Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning

Shinjini Kundu[a,b,1,2], Beth G. Ashinsky[c,3], Mustapha Bouhrara[c], Erik B. Dam[d], Shadpour Demehri[e], Mohammad Shifat-E-Rabbi[f], Richard G. Spencer[c], Kenneth L. Urish[a,g,h,i,j], and Gustavo K. Rohde[f,k]

[a]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213; [b]Medical Scientist Training Program, University of Pittsburgh, Pittsburgh, PA 15261; [c]Laboratory of Clinical Investigation, Magnetic Resonance Imaging and Spectroscopy Section, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224; [d]Department of Computer Science, University of Copenhagen, 2100 Copenhagen, Denmark; [e]Department of Radiology, The Johns Hopkins Hospital, Baltimore, MD 21287; [f]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908; [g]Arthritis and Arthroplasty Design Group, The Bone and Joint Center, Magee Womens Hospital of the University of Pittsburgh Medical Center, Pittsburgh, PA 15213; [h]Department of Orthopedic Surgery, University of Pittsburgh, Pittsburgh, PA 15213; [i]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA 15261; [j]Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA 15261; and [k]Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22908

Edited by Dennis A. Carson, University of California San Diego School of Medicine, La Jolla, CA, and approved July 20, 2020 (received for review October 12, 2019)

Many diseases have no visual cues in the early stages, eluding image-based detection. Today, osteoarthritis (OA) is detected after bone damage has occurred, at an irreversible stage of the disease. Currently no reliable method exists for OA detection at a reversible stage. We present an approach that enables sensitive OA detection in presymptomatic individuals. Our approach combines optimal mass transport theory with statistical pattern recognition. Eighty-six healthy individuals were selected from the Osteoarthritis Initiative, with no symptoms or visual signs of disease on imaging. On 3-y follow-up, a subset of these individuals had progressed to symptomatic OA. We trained a classifier to differentiate progressors and nonprogressors on baseline cartilage texture maps, which achieved a robust test accuracy of 78% in detecting future symptomatic OA progression 3 y prior to symptoms. This work demonstrates that OA detection may be possible at a potentially reversible stage. A key contribution of our work is direct visualization of the cartilage phenotype defining predictive ability as our technique is generative. We observe early biochemical patterns of fissuring in cartilage that define future onset of OA. In the future, coupling presymptomatic OA detection with emergent clinical therapies could modify the outcome of a disease that costs the United States healthcare system $16.5 billion annually. Furthermore, our technique is broadly applicable to earlier image-based detection of many diseases currently diagnosed at advanced stages today.

early diagnosis | classification | 3D transport-based morphometry | osteoarthritis | T2 imaging

Early diagnosis is a common goal in many diseases where timely intervention has the potential to modify disease outcome. For example, early detection of skin cancer (1) reduces mortality and timely diagnosis of retinopathy (2) can prevent blindness. Yet, in the early stages of a disease, observable cues are often not visually perceptible. Therefore, many diseases today are still diagnosed at advanced stages, after irreversible damage has occurred. The discovery and visualization of early markers of a disease before the emergence of symptoms is an area of immense interest.

Presymptomatic detection is a goal in osteoarthritis (OA), which affects as many as one in four adults (3). However, OA is currently not detectable until pain develops and an X-ray image confirms bone damage (4). Unfortunately, at this late stage, the disease is not reversible and joint destruction is an inevitable outcome (5). The mainstay of current treatments focuses on either palliation or invasive surgery. Despite current interventions, OA costs the US healthcare system $16.5 billion/y (6). Yet, there is recent evidence that preosteoarthritis may be a reversible process (7). In particular, early biochemical changes that occur in cartilage (8–10) precede classic pain and bone damage symptoms by several years (11, 12). In vivo T2-weighted magnetic resonance imaging (MRI) is a promising modality to investigate these changes. Specifically, the transverse relaxation time is highly sensitive to perturbations in collagen fiber isotropy and water content (13). However, the current gold standard relies on evidence of bone damage because both visual inspection and conventional statistics are insensitive to early signs of disease in the cartilage. Early biochemical shifts, superficial grooves, and cartilage fibrillation are often too subtle and complex to be readily perceptible (7, 13–16). Fig. 1 illustrates T2 texture maps of knee articular cartilage in healthy individuals. In 3 y time, one of these groups is known to progress to symptomatic knee OA with

Significance

Many diseases lack visual cues in early stages, eluding image-based detection. Today, osteoarthritis is diagnosed at an irreversible stage after bone damage has occurred. This research demonstrates that osteoarthritis detection in healthy individuals may be possible 3 y prior to symptoms or bone damage. We describe a technique combining mass transport theory with statistical pattern recognition, discovering sensitive cartilage phenotypes that predict future osteoarthritis progression with high accuracy. As our approach is based on mass transport principles, the cartilage phenotype defining predictive ability can be directly visualized, a key contribution of this work. This research demonstrates that transport-based learning might enable early osteoarthritis detection and visualization at a potentially reversible stage.

Author contributions: S.K., B.G.A., M.B., E.B.D., R.G.S., K.L.U., and G.K.R. designed research; S.K., B.G.A., M.B., S.D., and M.S.-E.-R. performed research; S.K., B.G.A., M.B., E.B.D., R.G.S., and G.K.R. contributed new reagents/analytic tools; S.K., S.D., and M.S.-E.-R. analyzed data; and S.K., B.G.A., M.B., E.B.D., M.S.-E.-R., R.G.S., K.L.U., and G.K.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: Department of Radiology, The Johns Hopkins Hospital, Baltimore, MD 21287.

[2]To whom correspondence may be addressed. Email: skundu2@jhmi.edu.

[3]Present address: Drexel University College of Medicine, Philadelphia, PA 19129.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1917405117/-/DCSupplemental.

First published September 21, 2020.

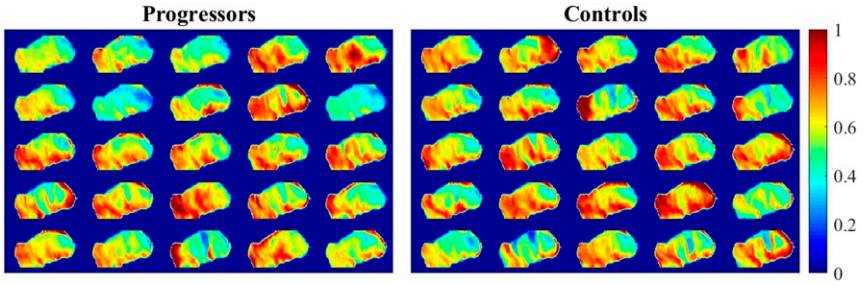

Fig. 1. Medial condyle images corresponding to individual subjects. Shown are T2 images of cartilage belonging to two groups of asymptomatic subjects. In 3 y, only one of these groups will go on to develop clinical OA, although it is currently not visually discernible. The latter motivates advanced pattern learning approaches to discover and visualize the abnormal variation leading to discrimination between the two groups. Scale bars correspond to the normalized T2 intensity for each cartilage image.

pain and bone damage. However, at the current timepoint, these groups are indistinguishable to a visual observer.

In fact, Fig. 1 illustrates a present challenge in early osteoarthritis detection today (17). Currently, a reliable method for early OA detection based on cartilage images does not exist (17), despite recent studies utilizing machine learning for image-based detection (11, 12, 18). However, the ability to detect and visualize early changes at a potentially reversible stage could enable additional investigational disease-modifying therapies (i.e., mesenchymal stem cells, platelet-rich plasma, etc.) (19–21). Furthermore, a visualizable early trace of biochemical injury could further motivate research into investigational modalities such as T1ρ to monitor the status of specific cartilage macromolecules (22, 23).

As Fig. 1 illustrates, visual cues of early OA are lacking. Prior data-driven approaches have trained machine-learning classifiers on ad hoc numerical texture features (e.g., Tamura texture statistics, histogram statistics, Zernicke polynomials, etc.) (11, 12, 18, 24, 25) to improve upon the prognostic capabilities of conventional T2 imaging (11, 12, 26). However, these approaches did not provide sufficient information to make robust prediction or prognostication (17). Furthermore, there has been a long-standing gap in biological interpretation, as there is no direct biological interpretation of Zernicke polynomials or wavelet coefficients into understandable biochemical tissue properties. If early cartilage signatures could inform risk of future symptomatic OA progression, these early signatures could motivate additional targets for disease-modifying osteoarthritis drugs (DMOADs) (17).

The goal of this research is to develop a technique to discover and visualize the early trace of biochemical cartilage injury that may enable OA detection at a potentially reversible stage. We develop an approach combining optimal mass transport theory and statistical pattern recognition to directly quantify the mass transport required in cartilage tissue to transform one morphology into another. We call our approach three-dimensional transport-based morphometry (3D TBM), and its theoretical foundations are described in prior studies (27–31). This paper focuses on extensions and modifications to this technique to enable the state of the art in early OA detection based on cartilage maps. In contrast to traditional supervised-learning approaches, our approach does not require a priori features. Furthermore, while traditional feature-based learning approaches are based on algorithms lacking provable guarantees, our approach is based on optimal mass transport (OT) theory and is generative. The latter enables direct visualization of the learned discriminant through inverse transformation (28).

This paper describes an application of 3D TBM for medical image classification. We apply our framework to develop and evaluate a blind predictive test for future symptomatic OA based on 86 healthy individuals. Comparing our classifier's predictions with gold standard diagnosis established at 3 y follow-up, we see that our approach may potentially enable preclinical OA detection at a reversible stage and outperforms traditional supervised-learning approaches at this task. Furthermore, our approach enables the discriminating phenotypes defining future OA progression to be directly visualized.

The following are contributions of this paper: a generative pattern learning technique for cartilage phenotype analysis that has foundations in optimal mass transport theory, 3D transport-based morphometry; a predictive approach that could potentially detect future symptomatic osteoarthritis up to 3 y in advance of symptoms at a reversible stage of the disease; evaluation of our predictive approach through precision-recall characteristics using balanced cross-validation and null hypothesis testing, as well as comparison to existing supervised-learning approaches; visualization of the cartilage texture phenotypes sensitive for future OA progression through inverse transport mapping, enabled by inverse 3D transport-based morphometry; and additional understanding of the water distribution patterns in cartilage underpinning presymptomatic detection.

This work highlights the potential of our approach to expand the diagnostic capabilities of conventional MRI in preclinical OA detection. The remainder of this paper is organized as follows. *Methods: 3D Transport-Based Morphometry* contains a brief overview of the mathematical foundations of our approach as well as experimental methodology. We summarize our results in *Results*. Discussion of key results is presented in *Discussion*. Finally, *Conclusion* explores the implications of this work and future directions.

Results

Subject Demographics. Between progressor and control cohorts, there were no statistically significant differences in age (P = 0.11), but differences in body mass index (BMI) were statistically significant ($P < 0.001$). There was no significant difference in the level of baseline physical activity between the two cohorts, as measured by the Physical Activity Scale for the Elderly (PASE) (32). Total Western Ontario and McMaster Universities Arthritis Index (WOMAC) (33) was significantly different while Kellgren–Lawrence (KL) scores (34) were not significantly different between progressor and nonprogressor cohorts at baseline. Table 1 summarizes subject demographics.

Age, BMI, and gender predicted OA progression with accuracy above random chance (sensitivity = 53.5%, specificity = 65.1%). However, the Cohen's kappa was 0.19. Thus, demographic metrics were not sufficient to enable discrimination with high sensitivity and specificity (Table 2).

Characterizing Principal Phenotypes. The variability in the dataset was represented by fewer principal components in the transport domain compared to the image domain. Fig. 2 shows the variance plot for cartilage texture. Thus, the underlying structure of the

Table 1. Subject demographics

Parameter	Progressors (n = 43)	Control (n = 43)	P
Age, y	60.0 ± 9.1	56.8 ± 9.1	0.11
BMI, kg/m^2	29.1 ± 3.8	25.8 ± 3.7	<0.001
Mean PASE	169.8 ± 95.8	189.8 ± 65.9	0.26
Mean initial WOMAC	3.6 ± 3.0	1.2 ± 2.2	<0.001
Mean initial KL	0.4 ± 0.5	0.3 ± 0.5	0.4
Mean 3-y ΔWOMAC	23.9 ± 10.0	−0.4 ± 2.0	<0.001
Mean 3-y ΔKL	0.4 ± 0.8	0.1 ± 0.6	0.2

PASE score from the OAI was missing for one control subject; KL score at 3 y follow-up from the OAI was missing for two control subjects and one progressor subject.

data was better captured in the transport domain than the image domain.

Classification. Although future OA progression was not readily differentiable from Fig. 1 through inspection, the classes were separable in the transport space. Fig. 3*A* shows the histogram indicating cohort separation based on cartilage texture maps when the test data were projected onto the learned discriminant. The training accuracy was 92.9%.

We evaluated precision-recall characteristics using receiver-operator characteristic (ROC) curves as the discrimination threshold was varied (Fig. 3*B*), with an area under the curve (AUC) 0.87.

In the testing phase, we evaluated the blind predictive accuracy in identifying future osteoarthritis progression from baseline cartilage texture maps using 3D TBM. Table 2 summarizes the overall accuracy, sensitivity, and specificity on the test data when iterating over all combinatorial partitions of leave-two-out balanced cross-validation. The penalized linear discriminant analysis (pLDA) classification scheme and complete leave-two-out cross-validation approach are described in *Supervised Learning* and *Balanced Cross-Validation*.

Furthermore, based on the results in Table 2, we rejected the null hypothesis; the probability that the predictive test developed in this paper had learned a model between baseline cartilage texture maps and future OA progression through chance alone was less than 1 in 1,000.

Comparison to Existing Supervised-Learning Approaches. Our predictive test utilizing a penalized linear discriminant classifier under the 3D TBM approach enables the highest predictive accuracy, as Table 3 indicates.

Collectively, supervised learning in the transport domain enabled the highest predictive accuracies, followed by the image domain and finally weighted neighbor distance using compound hierarchy of algorithms representing morphology (WND-CHRM) features (35) extracted from the original images. On the transport domain embeddings, nonlinear classifiers using support vector machine with Gaussian kernel and random forest classifiers did not lead to superior predictive ability over the linear discriminant classifier. Sensitive discrimination in the transport space is possible using Euclidean distances because 3D TBM computes a linearized version of the OT distance, described as a generalized geodesic (30). Therefore, complex, nonlinear, and spatially diffuse shifts in the image domain can be discriminated using simpler classifiers in the transport domain (28, 30, 36).

Visualizing the Discriminating Cartilage Phenotype. Our approach, 3D transport-based morphometry, is generative. The latter enables direct visualization of the discriminating shifts in cartilage phenotype learned by a classification model. The images in Figs. 4–6 were synthesized from the images in the study population by 3D TBM. By sampling along the discriminant direction in the transport space and inverting the classifier Eq. **2**, we visualized the characteristic shifts in T2 texture defining TBM classification. The cartilage texture maps were colorized for ease of visual interpretation.

We observed that the early T2 texture shifts sensitive for future symptomatic OA were largely localized to the center of the knee. In the medial compartment, the lateral aspect of the femoral cartilage demonstrated intensity increase, while the tibial cartilage demonstrated intensity diffusion (Fig. 4). In the lateral compartment, there were changes observed in the medial aspect of both the femoral and tibial cartilage as well. The pattern of brightness in intensity appeared to become more diffuse for progressor images compared to control images in the femoral cartilage and more focal in the tibial cartilage (Fig. 5). Finally, we observed focal intensity increase in the patellar cartilage of progressors compared to controls (Fig. 6).

Discussion

Many diseases have no visual cues in the early stages, eluding image-based detection. This paper focused on osteoarthritis, for which there exists no reliable detection method in the presymptomatic phase (17). Yet, with emergent DMOADs, early OA detection is a challenge of great interest (17). This research demonstrated an approach enabling osteoarthritis detection in presymptomatic healthy individuals at a potentially reversible stage. By developing a technique combining optimal mass transport theory and statistical pattern recognition, we identified risk of progression 3 y before the gold standard diagnosis using the cartilage texture phenotype. In addition to differentiating progressors and nonprogressors in an automated manner, a key contribution of our technique, 3D transport-based morphometry, is that it is generative. Inverse TBM transformation visualized the early biochemical shifts in cartilage texture that could underlie risk of progression to symptomatic OA. The latter is a contribution enabled by the fact that 3D TBM is a generative technique. Direct visualization enables understanding of the mechanisms of early OA in the preclinical phase. As recent studies indicate that OA in the stages preceding bone damage may potentially be reversible (7), this work adds to the growing body of literature suggesting that disease-modifying therapies for OA may become feasible in the future. Furthermore, we demonstrated the potential of 3D transport-based morphometry as a diagnostic technique capable of earlier detection, which may be applicable to other diseases.

Previous work using machine learning investigated whether numerical texture features (i.e., histogram statistics, entropy, Tamura features, etc.) (11, 12, 26) were sensitive for predicting future OA risk. However, approaches to date have not provided sufficient information to make robust prediction or prognostication (17). Using a priori features limits the search for

Table 2. Complete leave-two-out cross-validation results using demographic information and TBM

Region	Test accuracy, %	Sensitivity, %	Specificity, %	Cohen's kappa	P
Cartilage TBM	78.0	76.9	79.0	0.56	<0.001
Age, BMI, and gender	59.3	53.5	65.1	0.19	0.03

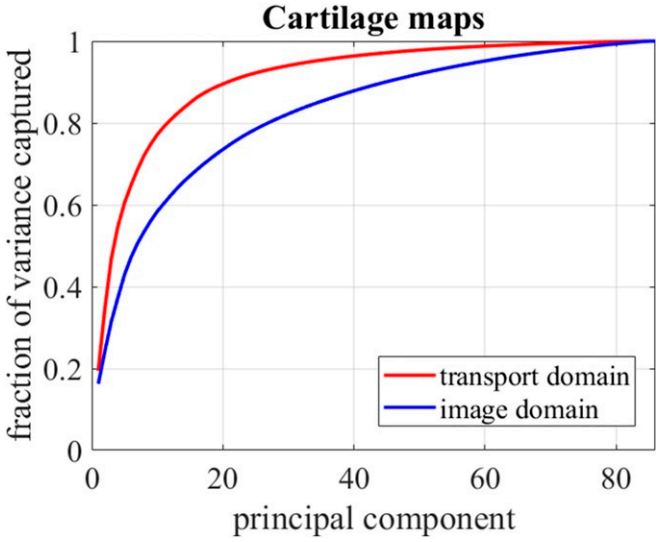

Fig. 2. The variance plot demonstrates that fewer components were needed to represent tricompartment knee cartilage images in the 3D TBM space than in the original image domain.

possible discriminating information. Furthermore, comparing images through sets of numerical features does not lead to an invertible model. Finally, there is no direct physical interpretation of a predictive model constructed in this manner (11, 28, 30). These algorithmic approaches also lack provable guarantees. In contrast to these methods, 3D TBM discovers the discriminating information in a fully automated manner without the need for a priori features. Three-dimensional TBM has a rich theoretical foundation with equations for analysis and synthesis (28), which we extend and modify for medical image classification. This study bridges the gap between classification ability and direct biochemical perturbations within the cartilage tissue defining future OA progression. A key contribution of our approach is the ability to directly visualize the texture shifts enabling sensitive classification for visual understanding of a process that is otherwise not well assessed visually or through conventional statistics.

Three-dimensional TBM is advantageous as a data-mining technique. First, it is a fully automated approach. It facilitates information extraction by representing high-dimensional, nonlinear perturbations in the image domain through more sparse distances in the transport domain (27–31). This research demonstrated that a linear discriminant in the transport domain differentiated progressors and nonprogressors to a clinically useful degree. Linear classifiers often have a favorable bias–variance tradeoff, especially in the classification of high-dimensional data like medical images (37). Furthermore, nonlinear classifiers may offer high classification accuracy but typically come at the tradeoff of variance and overfitting when the number of samples is small compared to the complexity of the model. Furthermore, as discussed above, the latter techniques are typically not generative. Because our approach is able to map complex, nonlinear shifts in texture profile (geodesic distances) to Euclidean distances in the transport domain (30), we demonstrated that linear classification in the transport domain could outperform existing supervised classifiers such as kernel support vector machine (SVM) and random forest to enable the state of the art in early OA detection.

Visually, we observe that our TBM approach identified water distribution in cartilage as key in enabling symptomatic prediction of future OA. By assessing water distribution, we achieved 78% accuracy on a validation set, higher than many studies in the literature to date (11, 12, 28). However, beyond classification ability, the greatest advance of our approach compared to existing studies is the ability to directly visualize structural changes in cartilage in vivo that enable sensitive classification. The ability to visualize markers of disease and health could identify treatments and interventions.

Our cohort consisted of baseline healthy patients from the Osteoarthritis Initiative (OAI) database (38), with a progressor subcohort determined by change in total WOMAC score on 3 y follow-up. We observed that in presymptomatic individuals, early texture shifts were localized to the center of the knee. We observed a focal increase and decrease in water concentration in the medial femoral and medial tibial cartilage, respectively. In the lateral compartment, we observed decrease in water concentration in the femoral cartilage and increase in the tibial cartilage. We also observed focal intensity increase in the patellofemoral cartilage. A possible mechanistic explanation involves fissuring of fibrous tissue and diffusion of water into areas of weakened cartilage, which is high intensity on the T2 images. Prior work supports this proposed mechanism. Both cartilage water content and collagen fiber anisotropy increase in early OA (10). The patterns of early cartilage injury found in this study are consistent with patterns of tibiofemoral bone damage seen on arthroscopic and cadaveric specimens in prior studies. Wear in the middle third of both the lateral and medial femoral articular cartilage has been implicated on osteoarthritic cadaveric knees (39). Furthermore, our findings corroborate known findings that biochemical changes in the tibiofemoral articulation correspond between surfaces (39). Imaging offers a noninvasive assessment of subtle biochemical and structural alterations

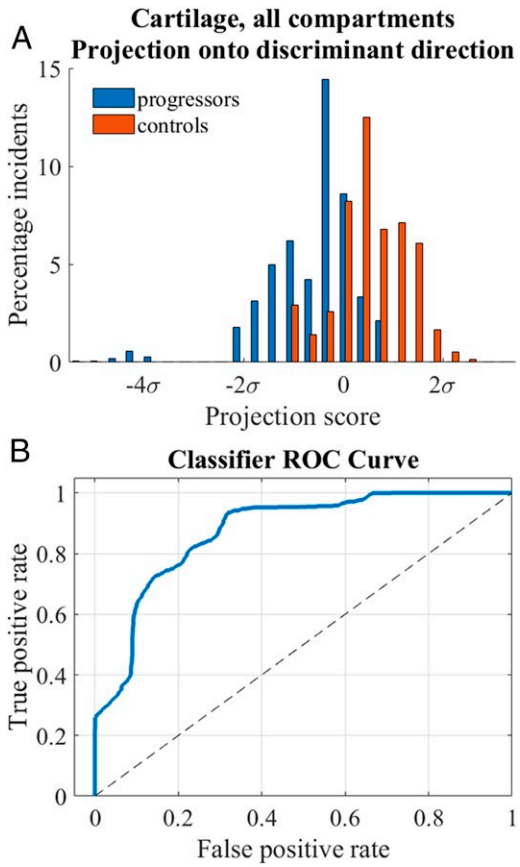

Fig. 3. Transport-based discrimination based on cartilage texture maps. (*A*) Histogram illustrating the projection of the test dataset onto the learned pLDA classifier boundary. (*B*) Corresponding receiver-operator characteristic curve (AUC = 0.87).

Table 3. Comparing classification results using complete leave-two-out cross-validation

Classifier for cartilage texture maps	Test accuracy, %	Sensitivity, %	Specificity, %	Cohen's kappa	*P*
Original image pLDA	64.5	61.5	67.4	0.29	< 0.001
Original image SVM (Gaussian kernel)	63.2	65.5	60.9	0.26	<0.001
Original image RF	54.9	39.0	70.7	0.10	<0.001
TBM pLDA	78.0	76.9	79.0	0.56	<0.001
TBM SVM (Gaussian kernel)	74.1	78.0	70.2	0.48	<0.001
TBM RF	70.2	65.0	75.3	0.40	<0.001
WNDCHRM features pLDA	60.8	60.6	61.0	0.22	0.02
WNDCHRM features SVM (Gaussian kernel)	53.3	62.7	43.9	0.07	0.19
WNDCHRM features RF	55.6	48.6	62.6	0.11	<0.001

Among the nine different classifiers, the 3D TBM features employing a linear discriminant classifier achieved the highest classification accuracy to date. SVM, support vector machine with radial basis function kernel.

caused by interactions of age, BMI, genetic factors, and prior knee injury (40). The T2 image in normal cartilage undergoes spatial intensity variation that corresponds to the changes in the distribution of water molecules, proteoglycans, and orientation of collagen fibers (8–10). The early perturbations in T2 spatial variation revealed by TBM enhance our understanding of early OA pathophysiology and may be used to prognosticate or monitor success of disease-modifying therapies in the future.

We ensured robustness of our automated predictive approach through several methods. First, by transforming the cartilage texture maps from the image domain to the transform domain, we represented the data more sparsely (28) (Fig. 2). As overfitting is a potential pitfall of high dimension, low sample size (HDLSS) data, a sparser representation of the data could mitigate incidental correlations of the cartilage dataset with the progression labels. Second, we utilized the principal component analysis technique to reduce the dimensionality of the feature vector using a basis of orthogonal components. Third, we eliminated noisy components of the data by including only the components contributing to the top 96% of the variance in the dataset. Fourth, we utilized a complete balanced cross-validation scheme iterating over the combinatorial space of all feasible leave-two-out partitions of the data. The advantage is that complete cross-validation yields the best generalization estimate of the classification accuracy (41). Fifth, we compared our TBM classification scheme to other classifiers, confirming that TBM outperformed existing supervised-learning approaches. Finally, we tested the null hypothesis by explicitly constructing an empirical estimate of the likelihood that the observed test accuracy in this paper could be generated from our proposed predictive test by random chance alone. The likelihood that our predictive model learned a relationship between

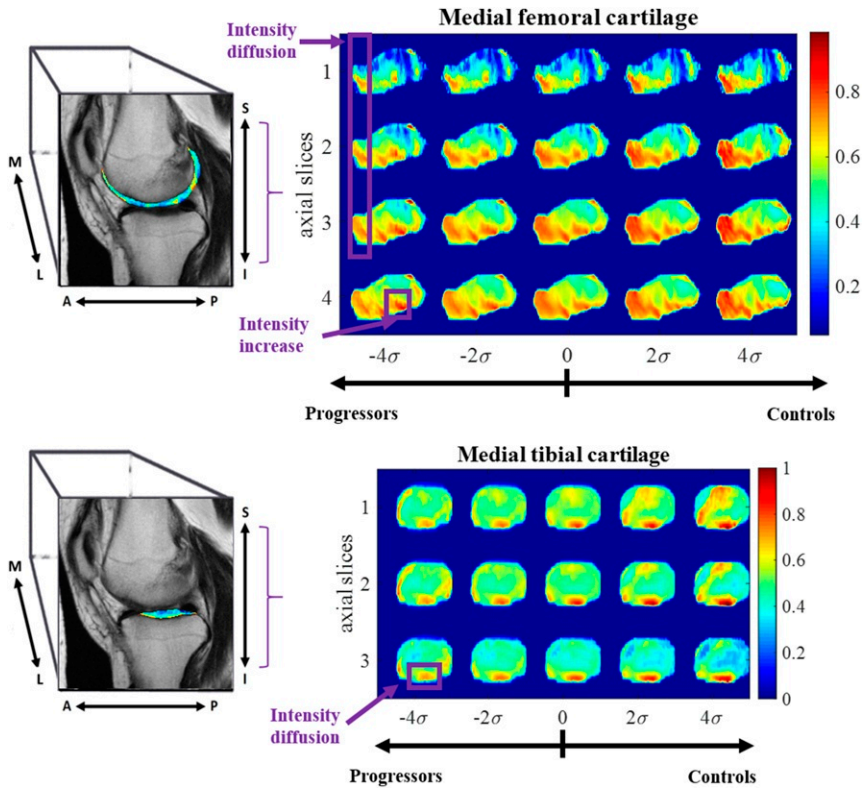

Fig. 4. Medial compartment images generated by inverse TBM transformation showing the cartilage texture changes enabling detection of future symptomatic OA progression. Here, σ represents standard deviations. These images are synthesized by TBM at the interface of the learned decision boundary spanning 4σ toward the control side to 4σ toward the progressor side. Scale bars correspond to the normalized T2 intensity for each generated 3D image.

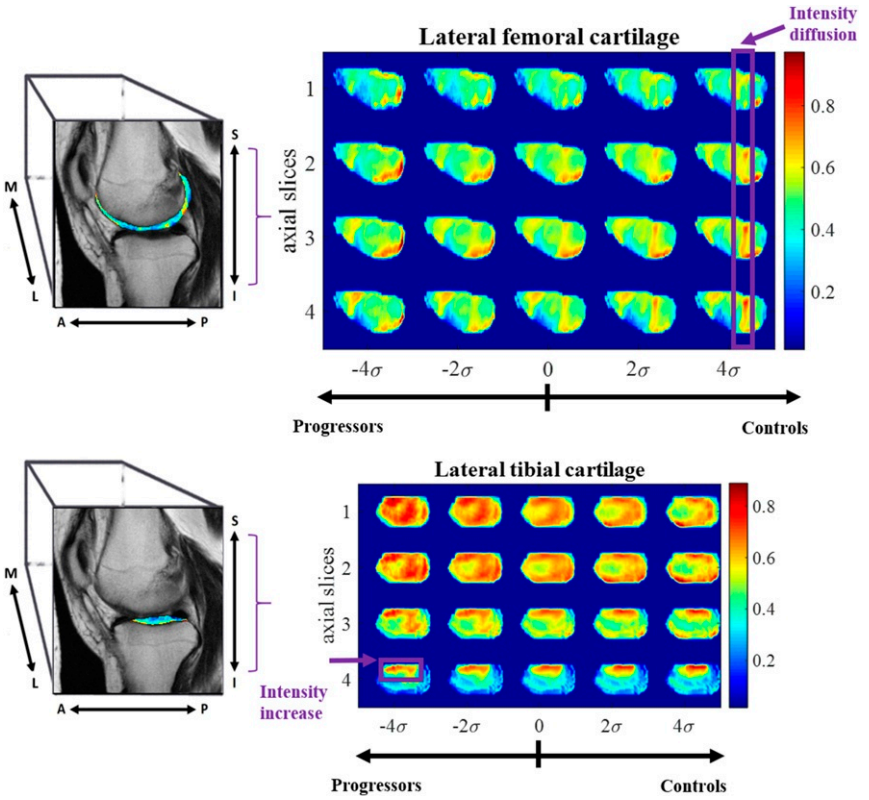

Fig. 5. Lateral compartment images generated by inverse TBM transformation showing the cartilage texture changes enabling detection of future symptomatic OA progression. Here, σ represents standard deviations. These images are synthesized by TBM at the interface of the learned decision boundary spanning 4σ toward the control side to 4σ toward the progressor side. Scale bars correspond to the normalized T2 intensity for each generated 3D image.

cartilage phenotype and future OA progression on the current cohort through chance alone was less than 1 in 1,000.

Next, we describe the limitations of this study. First, this is a retrospective study investigating whether T2 texture is a sensitive predictor of future OA. Prospective studies are needed to translate the findings in this study into a clinical tool. Second, this was an observational study on 86 subjects. With increasing sample size, machine-learning techniques become more robust to variations that may exist within a population. Third, there are several sources of variability in the data, including MRI inhomogeneity, segmentation, and registration procedure. We correct for many of these sources of variability in the data by eliminating the data dimensions containing little or no data with the principal components analysis (PCA) technique before a training a classifier. Finally, future studies are needed to understand the radiologic–pathologic correlation between biochemical imaging signatures on MRI and cartilage tissue specimen. Animal studies would be a first step toward this direction. The knee is one of the most common sites of osteoarthritis. Future research could also help elucidate how the present findings generalize to other joints in the body beyond the knee.

Currently, osteoarthritis is not easily detectable until pain symptoms occur and irreversible bone damage is demonstrated on plain radiography. However, even at this late stage of the disease, radiographic findings correlate poorly with pain symptoms and disease progression (4). In the future, our transport-based learning approach may enable more accurate image-based assessments of early disease development. Finally, the 3D transport-based morphometry technique is generally applicable to any imaging modality and has the potential to be implemented alongside other magnetic resonance (MR) contrast modalities such as $T_{1\rho}$ (23) and molecular imaging techniques.

Conclusion

Early diagnosis is a common goal in many diseases where timely intervention may have the potential to modify disease trajectory. Yet, many diseases remain undetectable until advanced stages. Our work describes an approach for automated pattern learning in magnetic resonance images. We demonstrated that our technique, 3D TBM, may enable automated discovery and visualization of discriminating patterns from articular cartilage maps, even when not easily assessed by visual interpretation. We demonstrated using TBM that future symptomatic osteoarthritis could potentially be predicted up to 3 y prior to the current gold standard diagnosis. By assessing the cartilage phenotype in healthy individuals using TBM, we designed a diagnostic approach that achieved clinically useful prognostic accuracy. In contrast to machine-learning techniques built on algorithms lacking explainability, our approach is based on the theoretical foundations of optimal mass transport which permits explainability and furthermore, direct physical interpretation of tissue shifts. TBM provided visualization of the cartilage phenotype underlying symptomatic OA prediction. As there is currently no reliable method to detect presymptomatic disease, OA detection at a potentially reversible stage of the disease could represent a step toward possible disease-modifying therapies. In the future, TBM may have broader application in earlier detection in conditions currently diagnosed at advanced stages today, such as cancers, retinopathy, and dementia.

Methods: 3D Transport-Based Morphometry

This section provides a brief overview of our technique, 3D TBM. Our approach is based on the theory of OT (28). Mathematical details, including the theoretical foundations of our approach,

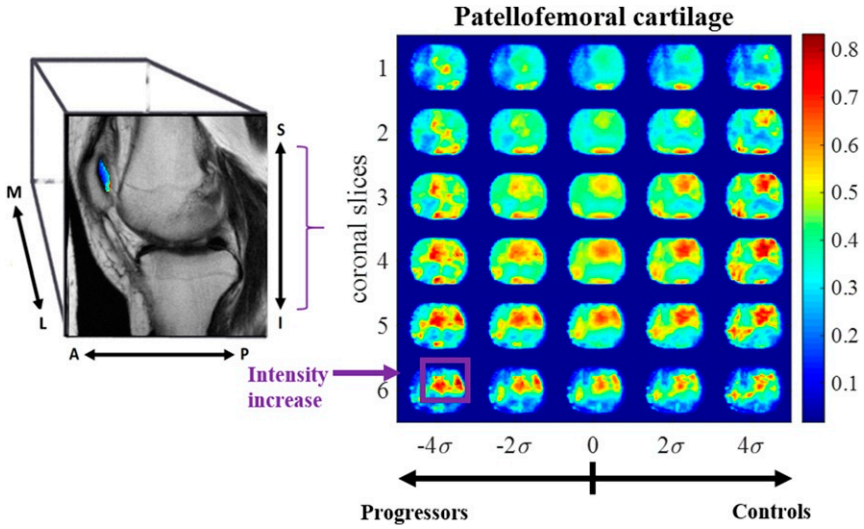

Fig. 6. Patellofemoral cartilage images generated by inverse TBM transformation showing the cartilage texture changes enabling detection of future symptomatic OA progression. Here, σ represents standard deviations. These images are synthesized by TBM at the interface of the learned decision boundary spanning 4σ toward the control side to 4σ toward the progressor side. Scale bars correspond to the normalized T2 intensity for each generated 3D image.

are further described in previous publications (27–30, 42). Here, we describe the modifications and extensions to 3D TBM to enable presymptomatic OA detection.

Overview. The basic principle of our approach is to quantify the degree of similarity, or distance, between cartilage texture maps by measuring the amount of effort (mathematically, mass transported over distance) to rearrange one image into the other. The latter leads to a mass-preserving mapping for each image. We select the mass-preserving mapping that is optimal in the sense of mass transport, which is a unique solution for a pair of images. As magnetic resonance images capture water content across the cartilage tissue, the 3D transport-based morphometry approach quantifies the mathematical shifts in water distribution from one image to another, generating a transport map for each image. Pattern recognition in the transport domain has been theoretically proved to render many classes of problems much more separable than in the image domain (43). Furthermore, our approach is fully automated and obviates the need for a priori features; OT defines a distance metric between images that considers intensity variations in the context of their locations (Lagrangian viewpoint) (42). Furthermore, a key contribution of the 3D TBM technique is that it is generative, which enables the learned classification model in the transport domain to be interrogated to visualize the biochemical shifts in cartilage tissue responsible for sensitive classification through inverse transformation. Therefore, 3D transport-based morphometry unifies tasks of discrimination and visualization in a single framework. It is an invertible Lagrangian transformation that produces a generative model for shape and texture variations in cartilage.

Analysis Equation. Total mass is first normalized across all images. Given a set of preprocessed MR images, $I_1(x),\ldots,I_N(x)$ corresponding to experimental subjects $1,\ldots,N$ and where x is a coordinate in 3D space, we can compute a unique transformation (28) for the image $I_i(x)$ according to the analysis equation

$$f_i^*(x) = argmin_{f_i \in MP} \int_\Omega |x - f_i(x)|^2 I_0(x)\,dx$$
$$\text{such that } det(Df_i(x)) I_i(f_i(x)) = I_0(x) \forall \in \Omega, \quad [1]$$

where $I_0(x)$ is the reference image. Here *MP* refers to the space of mass-preserving mappings. The operator D represents the Jacobian operator.

Eq. **1** computes the optimal mass-preserving mapping based on the $L2$-Wasserstein distance (29, 30) between each source $I_1(x)$ and template image $I_0(x)$. OT computes the most parsimonious way to rearrange the pixels in one image to match another (i.e., mass over distance moved). According to Brénier's theorem (44), under smoothness and domain conditions that our MR dataset satisfies, the optimal mass-preserving mapping is unique. Thus, the transformation is bijective.

The transport map $f_i^*(x)$ generated for each subject i quantifies the unique T2 spatial variation in each image $I_i(x)$ compared to the common reference image $I_0(x)$ (27–30). Under a mass-preserving mapping, the resultant transport map captures shape as well as texture variations between two images (28). The numerical solution of Eq. **1** was computed through the multiscale accelerated gradient descent optimization approach developed in ref. 28 to guide the solution toward the global optimum.

OT-Based Recognition. We emphasize that in contrast to previous data-driven approaches based on feature extraction (11, 12, 26), our approach is generative. Specifically, the metric space defined by the $L2$-Wasserstein distance has the geometric structure of a Riemannian manifold. Pairwise distances between images in the transform domain can be represented as the length of the shortest curve, or geodesic, connecting the images on the manifold (30). Computing pairwise distances between each image and the template image facilitates the computationally expensive task of calculating the space of all pairwise distances for classification. Our approach reduces the number of distance computations from $N(N-1)/2$ to N computations (30, 42) for all pairwise distances. The geodesic distances are projected onto the tangent space centered on the fixed template. Thus, we generate linear embeddings of the OT metric (29, 30), which approximates the geodesic.

In the transport domain, the model for shape and texture variations is generative (27–30, 43). The reason that linear analysis (e.g., linear discriminant analysis, principal component analysis) in the transport domain is highly sensitive for informative patterns is that discrimination based on Euclidean distances employs the linearized version of the geodesic. Therefore, linear

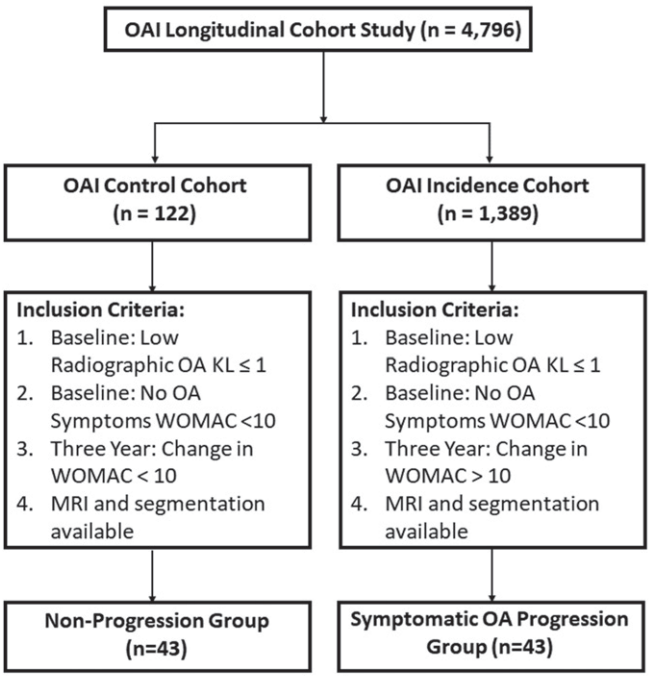

Fig. 7. Consort diagram of inclusion criteria.

analysis in the transport space corresponds to pattern recognition based on the OT metric in the image domain (29). In fact, modeling based on the 2-Wasserstein distance is highly sensitive to complex, high-dimensional, nonlinear image patterns in the dataset and can represent them more sparsely (28, 43).

Synthesis Equation. As our transport space is generative (*OT-Based Recognition*), inverse 3D TBM transformation can generate images defining the interface between tissue classes. We emphasize that direct visualization is not limited to the original images in the dataset. Rather, the underlying transport space can be interrogated at any arbitrary point to generate a synthetic image through inverse TBM transformation.

Particularly, we can choose to sample our transport space along the linear discriminant direction at the interface between classes. A sampled embedding from the transport domain $f(x)$ can be inverted to directly visualize the unique image to which it corresponds in the image domain $I(x)$, as the transform is bijective (28). The synthesis Eq. **2** defines the inverse TBM transformation:

$$I(x) = det(Df^{-1}(x))I_0(f^{-1}(x)) \quad \text{where } f^{-1}(x) \text{is the inverse mapping of } f(x). \qquad [2]$$

Therefore, inverse TBM transformation closes the gap between statistics defined by mass transport $f(x)$ and physical shifts in water distribution at the level of the tissue in $I(x)$.

Experimental Methodology

Subject Cohort. Subjects were selected from the Osteoarthritis Initiative public database (45). The OAI is sponsored by the National Institutes of Health and is a multicenter (38), longitudinal, prospective study of knee osteoarthritis (38, 45). Eight-six subjects (43 from the unexposed control subcohort and 43 from the incidence subcohort) were included in the study.

Nonprogressors were subjects for whom total WOMAC (33) score was <10 and KL score (34) was ≤1 at baseline, with 36-mo change in WOMAC score <10 and no risk factors for OA progression. The symptomatic OA progression cohort was selected based on the same initial baseline criteria, but with a change in WOMAC score >10 at 3 y follow-up indicating progression to symptomatic OA. The consort diagram is shown in Fig. 7. An equal number of control subjects were included in the analysis. These patients were used to discover and understand the cartilage morphology underlying future symptomatic OA progression.

OAI exclusion criteria were history of rheumatoid arthritis, bilateral total knee arthroplasty, or positive pregnancy test. The data was obtained from the OAI database. Institutional review board approval was obtained by participating institutions and informed consent was obtained by all participants in the study (45).

MRI Acquisition. Images were obtained according to the OAI protocol (38). Sagittal 2D T2-weighted images of the right knee at baseline were obtained using a standard extremity bird-cage coil on a MAGENTOM Tri 3T Multisite scanner (Siemens). A multiecho spin-echo (MESE) sequence was used with echo time (TE) linearly increased from 10 to 70 ms to acquire a total of seven images (repetition time = 2,700 ms, field of view = 120 × 120, matrix 384 × 384 with 27 slices of 3 mm thickness). Full details of MRI acquisition are in Peterfy et al. (38). Baseline knee radiographs were bilateral standing posterior–anterior fixed flexion view and were graded according to the KL scoring system (46).

Registration, Segmentation, and Preprocessing. For segmentation and generation of cartilage thickness maps, we used the Knee Imaging Quantification (KIQ) framework. KIQ combines rigid multiatlas registration with voxel classification and statistical shape modeling in a multistructure setting (47). The framework was trained on the OAI cohort MRIs with semiautomatic segmentations provided by iMorphics. These cartilage and meniscus training segmentations were supplemented with manual tibia bone training segmentations done by Biomediq. The KIQ registration method driven by normalized mutual information was also used for transfer of the multiecho T2 scan intensities to the double-echo steady-state (DESS) scan volume for each subject. The statistical shape model in KIQ was also trained on the training scan segmentations (48) and the resulting multiobject shape

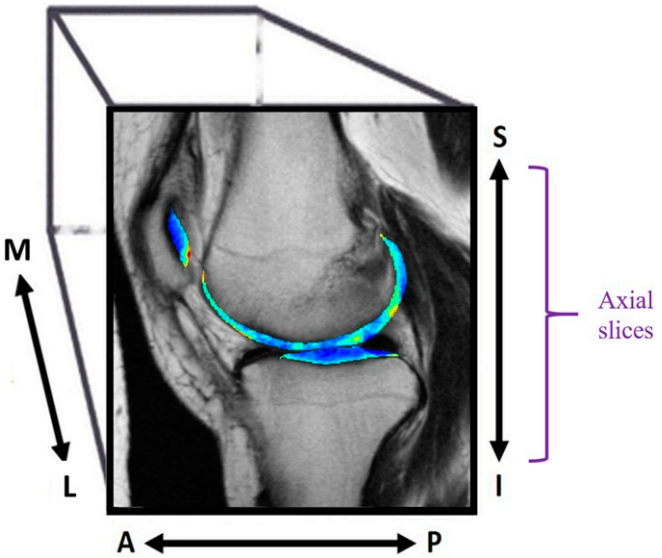

Fig. 8. Sagittal knee image showing segmented femoral, tibial, and patellofemoral cartilage maps. The cartilage T2 intensity has been colorized.

model provided a coordinate system with intersubject anatomical correspondence. This shape model coordinate system was used to transfer all scan intensities to the shape model mean shape using thin-plate spline warping steered by the shape model key points followed by trilinear interpolation (Fig. 8). Thereby, the T2 scan intensities could be analyzed in a common scan volume via the shape models. After segmentation, the sagittal hemispheric curvature of each condyle was mapped to a linear plane to facilitate visual interpretation of changes through the cartilage thickness using landmark-based thin plate spline (TPS) registration (49). More details are contained in *SI Appendix*, section 1.

A common reference template $I_0(x)$ was generated by obtaining an image from a subject with BMI 29.6 and age 56 y.

Cohort Statistics. After preprocessing, our final cohort consisted of baseline cartilage texture maps $I_i(x)$ corresponding to healthy individuals with no symptomatic or radiographic evidence of osteoarthritis $i = 1, \ldots, N$ for the $N = 86$ subjects in the study. Ground truth was established on 3 y follow-up to determine which of these subjects had gone on to progress to symptomatic osteoarthritis.

Whole-Organ Magnetic Resonance Imaging Score. As Fig. 1 indicates, visual inspection appeared to be insensitive to subtle biochemical changes reflected in the image domain. Whole-organ magnetic resonance imaging scores (WORMS) (50) were quantified on baseline cartilage maps and those on 3 y follow-up. These results are presented in *SI Appendix*, section 2.

Cartilage Texture Morphometry. Cartilage texture portrays molecular information about collagen fiber organization, water, and proteoglycan content through T2 intensity (8–10). We evaluate the biochemical organization of the articular cartilage tissue of the progressors and nonprogressors through texture morphometry.

In Fig. 1, the discriminating differences between the classes are not readily identified. The latter is the premise for our work in 3D TBM (28) to discover a hidden discriminant pattern from these images in an automated manner. In our prior work, we have demonstrated our approach discovers signature changes in the brains of older adults with age and aerobic fitness that are opaque to visual inspection and traditional morphometric techniques (28).

On the baseline cartilage maps, we quantified how the individual distribution of water (as implied by T2 intensity in $I_i(x)$) varied with respect to the common reference $I_0(x)$ (28) using TBM. The pairwise similarity between each source and template image was measured using the mathematics of OT (28) according to Eq. **1**. A transport map $f_i(x)$ was generated for each sample image in the dataset to characterize the optimal mass-preserving mapping. A transport map represents the transform of each image in the transport domain.

We developed all 3D TBM codes in MATLAB (MathWorks).

Principal Component Analysis. Given the high dimensionality of the data in the transport space relative to the number of samples (HDLSS), we seek techniques to reduce noise and avoid overfitting in subsequent training phases.

One such strategy to reduce noise in the dataset was to remove the dimensions containing little contribution to the overall variance. We applied the PCA technique to eliminate internal correlation among the feature variables in the transport space. A transport map is the transform of a particular cartilage texture map. We concatenated the transport maps to form a data matrix in the transport domain. Let us call the centered data matrix that results $D \in \mathbb{R}^{p \times m}$, where p corresponds to the number of elements in each transport map and m is the number of subjects.

The data were then projected onto the principal components corresponding to the top 96% of the variance for the cartilage images during each cross-validation iteration. Based on singular-value decomposition analysis, the data matrix can be written as $D = U \Sigma V^T$.

Here, the columns of $U \in \mathbb{R}^{p \times (m-1)}$ contain the orthonormal eigenvectors of the covariance matrix S_T. Because $p >> m$, the data can be summarized by $m-1$ total orthonormal eigenvectors. The singular values can be used to compute the fraction of the variance captured by each eigenvector. The topmost d vectors, where $d < m-1$, corresponding to the top 96% of the variance, were retained in this study. A reduced-dimension matrix $X \in \mathbb{R}^{d \times m}$ is determined by projecting the data matrix onto the top d columns of matrix U according to $X = U^T D$ during each cross-validation iteration.

Supervised Learning. After removing internal correlation among feature variables and denoising, a predictive classifier is trained during each cross-validation iteration in the transport domain on this reduced-dimension space. Our classifier is built using supervised machine learning and evaluated on held-out data over multiple iterations of a balanced cross-validation procedure.

Balanced Cross-Validation further describes how the data are partitioned into training and holdout sets over multiple iterations of balanced cross-validation. Here, we describe training and testing phases for a single iteration.

Training phase. From the training data, a discriminant is computed that differentiates the progressor and nonprogressor classes using pLDA in the transport space (31). The reason is that Euclidean distances in the transport domain correspond to geodesic OT distances on the Riemannian manifold (further discussed in *OT-Based Recognition*).

The pLDA technique computes the discriminant boundary that maximally separates the progressor and nonprogressor classes in the transport space according to Eq. **3** and ref. 31,

$$w_{pLDA} = argmax_{\|w\|=1} \frac{w^T S_T w}{w^T (S_w + \alpha I) w}, \qquad [3]$$

where w_{pLDA} is a vector in the transport domain describing the optimal separating hyperplane between the $C = 2$ classes. We denote the column m of matrix X as x_m. The covariance matrix S_T of the collection of reduced-dimension transport maps can be computed according to Eq. **4**. The parameter α controls the tradeoff between the traditional linear discriminant analysis direction and one that lies in the PCA subspace. We set α = 1:

$$S_T = \frac{1}{M} \Sigma_m (x_m - \bar{x})(x_m - \bar{x})^T \quad \text{such that } \bar{x} = \frac{1}{M} \Sigma_{m=1}^{M} x_m. \qquad [4]$$

Here, $S_W = \Sigma_C \Sigma_{n \in C} (x_n - \bar{x_c})(x_n - \bar{x_c})$ and represents the within-class scatter matrix, with C = 2 representing the number of classes.

Testing phase. We evaluate the blind predictive accuracy of our trained classifier on the held-out data in the testing phase. The test data are projected onto the discriminant direction w_{pLDA} computed from the training phase and evaluated against the ground truth. Further details regarding the cross-validation procedure are included in *Balanced Cross-Validation*.

All pLDA and PCA codes were implemented in MATLAB.

Balanced Cross-Validation. We use a balanced cross-validation approach to evaluate the generalizability of our predictive test in identifying future OA progression. We utilize a complete leave-two-out cross-validation scheme. Complete cross-validation yields the best estimation of the generalization accuracy on unseen data (41). Through complete cross-validation, we mitigate the potential influence of incidental correlations.

In each balanced cross-validation iteration, we partition the data to leave out a pair consisting of a nonprogressor subject and a progressor subject (leave-two-out). Under complete leave-two-out cross-validation, our learning approach iterates over the combinatorial space of all feasible partitions of the data when two subjects are left out in a holdout set.

We iterate over all 1,849 feasible partitions of the dataset (i.e., combinatorial pairings of progressor and control subject) under our balanced cross-validation approach to determine the overall accuracy, sensitivity, and specificity in blind prediction. ROC curves were generated by computing the sensitivity and specificity of classification as the discriminant threshold is varied across the range of projection scores that are obtained when images are projected onto the linear discriminant boundary.

Null Hypothesis Testing. In addition to complete cross-validation, we furthermore ensure the robustness of our training algorithm by estimating the probability that a test accuracy as high as the one observed in this paper can be obtained by chance alone, the null hypothesis.

Under the null hypothesis, a relationship between cartilage texture maps and future osteoarthritis progression can be learned by random chance. We assess the likelihood of obtaining a test accuracy as high as the one obtained in this paper through random chance alone. From the null hypothesis, it follows statistically that the conditional probability distributions of the cartilage texture maps would be identical regardless of progressor or nonprogressor class labels (51).

Permutation testing is a nonparametric technique (52) to empirically estimate the error distribution of our test accuracy under the null hypothesis. In permutation testing, we arbitrarily permute the assignment of labels to data over T = 1,000 permutation tests. We calculate the empirical likelihood of obtaining a test accuracy higher than the observed test accuracy by chance over these iterations. Thus, we are able to assess the probability of the null hypothesis for the predictive test we develop in this paper. We report the resultant statistical significance along with measures of accuracy, sensitivity, and specificity in the testing phase in Tables 2 and 3.

Comparison with Existing Supervised-Learning Techniques. For comparison, supervised learning is performed on the original image voxels and numerical features extracted using the WNDCHRM (35) software normalized for different feature scales. Prior studies employed WNDCHRM features (12, 26) to risk-stratify patients based on T2 cartilage texture maps. Images were smoothed and preprocessed using the same TBM preprocessing pipeline.

Furthermore, an SVM classifier using Gaussian kernel and a random forest classifier were also trained with the same cross-validation and null hypothesis testing procedure on the feature sets. In prior studies, Urish et al. (11) employed an SVM classifier to risk-stratify healthy patients based on T2 texture maps of cartilage. For the SVM classifier, the kernel coefficient parameter γ and the regularization parameter C were chosen to be $1/D$ and 1, respectively, where D denotes the number of features in a training sample. The tolerance value for the optimization stopping criteria was 10^{-3}. For the random forest (RF) classifier, 100 decision trees were used in the experiments. The subsamples in each decision tree were drawn from the training data with replacement. The number of subsamples in each decision tree was equal to the training set size. The scikit-learn package (53) of Python was used to implement both the SVM and RF classifier algorithms.

In total, we compared our 3D TBM approach to eight different linear and nonlinear classifiers.

Visualizing the Discriminant Cartilage Phenotype. We emphasize that in contrast to previous approaches focusing on numerical feature extraction to compare cartilage texture maps (11, 12, 26), the metric space defined by transport-based morphometry is generative. Therefore, 3D TBM provides the ability to interrogate the learned classifier at any desired point to visualize the unique corresponding image in the image domain. In the transport domain, the discriminant directions used to classify each pair of subjects were averaged over 100 iterations of cross-validation to yield a single direction in the transport space. Along this discriminant direction, five discrete points were sampled at the interface between control and progressor cohorts and inverted through inverse TBM transformation (Eq. **2**): at 0 (the mean image) as well as ± 2 and ± 4 standard deviations toward either side of the boundary. Figs. 4–6 were synthesized through inverse TBM transformation from the images in the study population. These images show the cartilage texture changes enabling future symptomatic OA detection. The latter closes the gap between statistical modeling and biologic interpretation of shifts at the tissue level.

Data Availability. Data are available at the publicly accessibility NIH, The Osteoarthritis Initiative database, https://nda.nih.gov/oai/. Code is available at GitHub, https://github.com/rohdelab.

ACKNOWLEDGMENTS. G.K.R. and M.S.-E.-R. were supported in part by NIH Awards GM130825 and GM090033. This work was supported in part by the NIH, National Institute on Aging, Intramural Research Program. K.L.U. is supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (K08AR071494), the National Center for Advancing Translational Science (KL2TR0001856), the Orthopedic Research and Education Foundation, and the Musculoskeletal Tissue Foundation.

1. M. A. Weinstock, Early detection of melanoma. *Jama* **284**, 886–889 (2000).
2. S. Ravishankar, A. Jain, A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images" in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, 2009), pp. 210–217.
3. A. O. Amoako, G. G. A. Pujalte, Osteoarthritis in young, active, and athletic individuals. *Clin. Med. Insights Arthritis Musculoskel. Disord.* **7**, CMAMD.S14386 (2014).
4. J. Bedson, P. R. Croft, The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. *BMC Muscoloskel. Disord.* **9**, 116 (2008).
5. J. N. Katz, B. E. Earp, A. H. Gomoll, Surgical management of osteoarthritis. *Arthritis Care Res.* **62**, 1220–1228 (2010).
6. Centers for Disease Control, CDC statistics. https://www.cdc.gov/arthritis/data_statistics/cost.htm. Accessed 31 December 2018.
7. C. R. Chu, A. A. Williams, C. H. Coyle, M. E. Bowers, Early diagnosis to enable early treatment of pre-osteoarthritis. *Arthritis Res. Ther.* **14**, 212 (2012).
8. Y. Xia, J. B. Moody, H. Alhadlaq, Orientational dependence of T2 relaxation in articular cartilage: A microscopic MRI (μMRI) study. *Magn. Reson. Med.* **48**, 460–469 (2002).
9. M. M. Harrison, T. D. Cooke, S. B. Fisher, M. P. Griffin, Patterns of knee arthrosis and patellar subluxation. *Clin. Orthop. Relat. Res.* **309**, 56–63 (1994).
10. B. S. Yulish *et al.*, Juvenile rheumatoid arthritis: Assessment with MR imaging. *Radiology* **165**, 149–152 (1987).
11. K. L. Urish *et al.*, T2 texture index of cartilage can predict early symptomatic OA progression: Data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* **21**, 1550–1557 (2013).
12. B. G. Ashinsky *et al.*, Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *J. Orthop. Res.* **35**, 2243–2250 (2017).
13. T. J. Mosher, S. W. Pruett, Magnetic resonance imaging of superficial cartilage lesions: Role of contrast in lesion detection. *J. Magn. Reson. Imag.***10**, 178–182 (1999).
14. W. D. Blackburn Jr., W. K. Bernreuter, M. Rominger, L. L. Loose, Arthroscopic evaluation of knee articular cartilage: A comparison with plain radiographs and magnetic resonance imaging. *J. Rheumatol.* **21**, 675–679 (1994).
15. Y. Kawahara *et al.*, Fast spin-echo MR of the articular cartilage in the osteoarthrotic knee: Correlation of MR and arthroscopic findings. *Acta Radiologica* **39**, 120–125 (1998).
16. M. Ochi, Y. Sumen, T. Kanda, Y. Ikutas, K. Itoh, The diagnostic value and limitation of magnetic resonance imaging on chondral lesions in the knee joint. *Arthrosc. J. Arthrosc. Relat. Surg.* **10**, 176–183 (1994).
17. A. Jamshidi, J.-P. Pelletier, J. Martel-Pelletier, Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nat. Rev. Rheumatol.* **15**, 49–60 (2019).

18. J. A. Swedberg, J. R. Steinbauer, Osteoarthritis-epidemiology, pathophysiology, diagnosis, and treatment. *Am. Fam. Physician* **45**, 557–568 (1992).
19. B. J. Cole *et al.*, Hyaluronic acid versus platelet-rich plasma: A prospective, double-blind randomized controlled trial comparing clinical outcomes and effects on intra-articular biology for the treatment of knee osteoarthritis. *Am. J. Sports Med.* **45**, 339–346 (2017).
20. G. Jamtvedt *et al.*, Physical therapy interventions for patients with osteoarthritis of the knee: An overview of systematic reviews. *Phys. Ther.* **88**, 123–136 (2008).
21. U. Nöth, A. F. Steinert, R. S. Tuan, Technology insight: Adult mesenchymal stem cells for osteoarthritis therapy. *Nat. Rev. Rheumatol.* **4**, 371–380 (2008).
22. R. Stahl *et al.*, T1rho, T 2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early OA patients—A 3.0-tesla MRI study. *Eur. Radiol.* **19**, 132–143 (2009).
23. Y.-X. J. Wáng *et al.*, T1ρ magnetic resonance: Basic physics principles and applications in knee and intervertebral disc imaging. *Quant. Imag. Med. Surg.* **5**, 858 (2015).
24. Y. Zhang, J. M. Jordan, Epidemiology of osteoarthritis. *Clin. Geriatr. Med.* **26**, 355–369 (2010).
25. C. Kauffmann *et al.*, Computer-aided method for quantification of cartilage thickness and volume changes using MRI: Validation study using a synthetic model. *IEEE Trans. Biomed. Eng.* **50**, 978–988 (2003).
26. B. G. Ashinsky *et al.*, Machine learning classification of OARSI-scored human articular cartilage using magnetic resonance imaging. *Osteoarthritis Cartilage* **23**, 1704–1712 (2015).
27. S. Basu, S. Kolouri, G. K. Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3448–3453 (2014).
28. S. Kundu *et al.*, Discovery and visualization of structural biomarkers from MRI using transport-based morphometry. *NeuroImage* **167**, 256–275 (2018).
29. W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, G. K. Rohde, A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *Int. J. Comput. Vis.* **101**, 254–269 (2013).
30. S. Kolouri, A. B. Tosun, J. A. Ozolek, G. K. Rohde, A continuous linear optimal transport approach for pattern analysis in image datasets. *Pattern Recogn.* **51**, 453–462 (2016).
31. W. Wang, Y. Mo, J. A. Ozolek, G. K. Rohde, Penalized Fisher discriminant analysis and its application to image-based morphometry. *Pattern Recogn. Lett.* **32**, 2128–2135 (2011).
32. R. A. Washburn, K. W. Smith, A. M. Jette, C. A. Janney, The physical activity scale for the elderly (PASE): Development and evaluation. *J. Clin. Epidemiol.* **46**, 153–162 (1993).
33. N. Bellamy, W. W. Buchanan, C. H. Goldsmith, J. Campbell, L. W. Stitt, Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J. Rheumatol.* **15**, 1833–1840 (1988).
34. M. D. Kohn, A. A. Sassoon, N. D. Fernando, Classifications in brief: Kellgren-Lawrence classification of osteoarthritis *Clin. Orthop. Relat. Res.* **478**, 1886–1893 (2016).
35. L. Shamir *et al.*, WNDCHRM–an open source utility for biological image analysis. *Source Code Biol. Med.* **3**, 13 (2008).
36. S. Kundu, A. Ghodadra, S. Fakhran, L. M. Alhilali, G. K. Rohde, Assessing postconcussive reaction time using transport-based morphometry of diffusion tensor images. *Am. J. Neuroradiol.* **40**, 1117–1123 (2019).
37. T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001), vol. 1.
38. C. G. Peterfy, E. Schneider, M. Nevitt, The osteoarthritis initiative: Report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis Cartilage* **16**, 1433–1441 (2008).
39. P. K. Raju, S. G. Kini, A. Verma, Wear patterns of tibiofemoral articulation in osteoarthritic knees: Analysis and review of literature. *Arch. Orthop. Trauma Surg.* **132**, 1267–1271 (2012).
40. Centers for Disease Control, OA basics. http://www.cdc.gov/arthritis/basics/osteoarthritis.htm. Accessed 28 October 2015.
41. R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection" in *International Joint Conferences on Artificial Intelligence* (Montreal, Canada, 1995), vol. 14, pp. 1137–1145.
42. S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, G. K. Rohde, Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Process. Mag.* **34**, 43–59 (2017).
43. S. R. Park, S. Kolouri, S. Kundu, G. K. Rohde, The cumulative distribution transform and linear pattern classification. *Appl. Comput. Harmon. Anal.* **45**, 616–641 (2018).
44. Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44**, 375–417 (1991).
45. G. Lester, M. C. Nevitt, D. T. Felson, *The Osteoarthritis Initiative: Protocol for the Cohort Study* (University of California, San Francisco; Boston University; National Institute of Arthritis, Musculoskeletal and Skin Diseases).
46. J. H. Kellgren, J. S. Lawrence, Radiological assessment of osteo-arthrosis. *Ann. Rheum. Dis.* **16**, 494–502 (1957).
47. E. B. Dam, M. Lillholm, J. Marques, M. Nielsen, Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. *J. Med. Imag.* **2**, 024001 (2015).
48. E. B. Dam, J. Runhaar, S. Bierma-Zienstra, M. Karsdal, Cartilage cavity—An MRI marker of cartilage lesions in knee OA with data from CCBR, OAI, and PROOF. *Magn. Reson. Med.*, **80**, 1219–1232 (2018).
49. F. L. Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).
50. C. G. Peterfy *et al.*, Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage* **12**, 177–190 (2004).
51. P. Golland, F. Liang, S. Mukherjee, D. Panchenko, "Permutation tests for classification" in *International Conference on Computational Learning Theory* (Springer, Bertinoro, Italy, 2005), pp. 501–515.
52. P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer Science & Business Media, 2013).
53. F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).