

Hybrid Convolutional Neural Network-Based Diagnosis System for Intracranial Hemorrhage

Sezin BARIN¹,
Murat SARIBAŞ²,
Beyza Gülizar ÇILTAŞ³,
Gür Emre GÜRAKSIN⁴,
Utku KÖSE⁵

¹Afyon Kocatepe University, Department of Biomedical Engineering, Afyonkarahisar, Turkey, sbarin@aku.edu.tr

²Afyon Kocatepe University, Department of Biomedical Engineering, Afyonkarahisar, Turkey, muratsaribas97@gmail.com

³Afyon Kocatepe University, Department of Biomedical Engineering, Afyonkarahisar, Turkey, gcltaas@gmail.com

⁴Afyon Kocatepe University, Department of Computer Engineering, Afyonkarahisar, Turkey, emreguraksin@aku.edu.tr

⁵Suleyman Demirel University, Department of Computer Engineering, Isparta, Turkey, utkukose@sdu.edu.tr

Abstract: *Early diagnosis of intracranial hemorrhage significantly reduces mortality. Hemorrhage is diagnosed by using various imaging methods and the most time-efficient one among them is computed tomography (CT). However, it is clear that accurate CT scans requires time, diligence, and experience. Computer-aided design methods are vital for the treatment because they facilitate early diagnosis of intracranial hemorrhage. At this point, deep learning can provide effective outcomes through an automated diagnosis way. However, as different from the known solutions, diagnosis of five different hemorrhage subtypes is a critical problem to be solved. This study focused on deep learning methods and employed cranial computed tomography scans in order to detect intracranial hemorrhage. The diagnosis approach in the study aimed to detect five subtypes of hemorrhage. In detail, EfficientNet-B3 and ResNet-Inception-V2 architectures were used for diagnosis purposes. Eventually, the study also proposed a two-architecture hybrid method for the diagnosis purpose. The obtained findings by the hybrid method were evaluated in terms of a comparative perspective. Results showed that the newly designed hybrid method was quite effective in terms of increasing classification rates of detecting intracranial hemorrhage according to the subtypes. Briefly, an accuracy of 98.5%, which is higher than those of the EfficientNet-B3 and the Inception-ResNet-V2, were obtained thanks to the developed hybrid method.*

Keywords: *Deep Convolutional Neural Network; Intracranial Hemorrhage; Computer Tomography; Inception-ResNet-V2; EfficientNet-B3.*

How to cite: Barin, S., Saribaş, M., Çiltaş, B.G., Güraksin, G.E., & Köse, U. (2021). Hybrid Convolutional Neural Network-Based Diagnosis System for Intracranial Hemorrhage. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 12(4), 01-27. <https://doi.org/10.18662/brain/12.4/236>

1. Introduction

Intracranial hemorrhage is a type of bleeding, which is caused by the rupture of a blood vessel in the brain. That situation results to the leaking of blood into the brain tissue or the space between the cortex and the skull. So, it is a serious health problem that requires rapid and intensive treatment as there may be different reasons (trauma, stroke, aneurysm, vascular malformations, high blood pressure, illegal drugs, and blood clotting disorders) for that and may result in serious health problems and even death (Gardner et al., 2012). There are five different subtypes of intracranial hemorrhage as intraventricular (bleeding into the ventricles inside the brain), intraparenchymal (bleeding into the brain tissue), subarachnoid (bleeding into the space between two of the membranes surrounding the brain), subdural (bleeding into the space between the outermost meninx and the arachnoid), and epidural (bleeding into the space between the outermost meninx and the skull) (Gardner et al., 2012).

Intracranial hemorrhage correspond to approximately ten percent of all strokes in the U.S. Stroke is the fifth-leading cause of death in the U.S, as resulting 129,000 deaths each year (Jauch et al., 2013). Essential imaging methods used in the diagnosis of intracranial hemorrhage are positron emission tomography (PET), cerebral angiography, computed tomography-angiography (CT-A), magnetic resonance imaging (MRI), and magnetic resonance angiography (MRA) (Muir et al., 2006). Although there is a remarkable variety of imaging methods as it can be seen, all of them take considerably long time to be processed. For example, cerebral angiography takes about three hours. Therefore, cranial computed tomography is the best method for the diagnosis of acute intracranial hemorrhage because it takes only 1 to 5 minutes (Phong et al., 2017). Computed tomography can detect hemorrhage for more than 98% of patients, as considering the first two days of hemorrhage. However, the duration of diagnosis depends on how long it takes the expert to complete and interpret a cranial CT scan (Medicine Hospital, n.d.).

In recent years, deep learning has been very popular in classification and segmentation of medical images. Since its first appearance in the context of artificial intelligence field, it has been proven to be very effective in medical image oriented applications. Therefore, there is also a growing body of research on the use of deep learning in classification of intracranial computed tomography scans (Web of Science, n.d.). Arbabshirani et al. (2018) used 46,583 non-contrast head CT scans of 31,256 patients and proposed a convolutional neural network (CNN) architecture (employing

five convolutional and two fully-connected layers) for intracranial hemorrhage detection. Arjun Majumdar et al. (2018) classified 4300 scans of 134 cases into six groups, which are respectively no hemorrhage, epidural, subdural, subarachnoid, intraparenchymal, and intraventricular. They designed a CNN model of nine blocks with one convolution layer in each block and used maximum pooling (for data reduction) and a 2x2 nearest neighbor augmentation method (after the first four blocks). Phong et al. (2017) focused on a dataset of 2000 CT scans (20 scans of each case in 115 hospitals in Vietnam) and developed a system in order to distinguish CT scans as normal or abnormal (presence of intracranial hemorrhage). They used the CNN methods of LeNet, GoogLeNet, and Inception-ResNet-V2 and reached to the accuracy values of 99%, 98%, and 99% respectively. Helwan et al. (2018) used a dataset of 2.527 CT scans (2,142 training and 385 tests) and developed an algorithm, which classified head CT scans as hemorrhage and no hemorrhage. They used an auto-encoder (AE) model with one hidden layer, a stacked auto-encoder (SAE) model with two hidden layers, and a CNN model to conclude that the SAE model had the highest accuracy (90.9%). Al-Ayyoub et al. (2013) developed a computer-aided diagnostic system to determine the presence of hemorrhage and three subtypes of hemorrhage (epidural, subdural, intraparenchymal). They used a data set of 76 CT (25 normal, 17 epidural hemorrhage, 20 subdural hemorrhage, and 14 intraparenchymal hemorrhages) scans from King Abdullah University Hospital. The system employed image preprocessing, image segmentation, feature extraction, and classification stages respectively. They used Otsu's method for segmentation and extracted the regions of interest (ROI) for classification. With the applications done via BayesNet, J48, Logistic, MLP (Multi-Layer Perceptron), and SVM (Support Vector Machine) classifiers in WEKA, they achieved 100% accuracy in hemorrhage detection and 92% accuracy in hemorrhage subtype classification (Al-Ayyoub et al. 2013). Chilamkurthy et al. (2018) developed a deep learning method for head CT scans to classify the presence of hemorrhage, hemorrhage subtypes, calvarial fractures, midline shift, and mass effect. For the verification purpose, they used some scans, which were randomly selected from the Qure25k data set (with 313318 scans) and employed the remaining scans for developing an algorithm. For the testing phase, they used the CQ500 dataset collected in two batches. The Qure25k dataset consists of 21095 scans, while the CQ500 dataset consists of 491 scans (first batch 214, second batch 277). In the Qure25k dataset, they achieved an accuracy of 92%, 90%, 96%, 92%, 93%, 90%, 92%, 93%, and 86% for intracranial, intraparenchymal, intraventricular, subdural, extradural, and

subarachnoid hemorrhage, and calvarial fractures, midline shift, and mass effect respectively. In the CQ500 dataset, they achieved an accuracy of 94%, 95%, 93%, 95%, 97%, 96%, 96%, 97%, and 92% for intracranial, intraparenchymal, intraventricular, subdural, extradural, and subarachnoid hemorrhage, and calvarial fractures, midline shift, and mass effect respectively. Salehinejad et al. (2021) developed a machine learning model by using SEResNeXt-50 and SEResNeXt-101 that was trained with a total of 21,784 scans from the RSNA Intracranial Hemorrhage CT dataset. They tested their model for 3528 scans from the RSNA Intracranial Hemorrhage CT dataset and achieved an accuracy of 98.3%, sensitivity of 98.8%, specificity of 98%, for five classes.

Objective of this study is to develop an alternative, effective deep learning method to diagnose hemorrhage, by considering detection of five subtypes. Moving from that, the study focused on a dataset of 752.803 head CT scans [an open dataset on Kaggle Kaggle (RSNA Intracranial Hemorrhage Detection, n.d.)] collected by the Radiology Society of North America (RSNA). The size of the dataset increased the reliability of the trained architectures. The study used the latest EfficientNet-B3 (Tan & Le, 2019) and ResNet-Inception-V2 (Szegedy et al., 2017) deep learning architectures for intracranial hemorrhage detection and its five subtypes. The study compared the high-dimensional dataset and the latest deep neural network architectures and eventually, proposed a two-architecture hybrid method. The developed hybrid method was compared with alternative architectures and the results showed that this newly designed method / architecture was more successful than the latest deep neural network architectures.

2. Methodology

2.1. Dataset

The dataset employed in this study was derived from Kaggle, which is an open-source database provided by the RSNA. The dataset consists of 752.803 head CT scans of six labels, which are respectively no hemorrhage and five subtypes of hemorrhage (subdural, epidural, subarachnoid, intraparenchymal, and intraventricular) (RSNA Intracranial Hemorrhage Detection, n.d.). Figure 1 shows some CT scans from the dataset and the distribution state according to the classes. The dataset was divided into two groups: training (90%) and test (10%). However, as seen in Figure 1, the dataset has imbalanced distribution. Epidural images are very few. For this reason, another test was conducted by reducing the number of images of

other classes. Five hundred images from each class except the epidural and 314 images from the epidural class were taken for this test. However, one image contains multiple hemorrhages; therefore, Test Dataset2 values in Figure 1 are different.

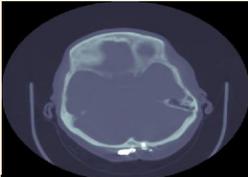
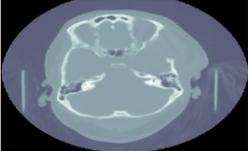
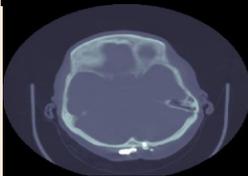
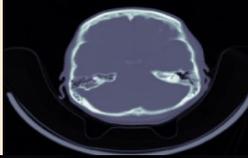
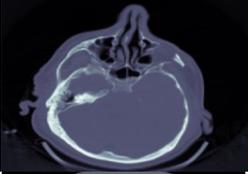
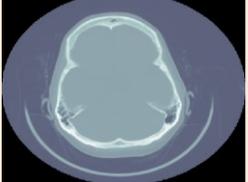
		Train Dataset	Test Dataset1	Test Dataset2
Epidural		1516	314	314
Intraparenchymal		32506	3612	964
Intraventricular		23584	2621	811
Subdural		32107	3568	929
Subarachnoid		42431	4785	893
Non-Hemorrhage		580382	64488	2500
Total		677522	75281	4871

Figure 1. CT Scans from Kaggle dataset and distribution of examination labels according to hemorrhage subtypes

2.2. Image Preprocessing

Image processing phase had several steps to prepare the images for better diagnosis. In this context, the scans were pretreated several times to train the deep learning architectures. First, they were resized to 300x300 for EfficientNet-B3 (Tan & Le, 2019) and 299x299 for Inception-ResNet-V2 (Szegedy et al., 2017) models, as considering the input layer dimensions of the deep neural network architectures. Following to that, as suggested by Chilamkurthy et al. (2018), the entire dynamic range of CT densities was defined as three separate windows, which are brain (length: 40 - weight: 80), subdural (length: 80 - weight: 200), and soft tissue (length: 40 - weight: 380). That is because a fracture in a bone window indicates the presence of an extra hemorrhage in a brain window, or a fracture in a subdural window indicates the presence of a hemorrhage, which is indistinguishable in a skull and normal brain window. Before the architectures were trained, the training images were flipped horizontally and vertically for the data augmentation.

2.3. Deep Learning with Transfer Learning

Deep learning is a branch of machine learning that is run with algorithms similar to the deep hierarchical architecture of the brain. Deep learning is based on deep artificial neural networks (Şeker et al., 2017). Neural networks have numerous layers and even layers within layers, hence owning the name of deep. There is a growing body of research on deep learning methods in the field of biomedicine (Figure 2) because they are more time- and cost-effective and also better in the detection and diagnosis of diseases than state-of-the-art methods (Web of Science, n.d.).

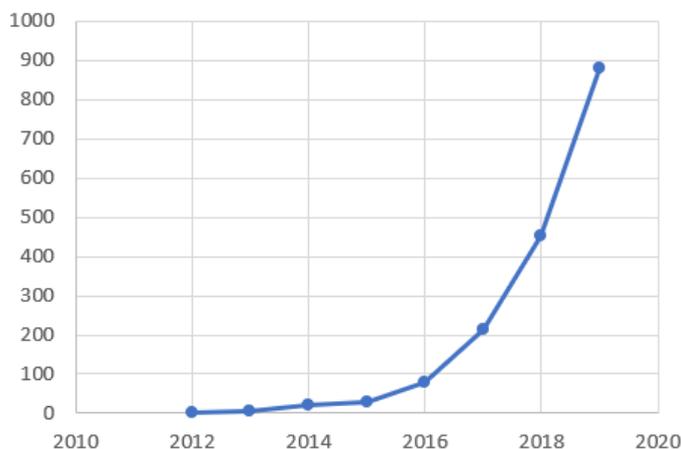


Figure 2. Distribution of studies on deep learning methods in biomedical field, by years (Web of Science, n.d.)

In the context of this study, two ready-to-use CNN-based deep neural network models were trained with ImageNet (n.d.) and eventually made ready accordingly for object classification. These deep CNN architectures were used to solve the problem characterized by a transfer learning method. By allowing for faster training and better performance, transfer learning has much better results because of the size of the dataset. Transfer learning is a solution that focuses on storing knowledge gained while solving a problem and applying that knowledge to a different but related problem. In other words, transfer learning is a deep learning method by which an artificial neural network model trained for a task is redesigned for a different but related task. In order to classify brain CT scans, this study employed two most recent CNN models trained with the ImageNet (n.d.), which is one of the largest image databases. The two CNN models employed in this context were EfficientNet-B3 (Tan & Le, 2019) and Inception-ResNet-V2 (Szegedy et al., 2017) respectively. After the training phase, the classification performance of the two models was separately compared, and the class-based average values of their classification probabilities were compared.

2.3.1. EfficientNet- B3

At the International Conference on Machine Learning (ICML) in 2019, Google introduced a CNN-based EfficientNet neural network architecture with a new structural approach (Tan & Le, 2019). A convolutional neural network (CNN) is known as a powerful class of deep neural networks, especially in image processing applications. A CNN architecture consists of input and output layers and intermediate layers, which are also known as hidden layers, as located between them. The main intermediate layers in a CNN architecture are convolution, pooling, and fully connected layers. Essential functions of these layers are as follows: The convolution layer is the layer where an activation map is generated by addition and multiplication operations via filters from the input data. In other words, it is the layer where feature extraction takes place (Stanford.edu, 2018). The pooling layer is the layer where nonlinear subsampling is performed, and the number of parameters is reduced for ensuring a simpler output (Stanford.edu, 2018). The fully connected layer converts the data from the previous layer into a one-dimensional matrix, in order to make the data fully connected to all neurons in the next layer (Stanford.edu, 2018). The fully connected layer generally precedes the classification layer, which is the last layer of a CNN architecture. To date, the depth, that is, the number of layers has been increased to improve the

performance of all architectures. However, that situation increased the cost of computing. So, because the accuracy reached saturation after a certain point, there was no increase in the success level.

The proposed EfficientNet- B3 structure is much better than other CNN architectures in increasing success without increasing the depth. However, the main feature of the architecture is that the model referred to as compound scaling increases not only the depth but also the parameters of width and resolution (Tan & Le, 2019). Figure 3 compares the proposed architecture with the other CNN architectures and shows that the EfficientNet (Tan & Le, 2019) performed better despite having fewer parameters.

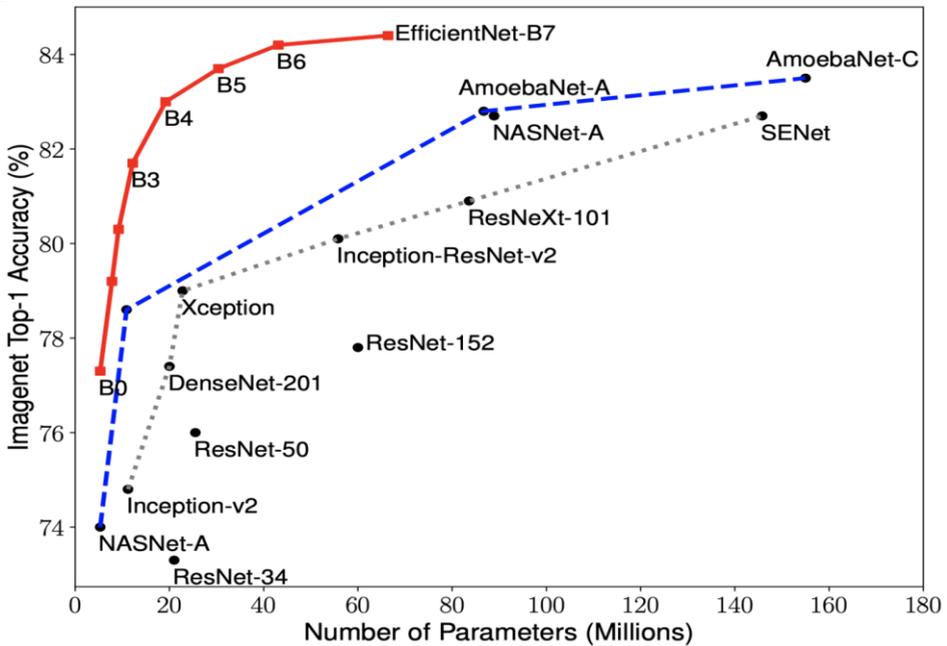


Figure 3. Accuracy Comparison of ImageNet top-1 (Tan & Le, 2019)

The EfficientNet (Tan & Le, 2019) consists of eight models from B0 to B7, with each subsequent model number indicating higher accuracy. Figure 4 shows the EfficientNet-B0 (Tan & Le, 2019) architecture. The model of choice in this study has been EfficientNet-B3(300x300) (Tan & Le, 2019) because it has input dimensions similar to those of the other CNN architecture: Inception-ResNet-V2 (299x299) (Szegedy et al., 2017), as allowing us to ignore the effect of resolution while comparing the architectures. Due to its input size and number of parameters, the EfficientNet-B3 (Tan & Le, 2019) model is different from the B0 model.

Both models have seven blocks, each of which has a different number of mobile inverted bottleneck convolution (MBCoV). Therefore, the parameter size of the B0 model is 5.3M while that of the B3 model is 12M.

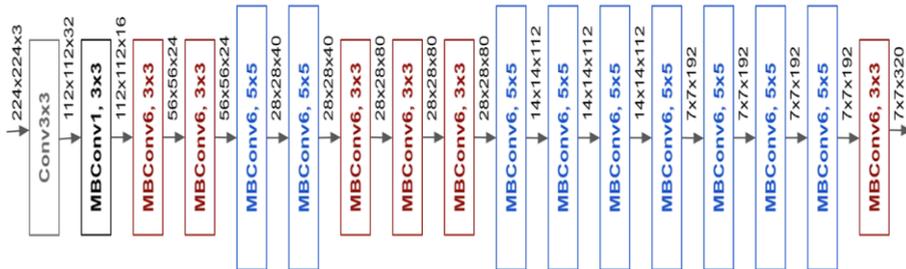


Figure 4. The Architecture of EfficientNet-B0 (Tan & Le, 2019)

In the training phase of the EfficientNet-B3 (Tan & Le, 2019) model, different parameters were used, and the most successful values were chosen accordingly. Here, the Adam was used as the optimization algorithm with a learning rate of 0.0001, batch size of 16, and binary cross-entropy as the loss function, whereas the sigmoid function was the activation function in the final classification layer. 10-fold cross-validation was also used within the related process.

2.3.2. Inception-ResNet-V2

The Inception-ResNet-V2 (Szegedy et al., 2017) is a CNN architecture based on the combination of Inception structure and the Residual connection, as trained with more than a million images in the ImageNet (n.d.) database. The network employs 164 layers deep and comes with the learned rich feature representations for different images, thanks to the diversity of the training set. The input image size for the model is 299x299. Figure 5 shows the basic architecture of the Inception-ResNet-V2 (Szegedy et al., 2017). In the Inception-Resnet-V2 (Szegedy et al., 2017) blocks, multi-dimensional convolution filters are combined with residual connections. Residual connections do not only prevent distortion caused by deep structures but also shorten the training time.

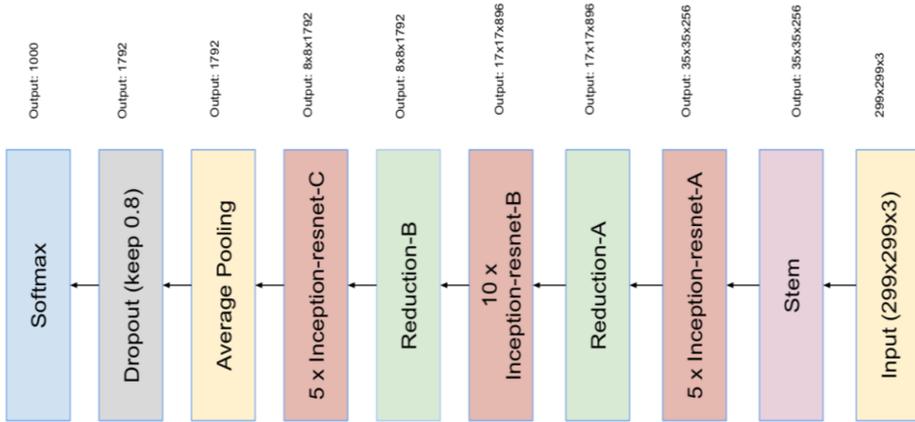


Figure 5. Main Architecture of Inception-ResNet-V2 (Szegedy et al., 2017.)

Parameters, which were selected for the training of the Inception-ResNet-V2 model, were similar to those in the EfficientNet-B3 model. The Adam was used as the optimization algorithm with a learning rate of 0.0001, and batch size of 32, with binary cross-entropy as the loss function, and the sigmoid as the activation function in the final classification layer. 10-fold cross-validation was also used within the process.

2.3.3. Recommended Inception-ResNet-V2 and EfficientNet-B3 based hybrid model

Both architectures were trained separately with the same training set. The mean probability values for each class in the final sigmoid classification layer of both architectures were calculated for obtaining new probability values. Figure 6 shows the flow chart for the proposed hybrid architecture. Additionally, Figure 7 shows the block diagram for the proposed hybrid model. As seen in Figure 7, EfficientNet-B3 has seven blocks, by also having a different number of mobile inverted bottleneck convolution. The Inception-ResNet-V2 has 3 Inception blocks and 2 Reduction blocks. On the other hand, input size of the EfficientNet-B3 is 300x300 while the Inception ResNet-V2 ensures the input size of 299x299.

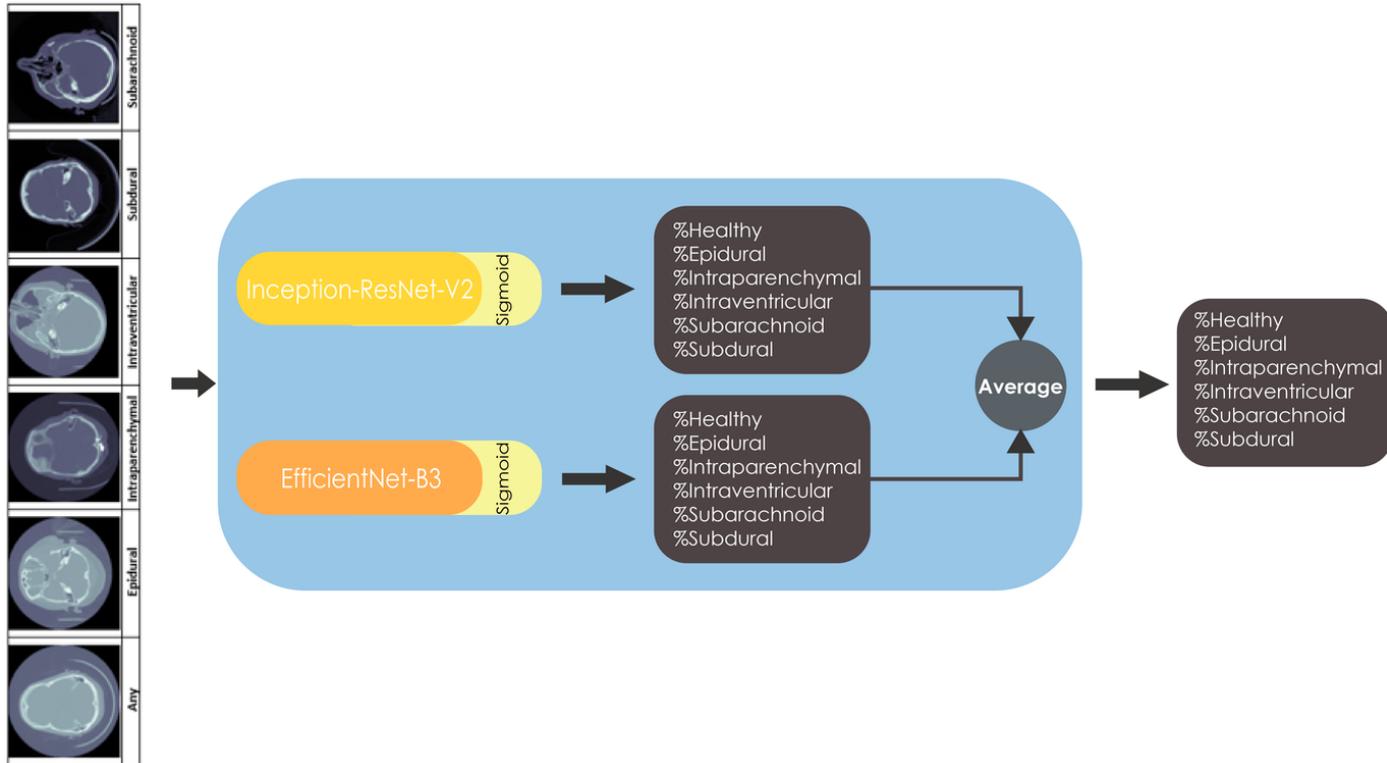


Figure 6. Recommended Inception-ResNet-V2 (Szegedy et al., 2017) and EfficientNet-B3 (Tan & Le, 2019) based hybrid model

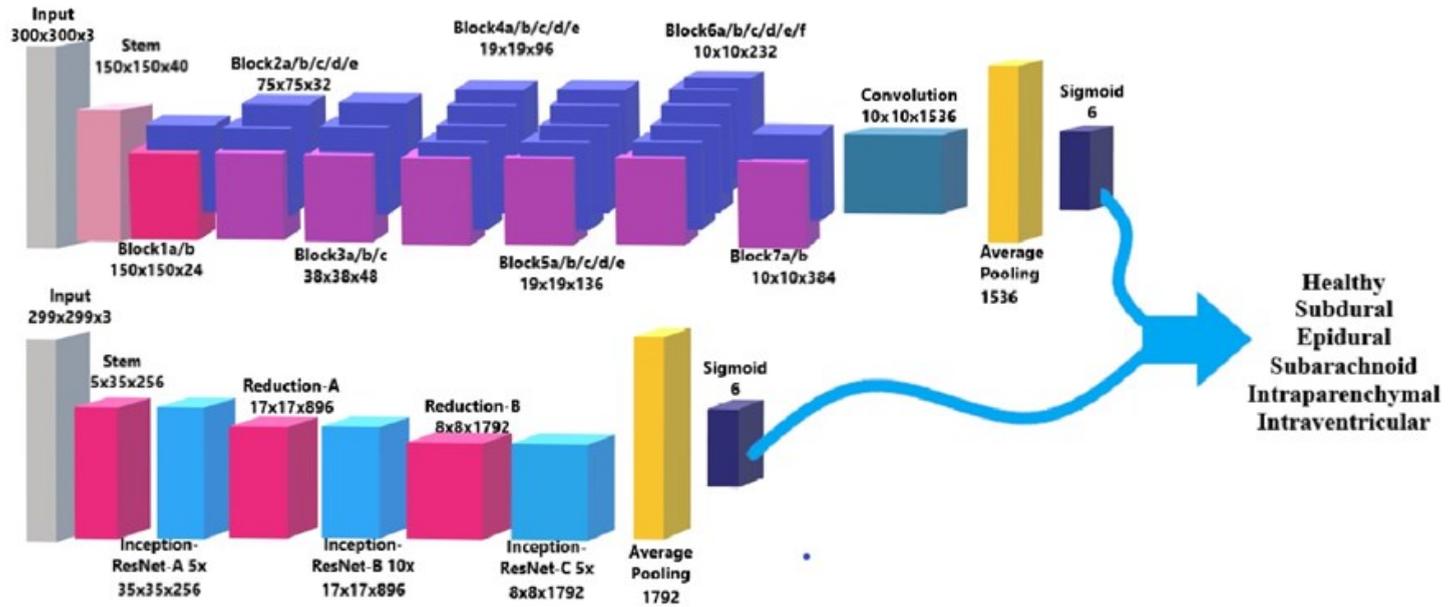


Figure 7. Recommended Inception-ResNet-V2 (Szegedy et al., 2017) and EfficientNet-B3 (Tan & Le, 2019) based hybrid model block diagram

3. Findings and Discussion

Table 1 compares the obtained results for the three methods of the study. 10-fold cross-validation was used, therefore the related results are the lowest, highest, and average accuracy values for 10-fold of each model. The success criteria of F1 score (equation 5), precision (equation 4), sensitivity (equation 3), specificity (equation 2), and accuracy (equation 1) were used to measure the success of the proposed hybrid architecture.

$$ACC = \left(\frac{TP+TN}{TP+TN+FP+FN} \right) * 100 \quad (1)$$

$$SPEC = \left(\frac{TN}{TN+FP} \right) * 100 \quad (2)$$

$$SENS = \left(\frac{TP}{TP+FN} \right) * 100 \quad (3)$$

$$PREC = \left(\frac{TP}{TP+FP} \right) * 100 \quad (4)$$

$$F_1 = 2 * \frac{PREC * SENS}{PREC + SENS} * 100 \quad (5)$$

The value of TP (True Positive) is used when a person with a disease is classified as a patient. FP (False Positive) is used when a healthy person is classified as a patient. Additionally, the value of TN (True Negative) indicates the healthy person classified as healthy, and the FN (False Negative) is used when a person with a disease is classified as healthy. Accuracy is the ratio of the number of correctly classified images to the total number of images. Precision refers to whether images classified belong to the class to which they are referred. Specificity is the correct classification rate of 'no hemorrhage' images. Sensitivity is the correct classification rate of 'hemorrhage' images. Based on these criteria, the deep learning models were applied to 75280 CT scans of six labels; no hemorrhage and subdural, epidural, subarachnoid, intraparenchymal, and intraventricular hemorrhage. Table 1 shows the three models' results, while Figure 8 and Figure 9 show the results' bar graph. Information in Figure 8 belongs to Test Dataset 1, and information in Figure 9 belongs to Test Dataset2.

Table 1. Results of Deep Learning Architectures

MODEL	Precision	Sensitivity	Specificity	F1 Score	Accuracy
EfficientNet-B3 (Min)	0.8596	0.7983	0.9922	0.8494	0.9812
EfficientNet-B3 (Max)	0.8838	0.7931	0.9937	0.8569	0.9824
EfficientNet-B3 (Avg)	0.8680	0.8061	0.9926	0.8531	0.982
Inception-ResNet-V2 (Min)	0.8446	0.8171	0.9910	0.8379	0.9811
Inception-ResNet-V2 (Max)	0.8585	0.8201	0.9919	0.8493	0.9822
Inception-ResNet-V2 (Avg)	0.8627	0.8106	0.9922	0.8441	0.9819
Recommended Hybrid Model (Min)	0.9068	0.8164	0.9950	0.8648	0.9848
Recommended Hybrid Model (Max)	<u>0.9119</u>	<u>0.8314</u>	<u>0.9952</u>	<u>0.8732</u>	<u>0.9859</u>
Recommended Hybrid Model (Avg)	0.9057	0.8269	0.9948	0.8676	0.9853

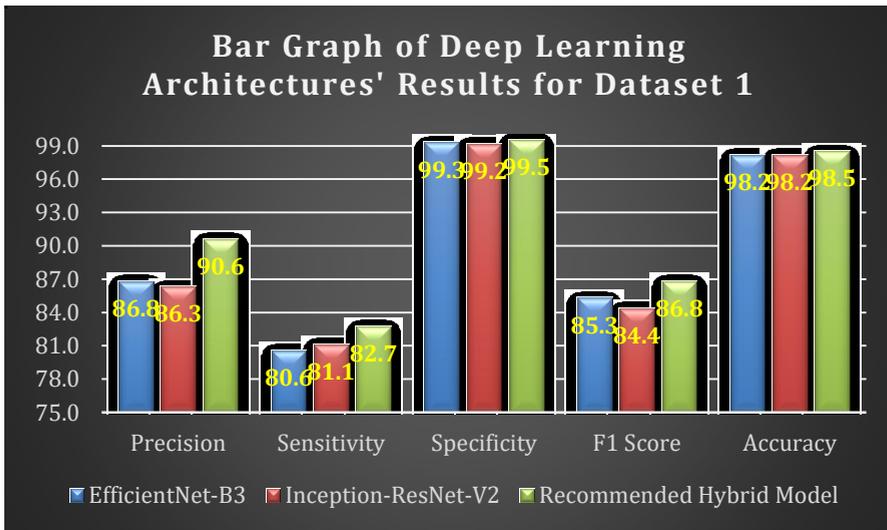


Figure 8. Bar Graph of CNN Architecture Results for Test Dataset 1

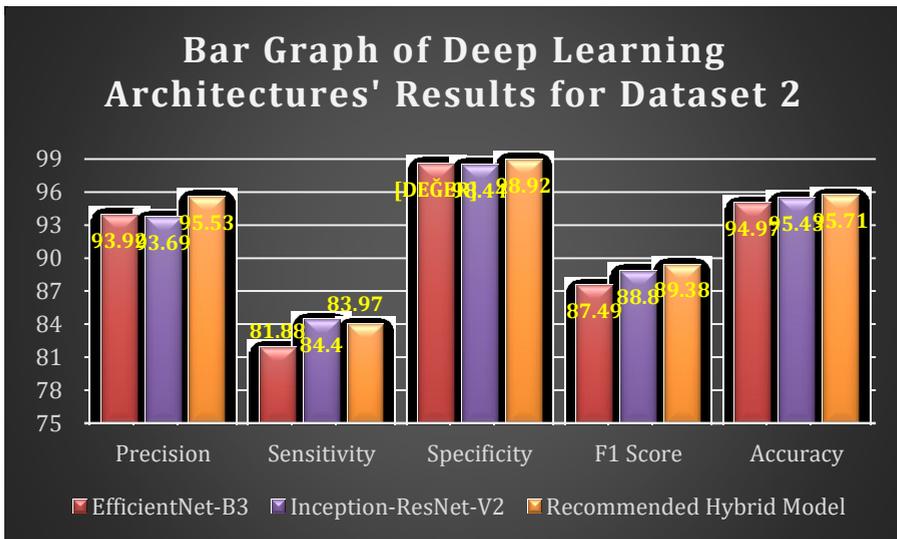


Figure 9. Bar Graph of CNN Architecture Results for Test Dataset 2

The EfficientNet-B3 (Tan & Le, 2019) and Inception-ResNet-V2 (Szegedy et al., 2017) models had similar results (Table 1, Figure 8, and Figure 9). However, the EfficientNet-B3 (Tan & Le, 2019) and Inception-ResNet-V2 (Szegedy et al., 2017) based hybrid model provided the best results with accuracy, F1 score, specificity, sensitivity, and precision of 0.9859, 0.8732, 0.9952, 0.8314, and 0.9119, respectively for Test Dataset 1 and had the best results with accuracy, F1 score, specificity, sensitivity, and

precision of 0.9571, 0.8938,0.9892,0.8397,0.9553, respectively for Test Dataset 2. Sensitivities were lower than the other values, which was expected because there are five hemorrhage subtypes, and an image in the dataset has more than one kind of hemorrhage. The bar graph in Figure 11 could support these results. The lower non-hemorrhage accuracy value in Figure 11 indicates that the sensitivity value is low. Specificities were higher than the other metrics because the sizes of the classes in the dataset were not equal. There were quite small differences between the minimum and maximum values, indicating that the models are consistent. Furthermore, in Test Dataset 2, sensitivity, f1 score and precision increased while accuracy and specificity decreased. No-changes have been made except balanced data distribution in Test Dataset 2. This result shows that this value change is due to the balanced data distribution.

Actual label	0	TN	FP
	1	FN	TP
		0	1
		Predicted label	

Figure 10 Confusion Matrix

Confusion matrixes were created for each class to analyze the class-based performance of the hybrid model, which was proposed in the study. A confusion matrix is a table that reports the ‘Actual’ and ‘Predicted’ class labels (presented in Figure 10). The results of the proposed hybrid model are presented separately for each class in Table 2, and the bar graph showing the accuracy values for each class is illustrated in Figure 11. The confusion matrix values in Table 2 correspond to the average of 10 folds.

Table 2. Hybrid model confusion matrix for Test Dataset 1

Non-Hemorrhage				Epidural			Intraparenchymal				
actual label	0	9401	1392	actual label	0	74940	27	actual label	0	71373	296
	1	710	63776		1	88	226		1	609	3003

		0	1			0	1			0	1
		predicted label				predicted label				predicted label	

Intraventricular				Subarachnoid				Subdural			
actual label	0	72418	241	actual label	0	71270	443	actual label	0	70149	415
	1	371	2250		1	954	2614		1	946	3771
		0	1			0	1			0	1
		predicted label				predicted label				predicted label	

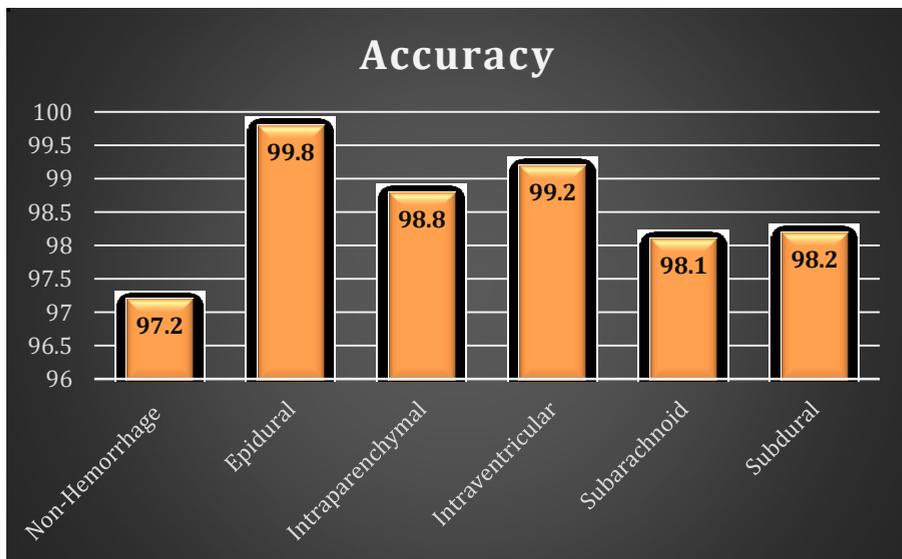


Figure 11. Bar graph of the hybrid model's classification results for Test Dataset 1

The results of Test Dataset1 in Table 2 and Figure 11 show that the hybrid model shows a lower performance in patients without hemorrhage while the hybrid model makes the estimation of patients with epidural hemorrhage close to 100%. This can be interpreted as the number of images without hemorrhage is much higher than the others, while the number of epidural test images is low. If the model created is evaluated in general for Test Dataset1, the model can diagnose non-hemorrhage images with 97.2% accuracy (Figure 11). When we examine the hemorrhage classes' success, epidural has the best accuracy. However, since the number of epidural images is low, it will not be correct to put the epidural performance in this

comparison. The non-epidural hemorrhage classes have a more balanced data distribution. Therefore, when an evaluation is made between them, the accuracy and sensitivity values are descending order intraventricular, intraparenchymal, subdural, and subarachnoid. The fact that all the obtained accuracy value is over 98% is a remarkable indicator of success.

Table 3. Hybrid model confusion matrix for Test Dataset 2

Non-Hemorrhage				Epidural			Intraparenchymal				
actual label	0	2122	249	actual label	0	4555	2	actual label	0	3848	59
	1	28	2472		1	88	226		1	147	817
		0	1			0	1			0	1
		predicted label				predicted label				predicted label	

Intraventricular				Subarachnoid			Subdural				
actual label	0	4030	30	actual label	0	3879	63	actual label	0	3913	65
	1	117	694		1	223	706		1	183	710
		0	1			0	1			0	1
		predicted label				predicted label				predicted label	

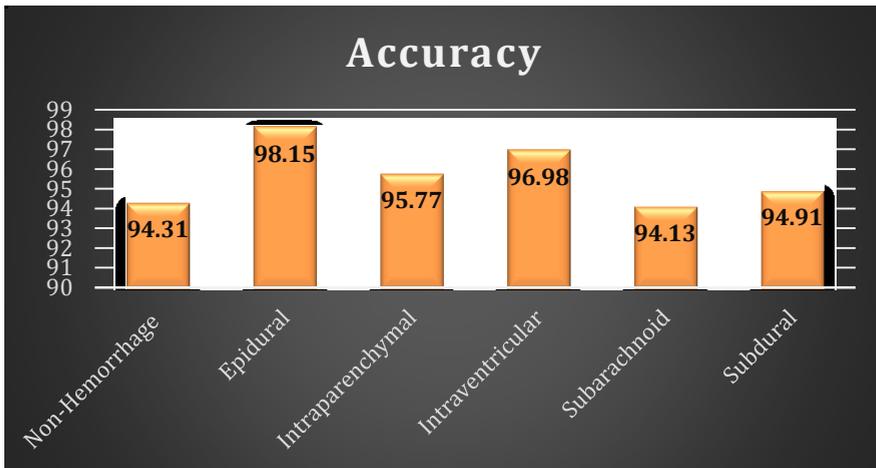


Figure 12. Bar graph of the hybrid model's classification results for Test Dataset 2

The results in Table 3 and Figure 12 show hybrid model performance for Test Dataset 2. As was mentioned above, Test Dataset 1 has imbalanced distribution. This situation prevents objective evaluation. For this reason, the system was tested again by creating a dataset with more balanced distribution. Test Dataset 2 consists of 314 epidural, 500 non-epidural hemorrhage, and 2500 non-hemorrhage images. Test Dataset 2 results show an average %2.8 fewer accuracies. However, when we compare the performance among classes, the same ranking is obtained with Test Dataset1. When we sort according to the highest to the lowest accuracy values, it is again epidural, intraventricular, intraparenchymal, subdural, subarachnoid, and non-hemorrhage. The best estimation is epidural hemorrhage with %98.15 accuracy. On the other hand, testing with a more balanced data set increased sensitivity values.

Table 4. Comparison with the literature

References	Output Size	Sensitivity	Specificity	Recall	Precision	Accuracy	AUC
Prevedello et al. (2017)	Subarachnoid Space Non-Subarachnoid Space	90% sensitivity	85% specificity				AUC of 0.91
Grewal et al. (2018)	Hemorrhage Non-Hemorrhage			88.64% recall	81.25% precision	81.82% accuracy	
Jnawali et al. (2018)	Hemorrhage Non-Hemorrhage	77% sensitivity			80 % precision		AUC of 0.87
Chilamkurthy et al. (2018) The CQ500 dataset	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	94.63% sensitivity	90.21% specificity				AUC of 0.94
Chilamkurthy et al. (2018) Qure25k dataset		90.06% sensitivity	90.04% specificity				AUC of 0.92
Arbabshirani et al. (2018)	Hemorrhage Non-Hemorrhage	73% sensitivity	80% specificity				AUC of 0.85

Ye et al. (2019)	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	98% sensitivity				99 % precision	AUC of 1
Shahangian et al. (2016)	Intracerebral Intraventricular Subdural Extradural Subarachnoid					92.46- 94.13% accuracy	
Chang et al. (2018)	Hemorrhage Non-Hemorrhage	95.1% sensitivity	97.3% specificity			97% accuracy	AUC of 0.98
Lee et al. (2019)	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	92.4-98% sensitivity	94.9-95% specificity				AUC of 0.96- 0.99
Cho et al. (2019)	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	97.91% sensitivity (binary classification)	98.76% specificity (binary classification)	82.15% recall (multi classification)	80.19% precision (multi classification)	98.28% accuracy (binary classification)	

Hybrid Convolutional Neural Network-Based Diagnosis System for Intracranial ...
Sezin BARIN, et al.

Phong et al. (2017)	Hemorrhage Non-Hemorrhage			100% recall	99.3% precision	99.7% accuracy	AUC of 1
Al-Ayyoub et al. (2013)	Non-Hemorrhage Intraparenchymal Subdural Extradural			92.2% recall	92.5% precision		AUC of 0.961
Salehinejad et al. (2021)	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	98.8% sensitivity	98.0% specificity			98.3% accuracy	
Burduja et al. 2020	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	75.6% sensitivity	97.2% specificity			94.7% accuracy	
Proposed Method	Non-Hemorrhage Intraparenchymal Intraventricular Subdural Extradural Subarachnoid	83.1% sensitivity	99.5% specificity		91.2% precision	98.6% accuracy	AUC of 1

In some aspects, this study is superior to previous ones (Table 4). For example, many classification studies use images for the diagnosis of only intracranial hemorrhage. Very few studies focus on diagnosing different subtypes of hemorrhage. Most of these studies have used a different dataset than ours. Therefore, it would not be correct to compare the performance results of all studies directly. However, Salehinejad et al. (2021) and Burduja et al. (2020) used the same data set in the context of their proposed studies. Salehinejad et al. (2021) developed a machine learning model using SEResNeXt-50 and SEResNeXt-101. They achieved 98.3%, 98.8%, and 98.0% for accuracy, sensitivity, specificity, respectively. When the results are compared with the method we propose, it is seen that our accuracy and specificity values are better than theirs. Burduja et al. (2020) developed a hybrid model using SEResNeXt-101 and Bidirectional LSTM. Their created model achieved average performance values of 94.7% accuracy, 75.6% sensitivity, and 97.2% specificity. Class-based performance values in the study were averaged for comparison. When the results are compared with the method we propose, it is seen that our accuracy, sensitivity, and specificity values are better than theirs.

If we evaluate the proposed model in general, we can say that this study's proposed hybrid model is one of the leading values in diagnosing brain hemorrhage.

4. Conclusion

Intracranial hemorrhage is an important public health problem, which is leading to high rates of death all over the world. Therefore, early detection of that health problem is too critical for reducing the mortality rate. Analysis of CT scans plays a crucial role in the diagnosis of intracranial hemorrhage. This study, therefore, proposed a system that uses CT scans to detect intracranial hemorrhage and subtypes of intracranial hemorrhage. The system has a hybrid model consisting of EfficientNet-B3 and Inception-ResNet-V2 and their combination. The proposed hybrid model has an accuracy of 98.5%, which is higher than those of the EfficientNet-B3 and Inception-ResNet-V2. The size of the dataset (752.803 images) was also instrumental in the success of the CNN architectures. The literature comparison also shows that the proposed method has great success.

In sum, the proposed deep learning method is successful and consistent, and the dataset used for its training is large. Therefore, it is a promising emergency diagnostic tool that could help healthcare professionals easily overcome clinical problems. This can be an essential step

in solving the problem of early detection of intracranial hemorrhage and can take the burden of CT reporting away from specialists. On the other, the related positive results have encouraged the authors to think about some future works to see if the study can open new doors. In this context, there will be some works to search for any alternative hybrid model formations improving the obtained results. Also, the current method will be used in alternative datasets to have deeper analyze of the solution. Finally, possibility of including the method in an Internet of Health Things (IoHT) system set-up will be considered in the context of a wider medical project.

Compliance With Ethical Standards

Funding: Funding information is not applicable / No funding was received.

Conflict of Interest: Author Murat Saribas declares that he has no conflict of interest. Author Beyza Guluzar Ciltas declares that she has no conflict of interest. Author Sezin Barin declares that she has no conflict of interest. Author Gur Emre Guraksin declares that he has no conflict of interest. Author Utku Kose declares that he has no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants performed by any of the authors. The used dataset is an open-source material provided in the context of RSNA Intracranial Hemorrhage Detection | Kaggle.

References

- Al-Ayyoub, M., Alawad, D., Al-Darabsah, K., & Aljarrah, I. (2013). Automatic detection and classification of brain hemorrhages. *WSEAS transactions on computers*, 12(10), 395-405.
<http://www.wseas.us/journal/pdf/computers/2013/b065705-365.pdf>
- Anevrizma Yirtılması Ve Beyin Kanamasına Dikkat.* (n.d). Medicine Hospital. Retrieved March 31, 2020, from <https://www.medicinehospital.com.tr/blog/anevrizma-yirtilmasi-ve-beyin-kanamasina-dikkat>
- Arbabshirani, M. R., Fornwalt, B. K., Mongelluzzo, G. J., Suever, J. D., Geise, B. D., Patel, A. A., & Moore, G. J. (2018). Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *Digital Medicine*, 1, 9.
<https://doi.org/10.1038/s41746-017-0015-z>
- Buduruja, M., Ionescu, R. T., Verga, N. (2020). Accurate and Efficient Intracranial Hemorrhage Detection and Subtype Classification in 3D CT Scans with

- Convolutional and Long Short-Term Memory Neural Networks. *Sensors*, 20(19), 5611. <http://dx.doi.org/10.3390/s20195611>
- Chang, P., Kuoy, E., Grinband, J., Weinberg, B., Thompson, M., Homo, R., Chen, J., Abcede, H., Shafie, M., Sugrue, L., Filippi, C. G., Su, M.-Y., Hess, C., & Chow, D. (2018). Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. *American Journal of Neuroradiology*, 39(9), 1609-1616. <https://doi.org/10.3174/ajnr.A5742>
- Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., Mahajan, V., Rao, P., & Warier, P. (2018). Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet*, 392(10162), 2388-2396. https://www.mlGdansk.pl/wp-content/uploads/2019/06/MLGdansk63_27.05.19_PIIS0140673618316453.pdf
- Gardner, M. A., Li, B. C., Wu, Y. W., & Slavotinek, A. M. (2012). Intraparenchymal hemorrhage in a neonate with cleidocranial dysostosis. *Pediatric Neurology*, 47(6), 455-457. <https://doi.org/10.1016/j.pediatrneurol.2012.08.009>
- Grewal, M., Srivastava, M. M., Kumar, P., & Varadarajan, S. (2018, April). Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 281-284). IEEE. <https://arxiv.org/pdf/1710.04934.pdf>
- Helwan, A., El-Fakhri, G., Sasani, H., & Uzun Ozsahin, D. (2018). Deep networks in identifying CT brain hemorrhage. *Journal of Intelligent & Fuzzy Systems*, 35(2), 2215-2228. <http://doi.org/10.3233/jifs-172261>
- ImageNet*. (n.d.). Retrieved March 31, 2020, from <http://www.image-net.org/>
- Jnawali, K., Arbabshirani, M. R., Rao, N., & Patel, A. A. (2018, February). Deep 3D convolution neural network for CT brain hemorrhage classification. In *Medical Imaging 2018: Computer-Aided Diagnosis*, 10575, (p. 105751C). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2293725>
- Jauch, E. C., Saver, J. L., Adams, H. P., Bruno, A., Connors, J. J. B., Demaerschalk, B. M., Khatri, P., McMullan, P. W., Qureshi, A. I., Rosenfield, K., Scott, P. A., Summers, D. R., Wang, D. Z., Wintermark, M., & Yonas, H. (2013). Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 44(3), 870-947. <https://doi.org/10.1161/STR.0b013e318284056a>
- Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S. H., Guerrier, C. E., Ebert, S.A., Pomerantz, S. R., Romero, J. M., Kamalian, S., Gonzalez, R. G., Lev, M. H., & Do, S. (2019). An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature*

- Biomedical Engineering*, 3, 173-182. <https://doi.org/10.1038/s41551-018-0324-9>
- Majumdar, A., Brattain, L., Telfer, B., Farris, C., & Scalera, J. (2018). Detecting Intracranial Hemorrhage with Deep Learning. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 583-587. <http://doi.org/10.1109/embc.2018.8512336>
- Muir, K. W., Buchan, A., von Kummer, R., Rother, J., & Baron, J. C. (2006). Imaging of acute stroke. *Lancet Neurology*, 5(9), 755-768. [https://doi.org/10.1016/S1474-4422\(06\)70545-2](https://doi.org/10.1016/S1474-4422(06)70545-2)
- Phong, T. D., Trong, N. T., Duong, H. N., Nguyen, V. H., Snasel, V., Nguyen, H. T., & Van Hoa, T. (2017). Brain hemorrhage diagnosis by using deep learning. *ACM International Conference Proceeding Series*, 34-39. <https://doi.org/10.1145/3036290.3036326>
- Prevedello, L. M., Erdal, B. S., Ryu, J. L., Little, K. J., Demirel, M., Qian, S., & White, R. D. (2017). Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology*, 285(3), 923-931. <https://pubs.rsna.org/doi/pdf/10.1148/radiol.2017162664>
- RSNA Intracranial Hemorrhage Detection*. (n.d.). Retrieved March 31, 2020, from <https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection>
- Salehinejad, H., Kitamura, J., Ditkofsky, N., Lin, A., Bharatha, A., Suthiphosuwat, S., Lin, H.-M., Wilson, J. R., Mamdani, M., & Colak, E. (2021). A Real-World Demonstration of Machine Learning Generalizability: Intracranial Hemorrhage Detection on Head CT, 1-15. <https://arxiv.org/ftp/arxiv/papers/2102/2102.04869.pdf>
- Şeker, A., Diri, B., & Balik, H. (2017). Derin Öğrenme Yöntemleri Ve Uygulamaları Hakkında Bir İnceleme [A Review of Deep Learning Methods and Applications]. *Gazi Mühendislik Bilimleri Dergisi (GMBD)*, 3(3), 47-64. <https://dergipark.org.tr/tr/download/article-file/394923>
- Shahangian, B., & Pourghassem, H. (2016). Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure. *Biocybernetics and Biomedical Engineering*, 36(1), 217-232. <http://doi.org/10.1016/j.bbe.2015.12.001>
- Stanford.edu. (2018). *Convolutional Neural Networks for Visual Recognition* [class course]. <http://cs231n.stanford.edu/>
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, 4278-4284.

<https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806/14311>

- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *36th International Conference on Machine Learning, ICML 2019, 2019-June*, 10691-10700. <http://proceedings.mlr.press/v97/tan19a/tan19a.pdf>
- Web of Science [v.5.34] - Web of Science Core Collection Results.* (n.d.). Retrieved March 31, 2020, from <http://proxy.afyon.deep-knowledge.net/MuseSessionID=0210h3diw/MuseProtocol=http/MuseHost=apps.webofknowledge.com>
- Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., Wang, X., Bai, J., Cao, K., Song, Q., Zhang, H., Chen, W., Guo, X., & Xia, J. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *European Radiology*, *29*(11), 6191-6201. <https://doi.org/10.1007/s00330-019-06163-2>