Student Research Abstract: Multimodal Deep Learning Based Approach for Cells State Classification

Paula Raissa Silva
PhD Candidate, Faculty of Engineering - University of Porto
Institute for Systems and Computer Engineering of Porto (INESC TEC)
Porto, Portugal
paula.r.silva@inesctec.pt

ABSTRACT

With the advances of the big data era in biology, deep learning have been incorporated in analysis pipelines trying to transform biological information into valuable knowledge. Deep learning demonstrated its power in promoting bioinformatics field including sequence analysis, bio-molecular property and function prediction, automatic medical diagnosis and to analyse cell imaging data. The ambition of this work is to create an approach that can fully explore the relationships across modalities and subjects through mining and fusing features from multi-modality data for cell state classification. The system should be able to classify cell state through multimodal deep learning techniques using heterogeneous data such as biological images, genomics and clinical annotations. Our pilot study addresses the data acquisition process and the framework capable to extract biological parameters from cell images.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → **Bioinformatics**;

KEYWORDS

Multimodal deep learning, bioimaging, bioinformatics

ACM Reference Format:
Paula Raissa Silva. 2020. Student Research Abstract: Multimodal Deep Learning Based Approach for Cells State Classification. In *The 35th ACM/SIGAPP Symposium on Applied Computing (SAC '20), March 30-April 3, 2020, Brno, Czech Republic.* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3341105.3374228

PROBLEM AND MOTIVATION

In the last years, with the advances in computational power and the growth of biological data available, new challenges have arisen such as store, process and analyse of the data. Two areas, from computer science field, have been proving to be useful to address these challenges: Computer Vision and Machine Learning. Computer Vision (CV) is an interdisciplinary subject that aims to bridge the gap between the level of images are represented by machines and the high level way that a human can interpret images and videos. Machine Learning (ML) is one of promising techniques that attempts to address the challenge to build and train machines for uncover underlying patterns, build models and make predictions based on the best fit model. The combination of these two techniques can build a bridge making what a human brain does with the retinal input and the brain in a artificial way.

Biological image analysis is a subfield of bioinformatics, where the representation and knowledge extraction methods are critical for understanding various features of cell biology, molecular biology and neuroscience. One of the most challenging problems in this field is the identification of cell malfunction from images. Changes in cellular mechanics are emerging as an important factor in numerous diseases [11]. The process of linking the disease to cell malfunction or misregulation remains challenging, because in numerous cases only the features which represent the cell shape are not sufficient to determine the cell function. For example, microglia is one type of cell in the central nervous system. It is considered the first line of defense within the brain and the major influencer of the brain inflammatory response. During the inflammatory process, the cell change their shape according to chemical influences and the micro-environment. It has been proved before that the cell shape and function are closely related [2]. Then, according to the phenomena that happen during the inflammation process and the forms that the cell takes on can detect cell malfunctions that cause disease like dementia [4], Alzheimer [5], brain cancer [19] and others neurodegenerative diseases [14].

Biological images are generally represented by the morphological parameters [7]. Considering that morphologies are closely related to their functional state, additional information like molecular, genome and clinical annotation may be important to improve the identification of cells malfunction and related diseases. This context characterises an heterogeneous scenario and can be addressed to multimodal learning problem, because the goal is combine different types of data and extract knowledge from them in a automatic way. Multimodal learning is one of the promising ML techniques with the aim of build models and make predictions in heterogeneous data scenarios. The conventional ML algorithms are high dependent of data representation and the problem complexity. Deep Learning (DL), a branch of machine learning, has recently emerged as the state-of-the-art multimodal processing is due to the fact that they have been outperforming previous traditional ML techniques in several tasks, and can model the abundance of complex data from different sources like visual, textual and numeric. The goal of this work is to build a deep multimodal learning based approach to

SAC '20, March 30-April 3, 2020, Brno, Czech Republic

ACM ISBN 978-1-4503-6866-7/20/03.
https://doi.org/10.1145/3341105.3374228

identify the different types of cells states, which are characterised by current health status data, biological annotation and cell image.

BACKGROUND AND RELATED WORK

In general terms, a modality refers to the way in which something happens or is experienced. Multimodal data sets consist of data from different such modalities which characterises a common phenomena. Multimodal machine learning is a multidisciplinary field which aims to use data in a complementary manner with the goal of building models to learn a complex task. The early researches in this area covered the problems in audio-visual speech recognition and still nowadays it is main application area. This research field brings some challenges for computational research given the heterogeneity of data. According to the taxonomy proposed by Baltrušaitis et al. [3] there are five core technical challenges surrounding multimodal learning: representation, translation, alignment, fusion and co-learning.

Deep learning has resulted in state-of-the-art performance for many problems regarding heterogeneous data scenarios, mainly in areas involving high dimensional unstructured data, because the hierarchically representation can be automatically learned for each modality [15]. Actually there are two important areas in deep multimodal learning: (i) methods that use regularization techniques to improve cross-modality learning and (ii) methods to attempt to find optimal architectures through search, optimization or some learning procedure.

For bioimaging data, the most common problem is the small number of subjects and large feature dimensions. Even after feature extraction, the dimension of feature continues high compared to the size of subject. In most cases, the irrelevant and redundant features need to be removed by feature selection process. In the literature, most existing feature selection methods are performed for each modality individually and ignore the potential relationship among different data modalities. Our proposed approach considers the information of all modalities and try to link the relationships across different modalities as extra information, addressing the problem of fusion structured learning. Previously researches tried to address the challenge of building models using biological data, such as image, clinical and genomics biomarkers using traditional machine learning methods [18].

The major challenges in applying deep learning based approaches for biological applications are: the obtaining of sufficiently labeled data and the problem of imbalance class. The first problem have been treated by applying patch-based training [8] and transfer learning [17]. To address the class imbalance problem we can apply weighted loss function techniques [10].

APPROACH AND UNIQUENESS

In this paper, we propose an approach that can fully explore the relationships across modalities and subjects through mining and fusing features from multi-modality data for cell function classification. Specifically, our proposed approach includes tree major steps: (i) data modalities representation; (ii) label-aligned multi-task feature selection; and (iii) multimodal classification. In order to achieve the stated goals, the proposed approach involves the following steps showed Figure 1 and explained in the next sections.

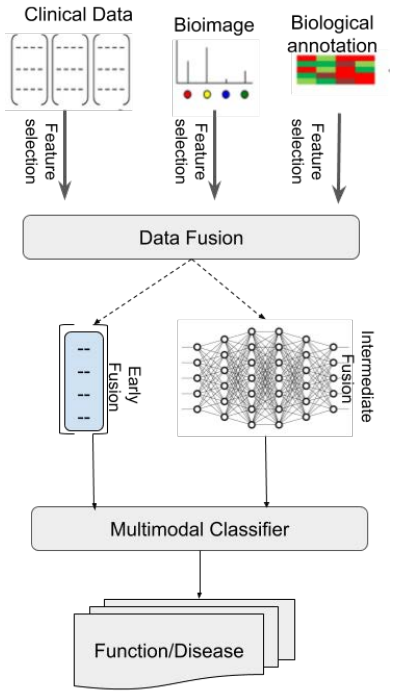

Figure 1: Multimodal approach for cell function classification

Data Representation

The data representation step is the most important step, because good representations are important for the performance of deep learning models. For this step the proposal involves the following tasks, subtasks and outcomes:

Task 1 : Morphological image analysis. This task involves the following subtasks: (i) definition of the morphological parameters; and (ii) a comparative study between the different tools to extract the parameters from images. The expected outcomes are: (a) the implementation of the framework to extract morphological and molecular parameters from cell image data ; and (b) a dataset composed by morphological and molecular parameters.

Task 2 : Biological data representation. This task involves the following subtasks: (i) extraction of bio-molecular annotations; (ii) extraction of genomics data; and (iii) gene enrichment analysis. The expected outcomes are: (a) the implementation/adoption of the framework to biological data enrichment and selection analysis; and (b) the data set of molecular and genomics data representation.

Task 3 : Clinical data representation. This task involves the following subtasks: (i) exploration of available data; (ii) grouping the different types of clinical data; (iii) selection of demographic data, if available; and (iv) natural language processing for textual features. The expected outcomes are: (a) health

status data set composed by numerical and categorical variables; and (b) word embedding representation for textural features.

Data Fusion

The data fusion step aims to join information from at least three modalities (image, biological and clinical) to perform a prediction. Architectures based in deep learning offer the possibility of implementing multimodal fusion either as early, intermediate and late fusion. Before fusing at the feature level, it is important to define the type of features representation. When non-hierarchical features are used, such as handcrafted features, features extracted from multiple modalities can be fused at only one level. Otherwise, if the task involves learning hierarchical representations from raw data, this gives rise to intermediate fusion. The late fusion is not applicable in this project, because this proposal involves the combination of multi-modalities before the building model process. For this step, the proposal involves the following tasks, subtasks and outcomes:

Task 1 : Feature level representation. This task involves the following subtasks: (i) non-hierarchical features; and (ii) hierarchical features. The expected outcomes are: (a) handcrafted features; and (b) learned representations.

Task 2 : Early fusion. This task involves the following subtasks: (i) handcrafted features concatenation; and (ii) dimensionality reduction. The expected outcome is the vector of multimodal features. An example of this process is showed in Figure 2.

Task 3 : Intermediate fusion. This task involves the following subtasks: (i) definition of the type of learned representation (2-D-convolution or fully connected); (ii) shared representation layer; and (iii) dimensionality reduction with stacked autoencoders. The expected outcome is the intermediate fusion architecture. Figure 3 presents an example of this process.

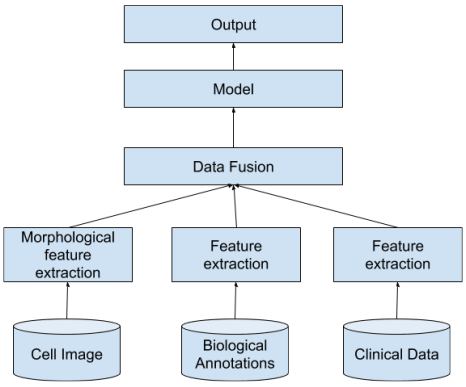

Figure 2: Early or data-level fusion

Multimodal Classification

The multimodal classification step aims to build models which are capable to predict the functions of cells based on current health

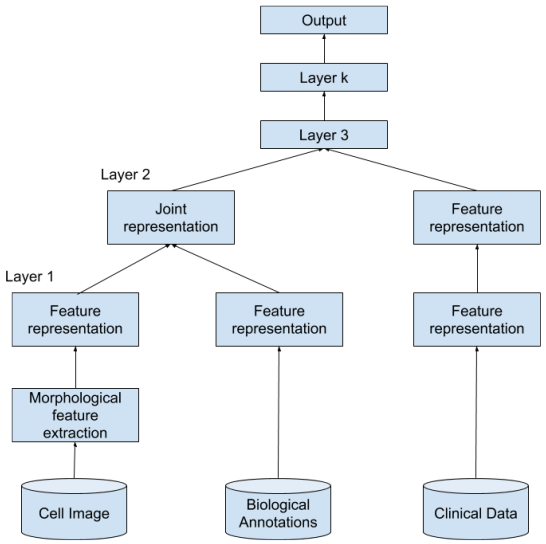

Figure 3: Intermediate fusion

status data, biological data and cell image. According the categorization proposed by Ramachandram and Taylor [15], there are three learning paradigms such as generative, discriminative and hybrid. In the supervised learning task which we have data X and labels Y, the generative model learn the joint probability $P(X, Y)$. Otherwise, the discriminative models learn the conditional $P(X|Y)$. Hybrid models combine both discriminative and generative models in a unified framework. The discriminative models generally require a large set of labeled data. Considering the most commons problems in bioimaging applications regarding the small number of subjects and unbalanced classes, the discriminative models are not applicable in our case. In this case, our proposal can be characterised as a generative approach.

For this step, the proposal involves the following tasks, subtasks and outcomes:

Task 1 : Test of deep architectures used in literature, such as DNN, ResNets and RNN. This task involves the following subtasks: (i) survey of the algorithms that exist for the three types of architecture; and (ii) choose one algorithm for each architecture which better fits in each fusion data method. The expected outcome is implemented pipelines with the selected algorithms.

Task 2 : Evaluation. This task involves the following subtasks: (i) definition of evaluation metrics; (ii) evaluate and compare each learning architecture chosen with different data sets; and (iii) build a pipeline including the algorithm that best fits the multimodal model. The expected outcome is an approach which is able to classify functions of cells and link to diseases using multimodal data.

RESULTS AND CONTRIBUTIONS

In the actual phase of the project, we are building the framework to deal with the extraction of morphological parameters. The image

acquisition process is happening during the project by a partner research center. From the set of available images, there are two types of filaments: actin[6] and myosin[6]. In this work, the cell images are acquired by a multi-channel microscope camera able to represent the cell into two channels, such as the first channel is the cell shape and the second channel is the molecular annotations. Hence, our approach has to be able to detect multi-channel images, split them and extract features for each channel. Actually, the framework is capable to extract fifteen morphological parameters from the first image channel according to Fernández-Arjona et al. [7], and the next phase is extract molecular annotations from the second channel. The source code is available at GitHub[1].

To expand our approach and test in different scenarios, in a literature review process, we selected 6 datasets with different types of modalities (Table 1). The selection criteria of the databases was that they had at least 3 types of data: images, clinical and genomics. The datasets TCGA-KIRC [1], TCGA-GBM [16], TCGA-LGG [13], TCGA-OV [9] and TCGA-BRCA [12] are available at The Cancer Image Archive [2] for public download. The IDNA dataset is available for free authorized researches through the LONI Image and Data Archive (IDA) [3].

Table 1: Heterogeneous Datasets

Dataset	Subjects	Modalities
TCGA-KIRC	267	Images, tissue slides, clinical, genomics
TCGA-GBM	262	Images, tissue slides, clinical, genomics
TCGA-LGG	199	Images, tissue slides, clinical, genomics
TCGA-OV	143	Images, tissue slides, clinical, genomics
TCGA-BRCA	139	Images, tissue slides, clinical, genomics
ADNI	1737	MRI, PET, clinical, genomics,

(a) MRI: Magnetic Resonance Image. PET: Positron Emission Tomography.

In this work, we showed how multimodal deep learning can be applied to this challenging task for identification of cells states from biological images, genomics and clinical data. There are several limitations that should be further considered in the planed approach. As we showed in the Table 1, the biggest dataset has 1737 subjects (study participants) and the other ones have from 139 to 267 subjects. When comparing with other heterogeneous scenarios, which authors are applying deep learning, the amount of data in this work looks small. However, it is important to consider that the complexity of the problem is more related to the combination of different data types than the number of subjects. Moreover, if the data are non very rich in information, we can improve the results of experiments with a gene enrichment process able to search and extract additional information from external annotation data bases such as *Ensembl*[4], *Uniprot*[5], *Gene Ontology (GO)*[6] and *KEGG*[7].

[1]https://github.com/paularaissa/DMLFramework
[2]https://www.cancerimagingarchive.net/
[3]http://adni.loni.usc.edu/data-samples/data-types/
[4]https://www.ensembl.org/
[5]https://www.uniprot.org/
[6]http://geneontology.org/
[7]https://www.genome.jp/kegg/

REFERENCES

[1] Heller M Jarosz R Erickson B J Kirk S Filippini J Akin O, Elnajjar P. 2016. Radiology Data from The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma [TCGA-KIRC] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR. https://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR

[2] Vasiliki Anesti and Luca Scorrano. 2006. The relationship between mitochondrial shape and function and the cytoskeleton. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1757, 5-6 (2006), 692–699.

[3] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 2 (2018), 423–443.

[4] Annachiara Cagnin, David J Brooks, Angus M Kennedy, Roger N Gunn, R Myers, Federico E Turkheimer, Terry Jones, and Richard B Banati. 2001. In-vivo measurement of activated microglia in dementia. *The Lancet* 358, 9280 (2001), 461–467.

[5] Brent Cameron and Gary E Landreth. 2010. Inflammation, microglia, and Alzheimer's disease. *Neurobiology of disease* 37, 3 (2010), 503–509.

[6] Geoffrey M Cooper and Robert E Hausman. 2016. *The Cell: A Molecular Approach*. Artmed Editora.

[7] María del Mar Fernández-Arjona, Jesús M Grondona, Pablo Granados-Durán, Pedro Fernández-Llebrez, and María D López-Ávalos. 2017. Microglia morphological categorization in a rat model of neuroinflammation by hierarchical cluster and principal components analysis. *Frontiers in cellular neuroscience* 11 (2017), 235.

[8] Yanrong Guo, Guorong Wu, Leah A Commander, Stephanie Szary, Valerie Jewells, Weili Lin, and Dinggang Shen. 2014. Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 308–315.

[9] Prior F Mutch D G Bhosale P Garcia K Erickson B J Holback C, Jarosz R. 2016. Radiology Data from The Cancer Genome Atlas Ovarian Cancer [TCGA-OV] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.NDO1MDFQ. https://doi.org/10.7937/K9/TCIA.2016.NDO1MDFQ

[10] Seong Gon Kim, Nawanol Theera-Ampornpunt, Chih-Hao Fang, Mrudul Harwani, Ananth Grama, and Somali Chaterji. 2016. Opening up the blackbox: an interpretable deep neural network-based classifier for cell-type specific enhancer predictions. *BMC systems biology* 10, 2 (2016), 54.

[11] Sanjay Kumar and Valerie M. Weaver. 2008. Mechanics, malignancy, and metastasis: The force journey of a tumor cell. *Cancer and Metastasis Reviews* 28 (2008), 113–127.

[12] Zuley M L Jarosz R Bonaccio E Filippini J Gruszauskas N Lingle W, Erickson B J. 2016. Radiology Data from The Cancer Genome Atlas Breast Invasive Carcinoma [TCGA-BRCA] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.AB2NAZRP. https://doi.org/10.7937/K9/TCIA.2016.AB2NAZRP

[13] Scarpace L Mikkelsen T Eschbacher J M Hermes B Ostrom Q Pedano N, Flanders A E. 2016. Radiology Data from The Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK. https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK

[14] V Hugh Perry, James AR Nicoll, and Clive Holmes. 2010. Microglia in neurodegenerative disease. *Nature Reviews Neurology* 6, 4 (2010), 193.

[15] Dhanesh Ramachandram and Graham W Taylor. 2017. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine* 34, 6 (2017), 96–108.

[16] Cha soonmee Rao S Tekchandani S Gutman D Pierce L J Scarpace L, Mikkelsen T. 2016. Radiology Data from The Cancer Genome Atlas Glioblastoma Multiforme [TCGA-GBM] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9. https://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9

[17] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging* 35, 5 (2016), 1299–1312.

[18] Yuria Utsumil, Ognjen Oggi Rudovicl, Kelly Petersonl, Ricardo Guerrero, and Rosalind W Picardl. 2018. Personalized gaussian processes for forecasting of alzheimerâĂŹs disease assessment scale-cognition sub-scale (adas-cog13). In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4007–4011.

[19] Shih-Ying Wu and Kounosuke Watabe. 2017. The roles of microglia/macrophages in tumor progression of brain cancer and metastatic disease. *Frontiers in bioscience (Landmark edition)* 22 (2017), 1805.