DISCRIMINATIVE FUSION OF MULTIPLE BRAIN NETWORKS FOR EARLY MILD COGNITIVE IMPAIRMENT DETECTION

Qi Wang[1] *Liang Zhan*[2] *Paul M. Thompson*[3] *Hiroko H. Dodge*[4,5] *Jiayu Zhou*[1]

[1] Computer Science and Engineering, Michigan State University, East Lansing, MI 48824
[2] Computer Engineering Program, University of Wisconsin-Stout, Menomonie, WI 54751
[3] Imaging Genetics Center, University of Southern California, Marina del Rey, CA 90032
[4] Michigan Alzheimer's Disease Center, University of Michigan, Ann Arbor, MI 48105
[5] Department of Neurology, Oregon Health & Science University, Portland, OR 97239

ABSTRACT

In neuroimaging research, brain networks derived from different tractography methods may lead to different results and perform differently when used in classification tasks. As there is no ground truth to determine which brain network models are most accurate or most sensitive to group differences, we developed a new sparse learning method that combines information from multiple network models. We used it to learn a convex combination of brain connectivity matrices from 9 different tractography methods, to optimally distinguish people with early mild cognitive impairment from healthy control subjects, based on the structural connectivity patterns. Our fused networks outperformed the best single network model, Probtrackx (0.89 versus 0.77 cross- validated AUC), suggesting its potential for numerous connectivity analysis.

Index Terms— Magnetic Resonance Imaging, Brain Connectome, Discriminative Fusion, Classification, Mild Cognitive Impairment

1. INTRODUCTION

With the development of diffusion-weighted magnetic resonance imaging (dMRI) techniques that can map patterns of connections in the brain, many researchers have begun to model the brain as a network of interconnected brain regions, or connectome [16]. The properties of these networks can then be studied mathematically with network theory. Mathematically, a brain network at the macro-scale is typically expressed by a connectivity matrix, in which each element represents some property of the connection between each pair of brain regions [15]. Based on applying analytical methods from graph theory, the brain exhibits significant group differences in network properties in various brain diseases such as bipolar disorder [2, 10], Alzheimer's Disease [20], and body dysmorphic disorder [3], among others. These network-derived features provide clues about how characteristic network disruptions occur and how they may progress in these brain diseases.

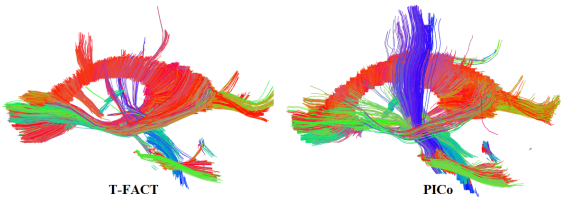

Fig. 1. Different tractography methods detect different sets of fibers. Here we show the fibers generated by two tractography algorithms (T-FACT [13] and PICo [14]), passing through the same brain slice.

Different MRI scanning modalities offer different information as a basis to generate maps of the brain's connectome. Here we focus on brain structural networks derived from diffusion MRI. Diffusion MRI is a variant of standard anatomical MRI that is sensitive to microscopic properties of the brain's white matter that are not detectable with standard anatomical MRI. The general process of reconstructing a structural brain network includes two main steps [21]. The first step extracts the dominant diffusion direction(s) at each voxel based on a diffusion MRI signal model. Some popular models include the diffusion tensor, the orientation distribution function (ODF), or a probabilistic mixture of tensors [11], among others. The next step is whole brain tractography based on these voxel- level diffusion direction(s). Currently, there are two main classes of tractography methods: deterministic and probabilistic approach. Based on whole brain tractography result, brain networks can be computed by combining the pattern of fiber tracts with some specific anatomical partitioning scheme, and measuring

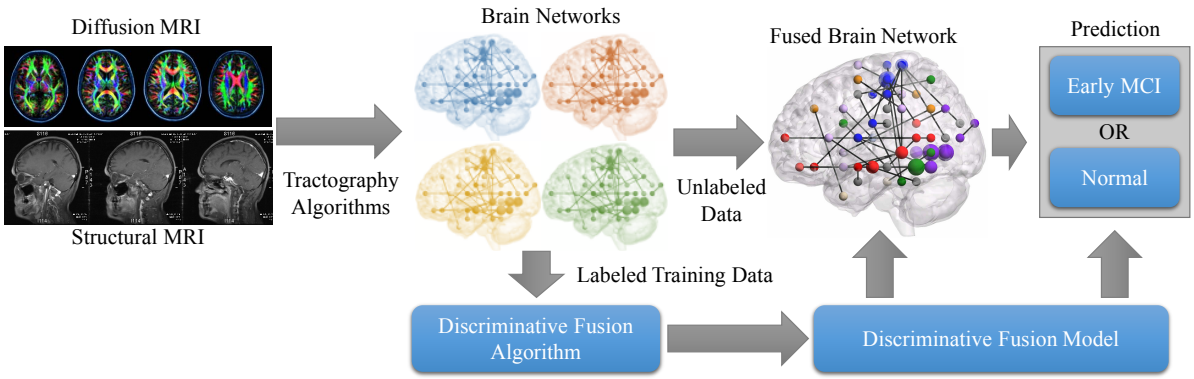

Fig. 2. Overview of our network fusion framework. Multiple types of brain networks are computed by applying different tractography methods to the participants' diffusion MRI data [21]. Different brain networks are combined using a sparse learning method and the optimal convex combination is used for classification. The combination coefficients and the classifiers are simultaneously learned from the training data and cross-validated.

some property of the connections between each pair of brain regions, such as their density or integrity.

Theoretically, different algorithmic methods to map structural connections should ultimately provide a consistent anatomical description of the brain. Even so, this may not be true in reality. Different tractography methods recover different sets of fibers (**Fig. 1**), and the fiber bundles that best differentiate patients from controls may be extracted by some algorithms but not others [19]. Different tract tracking methods vary in their ability to perform robustly on dataset of different quality. And there is no general principle to decide which tractography method or network model is most sensitive to disease effects in clinical research studies [21]. We therefore aimed to create a sparse learning framework to optimally fuse network methods, in an effort to combine different methods' advantages and boost statistical power in studies of brain disease.

2. METHODOLOGY

Overview: **Fig. 2** summarizes the overview of our fusion approach to build "consensus networks" based on fusing networks from multiple tract tracing methods. From diffusion MRI scans of multiple subjects, we extract different brain networks with whole brain tractography. Though our proposed fusion approach is not limited to structural networks computed from dMRI tractography, here we use the nine tractography methods studied in our previous work [21], which include methods that are classified as tensor-based deterministic, orientation distribution function (ODF)-based deterministic, and probabilistic approaches. Each network reconstruction method describes brain connectivity from a different perspective, and none is universally better than all others for diagnostic classifications tasks. Therefore when it comes to building models from diffusion MRI images, it is intuitive to fuse different brain networks and leverage the predictive information from all the networks. However, the key question is how to fuse the different networks and build effective predictive models from the fused models. As far as we know, there is no principled approach proposed to combine networks for use in predictive models. As shown in the experimental section, simple numerical averaging of nodal edge weights may not be able to boost the predictive performance. Instead, we propose to learn how to fuse the networks from data, such that the combination gives the optimal predictive performance. First, we study fused networks computed as a convex combination of different brain networks. We describe a new machine learning model to simultaneously learn the coefficients of the convex combination as well as the classifier parameters. As a result, the combination coefficients are learned to maximize the predictive performance of the classifier and meanwhile the classifier is learned specifically to use the combined network.

Discriminative Fusion: Our proposed discriminative fusion (DFUSE) is a data-driven model that includes a training stage and a prediction stage. In the training stage the DFUSE algorithm learns the optimal combination coefficients and a logistic regression classifier from a set of patients with known medical classification. In the prediction stage, the brain networks from a patient are combined according to the coefficients. The combined network is then used by the classifier to give a prediction for the medical classification problem.

Formulation. Given a set of diffusion MRI scans from N patients, we apply different tractography methods to obtain M brain networks for each participant. Let $\mathbf{x}_i^{(m)}$ denotes a vector representation of the m-th brain network for patient i ($i \in [1, N], m \in [1, M]$), in which each element is a numerical representation

of a connection property (e.g., density or integrity) between two brain regions. We would like to combine all networks for each participant into a single network using a convex combination, i.e., the combined network $x_i(\boldsymbol{\tau}) = \sum_{m=1}^{M} \tau_m \mathbf{x}_i^{(m)}$, where $\boldsymbol{\tau} = [\tau_1 \dots \tau_M]$ is the vector of combination coefficients, and the convex combination gives $\sum_{m=1}^{M} \tau_m = 1; \tau_m \geq 0, \forall \tau_m$. Convex combination is one type of linear combination that gives a clear interpretation on how much each original network contributes to the fused network. For the N subjects used for training, we also have diagnostic label information stored in $\mathbf{y} = [y_1, \dots, y_N]$, where $y_i = 1$ if the patient is case and -1 if control.

To learn the combination of the networks, we propose a machine learning formulation that jointly learns the classifier parameters and the combination coefficient, which solves the following optimization problem:

$$\min_{\mathbf{w},c,\boldsymbol{\tau}} \Sigma_{i=1}^{N} \ell(\mathbf{w}, c, \boldsymbol{\tau}; \mathbf{x}_i, y_i) + \lambda \|\mathbf{w}\|_1, \qquad (2.1)$$
$$\text{s.t.} \quad \Sigma_{m=1}^{M} \tau_m = 1; \tau_m \geq 0, \forall \tau_m$$

where $\mathbf{w}$ and c are classifier parameters, the constraints on $\boldsymbol{\tau}$ ensures a convex combination, the logistic loss is: $\ell(\mathbf{w}, c, \boldsymbol{\tau}; \mathbf{x}_i, y_i) = \log\left(1 + exp\left(-y_i(x_i(\boldsymbol{\tau})^T \mathbf{w} + c)\right)\right)$. The ℓ_1-norm induces sparsity in the parameters $\mathbf{w}$ [12, 25, 24, 23], such that the classifier learns a subset of predictive connections and only uses these connections in the classifier. The sparsity parameter λ controls the sparsity of the model. A smaller λ allows more connections to be involved in the model. The optimization problem in (2.1) can be solved by proximal block coordinate descent [5, 17, 18]. Once the optimization process has converged, we obtain the optimal combination coefficients $\boldsymbol{\tau}^*$ and classifier parameters $\mathbf{w}^*$ and c^*.

3. RESULTS AND DISCUSSIONS

Dataset. The imaging datasets analyzed for in this study were collected from 16 sites across the United States and Canada in the second stage of the Northern American Alzheimer's Disease Neuroimaging Initiative (ADNI2). In total, 124 subjects' diffusion MRI and structural MRI data were analyzed. Detailed subject inclusion, exclusion criteria and scanning protocols can be found in the ADNI2 website. These 124 subjects include 51 normal elderly controls (NCs), 73 individuals diagnosed with early mild cognitive impairment (eMCI).

Brain Networks. For each subject, we computed 9 brain networks using nine methods, including 4 tensor-based deterministic algorithms: FACT (T-FACT) [13], the second-order Runge–Kutta (T-RK2) method [4], the tensorline (T-TL) [9], and interpolated streamline (T-SL) methods [7], two deterministic tractography algorithms based on fourth order spherical harmonic derived ODFs – FACT (O-FACT) and RK2 (O-RK2), and three probabilistic approaches: "ball-and-stick model based probabilistic tracking" Probtrackx (Probt) [6], the Hough voting method [1] and the probabilistic index of connectivity (PICo) method [14]. Each brain network describes detected connections between 113 cortical and subcortical regions-of-interest (ROIs), which are defined by using the Harvard Oxford Cortical and Subcortical Probabilistic Atlas [8]. Therefore we can use a vector of dimension 6328 ($113 \times 112/2$) to represent all connections of distinct ROIs pairs in each network. Please see [21] for details of computing these nine brain networks.

Experiment Settings. In the first experiment we compared the predictive performance of individual networks, in terms of area under the ROC curve (AUC), sensitivity and specificity. These are standard metrics measuring algorithm performance in classification problems. We also provide two intuitive fusion methods for baseline comparisons. The first method concatenates vectors from all networks (B-CON), resulting in a feature vector of dimension 56952. The second method combines the networks by averaging of all of the individual networks; this can be considered as a special case of the general linear combination ($\tau_i = 1/9, \forall i$). For all the patients, we used 10-fold cross validation, i.e., each time we use the brain networks from 90% patients to train a classifier, and the 10% to test the classifier and compute performance metrics. For all individual brain networks as well as the two baseline methods, we use sparse logistic regression to train classifiers. For the proposed DFUSE, the classifier is trained using algorithms in Section 2. As the sample size is too small to generate extra validation data for model selection (the selection of hyper parameter λ in the sparse logistic regression), we report the best performance for all methods.

Results and Discussion. Averaged classification results over 10 iterations are given in **Table 1**. Our proposed DFUSE algorithm significantly outperformed all other competing methods (p-value < 0.001). DFUSE has an average AUC of 0.89, compared to 0.77 achieved by the best individual method, which used only the Probtrackx (Probt) networks. DFUSE also had the highest average sensitivity of 0.84 and specificity of 0.77, compared to the second highest sensitivity of 0.72 achieved by tensor-based FACT (T-FACT) and 0.69 by the Probtrackx networks. No individual brain network generation method had a predictive power that was even close to the one from the fused brain network. This significant improvement in predictive performance supports our hypothesis about the benefits of fusion for brain networks.

Two other baseline network combination methods

Table 1. Quantitative comparison of classifiers using different brain networks to predict the early MCI. We compare the performance of each individual brain networks from tractography, simple network combination, and our network fusion method (DFUSE). The average and variance of area under the ROC curve (AUC), sensitivity and specificity over 10 splittings are reported. The proposed DFUSE significantly outperforms all other methods on this problem (p-value < 0.001).

	AUC	Sensitivity	Specificity
DFUSE	$\mathbf{0.89 \pm 0.09}$	$\mathbf{0.84 \pm 0.16}$	$\mathbf{0.77 \pm 0.07}$
B-CON	0.58 ± 0.10	0.56 ± 0.21	0.50 ± 0.07
B-AVG	0.55 ± 0.15	0.58 ± 0.20	0.49 ± 0.08
T-FACT	0.59 ± 0.11	0.72 ± 0.25	0.44 ± 0.14
T-RK2	0.58 ± 0.11	0.56 ± 0.25	0.49 ± 0.10
T-SL	0.62 ± 0.14	0.48 ± 0.27	0.64 ± 0.26
T-TL	0.58 ± 0.14	0.60 ± 0.21	0.48 ± 0.07
O-FACT	0.62 ± 0.09	0.60 ± 0.19	0.51 ± 0.09
O-RK2	0.60 ± 0.13	0.60 ± 0.21	0.53 ± 0.07
PICo	0.58 ± 0.10	0.56 ± 0.21	0.50 ± 0.07
Hough	0.66 ± 0.11	0.64 ± 0.23	0.54 ± 0.11
Probt	0.77 ± 0.08	0.70 ± 0.22	0.69 ± 0.08

Table 2. Combination coefficients $\boldsymbol{\tau}$ of 9 networks.

network	τ	network	τ	network	τ
T-FACT	0.025	T-Rk2	0.014	T-SL	0.023
PICo	0.058	Hough	0.010	Probt	0.871
T-TL	0	O-FACT	0	O-RK2	0

also did not perform well: the predictive performance of the feature concatenation (B-CON) does not even perform as well as the best individual brain network. This may be because, for the B-CON method, there are too many features presented to the classifier (over 56k), relative to the number of subjects (samples) available to train it. Only $\sim$110 samples are available here to train the classifier at every iteration (90% of the total of 124 subjects). On the other hand, the AUC of the simple average brain network (B-AVG) is 0.55, which is even poorer than the worst performing brain network T-TL, at 0.58. Arbitrary combinations of brain networks may not help for the task of distinguishing early MCI from NCs. Task specific fusion as proposed in this paper may be more beneficial.

One attractive property of the proposed DFUSE approach is that we can obtain an interpretable combination coefficient $\boldsymbol{\tau}$, indicating how much each of the individual brain networks contributes to the final combined network. The average combination coefficients for all networks are given in Table 2. We see that in the combination, Probtrackx has the heaviest weight of 0.871 (all elements of $\boldsymbol{\tau}$ range from 0 to 1), averaged over 10 iterations. This is consistent with the finding that Probtrackx is also the best predictive individual network as shown in **Table 1**. On the other hand, the weights of T-TL, O-FACT, O-RK2 are consistently zeros, i.e., they do not contribute to the combined network. As such, the combination offers a guide to which tractography methods to run (clearly not all methods need to be run for problems where they are given zero weight). Moreover, the networks with zero weights are not the same as the least white individual networks (T-RK2, PICo, T-FACT). The inconsistency shows that networks with weak predictive power may still have valuable connection information to complement other better performed networks. It is possible to leverage clustering analysis [22] and explore different sub-modalities within the networks, and we will leave this interesting analysis in our future work.

Because of the sparsity introduced on the model $\mathbf{w}$, we are also able to inspect what are the important connections contributing to the final classifiers. By averaging the non-zero weights for each connection from different experiments, we can generate a ranked list of connections, many of which are previously known to be relevant to the progression of Alzheimer's. Here are a few connections that appear in the top of the list: *Right Temporal Pole* $\Leftrightarrow$ *Right Precentral Gyrus*, *Left Pallidum* $\Leftrightarrow$ *Left Caudate*, *Left Lingual Gyrus* $\Leftrightarrow$ *Left Thalamus*, *Left Cingulate Gyrus Anterior Division* $\Leftrightarrow$ *Left Frontal Medial Cortex*, *Right Planum Polare* $\Leftrightarrow$ *Right Hippocampus*.

4. CONCLUSIONS

In this paper, we developed a new method for discriminative fusion of multiple brain networks to detect early mild cognitive impairment (MCI). We simultaneously learned a convex combination of different brain networks to best detect early MCI, and a classifier that works with the combined brain network. As the networks are fused in a way that maximizes the discriminative power between normal controls and early MCI subjects, the results from the fused network significantly improve on single brain networks as well as simple fusion methods. The much better predictive performance in terms of detecting early MCI inspires us to explore network fusion for other prediction tasks such as progression from early to late MCI, or detecting of genetic effects on brain networks. We will also further develop more powerful non-linear fusion methods in our future work. Furthermore, although the technique we proposed here is demonstrated using brain networks, it can be applied to any type of network.

Acknowledgment. Funded in part by NIH ENIGMA Center grant U54 EB020403, supported by the Big Data to Knowledge (BD2K) Centers of Excellence program, and ONR grant N00014-14-1-0631.

5. REFERENCES

[1] I. Aganj, C. Lenglet, N. Jahanshad, E. Yacoub, N. Harel, P. M. Thompson, and G. Sapiro. A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography. *Medical Image Analysis*, 15(4):414–425, 2011.

[2] O. Ajilore, N. Vizueta, P. Walshaw, L. Zhan, A. Leow, and L. L. Altshuler. Connectome signatures of neurocognitive abnormalities in euthymic bipolar I disorder. *Journal of Psychiatric Research*, 68:37–44, 2015.

[3] D. Arienzo, A. Leow, J. A. Brown, L. Zhan, J. GadElkarim, S. Hovav, and J. D. Feusner. Abnormal brain network organization in body dysmorphic disorder. *Neuropsychopharmacology*, 38(6):1130–1139, 2013.

[4] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44(4):625–632, 2000.

[5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[6] T. Behrens, H. J. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155, 2007.

[7] T. E. Conturo, N. F. Lori, T. S. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences*, 96(18):10422–10427, 1999.

[8] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.

[9] M. Lazar, D. M. Weinstein, J. S. Tsuruda, K. M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. Haughton, A. Field, et al. White matter tractography using diffusion tensor deflection. *Human Brain Mapping*, 18(4):306–321, 2003.

[10] A. Leow, O. Ajilore, L. Zhan, D. Arienzo, J. GadElkarim, A. Zhang, T. Moody, J. Van Horn, J. Feusner, A. Kumar, et al. Impaired inter-hemispheric integration in bipolar disorder revealed with brain network analyses. *Biological Psychiatry*, 73(2):183–193, 2013.

[11] A. D. Leow, S. Zhu, L. Zhan, K. McMahon, G. I. de Zubicaray, M. Meredith, M. Wright, A. Toga, and P. Thompson. The tensor distribution function. *Magnetic Resonance in Medicine*, 61(1):205–214, 2009.

[12] J. Liu, J. Chen, and J. Ye. Large-scale sparse logistic regression. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 547–556. ACM, 2009.

[13] S. Mori, B. J. Crain, V. Chacko, and P. Van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269, 1999.

[14] G. J. Parker, H. A. Haroon, and C. A. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of mri diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18(2):242–254, 2003.

[15] O. Sporns. *Networks of the Brain*. MIT press, 2011.

[16] O. Sporns, G. Tononi, and R. Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005.

[17] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

[18] S. J. Wright, R. D. Nowak, and M. A. Figueiredo. Sparse reconstruction by separable approximation. *Signal Processing, IEEE Transactions on*, 57(7):2479–2493, 2009.

[19] L. Zhan, N. Jahanshad, Y. Jin, A. W. Toga, K. L. McMahon, G. de Zubicaray, N. G. Martin, M. J. Wright, P. M. Thompson, et al. Brain network efficiency and topology depend on the fiber tracking method: 11 tractography algorithms compared in 536 subjects. In *10th International Symposium on Biomedical Imaging (ISBI)*, pages 1134–1137. IEEE, 2013.

[20] L. Zhan, Y. Liu, Y. Wang, J. Zhou, N. Jahanshad, J. Ye, and P. M. Thompson. Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition. *Frontiers in Neuroscience*, 9(257), 2015.

[21] L. Zhan, J. Zhou, Y. Wang, Y. Jin, N. Jahanshad, G. Prasad, T. M. Nir, C. D. Leonardo, J. Ye, P. M. Thompson, et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. *Frontiers in Aging Neuroscience*, 7, 2015.

[22] J. Zhou, J. Chen, and J. Ye. Clustered multi-task learning via alternating structure optimization. In *Advances in neural information processing systems*, pages 702–710, 2011.

[23] J. Zhou, J. Liu, V. A. Narayan, J. Ye, A. D. N. Initiative, et al. Modeling disease progression via multi-task learning. *NeuroImage*, 78:233–248, 2013.

[24] J. Zhou, Z. Lu, J. Sun, L. Yuan, F. Wang, and J. Ye. Feafiner: biomarker identification from medical data through feature generalization and selection. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1034–1042. ACM, 2013.

[25] J. Zhou, L. Yuan, J. Liu, and J. Ye. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 814–822. ACM, 2011.