

## Research Article

# Pattern Classification of Signals Using Fisher Kernels

**Yashodhan Athavale,<sup>1</sup> Sridhar Krishnan,<sup>1</sup> and Aziz Guergachi<sup>2</sup>**

<sup>1</sup> Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3

<sup>2</sup> Ted Rogers School of Management, Ryerson University, Toronto, ON, Canada M5B 2K3

Correspondence should be addressed to Yashodhan Athavale, yashodhan15@gmail.com and Sridhar Krishnan, krishnan@ee.ryerson.ca

Received 20 May 2012; Accepted 6 August 2012

Academic Editor: Joao B. R. Do Val

Copyright © 2012 Yashodhan Athavale et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The intention of this study is to gauge the performance of Fisher kernels for dimension simplification and classification of time-series signals. Our research work has indicated that Fisher kernels have shown substantial improvement in signal classification by enabling clearer pattern visualization in three-dimensional space. In this paper, we will exhibit the performance of Fisher kernels for two domains: financial and biomedical. The financial domain study involves identifying the possibility of collapse or survival of a company trading in the stock market. For assessing the fate of each company, we have collected financial time-series composed of weekly closing stock prices in a common time frame, using Thomson Datastream software. The biomedical domain study involves knee signals collected using the vibration arthrometry technique. This study uses the severity of cartilage degeneration for classifying normal and abnormal knee joints. In both studies, we apply Fisher Kernels incorporated with a Gaussian mixture model (GMM) for dimension transformation into feature space, which is created as a three-dimensional plot for visualization and for further classification using support vector machines. From our experiments we observe that Fisher Kernel usage fits really well for both kinds of signals, with low classification error rates.

## 1. Fisher Kernels

Kernel functions may be of different forms, such as linear kernels, polynomial kernels, and so forth. One of the most popular forms of kernels is the Fisher kernels which were introduced in 1999 [1]. This paper highlights the functionality of Fisher kernels and how they were applied in our study.

Fisher kernels are known for their wide usage in generative and discriminative pattern classification, and recognition approaches [1–4]. Generative models are those which are

used for randomly generating observable data; whereas discriminative models are used in machine learning for assessing the dependency of an unobserved random variable on an observed variable using a conditional probability distribution function. They extract more information from a single generative model and not just its output probability [5]. The features obtained after applying Fisher kernels are known as *Fisher scores*. We analyze how these scores depend on the probabilistic model, and how they give us the information about the internal representation of the data items within the model.

The benefit of using Fisher kernels comes from the fact that they limit the dimensions of feature space in most cases thereby giving some regularity to visualization. This is very important when we are dealing with inseparable classes [6]. A typical Fisher kernel representation usually consists of numerous small gradients for higher probability objects in a model and vice-versa [7]. Fisher kernels combine the properties of generative models such as Hidden Markov model, and discriminative methods such as support vector machines [1]. The use of Fisher kernels for dimension transformation has the following two advantages [1, 5]:

- (1) Fisher kernels can deal with variable-length time series data;
- (2) discriminative methods such as SVMs, when used with Fisher kernels can yield better results.

Fisher kernels have been applied in various domains such as speech recognition [8–10], protein homology detection [3], web audio classification [11], object recognition [12], image decoding [13, 14], and so forth. For detailed information on Fisher kernels, refer to [5]. Following Table 1 indicates some recent applications of Fisher Kernels and their performances.

## 2. Applying Fisher Kernels to Time-Series Models

In this study, we investigate the application of Fisher kernels for identifying patterns emerging from a time-series model. We use a generative Gaussian mixture model [5, 21, 22] for our complex system. For a binary classification problem, there will be two Gaussian components for each time series signal. These Gaussian components for each signal help us in assessing the tendency of the signal to be categorized in either of the two classes. Let us make the following assumptions before we mathematically generate the Fisher score model. The methodology in this section has been adopted from [23].

- (i) The complex system to be modeled is a black box out of which a time series signal is emerging.
- (ii) In order to make the data distribution in the signal to be *i.i.d* (independent and identical distribution), we apply some mathematical relations such as log-normal relations or normalizing the dataset. By doing so, we generate another set of data which is nearly an *i.i.d* distribution, thus giving us a transformed time series signal.
- (iii) Upon applying the previous assumptions, we then generate the Fisher score model for the complex system as explained further.

Before we proceed with deriving the model and its equations, we assume the following variables and their meanings as shown in Table 2.

- (1) We use the assumptions mentioned in the beginning of this section for deriving the Fisher scores.

**Table 1:** Examples of Fisher kernel applications.

Application area	Performance
Abnormal activity recognition [15]	0.857—Maximum area under curve, best performance
Handwritten word spotting [16]	89% Average precision
Shape modeling and categorization [17]	Classification error of 0.45%
Signer independent sign language recognition [18]	Average classification rate of 93.08%
Image classification [19]	Average accuracy of 60%
Spam and image categorization [20]	Overall accuracy of 79%
Web audio classification [11]	Classification accuracy of 81.79%
Object recognition [12]	Average accuracy of 93%

**Table 2:** List of variables used for Fisher scores' computation.

Symbol	Indication
$X = x_1, \dots, x_N$	The input dataset, which consists of a number of vectors or time-series signals
$x$	An input vector or a time-series signal
$N$	Number of input vectors in the given input dataset
$N_d$	Dimensionality of the input vector
$N_c$	Number of samples available in an input vector $x_i$ . This value can be same for all input vectors within $X$ or can be of variable length.
$N_g$	Number of Gaussian components for clustering or pattern visualization = 2
$S_c$	$i$ th sample value in an input vector ( $i = 1, \dots, N_c$ )
$r_c$	Normalized value for the $i$ th sample
$\theta = [a_j, \mu_j, \sigma_j]$	Gaussian estimates for the $N_g$ components, with $a_j$ being the weight vector, $\mu_j$ being the mean vector and $\sigma_j$ being the variance vector.
$R(i, j)$	Gaussian mixture model for the $i$ th week's returns, built using $j = 2$ Gaussian components
$P(r_i   \theta)$	Probability density function for the $i$ th normalized sample value
$P(C   \theta)$	Probability density function for the entire input vector or time-series signal

- (2) Our study is on binary pattern classification, and in order to achieve this we use the Fisher scores generated as described further in this section, for plotting and visualizing between the two categories (e.g., active and dead companies, or abnormal and normal knee joints).
- (3) The length of each input vector (or the number of samples available) can be variable, or all the input vectors can be of same length. We assume each input vector to be a unique time series signal.
- (4) Our study based on binary pattern classification has been applied to two domains: the financial domain wherein we classify between the potentially dead and active companies; and the biomedical domain wherein we classify between abnormal and normal knee joints for assessing the risk of cartilage degeneration. In the financial study, we intend to find the log-normal stock returns using the weekly stock prices; whereas in the biomedical domain we normalize the knee angle signals between the interval  $0, \dots, 1$ . These stock returns or normalized knee angles are taken as our  $r_c$  values. We do this so as to make the distributions *i.i.d* in nature, as per the assumptions mentioned in the beginning of this section.

- (5) We first find the initial values of the Gaussian estimates,  $\theta = [a_j, \mu_j, \sigma_j]$  using the expectation maximization algorithm [24, 25] for  $j = 1, \dots, N_g$ . The expectation maximization algorithm is used for estimating the likelihood parameters of certain probabilistic models.
- (6) Using these estimates we create the Gaussian mixture model  $M$  so that  $R(i, j)$  is an  $N_c \times 2$  matrix

$$R(i, j) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{r_i - \mu_j}{\sigma_j}\right)^2\right) \quad \text{for } \begin{cases} i = 1, \dots, N_c \\ j = 1, \dots, N_g \end{cases}. \quad (2.1)$$

- (7) The diagonal covariance GMM likelihood is then given by the probability density function for the  $i$ th normalized sample value. Thus,  $P(r_i | M, \theta)$  is a  $N_c \times 1$  matrix

$$P(r_i | M, \theta) = \sum_{j=1}^{N_g} a_j R(i, j). \quad (2.2)$$

The global log-likelihood of an input vector's normalized values  $C = \{r_1, r_2, \dots, r_{N_c}\}$  is given using the probability density function as follows. Therefore,  $\log P(C | M, \theta)$  is a single value for each input time series signal:

$$\log P(C | M, \theta) = \sum_{i=1}^{N_c} \log P(r_i | M, \theta). \quad (2.3)$$

The Fisher score vector is composed of derivatives with respect to each parameter in  $\theta = [a_j, \mu_j, \sigma_j]$ .

The likelihood Fisher score vector for each signal is thus given as follows:

$$\Psi_{\text{fisher}}(C) = \left[ \frac{d}{da_j}, \dots, \frac{d}{d\mu_j}, \dots, \frac{d}{d\sigma_j} \right]^T \log P(C | M, \theta) \quad \text{for } j^* = 1, \dots, N_g. \quad (2.4)$$

Each of the derivatives comprises of two components, thus giving us a  $1 \times 2$  matrix for each derivative. Thus, for each input vector we get a  $6 \times 1$  Fisher score matrix. In order to plot the scores, we then add up each pair of Fisher scores (with respect to weights, mean, and variance), to get a three-dimensional scatter plot, as shown in equations below:

$$\begin{aligned} \text{SFS}_1 &= \left( \frac{d}{da_1} + \frac{d}{da_2} \right), \\ \text{SFS}_2 &= \left( \frac{d}{d\mu_1} + \frac{d}{d\mu_2} \right), \\ \text{SFS}_3 &= \left( \frac{d}{d\sigma_1} + \frac{d}{d\sigma_2} \right), \end{aligned} \quad (2.5)$$

where SFS—sum of Fisher scores.

The Fisher scores obtained for each input vector are then further used as input data for training and testing our SVM model. The SVM model basically performs binary classification, using which we can infer statements about the future state of the complex system taken into consideration. It should be noted that the datasets used our financial time-series experiments are balanced (256 active and 256 dead companies), whereas in biomedical time-series they are imbalanced (38 abnormal and 51 normal cases). In order to solve the problem of performance loss, we have used the Gaussian radial basis function as our kernel function, which creates a good classifier for nonoverlapping classes, and application of SMO (sequential minimal optimization) method for finding the hyperplane, which splits our large quadratic optimization problem into smaller portions for solving.

The correctness of the classification performed by SVM is further verified when we apply the same set of Fisher score data for linear discriminant analysis (LDA), along with the false positives versus the false negatives. As a note, our study is not intended to analyze the performance of LDA approach, or compare it with SVM. For detailed information on SVM concepts, the reader may refer to [26].

Following Section 3 describes our experiments with real-time data.

### **3. Experiments**

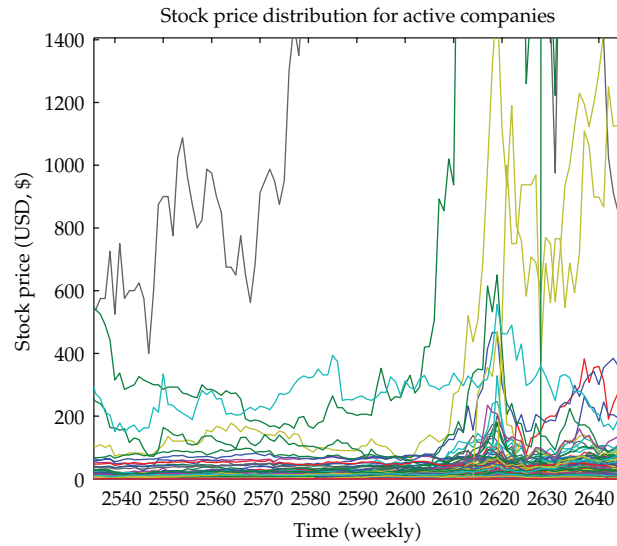
#### **3.1. Financial Time-Series**

In this study, we have considered the companies falling under the Pharmaceuticals and Biotechnology sectors listed in the TSX [27], NYSE [28], TSX-Ventures [27] and NASDAQ [29] stock exchanges. We collected the weekly stock price data for various companies from a common time frame of January 1950 to December 2008. Figures 1 and 2 illustrate the stock price distribution for the active and dead companies falling under the pharmaceuticals and biotechnology sector.

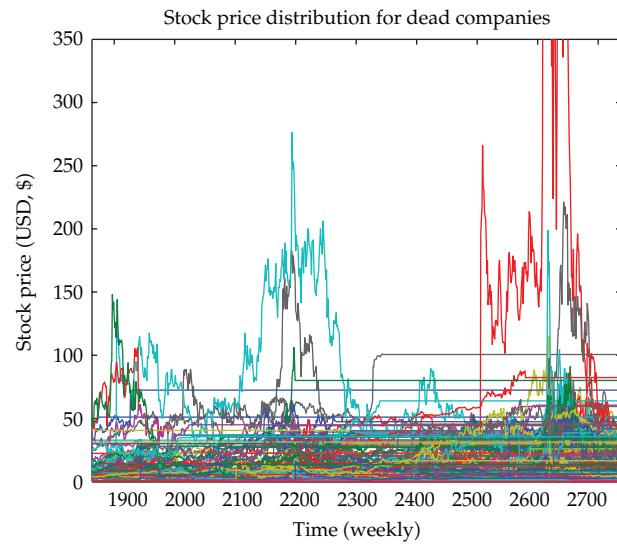
From observing the stock price distribution charts, it becomes clear that it is difficult or almost impossible to predict the next stock price or even the future state of the stock price, that is, whether the price will be high or low. In our study, by classifying between the active and dead companies, we have tried to infer statements about the performance of each company and whether it would be a potential survivor in the long run or not. Our experiments are not intended to predict an active company's rise or fall within a specific time range, but rather are developed to provide a qualitative measurement of its performance in the stock market with respect to the dead companies' cluster. In other words, based on cluster analysis we can infer that a company represented in three dimensions has more inclination to survive if it is nearer to an active cluster, or collapse if it is nearer to a dead cluster.

An "active" company in this context indicates that it is currently trading and is listed in a particular stock exchange. Whereas, a "dead" company indicates that the firm has been delisted from the stock exchange, and that it no longer performs stock trading. A company can be listed as "dead" for many reasons such as bankruptcy, mergers, or acquisitions. Thomson datastream [30] uses a flat plot or a constant value as an indication that the company has stopped trading in the exchange. This becomes clear from Figure 2.

As observed in Figures 1 and 2, the stock price distribution is not an i.i.d (independent and identical distribution). So in order to normalize the distribution, the datasets for various active and dead companies are then processed for getting the stock price returns using Black and Scholes theory [31–33]. Figures 3 and 4 illustrate the log-normal stock price returns' distribution. It might look trivial that, by definition, a dead company has no trading and



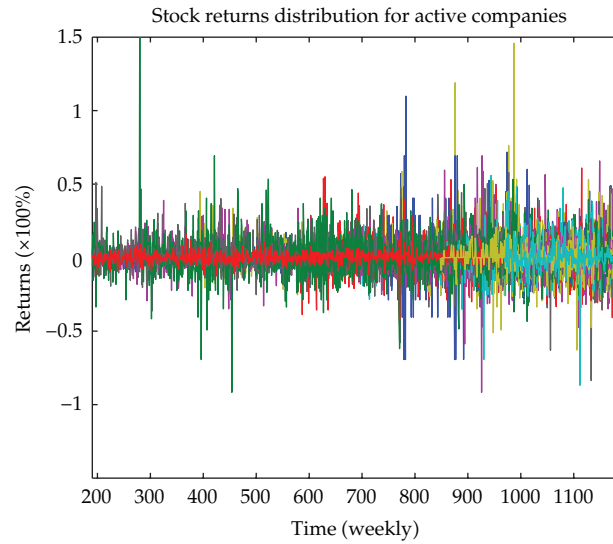
**Figure 1:** Stock price distribution of active companies.



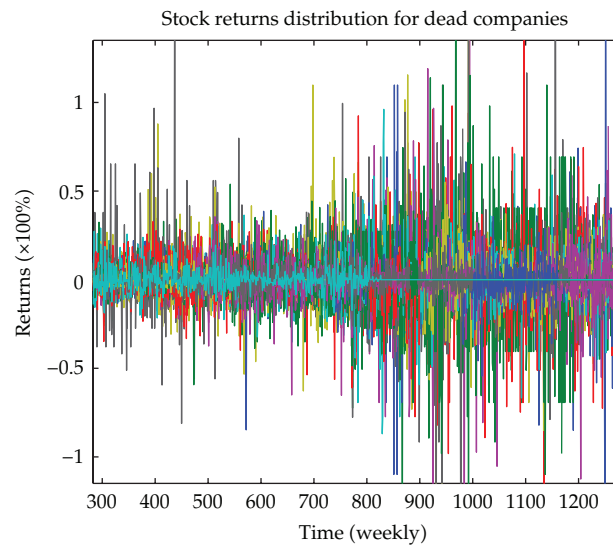
**Figure 2:** Stock price distribution of dead companies.

hence the stock returns must be close to zero. Our data collection and hence Figures 2 and 4 indicate that a constant stock price line indicate that the company stopped trading at that corresponding price, and there onwards the stock returns plot for each dead company indicates a convergence with the zero constant once the company stops trading.

The normalized stock returns data is then used for finding the initial estimates of mean, variance and weight vectors using the expectation maximization algorithm [24, 25]. The normalized dataset for each company is then processed using Fisher kernels implemented with a Gaussian mixture model in order to obtain the Fisher scores with respect to three parameters.



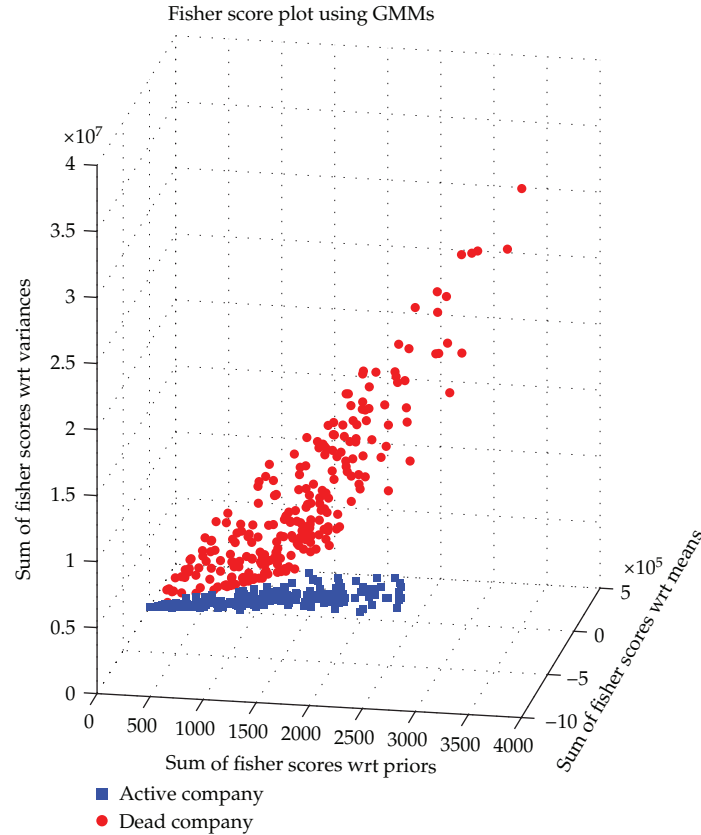
**Figure 3:** Log-normal stock returns distribution of active companies.



**Figure 4:** Log-normal stock returns distribution of dead companies.

These parameters are basically the derivatives of the global log-likelihood of each dataset with respect to each of mean, variance, and weight vectors. These Fisher scores when plotted in three dimensions provide a scope for visually classifying between the active and the dead companies, as shown in Figure 5. At this stage, we have basically performed a transformation of a financial time-series into six dimensions. That is, for each company we have processed its stock market data into a set of six Fisher scores. In order to plot these Fisher scores, we have summed up the Fisher score pairs for all the parameters.

SVMs were applied to both three-dimensional (Figure 5) and six-dimensional Fisher scores for classification and prediction. The results for this have been shown in Table 3. First,



**Figure 5:** Fisher score plot for visualizing financial time-series.

we randomly split the Fisher score dataset into training and testing groups. We then applied support vector machines for training and testing of the system, using our kernel functions as a Gaussian radial basis function (RBF). For finding the hyperplane separating the two classes, we used the method of sequential minimal optimization (SMO) [34].

In order to validate our results, we further used the method of linear discriminant analysis (LDA) [21] along with the leave-one-out cross validation technique, as shown in Tables 4 and 5.

- (i) In case of the three-dimensional Fisher scores, we obtained a classification accuracy of 95.9% in original grouped cases, and about 95.7% in cross validated cases.
- (ii) Similarly, in case of six-dimensional scores, the classification accuracy for both original grouped cases and cross validated cases was 95.7%.

### 3.2. Biomedical Time-Series

As mentioned, the Fisher kernel technique was also applied for classifying abnormal and normal knee joints. A database of 38 abnormal and 51 normal knee-joint case was used in our experiments. The knee-joint signal data was collected using vibration arthroscopy as described in [35]. Sample plots of the signals are shown in Figures 6 and 7.



**Table 3:** SVM performance results for financial time-series.

Dataset	Sensitivity (%)	Specificity (%)	Error rate (%)	Classification accuracy (%)
3D Fisher scores	92.8	100.0	3.57	96.43
6D Fisher scores	96.0	99.21	2.38	97.62

**Table 4:** LDA results for 3D Fisher scores: financial data.

Category		Predicted membership		
		Dead	Active	
Original data	Count of data classified			
		Dead	235	21
		Active	0	256
	% of data classified			
		Dead	91.8	8.2
		Active	0.0	100.0
Cross-validated data	Count of data classified			
		Dead	234	22
		Active	0	256
	% of data classified			
		Dead	91.4	8.6
		Active	0.0	100.0

In order to simplify our calculations, we normalize the dataset values for each case study between the interval  $0, \dots, 1$ , for generating the Fisher scores as shown in Figures 8 and 9.

Once the Fisher scores are generated using the method described in Section 2, we plot them (similar to the Fisher score plot for financial time-series) as shown in Figure 10.

SVMs were then applied to both three-dimensional and six-dimensional Fisher scores for classification and prediction. The results for this have been shown in Table 6.

The correctness of the classification performed by SVM is further verified when we apply the same set of Fisher score data for linear discriminant analysis (LDA) [21], as shown in Tables 7 and 8.

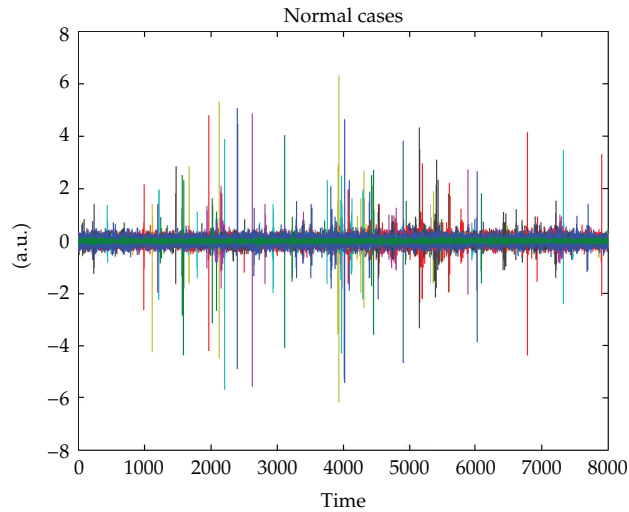
- (i) In case of the three-dimensional Fisher scores, we obtained a classification accuracy of 82.0% in original grouped cases, and about 75.3% in cross-validated cases.
- (ii) Similarly, in case of six-dimensional scores, we obtained a classification accuracy of 91.0% in original grouped cases, and about 88.8% in cross-validated cases.

#### 4. Discussions, Conclusions, and Future Works

In our previous work [36], Fisher kernels were not able to perform binary classification in two dimensions. But in this study, by introducing three-dimensional Fisher scores, we have been able to separate and visualize the two classes, more accurately. The intention of our research work in this study was to analyze the classification performance of time-series signals using Fisher kernels as feature extractors. Specifically when we classified active companies versus dead companies in a given economic sector, our intention was to see how good the dimension transformation of the time-series was. Also with regards to the separation between the two classes, we were attempting to predict the potential survival of a company using SVMs. In

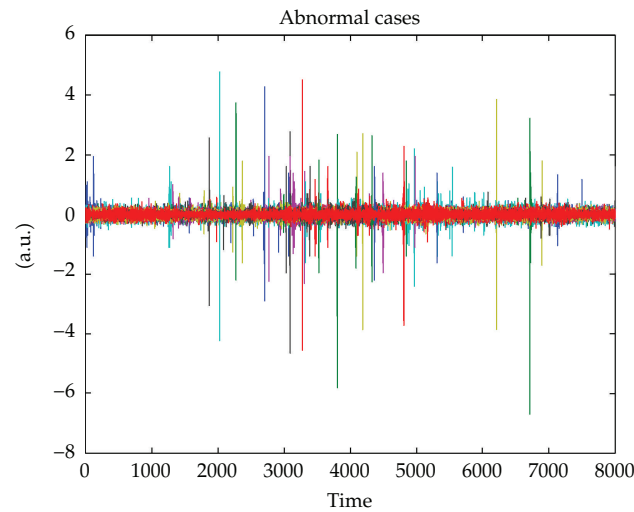
**Table 5:** LDA results for 6D Fisher scores: financial data.

Category		Predicted membership	
		Dead	Active
Original data	Count of data classified		
	Dead	234	22
	Active	0	256
	% of data classified		
	Dead	91.4	8.6
Cross-validated data	Count of data classified		
	Dead	234	22
	Active	0	256
	% of data classified		
	Dead	91.4	8.6
	Active	0.0	100.0

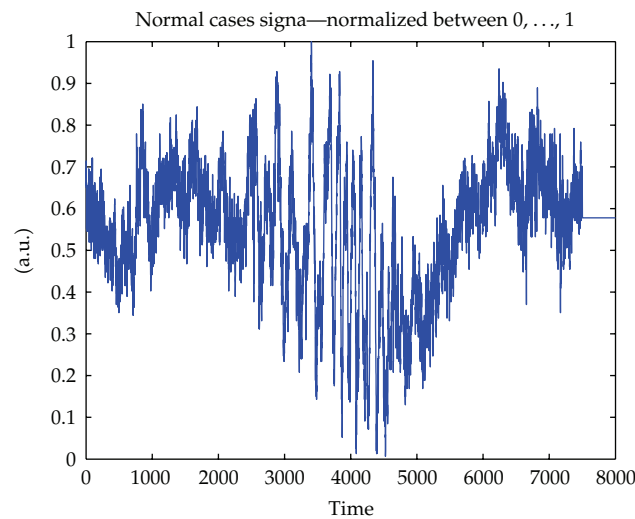
**Figure 6:** Normal knee signals.

other words, by visualizing the two clusters we observed that few active companies were more nearer to the dead cluster, which led to inferring that these companies could potentially collapse in the long run. This was evident from the observation that these active companies exhibited stock price changes similar to dead companies (before they collapsed). A similar observation could be derived from our biomedical time-series results.

A normal distribution is easier to analyze and model using GMM, as compared to a non-i.i.d distribution. In other words, we can say that Gaussian mixture models (GMM) give the best fit for normally distributed datasets. A qualitative observation of the time-series signals used in our experiments reveals that the distribution is Laplacian in nature. That is, although the histogram of a sample vector appears to be bell-shaped as is in a normal distribution, we actually observe that the curve is peaked around the mean value and has fat-tails on either sides. Upon further training, testing and cross-validation operations using SVMs and LDA, we have achieved a high classification rates in both the studies, as indicated in Tables 3 and 6. The highlighting factors behind such high classification rates could be as follows.



**Figure 7:** Abnormal knee signals.



**Figure 8:** Normal knee signals—normalized between 0, ..., 1.

- (i) The characteristic property of Fisher kernels—retaining the essential features during dimension transformation.
- (ii) The application of SMO method for finding the hyperplane, which splits out large quadratic optimization problem into smaller portions for solving.
- (iii) Other studies as indicated in Table 1, wherein Fisher kernels have performed exceptionally well.

Automation of our research work in order to yield dynamic outputs in the form of predictive statements and visualization plots can be pursued as a future study. Experimenting with variable-length time-series in this study has definitely opened doors for more research,

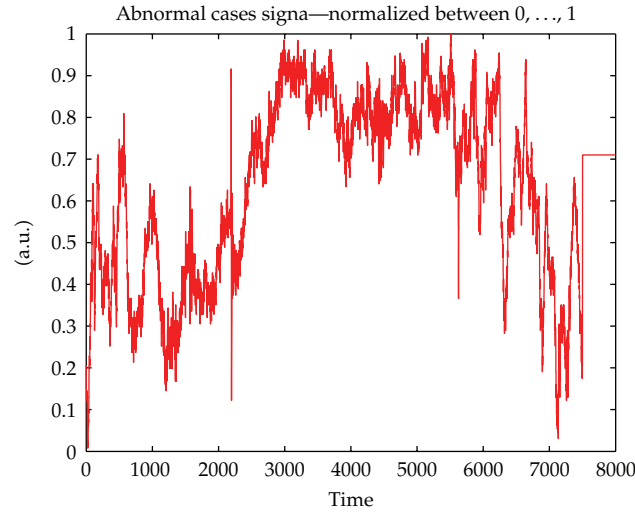


Figure 9: Abnormal knee signals—normalized between 0, ..., 1.

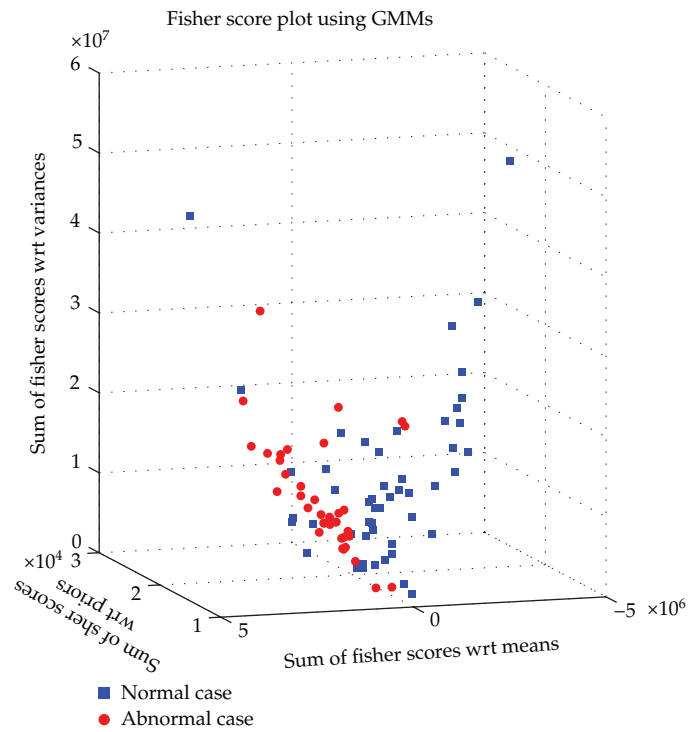


Figure 10: Fisher score plot for visualizing biomedical time-series.

Table 6: SVM performance results for biomedical time-series.

Dataset	Sensitivity (%)	Specificity (%)	Error rate (%)	Classification accuracy (%)
3D Fisher scores	94.7	84.0	11.37	88.63
6D Fisher scores	89.4	96.0	6.82	93.18

**Table 7:** LDA results for 3D Fisher scores: biomedical data.

Category		Predicted membership	
		Abnormal	Normal
Original data	Count of data classified		
	Abnormal	35	3
	Normal	13	38
	% of data classified		
	Abnormal	92.1	7.9
	Normal	25.5	74.5
Cross-validated data	Count of data classified		
	Abnormal	31	7
	Normal	15	36
	% of data classified		
	Abnormal	81.6	18.4
	Normal	29.4	70.6

**Table 8:** LDA results for 6D Fisher scores: biomedical data.

Category		Predicted membership	
		Abnormal	Normal
Original data	Count of data classified		
	Abnormal	34	4
	Normal	4	47
	% of data classified		
	Abnormal	89.5	10.5
	Normal	7.8	92.2
Cross-validated data	Count of data classified		
	Abnormal	34	4
	Normal	6	45
	% of data classified		
	Abnormal	89.5	10.5
	Normal	11.8	88.2

such as assessing the time-frame of cartilage degeneration and a scope for monitoring Osteoarthritis. Analyzing these issues in near future will be quite interesting and challenging.

## Acknowledgments

The authors extend their sincere thanks to NSERC Granting Council, CFI/OIT, and Canada Research Chairs program for funding the research.

## References

- [1] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," *Advances in Neural Information Processing Systems*, vol. 11, pp. 487–493, 1998.
- [2] K. Tsuda, S. Akaho, M. Kawanabe, and K. R. Müller, "Asymptotic properties of the fisher kernel," *Neural Computation*, vol. 16, no. 1, pp. 115–137, 2004.

- [3] T. Jaakkola, M. Diekhans, and D. Haussler, "Using the shier kernel method to detect remote protein homologies," in *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*, pp. 149–158, AAAI Press, Menlo Park, Calif, USA, 1999.
- [4] C. Saunders, J. Shawe-taylor, and A. Vinokourov, "String kernels, shier kernels and nite state automata," *Advances in Neural Information Processing Systems*, vol. 15, pp. 633–640, 2003.
- [5] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [6] M. Sewell, "The fisher kernel: a brief review," Research Note RN/11/06, 2011.
- [7] L. V. D. Maaten, "Learning discriminative shier kernels," in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, Wash, USA, 2011.
- [8] N. D. Smith and M. J. F. Gales, "Using SVMs and discriminative models for speech recognition," in *Proceedings of the IEEE International Conference on Acustics, Speech, and Signal Processing (ICASSP'02)*, pp. I/77–I/80, usa, May 2002.
- [9] N. Smith and M. Niranjan, "Data-dependent kernels in svm classification of speech patterns," in *Proceedings of the International Conference on Spoken Language Processing*, 2001.
- [10] N. Smith and M. Gales, "Speech recognition using svms," *Advances in Neural Information Processing Systems*, vol. 14, pp. 1197–1204, 2002.
- [11] P. J. Moreno and R. Rifkin, "Using the Fisher kernel method for web audio classification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00)*, pp. 2417–2420, June 2000.
- [12] A. D. Holub, M. Welling, and P. Perona, "Combining generative models and fisher kernels for object recognition," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05)*, pp. 136–143, October 2005.
- [13] J. Chen and Y. Wang, "Exploiting fisher kernels in decoding severely noisy document images," in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR'07)*, pp. 417–421, September 2007.
- [14] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, June 2007.
- [15] D. H. Hu, X. X. Zhang, J. Yin, V. W. Zheng, and Q. Yang, "Abnormal activity recognition based on HDP-HMM models," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, pp. 1715–1720, July 2009.
- [16] F. Perronnin and J. A. Rodriguez-Serrano, "Fisher kernels for handwritten word-spotting," in *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR'09)*, pp. 106–110, July 2009.
- [17] N. Bouguila, "Shape modeling and categorization using fisher kernels," in *Proceedings of the International Conference on Multimedia Computing and Systems (ICMCS'11)*, April 2011.
- [18] O. Aran and L. Akarun, "A multi-class classification strategy for Fisher scores: application to signer independent sign language recognition," *Pattern Recognition*, vol. 43, no. 5, pp. 1776–1788, 2010.
- [19] F. Perronnin, J. Snchez, and T. Mensink, "Improving the shier kernel for large-scale image classification," in *Proceedings of the European Conference on Computer Vision (ECCV'10)*, 2010.
- [20] N. Bouguila and O. Amayri, "A discrete mixture-based kernel for SVMs: application to spam and image categorization," *Information Processing and Management*, vol. 45, no. 6, pp. 631–642, 2009.
- [21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2nd edition, 2000.
- [22] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines : and other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
- [23] V. Wan and S. Renals, "Speaker verification using sequence discriminant support vector machines," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 2, pp. 203–210, 2005.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977, With discussion.
- [25] F. Dellart, "The expectation maximization algorithm," Tech. Rep. GIT-GVU-02-20, College of Computing, Georgia Institute of Technology, 2002.
- [26] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [27] "Toronto stock exchange," <http://www.tmx.com/>.
- [28] "New york stock exchange," <http://www.nyse.com/>.
- [29] "National association of securities dealers automated quotations," <http://www.nasdaq.com/>.
- [30] "Thomson reuters datastream," <http://online.thomsonreuters.com/datastream/>.

- [31] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *The Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
- [32] Brorsen, B. Wade, and S. R. Yang, "Nonlinear dynamics and the distribution of daily stock index returns," *Journal of Financial Research*, vol. 17, no. 2, pp. 187–203, 1994.
- [33] M. Staunton, *Deriving Black-Scholes from Lognormal Asset Returns*, Willmott, 2002.
- [34] J. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," Tech. Rep. MSR-TR-98-14, Microsoft Research, 1998.
- [35] R. M. Rangayyan, S. Krishnan, G. D. Bell, C. B. Frank, and K. O. Ladly, "Parametric representation and screening of knee joint vibroarthrographic signals," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 11, pp. 1068–1074, 1997.
- [36] Y. Athavale, S. Krishnan, P. Hosseinizadeh, and A. Guergachi, "Identifying the potential for failure of businesses in the technology, pharmaceutical and banking sectors using kernel-based machine learning methods," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*, pp. 1073–1077, October 2009.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

