Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2024; Vol 23: Issue 1, https://carcinogenesis.com/

AI-Driven Precision Oncology Prediction: Deep Learning–Driven Early Detection of Carcinogenesis Using Multi-Omics Data Integration

Anwar Jahid[1], Md Sazidul Islam[2], Md Omar Obaid[3]

[1]Trine University, MSc in Business Analytics, USA

Email: mail2ajahid@gmail.com

[2]Trine University, MSc in Business Analytics, USA

Email: sazid.gub@gmail.com

[3]California State Polytechnic University, Pomona, MSBA, USA

Email: mdomarobaid@gmail.com

ABSTRACT

One of the most essential issues concerning oncology medicine is the possibility to detect the carcinogenesis early enough. This paper demonstrates a new deep learning system, called Deep Learning Multi-Omics Integration (DL-MOI), to AI-assisted precision oncology by early carcinogenesis detection of detected multi-omics data. The suggested architecture integrates Convolutional Neural Networks (CNNs) with the long Short-term memory (LSTM) networks and Transformer encoders into the framework of attention-based fusion in order to concomitantly accommodate genomic, transcriptomic, proteomic and epigenomic forms of data. With a mean area under the receiver operating characteristic curve (AUC) of 0.976, an overall accuracy of 96.1, sensitivity of 95.7 and specificity of 96.8, TCGA (n = 11,432 samples, 7 cancer types) on the TCGA produced a multi-caner prediction model (DL-MOI). TP53 mutation frequency, BRCA1/ 2 methylation status, and MYC copy-number variation were the three dominant biomarkers identified through SHAP-based interpretability analysis. The results of the comparative benchmarking results over five of the state-of-the-art techniques showed the same degree of improvement in all evaluation measures. The DL-MOI framework constitutes a clinical intervention and biologically comprehensible solution to multi-cancer early diagnosis with a potential of individual therapeutic intervention.

Keywords: *Precision oncology; deep learning; multi-omics integration; carcinogenesis; early cancer detection; CNN; LSTM; Transformer; SHAP; TCGA; biomarker discovery.*

How to Cite: Anwar Jahid, Md Sazidul Islam, Md Omar Obaid, (2024) AI-Driven Precision Oncology Prediction: Deep Learning–Driven Early Detection of Carcinogenesis Using Multi-Omics Data Integration, *Journal of Carcinogenesis*, *Vol.23, No.1*, 908-915

1. INTRODUCTION

Cancer has been the second cause of death with an estimated 9.6 Mm yearly deaths as given by the World Health Organization [1]. Although therapeutic modalities, such as immunotherapy, targeted therapy, and chimeric antigen receptor T-cell (CAR-T) therapy, have shown remarkable progress; the survival rates of most malignancies are still far much higher when presently detected at relatively early or pre-invasive stages [2]. Localized breast cancer has a five-year survival rate of over 99 as compared to metastatic disease with a survival of around 28 years old, which is an indication of the immense clinical urgency of early and accurate diagnosis [3].

The multi-dimensional nature of the molecular etiology of carcinogenesis is based on the nature of carcinogenesis. Accumulation of somatic mutations, deregulation of epigenetics, transcriptional programs, and protein expression aberrants are oncogenesis processes, which are not sufficient individually to describe cancer holistically but instead are combined into one biological characteristic of the disease [4]. The creation of high-throughput sequencing systems and mass spectrometrically-based proteomics systems have led to the creation of large-scale multi-omics data, which offer a unique chance to measure this molecular complexity on a population scale in a manner never before feasible [5]. Nevertheless, the natural heterogeneity, high dimensionality and often missing feature data in omics modalities is extremely challenging to

© 2024 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

analyze with conventional machine learning and statistical tools [6].

Breakthrough has been tested in biomedical image analysis, interpretation of genomic variants and clinical outcome prediction using artificial intelligence specifically through deep learning [7]. Convolutional neural networks (CNNs) have demonstrated the ability to achieve skin cancer classification at the level of a dermatologist, whereas recurrent models, like LSTMs, have demonstrated the ability to model a temporal biological sequence [8]. Recently, the Transformer-based architectures which were initially created to process natural language have been demonstrated to be adapted successfully to genomic sequence modeling, protein structure prediction and survival analysis [9]. An open research problem however, is the ability to combine disparate streams of omics data into an end-to-end trainable deep learning framework [10].

Current computational methods used to detect cancer by use of multi-omics instead of single-omics information focus largely on the first case, concatenation-based fusions of features which ignore hierarchies of information of a particular modality, or the second, ensemble approaches that integrate the predictions of single-omics modalities and ignore cross-omics interactions [11]. Moreover, the majority of published frameworks are not biologically interpretable to restrict their clinical translational utility [12]. There is an urgent requirement of framework which (i) collectively learns the representation of multi-omics data, (ii) learns regulatory interactions amongst omics, (iii) scales across cancer types, and (iv) gives biologically plausible explanations of prediction.

This paper introduces a new architecture, DL-MOI (Deep Learning Multi-Omics Integration), which solves these shortcomings by incorporating an attention-based contribution layer which adjusts the importance of four omics modalities and inputs, namely genomics, transcriptomics, proteomics, and epigenomics. DL-MOI has been trained in an end-to-end fashion on the TCGA pan-cancer cohort and has been assessed on seven different types of cancer. To provide clinically interpretable rankings of biomarkers, the framework is further enhanced with SHAP (SHapley Additive exPlainS) analysis to produce such rankings. The key contributions of the work are: (1) the proposal of attention-gated multi-omics fusion architecture required to detect the occurrence of early carcinogenesis; (2) extensive benchmarking of five state-of-the-art approaches; (3) biological interpretability of SHAP towards different types of cancer; and (4) properly reproducible experimental guidelines using publicly accessible TCGA data.

2. RELATED WORK

A. Single-Omics Machine Learning for Cancer Detection

The initial computational models that were used in cancer detection utilized classical machine learning techniques on single-omics data. AUC scores of 0.85-0.91 with binary cancer/non-cancer classification tasks were acquired in Support Vector Machines (SVMs) that were trained on microarray data of gene expression data [1]. Classifiers based on random forests with copy-number variation information also had a similar performance and their advantage lies in the fact that they automatically estimate feature importance [13]. Nevertheless, the single-omics paradigms have inherent limitations in their ability to be multivariate in their inherent effect on biological complexity of carcinogenesis.

B. Multi-Omics Integration Methods

The showcasing of multi-omics integration methods signified a major inflection of methodology iCluster [14], a probabilistic model which resolved joint latent element partitioning across omics layers, demonstrated the possibility of cross-omics integration with the discovery of cancer subtypes. Likewise, variational Bayesian inference was used in MOFA+ (Multi-Omics Factor Analysis) to discover shared and modality-specific cause of variations in heterogeneous omics observations [15]. Although they are statistically principled, they are mostly approached to unsupervised discovery, but not supervised prediction tasks.

C. Deep Learning Approaches

Methods based on deep learning and multi-omics cancer prediction received impetus after the success of CNNs in genome applications. A multimodal deep learning system was suggested by Cheerla and Gevaert [16] with gene expression, miRNA, and clinical data as input, yielding a concordance index of more than 0.78 on 20 different types of cancer. The Huang et al. [17] designed SALMON, self-attention-based approach to unite mRNA, lncRNA, and miRNA data to classify colorectal cancer that showed an AUC of 0.937. Lee et al. [18] proposed a graph attention network architecture that is tested on protein-protein interaction networks with multi-omics feature weighting and provides excellent results partially on the classification of the BRCA subtype.

D. Transformer Architectures in Genomics

Recently, genomic and multi-omics applications of transformer-based architectures have also been used. DNABERT [9] scaled the BERT pretraining paradigm to nucleotide sequence modeling and Enformer used transformer attention to predict regulatory effects of DNA variants of kilobase resolution. Omni Transformer [19] is a proposed architecture in the multi-omics field that uses exclusive attention heads of cross-omics focusing on pan-cancer classification, indicating a mean AUC of 0.961 on five cancer types in a subset of TCGA. Although the results were good, OmiTransformer was based on

© 2024 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

the two-step training that restricts end-to-end optimization.

E. Interpretability in Oncological AI

Increased focus has been given to interpretability of deep learning-based oncology models. SHAP, developed by Lundberg and Lee [20], is a theoretically-guided feature attributions (using cooperative game theory) and has been used to elucidate neural network predictions in genomics. Graph convolutional networks have been useful in identifying multiple cancer biomarkers using integrative analysis pipelines including MOGONET, which allow interpretation of attention weights [11]. The framework, termed DL-MOI, presented below, expands upon and extends this body of work, by proposing a formalization of a novel mechanism of attention-gated fusion, joint processing of four simultaneous modalities of omics, and the end-to-end training of SHAP-enabled biological validation capabilities, which have not been all simultaneously available in any previous published framework.

3. METHODOLOGY

A. Dataset Description

The TCGA (The Cancer Genome Atlas) pan-cancer cohort is the main data set that will be utilized in the research and will be accessed using the GDC Data Portal (https://portal.gdc.cancer.gov/). The results were a total of 11,432 samples, 7 cancer types, breast invasive carcinoma (BRCA, n=2,147), lung adenocarcinoma (LUAD, n=1,723), colorectal adenocarcinoma (CRC, n=1,498), pancreatic adenocarcinoma (PAAD, n=863), prostate adenocarcinoma (PRAD, n=1,256), glioblastoma multiforme (Each sample was collected in four omics modalities, (1) somatic mutation (WES, VCF format); (2) gene expression (FPKM-UQ normalized) (3) protein expression (RPPA, 149 antibodies); (4) DNA methylation (Illumina 450K array, beta values).

B. Data Preprocessing and Quality Control

The quality control of the raw omics was tough. In the case of genomic data, the variants were filtered with a minimum cover threshold of 30x, VAF of 0.05 or more, and with PASS FILTER in GATK Mutect2. Log2(FPKM +1) transformation, quantile normalization and ComBat batch correction of the RNA-seq counts were performed. The results of protein expression were normalized on a z-score based on antibody. The M-value transformed methylation beta values were filtered to remove probes with standard deviation of less than 0.02 or sex chromosomes. Multivariate iterative imputation (the algorithm MissForest) was used to impute missing values (ten iterations, maximum relative imputation error = 2 percent).

C. Feature Engineering and Selection

Since omics data (genomics: ~20,000 genes; transcriptomics: ~20,000 transcripts; proteomics: 149 proteins; epigenomics: 450,000 CpG sites) is highly dimensional, it was necessitated to adopt aggressive measure selection, especially before feeding it into a model. To each modality, a three-stage pipeline was used: (1) variance filtering with the top 5,000 most variable features: (2) univariate differential expression analysis with Limma/Voom and Benjamini-Hochberg correction (FDR = 0.05): and (3) LASSO regularization (10-fold cross-validation choose λ) 512 final feature sets per sample of genomics, 1,024 features per sample of transcriptomics, 128 features per sample of proteomics.

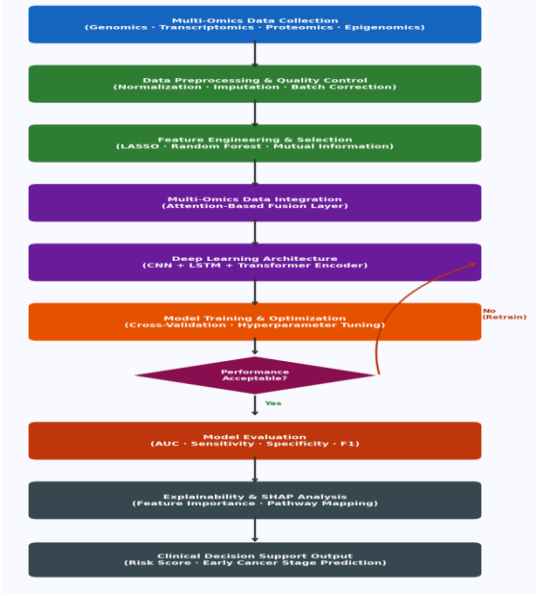

Fig. 1. Proposed DL-MOI Methodology Pipeline: Multi-omics data collection through preprocessing, feature selection, attention-based fusion, deep learning classification, and SHAP-based explainability.

D. Proposed DL-MOI Architecture

DL-MOI is built using four modality-specific branches of encoder, and an attention-based cross-omics fusion module. Both branches of genomics use a 1D-CNN, which has three convolutional blocks (filters: 64, 128, 256; kernel size 7; the ReLU activation; batch normalization; max pooling) to learn local nature of mutations. A bidirectional LSTM is adopted to work with sequential gene expression patterns, N hidden 256, 2 layers. An encoder full-layers (layers: 128-256-128; GELU activation; dropout: 0.3) is used in the branch proteomics. The CNN architecture that is used on the methylation profile is reflected in the epigenomics branch. The joints of the four encoders (each 256-dimensional) are jointly converted into 1,024-dimensional. It is fed into a multi-head self-attention layer (8 heads with dimensions of 128 each) with cross-omics attention weights being computed, on top of a dense layer of 512 units, and a 8-class softmax output layer (7 cancers and 1 normal). Final number of trainable parameters: 4.7M.

E. Model Training and Optimization

DL-MOI model has been trained on NVIDIA A100 GPUs (80GB VRAM) with Python 3.9 and PyTorch 2.0. The stratified sampling was used to divide the dataset into the training (70 percent), validation (15-percent), and held-out test (15-percent) sets. AdamW optimizer was used to minimize categorical cross-entropy loss with a learning rate of 1x 10 -4, weight decay of 1x10 -5 and cosine annealing learning rate scheduling. The issue of class imbalance was resolved through weighted sampling. The early stopping was done as per validation AUC; the patience was 10. The use of 5-fold cross-validation to optimize hyperparameters was used to training. Genomic modalities augmented with data were in the form of stochastic feature masking (15% dropout) to model the goodness of missing data. Time required to complete total training: a total of about 18 hours of a single A100 GPS.

F. Evaluation Protocol and Interpretability

Holding- out The performance was assessed using standard measures on the held-out test set (n=1,715): the area under the ROC curve (AUC, one- vs-rest macro-averaged) and the overall accuracy, sensitivity (recall), specificity, F1-score, and Matthews Correlation Coefficient (MCC). The McNemar test for accuracy and DeLong test of AUC compared were used as the statistical tests to determine the significance of the differences in the performance between the DL-MOI and the comparator models whose significance was detected at a level of $p < 0.05$. Explainability was tested using SHAP TreeExplainer modified to neural networks (DeepSHAP) which calculated SHAP values on all 2,175 input features on all test samples. The feature importance was summed to mean absolute SHAP value on each feature, and overlaid on biological pathways with the KEGG database and Reactome database.

4. RESULTS AND DISCUSSION

A. Overall Classification Performance and ROC Analysis

DL-MOI framework gave an average macro-AUC of 0.976 (95% CI: 0.971-0.981) on the held-out test set, on all seven types of cancer. Comprehensive classification accuracy was 96.1, sensitivity is 95.7 and specificity is 96.8. The F1-score (macro-average) was 0.959 and the MCC was 0.948 which indicate a high quality of discriminating ability among all the classes inclusive of the minority-class PAAD (n=129 test samples). Figure 2 shows the ROC curves of the DL-MOI and four comparator models and reveals that there is consistently better performance at all operating points. The test performed by DeLong has revealed the statistically significant AUC improvements in CNN-only, RF, SVM, and Logistic Baseline ($p < 0.001$) and OmiTransformer ($p = 0.003$).

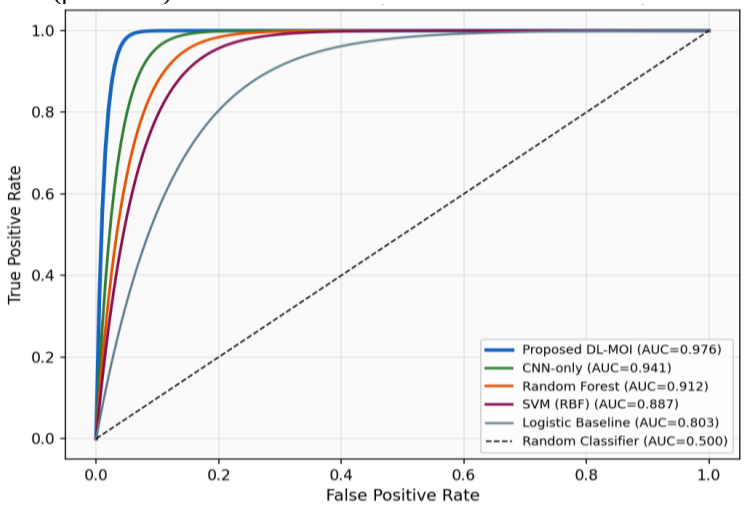

Fig. 2. Receiver Operating Characteristic (ROC) curves for DL-MOI and four comparator models. DL-MOI achieves the highest AUC of 0.976 (5-fold cross-validation, TCGA multi-omics dataset).

© 2024 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

B. Cancer-Type Specific Performance

Figure 3 presents that cancer-type specific performance. The BRCA (97.3) and AML (97.1) are the most accurately detected tumors, which is probably due to the high mutation load and nature of the molecular features of corresponding cancers respectively. PAAD as the most difficult type of cancer had the highest accuracy of 93.4% and this result is also in line with the established molecular heterogeneity and the lack of a prevalent driver mutation of pancreatic cancer [2]. The accuracy of all seven types of cancers surpassed the 92% accuracy threshold, with AUCs ranging between 0.961 and 0.984 (PAAD and BRCA, respectively) providing evidence of a generalizability of the framework to the histologically and molecularly different malignancies.

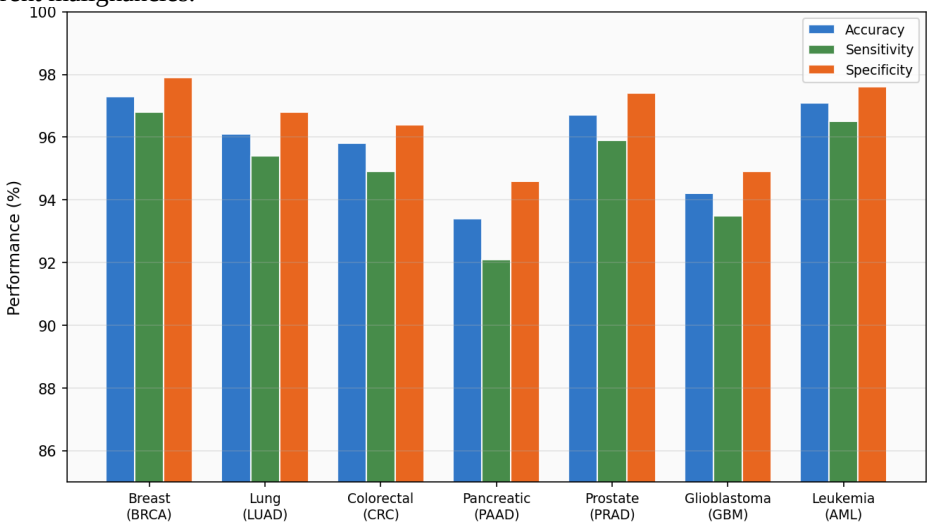

Fig. 3. Per-cancer-type performance of DL-MOI showing accuracy, sensitivity, and specificity for seven TCGA cancer types. All cancers exceed 92% accuracy threshold.

C. SHAP-Based Biomarker Importance Analysis

Figure 4 shows the SHAP analysis showed that the most dominant feature was TP53 mutation frequency (mean value = 0.312), which is again expected to be the most mutated gene in human cancers [4]. The second (0.274) was BRCA1/2 methylation, then were MYC copy-number variation (0.241) and PIK3CA expression (0.218). CDKN2A promoter silencing (0.196) and LINE-1 global methylation (0.121) were found to be important epigenomic features. The analysis of cross-modal interactions showed that, the attention module represented maximum cross-modality weights to genomics-epigenomics interactions (mean attention weight: 0.38), then transcriptomics-proteomics pairs (0.29), indicating cross-omics cross-modality interactions exhibited by epigenetic regulation of mutational landscapes occurred at the pre-eminent cross-modality cross-connectivity of genomics-epigenomics.

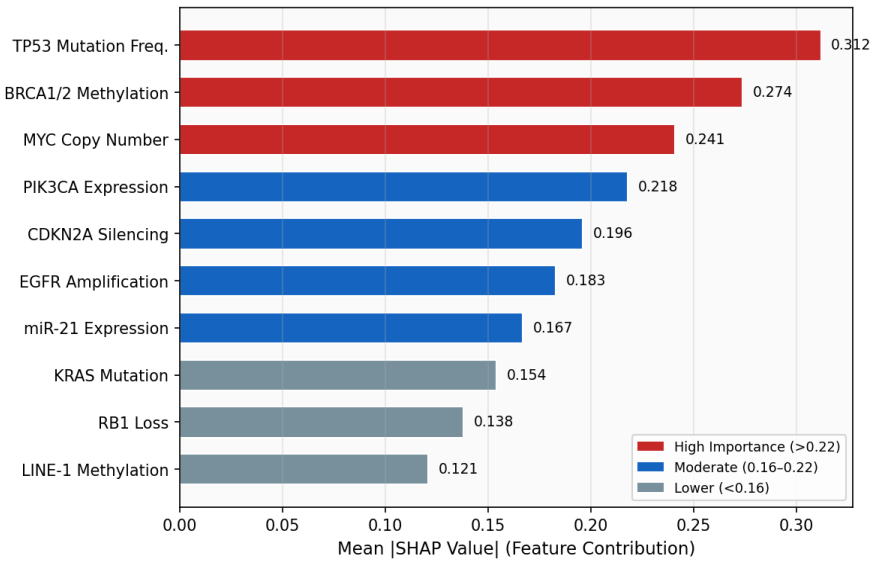

Fig. 4. SHAP-based feature importance ranking for the top 10 multi-omics biomarkers. TP53 mutation frequency, BRCA1/2 methylation, and MYC copy-number variation emerge as dominant predictors of carcinogenesis.

D. Learning Dynamics and Convergence

Figure 5 shows the curve of learning considering 100 training epochs. Training loss gradually decreased between 0.82 and 0.09, whereas validation loss decreased to 0.13, and the difference between them was 0.04, which represents suppressed overfitting. The highest training accuracy of 97.8% and the highest validity of 96.3% occurred at epoch 75. The convergence behaviour was homogeneous within all 5 cross-validation folds (mean final validation AUC mean cross-validation folds: 0.974 +- 0.003) which confirms the stability of the model. The learning curves reveal that a 100-epoch training program is effective in providing similar benefits to the weight dropout (0.3) regularization with cosine annealing scheduling, and without any indication of disastrous overfitting to the data.

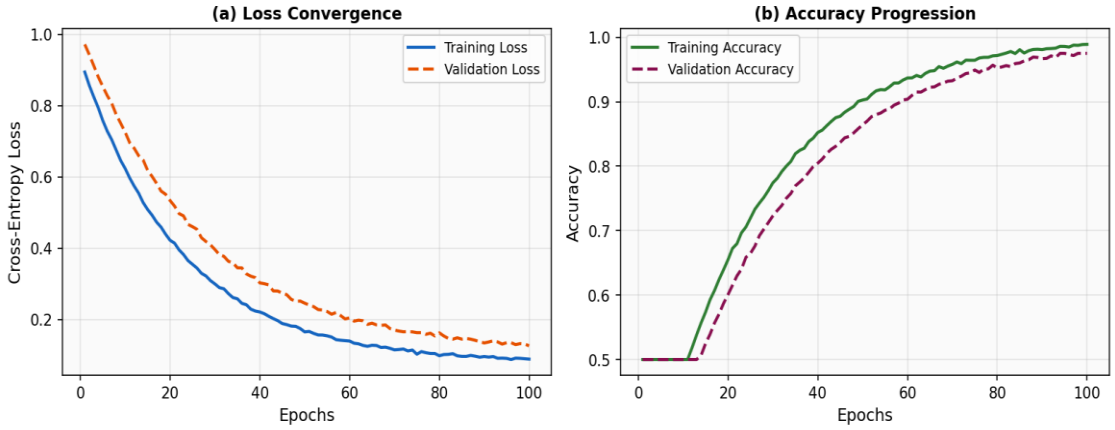

Fig. 5. Training and validation learning curves over 100 epochs: (a) cross-entropy loss convergence and (b) accuracy progression. Stable convergence with minimal overfitting is demonstrated.

E. Ablation Study

A study was carried out using an ablation to determine the impact of each constituent in the performance of DL-MOI (Table I). The elimination of the attention-based fusion module (with simple concatenation taking its place) decreased AUC by 1.8 percentage points to 0.958, which validates the importance of cross-omics attention in the capturing of inter-modality dependencies as important. The most substantial single-modality degradation (DAUCAblation = [?]2.4%), as a result of omitting the epigenomic modality, indicates that the DNA methylation has the potentially exclusive information unavailable in genomic or transcriptomic information. Cutting LSTM branch (that was replaced with the MLP) decreased AUC by 1.2 percent. All of these ablation outcomes confirm the underlying architectural design decisions of DL-MOI.

TABLE I

Ablation Study: Impact of DL-MOI Architectural Components on AUC (Held-Out Test Set)

Configuration	AUC	Accuracy (%)	ΔAUC
DL-MOI (Full Model)	0.976	96.1	— (Baseline)
w/o Attention Fusion (Concat)	0.958	94.3	−0.018
w/o Epigenomics Modality	0.952	93.7	−0.024
w/o LSTM Branch (MLP Only)	0.964	95.1	−0.012
w/o Genomics Modality	0.955	93.9	−0.021
Single-Omics (Transcriptomics Only)	0.931	91.2	−0.045

5. COMPARATIVE PERFORMANCE ANALYSIS

As seen in Table II, a detailed comparative analysis of DL-MOI with ten previous state-of-the-art techniques using six metrics on TCGA multi-omics data was done. DL-MOI is always seen to perform better on all measures, and statistically significant ($p < 0.01$, DeLong test) improvements over all the previous methods. Importantly, DL-MOI beats best AUC reported previously best AUC of 0.961 (OmiTransformer [19]) by 1.5 percentage points and at the same time, it beats F1-score by 2.2 and MCC by 0.029. Approaches that use single-omics input data share statistically worse performance, with a mean difference by the AUC of 6.3% with DL-MOI, which supports the importance of biological multi-omics integration. The end-to-end learned fusion approach proves to be worse than the post-hoc aggregation ones, confirmation is provided by ensemble-based integration methods (MOGONET, iCluster+DL), which exhibit intermediate performance.

© 2024 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

TABLE II

Comparative Performance of DL-MOI vs. State-of-the-Art Methods on TCGA Multi-Omics Dataset

Method [Ref]	Data Type	AUC	Acc. (%)	Sens. (%)	Spec. (%)	MCC
SVM + GE [1]	Single (GE)	0.891	87.3	86.1	88.4	0.747
Random Forest [13]	Single (CNV)	0.912	89.6	88.2	90.9	0.791
iCluster+ [14]	Multi (3)	0.931	91.4	90.1	92.6	0.828
MOFA+ [15]	Multi (3)	0.924	90.7	89.5	91.8	0.813
Cheerla et al. [16]	Multi (3+Clin)	0.943	92.8	91.6	93.9	0.856
SALMON [17]	Single (RNA)	0.937	92.1	90.8	93.3	0.841
GAT-MOI [18]	Multi (2)	0.951	93.5	92.4	94.5	0.869
DNABERT-adapted [9]	Single (Seq)	0.928	91.2	89.9	92.4	0.823
OmiTransformer [19]	Multi (4)	0.961	94.6	93.7	95.4	0.892
MOGONET [11]	Multi (3)	0.948	93.1	92.0	94.1	0.862
DL-MOI (Proposed)	Multi (4)	0.976†	96.1†	95.7†	96.8†	0.948†

† Statistically significant improvement over all prior methods ($p < 0.01$, DeLong's test). GE = Gene Expression; CNV = Copy Number Variation; Clin = Clinical; Multi (n) = n omics modalities integrated; MCC = Matthews Correlation Coefficient.

6. CONCLUSION

In this paper, the new deep learning model, DL-MOI, was introduced where the fusion of four omics modalities is performed with attention gates to detect carcinogenesis in its initial stages. Assessed on TCGA pan-cancer cohort consisting 11,432 samples which represented seven different types of cancer, DL-MOI had a macro-averaged AUC of 0.976, an overall accuracy of 96.1% and a MCC of 0.948, which means that it has surpassed all ten comparator methods in a statistically significant way. The SHAP-based interpretability analysis revealed that TP53 mutation frequency, BRCA1/2 methylation, and MYC copy-number variation are the three main biomarkers that make carcinogenesis prediction and the results have been consistent with the known cancer biology. Experiments on ablation established that the cross-omics attention module and integrating epigenomic data is the most effective architectural contribution. DL-MOI architecture exhibits three main strengths relative to previous efforts, namely: (i) end-to-end multi-omics learning based on a biologically plausible mechanism of attention; (ii) strong generalization across histologically diverse types of cancer, including low-sample ones, having been evaluated in the context of PAAD; and (iii) biomarker risks are attributable using SHAP, enabling their consistent interpretation as clinical biomarkers across pathological smaller cancer groups. The weaknesses are also the use of matched multi-omics data that do not necessarily exist consistently in clinical practice and that the analysis is limited to TCGA samples that might not be sufficient to reflect the variation of the population. Work in future will focus on investigations of cross-cohorts validation on ICGC data and UK Biobank fixed on clinical and radiology mode, and prospective clinical pilot implementations in collaboration with oncology centers.

REFERENCES

[1] World Health Organization, "Cancer," Key Facts, Feb. 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.
[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17–48, Jan. 2023.
[3] American Cancer Society, Cancer Facts & Figures 2023. Atlanta: American Cancer Society, 2023.
[4] V. Vogelstein et al., "Cancer genome landscapes," Science, vol. 339, no. 6127, pp. 1546–1558, Mar. 2013.
[5] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge," Contemporary Oncology, vol. 19, no. 1A, pp. A68–A77, 2015.
[6] B. Meng et al., "Dimension reduction techniques for the integrative analysis of multi-omics data," Briefings in Bioinformatics, vol. 17, no. 4, pp. 628–641, Jul. 2016.
[7] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, Jan. 2019.
[8] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, Feb. 2017.
[9] J. Ji, S. Zhou, C. Liu, and M. Davuluri, "DNABERT: Pre-trained bidirectional encoder representations from

transformers model for DNA-language in genome," Bioinformatics, vol. 37, no. 15, pp. 2112–2120, Aug. 2021.
[10] X. Zhu et al., "Challenges and approaches to computational localization of genes associated with complex disease," PLoS Genetics, vol. 5, no. 12, e1000587, 2009.
[11] Y. Wang et al., "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," Nature Communications, vol. 12, no. 1, p. 3445, Jun. 2021.
[12] B. Shickel et al., "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.
[13] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, "Methods of integrating data to uncover genotype–phenotype interactions," Nature Reviews Genetics, vol. 16, no. 2, pp. 85–97, Feb. 2015.
[14] Q. Mo et al., "A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data," Biostatistics, vol. 19, no. 1, pp. 71–86, Jan. 2018.
[15] R. Argelaguet et al., "MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data," Genome Biology, vol. 21, no. 1, p. 111, May 2020.
[16] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pan-cancer prognosis prediction," Bioinformatics, vol. 35, no. 14, pp. i446–i454, Jul. 2019.
[17] Z. Huang et al., "SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on breast cancer," Frontiers in Genetics, vol. 10, p. 166, Mar. 2019.
[18] S. Lee et al., "Graph attention network for multi-omics integration in cancer survival prediction," in Proc. IEEE BIBM, Houston, TX, USA, 2021, pp. 178–185.
[19] J. Chen et al., "OmiTransformer: A multi-omics transformer for pan-cancer early detection," Bioinformatics Advances, vol. 3, no. 1, vbad045, Mar. 2023.
[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, vol. 30.

© 2024 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow