only and the command line interface with PyTorch is comfortable only to IT specialists. It requires the extraction of the DICOM hospital image from the hospital system to a desktop computer. Our method has some higher metrics (ACC, TPR and PPV) than Aidoc FDA submission, bearing in mind that the amplitude of the confidence intervals of Aidoc FDA submission is wide.

Future research lines can also be considered. From the applications point of view, the deep learning technologies proposed in this paper can be also applied to obtain classification and visual explanation of other medical decisions based on medical images.

From a technical point of view, deep learning is continuously evolving, and other recent technologies on classification and visual explanation deserve to be explored, including 3D approaches.

6. Conclusions

We have developed a model that can detect bleeding in CT scans of a patient in DICOM 2D image files. This model has a higher AUC than the Aidoc one available commercially in a version submitted for approval by the FDA. The model has several adjustable parameters as threshold to classify as a slice with hemorrhage and percentage of the slices of the image of the patient to classify him/her as diagnosed with hemorrhage. The tool is Research Use Only for the moment.

Hemorrhages were divided into five classes in Kaggle competition. In clinical practice, the division is done typically between intracranial and subarachnoid. Subarachnoid ones are sometimes difficult to diagnose per slice, what can be seen in Figure 3 (d, e) for SAH. The concept of "time is brain", used for ischemic stroke, could be also applied in this devastating disease, so automatic diagnosis which takes seconds right after CT scan image reconstruction can prioritize the images for evaluation by a radiologist or help in complicated or debatable cases.

Despite the widespread use of deep learning in medical image classification, this paper makes two folded contributions to this field. On the one hand, the accuracy achieved is competitive with the AI techniques most commonly used in ICH classification. This represents an advance over many papers found in the literature. On the other hand, many deep learning tools are black boxes providing classification without explanation. In this paper, we propose to use a visual explanation technology to provide information on the key areas of CT. This additional information complements other tools of deep learning and makes our proposal a useful tool for clinicians.

Author Contributions: All authors contribute equally to this work and have read and agreed to the published version of the manuscript.

Funding: MAGN was funded by Ministerio de Ciencia e Innovación of Spain, grant number PID2019-107339GB-I00. For MB, the project leading to these results have received funding from "la Caixa" Foundation (ID 100010434), under the agreement CI19-0068 (Barcelona, Spain). The HUVR patients for this study were recruited thanks to a grant from Consejería de Igualdad, Salud y Políticas Sociales de Andalucía, Spain (PI-0136-2012).

Data Availability Statement: Data from Kaggle RSNA Intracranial Hemorrhage Competition are publicly available and can be obtained from [23].

Acknowledgments:

Conflicts of Interest: Juan José Egea-Guerrero and Marcin Balcerzyk are inventors of the patent-pending application WO2019025656.

Appendix A

This appendix includes some technical details of our approach.

Appendix A.1

Firstly, some technical details of the models ResNet and EfficientDet are provided. The main idea behind ResNet is to fit a residual mapping instead of hoping that each few stacked layers directly fit a desired underlying mapping. Formally, denoting the desired underlying mapping as $H(x)$, ResNet allowed the stacked non-linear layers to fit another mapping of $F(x) := H(x) - x$. Then, the original mapping is recast as $F(x) + x$. The authors [7] hypothesize that it is easier to optimize the residual mapping than to optimize the original one. For example, if an identity mapping were optimal, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers A scheme of the so-called *residual blocks* can be found in Figure A1.
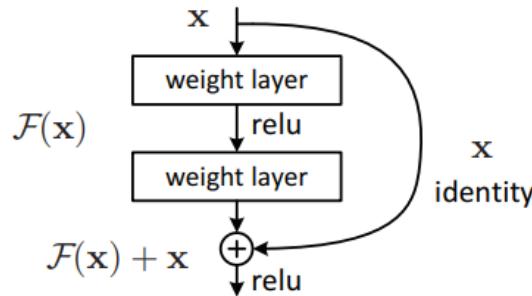
Figure A1. Residual learning: a building block. Source: [7].

If we have two identical architectures but we extend one of them with some layers, we expect to get the same performance at least since, if the network learns to propagate the result of the similar part, both networks should return the same result. However, as the authors stated in the paper, it is not easy to learn the identity function stacking nonlinear layers. With the skip connection, learning the identity function turns out to be easier, allowing the networks to learn more than the plain networks.

In Figure A2, VGG-19 [41], a plain 34-layers network and a ResNet34 neural network are shown. From a technical point of view, our implementation modifies the original architecture of EfficientDet, by means of a substitution of the final layers (the so-called classification and detection heads) by a ResNet18 with two units as outputs, representing the probability of ICH.

However, we cannot simply remove the heads and put a fully connected layer instead, since the BiFPN layer provides five sets of feature maps (the five units with colors shown in Figure A3 just before the two heads), each one at different scale: 64×64, 32×32, 16×16, 8×8 and 4×4, consisting of each set of 64 filters. It is possible to flatten the features up and put a fully connected layer with two neurons as output (i.e., create a list with the outputs of the Bi-FPN layer and connect them to two units, but this would probably not be enough processing for all these feature maps.

Our approach is to upsample every feature map to 224×224, since the ResNet family receives an image with a resolution of 224×224 as input) and passes it to a ResNet18, but we would have two problems:

- Upsampling approaches such as upsampling with nearest interpolation lead to the same problem since we are not adding any information.
- Even without adding any information, ResNet would learn from these features but, as ResNet receives $3 \times 224 \times 224$ (channels, height, width) as input, we would need an intermediate step: reduce 320 (5 sets of 64 filters each one) channels to only 3.

We solve these problems with two intermediate groups of layers between the Bi-FPN and the ResNet: transposed convolution layers as upsampling method and convolution layers to reduce the number of channels.

We call this neural network **EfficientClassification**. The implementation consists of the following steps:

1. **Load EfficientDet.** We get a pretrained EfficientDet-D0. In this step, we use the implementation from the timm library [42]. The timm library can be installed using Pypi.
2. **Cut the regression and classification heads.** We remove the regression and classification heads and only use the EfficientNet backbone and the BiFPN layers. Figure A4 shows this step and the previous one.
3. **Deconvolution.** Different deconvolution operations (transposed convolutions) are applied to the feature maps to get feature maps of 244×224 as dimension. 42 feature maps are passed to the convolution step. The number of channels returned by every deconvolution differs depending on the resolution of the feature maps, as can be seen in Figure A5.
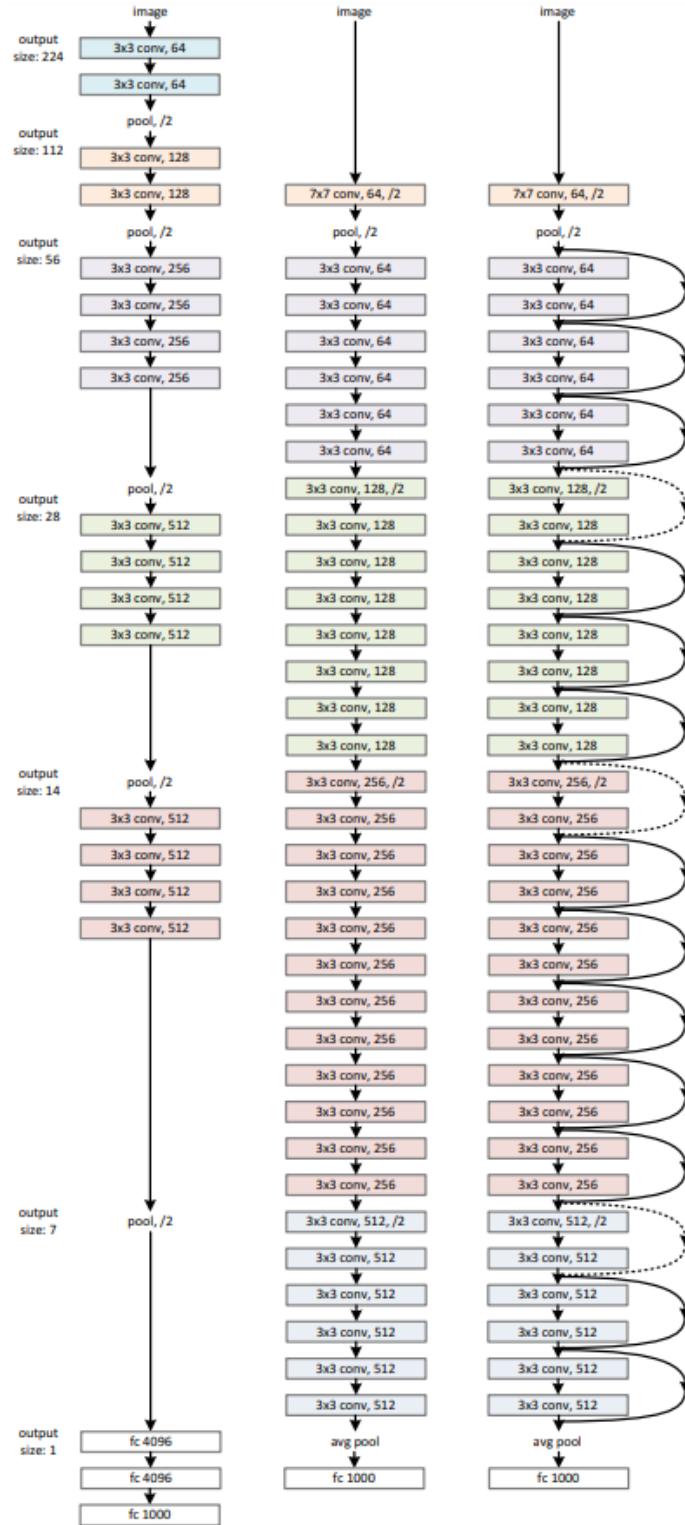
Figure A2. Example network architectures for ImageNet. Left: the VGG-19 model. Middle: a plain network with 34 parameter layers. Right: a residual network with 34 parameter layers. The dotted shortcuts increase dimensions. Source: ResNet paper [7]

4. **Convolution.** In this step, the number of channels is reduced to 3. Two convolution layers are applied as shown in Figure A6.
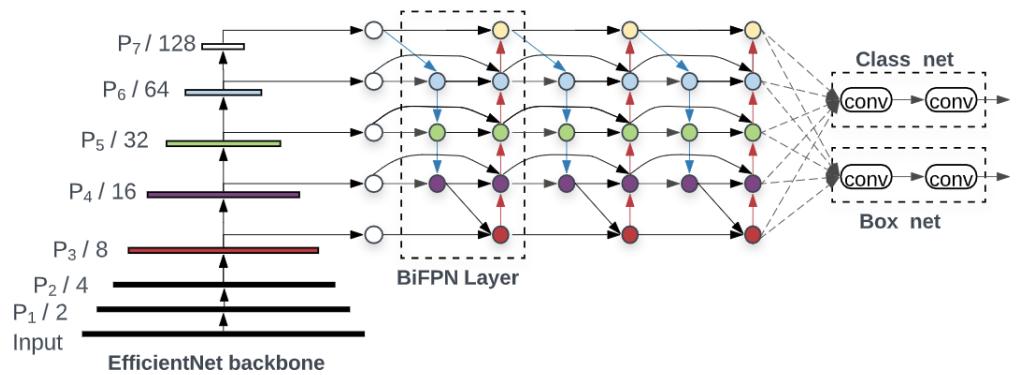
Figure A3. EfficientDet architecture. Source: EfficientDet paper [8].

5. **ResNet.** We call a pretrained ResNet18 from torchvision.models and cut the last layer as we did in the ResNet implementation section. The convolution layers output is passed to the ResNet as input.

```
from effdet import create_model
self.effdet = create_model(model_name='efficientdet_d0')
self.effdet.box_net = nn.Identity()
self.effdet.class_net = nn.Identity()
```

Figure A4. Removing the classification and the regression heads from the EfficientDet. Source: https://github.com/Keredu/Intracranial-Hemorrhage-Detection/blob/main/src/model.py

Figure A5. Deconvolution layers. Source: https://github.com/Keredu/Intracranial-Hemorrhage-Detection/blob/main/src/model.py

The previous steps result in the neural net architecture shown in the Figure A7.

Appendix A.2

Grad-CAM is applied to a trained neural network. Then, the neural network is fed with an image to calculate the Grad-CAM heatmap for that image for a chosen class

Figure A6. Convolution layers. Source: https://github.com/Keredu/Intracranial-Hemorrhage-Detection/blob/main/src/model.py
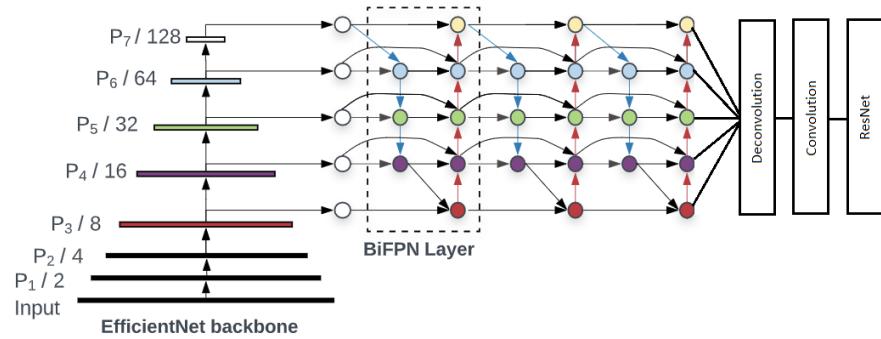
Figure A7. EfficientClassification architecture.

of interest. Grad-CAM has three steps: compute gradients, calculate alpha values by averaging gradients, and calculate the final Grad-CAM heatmap.

Step 1. Compute the gradient of y^c with respect to the feature map activations A^k of a convolutional layer, i.e., $\frac{\partial y^c}{\partial A^k}$, where y^c is the raw output of the neural network for class c, before the softmax is applied to transform the raw score into a probability.
Step 2. Apply global average pooling to the gradients over the width dimension (indexed by i) and the height dimension (indexed by j) to obtain the neuron importance weights a_k^c, producing k weights:

$$a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k} \tag{A1}$$

Step 3. Perform a weighted combination of the feature map activations A^k with the weights, a_k^c, calculated in the previous step:

$$L_{\text{Grad-CAM}}^c = ReLU \left(\sum_k a_k^c A^k \right) \tag{A2}$$

Typically, a resize operation is performed afterwards, since the Grad-CAM heatmap size is typically smaller than the original image size.

Appendix A.3

Since we needed to conduct many experiments due to both the problem complexity and the network architecture we used, we created a framework that controls the experiment pipeline.

This framework is built on top of the pytorch library [43] and consists of the following:

- **Configuration directory**. This directory contains the YAML files that govern the experiments. There are three types of configuration files corresponding to three different possible phases: training, validation, and testing.
- **Experiments directory**. If this directory does not exist, it is created automatically. It contains directories with the experiment code and, inside of them, there are loss

and accuracy graphs, the configuration file used in the experiment, and the metrics (ROC-AUC, confusion matrix, and so on).

- **Scripts directory**. This directory contains the different scripts that were used in the different phases. The most important one is the **config.py** file, which leads the execution and loads the configuration from the YAML config file.
- **Main file**. The main.py file just receives YAML config files and calls the config.py and the corresponding task script, allowing us to execute multiple experiments.

Figure A8 shows the framework repository tree in the experiment called R34_001.
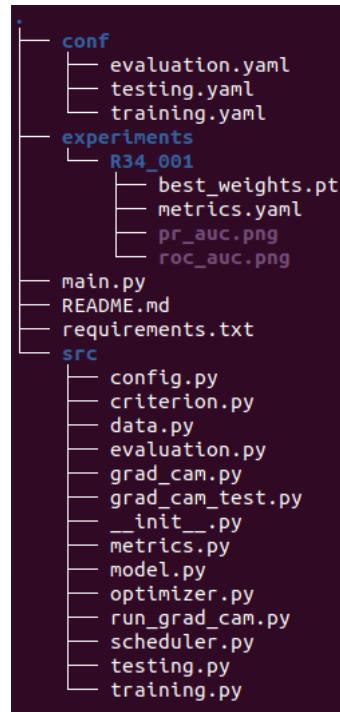
Figure A8. Framework repository tree.

References

1. Valery L Feigin *et al*. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Neurology* **2017**, *16*, 877–897. doi:https://doi.org/10.1016/S1474-4422(17)30299-5.
2. Caceres, J.A.; Goldstein, J.N. Intracranial hemorrhage. *Emerg Med Clin North Am* **2012**, *30*, 771–794.
3. Fogelholm, R.; Murros, K.; Rissanen, A.; Avikainen, S. Long term survival after primary intracerebral haemorrhage: a retrospective population based study. *J Neurol Neurosurg Psychiatry* **2005**, *76*, 1534–1538.
4. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI, 2019. doi:10.48550/ARXIV.1907.10597.
5. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* **2019**, *128*, 336–359. doi:10.1007/s11263-019-01228-7.
6. Burduja, M.; Ionescu, R.T.; Verga, N. Accurate and Efficient Intracranial Hemorrhage Detection and Subtype Classification in 3D CT Scans with Convolutional and Long Short-Term Memory Neural Networks. *Sensors (Basel)* **2020**, *20*.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *CoRR* **2015**, *abs/1512.03385*, [1512.03385].
8. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection, 2020, [arXiv:cs.CV/1911.09070].
9. Jaja, B.N.; Cusimano, M.D.; Etminan, N.; Hanggi, D.; Hasan, D.; Ilodigwe, D.; Lantigua, H.; Le Roux, P.; Lo, B.; Louffat-Olivares, A.; Mayer, S.; Molyneux, A.; Quinn, A.; Schweizer, T.A.; Schenk, T.; Spears, J.; Todd, M.; Torner, J.; Vergouwen, M.D.; Wong, G.K.; Singh, J.; Macdonald, R.L. Clinical prediction models for aneurysmal subarachnoid hemorrhage: a systematic review. *Neurocrit Care* **2013**, *18*, 143–153.
10. Etminan, N.; Chang, H.S.; Hackenberg, K.; de Rooij, N.K.; Vergouwen, M.D.I.; Rinkel, G.J.E.; Algra, A. Worldwide Incidence of Aneurysmal Subarachnoid Hemorrhage According to Region, Time Period, Blood Pressure, and Smoking Prevalence in the Population: A Systematic Review and Meta-analysis. *JAMA Neurology* **2019**, *76*, 588–597. doi:10.1001/jamaneurol.2019.0006.
11. Al-Kawaz, M.N.; Hanley, D.F.; Ziai, W. Advances in Therapeutic Approaches for Spontaneous Intracerebral Hemorrhage. *Neurotherapeutics* **2020**, *17*, 1757–1767.

12. Cordonnier, C.; Demchuk, A.; Ziai, W.; Anderson, C.S. Intracerebral haemorrhage: current approaches to acute management. *The Lancet* **2018**, *392*. doi:10.1016/S0140-6736(18)31878-6.

13. Cordonnier C, Demchuck A, Z.W.; CS., A. Erratum: Intracerebral haemorrhage: current approaches to acute management (The Lancet (2018) 392(10154) (1257–1268), (S0140673618318786) (10.1016/S0140-6736(18)31878-6)). *The Lancet* **2019**, *393*. doi:10.1016/S0140-6736(19)30159-X.

14. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016. http://www.deeplearningbook.org.

15. Fuad, M.T.H.; Fime, A.A.; Sikder, D.; Iftee, M.A.R.; Rabbi, J.; Al-Rakhami, M.S.; Gumaei, A.; Sen, O.; Fuad, M.; Islam, M.N. Recent Advances in Deep Learning Techniques for Face Recognition. *IEEE Access* **2021**, *9*, 99112–99142. doi:10.1109/access.2021.3096136.

16. Hernandez-Olivan, C.; Beltran, J.R. Music Composition with Deep Learning: A Review, 2021, [arXiv:cs.SD/2108.12290].

17. Pan, X.; Wang, M.; Wu, L.; Li, L. Contrastive Learning for Many-to-many Multilingual Neural Machine Translation, 2021, [arXiv:cs.CL/2105.09501].

18. Peixoto, F. A Simple overview of Multilayer Perceptron. Available online https://www.analyticsvidhya.com/blog/2020/12/mlp-multilayer-perceptron-simple-overview/ Accessed on 2021-01-25, 2020.

19. KDnuggets. Available online https://www.kdnuggets.com/2016/02/opening-deep-learning-everyone.html Accessed on 2021-01-25.

20. Shorfuzzaman, M.; Hossain, M.S. MetaCOVID: A Siamese neural network framework with contrastive loss for *n*-shot diagnosis of COVID-19 patients. *Pattern Recognit.* **2021**, *113*, 107700. doi:10.1016/j.patcog.2020.107700.

21. Fan, D.; Zhou, T.; Ji, G.; Zhou, Y.; Chen, G.; Fu, H.; Shen, J.; Shao, L. Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images. *IEEE Trans. Medical Imaging* **2020**, *39*, 2626–2637. doi:10.1109/TMI.2020.2996645.

22. Nair, T.; Precup, D.; Arnold, D.L.; Arbel, T. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. *Medical Image Anal.* **2020**, *59*. doi:10.1016/j.media.2019.101557.

23. Kaggle Competition: RSNA intracranial Hemorrhage Detection. Available online https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection Accessed on 2021-01-25.

24. Gildenblat, J.; contributors. PyTorch library for CAM methods. https://github.com/jacobgil/pytorch-grad-cam, 2021.

25. Viniavskyi, O.; Dobko, M.; Dobosevych, O. Weakly-Supervised Segmentation for Disease Localization in Chest X-Ray Images. Artificial Intelligence in Medicine; Michalowski, M.; Moskovitch, R., Eds.; Springer International Publishing: Cham, 2020; pp. 249–259.

26. Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos Solitons Fractals* **2020**, *140*, 110190.

27. Voter, A.F.; Meram, E.; Garrett, J.W.; Yu, J.J. Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Intracranial Hemorrhage. *J Am Coll Radiol* **2021**. doi:10.1016/j.jacr.2021.03.005.

28. Kuo, W.; Häne, C.; Mukherjee, P.; Malik, J.; Yuh, E.L. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proceedings of the National Academy of Sciences* **2019**, *116*, 22737–22745. doi:10.1073/pnas.1908021116.

29. Kishan Das Menon, H.; Janardhan, V. Intracranial hemorrhage detection. *Materials Today: Proceedings* **2021**, *43*, 3706–3714. International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology, doi:https://doi.org/10.1016/j.matpr.2020.10.982.

30. Sage, A.; Badura, P. Intracranial Hemorrhage Detection in Head CT Using Double-Branch Convolutional Neural Network, Support Vector Machine, and Random Forest. *Applied Sciences* **2020**, *10*. doi:10.3390/app10217577.

31. Gruschwitz, P.; Grunz, J.P.; Kuhl, P.J.; Kosmala, A.; Bley, T.A.; Petritsch, B.; Heidenreich, J.F. Performance testing of a novel deep learning algorithm for the detection of intracranial hemorrhage and first trial under clinical conditions. *Neuroscience Informatics* **2021**, *1*, 100005. doi:https://doi.org/10.1016/j.neuri.2021.100005.

32. Lee, J.Y.; Kim, J.S.; Kim, T.Y.; Kim, Y.S. Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm. *Scientific Reports* **2020**, *10*, 20546. doi:10.1038/s41598-020-77441-z.

33. Kim, J.S.; Cho, Y.; Lim, T.H. Prediction of the Location of the Glottis in Laryngeal Images by Using a Novel Deep-Learning Algorithm. *IEEE Access* **2019**, *7*, 79545–79554. doi:10.1109/ACCESS.2019.2923002.

34. AIDoc Medical Ltd. Available online https://www.aidoc.com/ Accessed on 2021-01-25.

35. Wang, X.; Shen, T.; Yang, S.; Lan, J.; Xu, Y.; Wang, M.; Zhang, J.; Han, X. A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. *NeuroImage: Clinical* **2021**, *32*. cited By 11, doi:10.1016/j.nicl.2021.102785.

36. Brinjikji, W.; Abbasi, M.; Arnold, C.; Benson, J.C.; Braksick, S.A.; Campeau, N.; Carr, C.M.; Cogswell, P.M.; Klaas, J.P.; Liebo, G.B.; Little, J.T.; Luetmer, P.H.; Messina, S.A.; Nagelschneider, A.A.; Schwartz, K.M.; Wood, C.P.; Nasr, D.M.; Kallmes, D.F. e-ASPECTS software improves interobserver agreement and accuracy of interpretation of aspects score. *Interventional Neuroradiology*. doi:10.1177/15910199211011861.

37. Dubosh, N.M.; Bellolio, M.F.; Rabinstein, A.A.; Edlow, J.A. Sensitivity of Early Brain Computed Tomography to Exclude Aneurysmal Subarachnoid Hemorrhage: A Systematic Review and Meta-Analysis. *Stroke* **2016**, *47*, 750–755. doi:10.1161/STROKEAHA.115.011386.

38. Arbabshirani, M.R.; Fornwalt, B.K.; Mongelluzzo, G.J.; Suever, J.D.; Geise, B.D.; Patel, A.A.; Moore, G.J. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digital Medicine* **2018**, *1*, 9. doi:10.1038/s41746-017-0015-z.

39. Brainscan.ai. Available online https://brainscan.ai/ Accessed on 2021-01-25.

40. Brzeski A. BRAINSCAN.AI, Private communication.

41. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* **2014**, *abs/1409.1556*.

42. Wightman, R. PyTorch Image Models. https://github.com/rwightman/pytorch-image-models, 2019. doi:10.5281/zenodo.4414861.

43. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H.; Larochelle, H.; Beygelzimer, A.; dAlché-Buc, F.; Fox, E.; Garnett, R., Eds.; Curran Associates, Inc., 2019; pp. 8024–8035.