Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

A New Application Based on GPLVM, LMNN, and NCA for Early Detection of the Stomach Cancer

Sevcan Aytaç Korkmaz & Furkan Esmeray

To cite this article: Sevcan Aytaç Korkmaz & Furkan Esmeray (2018) A New Application Based on GPLVM, LMNN, and NCA for Early Detection of the Stomach Cancer, Applied Artificial Intelligence, 32:6, 541-557, DOI: 10.1080/08839514.2018.1464285

To link to this article: https://doi.org/10.1080/08839514.2018.1464285

Published online: 27 Apr 2018.

APPLIED ARTIFICIAL INTELLIGENCE
2018, VOL. 32, NO. 6, 541–557
https://doi.org/10.1080/08839514.2018.1464285

Taylor & Francis
Taylor & Francis Group

A New Application Based on GPLVM, LMNN, and NCA for Early Detection of the Stomach Cancer

Sevcan Aytaç Korkmaz[a] and Furkan Esmeray[b]

[a]Electronic and Automation Department, Fırat University, Elazığ, Turkey; [b]Electric and Energy Department, Munzur University, Tunceli, Turkey

ABSTRACT

In this article, speeded-up robust features (SURF) for each image have been calculated. Discrete Fourier transform (DFT) method has been applied to these SURF. High dimensions of these SURF–DFT feature vectors are reduced to low dimensions with large-margin nearest neighbor (LMNN), Gaussian process latent variable models (GPLVM), and neighborhood component analysis (NCA). When size reduction process was done, effect on the GPLVM, LMNN, and NCA of the 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 feature numbers has been examined. These features are classified by naive Bayes (NB) classifier. Thus, SURF_DFT_GPLVM_NB, SURF_DFT_NCA_NB, and SURF_DFT_LMNN_NB methods for gastric histopathological images have been developed. Classification results obtained with these methods have been compared. According to the obtained results, the highest classification result was obtained as 90.24% by using 4 features by SURF_DFT_GPLVM_NB method for second group images.

Introduction

Stomach cancer is a type of cancer that spreads rapidly. And stomach cancer is a type of cancer that, when diagnosed late, causes death of the patients (Korkmaz et al. 2017b; Tannapfel, Schmelzer, and Benicke et al. 2001). Stomach cancer usually begins with gast`ritis ulcer and gastritis complaints. Stomach cancer can affect other peripheral organs and lymph nodes (Korkmaz et al. 2017b; Onishi, Takiguchi, and Ariki 2008). Also, stomach cancer may occur on the stomach tissue, and stomach wall.As a result of the Ministry of Health's work has been suggested that the second most common cancer type is stomach cancer. The most important factor that should be done first to provide early diagnosis of this disease is endoscopy.The stomach is examined by endoscopy. and then biopsy samples are taken and pathological examination is started. However, the patient is diagnosed as a result of histopathological examination. Early diagnosis is crucial for survival of the patient. Because if the cancer in the person is not diagnosed early, it becomes difficult for the

CONTACT Sevcan Aytaç Korkmaz sevcanaytackorkmaz@gmail.com Electronic and Automation Department, Fırat University, Elazığ, Turkey

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

patient to respond positively to the treatment applied by the doctors (Korkmaz et al. 2017b; Ural A. B., 2016; Lambert et al. 2002). Far East countries such as Japan and China are the countries where this disease is most seen. In Japan, about 30% of cancer diseases are gastric cancer patients. In America, the number of people who suffer from stomach cancer has increased every year (Brenner, Rothenbacher, and Arndt 2009; Hirayama et al. 2009; Ural A. B., 2016; Korkmaz et al. 2017b). According to world-wide research, to 26% of men and 11% of women have been diagnosed the gastric cancer in the last 10 years. Stomach cancer is located in the 2nd place after lung cancer in males and 3rd place after breast and lung cancer in females. The statistical studies carried out in our country suggest that the number of new gastric cancer patients is around 30,000 per year (Fujioka et al. 2004; Ural A. B. 2016). S. Yoshihiro and his colleagues (Yoshihiro et al. 2010) have developed a computer-based algorithm that predicts the risk factors for stomach cancer. In the algorithm, endoscopy images of the H. pylori bacteremia patients have been used. 3 parameters have been used to classify the gastric mucosa. The data obtained has been classified with Bayes theorem. D. Ahmadzadeh et al.(Ahmadzadeh 2013) have developed a stomach cancer diagnosis method using a local pattern algorithm and SVM (Support Vector Machine). 4 steps have been followed in developed system. These steps is the noise reduction, feature extraction, feature identification, and classification. 55 volunteer patients have been randomly chosened. Classification rate has been found as 91.8%. The motivation of this work is that both systems are a method that helps the expert doctor, save time and Money. Akbari et al. (Akbari et al. 2011) have performed a stomach cancer recognition procedure using an infrared ultra 3 spectral imaging technique. This study has been conducted by selecting 10 patients with gastric cancer. Spectral characteristics of cancerous and normal tissues have been obtained. And the necessary comparison has been made with the SVM method and the determination of cancerous regions by spectral diagram. In one study (Ural A. B., 2016), the detection of cancer sites in 25 patients from 30 patients was successfully performed. Fast Fourier Transformation (FFT) and Log transformation methods were used together in the stomach cancer recognition system. With this system, 83,3119% success has been achieved (Korkmaz et al. 2017b).

Methods of the study are explained in Section 2. Experiment results are given in Section 3. Discussions are done in Section 4, and conclusions are described in Section 5.

The purpose of this article is to add a different study to the literature to help early diagnosis of stomach cancer using images of histopathology. When studies in the literature are examined, it is seen that speeded-up robust features (SURF), DFT, GPLVM, LMNN, and NCA and NB methods are not used together to help diagnose early stomach cancer. Therefore, in this article have been developed a computer-aided algorithms to help diagnose early stomach cancer using in together the these methods. Also, the

classification performances of these algorithms have been compared according to the number of selected SURF.

Theory and method

In this article, gastric images in the two groups taken with the help of a light microscope from Fırat University are used. The number of the first group images is 180. The number of the second group images is 720. In this article, SURF have been used. The dimensions of these features have been reduced to lower dimensions with the help of the DFT-GPLVM, DFT-LMNN, and DFT-NCA methods. These lower size features have been classified by naive Bayes (NB) method. The steps of this article are shown in Figure 1.

Speeded-up robust features (SURF)

SURF is based on the sum of two-dimensional Haar wavelets and uses integral images. SURF uses Hessian matrices to determine the image attributes. Approximate determinants of the Hessian blob perceptors are calculated with the use of Haar wavelets. One of the important features that influence the performance of the SURF method is the use of integral images (Teke and Temizel 2010a; Viola and Jones 2001). Integral images speed the calculation of a given rectangular area. The area under the coordinates of the images is calculated by Equation (1) (Viola and Jones 2001):

$$I\sum(x,y) = \sum_{\substack{i\le x\\ i=0}}\sum_{\substack{j\le y\\ 0=j}} I(i,j) \tag{1}$$

When an integral image is calculated for the I image, the sum of the densities of any pixel coordinates is calculated with the three collecting process. The processing time reduces because the calculation process is independent of the size of the images (Alparslan 2013; Teke and Temizel 2010b).

Hessian matrix

The SURF blob sensor is based on the determinant of the Hessian matrix. The Hessian matrix is used to detect position blob type constants where the determinant is maximum. The Hessian matrix for the $I(x,y)$ image point is defined by Equation (2).

$$H(I(x,y)) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial y \partial x} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{2}$$

The determinant of the Hessian matrix is calculated by Equation (3):

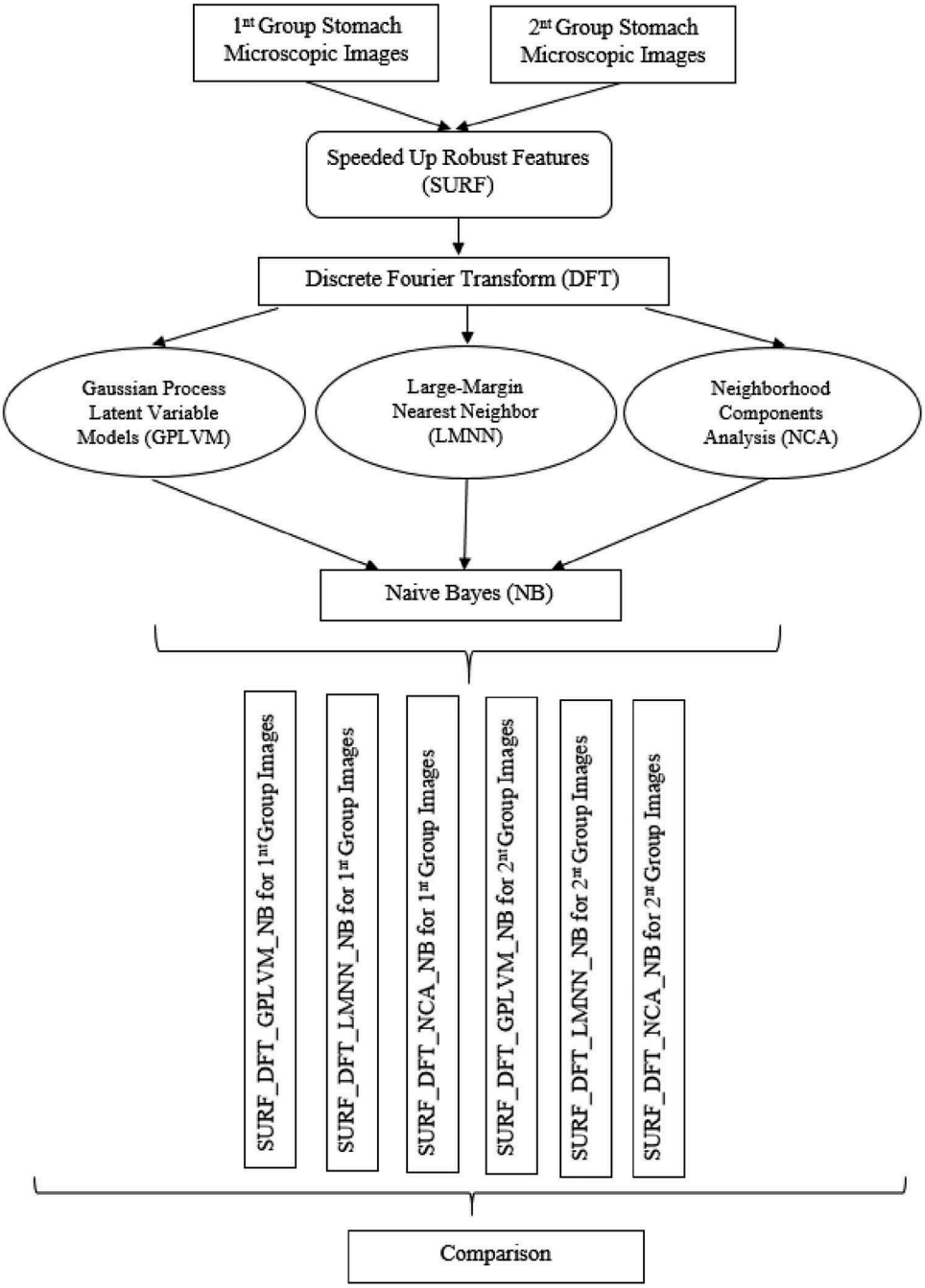

Figure 1. Steps to apply the SURF_DFT_LMNN_NB, SURF_DFT_GPLVM_NB, and SURF_DFT_NCA_NB algorithms.

$$\det(H(I(x,y))) = \frac{\partial^2 I}{\partial x^2}\frac{\partial^2 I}{\partial y^2} - \left(\frac{\partial^2 I}{\partial x \partial y}\right)^2 \tag{3}$$

The value of the determinant is used to determine the maximum and minimum points of the function by performing a second-degree function test. For the point $P(x,y)$ located at σ, the Hessian matrix is calculated as follows:

$$H(P,\sigma) = \begin{bmatrix} L_{xx}(P,\sigma) & L_{xy}(P,\sigma) \\ L_{xy}(P,\sigma) & L_{yy}(P,\sigma) \end{bmatrix} \tag{4}$$

$L_{xx}(P,\sigma)$ is the evolution at the point defined by $P(x,y)$ of the second-order Gaussian derivative $\frac{\partial^2 g(\sigma)}{\partial x^2}$. Similarly, L_{xx} and L_{xy} values are obtained. These derivatives are known as the Gauss Laplace operator (Laplacian of Gaussians). The formula for calculating the determinant of the Hessian matrix proposed by SURF (Alparslan 2013; Bellman 1961) is defined by Equation (5):

$$\det(H) \cong D_{xx}.D_{yy} - \left(wD_{xy}\right)^2 \tag{5}$$

The value w is calculated by the energy conversion in between the Gaussian kernels.

Scale space creation

The desired pairs of images may be at different scale ratios, or some of the attributes may be obtained at different scales. In computer vision applications, the scale spaces are obtained by folding with the original image Gaussian nuclei, and then the subsample is obtained by cyclic processing. SIFT successfully calculates Gaussian differences on this count (Lawrence 2004). Since the SURF method uses fixed box filters, the same filter is not applied to the previously filtered layer. As a result, an overall performance gain is achieved (Alparslan 2013; Teke and Temizel 2010a).

Orientation determination

For each pair of points of interest, a recursive orientation vector is assigned to provide rotation independence for matching of the images. Assuming that a point of interest is located at s, Haar wavelets for the $4s$ dimension are computed for the pixel located in the vicinity of $6s$ of the point of interest. Wavelet responses are weighted by a Gaussian $(\sigma = 2s)$ and are represented as points of space centered on the point of interest. The dominant orientation of responses is computed by a sliding window of $\frac{\pi}{3}$ size. The longest orientation vector is selected as the descriptor dominant orientation (Alparslan 2013; Baker, Nayar, and Murase 1998).

Descriptive components

For a point of interest found in s s scale at the point $P(x, y)$, a square region with an edge length of $20s$ is first created. The generated region is divided to 4×4 square subfields. Each subdomain is assigned four components. These four components are sampled at 5×5 equal intervals with Haar wavelet responses ($2s$ length) (Alparslan 2013). The total response of 25 points sampled for dx and dy of Haar wavelet components in directions x and y is found by Equation (6) (Alparslan 2013).

$$vsub = \left[\sum dx, \sum dy, \sum |dx|, \sum |dy|\right] \tag{6}$$

Discrete Fourier transform (DFT)

Fourier transformation is a mathematical method that allows a periodic signal to be expressed by sinus and cosine components at different frequencies (Cooley and Tukey 1965). The Fourier transformation is defined for an array of infinite lengths, and more importantly, a function of the w angular frequency, which is a continuous variable. When using MATLAB, we do not limit arrays and we need to evaluate for a limited number of points. Discrete Fourier transform (DFT) eliminates these problems. In this article, a DFT of multidimensional SURF property values is calculated using a fast Fourier transform (FFT) algorithm. The DFT is expressed as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk} \;\; 0 \leq k \leq N-1 \tag{7}$$

where $W_N = e^{-j\left(\frac{2\pi}{N}\right)}$. If the N sequence length is large, the direct presence of the DFT requires a large amount of processing. That is, as the N number increases, the number of transactions made increases rapidly and the number of transactions goes to an unacceptable level. In 1965, Cooley and Turkey developed a procedure to reduce the amount of processing required for DFT (Cooley and Tukey 1965). This procedure caused a sudden increase in DFT applications in digital signal processing and other fields. It has also been a pivotal step in the development of other algorithms. All these algorithms are known as FFT algorithms. These algorithms have greatly reduced the number of operations required for the DFT account, thereby ensuring ease of operation. FFT is an efficient and economical algorithm for DFT computation (http://web.itu.edu.tr/~baykut/lab/pdf/Deney_3.pdf).

Gaussian process latent variable models (GPLVM)

Gaussian process latent variable model (GPLVM) is the one of the dimensional reduction methods. GPLVM is based on the NxN inner product matrix of X. In GPLVM method have been focused to "RBF kernel" by considering a prior which allows for nonlinear processes (Lawrence 2004). "RBF kernel" is calculated as follow

$$k_{n,a} = \alpha exp(-.\frac{\gamma}{2}(x_n\text{-}x_a)^{\mathrm{d}}(x_n\text{-}x_a)) + \delta_{n,a}\,\beta^{-1} \tag{8}$$

where $k_{n,}$ is the element of n^{th} row and a^{th} column of K, γ is a scale parameter, and δ_n denotes the Kronecker delta. Gradients of the latent points can be found by Equation (9):

$$\partial L/\partial K = K^{-1}YY^{d}K^{-1} - DK^{-1} \tag{9}$$

Gradients of the latent points have been found as $\frac{\partial K}{\partial n,j}$ via the chain rule. These gradients may be used in combination with a nonlinear optimizer such as scaled conjugate gradients (SCGs) (Nabney 2002) to obtain a latent variable representation of the data. Furthermore, gradients with respect to the parameters of the kernel matrix may be computed and used to jointly optimize X,α, γ, and β. The solution for X will not naturally be unique; even for the linear event described above, the solution is a random transformation, where we can expect a multiple local minimum (Lawrence 2004).

Neighborhood component analysis (NCA)

NCA was acquainted by Equation (10), and here we explain the details of this method (Singh-Miller, Collins, and Hazen 2007). The NCA learns to flexibly reflect vectors in a field that optimizes a criterion for dropout-validation accuracy of the nearest neighbor classifier in a training set. In particular, NCA receives a training set entry containing $\{a_1,a_2,\ldots,a_n\}$ vectors, where $a_i \epsilon R^p$ and a set of associated labels are $\{b_1,b_2,\ldots,b_n\}$ where $b_i \epsilon L$. For example, in our experiments, a_i consists of the combined vortices of the Mel-frequency cepstrum coefficients (MFCC) measurements and specifies the class of the phonetic event identified by the vector, such as b_i. The method then learns a projection matrix H of size $k\ x\ p$ that reflects the training vectors a_i of the nearest neighbor classifier Z_i'=Ha_i that are effective in distinguishing between classes. This projection matrix H describes a Mahalanobis distance metric that can be used by the nearest neighbor classifier in the planned area.

$$d(a_i,\ a_j) = \left(Ha_i - Ha_j\right)^T\left(Ha_i - Ha_j\right) \tag{10}$$

We can learn to recalculate the original acoustic vectors in a lower dimension by choosing $k < p$. The goal of the method is to learn a projection H that raises the closeness of the nearest neighbor classifier. To describe a differentiable optimization criterion, the method assigns "soft neighbors" directly to the d nearest neighbors. In particular, every j-point in the training set has the possibility k_{ij} to assign the label to an i-point that becomes corrupted as the distance between i and j points increases.

$$k_{ij} = \frac{exp\left(-||Ha_i - Ha_j||\right)^2}{\sum_{b \neq i} exp\left(-||Ha_i - Ha_j||\right)^2} \tag{11}$$

The method tries to maximize the number of points expected to be correctly classified in the one-time-drop setting on the training set. This optimization criterion can be defined using the soft-neighbor assignments. First, an amount k_i is defined that indicates the probability that a i point is assigned to the correct class tag.

$$k_i = \sum_{j \in B_i} k_{ij} \; B_i = \{\mathrm{j} | b_j = b_i\} \tag{12}$$

The final optimization criterion $f(H)$ can then be defined simply as the sum of the probabilities of classifying each point correctly.

$$f(H) = \sum_i k_i \tag{13}$$

This criterion gives rise to a gradient rule that can be used to optimize the matrix H. (Note that a_{ij} is short hand for $a_i - a_j$.)

$$\frac{\partial f}{\partial H} = 2\mathrm{H} \sum_i \left(k_i \sum_d k_{id} a_{id} a_{id}^T - \sum_{j \in B_i} k_{ij} a_{ij} a_{ij}^T\right) \tag{14}$$

This function can be optimized using a series of degrade methods (Singh-Miller, Collins, and Hazen 2007).

Large-margin nearest neighbor (LMNN)

Mahalanobis distance metric parameterized by is learned with LMNN. So, given a set of feature vectors $a_i \in R^D$ (i = 1,...,n) along with their labels b_i and their target (right) neighbours, i.e.

$$D\left(a_i a_j\right) = ||\, L\left(a_i - a_j\right)||^2 = \left(a_i - a_j\right)^T L^T L \left(a_i - a_j\right) \tag{15}$$

which encourages the classification of the right kNN. In particular, the L parameter must minimize the distance between the pairings of a vector to its K neighbors. $N_{ij} \in \{0, 1\}$ specifies whether a_i belonging to a_j is a target neighbor. Thus,

$$\Psi_1(L) = \sum_{ij} N_{ij} ||L(a_i - a_j)||^2 \tag{16}$$

should be selected to reduce the most. Moreover, the distance of the a_i from a target neighbor a_j should be less than the distance from a fraudulent (false) neighbor a_l. For this purpose, Weinberger et al. (Kumar, Torr, and Zisserman 2007; Weinberger and Saul 2009) suggest minimizing the sum of the standard hinge loss over triplest of input vectors, i.e.

$$\Psi_2(L) = \sum_{ijl} \vartheta_{ijl}[1 + ||L(a_i - a_j)||^2 - ||L(a_i - a_l)||^2] \tag{17}$$

where $\vartheta_{ijl} = N_{ij}\delta(b_i \neq b_l)$ ($\delta(.)$ is 1 if and only if the argument is true) and the function $[m]_+ = max\{m, 0\}$. The hinge loss will be zero only when $D(a_i, a_l) \geq 1 + D(a_i, a_j)$, i.e., when there is a margin of width 1 between target and swindler neighbors. The total cost function is given by $\Psi(L) = \Psi_1(L) + \lambda_h \Psi_2(L)$ where $\lambda_h \geq 0$ is some constant. Since $M{=}L^T L \geq 0$, the minimization of the cost function $\Psi(L)$ can be formulated as a convex SDP (Kumar, Torr, and Zisserman 2007; Weinberger and Saul 2009), i.e.,

$$\begin{aligned} \min \sum_{ij} N_{ij} d_{ij} + \lambda_h \sum_{ijl} \vartheta_{ijl} \xi_{ijl},\ & (a_i - a_j)^T M (a_i - a_j) = d_{ij}, \\ & (a_i - a_l)^T M (a_i - a_l) - d_{ij} \geq 1 - \xi_{ijl} \\ & M \geq 0, \xi_{ijl} \geq 0, \forall i, j, l \end{aligned} \tag{18}$$

The variables d_{ij} and ξ_{ijl} represent the distance and hinge loss between a_i and a_j, respectively.

Note that the above problem is convex with respect to the variable M, but not according to the L variable. So the methods described in (Bar-Hillel et al. 2005; Goldberger et al. 2005; Holub, Liu, and Perona 2007; Kumar, Torr, and Zisserman 2007; Yang et al. 2006) (which make L the best) tend to be local minimum.

Naïve Bayes (NB)

A probabilistical classifier based on the Bayes theorem and based on independent assumptions is the NB classifier. A classifier method that specifies a random variation that shows the relationship between conditional probabilities and marginal probabilities is called Bayes theorem (McCallum and Nigam 1998).

$$P(X \backslash Y) = \frac{P(Y \backslash A) P(X)}{P(Y)} \tag{19}$$

$P(X \backslash Y)$ is the probability of occurrence of the X event in the case of Y event as to be in Equation (19). The likelihood of the Y event occurring when the X event occurs is $P(Y \backslash A)$. The prior probability values of X and Y events are

the $P(X)$ and $P(Y)$ probabilities (Nabney 2002). Many attributes and a target variable create the classification process.

$$p(T\backslash K_1 \ldots..K_k) = \frac{P(T)p(K_1 \ldots\ldots K_n\backslash T)}{p(K_1 \ldots..K_n)} \tag{20}$$

T represents the given target and K properties. NB classifier is generally the product of all conditional probabilities (McCallum and Nigam 1998; Yongkui et al. 2014).

Experimental results and discussion

SURF of gastric microscopy images have been obtained. The sizes of the SURF for each image are very high and have been obtained differently. SURF feature sizes obtained from normal, benign and malign stomach images are different from each other. And the sizes of the obtained SURF features are very high. By applying the DFT method to the SURF obtained from this aspect, the cells obtained for the SURF are converted into vectors, and the same dimensional values are obtained for all the microscopic images. As shown in Table 1, dimensions of the SURF obtained for 180 microscope image are different from each other. We have applied the DFT method to these obtained feature values because the numbers of these features are different for each microscope image. When the DFT method is applied, the feature size obtained for normal images is 60 × 147,712. It is 60 × 286,592 for benign images and it is 60 × 542,464 for malign images. As shown in Table 2, dimensions of the SURF obtained for 720 microscope image are different from each other. We have applied the DFT method to these obtained feature values because the numbers of these features are different for each microscope image. When the DFT method is applied, the feature size obtained for normal images is 240 × 40,064. It is 240 × 99,328 for benign images, it is 240 × 201,600 for malign images. Finally, GPLVM, LMNN, and NCA size

Table 1. The examples ın the different dimensıonal found for SURF of the 180 microscope images.

SURF of the normal microscopic images	SURF of the benign microscopic images	SURF of the malign microscopic images
39552x1	140288x1	76416x1
31680x1	40896x1	79488x1
53184x1	56704x1	88576x1
54208x1	40640x1	91776x1
43072x1	68672x1	84480x1
35776x1	101632x1	74752x1
48832x1	79872x1	67072x1
52416x1	41728x1	56640x1
50752x1	31488x1	58048x1
34816x1	36352x1	64192x1
44864x1	34240x1	80896x1
31936x1	73280x1	175616x1
60864x1	31296x1	93376x1
69184x1	30272x1	104512x1
62400x1	39936x1	114304x1
47168x1	30016x1	121280x1

Table 2. The examples ın the different dimensıonal found for SURF of the 720 microscope images.

SURF of the normal microscopic images	SURF of the benign microscopic images	SURF of the malign microscopic images
7296x1	11520x1	20736x1
7232x1	7680x1	12800x1
5952x1	12352x1	22144x1
8896x1	6912x1	18432x1
14592x1	15232x1	13120x1
7872x1	12864x1	16640x1
15872x1	9600x1	20800x1
10368x1	14976x1	16704x1
13568x1	5952x1	15680x1
11328x1	6656x1	31360x1
9792x1	13120x1	28992x1
14144x1	11072x1	36416x1
7552x1	24000x1	22336x1
15040x1	22464x1	46720x1
8256x1	6464x1	14208x1
9152x1	10496x1	21312x1

reduction methods have been applied to feature values of these images. Feature values of 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 for GPLVM, LMNN, and NCA size reduction methods were discussed, respectively. In the first group microscopic images, 30 normal + 30 benign + 30 malign microscopic images have been used for the testing purposes, while 30 normal + 30 benign + 30 malign have been used for the training purposes. In the second group microscopic images, 120 normal + 120 benign + 120 malign microscopic images have been used for the testing purposes, while 120 normal + 120 benign + 120 malign microscopic images have been used for the training purposes. The test microscopic images have been classified with the NB classifier. According to the feature number which is selected between 1 and 10 as shown in Table 3, the classification accuracy rate obtained with the NB classifier is shown. Three important results according to Table 3 have been obtained. First, the highest classification accuracy rate obtained with GPLVM is found as 87.77% for 180 microscopic images. This ratio was obtained by selecting 4 features for 180 microscopic images. Second, classification

Table 3. Classification accuracy rate according to LMNN, GPLVM, and NCA methods for 180 microscopic images.

	Classification accuracy rate (CAR) for 180 microscopic images		
Selected feature number (SFN)	SURF_DFT_LMNN_NB	SURF_DFT_GPLVM_NB	SURF_DFT_NCA_NB
SFN-1	68.88	56.66	50.00
SFN-2	68.88	75.55	54.44
SFN-3	68.88	86.66	55.55
SFN-4	68.88	87.77	71.11
SFN-5	68.88	81.11	70.00
SFN-6	68.88	86.66	65.55
SFN-7	68.88	84.44	70.00
SFN-8	68.88	82.22	81.11
SFN-9	68.88	83.33	77.77
SFN-10	68.88	83.33	78.88

accuracy rate obtained with LMNN is found as 68.88% by selecting all features among 1–10 for 180 microscopic images. Third, classification accuracy rate obtained with NCA is found as 81.11% by selecting 8 features for 180 microscopic images. Three important results according to Table 4 have been obtained. First, the highest classification accuracy rate obtained with GPLVM is found as 90.27% for 720 microscopic images. This ratio was obtained by selecting 4 features for 720 microscopic images. Second, classification accuracy rate obtained with LMNN is found as 71.94% by selecting all features among 1–10 for 720 microscopic images. Third, classification accuracy rate obtained with NCA is found as 79.16% by selecting 9 features for 720 microscopic images. Based on these results, the GPLVM size reduction methods over the SURF_DFT property values have showed higher classification performance than the LMNN and NCA size reduction methods. In Figure 2, Figure 3, and Figure 4, classification accuracy rate obtained for 180 and 720 microscopic images have been comprised for each selected feature number, respectively. According to Figure 2, using the SURF_DFT_LMNN_NB algorithm, the success rate obtained for 180 image sets is lower than the classification success rate obtained for 720 image sets. As the number of images increases, the success rate achieved with LMNN has increased. In Figure 2, effect on LMNN of the image sets for SURF_DFT_LMNN_NB algorithm has been compared. In Figure 3, effect on GPLVM of the image sets for SURF_DFT_GPLVM_NB algorithm has been compared. Generally, according to Figure 3, using the SURF_DFT_GPLVM_NB algorithm, the success rate obtained for 720 image sets is lower than the classification success rate obtained for 180 image sets. But, using the SURF_DFT_GPLVM_NB algorithm in selected 1, 4, and 5 features, the success rate obtained for 180 image sets is lower than the classification success rate obtained for 720 image sets. Moreover, the highest classification accuracy rate obtained with GPLVM is found as 90.27% by selecting 4 features for 720 microscopic image sets. In Figure 4, effect on NCA of the image sets for SURF_DFT_NCA_NB algorithm has been compared. Generally, according to Figure 4, using the SURF_DFT_NCA_NB algorithm, the success rate

Table 4. Classification accuracy rate according to LMNN, GPLVM, and NCA methods for 720 microscopic images.

	Classification accuracy rate (CAR) for 720 microscopic images		
Selected feature number (SFN)	SURF_DFT_LMNN_NB	SURF_DFT_GPLVM_NB	SURF_DFT_NCA_NB
SFN-1	71.94	62.50	60.00
SFN-2	71.94	71.11	62.50
SFN-3	71.94	86.38	71.11
SFN-4	71.94	90.27	68.88
SFN-5	71.94	81.94	75.00
SFN-6	71.94	82.50	66.11
SFN-7	71.94	81.94	66.11
SFN-8	71.94	77.22	72.77
SFN-9	71.94	76.66	79.16
SFN-10	71.94	75.83	78.56

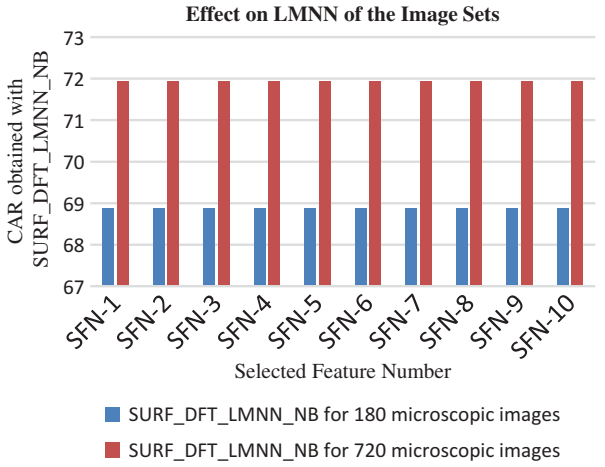

Figure 2. Comparison of the effect on LMNN of the image dataset numbers for SURF_DFT_LMNN_NB algorithm.

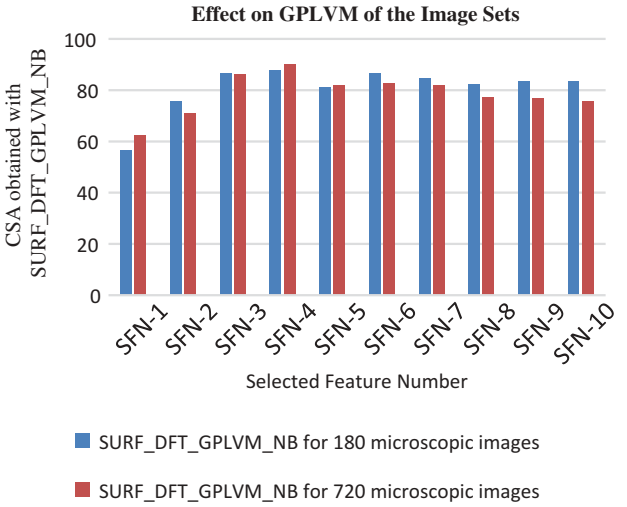

Figure 3. Comparison of the effect on GPLVM of the image sets for SURF_DFT_GPLVM_NB algorithm.

obtained for 180 image sets is lower than the classification success rate obtained for 720 image sets. But, using the SURF_DFT_GPLVM_NB algorithm in selected 4, 7, 8, and 10 features, the success rate obtained for 180 image sets is higher than the classification success rate obtained for 720 image sets. Moreover, the highest classification accuracy rate obtained with GPLVM in Figure 4 is found as 81.11% by selecting 8 features for 180 microscopic image sets. In Table 5, GLCM is gray-level co-occurrence matrix features. LBP is local binary pattern features. LPP is locality preserving projections for dimensional reduction. HOG is histograms of oriented gradient feature. LDA is linear discriminant analysis. ANN is artificial neural network. In Table 5, HOG_LDA_ANN accuracy rate was found as 88.9% (Korkmaz et al. 2017b). GLCM_LPP_ANN accuracy rate was found

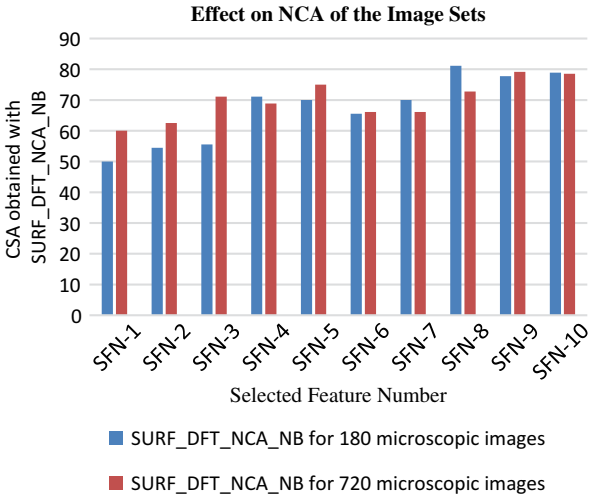

Figure 4. Comparison of the effect on NCA of the image sets for SURF_DFT_NCA_NB algorithm.

Table 5. Comparison of different results.

Compared	
Method	Accuracy (%)
Our Method SURF_DFT_GPLVM_NB for 720 Images	90.27
Our Method SURF_DFT_GPLVM_NB for 180 Images	87.77
HOG_LDA_ANN (Korkmaz et al. 2017b)	88.9
GLCM_LPP_ANN (Korkmaz et al. 2017b)	80.12
LBP_LPP_ANN (Nabney 2002)	85.56

as 80.12% (Korkmaz et al. 2017b). Accuracy rate with our method SURF_DFT_GPLVM_NB has been found as 90.27% for 720 image set. Accuracy rate with our method SURF_DFT_GPLVM_NB has been found as 87.77% for 180 image set.

Conclusion

Until today, many classification methods have been used in the field of health (Korkmaz 2018a, 2018b; Korkmaz et al. 2017a; Korkmaz and Binol 2018; Korkmaz and Eren 2013; Korkmaz and Korkmaz 2015; Korkmaz, Korkmaz, and Poyraz 2016; Korkmaz et al. 2016; Korkmaz and Poyraz 2014, 2015; Sengur 2012, 2008, 2009; Sengur, Turkoglu, and Ince 2007). 180 gastric images taken with the help of the light microscope in this article have been classified. In this article, gastric images in the second group taken with the help of a light microscope are used. The number of the first group images is 180. The number of the second group images is 720. SURF for each images have been calculated. DFT methods have been applied to these SURF. High dimensions of these SURF-DFT feature vectors are reduced to low dimensions with LMNN, GPLVM, and NCA. When size reduction process was done, effect on the GPLVM, LMNN, and NCA of the 1, 2, 3, 4, 5, 6, 7, 8, 9,

and 10 feature numbers has been examined. These features are classified by NB classifier. Thus, SURF_DFT_GPLVM_NB, SURF_DFT_NCA_NB, and SURF_DFT_LMNN_NB methods for gastric histopathological images have been developed. Classification results obtained with these methods have been compared. Also, classification results have been compared for the number of the first group and second group images. According to the obtained results, the highest classification result was obtained as 90.24% by using 4 features by SURF_DFT_GPLVM_NB method for second group images. Success rate obtained with these methods has been compared with other classification results in the literature. According to the other methods, our success rate for stomach microscopic images has been seen to be higher. In future studies, an analysis will be performed by applying different feature extraction methods to different cancer images.

Acknowledgments

The authors would like to thank Prof. Dr. İbrahim Hanifi ÖZERCAN in the pathology department of the Fırat University Hospital.

Conflicts of interest

Sevcan Aytaç Korkmaz declares that she has no conflict of interest. Furkan Esmeray declares that he has no conflict of interest.

References

http://web.itu.edu.tr/~baykut/lab/pdf/Deney_3.pdf

Ahmadzadeh, D., M. Fiuzy, and J. Haddadnia. 2013. Stomach cancer diagnosis by using a combination of image processing algorithms, local binary pattern algorithm and support vector machine. *Journal of Basic and Applied Scientific Research* 3 (2):243–51.

Akbari, H., K. Uto, Y. Kosugi, K. Kojima, and N. Tanaka. 2011. Cancer detection using infrared hyperspectral imaging. *The Official Journal of the Japanese Cancer Association* 102 (4):852–57.

Alparslan, N. 2013. *New approaches to gradient based heterogeneous feature extraction methods*. Turkey: Institute of Science, İnönü Üniversity.

Baker, S., S. K. Nayar, and H. Murase. 1998. Parametric feature detection. *International Journal of Computer Vision* 27 (1):27–50.

Bar-Hillel, A., Hertz T., Shental N., & Weinshall D., 2005. Learning a mahalanobis metric from equivalence constraints. Journal of Machine Learning Research, 6 (Jun): 937–965. Greg Ridgeway (ed.), MA: MIT Press, Microtome Publishing.

Bellman, R. 1961. *Adaptive control processes: A guided tour*, 255. Princeton, New Jersey, USA: Princeton University Press.

Brenner, H., Rothenbacher D., & Arndt, V., 2009. Epidemiology of stomach cancer. Mukesh Verma (ed.), In Cancer Epidemiology (pp. 467–477). USA: Humana Press, Springer.

Cooley, J. W., and J. W. Tukey. 1965. An algorithm for the machine computation of complex Fourier series, vol. 19. Mathematics of Computation. USA: American Mathematical Society.

Fujioka, N., Y. Morimoto, T. Arai, and M. Kikuchi. 2004. Discrimination between normal and malignant human Stomach tissues by Fourier transform infrared spectroscopy. *Cancer Detection and Prevention* 28 (1):32–36.

Goldberger, J., G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov. 2005. Neighbourhood components analysis. In Yair Weiss, Bernhard Schölkopf, and John C. Platt (eds.), *Advances in neural information processing systems*, 513–20. USA: NIPS Proceedings.

Hirayama, A., K. Kami, M. Sugimoto, M. Sugawara, N. Toki, H. Onozuka, … H. Esumi. 2009. Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Research* 69 (11):4918–25.

Holub, A., Y. H. Liu, and P. Perona. 2007, June. On constructing facial similarity maps. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1–8). USA: IEEE.

Korkmaz, S. A. 2018a. LBP Özelliklerine Dayanan Lokasyon Koruyan Projeksiyon (LPP) Boyut Azaltma Metodunun Farklı Sınıflandırıcılar Üzerindeki Performanslarının Karşılaştırılması. *Sakarya University Journal of Science* 22 (4):1–1.

Korkmaz, S. A. 2018b. Detecting cells using image segmentation of the cervical cancer images taken from scanning electron microscope. *The Online Journal of Science and Technology-January* 8 (1):47–53.

Korkmaz, S. A., A. Akçiçek, H. Bínol, and M. F. Korkmaz. 2017a, September. Recognition of the stomach cancer images with probabilistic HOG feature vector histograms by using HOG features. In Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on (pp. 000339–42). USA: IEEE.

Korkmaz, S. A., and H. Binol. 2018. Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. Journal of Molecular Structure, 1156, 255–263.

Korkmaz, S. A., H. Bínol, A. Akçiçek, and M. F. Korkmaz 2017b, September. A expert system for stomach cancer images with artificial neural network by using HOG features and linear discriminant analysis: HOG_LDA_ANN. In Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on (pp. 000327–32). USA: IEEE.

Korkmaz, S. A., and H. Eren 2013, December. Cancer detection in mammograms estimating feature weights via Kullback-Leibler measure. In Image and Signal Processing (CISP), 2013 6th International Congress on (Vol.2, pp. 1035–40). USA: IEEE.

Korkmaz, S. A., and M. F. Korkmaz. 2015. A new method based cancer detection in mammogram textures by finding feature weights and using Kullback–Leibler measure with kernel estimation. *Optik-International Journal for Light and Electron Optics* 126 (20):2576–83.

Korkmaz, S. A., M. F. Korkmaz, and M. Poyraz. 2016. Diagnosis of breast cancer in light microscopic and mammographic images textures using relative entropy via kernel estimation. *Medical & Biological Engineering & Computing* 54 (4):561–73.

Korkmaz, S. A., M. F. Korkmaz, M. Poyraz, and F. Yakuphanoglu. 2016. Diagnosis of breast cancer nano-biomechanics images taken from atomic force microscope. *Journal of Nanoelectronics and Optoelectronics* 11 (4):551–59.

Korkmaz, S. A., and M. Poyraz. 2014. A new method based for diagnosis of breast cancer cells from microscopic images: DWEE—JHT. *Journal of Medical Systems* 38 (9):92.

Korkmaz, S. A., and M. Poyraz. 2015. Least square support vector machine and minimum redundancy maximum relevance for diagnosis of breast cancer from breast microscopic images. *Procedia-Social and Behavioral Sciences* 174:4026–31.

Kumar, M. P., P. H. Torr, and A. Zisserman. 2007, October. An invariant large margin nearest neighbour classifier. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1–8). USA: IEEE.

Lambert, R., A. Guilloux, A. Oshima, V. Pompe-Kirn, F. Bray, M. Parkin, W. Ajiki, and H. Tsukuma. 2002. Incidence and mortality from Stomach cancer in Japan, Slovenia and the USA. *International Journal Of Cancer* 97:811–18.

Lawrence, N. D. 2004. Gaussian process latent variable models for visualisation of high dimensional data. In Yair Weiss, Bernhard Schölkopf, and John C. Platt (eds.), *Advances in neural information processing systems*, 329–36. USA: NIPS Proceedings.

McCallum, A., and K. Nigam. 1998. A comparison of event models for naive Bayes text classification. *AAAI-98 Workshop on Learning for Text Categorization* 752 (July):41–48.

Nabney, I. 2002. NETLAB: Algorithms for pattern recognition. *Springer Science & Business Media*. 6:1-59.

Onishi, K., T. Takiguchi, and Y. Ariki. 2008, December. 3D human posture estimation using the HOG features from monocular image. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1–4). USA: IEEE.

Sasaki, Y., Hada R., Yoshimura, T., Hanabata, N., Mikami, T., and Fukuda, S.. 2010. Computer-aided estimation for the risk of development of gastric cancer by image processing. In IFIP International Conference on Artificial Intelligence in Theory and Practice (pp. 197–204). Springer, Berlin, Heidelberg.

Sengur, A. 2008. An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. *Computers in Biology and Medicine* 38 (3):329–38.

Sengur, A. 2009. Multiclass least-squares support vector machines for analog modulation classification. *Expert Systems with Applications* 36 (3):6681–85.

Sengur, A. 2012. Support vector machine ensembles for intelligent diagnosis of valvular heart disease. *Journal of Medical Systems* 36 (4):2649–55.

Sengur, A., I. Turkoglu, and M. C. Ince. 2007. Wavelet packet neural networks for texture classification. *Expert Systems with Applications* 32 (2):527–33.

Singh-Miller, N., M. Collins, and T. J. Hazen. 2007. Dimensionality reduction for speech recognition using neighborhood components analysis. In Eighth Annual Conference of the International Speech Communication Association. USA: MIT.

Tannapfel, A., Schmelzer, S., Benicke, M., Klimpfinger, M., Kohlhaw, K., Mössner J., Engeland K., & Wittekind C. (2001). Expression of the p53 homologues p63 and p73 in multiple simultaneous gastric cancer. *The Journal of Pathology* 195 (2), 163–70.

Teke, M., and A. Temizel. 2010a, August. Multi-spectral satellite image registration using scale-restricted SURF. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 2310–13). USA: IEEE.

Teke, M., and A. Temizel. 2010b, April. Registration of multi-spectral satellite images with scale-restricted SURF. In Signal Processing and Communications Applications Conference (SIU), 2010 IEEE 18th (pp. 356–59). IEEE.

Ural, A. B., 2016. "Determining gastric cancer non-invasively with image processing methods", Gazi University, Institute of Science and Technology, (pp. 1–47).

Viola, P., and M. Jones. 2001. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I–I). USA: IEEE.

Weinberger, K. Q., and L. K. Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research* 10 (Feb):207–44.

Yang, L., R. Jin, R. Sukthankar, and Y. Liu. 2006, July. An efficient algorithm for local distance metric learning. In American Association for Artificial Intelligence, Vol. 2, 543–48.

Yongkui, S., L. Pengrui, W. Ying, Z. Jingyu, and L. Meijie. 2014. The prediction of the caving degree of coal seam roof based on the naive bayes classifier. *Electronic Journal of Geotechnical Engineering* 19:Z2.