# Addressing imaging accessibility by cross-modality transfer learning

Zheng, Zhiyang, Su, Yi, Chen, Kewei, Weidman, David, Wu, Teresa, et al.

Zhiyang Zheng, Yi Su, Kewei Chen, David A. Weidman, Teresa Wu, Ben Lo, Fleming Lure, Jing Li, "Addressing imaging accessibility by cross-modality transfer learning," Proc. SPIE 12033, Medical Imaging 2022: Computer-Aided Diagnosis, 120330X (4 April 2022); doi: 10.1117/12.2611791

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

# Addressing imaging accessibility by cross-modality transfer learning

Zhiyang Zheng[a], Yi Su[b], Kewei Chen[b], David A. Weidman[b], Teresa Wu[c], Ben Lo[d], Fleming Lure[d], Jing Li[a]

[a]School of Industrial and Systems Engineering, Georgia Institute of Technology, GA USA; [b]Banner Alzheimer's Institute, AZ USA; [c]School of Computing, Informatics and Decision Systems Engineering, Arizona State University, AZ USA; [d]MS Technologies Corp, Rockville, MD USA

## ABSTRACT

Multi-modality images usually exist for diagnosis/prognosis of a disease, such as Alzheimer's Disease (AD), but with different levels of accessibility and accuracy. MRI is used in the standard of care, thus having high accessibility to patients. On the other hand, imaging of pathologic hallmarks of AD such as amyloid-PET and tau-PET has low accessibility due to cost and other practical constraints, even though they are expected to provide higher diagnostic/prognostic accuracy than standard clinical MRI. We proposed Cross-Modality Transfer Learning (CMTL) for accurate diagnosis/prognosis based on standard imaging modality with high accessibility (mod_HA), with a novel training strategy of using not only data of mod_HA but also knowledge transferred from the model based on advanced imaging modality with low accessibility (mod_LA). We applied CMTL to predict conversion of individuals with Mild Cognitive Impairment (MCI) to AD using the Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets, demonstrating improved performance of the MRI (mod_HA)-based model by leveraging the knowledge transferred from the model based on tau-PET (mod_LA).

**Keywords:** Multi-modality images, transfer learning, knowledge distillation, Alzheimer's disease, mild cognitive impairment

## 1. INTRODUCTION

Imaging has become an important tool to facilitate clinical diagnosis and prognosis. Focus on a specific disease, different imaging modalities usually exist but with different levels of accessibility and accuracy. Taking Alzheimer' Disease (AD) as an example, an important task is to predict, while individuals are still at the early Mild Cognitive Impairment (MCI) stage, who will convert to AD dementia in several years. This would allow early intervention for these individuals to slow down their progression to AD[1]. Imaging has shown great promise to facilitate the prediction of conversion to AD for MCI subjects. Among the different imaging modalities, MRI is used in the standard of care, thus having high accessibility to patients. On the other hand, imaging of pathologic hallmarks of AD such as amyloid-PET and tau-PET[2] has low accessibility due to cost and other practical constraints, even though they are expected to provide higher diagnostic/prognostic accuracy than standard clinical MRI.

The goal of this work is to develop a Cross-Modality Transfer Learning (CMTL) model for accurate diagnosis/prognosis based on standard imaging modality with high accessibility (mod_HA) such as MRI, with a novel training strategy of not only using the data of mod_HA but also leveraging knowledge transferred from the diagnostic/prognostic model based on advanced imaging modality with low accessibility (mod_LA) such as amyloid-PET or tau-PET.

The design of CMTL adapts the Knowledge Distillation[3] (KD) process, a Deep Neural Networks (DNN)-based framework for network compression. The purpose of KD is to transfer/distil knowledge from a "teacher" model to a "student" one. The teacher model typically has higher knowledge capability and better generalizability. On the other hand, the teacher model also has a complicated structure with a large number of parameters. This makes it computationally expensive and is not desirable in some applications, such as when deployed on a mobile device. This drives the need of transferring the knowledge learned by the teacher model about the problem domain to a student model which typically has a less complex structure with fewer parameters. By distilling the knowledge from the teacher model, the student model is expected to have better performance than its naïve version without KD, while at the same time being computationally efficient. Borrowing

Medical Imaging 2022: Computer-Aided Diagnosis, edited by Karen Drukker, Khan M. Iftekharuddin, Hongbing Lu, Maciej A. Mazurowski, Chisako Muramatsu, Ravi K. Samala, Proc. of SPIE Vol. 12033, 120330X · © 2022 SPIE · 1605-7422 · doi: 10.1117/12.2611791

the teacher-student concept of KD to the design of CMTL, we consider the diagnostic/prognostic model based on mod_LA (e.g., tau-PET) as the teacher model, and that based on mod_HA (e.g., MRI) as the student model. However, different from the original KD, the teacher/student model in CMTL does not correspond to the model with more/less structural complexity and parameter size but instead corresponds to imaging modalities with low/high accessibility in serving the diagnostic or prognostic purpose of disease.

Most previous studies focused on predicting MCI to AD conversion using single modality[4-5] or combined multiple modalities[6-7] but did not consider the possibility to use the mod_LA to enhance the performance of the model using mod_HA. This paper is the first work using KD for transfer learning from the diagnostic/prognostic model based on advanced but limitedly-accessible imaging modality to the model based on standard but broadly-accessible imaging modality, in the application of predicting the conversion of individuals with MCI to AD dementia. By doing this, one can build the best possible diagnostic/prognostic model based on standard imaging modality, so that the general patient population can benefit from this capacity instead of just a small portion of people who have access to the advanced modality. Ultimately, CMTL can help improve the access of patients to advanced diagnostic/prognostic capacity.

# 2. DATA

## 2.1 Patient inclusion

We used data from the Alzheimer's Disease Neuroimaging Initiative[8] (ADNI) datasets. MRI and tau-PET were considered as mod_HA and mod_LA, respectively. This study included 936 MCI subjects who have MRI, with 213 subjects converting to AD in 36 months—namely converters (CON), while the remaining subjects are considered as non-converters (N-CON). Among these subjects, 49 also have tau-PET (13 CON and 36 N-CON). 67 AD and 393 normal controls (NC) who have tau-PET were also included.

## 2.2 Image processing and feature extraction

We used a regional standardized update value ratio (SUVR) of tau-PET images extracted by the ADNI PET core, among which 84 cortical and sub-cortical SUVR features were included. All features were normalized using the inferior cerebellum as the reference region. MRI was processed using FreeSurfer v7. 194 regional thickness and volumetric features were included. Volumetric features were normalized by estimated total intracranial volume.

## 2.3 Other clinical features

In addition to imaging features, commonly used demographic and cognitive measures such as gender, age, education, MMSE score, and ADAS score were also included.

# 3. METHODOLOGY

## 3.1 Training of CMTL

Two steps were included in the training process of the CMTL model (Fig.1):

*(1). Pre-training the teacher model.* The success of KD requires a superior teacher model. In our case, the teacher model is one using tau-PET to classify CON and N-CON. In related work, we found that classifiers trained using the tau-PET of AD and NC can achieve high accuracy in classifying CON and N-CON. Using the same strategy, we trained a feed-forward neural network (NN) with 4 layers and 800 neurons per layer based on the tau-PET of 67 AD and 393 NC. A separate dataset consisting of 33 CON and 47 N-CON was used as the validation set. Early stopping was used to prevent overfitting. When stopped, the trained model achieved 0.98 sensitivity and 0.99 specificity on the validation set.

*(2) Training the teacher model with KD.* The student model is one using MRI only. Even though the input imaging modalities to the teacher and student models are different, the NN structure of the student model is kept the same as the teacher to facilitate KD. To train the parameters of the student model using KD, one part of the input is the MRI of MCI subjects. 80% of the 936 MCI subjects with MRI were included in training while the remaining 20% were left for testing. The other part of the input is distilled knowledge from the pre-trained teacher model in step (1). To achieve this, the 49 MCI subjects with both MRI and tau-PET were put in a "transfer data set". Then, they were input into both the teacher

model which uses only the tau-PET of these subjects and the student model which only uses their MRI. KD is enabled by encouraging the outputs of the same layer and the prediction by the student and teacher models on the same subject to be similar. Mathematically, the student model with KD is trained to minimize the following loss function:

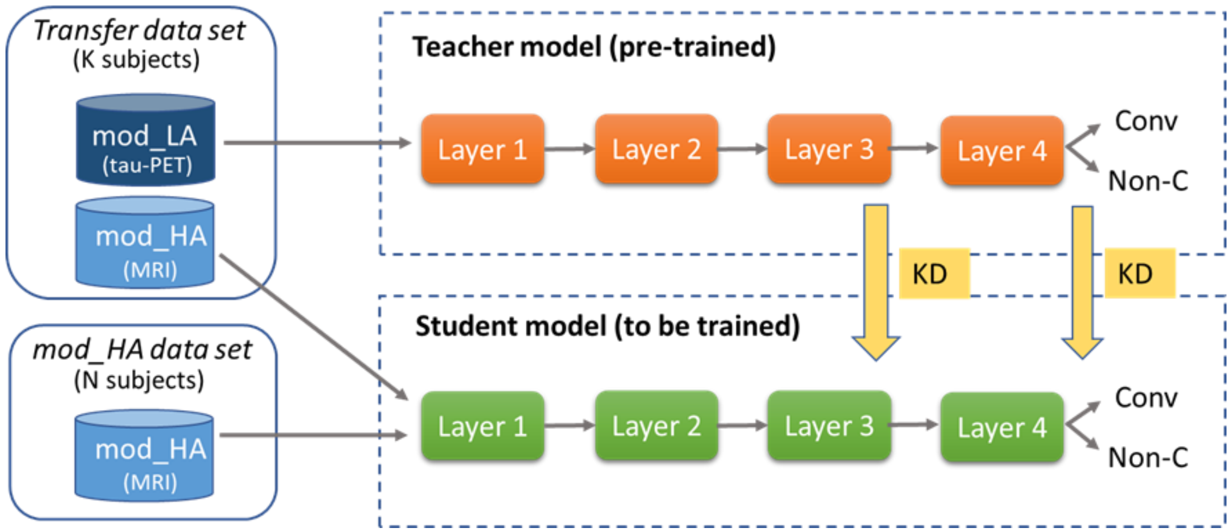


Figure 1. Proposed cross-modality transfer learning (CMTL) through knowledge distillation (KD).

$$L = \sum_{i=1}^{N} L_{CE}^{(i)} + \alpha \sum_{j=1}^{K} L_{KD-pred}^{(j)} + \beta \sum_{j=1}^{K} L_{KD-layer}^{(j)}. \tag{1}$$

$N$ is the number of subjects with MRI in the training set. $K$ is the number of subjects with both MRI and tau-PET, i.e., subjects in the transfer dataset. $L_{CE}^{(i)}$, $i = 1, \dots, N$, is the cross-entropy loss by comparing the predicted probability of subject $i$ to the true class label (i.e., CON and N-CON). $L_{KD-pred}^{(j)}$ is the KL-divergence between the soft predictions from the teacher and student models on subject $j$. The soft prediction from a model is defined as a softmax activated probability on class c, $q_c^{(j)} = \frac{exp\,(z_c^{(j)}/T)}{\sum_l exp\,(z_l^{(j)}/T)}$, where $z_c^{(j)}$ denotes the logit of class c for subject j and T denotes the temperature parameter which controls the softness of the prediction. $L_{KD-layer}^{(j)}$ is the mean square error loss between the same-layer outputs of the student and teacher models. In this study, layer 3 of the student and teacher NNs was chosen to impose this loss, which led to the best performance. $\alpha$, $\beta$, and $T$ are tuning parameters.

### 3.2 Training of the student model without KD (comparison method)

The student model is re-trained without KD, equivalent to setting α, β to be zero in Eq. (1), to evaluate the benefit of KD.

### 3.3 Uncertainty quantification (UQ)

The relatively limited sample size and inherent randomness in NN training make it important for the user to know how uncertain the model is for its prediction on a new subject. This helps the user/clinician make a more informed decision. For both the CMTL model and the student model without KD, we use Bayesian dropout to quantify the uncertainty of the model prediction. With dropout enabled, we can obtain $T$ predictions for each subject, $p_t$, i.e., the predicted probability of the subject belonging to one class, $t = 1, \dots, T$. Then, we can compute the mean and variance of the $p_t$'s as the prediction result and uncertainty score, respectively.

## 4. RESULTS

Based on 50 times random training-test split of the dataset, the student model without KD has, on average, 0.668 sensitivity and 0.841 specificity on the test set in the prediction of MCI conversion (43 CON and 139 N-CON in the test set). Using

CMCT, the sensitivity improves to 0.702 (p=0.036), while the specificity has no significant change. Furthermore, we ordered the converters according to the uncertainty scores of their predictions by each model from the highest to the lowest. We removed subjects whose uncertainty scores rank in the top x%, with x%=10%, ...,50%, and computed the sensitivity of the remaining subjects with more certain predictions. The rationale behind this is that if the model is uncertain about its prediction on a subject, the prediction result should not be trusted. The same was done for non-converters. Fig. 2 shows that the sensitivity and specificity of both models improve as more subjects with uncertain predictions are removed. For example, with the top 50% uncertain predictions removed, the sensitivity and specificity of CMCT increase to 0.786 and 0.921. CMCT significantly outperformed the student model without KD in sensitivity when the same percentage of uncertain predictions was removed from each model. The specificity of the two models remained statistically the same.

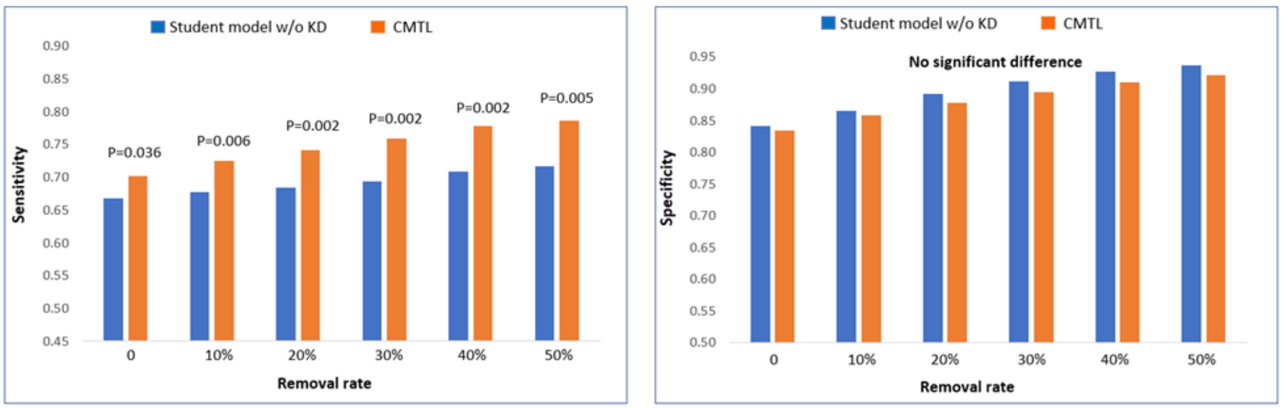


Figure 2. Comparison of CMTL and student model without KD on sensitivity (left) and specificity (right) with respect to subject removal rate according to prediction uncertainty score.

## 5. CONCLUSIONS

We proposed a CMCT model for accurate diagnosis/prognosis based on standard imaging modality with high accessibility (mod_HA), with a novel training strategy of using not only data of mod_HA but also knowledge transferred from the model based on advanced imaging modality with low accessibility (mod_LA). We applied CMTL to the prediction of conversion to AD for individuals with MCI. The result indicated that CMCT improved the performance of the MRI (mod_HA)-based model by leveraging the knowledge transferred from the model based on tau-PET (mod_HA). One limitation of the present study is the small sample size in the transfer data set for the student model to distil knowledge from the teacher model. With a larger sample size in a future study, we expect that more significant improvement of CMCT compared to the student model without KD can be obtained.

## 6. DECLAIMERS

This work is not being, or has been, submitted for publication or presentation elsewhere. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number 2R42AG053149-02A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES

[1] Tanveer, M., et al. "Machine learning techniques for the diagnosis of Alzheimer's disease: A review." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16.1s: 1-35 (2020).

[2] Johnson KA., Schultz A., Betensky RA., Becker JA., Sepulcre J., Rentz D., et al., “Tau positron emission tomographic imaging in aging and early Alzheimer disease,” Ann Neurol 79:110–9 (2016).
[3] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531 (2015).
[4] Jo, Taeho, et al. "Deep learning detection of informative features in tau PET for Alzheimer’s disease classification." BMC bioinformatics 21.21: 1-13 (2020).
[5] Spasov, Simeon, et al. "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease." Neuroimage 189: 276-287 (2019).
[6] Liu, Yanbei, et al. "Incomplete multi-modal representation learning for Alzheimer’s disease diagnosis." Medical Image Analysis 69: 101953 (2021).
[7] Punjabi, Arjun, et al. "Neuroimaging modality fusion in Alzheimer’s classification using convolutional neural networks." PloS One 14.12: e0225759 (2019).
[8] Weiner, Michael W., et al. "The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement." Alzheimer's & Dementia 13.5: 561-571 (2017).