Inception Architecture for Brain Image Classification

R Tamilarasi* and S Gopinathan
Department of Computer Science, University of Madras, Chennai, Tamil Nadu, India
Email: *tamil.ramamoorthy@gmail.com

Abstract. A non-invasive diagnostic support system for brain cancer diagnosis is presented in this study. Recently, very deeper convolution neural networks are designed for computerized tasks such as image classification, natural language processing. One of the standard architecture designs is the Visual Geometric Group (VGG) models. It uses a large number of small convolution filters (3x3) connected serially. Before applying max pooling, convolution filters are stacked up to four layers to extract features' abstraction. The main drawback of going deeper is overfitting, and also updating gradient weights is very hard. These limitations are overcome using the inception module, which is wider rather than deeper. Also, it has parallel convolution layers with 3x3, 5x5, and 1x1 filters that reduce the computational complexity due to stacking, and the outputs are concatenated. This study's experimental results show the usefulness of inception architecture for aiding brain image classification on Repository of Molecular Brain Neoplasia DaTa (REMBRANDT) Magnetic Resonance Imaging (MRI) images with an average accuracy of 95.1%, sensitivity of 96.2%, and specificity of 94%.

Keywords: Deep learning, neural networks, inception architecture, brain cancer, image classification

1. Introduction

The most important complex organ is the brain in the human body. The uncontrolled cell growth, often referred to as brain cancer, leads to sudden death. It spreads very quickly throughout the body. Thus an efficient approach is required which detects brain cancer at the earliest. Though the recently developed deep learning is applied to medical image analysis to provide promising results, many conventional approaches are also used for brain image classification. This section discusses the latest approaches available for brain image classification.

A series of operations are performed for brain image classification in [1]. The MRI image use Gaussian kernel for smoothing, and then segmented using a clustering approach. Features such as run-length matrix, color, and texture are extracted, and features are reduced by singular value decomposition and Principal Component Analysis (PCA). Finally, the residual neural network is employed for classification. Convolution Neural Network (CNN) is employed successfully in [2] for brain image classification. Three different architectures, Google Net, Alex Net, and VGG Net, are employed.

Different frequency domain transformations are analyzed in [3-6] for brain image classification. Wavelet-based thresholding is applied for de-noising, and then Otsu thresholding is applied in [3] to get the region of interest. Different wavelet filters [4]. Dual-Tree M-band Wavelet Transform [5] and tetrolet transform [6] are employed to extract features. Support Vector Machine (SVM) is employed in
[3-6] to diagnose brain cancer. Morphological features such as region properties, intensity features are extracted from the segmented region in [7] to diagnose brain cancer using Naive Bayes (NB) classification. CNN is implemented in the curvelet domain for brain image classification [8]. After decomposing, the curvelet coefficients are given to the CNN for classification. Alex Net is used for classification [20].

A combination of deep and handcrafted features is employed in [9] for brain image classification. The co-occurrence features from the modified greyscale matrix are extracted. The combined features are classified using an SVM classifier. A modified CNN architecture is implemented in [10] for brain image classification. Complex networks based on graph algorithms and a modified activation function are designed to improve CNN-based systems' classification accuracy.

A Hybrid feature-based system is described in [11] for brain image classification. Before extracting features, dynamic histogram method is used for equalization to enhance the MRI image, and unwanted region of skull is also removed in the pre-processing. A set of features, intensity-based, local binary pattern, co-occurrence, gradient-based features are extracted, and then PCA is employed for feature reduction [27]. Finally, particle swarm optimization with SVM is used for classification. Wavelet features and 1st and 2nd order statistical features are employed in [12] for brain image classification. The combined features are classified using multi-layer perceptron.

A whole image classification that does not require any pre-processing or segmentation is described in [13]. It uses CNN architecture for classification using the extracted deep features. A fast brain cancer diagnosis system is discussed in [14] using the CNN method. It uses thresholding to segment the brain tumor, which increases the network speed with root mean square loss function.

Deep wavelet auto encoder architecture is discussed in [15] for brain image classification. It uses wavelet-based decomposition as a feature extraction approach and an auto encoder-based feature reduction approach for effective classification. The well-developed deep learning architectures are employed in many image analysis approaches to achieve better performances [16-19]. The most commonly used architectures are VGG-16, Alex Net, and Google Net [21].

This study aims to design the inception module and then show its applicability in brain image classification. The rest of the paper is as follows: Section 2 describes the inception architecture used in this study to extract feature maps using convolution filters and classification by fully connected layers in the deep learning architecture. Section 3 evaluates the inception architecture applied to REMBRANDT data by several experiments, and the last section summarizes the study.

2. Methods and Materials
MRI images give a unique perspective view of the human brain. Though many images are acquired, the whole information in these images is not utilized fully for cancer classification. The main reason for this is the difficulty of quantifying the images' information in a meaningful form. Extracting basic texture features and other information from these MRI brain images by image interpretation technique is a quantitative approach to the problem. Deep learning, the direct technique that includes feature extraction and classification module, the analysis proves to be successful to varying degrees in many applications, especially in pattern recognition applications. One of the best image localization models and classification is the VGG model [22], which uses smaller-sized kernels. Figure 1 shows a simple VGG architecture.
The size and area occupied by the tumors and the brain images' tumor location are different in each image. To effectively extract this information, the selection of kernel size is very important. The locally distributed information is extracted when using the smaller kernels, whereas the VGG model uses a smaller sized kernel and computationally expensive as it is going deeper and deeper. Thus, the inception module is used [23]. Figure 2 shows the arrangement of convolution layers in the naive inception module (NIM).

An efficient inception module for brain image classification is developed. Figure 3 shows the stacked NIM for the effective classification of brain images. In any deep learning architecture, there are two simple elements; convolution and pooling layers. However, deep learning for a given problem mainly depends on the arrangement of these two layers. The performance and model stability are improved, the input data is rescaled to small values between [0,1] before applying the convolution. Due to small weights in the model, the un-scaled data results in an unstable learning process.
The data normalization is employed to rescale the data, which is defined by

\[ \text{normalized}_x = \frac{x - \text{min}}{\text{max} - \text{min}} \]  

(1)

where min and max are the minima, and maximum values in the available data and normalized \( x \) are the normalized value of \( x \). Table 1 shows the data normalization on sample data.

| Input data (x) | Min | Max | Normalization | normalized \( x \) |
|----------------|-----|-----|---------------|-------------------|
| 15             | 5   | 27  | 15 – 5/27 – 5 | 0.45              |
| 22             | 5   | 27  | 22 – 5/27 – 5 | 0.77              |
| 27             | 5   | 27  | 27 – 5/27 – 5 | 1                 |
| 5              | 5   | 27  | 5 – 5/27 – 5  | 0                 |
| 15             | 5   | 27  | 15 – 5/27 – 5 | 0.45              |
| 24             | 5   | 27  | 24 – 5/27 – 5 | 0.86              |

It is inferred from the normalization Table 1 that the input data is rescaled in 0-1. After scaling, the normalized input data is convolved with the convolution filters to obtain the feature map. The convolution filters are also called feature detectors. The mathematical definition for two-dimensional convolution is given by

\[ y(i, j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(m, n) \cdot x(i-m, j-n) \]  

(2)
After flipping, simple multiplication is performed by placing the feature detector on the top-left corner of the input image, as shown in Figure 4. The feature detector is then moved in the right direction by one pixel, i.e., a stride of 1, and does the same multiplication operation to get the feature map. After moving the feature detector through the first row completely, the same operation is repeated for the second row and so on to get the feature map. There are no zero paddings applied in Figure 5. Thus the convolution gives the feature map of only 3x3. in obtaining the feature map with the same size as the input image, zero paddings are applied. Figure 6 shows the 3x3 convolution with 32 filters at the first block of VGG architecture.
Figure 6: Extracted feature map

Multiple feature detectors are used in each convolution layer to detect multiple feature maps that increase the feature map's dimension. Thus, the max-pooling layer is introduced to reduce the feature map's dimension after extracting the feature map. Figure 7 shows the pictorial representation of how the max-pooling layer works.

![Figure 7: Pictorial representation of the max-pooling layer](image)

The final feature maps obtained via the stacked convolution layers and max-pooling layers are passed to the fully connected or dense layers for classification. The first layer, which is the input layer, receives the feature maps. To back-propagate the errors in the neural network, a linear activation function is needed. Simultaneously, it acts as a non-linear function to characterize the complex relationship in the data. Thus, Rectified Linear Activation Unit (ReLU) is used in the hidden layers. Table 2 shows how the activation function ReLU works.
Table 2: Activation function @ Input layer

| Input (x) | Output of ReLU function |
|-----------|-------------------------|
| 0.8       | 0.8                     |
| -0.2      | 0                       |
| 0.5       | 0.5                     |
| 1.2       | 1.2                     |
| -0.8      | 0                       |
| -0.2      | 0                       |

From Table 2, it is observed that the ReLU is linear for positive values and non-linear for negative values. It remains very close to linear as it is a piecewise linear function, making the optimization very easier for neural networks. During training, the model weights are updated to provide better performance through an error function named cross-entropy loss. It is defined by

$$CrossEntropyLoss = -\sum_{n=1}^{m} t_n \log(s_n)$$

$t_n$, $s_n$ The true label and the sigmoid function output for the class $n$ and $m$ are the total numbers of classes.

The deep learning architecture is used for a two-class problem (normal/abnormal) and thus sigmoid activation function is employed at the output layer. The class is assigned to 1 if the sigmoid layer outputs over 0.5 else assigned to class 0. Table 3 shows how the sigmoid activation function works for several input values, and Table 4 summarizes the training parameters for VGG and Inception architectures.

Table 3: Activation function @ Output layer

| Input (x) | The output of the sigmoid function $\sigma(x) = \frac{1}{1 + e^{-x}}$ |
|-----------|---------------------------------------------------------------|
| 0.2       | 0.55                                                          |
| 0         | 0.5                                                           |
| 0.3       | 0.574                                                         |
| 0.5       | 0.622                                                         |
| 1         | 0.731                                                         |
| 0.4       | 0.599                                                         |

Table 4: Training parameters for VGG and Inception architectures

| Parameters                                  | Values               |
|---------------------------------------------|----------------------|
| Optimizer                                   | Stochastic gradient descent |
| Loss function                               | Cross entropy loss    |
| Activation function @ input layer           | ReLU                 |
| Batch size                                  | 64                   |
| Learning rate                               | 0.01                 |
| Iteration                                   | 20                   |
| Activation function @ output layer          | sigmoid              |

In Table 4, batch size refers to the number of samples used to compute the error in the loss function, and the learning rate refers to the number of parameters to be updated at each iteration. The overall algorithm for the classification of brain image into normal or abnormal is as follows:
Algorithm 1: Classification of brain image by Inception-V3 model

| Input: | Training images $I = \{I_1, I_2, \ldots, I_n\}$ and testing images |
|--------|---------------------------------------------------------------|
| Output: | Predicted class of the testing image                          |

- Normalize the range 0 to 1 using Eqn.1
- Apply the two-dimensional convolution operation using Eqn. 2 with the desired number of convolution filters to obtain the feature map
- Apply the max-pooling layer to reduce the feature's map dimension as described in Figure 7.
- Steps 2 and 3 are repeated many times as per the design of the architecture
- Initialize the training parameters, which is shown in Table 4
- Train the network using the extracted feature maps
- Update the model weights using cross-entropy loss using Eqn. 3
- Repeat the training until the required number of epochs is reached.
- Save the trained model for testing.
- Test the samples with the trained model using the sigmoid function, which is shown in Table 3.
- If the model's output is over 0.5, the predicted class is '1'; otherwise, it is assigned to class '0.'

3. Results and Discussions

This section evaluates comprehensively the effect of inception architecture designed for brain image classification on interpreting a standard set of brain tumor images in the REMBRANDT database [24, 25]. A general requirement for the images used in this study is that the images are MRI brain images. The selected images include normal and abnormal classes of interest since this study aims to assess the performances of inception architecture in differentiating between normal and abnormal classes. Thus, 100 images in each category (normal/abnormal) are selected from the database of 130 subjects. Figure 8 shows MRI sample images.

![Sample REMBRANDT database images](image-url)

**Figure 8:** Sample REMBRANDT database images
The system’s performance is analyzed and compared with the standard VGG-16 architecture. The Classification Accuracy (CA) is defined by

\[
CA(\%) = \frac{\text{#images correctly classified by the deep learning architecture}}{\text{#images used for testing the deep learning architecture}} \times 100
\]  

The obtained CA reveals the overall accuracy of the system. The system performance on normal and abnormal images is independently performed. Two important metrics, Sensitivity (Sn) and specificity (Sp) are commonly used in the medical domain for a classification system. Their definitions are as follows:

\[
Sn(\%) = \frac{TP}{TP + FN} \times 100
\]
\[
Sp(\%) = \frac{TN}{FP + TN} \times 100
\]

The four parameters; TP, FN, TN, and FP in the above Eqns. (5) and (6) can be computed based on the classifier’s outputs for abnormal and normal images. The correct classification of abnormal images is considered True Positive (TP), and incorrect classification is considered False Negative (FN). Similarly, the correct classification of normal images is considered True Negative (TN), and incorrect classification is considered False Positive (FP).

The CA in Eqn. 4 can also be defined using the above four parameters and defined by

\[
CA(\%) = \frac{TP + TN}{TP + FN + TN + FP} \times 100
\]

The designed architecture needs validation after training as there is no guaranteed performance on testing samples. A simple random spilled cross-validation is employed with 50% of data in each category for training and remaining for testing. Table 5 shows the performances of VGG-16 and Inception-V3 on REMBRANDT MRI brain images.

**Table 5:** Performances of VGG-16 and Inception-V3 on REMBRANDT MRI brain images

| #Run | CA (%) | VGG-16 | Inception-V3 | CA (%) | VGG-16 | Inception-V3 | CA (%) | VGG-16 | Inception-V3 | CA (%) | VGG-16 | Inception-V3 |
|------|--------|--------|--------------|--------|--------|--------------|--------|--------|--------------|--------|--------|--------------|
| 1    | 91     | 93     | 92           | 94     | 90     | 92           | 94     | 90     | 92           | 94     | 90     | 92           |
| 2    | 95     | 93     | 96           | 94     | 94     | 92           | 96     | 92     | 92           | 94     | 90     | 92           |
| 3    | 92     | 94     | 94           | 96     | 92     | 94           | 96     | 92     | 92           | 96     | 92     | 92           |
| 4    | 92     | 95     | 94           | 96     | 92     | 94           | 96     | 92     | 92           | 96     | 92     | 92           |
| 5    | 94     | 94     | 96           | 96     | 92     | 94           | 94     | 92     | 94           | 96     | 92     | 92           |
| 6    | 93     | 96     | 94           | 96     | 92     | 94           | 94     | 96     | 92           | 94     | 90     | 96           |
| 7    | 88     | 96     | 88           | 98     | 88     | 94           | 98     | 90     | 98           | 94     | 90     | 96           |
| 8    | 90     | 97     | 90           | 98     | 90     | 96           | 98     | 96     | 98           | 94     | 90     | 96           |
| 9    | 96     | 97     | 98           | 98     | 94     | 96           | 98     | 96     | 98           | 94     | 90     | 96           |
| 10   | 97     | 96     | 98           | 96     | 96     | 96           | 96     | 96     | 96           | 94     | 90     | 96           |

It is inferred from Table 5 that the average CA of inception architecture has an improvement of 2.3% over the standard VGG-16 architecture. Though the maximum CA obtained by both architectures is 97%, the performance of inception architecture is better than VGG-16 architecture. The inception architecture provides consistent performance of over 93%. The increase in performance may be due to the extraction of globally distributed information by the larger kernel. It is also noted that the Inception architecture has more Sn and Sp than VGG-16 architecture, which means the more number of normal and abnormal images are correctly classified than VGG-16. At each iteration, the
performances of the loss function used during training are plotted in Figure 9, and their corresponding CA is shown in Figure 9.

(a) Plots for 97% accuracy
(b) Plots for 96% accuracy
(c) Plots for 95% accuracy
(d) Plots for 94% accuracy

Figure 9: Cross entropy loss and accuracy curves for 20 iterations

Table 6: Performances of Inception-V3 with other systems

| Techniques                        | Performance measures |
|-----------------------------------|----------------------|
| DWT+SVM [4] (Daubechies 8 filter) | CA       | Sn  | Sp  |
| DWT+SVM [4] (Symlet 8 filter)    | 93.5     | 95  | 92  |
| DWT+SVM [4] (bi-orthogonal 3.7 filter) | 88    | 88  | 88  |
| DTMBWT+SVM [5]                    | 97.5     | 95  | 100 |
| Tetrolet + SVM [6]                | 98       | 96  | 100 |
Table 6 shows a comparative study of the system designed in this study with existing systems. It is inferred that the Inception-V3 model gives reliable performance for both normal and abnormal brain image classification. Though the systems in [5-7] provide higher specificity, their obtained sensitivity is an important indicator for the malignant tumor classification is lower than the Inception-V3 system.

4. Conclusion

In this study, inception architecture for brain image classification is analyzed. The deep learning architecture of VGG-16 is going deeper and deeper, whereas the inception module is wider. This study is concerned mainly with the REMBRANDT database image analysis and feature extraction mapping by the inception architecture that may describe the dominant patterns for brain image classification into normal or abnormal. Results show that the inception architecture has an improvement over the VGG architecture. Both systems' performances are analyzed using the same set of training and testing images and the same parameter settings. It is observed that the average CA of inception architecture is 95.1%, whereas VGG-16 is 92.8%.

References

[1] Z. A. Al-Saffar and T. Yildirim, "A Novel Approach to Improving Brain Image Classification Using Mutual Information-Accelerated Singular Value Decomposition," in IEEE Access, vol. 8, pp. 52575-52587, 2020.

[2] V. Bhanumathi and R. Sangeetha, "CNN Based Training and Classification of MRI Brain Images," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 129-133.

[3] M. Gurbină, M. Lascu, and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines," 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 2019, pp. 505-508.

[4] S, Mohan Kumar (2016). Analysis of Different Wavelets for Brain Image Classification Using Support Vector Machine. International Journal of Advances in Signal and Image Sciences, 2(1), 1-4.

[5] Ayalapogu, R.R., Pabboju, S. and Ramisetty, R.R., 2018. Analysis of dual-tree M-band wavelet transform-based features for brain image classification. Magnetic resonance in medicine, 80(6), pp.2393-2401.

[6] Babu, B.S. and Varadarajan, S., 2017. Detection of a brain tumor in MRI scan images using tetrolet transform and SVM classifier. Indian J Sci Technol, 10(19), pp.1-10.

[7] Zaw, H.T., Maneerat, N. and Win, K.Y., 2019, July. Brain tumor detection based on Naïve Bayes Classification. In 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST) (pp. 1-4). IEEE.

[8] P. M. Krishnammaal and S. S. Raja, "Convolutional Neural Network-based Image Classification and Detection of Abnormalities in MRI Brain Images," 2019 International Conference on Communication Signal Processing (ICCSP), Chennai, India, 2019, pp. 0548-0553.

[9] Z. Huang et al., "Convolutioanal Neural Network Based on Complex Networks for Brain Tumor Image Classification With a Modified Activation Function," in IEEE Access, vol. 7, pp. 79959-79967, 2019.

[10] G. Song et al., "A Noninvasive System for the Automatic Detection of Gliomas Based on Hybrid Features and PSO-KSVM," in IEEE Access, vol. 7, pp. 13842-13855, 2019.
[11] G. Latif, D. N. F. A. Iskandar, J. M. Alghazo and N. Mohammad, "Enhanced MR Image Classification Using Hybrid Statistical and Wavelets Features," in IEEE Access, vol. 7, pp. 9634-9644, 2019.

[12] Badža, M.M. and Barjaktarović, M.Č., 2020. Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. Applied Sciences, 10(6), p.1999.

[13] El Boustani, A., Attila, M., El Bachari, E. and El Oirrak, A., 2019, July. MRI brain image classification using convolutional neural networks. In International Conference on Advanced Intelligent Systems for Sustainable Development, pp. 308-320. Springer, Cham.

[14] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network," in IEEE Access, vol. 7, pp. 46278-46287, 2019.

[15] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," 2019 International Conference on Information Technology (ICIT), Bhubaneswar, India, 2019, pp. 94-98, DOI: 10.1109/ICIT48102.2019.00023.

[16] R, Karthick (2018). Deep learning for age group classification system. International journal of advances in signal and image sciences, 4(2), 16-22.

[17] Khan, M.A., Ashraf, I., Alhaisoni, M., Damaševičius, R., Scherer, R., Rehman, A. and Bukhari, S.A.C., 2020. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. Diagnostics, 10(8), p.565.

[18] Shekainah Justin, & Manjula Pattnaik. (2020). Skin lesion segmentation by pixel by pixel approach using deep learning. International journal of advances in signal and image sciences, 6(1), 12-20.

[19] Narayanan, K. L., & Ramesh, G. P. (2017), "Discrete Wavelet Transform Based Image Compression using Frequency Band Suppression and Throughput Enhancement." International Journal of MC Square Scientific Research, Vol. 9(2), pp.176-182.

[20] V L, Raja. (2018). Surface grading using deep learning approach for ceramic tiles. International journal of advances in signal and image sciences, 4(1), 1-7. 11.

[21] Simonyan, K., Zisserman, A. “Very deep convolutional networks for large-scale image recognition”, 3rd International Conference on Learning Representations, 2015.

[22] Murugan, S., Anjali, B., & Ganeshbabu, T. R. (2015). "Object recognition based on empirical wavelet transform," International Journal of MC Square Scientific Research, Vol. 7(1), pp.77-83.

[23] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, Volume 26, Number 6, December 2013, pp 1045-1057. (paper).

[24] Madhavan S, Zenklusen JC, Kotliarov Y, Sahni H, Fine HA, Buetow K. Rembrandt: helping personalized medicine become a reality through integrative translational research. Mol Cancer Res. 2009 Feb;7(2):157-67.

[25] REMBRANDT:https://wiki.cancerimagingarchive.net/display/Public/REMBRANDT.