November 11, 2022

Dr. Kostadinka Bizheva

Associate Editor, Biomedical Optics Express

OSA

Re: Revised Manuscript ID **477311**

Dear Dr. Bizheva:

We would like to thank you and the reviewers for the constructive comments and suggestions on our manuscript entitled **“Deep learning-based optical coherence tomography image analysis of human brain cancer,”** and for the opportunity to further improve our manuscript. We assure that all the comments and suggestions have been carefully read and addressed as adequately as we can in the revised manuscript.

For your convenience, we have highlighted our revisions in red in the revised manuscript, and also provided point-by-point responses to the reviewers’ comments below.

Again, we would like to thank you for your kind consideration.

Sincerely,

Xingde Li, PhD

Professor of BME, ECE, and Oncology

Johns Hopkins School of Medicine

[Xingde@jhu.edu](mailto:Xingde@jhu.edu)

**Comments and Responses**

**Responses to Comments from Reviewer 1**

**Comment 1**:

What is the key innovation of this study, the use of texture and Bscan features for CNN?

**Response**:

We thank the reviewer for the question. Our use of an ensemble neural network which incorporates both B-frames and textures is novel, improving accuracy and robustness. Importantly, we report high performance without the need for reference phantom normalization. Overall, we believe that this study is an important step to show that deep networks can detect cancer robustly. Hence, deep networks have the potential to overcome challenges such as motion artifacts and image defocusing. Regarding on the reviewer’s comment, we revised the manuscript elaborating the innovations in this work.

Page 1, line 25, Abstract:

“… without the need for beam profile normalization using a reference phantom.”

Page 2, line 55-57, Introduction:

“Previous studies have employed machine learning methods on hand-crafted textures or deep neural networks for prediction on the B-frame, but our proposed method synergizes both.”

**Comment 2**:

Compared to existing machine learning methods including other deep CNNs, how is the

method described by the author more advantageous?

**Response**:

We appreciate the Reviewer's excellent comment. Our paper puts our work in perspective with respect to the literature. We cite a closely related study which classifies human brain cancer using an SVM on hand-crafted texture features. In contrast, deep learning is an attractive paradigm which can implicitly learn feature representations by training. These feature representations can be robust and have important potential for this application, please also refer to our response to Comment 1. Another related study employs a convolutional neural network for detecting cancer infiltration in white matter with 91% accuracy (cited on line 57 in the revised manuscript). In contrast, our approach is more advanced in terms of its hybrid architecture, superior accuracy, and increased level of confidence.

**Comment 3**:

It seems the samples were ex vivo specimens taken from patients. Please specify the sample acquisition procedure and the criterion used to recruit the subjects. How fresh were the samples? How were the samples preserved, in salient solution or in formalin? How were the samples acquired, through biopsy procedures or through surgery? Did all the patients have the same cancer

**Response**: We would thank the Reviewer for this important comment. The fresh human tissue samples were acquired from patients undergoing brain cancer surgery at the Johns Hopkins Hospital under an approved Institutional Review Board protocol. The ex vivo images were taken within an hour after the surgical removal while kept in cold salient solution to keep it fresh. The above information was added to the revised manuscrips.

Page 2-3, lines 79-82, Methods and Materials:

“Fresh human brain tissue samples were obtained at the Johns Hopkins Hospital under an approved Institutional Review Board protocol. The samples were contained in a cold salient solution and ***ex vivo*** images were taken within one hour of surgical removal of the tissues.”

**Comment 4**:

Data processing. Please specify if logarithm data or linear data is used for the training.

**Response**:

We thank the reviewer for the helpful suggestion. We have clarified that logarithmic data is used on line 82 under section 2.1.

Page 3, line 94, Methods and Materials:

“In this study, logarithmic (base 10) B-frame data is used.”

**Comment 5**:

There are many methods for texture analysis. Please explain why grey level co-occurrence matrix was selected to analyze OCT images of brain tissue. Or refer to published results.

**Response**:

We thank the reviewer for the helpful suggestion. We have added some clarification and cited an OCT study which successfully utilized GLCM features for skin disease diagnosis.

Page 3, lines 12-114, Methods and Materials:

“The GLCM is a classic approach for quantifying the distribution of co-occurring pixel values that has been applied to extract discriminative features in many medical imaging applications, including OCT”

**Comment 6**:

Please justify the selection of CNN architecture. Is there a particular reason to choose the specific layers to build the CNNs?

**Response**:

We thank the reviewer for the question and comment. Subsection 2.3 describes the general principle we considered when designing the networks (e.g., smaller feature maps are balanced by a greater number of them to preserve visual information). We have also added a statement that hyperparameter tuning was done through empirical experimentation.

Page 4, lines 160-162, Methods and Materials:

“All hyperparameter values (layer size, feature maps, optimizer, loss, learning rate, etc.) were selected through empirical experimentation following general good practices as briefly described in this section”

**Comment 7**:

In addition the CNN design, the training strategies are also critical for the performed of the network. Please provide detailed information on the training (learning rate, choice of solver, …).

**Response**:

We really appreciate the reviewer’s valuable comment. We have included the optimizer and learning rate of the network starting.

Page 3, lines 152-153, Methods and Materials:

“Stochastic gradient descent was used as the optimizer with a learning rate of 0.001 for the “B-frame slices CNN” and a learning rate of 0.01 for the “texture CNN” and the overall ensemble”

**Comment 8**:

What is the software platform used for training? How long did the training take?

**Response**:

We thank the reviewer for the question. The study was conducted in Python/Tensorflow and our code is available upon request and will be shared on GitHub. Our hardware is detailed on lines 162-163 and 167-168.

**Responses to Comments from Reviewer 2**

**General Comment**:

This paper presents a novel ensemble based deep learning model to classify cancerous and non-cancerous brain tissues. The paper trained the network using 5831 images and corresponding texture images generated by grey level co-occurrence matrix (GLCM) from 7 patients. The trained model was then tested on an independent dataset from 4 patients with impressive accuracy (93% sensitivity and 97% specificity). The study design is rigorous, and the organization of paper is clear. The clinical results are significant in improving neurosurgery in real time. This paper can be accepted after addressing the following minor comments:

**Response:**

Thank you so much for your encouragement and summary!

**Comment 1**:

In this study, how did the deep learning hyperparameters be tuned?

**Response**:

We would like to thank the reviewer for bringing up the question. We have clarified that hyperparameter tuning was made empirically (lines 160-162). Specifically, the neural networks were designed according to well-established general principles (such as increasing the number of feature maps as the size of each feature map becomes smaller), which are described in Subsection 2.3.

See our response to Reviewer 1 Comment 6.

**Comment 2**:

In section 2.3, ensemble learning, is it stacking ensemble? If yes, it should be clearly mentioned in the manuscript.

**Response**:

We thank the reviewer for the comment. In our study, we have two separate cancer-classifying networks (one trained on B-frames, the other trained on textural features). Then, their feature embedding (the penultimate layer) is concatenated and fed into a multi-layer perceptron which predicts the likelihood of cancer. This is a commonly employed strategy in deep learning which is analogous to (but not exactly the same as) the stacking ensemble.

**Comment 3**:

Have authors compared the ensemble results with regular CNN without stacking? How much improvements can be achieved with/without ensemble?

**Response**:

We thank the reviewer for the question. In table 1, we provide a statistical analysis of the ensemble model compared to its individual network components. In Subsection 3.2, we have reported that the ensemble achieves a statistically significant increase in confidence on the predictions (quantified using dropout) and also significantly greater accuracy when early stopping was used on the texture CNN.

**Comment 4**:

Since two different CNNs with different input dimensions fed to the network, I assume the number of trainable parameters for each model should differ. Have authors used same #epochs, batch size, … for both models?

**Response**:

We thank the reviewer for the helpful comment. Yes, the size of the models (number of layers and parameters) is different. The batch size is the same, but the number of training epochs is not. Figure 2A shows the training curve for the relevant networks which shows the final epoch for each model. We have also clarified that training was stopped if no improvement in validation accuracy was observed for 10 epochs.

Page 5, lines 180-182, Results and Discussion:

“As seen in Figure 3A, not all networks were trained for the full 50 epochs as early stopping was implemented if there was no improvement in validation accuracy for 10 epochs.”

**Comment 5**:

What learning rate and optimizer were used for models?

**Response**:

We thank the reviewer for the comment. We have included the optimizer and learning rate of the network starting on lines 152-153.

See our response to Reviewer 1 Comment 7.

**Comment 6**:

The performance of the proposed model should be compared with the similar works in the literature, especially the prior study reported by the authors.

**Response**:

We thank the reviewer for the comment. We have expanded our analysis in the conclusion (starting on line 237) to explicitly compare our prior method that is based on attenuation mapping to our deep network model presented in this paper. In essence, our deep learning model takes image spatial and textural features into account and enjoys several merits such as real-time processing and not requiring reference phantom normalization. Along this direction, we plan to perform a follow-up study with more patient data under diverse conditions (image defocusing, motion artifacts, etc.) in the near future. Additionally, please see our response to reviewer 1 comment 2 where we compare our method to another deep learning study.