ISSN 2664-4150 (Print) & ISSN 2664-794X (Online)
South Asian Research Journal of Engineering and Technology
Abbreviated Key Title: South Asian Res J Eng Tech

| Volume-8 | Issue-5 | Nov-Dec- 2026 | **DOI: https://doi.org/10.36346/sarjet.2026.v08i05.003**

Original Research Article

Capsule-Based Brain Tumor Classification with Residual and Pyramid Spatial Features

Mustafa Challoob Albazoone[1]*

[1]Technology and Science Incubator Department, University of Misan

***Corresponding Author:** Mustafa Challoob Albazoone
Technology and Science Incubator Department, University of Misan

Article History
Received: 18.09.2026
Accepted: 10.11.2026
Published: 20.11.2026

Abstract: Accurate classification of brain tumors from Magnetic Resonance Imaging (MRI) may help with the clinical evaluation process. However, many automated classification systems have demonstrated sensitivity to variation in tumor appearance, spatial relationships and the distribution of the training data. This study evaluates the classification performance of capsules in three different configurations: the baseline HitNet configuration, a Res-HitNet configuration with residual connections, and a PSP-HitNet configuration with hierarchical spatial analysis. All of these configurations use a Hit-or-Miss (HoM) capsule layer that maps each class-specific capsule vector to the extracted image features. The Res-HitNet configuration utilizes shortcut connections that allow for integration of features throughout the network. Processing stages, while PSP-HitNet incorporates multi-scale branches with different spatial filter sizes to represent patterns at more than one scale. The best-performing architecture is then evaluated in an uncertainty-guided deep active-learning setting, in which low-confidence unlabeled examples are selected for annotation and added incrementally to the labeled training set. The paper reports experiments on 3,264 MRI images carried out using a Kaggle dataset. The dataset characterization lists 926 glioma images, 937 meningioma images, 901 pituitary-tumor images, and 500 images without a tumor. Five-fold cross-validation is reported for the main model comparisons. The paper reports a best overall accuracy of 95.98% for the proposed approach and describes faster improvement from uncertainty-based sample selection than from random selection during early active-learning rounds. Additionally, this study demonstrated that the decision verification performance varied from approximately 89.3% to 93% when the capsule-based classification system underwent a longer period of reinforcement learning. The findings suggest that capsule-based classification, along with multi-scale spatial processing, has potential as an effective method for evaluating the performance of capsule-based systems.

Keywords: Brain Tumor Classification, Magnetic Resonance Imaging, Capsule Networks, Hit-Or-Miss Capsules, Residual Connections, Pyramid Spatial Pooling, Active Learning, Uncertainty Sampling.

1. INTRODUCTION

Brain tumors come in many forms. Their appearance changes with each tumor's type, where it is located, how large it is and what types of tissues surround it. Because MRIs can create images of soft tissues within the skull, they are frequently used to study the structure of the brain. However, reviewing these images is time-consuming and demands specialized expertise [1]. For this reason, interest in computer-aided image analysis techniques has begun to grow, as a means of supporting clinical decisions, not replacing them [2].

Deep learning has improved image classification by learning multi-layer representations directly from image data. CNN's have proven to be extremely successful at identifying specific types of visual information including edges, textures and component parts of shapes [3]. However, traditional convolutions and grouping (pooling) processes may result in loss of fine detail, and the classifier may learn visual patterns that are not suitable for application outside of its training data. This issue is of great importance in the field of medical imaging, as tumors can vary greatly in size, location, contrast, and shape.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

The idea of capsule networks (Capsule networks) was introduced in part with the aim of representing entities using vectors, rather than relying on a single standard value (scalar) for activation [4]. In this type of representation, the vector can capture multiple features of the detected pattern. The idea is based on the fact that relationships between parts may help distinguish between visually similar categories [5]. This research is investigating this idea using a capsule layer based on the Hit-or-Miss (HoM) algorithm, and has developed three different versions of the same model which have been designed to categorize brain MRI images [6].

The initial model, Base-HitNet, offers the foundational architecture. The Res-HitNet framework incorporates residual connection links (residual shortcuts) to facilitate the transmission of information across the convolutional layers (convolutional stages). The PSP-HitNet model includes parallel branches with different spatial scales, inspired by hierarchical scene analysis techniques (pyramid-based scene parsing) [7]. Furthermore, the study combines the PSP-HitNet model with deep active learning (deep active learning), instead of classifying each available image before starting training, the uncertainty measurement-based query strategy (uncertainty-driven query strategy) prioritizes examples for which the current model shows the least confidence [8].

This work addresses two practical questions. First, do residual and hierarchical feature processing modifications improve the performance of the HoM capsule classifier on the selected MRI dataset? Second, can uncertainty-based sample selection improve verification performance faster than random selection when only part of the training data is classified? The message indicates that PSP-HitNet performed best among the three architectures, and that uncertainty-based sampling improved learning overall faster during the first query rounds [8, 9].

1.1 Contributions

This study makes four main contributions:

1. Apply a hit or miss capsule-based classification methodology to brain MRI images and three categories of tumors.
2. Compare the basic HitNet model with the variables used for both hierarchical spatial analysis and residual analysis.
3. Evaluate each class for its classification performance in terms of accuracy, recall and F1-score. Additionally evaluate verification performance.
4. Use active learning guided by uncertainty to select the most relevant samples for classification from a large pool of unlabeled data.

1.2 Scope and Interpretation

The experiment was to classify MRI images as a single image. While the study indicated that spatial relationships could be clinically relevant and that the pyramid unit (pyramid module) may have potential application for localized identification, it did not demonstrate an accurate and reliable method for segmenting or localizing tumors. Additionally, the accuracy rates evaluated do not by themselves indicate that they will provide clinically relevant information, be applicable to multiple hospital settings, or result in improved patient care.

2. Background and Related Work

2.1 CNN-Based Image Classification

The CNN uses a set of learned filters that are applied to a series of localized image areas to create map features [10]. In general, early filter stages will react to relatively simple visual structures, and later stages will synthesize those reactions into more complex abstractions. To decrease the size of spatial dimensions and the amount of computation required, either pooling or stride convolutions may be utilized. Finally, a classification layer takes the developed representation and assigns it class scores [11].

The advantages of using CNNs include sharing parameters, integrating local perspectives, and learning feature representation based on labeled data. On the other hand, there are disadvantages that vary with architecture and training conditions[3]. For instance, the use of repeated down sampling can eliminate information regarding small structures within an image, further, a CNN's performance may be affected by the manner in which images were acquired and preprocessed, the degree of class imbalance in the dataset, as well as the size and composition of the dataset[12]. Therefore, good performance on a given dataset needs to be supported by appropriate splitting and external validation. Anwar *et al*., [13]. examined the use of convolutional neural networks (CNNs) to detect and scope brain tumors and demonstrated the models were strong ability to classify images[14]. However, the study also highlighted challenges such as attributing the loss during the reduction process and the need for more efficient attribute representations. Kumar *et al*., [15], use a convolutional neural network that classifies brain tumors using two public datasets, where the methodology employs a 25-layer CNN model using T1-weighted Magnetic Resonance Imaging (MRI) images to evaluate the method's performance against previously published approaches in the field. The proposed algorithm has produced impressive results in the classification and segmentation of MRI brain images[16].

2.2 Residual Connections

Residual networks create direct connections between layers of a neural network so that each layer can learn a residual mapping in addition to its own input. A common version of this is when the output of a layer is represented as: **y = F(x) + x, with F** representing the operation performed by the layer.

If there are different sizes for the two parts, then a projection can be used to make sure they match up. This allows for better optimization in deeper neural networks and provides ways for information to flow around one or more operations[17].

In the Res-HitNet model, the shorthand link (shortcut connection) is used to combine both low-level and high-level features prior to entering the HoM capsule layer[18]. The reason behind using the shorthand link is that it can possibly help the classification process with both local patterns and larger context. The results from the ablation study indicate that when we remove each subsequent shortcut link from the Res-HitNet architecture, the performance of the model on the verification set declines by approximately 2% in certain configurations. Unfortunately, the source did not give us enough numerical detail to reproduce all of the ablation studies. S. Hong *et al.*, [19]. An improved algorithm based on vision transformation technology to classify human brain tumors. To overcome the limitations of current small datasets, symmetric filtering techniques, adaptive histogram equalization limited by channel variance, and blur masking techniques were applied to enrich the dataset images, enhancing the information and improving the model's generalization ability

2.3 Capsule Representations and Hit-or-Miss Capsules

Due to the recent emergence of capsule networks (CapsNets), good results on large datasets such as ImageNet [20] and Pascal VOC2012 have not yet emerged in research circles[21]. However, some researchers working in medical imaging have adapted capsule networks or created their own versions of them. There are two main cases in which capsules have begun to be adapted to new problems other than recognizing handwritten numbers. In addition, there is a network that developed the idea of capsules to add a new type of layer to each type of classification network.

Afshar and colleagues[22], applied a simple and direct modification of the capsule network (CapsNet) in classifying brain tumor types between meningiomas, gliomas, and pituitary tumors. Minor modifications have been made to the original capsule network to make it more user-friendly. In general, the structure used was practically identical to the capsule network presented by Sabour and colleagues [23]. Researchers have shown that the new architecture promises increased accuracy in medical imaging problems. In the field of object segmentation, Lalonde and Bagsey[24], introduced the concept of disassembly capsules. [25]. Combining this with the jump links he and his colleagues implemented in ResNet , Lalonde and Bagsey[24], were able to create a CapsNet that is at least four times deeper than the original CapsNet. Unlike bypass capsules, which may reduce the image size from one capsule layer to another, disassembly capsules sample images, eventually returning outputs the same size as the input images using shortened links. These images contain information that enables CapsNet to successfully classify and separate diseased lungs from CT images[26]. In addition, the idea of locally limited capsules was introduced, allowing CapsNet to receive input images with dimensions of 512 × 512 while maintaining a few trainable parameters.

2.4 Pyramid Spatial Processing

The PSP-HitNet employs four branches with filter size 1x1, 2x2, 3x3 and 6x6. Each of these branches is intended to identify image features at varying levels of spatial scale[27]. Once each of the image features have been identified, they are then combined before they are used as input into a classification system. Filters that are smaller will enable the detection of finer details, whereas, filters that are larger will provide the opportunity to integrate information over a larger area[28].

A pyramid architecture for classification was used as opposed to a full segmentation pipeline[19]. Each of the four branches of the pyramid will be part of the feature encoder and their respective outputs will be combined into a single representation that will then be passed to the HoM capsule layer. The use of multi-scale design should enable the model to identify tumors by both size and appearance[2].

2.5 Active Learning for Medical Image Classification

Although many approaches to solving deep learning problems involve creating new structures, active learning is another way to improve the model by allowing it to choose the data it wants to train on[29]. The basic idea is that a machine learning algorithm can achieve greater accuracy when it is allowed to choose which data it learns from[30]. A comprehensive review of active learning can be found in Settles. Cheeta *et al*., [31], proposed deep probability sets (DPEs) that use regular sets to approximate the Bayesian neural network. Although it is one of many active learning approaches, these have been shown to enable models to perform competitively with less training data required. In practice, Algorithm 2 shows that after the initial training cycle, the system adds subcategories to the classified data set and trains the model set on these subcategories. This process is repeated until we arrive at B, a total of examples added to the set. The algorithm is implemented for each model in the set, where:

$U = \{X_u^j\}_{j=1}^{Ni}$

Where: A large unclassified set consists of Nu sample, where each xj ∈ X represents an input data point.

$L = \{(X_u^j, Y_l^j\}_{j=1}^{Ni}$

3. METHODS

3.1 Model Architectures

3.1.1 Base-HitNet

This section will discuss the base HitNet architecture and present two alternate architectures based on the HitNet architecture. The alternate architectures are developed from the base HitNet architecture using the methods outlined below. Specifically, the base HitNet architecture is the first of the three architectures presented, followed by the remaining HitNet architecture, and finally, the combined spatial hierarchical HitNet architecture[32], that Figure1 illustrates. The base HitNet architecture includes two convolutional layers that precede a Hit-or-Miss (HoM) layer. The HoM layer then produces a reconstruction subnet or prediction vector, as indicated in Figure 2. Additionally, only 28×28 grayscale images were fed into the proposed structure during the study. The HoM layer consists of ten classes that each have 16 values. Each class represents one of the numbers in the MNIST dataset. The reconstruction subnet receives both the output of the HoM layer and the actual category classification of the sample as inputs. The reconstruction subnet attempts to reproduce the original image based on these inputs and calculates the average squared error between the reproduced image and the original image. Low average squared errors indicate that the model and submodels effectively encoded the information. In general, the HitNet model is essentially a HoM layer that has been constructed on top of feature maps and is utilized in pairs with a center loss function[8]. Figure 1 illustrates the overall HitNet model architecture, which includes the following steps:

1. To create feature maps, there are two 9x9 convolution layers with stride lengths of (1,1) and (2,2) respectively, containing 256 channels each and a ReLU activation function.
2. A layer that is connected to the whole K x N matrix, followed by a Batch Normalization (BN) layer and a Sigmoid activation function applied to each individual component, this results in a High-Order Model (HoM) composed of K capsules of size N.
3. For each HoM capsule, calculate the Euclidean distance to the central capsule C: (0.5,…, 0.5) to produce the prediction vector of the model ypred.
4. The Euclidean distance from the central capsule C: (0.5,..., 0.5) is determined for each HoM capsule. This gives us our prediction vector of the model ypred. 4. Except for the HoM capsule which represents the true class (predicted during test), all HoM capsules will be hidden (assigned 0). Then, we combine these HoM capsules and send them to a decoder to generate an Xrec output image. The objective of this step is to reconstruct the original image. The decoder is composed of two fully connected layers of size 512 and 1024 with ReLU activation functions, and a fully connected layer with an array of the same dimensions as the input image, with a sigmoid activation function.

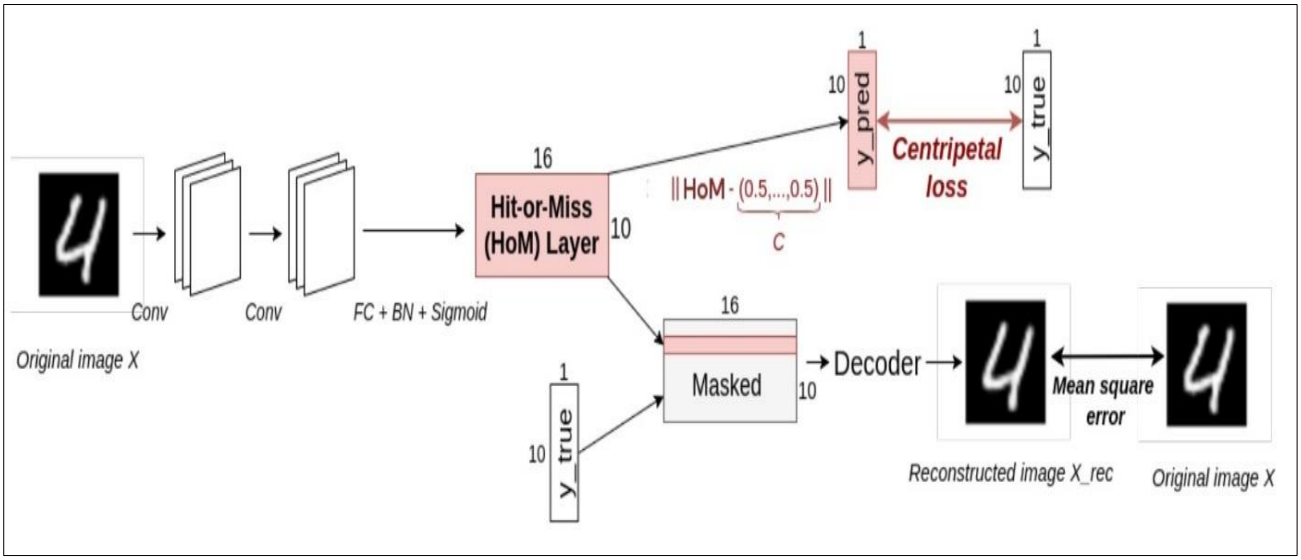


Figure 1: Original architect HitNet

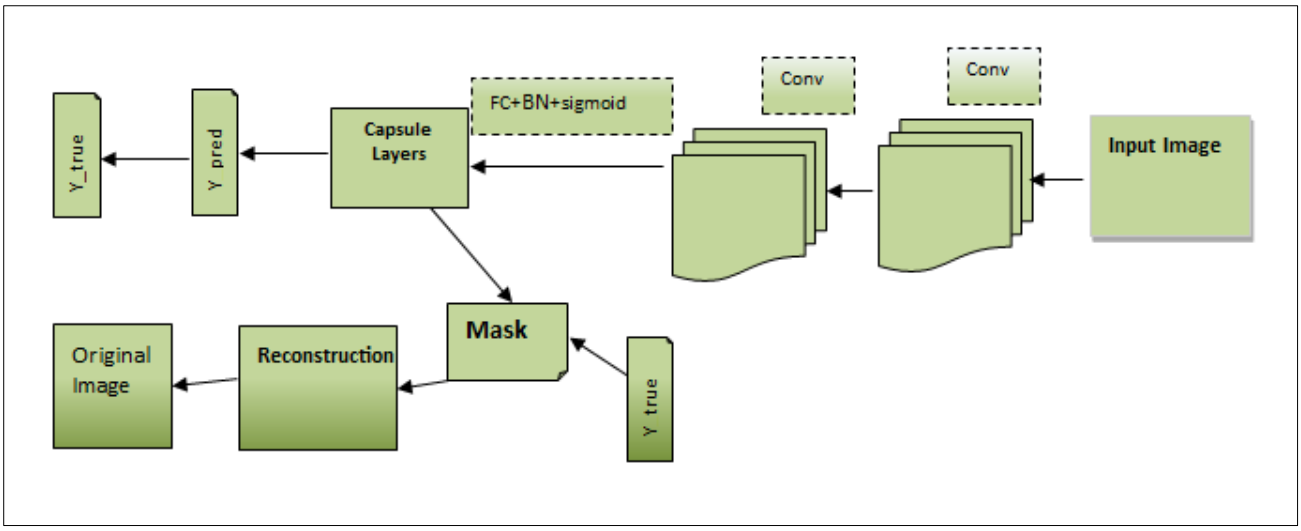


Figure 2: Flowchart of the Proposed Method

The fundamental model is composed of three convolutional layers that will represent the input data. Table 1 lists the parameters of every single layer. Also, Figure1 provides a visualization of the composition of the basic model being presented here. Features were selected to be the most consistent with the HitNet main network. Since HitNet takes 256x256 image inputs, it has an additional convolutional layer to encrypt the data and reduce the dimensions of the images. Additionally, there are only three categories within the basic model, and every category is represented by a vector of length 20. For all of the methods discussed in this paper (the basic model and the two other methods), the length of the HoM vectors was randomly assigned to be 20. Generally speaking, HitNet uses a HoM layer that operates over feature maps and is paired with a center loss function. To demonstrate the benefits of using a HoM layer, a surface method to produce the feature maps used in our study. Figure 1 displays the entire HitNet network design including all components.

A two-layer 9 x 9 network with steps (1,1) and (2,2) with 256 channels and a ReLU activation function is used to produce the feature maps. After that, there will be a fully-connected layer with K × n matrix, followed by a Batch Normalization (BN) layer, and each element will have a sigmoid activation function. The result will be a HoM model with K capsules of size n. The Euclidean distance to the center capsule C will be determined, i.e., (0.5,..., 0.5) per HoM capsule, which will provide the model's prediction vector y_{pred}. All HoM capsules will be blocked (set to zero) except for the capsule that represents the correct class (the class that is expected during testing), then combined and passed onto the decoder to create the X_{rec} output image. The objective of the decoder is to reconstruct the original image. The decoder is made up of two fully connected layers of sizes 512 and 1024 with ReLU activation functions, and a fully connected layer with the same dimensions as the input image, with a sigma activation function.

In this way, if X is the original image, and y_{true} is the encoded classification, then y_{true} and y_{pred} will produce the L1 loss using the centrality loss function, while X and X_{rec} will produce the L2 loss using the mean square error function. The total loss function for X is defined as L = L1 + αL2, with α equal to 0.392 [5-22]. When performing a classification task, HitNet's classification prediction will be the smallest value of ypred. The parameters for L1 were selected as follows:

l = l0 = 0.1, h = h0 = 0.2, m = 0.1, m0 = 0.9, n = 16, and λ = 0.5 as in [22].

3.1.2 Residual HitNet (Res-HitNet)

A second variant of HitNet is formed with abbreviated representations of local and global knowledge in order to allow the higher-level layers of HitNet to access the knowledge from lower levels. One of the reasons why HitNet is formed in such a way is because it is most effective at predicting the correct tumor classification for each input image if both local and global knowledge are available. HitNet's short cuts enable the network to retain global knowledge, which can then be combined with local knowledge. It is anticipated that the remaining short cuts used within HitNet will improve overall network performance relative to the base model. Additionally, a third bypass layer has been incorporated into HitNet prior to the HoM layer in order to create a greater variety of short cuts.

The proposed model has 33,767,168 parameters in its reconstruction network. Most of those parameters come from the fact that it is reconstructing an image that is 256 x 256 pixels and has 65,536 pixels. There are two fully connected layers (each with a size of 256, then 512 neurons) before the reconstructed image. If we use a fully connected layer with 512 neurons, the reconstruction subnet would have around 33.7 million parameters, as shown in Figure 3. The transaction counts given in Table 1 show what happens when we subtract the reconstruction subnet from the entire model, that leaves only the convolutional coding layers and the HoM layer.

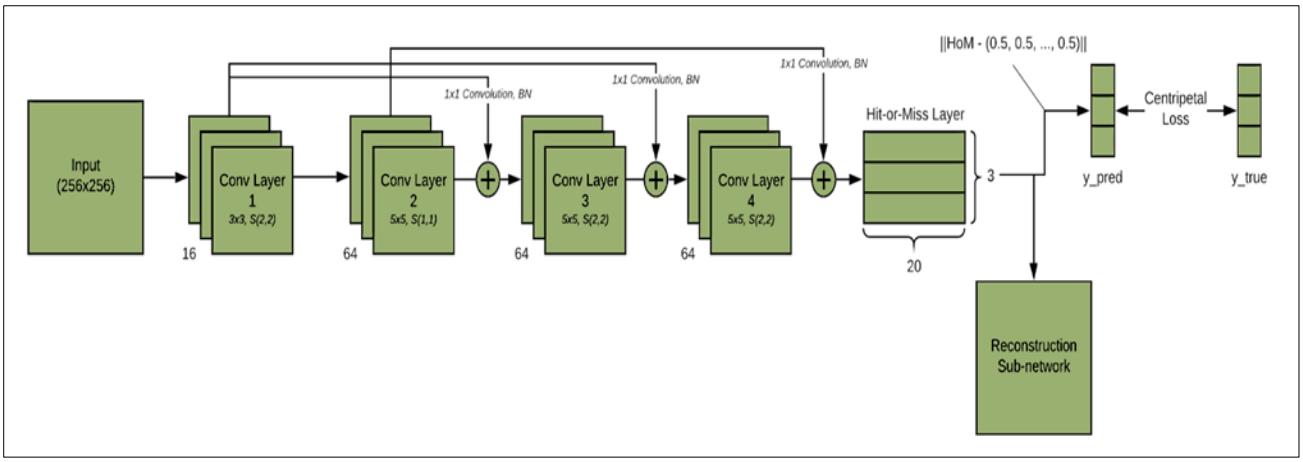


Figure 3: Res-HitNet Implementation

Table 1: Encoding Layers for the Baseline and Proposed HitNet Models

Model	Layer	Features	Filter Size		# parameter
HitNet [16, 17]	Conv. Layer 1	42	5x5	S(2,2)	——
	Conv. Layer 2	42	5x5	S(2,2)	
Baseline HitNet	Conv. Layer 1	16	3x3	S(2,2)	4,190,844
	Conv. Layer 2	80	5x5	S(2,2)	
	Conv. Layer 3	80	5x5	S(2,2)	
Residual HitNet	Conv. Layer 1	16	3x3	S(2,2)	4,170,156
	Conv. Layer 2	64	5x5	S(1,1)	
	Conv. Layer 3	64	5x5	S(2,2)	
	Conv. Layer 4	64	5x5	S(2,2)	
Pyramid Scene Parsing HitNet	Conv. Layer 1	16	3x3	S(2,2)	4,358,040
	Conv. Layer 2	32	5x5	S(1,1)	
	Max Pooling	32	2x2	S(2,2)	
	Branch 1	32	1x1	S(2,2)	
	Branch 2	32	2x2	S(2,2)	
	Branch 3	32	3x3	S(2,2)	
	Branch 4	32	6x6	S(2,2)	

3.1.3 Pyramid Scene Parsing HitNet (PSP-HitNet)

The PSP-HitNet is designed similarly to a series of four networks that are each trained on a different size of convolutional filters (1x1, 2x2, 3x3, and 6x6) that have been used previously by Zhao *et al*., for analyzing city and room scenes [61]. These varying sized filters enable the network to learn how to encode features from various size objects in an image as illustrated in the figure. It appears that the ability to encode these features will allow the network to make decisions about how to categorize each pixel in an image because of the larger area of influence created by the filters. Also included in the PSP-HitNet is a global prototype that transmits information from the first convolutional layer to the output of the cluster branches.

The original PSPNet was trained on images sampled from the set of ascending convolution branches, however, the PSP-HitNet used for testing was trained on images that were not split into grids, and thus did not have to predict the class of each individual pixel as in the original PSPNet. All but the first convolutional layer in the PSP-HitNet (and all branches within them) are linked together using a HoM layer that corresponds to that particular branch. At the end of the process, a voting mechanism is used to combine the five separate HoM branches into a single HoM layer, which then classifies each test sample.

4. Implementation and Evaluation

4.1 Dataset

The MRI dataset was obtained and is a collection from Kaggle, which contains 3,264 images. The reported class distribution is summarized in Table 2. The source describes the dataset as containing four categories: glioma, meningioma, pituitary tumor, and no tumor. While figure 4 shows MRI images according to the different types of brain tumors.

Table 2: The source describes the dataset

Category reported in study	Number of images	Share of dataset
Glioma	926	28.37%
Meningioma	937	28.71%
Pituitary tumor	901	27.60%
No tumor	500	15.32%
Total	3,264	100.00%

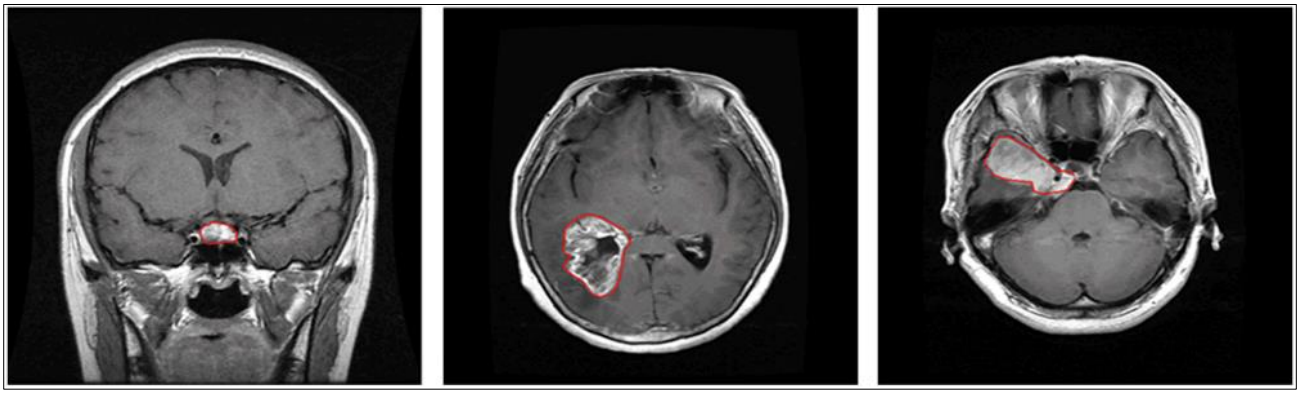


Figure 4: Sample images of three common brain tumors: (a) Meningioma, (b) Glioma, and (c) Pituitary Tumor

4.2 Implementation

To implement a system for the detection and classification of brain tumors through Python using the Kaggle dataset, Python libraries like TensorFlow and Keras were utilized to create Convolutional Neural Networks (CNNs). CNNs can identify brain tumors in MRI images. The process begins with importing the dataset from Kaggle, then preprocessing the images. After the images have been preprocessed they will serve as input to the neural networks. The neural networks will then be trained. After the model has been trained, it can be utilized to locate brain tumors within new MRI images.

4.3 Evaluation

Accuracy, recall, and F1 score are other metrics displayed in addition to the verification accuracy for every category. Recall is used to determine how many times a model makes a false negative error. Conversely, accuracy is used to determine how many times a model makes a false positive error. An F1 score that is high means the model produces very few false positive errors as well as false negative errors. Thus, you want to increase your F1 score when evaluating your models. You can find the equations for accuracy, recall, and F1 score listed below. True positive, false positive, true negative, and false negative are represented by TP, FP, TN, and FN, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

4.4 RESULTS

The five-fold cross validation with the regular training data set and the matching data set were used to present the results for each of the three proposed models, and then compared the performances of these models. The results were also presented when each model was used without its structure. Based on these results, we chose the best model for active deep learning testing.

The removal of the convolutional layers in succession from Base-HitNet resulted in similar performance levels to those of Base-HitNet prior to modifications. To be precise, the average verification accuracy of Base-HitNet when the first layer was removed was 84.59%, and when both the second and third layers were removed was 84.52%. It is not necessary to obtain additional data for the removal of the third layer since both the second and third layers share the same filtering specifications.

For Res-HitNet, there was a greater reduction in the overall performance of the original model when shortcut links were removed. The removal of the second or third shortcut link led to a total loss of about 2% from the original model's results. It appears that shortcut links added after the initial model development had a larger effect on model performance than previous shortcuts. As with the base HitNet results, modifications made to PSP-HitNet had virtually no effect. The

lack of effect could possibly be attributed to the random initial weight value, as some modified architectures performed slightly better than the original PSP-HitNet. One such modification was the removal of the branch containing 3×3 deviations, this resulted in an average verification accuracy of 86.63%, whereas the original PSP-HitNet had an average verification accuracy of 86.55%. In general, the results did not indicate any degradation in performance. A side-by-side comparison of the architectural changes to each of the three HitNet models is provided in Table3.

Table 3: Average Accuracy, Recall, and F1-Score for All Models after Training on the Original Dataset for 30 Epochs

Class	HitNet Model	Accuracy	Recall	F1-Score
Meningioma	Base	78.3%	75.4%	75.8%
	Res	79.6%	74.3%	76.6%
	PSP	78.5%	79.2%	77.8%
Glioma	Base	77.3%	78.6%	82.6%
	Res	87.3%	88.8%	89.8%
	PSP	93.3%	91.4%	90.1%
Pituitary	Base	92.3%	93.9%	91.4%
	Res	87.2%	88.3%	89.3%
	PSP	94.4%	95.8%	89.6%

To evaluate the effect of changing parameters on the average verification accuracy of networks, the number of parameters is also included. All parameters were calculated without counting the reconstruction subnet.

Table 4: Average Accuracy, Recall, and F1-Score for PSP-HitNet after Training on the Balanced Augmented Dataset (Up to 165 Epochs)

CLASS	ACCURACY (%)	RECALL (%)	F1-SCORE (%)
MENINGIOMA	91.3	90.8	89.3
GLIOMA	96.1	90.2	92.7
PITUITARY	92.2	95.4	96.9

4.4.1 Use Active Deep Learning on the Best Model

The highest performing model (PSP-HitNet) was trained using the batch active learning approach. For every iteration of the active deep learning process, the model was trained for at least 200 training epochs, however, if the model did not show improvements in validation accuracy for 25 consecutive training epochs, the training was discontinued. There were slight performance enhancements in comparison to the random selection method based on the use of uncertainty estimations. Unlabeled samples that have low confidence in their classification are selected from the unlabeled dataset, which allows the network to learn more quickly than if it were using randomly selected samples.

The distinction between the two query methods can be seen in the way the model selects from the unclassified data set 100 of the lowest confidence, randomly selected, or 100 of the highest confidence samples during each iteration. To choose 100 samples per iteration, a minimum of 13 iterations will need to take place. Therefore, on average, approximately 6 percent of the total training set is classified. As illustrated in Table 5, selecting the first 12-24 percent of the training data resulted in greater improvements as compared to selecting fewer 100 reliability samples per iteration. As illustrated in figure 5.

Table 5: Mean Validation Accuracy at Each Iteration of Active Learning Using SGD Optimizer with L2 Regularization

ITERATION	80% UNCERTAINTY	80% RANDOM
0	74.503%	74.793%
1	86.670%	82.302%
2	87.027%	87.993%
3	89.512%	90.051%
4	90.901%	90.048%
5	90.597%	89.686%

The initial phase of Active Learning (AL), i.e., labeling the 100 most unreliable samples did not result in as much improvement as other techniques, perhaps as a consequence of using Random Initialization for each iteration when the model was re-trained from scratch. Also, because of Stochastic Gradient Descent (SGD) optimization algorithm, some AL iterations may employ Query Strategies in which the model can become trapped in a Local Minimum. On the contrary, for the second, third, and fourth iterations of AL, labeling less reliable samples outperformed.

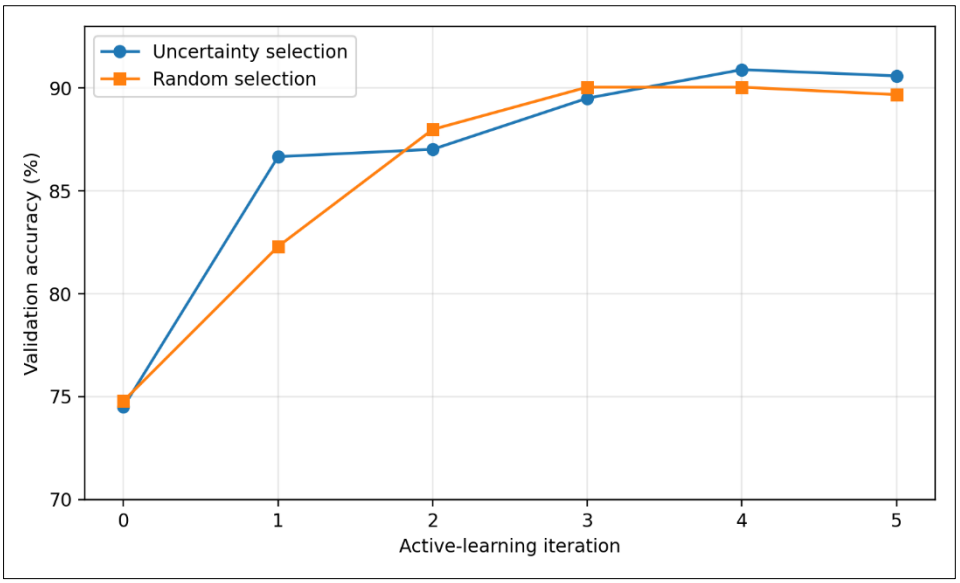


Figure 5: Active-learning validation accuracy

Labeling random samples and labeling more reliable samples. This technique provided the highest amount of information for PSP-HitNet training. Since many of the images within the training set provide less information to the model than others, labeling less reliable images produces an average verification accuracy of approximately 90% as illustrated in Table 6. Overall, with an increase in the number of labeled samples in the training set, all three Query Strategies produce similar results with an average verification accuracy of approximately 90%.

Table 6: Comparison of the Proposed Method with Other Studies

REF.	METHOD	ACCURACY (%)
(KUMAR,2023)[15]	CNN (classification and segmentation)	86.23%
(S. HONG,2024)[19]	Residual MLP	91.36%
(GOCERI,2020) [20]	Capsule Networks	92.65%
(RAYTHATHA,2023) [26]	Capsule Networks based CNN	93.55%
PROPOSED METHOD	**PSP-HitNet**	**95.98**

With respect to the results obtained and the comparative analysis of the accuracy of the comprehensive diagnosis of brain tumor types in the proposed method, it can be stated that the proposed method demonstrates adequate and relevant performance compared to the other methods examined. An accuracy level of 95.98% is deemed to be acceptable relative to other available methods. As illustrated in figure 6.

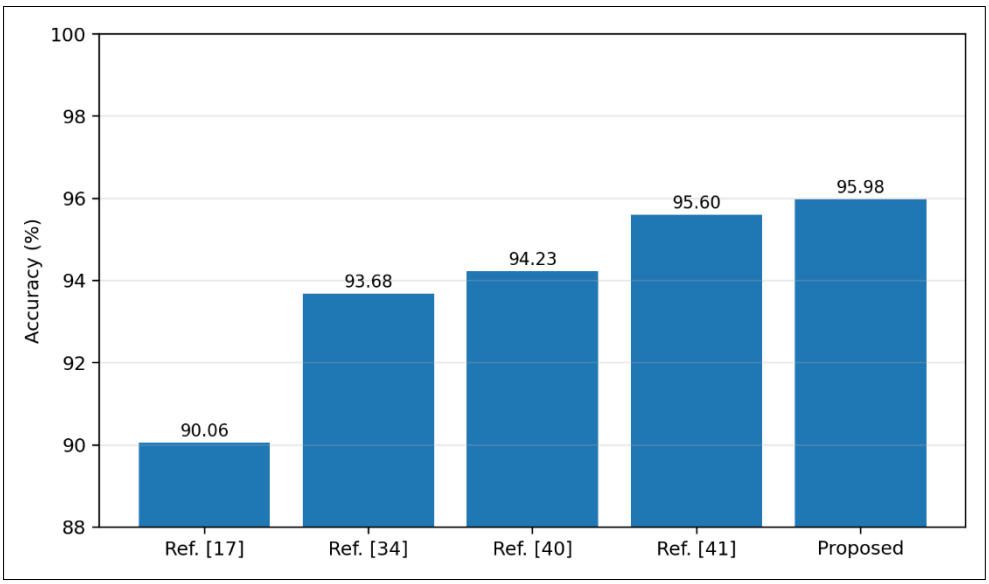


Figure 6: Accuracy comparison reported in the thesis

5. DISCUSSION

This study was conducted in order to determine if capsule-based networks can classify brain tumors using MRIs more effectively and reliably than conventional CNNs and if using active learning can decrease the number of labeled images needed to reach high levels of classification effectiveness. Overall, the results showed both positive and negative outcomes. The use of capsules was effective in classifying brain tumors. One of the three capsule-based models performed significantly better than the other two models. Active learning also improved the results, however, it did not improve the results to the degree that we would have preferred.

5.1 Compare the Three Structures

The design goal of the three different HitNet models was to make them as similar as possible. The coding layers (encoding layers) in all three of the models use roughly the same amount of parameters (approximately 4.19 million for the basic parameter model, 4.17 million for the Res-HitNet model and 4.36 million for the PSP-HitNet model, without including the reconstruction subnet, as illustrated in Table 7. therefore, any observed differences are most likely not attributable to the size of the model, but instead to the manner in which every network receives and combines information.

Table 7: Encoder parameter counts

ARCHITECTURE	ENCODER PARAMETERS
BASE-HITNET	4,190,844
RES- HITNET	4,170,156
PSP- HITNET	4,358,040

The PSP-HitNet model performed best, It had the highest results for every single category of glioma (glioma) and pituitary tumor (pituitary tumors), and its highest performing setting had an accuracy of 95.98%. The Res-HitNet model came in second position, and the basic model was the least successful. The largest increase in accuracy occurred in the glioma category. Accuracy (precision) went from 77.3% in the basic model to 87.3% with the use of additional shortcut connections (residual shortcuts), and was further improved to 93.3% with the implementation of a pyramid unit (pyramid module). This can be attributed to the variability in size and shape of glioma border regions. Thus, the branches of the pyramid which analyze images using different-sized filters provide greater capability to recognize these variations. Additional shortcut connections appear to contribute to the same end: they convey broad, overall information (global information) in conjunction with local characteristics.

Meningiomas represented the least precise area for classification by all models. The F1 scores on this scale were between about 76% and 79% for each model, which is significantly less than the F1 scores obtained for both of the other categories of tumors. Although we did not fully investigate why meningiomas were the most problematic category for classification, we suspect that meningiomas are typically located close to bone in the skull and, therefore, can look very similar to other types of lesions on a single two-dimensional (2D slice). Therefore, a model that relies solely upon analysis of a single slice has limited ability to provide enough information to accurately classify meningiomas.

6. Limitations

Limitations of the research include the fact that the research is based on a single publicly available database of 3,264 images and does not include external validation through surveys conducted with image data from other manufacturers or healthcare facilities. Therefore, it cannot be determined whether or not the results of this research can be generalized to other populations. In addition, all of the images included in the study were collected as 2D sections. As a result, the images lose the 3D perspective that radiologists typically use during patient diagnosis. The (no tumor) class is much smaller than the other classes (only 500 images), and the analyses of each class separately focused on the three different types of tumors. The reconstruction subnet has about 33.7 million parameters. This is substantially larger than the encoder alone. The number of parameters in the reconstruction subnet increases the complexity of training the model and also requires more computational power than training only for classification purposes. Finally, the proposed model will provide a second opinion for clinicians but will not replace clinicians' decision-making.

7. CONCLUSION

Rapid diagnosis of brain tumors using MRI can contribute to increasing patients' chances of survival. Some researchers use trained machine learning models as second opinions to aid diagnosis. These techniques include optimizing the target tumor area, capsule networks, and convolutional neural networks. In this study, several HitNet networks were proposed and trained on a dataset of digital images from brain MRI scans, containing meningiomas, gliomas, and pituitary tumors. Through experiments on original and homogeneous datasets, PSP-HitNet was shown to perform better than Base-HitNet and Res-HitNet, achieving a maximum verification accuracy of about 95.8%, improved by up to 2% compared to Base-HitNet. Training the best model using incremental data for up to 165 sessions showed that PSP-HitNet achieved a maximum validation accuracy of 94% per training unit.

While improving and building new deep learning architectures is one way to increase overall test validity and accuracy, other techniques, such as active deep learning, work with datasets to increase verification accuracy. Access to large amounts of data is not always available, and active deep learning helps train models faster with less data. This work proposes the use of PSP-HitNet with active deep learning, allowing it to perform faster training by selecting examples with the lowest level of confidence or the highest level of uncertainty. Using this technique, the model was able to achieve better performance than random selection. Even with a small dataset, the network was able to quickly achieve efficiency by classifying unclassified examples with the lowest confidence level and adding them to the main training dataset.

8. Future Work

To expand the scope of this work and improve the active learning technique offered, there are several possible paths. Medical images may contain overlapping information, making it difficult to select a subset of unclassified data based on uncertainty estimates. The researcher propose the Fisher Information Matrix and an effective matrix reduction method to select the most useful samples from unclassified data.

Some features can also be extracted from MRI slides using wavelet conversion. Wavelet conversion has also been applied to detect cancer in ultrasound images. Separate or fixed waves can help filter out unnecessary information in an MRI, allowing the network to focus on remaining brain tumor features.

In the context of deep active learning, another approach is to resize the classified data set into smaller, more precise parts to better understand model behavior as each sample is added. For example, instead of annotating 100 samples from an unclassified data set, the system annotations at least 50, 25, 10, or 5 samples in each iteration of active deep learning. The latest proposed expansion is to re-implement the segmentation portion of the PSP module in PSP-HitNet. This model will enable tumors to be accurately classified and precisely located in the brain. Therefore, it will make this model more powerful than it is now.

REFERENCES

1. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing,* vol. 39, no. 2, pp. 757-775, 2020.
2. H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," *IEEE access,* vol. 7, pp. 69215-69225, 2019.
3. W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification: W. Ayadi et al," *Neural processing letters,* vol. 53, no. 1, pp. 671-700, 2021.
4. P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries," in *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2019: IEEE, pp. 1368-1372.
5. D. Saravanan, G. Arunkumar, T. Ragupathi, and P. R. Rao, "Weighted Majority Voting Ensemble for MRI-Based Brain Tumor Classification Using Capsule Networks and XGBoost," *Applied Magnetic Resonance,* vol. 56, no. 3, pp. 395-425, 2025.

A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: a survey," *Journal of imaging,* vol. 8, no. 8, p. 205, 2022.

6. E. Wahyudi and B. Imran, "HCViT: A hybrid convolutional vision transformer with spatial attention bias for brain tumor classification in MRI," *Journal of Soft Computing Exploration,* vol. 7, no. 3, pp. 695-715, 2026.
7. S. J. Chang, "Brain Tumor Classification Using Hit-or-Miss Capsule Layers," California Polytechnic State University, 2019.
8. R. Sheergojri and P. Iqbal, "MRI Brain Tumor Classification and Uncertainty Estimation using Hybrid CNN-ELM Model," in *2022 3rd International conference on smart electronics and communication (ICOSEC)*, 2022: IEEE, pp. 978-983.
9. S. Deepak and P. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in biology and medicine,* vol. 111, p. 103345, 2019.
10. N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World Congress on Medical Physics and Biomedical Engineering 2018: June 3–8, 2018, Prague, Czech Republic (Vol. 1)*, 2018: Springer, pp. 183-189.
11. R. Vankdothu, M. A. Hameed, and H. Fatima, "A brain tumor identification and classification using deep learning based on CNN-LSTM method," *Computers and Electrical Engineering,* vol. 101, p. 107960, 2022.
12. S. Iqbal *et al.*, "Hybrid parallel fuzzy CNN paradigm: Unmasking intricacies for accurate brain MRI insights," *IEEE Transactions on Fuzzy Systems,* vol. 32, no. 10, pp. 5533-5544, 2024.
13. S. Srinivasan, D. Francis, S. K. Mathivanan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, "A hybrid deep CNN model for brain tumor image multi-classification," *BMC Medical Imaging,* vol. 24, no. 1, p. 21, 2024.

14. S. Kumar and D. Kumar, "Human brain tumor classification and segmentation using CNN," *Multimedia Tools and Applications,* vol. 82, no. 5, pp. 7599-7620, 2023.
15. S. Boehringer, A. Sanaat, H. Arabi, and H. Zaidi, "An active learning approach to train a deep learning algorithm for tumor segmentation from brain MR images," *Insights into Imaging,* vol. 14, no. 1, p. 141, 2023.
16. S. A. A. Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artificial intelligence in medicine,* vol. 102, p. 101779, 2020.
17. S. M. Vijithananda *et al.*, "Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques," *Biomedical engineering online,* vol. 21, no. 1, p. 52, 2022.
18. S. Hong, J. Wu, L. Zhu, and W. Chen, "Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP," *Plos one,* vol. 19, no. 7, p. e0298102, 2024.
19. E. Goceri, "CapsNet topology to classify tumours from brain images and comparative evaluation," *IET Image Processing,* vol. 14, no. 5, pp. 882-889, 2020.
20. R. Rekha, P. Shruti, M. Deekshitha, and J. Akash, "DCSwin-UNet: Dual Encoder U-Net based on CNN and Swin Transformer with Trainable Multiplication Layer for brain tumor segmentation from MRI images," *Biomedical Signal Processing and Control,* vol. 110, p. 108325, 2025.
21. Shahroudnejad, P. Afshar, and A. Mohammadi, "EXPLAINABILITY OF CAPSULE NETWORKS AND THEIR APPLICATION FOR BRAIN TUMOR CLASSIFICATION."
22. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *2009 ieee conference on computer vision and pattern recognition*, 2009: IEEE, pp. 2372-2379.
23. R. LaLonde, P. Kandel, C. Spampinato, M. B. Wallace, and U. Bagci, "Diagnosing colorectal polyps in the wild with capsule networks," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020: IEEE, pp. 1086-1090.
24. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction," *Sensors,* vol. 22, no. 3, p. 1184, 2022.
25. Y. Raythatha and M. Vani, "A Paradigm Shift in Brain Tumor Classification: Harnessing the Potential of Capsule Networks," in *2023 IEEE 2nd International Conference on Data, Decision and Systems (ICDDS)*, 2023: IEEE, pp. 1-6.
26. S. b. Nuthalapati, A. Bushara, and K. Abubeker, "SPP_CNN: Spatial Pyramid Pooling for Optimizing Brain Tumor Classification," in *International Conference on Electrical and Electronics Engineering*, 2024: Springer, pp. 1-16.
27. S. babu Nuthalapati, A. Bushara, and K. Abubeker, "SPP_CNN: Spatial Pyramid Pooling for Optimizing Brain Tumor," *Innovations in Electrical and Electronics Engineering: Proceedings of ICEEE 2024, Volume 2,* vol. 2, p. 1, 2025.
28. M. I. Sharif, J. P. Li, M. A. Khan, and M. A. Saleem, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," *Pattern Recognition Letters,* vol. 129, pp. 181-189, 2020.
29. S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, 2017: IEEE, pp. 835-838.
30. Ma *et al.*, "Current treatment outcomes of congenital heart disease and future perspectives," *The Lancet Child & Adolescent Health,* vol. 7, no. 7, pp. 490-501, 2023.
31. F. H. Shajin, S. P, P. Rajesh, and V. K. Nagoji Rao, "Efficient framework for brain tumour classification using hierarchical deep learning neural network classifier," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 11, no. 3, pp. 750-757, 2023.