2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-7358-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISBI53787.2023.10230698

# MOAB: MULTI-MODAL OUTER ARITHMETIC BLOCK FOR FUSION OF HISTOPATHOLOGICAL IMAGES AND GENETIC DATA FOR BRAIN TUMOR GRADING

*Omnia Alwazzan*[⋆†] , *Abbas Khan*[⋆†], *Ioannis Patras*[⋆†] , *Gregory Slabaugh*[⋆†]

[⋆]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
[†]Queen Mary's Digital Environment Research Institute (DERI), London, UK

## ABSTRACT

Brain tumors are an abnormal growth of cells in the brain. They can be classified into distinct grades based on their growth. Often grading is performed based on a histological image and is one of the most significant predictors of a patient's prognosis; the higher the grade, the more aggressive the tumor. Correct diagnosis of the tumor's grade remains challenging. Though histopathological grading has been shown to be prognostic, results are subject to interobserver variability, even among experienced pathologists. Recently, the World Health Organization reported that advances in molecular genetics have led to improvements in tumor classification. This paper seeks to integrate histological images and genetic data for improved computer-aided diagnosis. We propose a novel Multi-modal Outer Arithmetic Block (MOAB) based on arithmetic operations to combine latent representations of the different modalities for predicting the tumor grade (Grade II, III and IV). Extensive experiments evaluate the effectiveness of our approach. By applying MOAB to The Cancer Genome Atlas (TCGA) glioma dataset, we show that it can improve separation between similar classes (Grade II and III) and outperform prior state-of-the-art grade classification techniques.

***Index Terms*—** Multi-modal fusion, Outer-arithmetic fusion, Cancer grade classification, Channel fusion, Brain tumor

## 1. INTRODUCTION

With the advent of artificial neural networks, many advanced medical imaging algorithms have been proposed to analyze histopathological cancer images for grade classification [1, 2] and survival prediction [3, 4]. Digitized histopathology provides whole slide images (WSI) and related phenotypic information. Literature suggests both single-modality [5] and multi-modality [6] methods for histopathological image analysis. However, approaches based on a single modality are limited and unable to exploit the complementary information present in other modalities for cancer prognosis prediction [7, 8]. This provides an opportunity to integrate multi-modal data for more precise diagnosis.

A recent study [9] integrated 'omic data with image modalities using attention gating and a combination of concatenation and arithmetic operation-based fusion approaches. Braman et al. [9] presented an intermediate fusion deep multi-modality network that integrated radiology, pathology, genetics, and clinical data to predict glioma patients. Deep learning models were used in [9] to extract features from each modality and generate its corresponding embedding layer. Using the attention gated mechanism and an outer product operator, the authors were able to fuse significant features to improve prediction.

A multi-modal fusion method was proposed by [10] to learn the combined feature representations of histopathological images and mRNA expression, through ResNet-152 and a sparse graph convolutional network respectively. The features were merged using a shared multilayer perceptron to predict survival analysis and cancer grade. Simailarly, Pathomic Fusion [6], an integrative image-omic analysis method used a gated attention mechanism and a Kronecker product to combine features of different modalities. The image features were extracted using a combination of convolutional and graph convolutional neural networks, and a feed-forward network was trained separately to learn genomics features. Mobadersanya et al. [11] proposed survival convolutional neural networks (SCNNs) to merge a histology image and genomic biomarkers.

We note that previous work [6, 10] focused on Grade IV and did not consider Grade II and III. These two grades are minorities in the dataset compared to Grade IV and their similarity hinders typical classification techniques from detecting them accurately. However, Grade III is considered malignant and as invasive as Grade IV [12] which could make its classification as crucial as Grade IV. Hence, in this study, we show how to integrate rich features from genetic data and histology images to improve grading of tumors.

We propose a novel fusion method that captures the interrelated features between disparate modalities for improved classification. An overview of our approach is presented in Fig. 1. Our paper makes the following contributions:

- We introduce a novel Multi-modal Outer Arithmetic Block (MOAB) that fuses latent representations of different modalities using arithmetic operations.
- We demonstrate MOAB's ability to capture rich representations of fused data in a brain cancer application using a histology image and genetic data. Our method outperforms the previous state-of-art, by producing more discriminative fused features for classification. This is also demonstrated with t-SNE plots which visualize the improved separability of classes thanks to MOAB.
- Source code for all proposed fusion methodologies are available at https://github.com/omniaalwazzan/MOAB

## 2. METHOD

From paired histology and genetic data with known cancer grades, we aim to learn and combine informative features that outperform single-modality classifiers in supervised learning. For this purpose, we propose a novel method for effective integration of multi-modal data.

We first provide a brief overview of the single modality backbone that we use, followed by in-depth explanations of the architecture of two components: outer arithmetic fusion methodologies and our proposed MOAB.

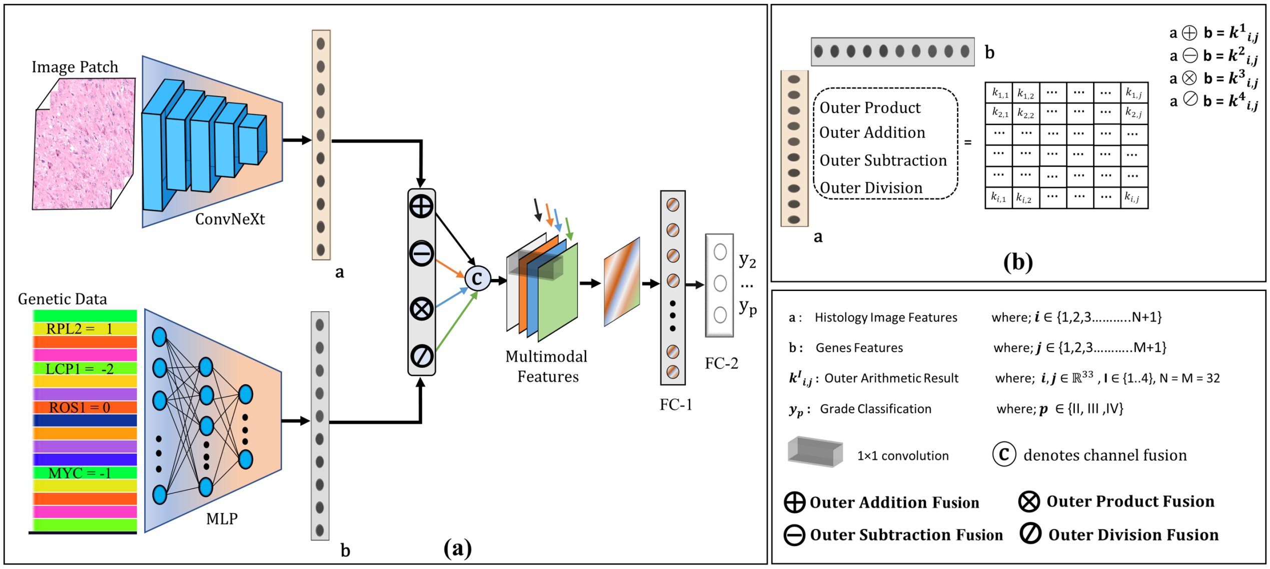
Image Patch
ConvNeXt
a
Genetic Data
RPL2 = 1
LCP1 = -2
ROS1 = 0
MYC = -1
MLP
b
c
Multimodal Features
FC-1
FC-2
y2
...
yp
(a)
b
a
Outer Product
Outer Addition
Outer Subtraction
Outer Division
=
$k_{1,1}$ $k_{1,2}$ ... $k_{1,j}$
$k_{2,1}$ $k_{2,2}$ ... $k_{2,j}$
$k_{i,1}$ $k_{i,2}$ ... $k_{i,j}$
a ⊕ b = $k^1_{i,j}$
a ⊖ b = $k^2_{i,j}$
a ⊗ b = $k^3_{i,j}$
a ⊘ b = $k^4_{i,j}$
(b)
a : Histology Image Features where; $i \in \{1,2,3..........N+1\}$
b : Genes Features where; $j \in \{1,2,3..........M+1\}$
$k^I_{i,j}$ : Outer Arithmetic Result where; $i,j \in \mathbb{R}^{33}$ , I ∈ {1..4}, N = M = 32
$y_p$ : Grade Classification where; $p \in$ {II, III ,IV}
1×1 convolution
C denotes channel fusion
⊕ Outer Addition Fusion
⊗ Outer Product Fusion
⊖ Outer Subtraction Fusion
⊘ Outer Division Fusion

**Fig. 1**. Architecture of our proposed approach Multi-modal Outer Arithmetic Block (MOAB) fusion model. (a) MOAB (b) Generic outer arithmetic approach.

### 2.1. Single modality classifiers

We initially created a single modality classifier for each data type (i.e. histology image and genetic data).

For the histology images, a pre-trained ConvNeXt initialized with ImageNet weights was used. ConvNeXt is a recently proposed [13], state-of-the-art CNN demonstrating how convolutional backbones can outperform transformers in image classification [14]. We replace the LayerNorm2D of the ConvNeXt classifier layer with a dropout = 0.2 followed by a fully connected (FC) layer that outputs three values representing the brain tumor grades (II, III and IV). We fine-tune ConvNeXt on the histology images. We observe that ConvNeXt is more robust than the VGG19 backbone implemented in [6] in terms of generalizing well to the histology images after a few iterations.

A 3 consecutive blocks of FC layers with dimensions [80, 40, 32] forming our multi-layer perceptron (MLP) was utilized to classify the genomic data. Each layer in the MLP is followed by a rectified linear unit (ReLU) activation and a layer normalization. After the second and third layers, a gentle dropout of 0.2 was applied. According to the results presented in Table 1, the MLP achieves comparable performance to the ConvNeXt applied to the histological images.

### 2.2. Outer Arithmetic Operations

Our work is based on Outer Product Fusion (OPF), which has been used previously [9] to combine feature vectors of different modalities. Consider two feature vectors $\mathbf{a} \in \mathbb{R}^{N\times 1}$ and $\mathbf{b} \in \mathbb{R}^{M\times 1}$. We first append a 1 to each feature vector, i.e. $\mathbf{a}_1 = [1;\mathbf{a}]$ and $\mathbf{b}_1 = [1;\mathbf{b}]$. The OPF is expressed as

$$(\mathbf{a}_1 \otimes \mathbf{b}_1)_{ij} = a_{1i} * b_{1j}, \tag{1}$$

for $i \in [1...N+1]$ and $j \in [1...M+1]$ resulting in an $(N+1)\times(M+1)$ matrix combining all pairs of features. The appended 1 in both $\mathbf{a}_1$ and $\mathbf{b}_1$ ensures the original $\mathbf{a}$ and $\mathbf{b}$ vectors appear in the outer product matrix.

We extend the OPF by proposing the use of three new additional arithmetic operations: *Outer Addition Fusion (OAF)*, *Outer Subtraction Fusion (OSF)* and *Outer Division Fusion (ODF)*. The ODF is expressed as:

$$(\mathbf{a}_1 \oslash \mathbf{b}_1)_{ij} = a_{1i}/(b_{1j}+\epsilon), \tag{2}$$

where $\epsilon$ is a small number to avoid division by zero.

For the outer addition and outer subtraction fusion, we append a 0 to each feature vector, i.e. $\mathbf{a}_0 = [0;\mathbf{a}]$ and $\mathbf{b}_0 = [0;\mathbf{b}]$. Then, the OAF is defined as

$$(\mathbf{a}_0 \oplus \mathbf{b}_0)_{ij} = a_{0i} + b_{0j}, \tag{3}$$

and OSF as

$$(\mathbf{a}_0 \ominus \mathbf{b}_0)_{ij} = a_{0i} - b_{0j}, \tag{4}$$

As before, the appended 0 in both $\mathbf{a}_0$ and $\mathbf{b}_0$ ensures the original $\mathbf{a}$ and $\mathbf{b}$ vectors appear in the outer addition and subtraction matrices. Each of the four outer arithmetic fusion operators will produce a matrix of size $(N+1)\times(M+1)$.

### 2.3. Multi-modal Outer Arithmetic Block Fusion

The MOAB fusion model uses the previously explained outer arithmetic fusion operations to capture essential features from the genomic and histology data. We perform the fusion on a four branch scheme. First, we obtain latent representation feature vectors $\mathbf{a}$ and $\mathbf{b}$ from histology and genomic data, respectively. As illustrated in Fig 1(a), $\mathbf{a}$ was extracted from the customized ConvNeXt network used previously for histology image classification. Inspired by [6], we obtained 32 features by modifying the number of the output dimension of the last FC layer from the ConvNeXt. Similarly, 32 features were extracted from the last FC layer from our MLP to represent $\mathbf{b}$.

Next, each branch will take $\mathbf{a}$ and $\mathbf{b}$, perform the four arithmetic operations to produce four multi-modal feature maps, $A_{i,j}$, $S_{i,j}$, $P_{i,j}$, and $D_{i,j}$, that represent the OAF, OSF, OPF and ODF respectively. Moreover, we utilized a sigmoid function to compress our feature maps between [0,1]. The sigmoid is employed due to the addition of the $\epsilon$ = 1.2e−20 avoiding zero division in the ODF branch. The sigmoid was applied to all four branches to maintain compatible fusion between all branches.

The four matrices $A_{i,j}$, $S_{i,j}$, $P_{i,j}$, and $D_{i,j}$, are concatenated along the channel dimension into a multi-modal tensor $M_{ij} \in \mathbb{R}^{4\times33\times33}$. Then, a 2D convolution layer is applied with a filter $f = 1$ and a stride $s = 1$ to take advantage of interrelated interactions and produce one condensed multi-modal feature tensor $M^{*}_{ij} \in \mathbb{R}^{1\times33\times33}$. We hypothesize that channel fusion will maintain the proximity of closer points and will use fewer parameters compared to a typical concatenation. By combining features across the channel dimension, we massively decrease the dimension of the final FC layer from 4356 to 1089 in terms of concatenation. It has been noted that $f = 3, 5$ leads to a decrease in the model's performance. Hence, we avoid using filter $f > 1$.

Finally, $M^{*}_{ij}$ is passed to two FC layers to perform the final classification. A mild dropout=0.1 is employed before the final layer to avoid overfitting. We also adopt the orthogonal weight initialization introduced in [6] for the MLP component in the proposed architecture.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset

The glioma dataset was preprocessed as proposed by Chen et al. [6], and provides paired histology images and gene expression derived from The Cancer Genome Atlas (TCGA) repository. The TCGA is a cancer database containing paired high-throughput genome diagnostic whole slide images with ground-truth histological grade labelling. There are 769 patients belong to 1505 histological region-of-interests (ROIs) for astrocytomas and glioblastomas in the TCGA-GBMLGG project. The selection of ROI derived from histopathological whole slide images (WSIs) was generated, revised, and reviewed by [11]. Each patient had 1-3 ROIs from diagnostic WSI. For the genomic data, 80 features include 79 copy number variations (CNVs) and one mutation status. Similar to [6], we removed cases with missing grades, resulting in 396, 408, and 654 ROIs for Grade II, III and IV, respectively, and corresponding to 736 cases. As in [6], ROIs are used for training models, while 9 overlapping patches extracted per each ROI derived from the testing cohort were used for testing. Each image was regarded as a single data point as in [6], with genetic and ground-truth label data copied over.

### 3.2. Implementation Details

Proposed algorithms were implemented in PyTorch using the Adam optimizer and a cross-entropy loss function for all experiments. For the single modalities, we empirically found the best learning rate of 0.001. While for fusion models, to avoid overfitting, we regularized our networks with a weight decay of size 0.0005 and a 0.005 learning rate. A batch size of 8 was set for all experiments. All models were trained for 10 epochs. The parameters for ConvNeXt, MLP and MOAB are 87M, 11K and 88M respectively. A cluster of NVidia A100 GPUs was utilised for all experiments.

### 3.3. Ablation Study

As shown in Table 1, we validate the performance of MOAB fusion through multiple ablation studies. In an identical Monte Carlo 15-fold cross-validation train-test split obtained from [6, 11], our ablation studies compare different model configurations and fusion methodologies. In addition, as highlighted in Fig. 2, a (t-SNE) algorithm [15] was applied to the inference split for ablated fusion models to visualize the high-dimensional relationships between the three learned grades (II, III and IV) in three-dimensional space. In t-SNE plots, nearby dots represent comparable samples, whereas distant dots reflect dissimilar ones. We empirically set an optimal perplexity and a max iteration to 80 and 1000.

**Table 1**. Ablation studies.

| Method | F1 (Grade IV) | F1-Micro | F1-Macro |
|---|---|---|---|
| CNN (Image) | 0.888 | 0.715 | 0.586 |
| MLP (Genes) | 0.901 | 0.700 | 0.626 |
| Concatenation | 0.910 | 0.726 | 0.658 |
| OAF | 0.944 | 0.740 | 0.675 |
| DBF | 0.916 | 0.727 | 0.660 |
| Standard Addition* | 0.933 | 0.730 | 0.665 |
| MOAB (VGG19_bn+MLP)) | 0.945 | 0.752 | 0.689 |
| **MOAB (ConvNext+MLP)** | **0.956** | **0.766** | **0.697** |

At first, single classifiers for both gene and image data were constructed to draw a fair comparison to multi-modality fusion. Next, we show results for a simple concatenation fusion model to further investigate the ability of other fusion methodologies. To explore the capability of MOAB, we produced two ablated fusion models: the Outer Addition Fusion (OAF) and a Dual-branch channel Fusion (DBF). Note that OAF has no channel fusion, features are combined as shown in Fig. 1(b) performing $A_{ij} \in \mathbb{R}^{33\times33}$ and followed by FC layers similar to MOAB. However, DBF imitates the same configuration as MOAB but with two branches: OAF and OPF. From Table 1 and Fig. 2(b) we observe that OAF performs better than the DBF even with a simple integration. This might occur due to the redundant features obtained from OAF, and OPF. Yet, DBF achieves similar metric scores (F1 and F1-Micro) to that of concatenation, the separated cluster shown in Fig. 2(c) is slightly better than those of Fig. 2(d), which might reflect the effect of OAF operation on capturing correlated features.

As demonstrated in Table 1, the best results were obtained utilising MOAB. This is further supported by Fig. 2(a), which shows MOAB fusion has the most separated classes among all other ablated fusion models. By comparing Fig. 2(a) with Fig. 2(f) we can clearly inspect the positive effect of integrating features from genetic data and histology images. Table 1 summarizes that utilizing all four branches in the channel fusion has significant improvements on the classifier performance namely F1-Micro and F1-Macro. F1-micro score is calculated from the micro-averaged precision (miPr), and a micro-averaged recall (miRe) and defined as the following:

$$\begin{aligned} miPr &= \frac{\sum_{i=1}^{r} TP_i}{\sum_{i=1}^{r}(TP_i + FP_i)} \\ miRe &= \frac{\sum_{i=1}^{r} TP_i}{\sum_{i=1}^{r}(TP_i + FN_i)} \end{aligned} \tag{5}$$

The $TP_i$, $FP_i$ $FN_i$ denote true positive, false positive and false negative rates at a cell $i$ in the confusion matrix. Then the micro-averaged F1 score (miF1) is defined as the mean of $miPr$, $miRe$ values.

We use the F1-micro score as the primary metric to evaluate our results, following [10],[6], as it is an appropriate evaluation metric

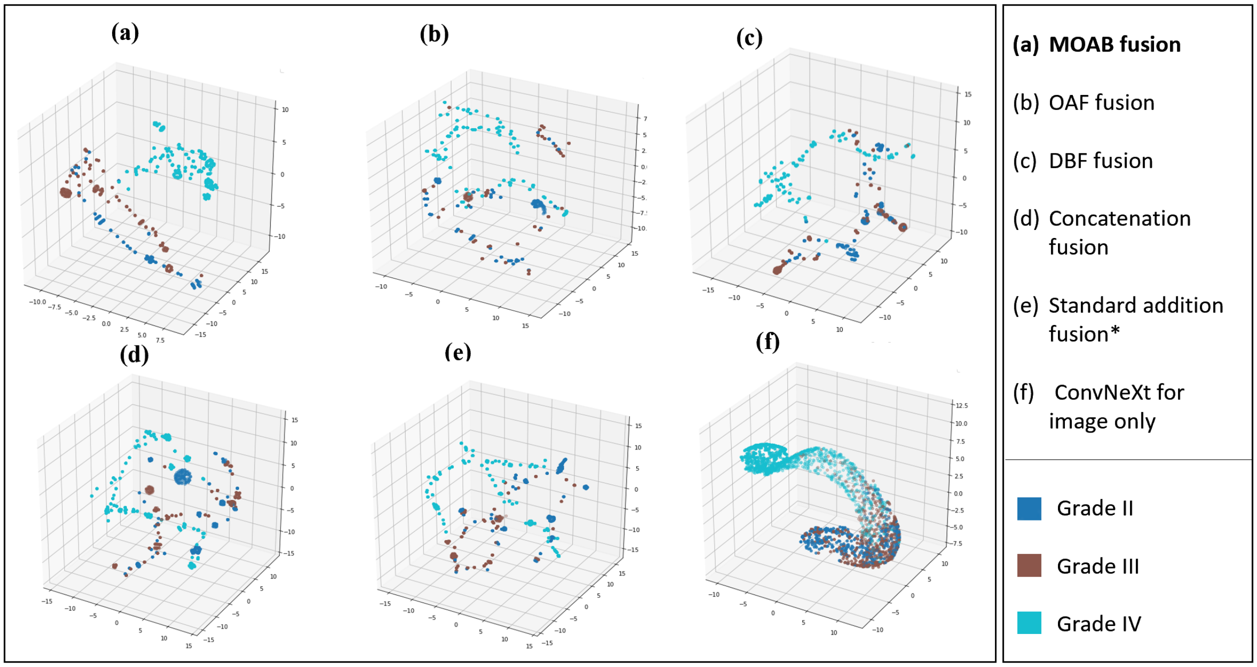

**Fig. 2**. t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization for fusion models. Note in (a), the MOAB fusion provides better separability between brain tumor grades.

for imbalanced class data. Notably, F1-Grade IV is an easier classification task given that it dominates the majority of our dataset; still, we include results for comparison purposes. Furthermore, to quantify the performance of MOAB, we calculate the F1-macro average which treats classes equally and gives a sense of MOAB's effectiveness on minority classes. The macro averaging method is more challenging for our task because it is more strongly influenced by the performance of rare classes [16]. Table 1 proves that MOAB manages to achieve strong performance. Finally, to conclusively say OAF is preserving the interlarded features between modalities, we conduct an experiment for a standard addition fusion model shown in Table 1, where we increase the parameters of the standard addition (denoted as an *) to have roughly the same learnable parameters as the OAF. Results of the standard addition exhibited in Table 1 and Fig. 2(e) indicate that OAF outperforms standard addition fusion*.

We have also switched the backbone of the MOAB from ConvNxet to the VGG19 with batch normalization used in [6], again, our obtained results outperform the Pathomic fusion [6] approach. We conclude that MOAB is improving the classifier thoroughly compared to all other configurations presented in Fig. 2. We observed that MOAB is simple to implement and extracts rich interaction combining modalities.

### 3.4. Comparison Experiments

We compared our method with the Pathomic fusion [6] and a MultiCoFusion [10] on the glioma dataset. The results are shown in Table 2. The TCGA glioma dataset has been used with 15-folds in both [6] and [10], however, the approach used in [10] is slightly different than ours, in which they additionally use mRNA expression data using a graph CNN. According to the ablation studies shown in Table 1 and the comparison presented in Table 2, the F1-Micro of MOAB is superior to that of the other models. Experiments were performed 15 times; the results shown in the tables represent the mean of the total of all evaluation metrics for 15-folds, as demonstrated in [6] and [10].

**Table 2**. Comparison experiments.

| Model | F1 (Grade IV) | F1-Micro |
|---|---|---|
| Pathomic SNN [6] | 0.857 ± 0.017 | 0.652 ± 0.015 |
| Pathomic (CNN+SNN) [6] | 0.913 ± 0.011 | 0.730 ± 0.019 |
| MultiCoFusion [10] | 0.998 ± 0.005 | 0.759 ± 0.032 |
| MLP (Genes) | 0.901 ± 0.051 | 0.700 ± 0.042 |
| MOAB (VGG19_bn+MLP)) | 0.945 ± 0.012 | 0.752 ± 0.002 |
| **MOAB (ConvNext+MLP)** | 0.956 ± 0.000 | **0.766 ± 0.001** |

### 4. CONCLUSION

In this paper, we proposed a novel Multi-modal Outer Arithmetic Block (MOAB) that fuses features from histology and genetic data. Involving features from all arithmetic operations encourages the classifier to capture intrinsic relations between modalities, ultimately enhancing the overall prediction. The channel fusion strategy enables MOAB to reduce the parameters of conventional fusion methods and enhances the classifier's knowledge. All conducted experiments demonstrate the effectiveness of proposed approach. Last, we note that the architecture of MOAB is transferrable to other multi-modal problems as it can combine modalities of any type with any CNN backbones. In the future, We plan to experiment MOAB with additional datasets and expand it to integrate more modalities.

## 5. ACKNOWLEDGMENTS

The authors would like to thank University of Jeddah and Saudi Arabia Cultural Bureau for their continued support of this study.

## 6. COMPLIANCE WITH ETHICAL STANDARDS

The licence related to the open-access data ensured no ethical approval was necessary.

## 7. REFERENCES

[1] M Khalid Khan Niazi, Keluo Yao, Debra L Zynger, Steven K Clinton, James Chen, Mehmet Koyutürk, Thomas LaFramboise, and Metin Gurcan, "Visually meaningful histopathological features for automatic grading of prostate cancer," *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 1027–1038, 2016.

[2] Tao Wan, Jiajia Cao, Jianhui Chen, and Zengchang Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomputing*, vol. 229, pp. 34–44, 2017.

[3] Sairam Tabibu, PK Vinod, and CV Jawahar, "Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[4] Zhi Huang, Xiaohui Zhan, Shunian Xiang, Travis S Johnson, Bryan Helm, Christina Y Yu, Jie Zhang, Paul Salama, Maher Rizkalla, Zhi Han, et al., "Salmon: survival analysis learning with multi-omics neural networks on breast cancer," *Frontiers in genetics*, vol. 10, pp. 166, 2019.

[5] Suzanne C Wetstein, Vincent MT de Jong, Nikolas Stathonikos, Mark Opdam, Gwen MHE Dackus, Josien PW Pluim, Paul J van Diest, and Mitko Veta, "Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images," *Scientific reports*, vol. 12, no. 1, pp. 1–12, 2022.

[6] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood, "Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Transactions on Medical Imaging*, 2020.

[7] Nikhilanand Arya and Sriparna Saha, "Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model," *IEEE/ACM transactions on computational biology and bioinformatics*, 2020.

[8] Rui Yan, Fa Zhang, Xiaosong Rao, Zhilong Lv, Jintao Li, Lingling Zhang, Shuang Liang, Yilin Li, Fei Ren, Chunhou Zheng, et al., "Richer fusion network for breast cancer classification based on multimodal data," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–15, 2021.

[9] Nathaniel Braman, Jacob WH Gordon, Emery T Goossens, Caleb Willis, Martin C Stumpe, and Jagadish Venkataraman, "Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 667–677.

[10] Kaiwen Tan, Weixian Huang, Xiaofeng Liu, Jinlong Hu, and Shoubin Dong, "A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction," *Artificial Intelligence in Medicine*, vol. 126, pp. 102260, 2022.

[11] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, and Lee AD Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018.

[12] Farina Hanif, Kanza Muzaffar, Kahkashan Perveen, Saima M Malhi, and Shabana U Simjee, "Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment," *Asian Pacific journal of cancer prevention: APJCP*, vol. 18, no. 1, pp. 3, 2017.

[13] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976–11986.

[14] Zhimeng Han, Muwei Jian, and Gai-Ge Wang, "Convunext: An efficient convolution neural network for medical image segmentation," *Knowledge-Based Systems*, vol. 253, pp. 109512, 2022.

[15] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[16] Yiming Yang and Xin Liu, "A re-examination of text categorization methods," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 1999, pp. 42–49.