# SUPPORT VECTOR DRIVEN MARKOV RANDOM FIELDS TOWARDS DTI SEGMENTATION OF THE HUMAN SKELETAL MUSCLE

*R. Neji*[1,2,3], *G. Fleury*[3], *J-F. Deux*[4], *A. Rahmouni*[4], *G. Bassez*[4], *A. Vignaud*[5], *N. Paragios*[1,2]

[1]Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France
[2]Equipe GALEN, INRIA Saclay - Île-de-France, Orsay, France
[3]Département SSE, Ecole Supérieure d'Electricité, Gif-sur-Yvette, France
[4]Centre Hospitalier Universitaire Henri Mondor, Créteil, France
[5]Siemens Medical Solutions, Saint Denis, France

## ABSTRACT

In this paper we propose a classification-based method towards the segmentation of diffusion tensor images. We use Support Vector Machines to classify diffusion tensors and we extend linear classification to the non linear case. To this end, we discuss and evaluate three different classes of kernels on the space of symmetric definite positive matrices that are well suited for the classification of tensor data. We impose spatial constraints by means of a Markov random field model that takes into account the result of SVM classification. Experimental results are provided for diffusion tensor images of human skeletal muscles. They demonstrate the potential of our method in discriminating the different muscle groups.

***Index Terms*—** Diffusion Tensor Imaging, Support Vector Machines, Kernels, Markov Random Fields, Human Skeletal Muscle

## 1. INTRODUCTION

Diffusion tensor imaging (DTI) is a modality that gives insight in the diffusion properties of the water molecules within anatomical tissues [1]. Several studies have investigated the possibility of using DTI in order to separate different structures according to scalar criteria such as the principal direction of diffusion and the fractional anisotropy [2]. However the problem of segmentation and classification of diffusion tensors is still open and investigating the use of common machine learning techniques like Support Vector Machines (SVMs) can lead to include a priori knowledge about the structure of the tensor field and to further assess the potential of DTI in discriminating different anatomical regions within the human skeletal muscle, leading to a better understanding and quantification of muscular diseases (myopathies) and their effects on water diffusion. While previous studies focused mainly on using DTI to distinguish between the regions of the brain, DTI muscle segmentation has lacked attention and remains a difficult issue since muscular fibers belonging to some groups have approximately the same direction [3].

The existing methods of diffusion tensor segmentation and classification rely broadly on three techniques. Manifold learning techniques [4] have been proposed in order to reduce dimensionality and do the clustering in the induced low-dimensional space. However most of these methods suffer from two shortcomings: they are unable to provide an explicit mapping from the high-dimensional space to the low-dimensional one limiting their use in the testing phase. Besides, classification in the relatively high dimensional space may yield better results as the data incur an information loss after dimensionality reduction. Moreover, the unsupervised versions of these algorithms may have unsatisfactory separation capability. Another class of approaches consists in building a graph over the tensor field using different variations of spectral clustering where segmentation is based on the construction and the eigenanalysis of an affinity matrix between fiber tracts [5, 6] or graph cuts using seed points [7]. However, while these methods try to learn the structure of the tensor field from the data, they do not include prior knowledge or they require fiber extraction. The third class mainly relies on level-sets and active contours and may lead to local minima [8, 9]. In this work, we learn the diffusion tensors using linear and non linear SVMs. We define specific kernels on the space of symmetric positive definite matrices and use them for non linear classification. A Markov random field (MRF) captures the scores resulting from the SVMs and imposes spatial regularity. This is essential to account for the subtle differences in the muscular fibers characteristics. We show experimental results on diffusion tensor images towards segmentation of the human skeletal muscle (calf).

## 2. OVERVIEW

Given diffusion tensor data, we train a Support Vector Machine classifier using segmented data to learn the diffusion tensors of different muscle groups. To better capture the structure of the different groups, we use specific kernels in the non-linear SVM framework. During the testing phase, the SVM results are incorporated in a MRF model to take into account spatial information as well as noise. The remainder of this paper is organized as follows: in section 3, we detail the SVM classification of tensors and discuss the link between linear/non linear classification of diffusion tensors and the classification on the statistical manifold of normal multivariate distributions which leads us to define a family of kernels on the tensor space. In section 4, we introduce the MRF framework and show how the SVM scores can guide the segmentation while respecting the structure of the tensor field. Section 5 is dedicated to the assessment of the kernels and to the experimental results for the segmentation of the human skeletal muscle. In section 6, we discuss the perspectives of this work.

This work was partially supported by Association Française contre les Myopathies (AFM: `http://www.afm-france.org`) under the DTI-MUSCLE project.

## 3. SVM CLASSIFICATION AND KERNELS ON TENSORS

### 3.1. Two-class Support Vector Machines

We briefly review the principles of two class SVMs [10]. Given $N$ points $\mathbf{x}_i$ with known class information $y_i$ (either $+1$ or $-1$), SVM training consists in finding the optimal separating hyperplane described by the equation $\mathbf{w}^t\mathbf{x} + b = 0$ with the maximum distance to the training examples. It amounts to solving a dual convex quadratic optimization problem and each data point $\mathbf{x}$ is classified using the SVM output function $f(x) = (\sum_i^N \alpha_i y_i \mathbf{x}\mathbf{x}_i) + b$. The algorithm is extended to achieve non linear separation using a kernel function $K(\mathbf{x}, \mathbf{y})$(symmetric, positive definite) that is used instead of the standard inner product.

### 3.2. Linear classification of tensors

The binary linear classification of tensors belonging to two classes $C_1$ and $C_2$ consists in finding a linear form $\phi_{\mathbf{M}} : \mathcal{M}_3(\mathbb{R}) \rightarrow \mathbb{R}$, $\phi_{\mathbf{M}}(\mathbf{D}) = tr(\mathbf{M}^t\mathbf{D})$ such that $\phi_{\mathbf{M}}(\mathbf{D}) \leq b, \forall \mathbf{D} \in C_1$ and $\phi_{\mathbf{M}}(\mathbf{D}) > b, \forall \mathbf{D} \in C_2$. We can understand this linear classification as stemming from a marginalized kernel Fisher classification over the space of random water displacements.

Indeed, we consider the diffusion tensor field as a family of Gaussian probability distributions with zero mean over the set $\mathbb{R}^3$ of displacements. Thus the Fisher score [11] is given by $\nabla_{\mathbf{D}} \log P(\mathbf{x}|\mathbf{D})$ and indicates the direction in which $\mathbf{D}$ should be moved to explain the displacement $\mathbf{x}$ and yields the following kernel (on the set $\mathbb{R}^3$) using the standard scalar product:

$$K(\mathbf{x}, \mathbf{y}) =< \nabla_{\mathbf{D}} \log P(\mathbf{x}|\mathbf{D}), \nabla_{\mathbf{D}} \log P(\mathbf{y}|\mathbf{D}) >$$

We can verify that $\nabla_{\mathbf{D}} \log P(\mathbf{x}|\mathbf{D}) \propto \mathbf{D}^{-1}\mathbf{x}\mathbf{x}^t\mathbf{D}^{-1}$. Starting from an isotropic distribution $\widehat{\mathbf{D}} = \alpha\mathbf{Id}$, we have $\nabla_{\widehat{\mathbf{D}}} \log P(\mathbf{x}|\mathbf{D}) \propto \mathbf{x}\mathbf{x}^t$. In order to define a kernel over the diffusion tensor field, we average over all the possible displacements:

$$\begin{aligned} K(\mathbf{D}_1, \mathbf{D}_2) &= \int\int K(\mathbf{x}, \mathbf{y}) p(\mathbf{x}|\mathbf{D}_1) p(\mathbf{y}|\mathbf{D}_2) d\mathbf{x} d\mathbf{y} \\ &\propto < E(\mathbf{x}\mathbf{x}^t), E(\mathbf{y}\mathbf{y}^t) > \end{aligned}$$

Knowing that $E(\mathbf{x}\mathbf{x}^t) = \mathbf{D}_1$ and $E(\mathbf{y}\mathbf{y}^t) = \mathbf{D}_2$ we conclude that the kernel in the tensor space is given simply by $K(\mathbf{D}_1, \mathbf{D}_2) =< \mathbf{D}_1, \mathbf{D}_2 >$, that is the Euclidean scalar product that defines the linear classification.

We now extend this formulation to the non linear case by defining kernels that are specific to the space of symmetric positive definite matrices.

### 3.3. Information diffusion kernel

In order to define a kernel on the set of symmetric positive definite matrices, we can propagate class and structure information using its geometry as a Riemannian manifold [12]. Intuitively, we can see the construction of this kernel as diffusing the labels of the training set to the whole set of symmetric positive definite matrices. Therefore, similarly to heat diffusion on a Euclidean space, where the solution is given by the convolution of the initial condition by a Gaussian kernel, heat diffusion on a Riemannian manifold is driven by a kernel function $K_t$ and given by the following asymptotic series expansion [13]:

$$K_t(\mathbf{D}_1, \mathbf{D}_2) \propto \exp(-\frac{d^2(\mathbf{D}_1, \mathbf{D}_2)}{4t}) \sum_{n=0}^{\infty} a_n(\mathbf{D}_1, \mathbf{D}_2) t^n$$

where $d$ corresponds to the geodesic distance induced by the Riemannian metric, $a_n$ are the coefficients of the series expansion and $t$ is the diffusion time, which is a parameter of the kernel. We use a first order approximation in $t$ of the previous expression that yields

$$K_t(\mathbf{D}_1, \mathbf{D}_2) \propto \exp(-\frac{d^2(\mathbf{D}_1, \mathbf{D}_2)}{4t})$$

In our case, $d$ has an explicit expression given by $d(\mathbf{D}_1, \mathbf{D}_2) = \sqrt{\sum_i (\log(\lambda_i))^2}$ where $\lambda_i$ are the generalized eigenvalues of $\mathbf{D}_1$ and $\mathbf{D}_2$ [12].

### 3.4. Bregman divergence kernels

Instead of using the geodesic distance in the information diffusion kernel, one can instead use the Bregman divergence framework [14] to define a rich class of kernels parametrized by a convex scalar function $\phi : S_+^3 \rightarrow \mathcal{R}$ that extend in a natural way the Euclidean distance and therefore the standard Gaussian radial basis function kernel. Knowing $\phi$, one can define the corresponding Bregman divergence $\mathcal{D}_\phi$ between two matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ as follows :

$$\mathcal{D}_\phi(\mathbf{D_1}, \mathbf{D_2}) = \phi(\mathbf{D_1}) - \phi(\mathbf{D_2}) - tr(\nabla\phi(\mathbf{D_2})^t(\mathbf{D_1} - \mathbf{D_2}))$$

The symmetrization of the divergence gives the following similarity measure $\widehat{\mathcal{D}}$:

$$\widehat{\mathcal{D}}_\phi(\mathbf{D_1}, \mathbf{D_2}) = tr((\nabla\phi(\mathbf{D_1}) - \nabla\phi(\mathbf{D_2}))^t(\mathbf{D_1} - \mathbf{D_2}))$$

It is clear that choosing $\phi(\mathbf{D}) = \frac{1}{2}\|\mathbf{D}\|_F^2$ where $\|.\|_F$ is the Frobenius norm yields the standard Euclidean distance. Therefore, we extend the Gaussian radial basis function (RBF) kernel using the exponential embedding:

$$K(\mathbf{D}_1, \mathbf{D}_2) = \exp(-\gamma\widehat{\mathcal{D}}_\phi(\mathbf{D_1}, \mathbf{D_2}))$$

Two interesting cases of $\phi$ are the Burg entropy [Eq. 1] and the Von Neumann Entropy [Eq. 2]:

$$\phi_1(\mathbf{D}) = -\log(det(\mathbf{D})) \quad (1)$$

$$\phi_2(\mathbf{D}) = tr(\mathbf{D}\log(\mathbf{D}) - \mathbf{D}) \quad (2)$$

They induce the following kernels:

$$\begin{aligned} K_1(\mathbf{D}_1, \mathbf{D}_2) &= \exp(-\gamma tr((\mathbf{D_1} - \mathbf{D_2})(\mathbf{D_1}^{-1} - \mathbf{D_2}^{-1})) \\ K_2(\mathbf{D}_1, \mathbf{D}_2) &= \exp(-\gamma tr((\mathbf{D_1} - \mathbf{D_2})(\log(\mathbf{D_1}) - \log(\mathbf{D_2}))) \end{aligned}$$

These kernels provide global similarity measures that quantify the differences between tensors both in eigenvalues and eigenvectors. Note that the divergence that derives from Burg entropy can also be obtained from a Kullback-Leibler divergence between Gaussian distributions with zero mean [8].

### 3.5. Probability product kernels

The third kernel we will study is a probability product kernel [15]. We consider again the set of Gaussian probability distributions of zero mean, two elements $p_1$ and $p_2$ of this set with $\mathbf{D}_1$ and $\mathbf{D}_2$ their covariance matrices and the corresponding probability product kernel:

$$K(p_1, p_2) = \int p_1(\mathbf{x})^\rho p_2(\mathbf{x})^\rho d\mathbf{x} =< p_1^\rho, p_2^\rho >_{L^2}$$

where $\rho$ is a positive constant. Note that the special case $\rho = 1/2$ coincides with the well known Bhattacharyya kernel. Replacing the

probabilities $p_1$ and $p_2$ by their expressions gives the following kernel:

$$K(\mathbf{D}_1, \mathbf{D}_2) = \det(\mathbf{D}_1)^{-\rho/2} \det(\mathbf{D}_2)^{-\rho/2} \det(\mathbf{D}_1^{-1} + \mathbf{D}_2^{-1})^{-\rho}$$

This defines a Mercer kernel as the probability product kernel on Gaussian distributions is based on the scalar product in $L^2$ which is a Hilbert space. This is an advantage over the two above-cited classes of kernels which are not necessarily positive definite. Note however that in practice, a thorough choice of the parameter $\gamma$ will ensure the positive definiteness in a statistical sense, i.e. the property will hold with high probability [16].

## 4. MRF REGULARIZATION

The goal behind the use of an MRF model is two-fold: we aim at including spatial information, i.e. tensors along the same fiber should belong to the same class and we also try to minimize the effect of noise during segmentation. Besides, the MRF framework allows to use all the scores given by the SVMs, instead of making the labeling decision by simply taking the maximum score. Therefore, we define the following energy $E$ to minimize:

$$E = \sum_{i \in \Omega} u^s(l(i)) + \lambda \sum_{i \in \Omega, j \in \mathcal{N}(i)} u^p(l(i), l(j)) \qquad (3)$$

where $\Omega$ is the image domain, $l(i)$ is the label of the voxel $i$, $\mathcal{N}(i)$ is the considered neighborhood, $u^s$ is the potential given by the SVM scores and $u^p$ is a pairwise potential that imposes spatial regularity. We choose $u^s(l(i)) = \exp(-\alpha f_l(\mathbf{D}(i)))$ which is a decreasing potential in the score given by a one-against-all SVM classifier $f_l$. If the voxels $i$ and $j$ belong to the same axial slice, the pairwise potential $u^p$ is set to

$$u^p(l(i), l(j)) = c(1 - \delta_{l(i),l(j)})$$

where $c$ is a constant and $\delta$ is the Kronecker delta. If not, we choose

$$u^p(l(i), l(j)) = (1 - \frac{1}{2}\frac{\mathbf{v}^t\mathbf{D}(i)\mathbf{v}}{\lambda_{max}(i)} - \frac{1}{2}\frac{\mathbf{v}^t\mathbf{D}(j)\mathbf{v}}{\lambda_{max}(j)})(1 - \delta_{l(i),l(j)})$$

where $\mathbf{v} = \frac{i-j}{\|i-j\|}$ and $\lambda_{max}(i)$ is the largest eigenvalue of $\mathbf{D}(i)$, $\lambda_{max}(i) = \max_{\|\mathbf{z}\|=1} \mathbf{z}^t\mathbf{D}(i)\mathbf{z}$ . This potential is low for tensors belonging to the same fiber. We propose two different costs because muscular fibers have a privileged direction since they follow the direction of the leg with a pennation angle. To minimize the energy defined in [Eq. 3], we use the optimization algorithm proposed in [17].

## 5. EXPERIMENTS AND RESULTS

### 5.1. Kernels comparison

Magnetic Resonance DTI of human skeletal muscle (calf) was performed on a 1.5T scanner with the following imaging parameters: repetition time (TR)= $3600\,ms$, echo time(TE) = $70\,ms$, slice thickness = $7\,mm$ and $b$ value of $700\,s.mm^{-2}$ with 13 repetitions. Simultaneously, a high-resolution T1-weighted image was acquired and manually segmented. To assess the behavior of the defined kernels, we consider two muscle groups: the soleus (SOL) and the medial gastrocnemius (MG). SVM classification was performed both in a linear and a non linear fashion using the above-defined kernels.

We motivate the use of kernel-based SVMs by focusing on the architecture of the soleus muscle. While the medial gastrocnemius is a unipennate muscle (the fibers have one line of action), the soleus is a bipennate one (the fibers have two lines of action) and exhibits a richer structure. As can be seen in [Fig.1] where the principal directions of diffusion in the MG and SOL muscles are displayed as points on the unit sphere, it is more natural and mathematically more sound to trace the decision boundary while respecting the manifold structure, rather than using a hyperplane to separate the different classes or flattening the manifold using the Gaussian RBF kernel. We compared the behavior of the defined kernels in separating the MG from the SOL using a set of 9904 diffusion tensors with approximately the same number of tensors for each class (4976 tensors belonging to MG, 4928 tensors from the SOL muscle). We subdivided this set into a training set and a testing one to evaluate the generalization errors (50% of the set for the training and 50% for the testing) . As shown in [Table 1], the kernels that are specific to the space of symmetric positive definite matrices perform better than the linear classification both in the training and testing phases. The best result was obtained for the information diffusion kernel with approximately 3% of classification error. Note that the number of support vectors has not increased much with respect to a linear classification.

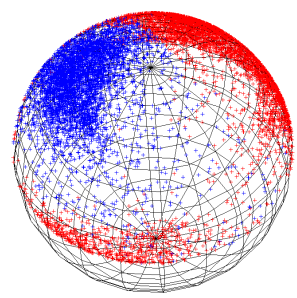

**Fig. 1**. Principal directions of diffusion of SOL and MG in red and blue respectively

| Method | Training error | Support vectors | Testing error |
|---|---|---|---|
| Linear | 5.27% | 741 | 5.81% |
| Diffusion | 0.26% | 999 | 2.89% |
| Probability product | 3.77% | 547 | 4.94% |
| Burg | 3.23% | 482 | 3.94% |
| V. Neumann | 3.25% | 494 | 4.05% |

**Table 1**. Performance of the different kernels

### 5.2. Segmentation of the calf muscle with SVM-driven MRFs

Two different diffusion datasets were considered to evaluate the segmentation algorithm, one for training and the other for testing. We used SVMs in a one against all fashion to learn the diffusion tensors of three major groups in the calf muscle: the soleus, the medial gastrocnemius and a third group consisting of the lateral gastrocnemius (LG) as well as the muscles of the anterior compartment of the calf. The learning was done on a manually segmented volume and we tested the performance of the SVM+MRF algorithm using the information diffusion kernel on another volume. As can be seen in

[Table 2], the MRF regularization improves significantly the correct classification rates with respect to a labeling decision based on the maximum score, achieving a correct classification rate of approximately 90% for each of the three classes. Qualitative results are provided in [Fig.2] (MG, SOL, LG and the muscles of the anterior compartment of the leg are displayed in red, yellow and green respectively). We can see that while SVM classification can be misled by noise, the combination of SVMs and MRF provides a smoother result and decreases the number of misclassifications.

| | class 1 | class 2 | class 3 |
|---|---|---|---|
| SVM | 85.71% | 82.53% | 76.4% |
| SVM+MRF | 88.64% | 90.22% | 88.42% |

**Table 2**. Comparison of correct classification rates with and without MRF regularization

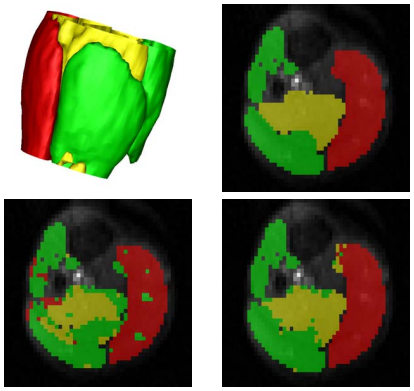

**Fig. 2**. Obtained 3D segmentation in three groups with some misclassifications visible, axial slice of the baseline image with overlaid segmentation for the ground truth, the SVM classification and the SVM+MRF algorithm respectively

## 6. CONCLUSION

We proposed a framework for the segmentation of the human skeletal muscle. It introduces a priori knowledge about the diffusion tensors belonging to different muscle groups combined with a spatial regularization based on MRFs. It is possible to further extend this framework to build kernels on entire fibers if we are dealing with a tensor field with a more complex structure (brain fibers for example). Considering naturally a fiber as a set of tensors one can use the summation kernel to define a kernel on the fiber space or more elaborate kernels on sets of features where mismatchings are taken into account.

Another interesting direction of research is to learn the diffusion tensors for different populations (healthy patients, ill patients affected by different kinds of myopathies) and apply the proposed classification framework to detect alteration in diffusion due to illness.

## 7. REFERENCES


[1] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, pp. 534–546, 2001.

[2] Craig J. Galban, Stefan Maderwald, Kai Uffmann, Armin de Greiff, and Mark E. Ladd, "Diffusive sensitivity to muscle architecture: a magnetic resonance diffusion tensor imaging study of the human calf," *European Journal of Applied Physiology*, vol. 93, no. 3, pp. 253 – 262, Dec 2004.

[3] D. C. Martin, M. K. Medri, R. S. Chow, V. Oxorn, R. N. Leekam, A. M. Agur, and N. H. McKee, "Comparing human skeletal muscle architectural parameters of cadavers with in vivo ultrasonographic measurements," *J Anat*, vol. 199, no. Pt 4, pp. 429–34, 2001.

[4] Andy Tsai, Carl-Fredrik Westin, Alfred O. Hero, and Alan S. Willsky, "Fiber tract clustering on manifolds with dual rooted-graphs," in *CVPR*, 2007.

[5] A. Brun, H. Knutsson, H. J. Park, M. E. Shenton, and C.-F. Westin, "Clustering fiber tracts using normalized cuts," in *MICCAI*, 2004.

[6] L. O'Donnell and C.-F. Westin, "White matter tract clustering and correspondence in populations," in *MICCAI*, 2005.

[7] Yonas T. Weldeselassie and Ghassan Hamarneh, "DT-MRI segmentation using graph cuts," 2007, vol. 6512, p. 65121K, SPIE.

[8] Zhizhou Wang and B.C. Vemuri, "DTI segmentation using an information theoretic tensor dissimilarity measure," *IEEE Transactions on Medical Imaging*, vol. 24, no. 10, pp. 1267–1277, 2005.

[9] L. Jonasson, P. Hagmann, C. Pollo, X. Bresson, C. Richero Wilson, R. Meuli, and J. Thiran, "A level set method for segmentation of the thalamus and its nuclei in DT-MRI," *Signal Process.*, vol. 87, no. 2, pp. 309–321, 2007.

[10] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.

[11] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," Tech. Rep., 1998.

[12] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, January 2006.

[13] John Lafferty and Guy Lebanon, "Diffusion kernels on statistical manifolds," *Journal of Machine Learning Research*, vol. 6, pp. 129–163, 2005.

[14] K. Tsuda, G. Rtsch, and M. Warmuth, "Matrix exponentiated gradient updates for on-line learning and Bregman projection," *Journal of Machine Learning Research*, vol. 6, pp. 995–1018, 06 2005.

[15] Tony Jebara, Risi Kondor, and Andrew Howard, "Probability product kernels," *Journal of Machine Learning Research*, vol. 5, pp. 819–844, 2004.

[16] S. Boughorbel, J.-P. Tarel, and F. Fleuret, "Non-Mercer kernels for SVM object recognition," in *BMVC*, 2004.

[17] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios, "Fast, approximately optimal solutions for single and dynamic MRFs.," in *CVPR*, 2007.