PROCEEDINGS OF
INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND NEW APPLICATIONS
https://proceedings.icisna.org/
2nd International Conference on Intelligent Systems and New Applications (ICISNA'24), Liverpool, April 26-28, 2024.

A Novel Hybrid Approach: Capsule Network Enhanced with Residual Block and CBAM Module for Medical Image Classification

Sumeyra Busra SENGUL[1], Ilker Ali OZKAN[2]

[1]Department of Computer Engineering, Selcuk University, Konya, Turkey
sumeyra.sengul@selcuk.edu.tr, ORCID: 0000-0003-1385-0920

[2] Department of Computer Engineering, Selcuk University, Konya, Turkey
ilkerozkan@selcuk.edu.tr, ORCID: 0000-0002-5715-1040

***Abstract*— Convolutional neural networks (CNN) have been successfully used in the classification of medical images. However, CNNs extract the features of the image, neglecting the spatial information between the objects since they evaluate each object separately and this makes it difficult to analyse images reliably. CapsNets partially overcomes these limitations by encapsulating the objects in the image and their feature parameters without losing the spatial information between objects. However, this advantage of CapsNet can become a disadvantage due to the fact that CapsNet stores a lot of information, which can lead to an increase in the number of parameters and the complexity of the network. In this study, a new method is proposed in which residual block and CBAM (Convolutional Block Attention Module) modules are used hybridly with CapsNet in order to classify medical images with high accuracy and reliability and, at the same time, to overcome this disadvantage of CapsNet by reducing the number of parameters. In this study, more features of the image are extracted by using the residual block instead of the convolutional layer of CapsNet. The attention mechanism is used to increase the effectiveness of important features and reduce the effect of unnecessary features. In addition, the residual block and CBAM module reduced the number of parameters in CapsNet by approximately 2 million. Brain tumour images and OCT (Optical Coherence Tomography) images were classified with the proposed model. The results of the study show that the proposed model achieves significant results when compared to other CapsNet-based studies in the literature.**

***Keywords*— CapsNet, CBAM, Classification, Medical Images, Residual Block**

I. INTRODUCTION

Medical imaging is an important resource in disease diagnosis. Imaging techniques such as Computed Tomography, X-Ray, MRI (Magnetic Resonance Imaging), OCT are used for the automatic and rapid diagnosis of diseases. In order to analyse these images with the naked eye, a trained expert's perspective is needed.

Deep learning methods fulfil this need for trained experts by analysing images comprehensively in a shorter time, while also noticing details that may be overlooked. The main application areas of medical image analysis are segmentation, classification, and abnormality detection [1]. Convolutional neural networks (CNN), a deep learning technique, are widely employed in the classification of medical pictures [2-4].

CNN needs a large amount of data when classifying medical data. In cases where there is not enough data, it encounters the problem of overfitting. In addition, the pooling layer in CNN causes the spatial information of the image to be lost, and the performance decreases [5]. In order to solve these limitations of CNN, Sabour et al. [6], proposed the CapsNet architecture. In CapsNet, a dynamic routing algorithm is used instead of the pooling layer, which causes the loss of spatial information in CNN, and capsules with vector output are used instead of scalar output feature maps. These capsules consist of neuron groups. In the neuron groups, the features of the image and the properties of these features are stored [7]. These properties include direction, size, and color information, as well as the spatial position of the feature. Especially in medical images where location is important, this characteristic of CapsNet provides a significant advantage. Since every detail of the image is taken into account, CapsNet can perform better with less data [8], [9]. However, although it is an advantage for CapsNet to evaluate every detail of the image for a reliable analysis of the medical image, this may become a disadvantage in complex images. The high number of details to be evaluated in the image may cause the number of parameters to increase with the increase in the number of operations performed in CapsNet. In addition, CapsNet needs more convolution layers to extract more features; in other words, the network needs to be deepened. When the studies in the literature are examined, this is usually achieved by using more layers or hybrid models instead of the single convolution layer of CapsNet[10-13].

In this study, a new hybrid model is proposed to classify medical images with high accuracy and reliability. In the proposed model, the residual block is used instead of the convolution layer of CapsNet to deepen the network and extract more features of the image. The CBAM attention mechanism is used to increase the efficiency of important features, eliminate unnecessary features, and reduce the number of parameters [14]. Thanks to the proposed model, the number of parameters is reduced by approximately 2 million compared to CapsNet. Brain MRI images and retinal OCT images were used to evaluate the performance of the proposed model.

In the second section, the datasets used in the study, the details of the CapsNet architecture, the modules used, and the proposed model are given. In Section 3, the results of the study are presented, and in the last section, an overall evaluation of the study is presented.

II. Material and Methods

A. Dataset Details

1) Brain Tumour Dataset

The dataset includes 2D T1-weighted contrast images. The dataset consists of images from a variety of angles: axial, coronal, and sagittal. It consists of 4 classes[15]. These classes are glioma, meningioma, no tumor, and pituitary. The images were scaled to 50x50 dimensions. Table 1 shows the number of train data and test data for the classes. The dataset was downloaded Kaggle.com [16].

TABLE I
Brain Tumor Dataset Image Numbers

Classes	Train Set	Test Set	Total
Glioma	712	214	926
Meningioma	755	182	937
No Tumor	416	84	500
Pituitary	728	173	901

2) OCT Dataset

The OCT dataset contains a total of 84,495 jpeg photos divided into four categories: Normal, CNV, DME, and DRUSEN [17]. The images were scaled to 50x50 dimensions. The train, test and validation numbers of the dataset are given in Table 2. Dataset was downloaded Kaggle.com [18].

TABLE II
OCT Dataset Image Numbers

Classes	Train Set	Test Set	Validation Set	Total
CNV	8,616	242	242	9,100
DME	8,616	242	242	9,100
DRUSEN	8,616	242	242	9,100
NORMAL	8,712	242	242	9,196
TOTAL	34,560	968	968	36,496

B. CapsNet

CapsNet is an architecture proposed in 2017 by Sabour et al. [6]. The most basic unit is the capsule which consists of neurons. The capsule stores the properties of the image such as pose (position, direction and size), spatial position, color, and texture. CapsNet architecture basically consists of 3 layers. The architecture is shown in Figure 1. The first layer is the convolution layer. In this layer, the basic features of the image are extracted. The second layer is the primary capsule layer. In this layer, a more detailed feature map is achieved by applying convolution process to the inputs from the previous layer. The feature map is passed through the squash activation function to obtain feature vectors (Equation 1). This function limits the length of the incoming input between 0 and 1 and outputs it as a unit vector.

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \tag{1}$$

Following this layer, the output is routed via the dynamic routing method to the digit capsule layer (Algorithm 1). The dynamic routing algorithm is the algorithm that provides the relationship between the lower-level capsules and the upper-level capsules.

Algorithm1. Dynamic Routing Algorithm [6]

procedure ROUTING ($\hat{u}_{j|i}$, r,l)
for all capsules i in layer l and capsule j in layer (l+1): $b_{ij} \leftarrow 0$.
for r iterations **do**
for all capsule i in layer l: $c_i \leftarrow$ **softmax** (b_i)
for all capsule j in layer (l+1): $s_j \leftarrow \sum_i c_{ij}\hat{u}_{j|i}$
for all capsule j in layer (l+1): $v_i \leftarrow$ **squash** (s_j)
for all capsule i in layer l and capsule j in layer (l+1): $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i}v_{ij}$
return v_j

After this stage, the image is reconstructed in the decoder layer and the loss is calculated. For the loss operation, the margin loss function is used (Equation 2). In the equation, if there are k classes, Tk=1, otherwise Tk=0. The lambda value is taken as 0.5, m^{+}= 0,9, m^{-} = 0,1. Total loss is found by adding the losses of the capsules.

$$L_k = T_k max(0, m^{+} - \|v_k\|)^2 + \lambda(1 - T_k) max(0, \|v_k\| - m^{-})^2 \tag{2}$$

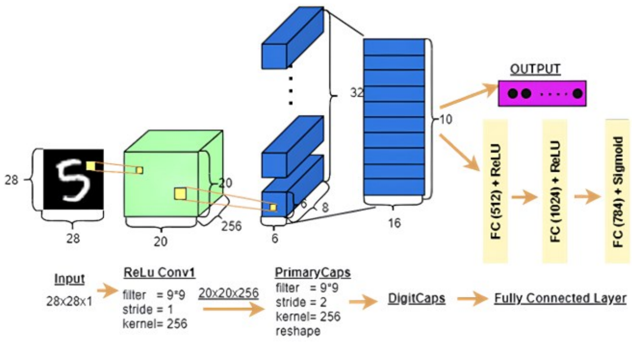

Fig.1. CapsNet Architecture [7]

C. Residual Block

Residual Network (ResNet) is an architecture proposed by He et al. [19]. The aim of the ResNet architecture is to avoid the problem of vanishing and exploding gradient in deep network architectures. The architecture achieves this with residual blocks. The residual block unit used in the study is shown in Figure 2.

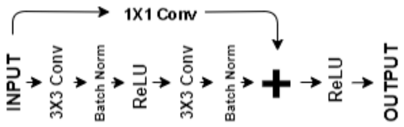

Fig.2. Residual Block

D. CBAM

CBAM has two submodules. They are the Channel Attention Module (CAM) and the Spatial Attention Module (SAM) [14]. The CBAM module generates refined feature maps by highlighting important features from the input feature maps in deep networks. After extracting the cross-channel information of the input features, the module extracts the spatial information with average and maximum pooling. Thus, it provides an efficient information flow within the network by answering the questions of which information should be highlighted or the impact of which information should be minimised. The CBAM module is shown in Figure 3. The sub-modules CAM and SAM of the module are shown in Figure 4.

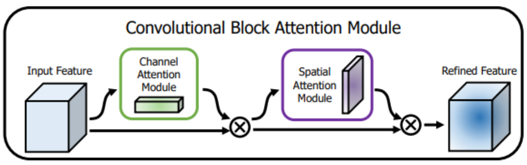

Fig 3. CBAM Module[14]

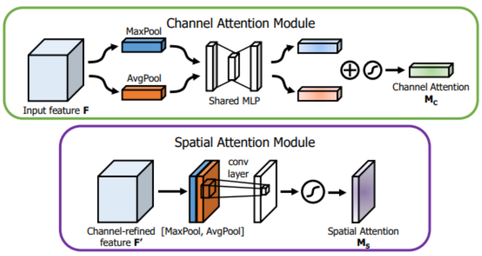

Fig 4. Sub-modules of the CBAM module [14]

E. Proposed Method

The proposed method consists of a residual block layer, a CBAM layer, a primary capsule layer and a digit capsule layer. The residual block and CBAM module are used instead of the convolution layer, which is the first layer in the classic CapsNet architecture, to achieve a feature map including detailed and significant image features while also reducing the amount of architecture parameters. The architecture of the proposed method is shown in Figure 5.

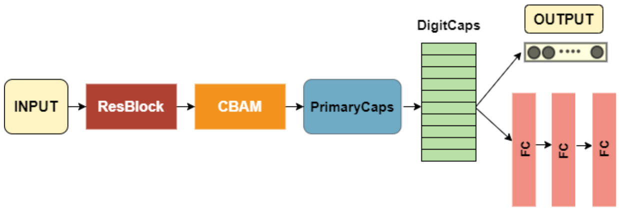

Fig 5. Architecture of the Proposed Method

III. Experimental Results

In this study, the models were developed using Python 3.8 programming language and Tensorflow 2.3. libraries. The device used has Intel Xeon Gold 6226R processor and Nvidia Grid RTX8000-12Q graphics card. In the classification of the proposed modelled brain tumour images, batch size was set to 8, learning rate was set to 1 x 10^(-4), number of epochs was set to 50 and Adam Optimizer was used. For the classification of OCT images, the batch size is set to 64 since the number of data is more.

The ACC (accuracy) metric and parameter numbers of the models were used to evaluate the performance of the proposed model. The ACC is calculated by dividing the number of accurately predicted data by the total amount of data (Equation 3). ACC value is calculated with the true positive (TP), true negative (TN), false positive (FP), false negative (FN) values given in the complexity matrix. The complexity matrix is shown in Table 3.

TABLE III
Confusion Matrix

	Predicted	
Actual	True Positives (TP)	False Negatives (FN)
	False Positives (FP)	True Negatives (TN)

$$ACC = (TP + TN / TP + FP + FN + TN) *100 \quad (3)$$

A. Results of Brain Tumour Dataset

In order to compare the performance of the proposed model, CapsNet-based models in the literature where brain tumour images are classified and the original CapsNet were used. While classifying brain tumour images, an accuracy of 78.10% was obtained with the original CapsNet architecture. With the proposed model, this accuracy rate increased by approximately 12% and reached 90.50%.

TABLE IV
BRAIN TUMOUR DATASET LITERATÜR KARŞILAŞTIRMALI SONUÇLAR

Network Model	ACC (%)	Numbers of Parameter
Original CapsNet	78.10	13,185,732
Afshar ve ark. [20]	86.56	-
Goceri [21]	92.65	-
Proposed Method	90.50	11,020,487

Table 4 also shows the parameter numbers of the CapsNet architecture and the proposed model. The original CapsNet has 13,185,732 parameters, whereas the new proposed model has 11,020,487 parameters.

B. Results of OCT Dataset

The OCT dataset was classified with the original CapsNet architecture and the proposed model. In order to evaluate the success of the proposed model on OCT images, CapsNet-based models in the literature and the original CapsNet were used. The results of these models are given in Table 5. When the table is analysed, while 88.73% accuracy is obtained with the original CapsNet, 96.28% accuracy is obtained with the proposed model. Furthermore, the original CapsNet model has 13,185,732 parameters, whereas the suggested model has 11,020,487.

TABLE V
OCT DATASET LITERATÜR KARŞILAŞTIRMALI SONUÇLAR

Network Model	ACC (%)	Numbers of Parameter
Original CapsNet	88.73	13,185,732
CLAHE-CapsNet [22]	97.7	-
SFFT-CapsNet [23]	99.0	-
Proposed Method	96.28	11,020,487

IV. CONCLUSIONS

In this study, a hybrid CapsNet model using residual block and CBAM module is proposed for the classification of medical images. The results show that the residual block and CBAM module increase the success rate and reduce the number of parameters. In addition, a performance very close to the studies in the literature has been obtained.

Residual block enhanced the performance of CapsNet by improving the training of the network. The CBAM module, on the other hand, not only enhanced the performance of CapsNet by focusing on important features in the image, but also reduced the number of parameters by eliminating non-essential features. For CapsNet, which takes into consideration every detail in complex images, it is very important that important features are selected in the feature map coming to the primary capsule layer. Thus, fewer features will be selected, but more necessary features will be selected for the analysis and the processing load of CapsNet will be reduced.

This study will contribute to the use of CapsNet not only in medical image analysis but also in other fields. The proposed model can be used with different numbers of layers or modules. In addition, attention modules can be used in the capsule layers to make the capsules more effective.

ACKNOWLEDGMENT

We would like to thank the Scientific Research Coordinatorship of Selcuk University for their support with the project titled "Development of A New Capsule Network-Based Deep Learning Architecture For Analysing Medical Images" numbered 22111010.

REFERENCES

[1] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, ve M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review", *J Med Syst*, c. 42, sy 11, s. 226, Eki. 2018, doi: 10.1007/s10916-018-1088-1.

[2] M. H. Rezazade Mehrizi, P. van Ooijen, ve M. Homan, "Applications of artificial intelligence (AI) in diagnostic radiology: a technography study", *Eur Radiol*, c. 31, sy 4, ss. 1805-1811, Nis. 2021, doi: 10.1007/s00330-020-07230-9.

[3] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, ve H. J. W. L. Aerts, "Artificial intelligence in radiology", *Nat Rev Cancer*, c. 18, sy 8, ss. 500-510, Ağu. 2018, doi: 10.1038/s41568-018-0016-5.

[4] C. Martin-Isla *vd.*, "Image-Based Cardiac Diagnosis With Machine Learning: A Review", *Front Cardiovasc Med*, c. 7, s. 1, Oca. 2020, doi: 10.3389/fcvm.2020.00001.

[5] S. B. Şengül ve İ. A. Özkan, "New Architecture in Deep Learning: Capsule Networks (CapsNet)", içinde *International Research in Engineering Sciences III*, Egitim Publishing, 2022, ss. 245-262.

[6] S. Sabour, N. Frosst, ve G. E. Hinton, "Dynamic routing between capsules", içinde *Proceedings of the 31st International Conference on Neural Information Processing Systems*, içinde NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Ara. 2017, ss. 3859-3869.

[7] S. B. Şengül ve İ. A. Ozkan, "Performance Evaluation of Basic Capsule Network Architecture in Classification of Biomedical Images", *GJES*, c. 9, sy 2, Art. sy 2, Ağu. 2023.

[8] J. Rajasegaran, V. Jayasundara, S. Jayasekara, H. Jayasekara, S. Seneviratne, ve R. Rodrigo, "DeepCaps: Going Deeper with Capsule Networks". arXiv, 21 Nisan 2019. doi: 10.48550/arXiv.1904.09546.

[9] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, ve B. Y. Edward, "Capsule Networks – A survey", *Journal of King Saud University - Computer and Information Sciences*, c. 34, sy 1, ss. 1295-1310, Oca. 2022, doi: 10.1016/j.jksuci.2019.09.014.

[10] Y. Xiong, G. Su, S. Ye, Y. Sun, ve Y. Sun, "Deeper Capsule Network For Complex Data", içinde *2019 International Joint Conference on Neural Networks (IJCNN)*, Tem. 2019, ss. 1-8. doi: 10.1109/IJCNN.2019.8852020.

[11] S. Chang ve J. Liu, "Multi-Lane Capsule Network for Classifying Images With Complex Background", *IEEE Access*, c. 8, ss. 79876-79886, 2020, doi: 10.1109/ACCESS.2020.2990700.

[12] X. Ai, J. Zhuang, Y. Wang, P. Wan, ve Y. Fu, "ResCaps: an improved capsule network and its application in ultrasonic image classification of thyroid papillary carcinoma", *Complex Intell. Syst.*, c. 8, sy 3, ss. 1865-1873, Haz. 2022, doi: 10.1007/s40747-021-00347-4.

[13] J. Ni, Z. Yan, ve J. Jiang, "TongueCaps: An Improved Capsule Network Model for Multi-Classification of Tongue Color", *Diagnostics (Basel)*, c. 12, sy 3, s. 653, Mar. 2022, doi: 10.3390/diagnostics12030653.

[14] S. Woo, J. Park, J.-Y. Lee, ve I. S. Kweon, "CBAM: Convolutional Block Attention Module", içinde *Computer Vision – ECCV 2018*, c. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, ve Y. Weiss, Ed., içinde Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, ss. 3-19. doi: 10.1007/978-3-030-01234-2_1.

[15] A. Kadam, "Brain Tumor Classification using Deep Learning Algorithms", *IJRASET*, c. 9, sy 12, ss. 417-426, Ara. 2021, doi: 10.22214/ijraset.2021.39280.

[16] "Brain Tumor Classification (MRI)". Erişim: 23 Nisan 2024. [Çevrimiçi]. Erişim adresi: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

[17] D. S. Kermany *vd.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning”, *Cell*, c. 172, sy 5, ss. 1122-1131.e9, Şub. 2018, doi: 10.1016/j.cell.2018.02.010.
[18] “Retinal OCT Images (optical coherence tomography)”. Erişim: 23 Nisan 2024. [Çevrimiçi]. Erişim adresi: https://www.kaggle.com/datasets/paultimothymooney/kermany2018
[19] K. He, X. Zhang, S. Ren, ve J. Sun, “Deep Residual Learning for Image Recognition”, içinde *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Haz. 2016, ss. 770-778. doi: 10.1109/CVPR.2016.90.
[20] P. Afshar, A. Mohammadi, ve K. N. Plataniotis, “Brain Tumor Type Classification via Capsule Networks”, içinde *2018 25th IEEE International Conference on Image Processing (ICIP)*, Eki. 2018, ss. 3129-3133. doi: 10.1109/ICIP.2018.8451379.
[21] E. Goceri, “CapsNet topology to classify tumours from brain images and comparative evaluation”, *IET Image Processing*, c. 14, sy 5, ss. 882-889, 2020, doi: 10.1049/iet-ipr.2019.0312.
[22] M. Opoku, B. A. Weyori, A. F. Adekoya, ve K. Adu, “CLAHE-CapsNet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization”, *PLoS One*, c. 18, sy 11, s. e0288663, Kas. 2023, doi: 10.1371/journal.pone.0288663.
[23] M. Opoku, B. A. Weyori, A. F. Adekoya, ve K. Adu, “SFFT-CapsNet: Stacked Fast Fourier Transform for Retina Optical Coherence Tomography Image Classification using Capsule Network”, *IJACSA*, c. 14, sy 9, 2023, doi: 10.14569/IJACSA.2023.0140932.