

Toward a Multimodal Multitask Model for Neurodegenerative Diseases Diagnosis and Progression Prediction

Sofia Lahrichi, Maryem Rhanoui, Mounia Mikram and Bouchra El Asri

IMS Team, ADMIR Laboratory, Rabat IT Center, ENSIAS, Mohammed V University, Rabat, Morocco

Keywords: Alzheimer's Disease, Multimodal Multitask Learning, Machine Learning, Deep Learning, Progression Detection, Time Series.

Abstract: Recent studies on modelling the progression of Alzheimer's disease use a single modality for their predictions while ignoring the time dimension. However, the nature of patient data is heterogeneous and time dependent which requires models that value these factors in order to achieve a reliable diagnosis, as well as making it possible to track and detect changes in the progression of patients' condition at an early stage. This article overviews various categories of models used for Alzheimer's disease prediction with their respective learning methods, by establishing a comparative study of early prediction and detection Alzheimer's disease progression. Finally, a robust and precise detection model is proposed.

1 INTRODUCTION

Alzheimer's disease (AD) is a progressive, irreversible neurodegenerative disease characterized by an abnormal build-up of amyloid plaques and neurofibrillary tangles in the brain, resulting in memory, thinking and behavior issues. AD is the most common form of dementia characterized by a slow and asymptomatic progress of the disease.

AD is clinically very heterogeneous, varying from patient to patient in terms of cognitive symptoms, test results, and rate of progression. Indeed, several recent therapeutic trials have shown variable efficacy from one subset of patients to another. Currently available treatments only slow the progression of AD, and no definitive cure has been developed yet. When it comes to AD, patient data (Zhang et al., 2012) is heterogeneous, but complementary, of different types; magnetic resonance imaging (MRI), positron emission tomography (PET), genetics, cerebrospinal fluid (CSF), etc. The combination of multimodalities (Weiner et al., 2013) facilitates the detection of changes in the patient's states and constitutes a reliable diagnosis.

Patient data is gathered from different visits and from continuous patient monitoring. The state of the disease at any given time is not independent of the condition at a previous time. Therefore, AD data are not only multimodal but could also be considered as time series and longitudinal series.

In the practical diagnosis of AD (Jo et al., 2019), the most widely used neuroimaging dataset comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) which contains (Jack Jr et al., 2008) socio-demographic data (gender, education level), the APOE genotype and five neuropsychological test results: MMSE (mini-mental state examination), CDR-SB (the sum of boxes of clinical dementia rating), ADASCog (Alzheimer's Disease Assessment Scale cognitive sub-scale), LMT (Logical Memory Test) and RAVLT (Rey Auditory Verbal Learning Test). Several learning methods have been developed for the classification and prediction of MCI (Mild Cognitive Impairment) to AD conversion, namely Machine Learning and Deep Learning.

After extensive bibliographical and analytical work on various articles, we will give an overview of the categories of approaches used for the prediction of Alzheimer's disease, then we will study the different learning techniques for each approach and establish a comparative table of this last. This categorization was essential to us because it will allow us at the end of the study to propose a stable and precise detection model.

The remainder of the article will consist of three main parts; the first will be the most important because through it, we will establish the bases of our study by defining the concepts of multimodal and multitask learning. The second part will be devoted

to the comparative study of the different models for predicting the progression of AD. And in the last part, we will propose a model grouping together all the adequate criteria for a better detection of the disease.

2 BACKGROUND

In this section, we introduce the main notions necessary to understand the context of our proposal.

2.1 Multi-modal Learning

Multimodal learning combines (Zhang et al., 2020) data from multiple sources that are semantically correlated and sometimes provide complementary information to each other, resulting in more robust predictions.

The multimodal learning model not only captures the correlation structure between different modalities but also allows to recover missing modalities given those observed.

The multimodal learning model for example (Baheti, 2020) can combine two deep Boltzmann machines, each corresponding to a modality. An additional hidden layer is placed on top of the two Boltzmann machines to give the common representation.

Multimodal learning is divided into four stages:

- (1) Representing inputs and summarize data in a way that expresses multiple modalities.
- (2) Translating (map) the data from one modality to another.
- (3) Extracting important features from information sources while creating models that best match the type of data.
- (4) Fusion and co-learning. Consists in combining the information of two or more modalities to make a prediction. This combination should be normally weighted.

2.2 Multitask Learning

Multitask learning (MTL) aims to extract useful information from multiple tasks and use their correlations to help learn a more accurate pattern for each task.

Multitask learning has been used successfully in all applications of machine learning, natural language processing (Worsham and Kalita, 2020), speech recognition (Pironkov et al., 2016), computer vision, and drug discovery.

MTL (Zhang and Yang, 2018) can be categorized into several categories, including supervised multitask learning (MTSL), unsupervised multitask learning, semi-supervised multitask learning...

MTSL aims to learn n functions for the n tasks of the labeled training set, (Each task can be a classification or regression problem) after it uses f_i (\bullet) to predict the labels of the data instances i from the j task. There are three categories of MTSL patterns to indicate the relationship between tasks:

Feature-based MTSL. Which learns common features of different tasks in order to share knowledge and avoid using original representations directly. It can easily be affected by outliers.

Parameter-based MTSL. Which finds how the model parameters of different tasks are related, which leads to ranking. It is more robust to outliers.

Instance-based MTSL. Which aims to find in one task the instances useful for other tasks. It is not used too much.

In unsupervised multitask learning, each task, can be a clustering problem, aims to identify useful patterns contained in a training dataset composed only of data instances.

In multitask semi-supervised learning, each task aims to predict the labels of data instances from the labeled and also unlabeled data.

3 COMPARATIVE STUDY

3.1 Models Categories

After an in-depth bibliographical study on the progression of AD, and after consulting a large number of articles that seemed relevant to us, two main ideas emerge: Research on modeling disease progression has received special attention from researchers around the world, and the modeling has been divided into four categories.

Mono-modal Mono-task Learning. Where the model optimizes only one objective function based on a type of data. Through this model, neither the correlation between the tasks, nor the collective information between the modalities are explored.

Mono-modal Multitask Learning. Where the tasks share a few training instances. The relationship between the spots is modeled assuming that they share a common representation space or they share certain parameters. However, they don't take into account the relationship between the different modalities of the same task.

Multi-modal Mono-task Learning. Where several modalities are taken into consideration in the prediction of a state while ignoring information from other tasks.

Multi-modal Multitask Learning. Where each task of the problem has characteristics from several

modalities and or several tasks are linked to each other in a chronological sequence.

Artificial intelligence is advancing in understanding the world around us and the operating model that most closely matches our real environment and provides greater precision is multitask multimodal learning

3.2 Related Works

In our bibliographical study, we focused on the following articles which seem to us the most relevant and which meet our expectations.

Mono-modal Mono-task Learning. (Cui et al., 2019) proposed a single-task, monomodal model based on six-step MRI time series data for the detection of AD progression. They use a stacked CNN-BGRU (Convolutional Neural Network Bidirectional Gated Recurrent Unit) pipeline. It achieves an accuracy of 91.33% for AD vs CN (normal cognitive), and 71.71% for pMCI (progressive MCI) vs sMCI (stable MCI). However, relying on MRI alone is insufficient in the medical field.

Mono-modal Multitask Learning. (Liu et al., 2018) proposed a deep multi-channel multi-task convolutional neural network for classification (Multi-class classification) and regression using MRI data and BG demographic information.

(Lopez-de Ipinia et al., 2018) have developed a new nonlinear multitasking (three tasks: animal naming (AN), picture description (PD) and spontaneous speech (SS)) approach based on automatic speech analysis. They introduced linear features, perceptual features, Castiglioni fractal dimension and Multiscale Permutation Entropy into their analysis.

Based on the MRI data, they performed a classification, using Multilayer Perceptron (MLP) and Deep Learning by means of Convolutional Neural Networks (CNN) (biologically- inspired variants of MLPs) which led to promising results.

Multi-modal Mono-task Learning. (Pan et al., 2018) proposed a two-step deep learning framework for the diagnosis of AD using both MRI and PET data. In the first step, they assign the corresponding MRI data to the missing PET data using 3D-cGAN (3D Cycle-consistent Generative Adversarial Networks) to capture their underlying relationship. In the second step, they develop LM3IL (Landmark-based Multi-modal Multi-Instance Learning Network) which learns and merges the characteristics necessary for the diagnosis of AD and the prediction of MCI.

Using the ADNI dataset, (Shi et al., 2017) proposed a method of transforming nonlinear feature space into more linearly separable data using SVM. They chose TPS (Thin-platespline) as a geometric model because of its power of representation. They also adopted a feature fusion strategy based on a deep network by SDAE (Denoising Sparse Auto-Encoder) to merge the transverse and longitudinal features estimated from MRI brain images.

(Lee et al., 2019) proposed a one-task multimodal deep learning approach by incorporating multi-domain longitudinal data. They applied a GRU (Gated Recurrent Unit) for each modality (4Gru) to produce feature vectors of fixed size which will be concatenated to form an input for the final prediction where regularized logistic regression is used for the classification of MCI- C and MCI-NC (MCI converter and MCI non-converter). The model on a very small number of features (no addition of BG data) has better prediction, it only optimizes binary classification tasks.

Multi-modal Multitask Learning. (Lahmiri and Shmuel, 2019) proposed M3T multimodal multitask model for the prediction of AD over two years. The latter has two main stages: (1) Selecting multitask features using the Lasso (2) Using an SVM (Support Vector Machine) model for separate classification and regression.

Multi-source multitask learning (MSMT) simultaneously considers two types of prior knowledge. 1) Source consistency 2) Slow temporal evolution (Nie et al., 2016) proposed a linear MSMT model that predicts the future disease status over 2 years (M06-M48) of new patients, based on their health information at the first moment (Baseline).

(Tabarestani et al., 2020) propose a distributed multitask multimodal model to predict MMSE cognitive measures of AD progression, the latter individually exploits several multitask regression coefficient matrices for each modality, then It concatenates the risk factors with the predicted y of each modality then it goes through gradient boosting to group the results of different modalities and reduce their prediction error.

(Nie et al., 2015) proposed an adaptive multimodal multitask linear learning model (aM2L) to regularize the modality agreement for the same task, the temporal progression on the same modality and the weight of the modalities.

(Nie et al., 2016) and (Nie et al., 2015) in their predictions, don't incorporate the follow-up observations. Example: To predict her condition at M24 (24 months after Baseline), they don't merge the observations of Baseline, M06 AND M12.

(Li et al., 2015) presented a robust deep learning system to identify the different stages of progression of patients with AD based on MRI and PET. They used the dropout technique to improve classical deep learning by preventing its weight co-adaptation, which is a typical cause of deep learning overfitting. They stacked multiple RBMs (Restricted Boltzmann machine) to build a robust deep learning framework, which integrates stability selection and multi-task learning strategy.

(El-Sappagh et al., 2020) proposed a robust ensemble deep learning model based on a stacked convolutional neural network (CNN) and a bi-directional long-term memory network (BiLSTM). This multimodal multitask model jointly predicts several variables based on the fusion of five types of multimodal time series data plus a basic knowledge set (BG). The predicted variables include the multi-class task and four critical cognitive score regression tasks. This model gave equal weights for the classification and regression tasks.

The table below 1 is a summary of some works seen previously, it lists the advantages and limitations of its different models.

3.3 Synthesis of Models

All the models seen previously for the study of the prediction of the progression of AD, use the ADNI database which can contain missing data. In order to overcome this problem, among the methods that were used:

DEL: We simply eliminated the subjects with either missing sources or missing labels.

ZERO: We assigned zero value to any element that is missing.

KNN: The k-nearest neighbor (KNN) method replaced the missing value in the data matrix with the corresponding value from the nearest column ...

In recent years, machine learning algorithms (such as SVM and random forest, SDAE) and deep learning (CNN, GRU, LSTM, BGRU, BiLSTM,...) have been used for the design of a predictive model of AD progression. And for the generalization of these models to avoid overfitting ,among the regularizers that have been used ,we find: lasso ,sparse group lasso,l2,1 norm, ridge regression ,fused group lasso(convex and non-convex),dropout

(Liu et al., 2013) and (Duchesne et al., 2009) used regular machine learning techniques to study multimodal single-task classification and regression, respectively.

To detect AD progression based on a multimodal single-task deep learning model, we find; (Spasov

et al., 2019) who proposed a classification model based on a CNN ,(Lee et al., 2019) who also proposed a binary classification model based on GRU and used logistic regression for regularization.

In a real medical environment, many modalities are analyzed and multiple clinical variables must be predicted. Multimodal single-task models will not provide in this case sufficient information for the study of AD progression, hence the interest of the multimodal multitask model.

This model has been used by (Lahmiri and Shmuel, 2019) for classification and regression using SVM and SVR respectively. In order to generalize their model, they used the Lasso regularizer.

It was also used by el sappagh who added to their model the time series constraint which is consistent with the longitudinal nature of AD because the patient's state at a given time is not independent of his state at an earlier time.

They used the CNN-BILTSM for multiclass classification and for regression of 4 cognitive scores. For the regularization of their model, they opted for the dropout.

(El-Sappagh et al., 2020) consider that the modalities have the same weight of importance which is not always true in the medical field.

4 PROPOSED MODEL

Alzheimer's disease (AD) is a longitudinal disease, that is to say the state of a patient at a time t depends strongly on his state at $t-1$. AD cannot be triggered spontaneously; there must be a prior appearance of signs that predict the manifestation and evolution of the disease. Most models studied in the previous section do not take this notion into consideration.

Several factors can influence the diagnosis outcome of AD, and even the most indirect symptoms may have in some cases an effect on the detection of AD, hence the interest of introducing Multitask Learning. Multitask Learning will allow us to find a relationship between the features, the instances and the parameters involved in the detection of the disease.

Moreover, patients' data is considered multimodal since they're gathered from various sources such as MRI, PET, CSD, ASD, NPD etc... These modalities can also interact with each other, because the information extracted from a modality can be complementary to other modalities, which opens the possibility of improving the performance of the model for AD progression prediction. Additionally, it is important to note that the data retrieved via different modalities

Table 1: Comparative table of models for predicting the progression of Alzheimer’s disease.

Study	Data	Modality	Merging	Time-series	Performance	Model	Task
(Lee et al., 2019)	ADNI (Database, 2004)	Demographic, MRI, CSD, CSF(LCR)	NO	YES (4 Steps)	Accuracy: 81% (MCI/AD)	GRU	Classification
(Lahmiri and Shmuel, 2019)	ADNI	MRI, FDG-PET, LCR	NO	NO	Classification Accuracy: 93,3%(CN/AD), 83,2%(CN/MCI), 73,9%(sMCI/pMCI). Regression Accuracy: 0,697(MMSE), 0,739(ADAS)	SVM	Classification and regression
(Nie et al., 2016)	ADNI	MRI,PET, CSF,PROPT, META	NO	NO	ADAS-Cog =90,94% MMSE=87,94%	MSMT	Regression
(Cui et al., 2019)	ADNI	MRI	NO	YES(6 Steps)	Classification: Accuracy: 91,33% (AD/ NC), 71,71% (pMCI:progressive MCI / sMCI:static MCI)	Stacked CNN-BGRU	Classification
(Tabarestani et al., 2020)	ADNI	MRI, PET, COG, CSF	YES	NO	MMSE: 70.1%	distributed multimodal multitask learning	Regression
(Liu et al., 2018)	ADNI	MRI	YES	NO	Classification Accuracy: 51,8% (CN/sMCI/pMCI/AD). Regression: (CDRSB, ADAS11, ADAS13, MMSE): 1.666, 6.2, 8.537, 2.373	CNN	Classification and regression
(Shi et al., 2017)	ADNI	MRI, Age	YES	NO	Classification: TML(Theoretic Learning)-SVM Metric Accuracy:AD/NC:91.95% MCI/NC:83,72% Multi-modal S-DSAE Accuracy:80.91%(MCI/NC) 88.73%(AD/NC)	TML-SVM SDAE (Stacked Denoising Sparse AE)	Classification
(Pan et al., 2018)	ADNI	MRI, PET	YES	NO	Classification Accuracy: 92,5%(HC/AD) 79,06%(sMCI/pMCI)	3D-CNN+GAN LM3IL	Classification
(Li et al., 2015)	ADNI	MRI,PET, CSF	NO	NO	Classification Accuracy: 91.4%(AD /HC), 77.4%(MCI/HC), 70.1% (AD/MCI),and 57.4%(MCI.c/MCI.NC)	RBM	Classification
(Nie et al., 2015)	ADNI	MRI, PET, CSF, PROPT, META	NO	NO	MMSE:89,01 % ADAS-Cog: 91,68%	aM2L	Regression
(El-Sappagh et al., 2020)	ADNI	MRI, PET, CSD, ASD, NPd	YES	YES(15 Steps)	ACC: 92.62%, PRE: 94.02%, F1: 92.56%, REC: 98.42. MAE: 0.107, 0.076, 0.075, and 0.085, (FAQ, ADAS, CDR, MMSE)	StackedCNN-BiLSTM	Classification and regression

may vary in terms of importance. For instance, MRI and PET are considered a more reliable and relevant data sources, hence the interest in introducing the idea of a weighted model.

According to the results of the comparative study seen in the previous section, which follows closely the same direction mentioned previously, we can conclude that the most adequate, precise and stable model is the Multimodal Multitask Model based on time series.

The research axis that can be developed from this study, will center around a Multimodal Multitask system based on time series, which will be able to simultaneously regularize the modality weighting, the temporal progression as well as the modality agreement, i.e. the status of the patient estimated by different modalities must be consistent. The model will also be capable of taking into consideration the relationship between AD progression and the patient's comorbidities (cardiovascular disease, depression, genitourinary renal metabolism, endocrine, etc.). Such model has not yet been addressed in the literature.

To advance on this research axis, we propose a Machine Learning or a Deep Learning architecture, multitask, multimodal (using MRI, PET, CSD, ASD, NPD), trained on the ADNI database and the patients' demographic data. This model predicts AD progression status of a through a multiclass classification task, and the values of two cognitive scores ADAS, MMSE which will be implemented as two regression tasks.

In order to meet the different expectations as well as possible, two possible scenarios seem to open up at the moment:

- The first approach consists in extracting separately the temporal characteristics of each modality (MRI, PET, CSD, ASD, NPD), then merged with the demographic data to extract the common characteristics that respond to each task by applying a multitask learning.

- Or by applying multitasking to each modality and merging the initial results with demographic data, which is assumed to be a time-invariant information. This model has the ability to stop the propagation of an error from one modality to another.

The loss function in both approaches will include a first term for classification, a second term for regression and 2 last terms to regularize the weight and the agreement of the modalities.

5 CONCLUSION

Alzheimer's disease (Thung et al., 2017) is currently calling the attention of many researchers. A considerable amount of effort is being put to understand the biological and physiological mechanisms of AD as well as its monitoring and early detection. The presented works in this paper have raised several approaches for the detection of AD, namely Monomodal monotask, Monomodal multitask Multimodal monotask and Multimodal multitasking. Following our comparative study of different works and papers, we believe that a major research axis has been revealed in which the Multimodal Multitask based on time series could at the same time to regularize the modality weighting, the temporal progression, the modality agreement and which will even take into consideration the relationship between the progression of AD and the patient's comorbidities. This model could bring a considerable advance in the field of medical research.

REFERENCES

- Baheti, P. (2020). *Introduction to Multimodal Deep Learning*. <https://heartbeat.fritz.ai/introduction-to-multimodal-deep-learning-630b259f9291>.
- Cui, R., Liu, M., Initiative, A. D. N., et al. (2019). Rnn-based longitudinal analysis for diagnosis of alzheimer's disease. *Computerized Medical Imaging and Graphics*, 73:1–10.
- Database (2004). *Introduction to Multimodal Deep Learning*. <http://adni.loni.usc.edu>.
- Duchesne, S., Caroli, A., Geroldi, C., Collins, D. L., and Frisoni, G. B. (2009). Relating one-year cognitive change in mild cognitive impairment to baseline mri features. *Neuroimage*, 47(4):1363–1370.
- El-Sappagh, S., Abuhmed, T., Islam, S. R., and Kwak, K. S. (2020). Multimodal multitask deep learning model for alzheimer's disease progression detection based on time series data. *Neurocomputing*, 412:197–215.
- Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691.
- Jo, T., Nho, K., and Saykin, A. J. (2019). Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Frontiers in aging neuroscience*, 11:220.
- Lahmiri, S. and Shmuel, A. (2019). Performance of machine learning methods applied to structural mri and adas cognitive scores in diagnosing alzheimer's dis-

- ease. *Biomedical Signal Processing and Control*, 52:414–419.
- Lee, G., Nho, K., Kang, B., Sohn, K.-A., and Kim, D. (2019). Predicting alzheimer’s disease progression using multi-modal deep learning approach. *Scientific reports*, 9(1):1–12.
- Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., and Li, J. (2015). A robust deep model for improved classification of ad/mci patients. *IEEE journal of biomedical and health informatics*, 19(5):1610–1616.
- Liu, F., Zhou, L., Shen, C., and Yin, J. (2013). Multiple kernel learning in the primal for multimodal alzheimer’s disease classification. *IEEE journal of biomedical and health informatics*, 18(3):984–990.
- Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Joint classification and regression via deep multi-task multi-channel learning for alzheimer’s disease diagnosis. *IEEE Transactions on Biomedical Engineering*, 66(5):1195–1206.
- Lopez-de Ipina, K., Martinez-de Lizarduy, U., Calvo, P. M., Mekyska, J., Beitia, B., Barroso, N., Estanga, A., Tainta, M., and Ecay-Torres, M. (2018). Advances on automatic speech analysis for early detection of alzheimer disease: a non-linear multi-task approach. *Current Alzheimer Research*, 15(2):139–148.
- Nie, L., Zhang, L., Meng, L., Song, X., Chang, X., and Li, X. (2016). Modeling disease progression via multisource multitask learners: A case study with alzheimer’s disease. *IEEE transactions on neural networks and learning systems*, 28(7):1508–1519.
- Nie, L., Zhang, L., Yang, Y., Wang, M., Hong, R., and Chua, T.-S. (2015). Beyond doctors: Future health prediction from multimedia and multimodal observations. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 591–600.
- Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., and Shen, D. (2018). Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer’s disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 455–463. Springer.
- Pironkov, G., Dupont, S., and Dutoit, T. (2016). Multi-task learning for speech recognition: an overview. In *ESANN*.
- Shi, B., Chen, Y., Zhang, P., Smith, C. D., Liu, J., Initiative, A. D. N., et al. (2017). Nonlinear feature transformation and deep fusion for alzheimer’s disease staging analysis. *Pattern recognition*, 63:487–498.
- Spasov, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., Initiative, A. D. N., et al. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer’s disease. *Neuroimage*, 189:276–287.
- Tabarestani, S., Aghili, M., Eslami, M., Cabrerizo, M., Barreto, A., Rische, N., Curiel, R. E., Loewenstein, D., Duara, R., and Adjouadi, M. (2020). A distributed multitask multimodal approach for the prediction of alzheimer’s disease in a longitudinal study. *NeuroImage*, 206:116317.
- Thung, K.-H., Yap, P.-T., and Shen, D. (2017). Multi-stage diagnosis of alzheimer’s disease with incomplete multimodal data via multi-task deep learning. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 160–168. Springer.
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., et al. (2013). The alzheimer’s disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer’s & Dementia*, 9(5):e111–e194.
- Worsham, J. and Kalita, J. (2020). Multi-task learning for natural language processing in the 2020s: Where are we going? *Pattern Recognition Letters*, 136:120–126.
- Zhang, D., Shen, D., Initiative, A. D. N., et al. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer’s disease. *NeuroImage*, 59(2):895–907.
- Zhang, L., Wang, M., Liu, M., and Zhang, D. (2020). A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in neuroscience*, 14.
- Zhang, Y. and Yang, Q. (2018). An overview of multi-task learning. *National Science Review*, 5(1):30–43.