MRI brain tumor classification Employing transform Domain projections

Maedeh Sadat Fasihi and Wasfy B. Mikhael, Fellow, IEEE
Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL, 32816
Emails: fasihi_m@knights.ucf.edu, Wasfy.Mikhael@ucf.edu

***Abstract*—Brain tumors are among the common types of cancers; which affect all ages, genders, and ethnicities. Early detection of the tumors increases the survival chance of the patients as well as life quality. Brain tumor treatment options depend on the type of tumors, as well as their size and location. Therefore, there is a high demand for accurate tumor classification methods. Recently, Convolutional Neural Networks (CNNs) were used for the problem of brain tumor type classification. However, CNNs typically require large amount of training data, and have large computational cost.**

In this paper we propose a fast and accurate brain tumor classifier implementation, which employs a feature extractor realized as an appropriately combined Wavelet and Discrete Cosine Transform (DCT) domain representations of the data, that outperforms the existing overly parameterized CNN. The proposed efficient, low dimensional features exploit the power of shallow deep learning models to achieve higher performance with lower computational cost. In Contrast with CNN feature extractors that require parallel computational units such as GPU, our proposed method can be efficiently run on a CPU and reduces the computational time, 10 times compared to a CNN equivalent. We evaluated the proposed method on brain tumor data set and achieved the state of the art performance with the resolution of 256×256. We also conducted a comprehensive set of experiments to analyze the effect of each component on the performance.

keywords: Brain tumors classification, Medical image classification, Machine Learning, Deep Learning

I. Introduction

Brain tumors are some of the most common and also deadliest types of tumors both in adults and children. They account for 85% to 90% of all primary CNS(Central Nervous System) tumors in the U.S. [1]. Early detection of brain tumors will facilitate the medical interventions and increase the survival rate while improving the life quality simultaneously. Survival rate is hampered by the lack of accurate diagnosis and reliable detection methods. This demands for powerful methods to address the situation and guarantee the accurate and early detection. The first step in tumor therapy is to determine the type and it's whereabouts. The signs and symptoms of a brain tumor vary greatly and depend on the brain tumor's size, location and rate of growth. The type of the tumor plays a key role in the selection of suitable therapy method. Since each type of tumor needs different attention and care, therefore algorithms which can reliably and precisely classify the tumors are of interest. Moreover, there are different types of tumors(Glioma, Meningioma, Pituitary,etc.) which makes the early detection of the tumor type a challenging task. On the other hand, classification of different brain tumors is a time consuming process, and requires expert knowledge to accurately distinguish the type of tumors. Hence, automatic classification methods have gained interest in recent years [2]–[4]. There are different approaches to tumor classification problem. These methods can be categorized into two major groups. The first group, combines the feature extraction and classification part together(e.g. most deep learning methods). The second group, separates feature selection and classification part. Our proposed method fits in the second category. First approach methods require huge number of images, which is not the case in medical applications. So the best way to tackle this problem is to go with the second approach.

In general, for tumor classification, one has to first identify the tumor region. Next, extract some features from the region of interest (tumor region). Finally, these features will be given to a classifier to determine the type of tumor. Several tumor classification methods have been proposed. In [4], a 3-level decomposition of Haar wavelet was used for feature extraction. Next, the PCA is applied on the extracted features to reduce the feature dimensions. The selected features were then given to a DNN (Deep Neural Network) to classify the tumors. This paper presents a tumor classification system based on features extracted from two domains(Discrete wavelet transform and Discrete cosine transform). The most important aspect of the selected transforms is that they provide representative (i.e., a feature which can be used to represent many other features) and sparse features. While there are many variations of neural network architectures like [5], [6], here, we use a simple but efficient MLP (Multi Layer Perceptron) network classifier to discriminate tumor types.

The rest of the paper is organized as follows: in section II a background on feature extraction algorithms is provided. The proposed approach is described in section III followed by the experimental setup and results in section IV. Finally, section V concludes the paper.

II. Background

In our method some hand crafted features are extracted from images. Next, these features are given to a classifier. The classifier maps the selected features into a high dimensional space to find a nonlinear decision boundary to precisely categorize the tumors.

The main focus of this paper is on the feature selection part. There are many hand crafted features that can be used, such as wavelet, DCT, Haralick texture features [7] to name a few. Features are selected based on their sparsity and representativity. These features serve as coefficients of the basis functions of the transformed image, where the basis functions are simply determined by the feature domain. One can project image on these basis functions and the corresponding coefficients will be the features. One challenging aspect of medical image processing is that they do not contain sharp edges. This means the organ edges and boundaries are not sharp enough to use an edge detector and this emphasizes on using texture features. Among different transforms that were used for medical image processing, wavelet transform has gained a lot of attention in the past few years [4]. On the other hand, transforms that summarize information and allow for compression are of interest. One of the most popular transformations capable of doing so is DCT. Thus combining the abilities of these two transformations together for feature extraction will be of value. The following two subsections provide more detail about these two transforms.

A. Discrete Wavelet Transform (DWT)

While the time information is lost in the Fourier transform, the DWT was found to keep both time and frequency information simultaneously. The DWT, is based on multi-resolution analysis (MRA); different frequency bands provide useful information for image processing. For medical images, global characteristics are localized in the low frequency band while the sharp edges of the image are located in high frequency bands. Since, most medical images does not contain sharp edges and boundaries, the high frequency bands are omitted and only low frequency bands are retained. When decomposing an image using DWT, a combination of low pass and high pass filters are used on rows and columns of the image. As a result, the image is divided into four equally sized sub-images(bands): one approximation, and 3 detailed images. Fig.1 shows two-level of decomposition for one of the images in the dataset. As it is shown in this figure, the majority of information is kept at the approximation image, therefore only the approximation band is kept and the rest is discarded.

B. Discrete Cosine Transform (DCT)

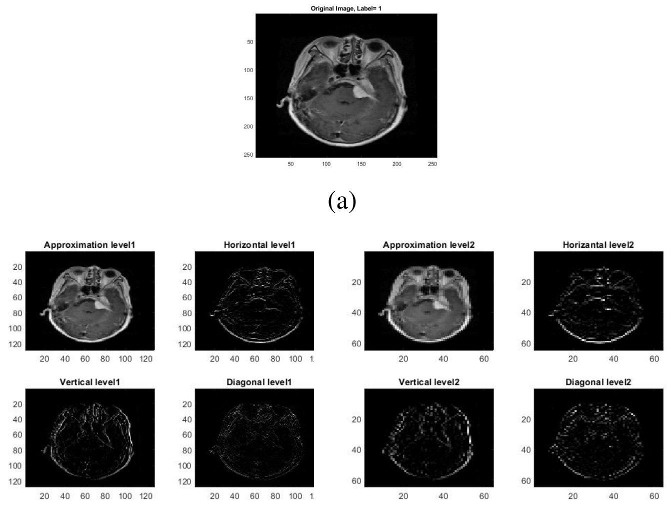
The DCT represents an image in terms of a sum of sinusoids with different frequencies and amplitudes. This transform concentrates most of the signal power in a small part of the domain. Thus, fewer coefficients are sufficient to approximate the original signal, so it is sparse in nature.

The 2D DCT can be calculated using the following equation:

$$H(m,n)=\frac{2}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}h(x,y)c(m)c(n)\cos(\alpha)\cos(\beta) \tag{1}$$

where $h(x,y)$ is the original image with dimension $M\times N$ and $H(m,n)$ is the mth row, nth column DCT coefficient for $x=0,1,...M-1$ and $y=0,1,...N-1$. Also, α,β, and $c(\gamma)$ are defined as:

$$\alpha=\frac{m(2x+1)\pi}{2M} \tag{2}$$



(a)

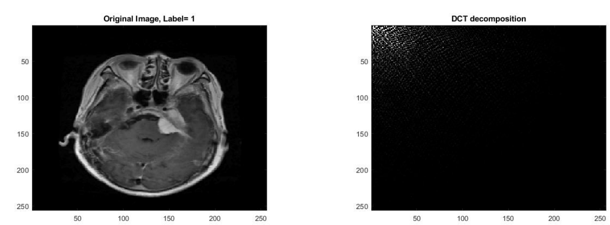
(b) (c)

Fig. 1. a) original image b)1st level DWT decomposition c) 2nd level dwt decomposition on approximation image of the first level

$$\beta=\frac{n(2y+1)\pi}{2N} \tag{3}$$

$$c(\gamma)=\begin{cases}\frac{1}{\sqrt{2}} & \text{for } \gamma=0\\ 1 & \text{otherwise}\end{cases} \tag{4}$$

Fig.2 shows the 2D DCT for an example image in the dataset.



(a) (b)

Fig. 2. a) original image b) 2D DCT decomposition

III. PROPOSED SYSTEM

The focus of this work is mostly on selecting the best sparse features using combination of transforms. Next, these features are given to a neural network classifier to determine the type of tumors. Several structures of neural network were also investigated which will be explained later in section IV.

Among different existing transformations, DWT extracts texture features from an image. Also, it provides lossless compression when applied consecutively on an image. Therefore, it is suitable for extracting the compressed textures features from medical images. Moreover, the location of the features are preserved when DWT is applied, and is useful for brain tumor localization in our application.

DCT, on the other hand is a very powerful compression transform, which supplies sparse features with less number

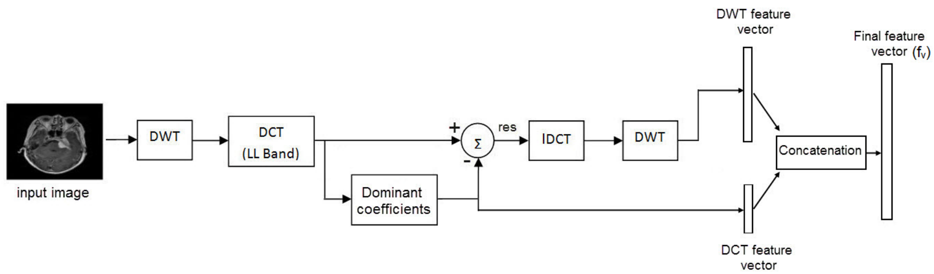


Fig. 3. Block Diagram of the proposed feature selection algorithm. This figure shows when DWT is only applied once on inverse DCT residual

of coefficients. Unfortunately, the localization property is lost utilizing DCT. Hence, in our method we are applying the DWT on the input image first, followed by DCT on top of that (in order to extract the important features of the first level). Then, to capture the rest of the textural features, DWT is applied on the residual image consecutively.

The block diagram of the proposed feature selection algorithm is shown in Fig.3. First, 2D discrete wavelet transform is applied on each image using Haar wavelet function. Then, 2D Discrete cosine transform is applied only on the LL band (low frequency band) of the wavelet transform. The dominant coefficients of the DCT are kept and the rest of the features(residual) are transformed back into the wavelet domain. The wavelet transform is then applied several times to get the DWT feature matrix. The final feature vector is the concatenation of the vectorized DCT and DWT (last LL band) matrices.

Several combinations for the dimension of feature vectors were considered. Also, the effect of applying only one transform individually in this consecutive manner was studied. Further details are provided in section IV.

After forming the feature vector, it is fed to a simple MLP neural network to classify the images.

IV. Experiments

A. Data-set

To evaluate the proposed method, we used the brain tumor data set presented in [8]. This data set contains 3064 MRI images (T1-weighted contrast-enhanced) with the resolution of 512×512 from 233 patients. For the sake of reducing the computational cost, we resized the images to the resolution of 256×256. Each patient is diagnosed with one of the three kinds of brain tumors; Meningioma, Glioma, and Pituitary tumor. Since these 3 types of tumors occur in different locations of brain, and the size and shape of the tumors vary among patients; Therefore, the tumor classification is a challenging task.

Our experiments consists of two parts: Feature selection, and choosing the appropriate classifier.

B. Feature selection

For the sake of clarity we define the notation used through out the section. The feature vectors are referred to as C_pW_q where C stands for DCT transform and W stands for DWT transform; p, and q are the lengths of the transform vectors in DCT and DWT respectively. Also, $p \in \{25, 100\}$, and $q \in \{64, 256\}$, We also use N in the case that a particular transform is not applied. For instance, $C_{25}W_N$ denotes the length of the DCT vector is 25, and no DWT is applied.

First, we investigate the effect of each of DCT, DWT and the combination of both of them; as shown in Fig.3; on the performance. Table I represents the classification accuracy based on each of these transformations. As it can be seen the combination of transformation improves the classification accuracy compared to one of the transformations.

C. Classifier

In the second part, we have utilized dense multilayer neural network classifier. The model has been trained using Adam optimizer [9] with learning rate of 0.005 with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Our model is implemented based on Keras framework [10] using Python 2.7. Rectified linear unit (ReLu) is applied as the activation function on all layers except the last one. To avoid over-fitting on the training data, drop out with ratio of 0.2 is adopted on the first 2 dense layers. The performance is reported based on 5 fold cross validation on the whole dataset.

For the best combination of DWT and DCT feature ($C_{100}W_{64}$) different architectures of classifier in terms of number of neurons, layers, and dropout ratio were studied. The result of our ablation study on the number of layers;(3-layer, 4-layer, and 5-layer); are summarized in the Table II. The best accuracy is achieved using the $C_{100}W_{64}$ feature vector as input to a 3 layer dense neural network as the classifier. The 3-layer dense neural network is shown in Fig.4. The architecture of this classifier is as follows:

- Two Fully connected layers with 150, 90 neurons for first and second layer respectively. Each layer is then followed by dropout and batch normalization
- The last layer, is fully connected layer with 3 neurons, to classify the brain tumors.
- number of epochs for training was set to 200, and the learning rate 0.005 is selected.

Moreover, we compared our method with an end to end CNN model proposed in [11]; also used for the same brain tumor classification problem. This model served as the baseline in our experiments. The architecture of the model in [11], is as follows:

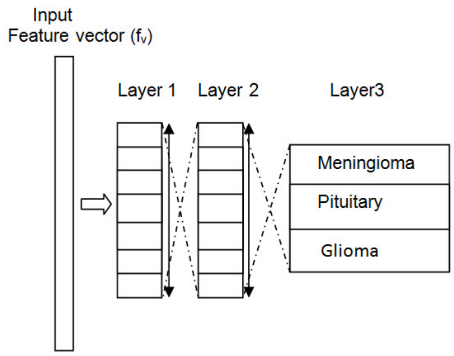


Fig. 4. Architecture of the NN classifier (best result)

TABLE I
BRAIN TUMOR CLASSIFICATION ACCURACY BASED ON DIFFERENT FEATURE VECTORS

Features	Average Accuracy	Best Accuracy
$C_N W_{64}$	78.07%	81.9%
$C_{100} W_N$	78.39%	80.99%
$C_{100} W_{64}$	84.12%	86.98%

- Two convoloutional layers each with $64 \times 5 \times 5$ filters and stride of 1, each layer is then followed by 2×2 Max-pooling layer
- Two Fully connected layers each with 800 neurons
- The last layer(fifth layer) is fully connected layer with 3 neurons, to classify the brain tumors.

TABLE II
EFFECT OF CHANGING CLASSIFIER ARCHITECTURE ON ACCURACY

Architecture	parameters	Average Accuracy	Best Accuracy
3 layer	38370	84.12%	86.98%
4 layer	46290	82.42%	83.75%
5 layer	51690	80.63%	82.41%
CNN [11]	$210 * 10^6$	61.97%	65.39%

Relu activation function is used for layers and Adam optimizer is utilized for training the Neural Network using the back propagation method; Batch normalization has been applied to stabilize the training process and facilitate the convergence.

The number of parameters of our proposed classifier is about 3 orders of magnitude less than the baseline, and the relative processing time is reduced by 90% for our method compared to that. The relative processing time reduction equation is shown in equation 5. Python environment is used to setup experiments for computing processing time.

$$\text{Relative Processing Time Reduction} = 1 - \frac{\text{Processing Time of proposed System}}{\text{Processing Time of Baseline}} \quad (5)$$

This proves that our method not only outperforms the baseline in terms of accuracy, but is also computationally efficient, and considerably faster. Moreover, based on the experiments, due to the limited number of samples, when the feature vector becomes large the classifier accuracy drops. Besides, when we increase the number of classifier layers, the accuracy will also drop, since the number of parameters to be estimated exceeds the available information. The best result was obtained when $C_{100} W_{64}$ feature vector along with 3 layer dense neural network was used. The DCT part in the feature vector provides the important features of the approximation image, while the DWT part of the feature vector provides texture features not captured by DCT.

V. CONCLUSION

In this paper, a brain tumor classification system was proposed. The system is based on utilizing an optimal and compressed set of features from DWT and DCT domains. This set is then fed into the NN classifier. Various options were investigated both in feature selection and classifier architecture. The results of the study indicates that the proposed system that is using the combination of DWT and DCT increases the accuracy without significant time overhead compared to employing each of the transforms individually. Moreover, the proposed system outperforms the baseline [11] both in accuracy and processing time, being 10 times faster. The significant improvement in processing time was achieved because of the feature selection algorithm which provides sparse and optimal features. In addition, this reduces the complexity of the NN classifier to 3 layers(which is comparatively small). Overall, the proposed method is fast, accurate, and considerably reduces parameters. Moreover, the technique presented here can be implemented on handheld devices for medical applications.

REFERENCES

[1] PDQ Adult Treatment Editorial Board, "Adult central nervous system tumors treatment (pdq®)," in *PDQ Cancer Information Summaries [Internet]*. National Cancer Institute (US), 2019.

[2] Taranjit Kaur, Barjinder Singh Saini, and Savita Gupta, "Quantitative metric for mr brain tumour grade classification using sample space density measure of analytic intrinsic mode function representation," *IET Image Processing*, vol. 11, no. 8, pp. 620–632, 2017.

[3] Sahar Tavakoli and Emad Fatemizadeh, "Decoding the long term memory using weighted thresholding union subspaces based classification on magnetoencephalogram," in *International Symposium on Artificial Intelligence and Signal Processing*. Springer, 2013, pp. 164–171.

[4] Heba Mohsen, El-Sayed A El-Dahshan, El-Sayed M El-Horbaty, and Abdel-Badeeh M Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.

[5] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, 2017, pp. 3856–3866.

[6] Marzieh Edraki, Nazanin Rahnavard, and Mubarak Shah, "Subspace capsule network.," in *AAAI*, 2020, pp. 10745–10753.

[7] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, , no. 6, pp. 610–621, 1973.

[8] Jun Cheng, Wei Yang, Meiyan Huang, Wei Huang, Jun Jiang, Yujia Zhou, Ru Yang, Jie Zhao, Yanqiu Feng, Qianjin Feng, et al., "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS one*, vol. 11, no. 6, pp. e0157112, 2016.

[9] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *3rd International Conference for Learning Representations, San Diego*, 2015.

[10] Marcus D Bloice and Andreas Holzinger, "A tutorial on machine learning and data science tools with python," in *Machine Learning for Health Informatics*, pp. 435–480. Springer, 2016.

[11] Justin S Paul, Andrew J Plassard, Bennett A Landman, and Daniel Fabbri, "Deep learning for brain tumor classification," in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*. International Society for Optics and Photonics, 2017, vol. 10137, p. 1013710.