Lorentzian Distance Classifier for Multiple Features

Yerzhan Kerimbekov[1] and Hasan Şakir Bilge[2]
[1]*Department of Computer Engineering, Ahmet Yesevi University, Ankara, Turkey*
[2]*Department of Electrical-Electronics Engineering, Gazi University, Ankara, Turkey*
kerimbekov@ogr.yesevi.edu.tr, bilge@gazi.edu.tr

Keywords: Classification, Lorentzian Distance, Feature Selection.

Abstract: Machine Learning is one of the frequently studied issues in the last decade. The major part of these research area is related with classification. In this study, we suggest a novel Lorentzian Distance Classifier for Multiple Features (LDCMF) method. The proposed classifier is based on the special metric of the Lorentzian space and adapted to more than two features. In order to improve the performance of Lorentzian Distance Classifier (LDC), a new Feature Selection in Lorentzian Space (FSLS) method is improved. The FSLS method selects the significant feature pair subsets by discriminative criterion which is rebuilt according to the Lorentzian metric. Also, in this study, a data compression (pre-processing) step is used that makes data suitable in Lorentzian space. Furthermore, the covariance matrix calculation in Lorentzian space is defined. The performance of the proposed classifier is tested through public GESTURE, SEEDS, TELESCOPE, WINE and WISCONSIN data sets. The experimental results show that the proposed LDCMF classifier is superior to other classical classifiers.

1 INTRODUCTION

Nowadays, machine learning techniques are used in different domains such as data mining, pattern recognition, image processing and artificial intelligence (Louridas and Ebert, 2016), (Wang et al., 2016). Generally, a machine learning algorithm has two stages: training and testing. The main purpose of machine learning is to train a computer system by studying a training samples and use it in test samples. Two Learning Strategies as supervised (classification) and unsupervised (clustering) learning are existed in literature (Bkassiny and Jayaweera, 2013). In supervised learning a training is used over the labelled data and a model is built to classify the new samples. Unsupervised learning is the clustering of unlabeled samples which have similar properties (Bkassiny and Jayaweera, 2013). One of the most solved problems in machine learning is a classification problem. As known, Bayes, k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) classifiers are the commonly used machine learning algorithms (Theodoridis and Koutroumbas, 2009).

In this study, a classification problem was investigated in Lorentzian space for data sets that have more than two features. Lorentzian space is one of the main issues of the General Relativity Theory (Kerimbekov et al., 2016). In this context, for obtaining the best classification result a feature selection method and pre-processing step were developed. As known every feature selection method needs a discriminative criterion (Theodoridis and Koutroumbas, 2009). For this purpose, in this study, unlike the criteria that commonly used in pattern recognition as Divergence, Bhattacharyya Distance, Scatter Matrix, Fisher's Discriminant Ratio (FDR) (Theodoridis and Koutroumbas, 2009), a new criterion was improved based on Lorentzian metric.

In this study, the Lorentzian metric is used for feature selection and classification. This metric is non-positive definite. The use of such a metric is an interesting contribution of our study. For two dimensional features, one of the features has a negative effect on the distance measure. This property gives us a special opportunity to increase the success rate of the classification in Lorentzian space. The statement that mentioned above gives us the idea to use the Lorentzian metric as a discriminative criterion and use it in feature selection. Thus, in this study, the new classifier for more than two features data in Lorentzian space was developed.

Kerimbekov, Y. and Bilge, H.
Lorentzian Distance Classifier for Multiple Features.
DOI: 10.5220/0006197004930501
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 493-501
ISBN: 978-989-758-222-6

2 THE SPECIAL PROPERTIES OF LORENTZIAN SPACE

The Lorentzian space is also recognized as a non-Euclidean space and known as special case of Riemannian space. Because of positive definiteness condition an inner product operation in Lorentzian space is different than the analogue in Euclidean space (Gündogan and Kecilioglu, 2006). Also, a distance between points in Lorentzian space is different from commonly used Euclidean distance. The group of points with the same distance occurs a circle in Euclidean space. However, because of the neighborhood structure dissimilarity according to Euclidean space the shape of the same distance points in Lorentzian space is different. The only way to find out the neighborhood structure in Lorentzian space is possible by clearly understanding the concept of the distance between two points in this space. In every defined space in art the metrics are existed to compute the distance between points. Thus, the distance *d* between two points (*U* and *Y*) in Lorentzian space can be computed by the following formula.

$$d(U,Y) = \sqrt{\left(\sum_{i=1}^{l-1} |u_i - y_i|^2 - |u_l - y_l|^2\right)} \quad (1)$$

where *l* is the dimension of the space (the number of features). This value also defines that the last dimension has negative signature (Kerimbekov et al., 2016).

As it can be clearly seen from (1), the Lorentzian metric has a minus sign in the second term, which corresponds to time axis. The main difference in Lorentzian metric is that the distance between two points can be zero. To demonstrate this case, the calculation of distances between two points are done according to both Lorentzian and Euclidean metrics. For this, two points: (-2, -1) and (0, 1) are selected. The places of these points visually can be seen from 2 dimensional Lorentzian space that shown in Figure 1. The first coordinate belongs to the first feature, the second one belongs to the second feature. If we accept that these points are in Euclidean space:

$$E_d = \sqrt{(-2-0)^2 + (-1-1)^2} = \sqrt{8}$$

then the distance is $\sqrt{8}$. If we accept that these points are in Lorentzian space:

$$L_d = \sqrt{(-2-0)^2 - (-1-1)^2} = \sqrt{0}$$

then the distance becomes zero according to the Formula-1.

In the Lorentzian space, the Lorentzian distance between two points over the lines parallel to cross direction with 45° degree (cone edges or cone lines or forward/backward light rays or null like lines) is zero. Thus the neighborhood is different in Euclidean and

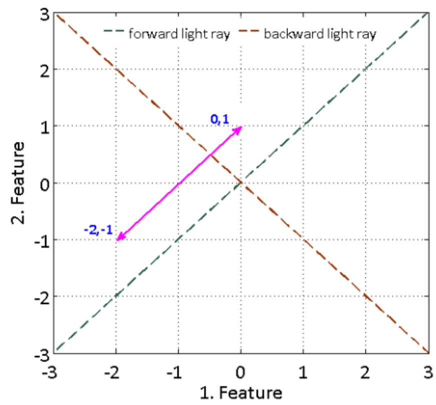

Figure 1: The difference between Euclidean and Lorentzian distances.

Lorentzian spaces. The other attribute of the Lorentzian space is the matrix multiplication operation that different than the analog in Euclidean space. Namely, for $A = [a_{ij}] \in \mathbb{R}_u^t$ and $B = [b_{jl}] \in \mathbb{R}_p^u$ matrix a multiplication operation can be calculated with the formula below:

$$A \cdot_L B = \left[\sum_{j=1}^{u} a_{ij} b_{jl} - a_{iu} b_{ul}\right] \quad (2)$$

Where, the notation '$\cdot_L$' is define the multiplication in Lorentzian space (Gündogan and Kecilioglu, 2006). For example, the multiplication of two matrix A, B in Lorentzian space with 2×2 dimensions is obtained by following expression:

$$A \cdot_L B = \begin{bmatrix} a_{11}b_{11} - a_{12}b_{21} & a_{11}b_{12} - a_{12}b_{22} \\ a_{21}b_{11} - a_{22}b_{21} & a_{21}b_{12} - a_{22}b_{22} \end{bmatrix} \quad (3)$$

3 PROPOSED METHOD

3.1 Feature Subsets and Selection

In classification problem the requested classification results can be produced in case of using the most important features from data set. The extracting or selecting the most significant features from data set is the main purpose of the Data Mining (DM) algorithms and it is also considerably decreases the computational complexity of classifier. In this study, first of all, the properties (metric) of Lorentzian space were investigated in term of selecting the best feature subsets that represent the data set ideally. Also, the diverse number of selected feature subsets were tested in obtaining better classification success rate. In our previous research we found out that the classification success rate can be increased by using less number of best feature subsets (Kerimbekov et al., 2016). Hence, in this study, from original data sets the feature pair subsets were generated according to the well know combination formula (4) which is commonly used in statistics (Brualdi, 2010). In this formula feature combination subsets are occurred by rule as one feature and other ones. For example, in three dimensional data case all feature pair subsets looks as $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$. The position and order of the feature subsets in cluster are not important. Generally, in n dimensional data set the total number feature subsets defined as $S(n,r)$ and calculated by expression below:

$$S(n,r) = \binom{n}{r} = \frac{n!\,(n-r)}{r!} \tag{4}$$

where, r is the dimension of subsets. Thus, by (4) formula we can obtain the feature pair subsets that include all features in original data set. In this study, the dimension of subsets was taken as two. Because of smallest dimension the computational complexity of classification process is acceptable. Furthermore, the two dimensional Lorentzian space classifier was introduced in our study (Kerimbekov et al., 2016) and the superiority of that algorithm was also proved. Thus, for data set with 50 features and dimension of subsets as $r = 2$ totally 1225 feature pair subsets are produced according to (4). However, as seen from this example the number of these subsets in high dimensional data set will be huger and it is costly to use all of them. Hence, in this study, we propose the novel feature selection method that selects optimal feature subsets according to Lorentzian metric.

The main aim of feature selection method is to increase the classification success rate by using less number of feature and decrease the computational complexity of classifier. The methods based on statistics like mean, variance, correlation are commonly used in pattern recognition (Theodoridis and Koutroumbas, 2009). These criteria serve in feature selection process as a determinative criterion in measuring the relation among the features and the discrimination for best or worst feature subsets is made. In Euclidean space we have the discriminative criterion J which based on within and between class scatter matrices of samples:

$$J = trace(S_w^{-1} S_m) \tag{5}$$

Where, S_w is the within class scatter matrix of M class data set. The within class scatter matrix of samples consists from multiplication of a prior probability value P_h and the covariance matrix Σ_h for μ_h class. The subtraction of feature vector x and within class mean μ_h for every w_i class from data set is established covariance matrix Σ_h. Hence, the covariance matrix Σ_h can be occurred as:

$$\Sigma_i = E[(x-\mu_h)(x-\mu_h)^T] \tag{6}$$

Thus, according to the statement mentioned above a scatter matrix of within class samples S_w takes form like:

$$S_w = \sum_{h=1}^{M} P_h \Sigma_h \tag{7}$$

The other S_m value in (5) formula is the Mixture Scatter Matrix of samples (Theodoridis and Koutroumbas, 2009). This matrix is calculated as covariance matrix of feature vector x and general mean μ_h subtraction and can be calculated by formula below:

$$S_m = E[(x-\mu_0)(x-\mu_0)^T] \tag{8}$$

The discriminative criterion J that given by (5) is valid only in Euclidean space and this criterion was restructured according to Lorentzian metric. As we can see from (7) and (8) expressions the criterion J includes the covariance matrix calculation. Furthermore, a covariance matrix is based on matrix multiplication operation. However, as explained in section II above a matrix multiplication operation in Lorentzian space is different than Euclidean analogue and dependent to rule (2). Hence, redesigning of the (7) and (8) expressions in Lorentzian space according to rule (2) gives us next formulas:

$$(\Sigma_i)_L = E[(x-\mu_h)\cdot_L (x-\mu_h)^T] \tag{9}$$

And

$$(S_m)_L = E[(x-\mu_0)\cdot_L (x-\mu_0)^T] \tag{10}$$

As a result of this restructuring the covariance matrix calculation path in Lorentzian space is suggested as (9). Thus, the novel LJ (Lorentzian J) discriminative criterion in Lorentzian space based on (9) and (10) expressions was suggested. The LJ criterion defines a significance rate of features in Lorentzian space and according to (5) can be formulated as below:

$$LJ = trace[(S_w^{-1})_L (S_m)_L] \tag{11}$$

Eventually, the new Feature Selection in Lorentzian Space (FSLS) method based on LJ discriminative criterion was proposed. The new FSLS method selects optimal feature subsets according to Lorentzian metric.

3.2 Pre-processing and Optimal Parameters

In classification problem occasionally a preprocessing step is necessarily. Because of better representing and making usable a data set this operation can enhance the classification success rate. In this study, the preprocessing step is composed only from matrix multiplication (compression) (Marcus and Minc, 1992). This transformation matrix is used with the aim to make the data meaningful in Lorentzian space. Thus, after doing compression over n-dimensional $X = (x_1, x_2, \dots, x_n)$ training set in Euclidean space it is transformed as $X' = (x_1', x_2', \dots, x_l')$ and becomes suitable for training and classification in Lorentzian space. This preprocessing step can be defined as the following expression:

$$X' = X\lambda \tag{12}$$

Where, λ is the diagonal matrix which can be expressed by $\lambda_{ij} = 0,\ \text{if } i \neq j\ \forall\ i,j \in \{1,2,\dots,n\}$. Hence, the transformation matrix that forms the preprocessing step for two dimensional data is determined as following formulas:

$$\lambda = \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix} \text{or } \lambda = \begin{pmatrix} 0 & w \\ q & 0 \end{pmatrix} \tag{13}$$

where, $w, q \in R$.

In this study, the first form of transformation matrix was used. The relation between the parameters w, q of this matrix λ is as $w = 20 * q$. Hence, the primary case is assumed as:

$$\lambda = \begin{pmatrix} 2 & 0 \\ 0 & 0.1 \end{pmatrix}$$

However, our research shows us that these parameters meanings are significant in term of classification success. Because of this the optimal meanings of parameters which produce the best classification output were also investigated in experiments.

4 LORENTZIAN CLASSIFIER

Generally, a classification process consists from training and test steps. In this study, preparing the data for training is done in two steps. First of all, the optimal feature pair subsets are selected by new proposed FSLS method. Subsequently, over these feature subsets the pre-processing operation is applied that mentioned in third section. For training of selected and transformed feature subsets the Classification via Lorentzian Metric (CLM) (Kerimbekov et al., 2016) method was improved. The classification algorithm CLM is valid in two dimensional Lorentzian space and based on Lorentzian distance. The CLM classifier assigns the class label of new sample according to Lorentzian distances that explained by formula (1). It means that, the k nearest pairs are selected by Lorentzian metric. These pairs define the relation of a test sample between k training set samples and finally the classification can be done by using the majority rule. The CLM method was described as a classifier in two dimensional Lorentzian space. However, in our research, we use the multidimensional data sets. Therefore, the CLM method was improved by adding the supplementary decision rule and hereinafter referred to as the Lorentzian Distance Classifier for Multiple Features (LDCMF).

The proposed novel LDC method is the aggregate of next stages. The novel LDC method takes as the inputs $X, Y \in \mathbb{R}$ training and test sets. However, as mentioned before, the training data sets are separated to feature pair subsets by (4). Namely, in first step from the X training set all possible $S(n, 2)$ feature pair subsets are occurred as $X' = (\{x_1, x_2\}, \{x_1, x_3\}, \dots, \{x_{n-1}, x_n\})$. Subsequently, the produced $S(n, 2)$ feature pair subsets are weighted by LJ criterion. Thereafter, the $k = (1, S)$ number X'' optimal feature pair subsets are selected by FSLS method that based on Lorentzian metric. Here, S defines the total number of feature combination (*fc*) pairs. The selected feature pair subsets are compressed by (12) formula and becomes ready for training. The new LDC classifier has iteration in length k. This value is also used as a threshold for stopping in the proposed algorithm. According to how will be defined the meaning of k less or more the computational time of proposed algorithm is changed. Furthermore, was found that the selected feature pair subsets X'' by including the efficient features represents the original data set in best way. Thus, the selected feature pair subsets X'' are used in proposed LDC classifier as training data set.

For new sample coming from Y'' test set feature selection and preprocessing step that explained before are applied as like in training samples case. Subsequently, the class labels of test samples are assigned as $c_i, i = (1, k)$. The determined c_i is the class label of i. feature pair from Y'' which respective to X''. It means that, the new proposed LDC classifier in testing stage of new coming sample is iterated k times. In every iteration the new proposed classifier produces a combined class label C_i which includes the class labels of each selected feature pairs c_i. The combined class label C_i represents one test sample and defines the class affiliation. In first step of iteration the combined class label is defined as $C_0 = [c_1]$. In the other iteration it continues as $C_i = [C_{i-1}, c_i]$. The classification ratio obtains according to majority rule. It means that, in two class sample case if the number of selected optimal feature pairs will be 3 than the proposed classifier produces class label as $C_0 = [c_1], C_1 = [c_1, c_2], C_2 = [c_1, c_2, c_3]$. All steps that mentioned before compose the new Lorentzian Distance Classifier for Multiple Features (LDCMF) method. Finally, the LDC method can be defined as Algorithm-1 in the following processes in order:

Algorithm-1. Lorentzian Distance Classifier (LDC)

Input: $X, Y \in \mathbb{R}$ training and test datasets
Step 1: Create X' fc pairs with $S(n, 2)$
Step 2: From X' select k # feature subset X'' using LJ
Step 3: Do compression $X''' = X''\lambda$
Step 4: For new sample Y from test set,
Generate Y''' and find K nearest pairs
Assign class label c_i by using the majority rule
Obtain $C_i = [C_{i-1}, c_i]$
Step 5: Compute classification rate using C_i

5 EXPERIMENTAL RESULTS AND DISCUSSIONS

5.1 Data Sets

In this study, for purpose of testing the new suggested classifier performance some public data sets were used as: GESTURE, SEEDS, TELESCOPE, WINE and WISCONSIN (Lichman, 2013). The number of features in the selected data sets varies in interval of 7-33. There is some statistical information about these data sets in Table 1. The samples in training and test set were selected randomly from original data set. In experiments the 30% of the data was used for training and the rest 70% for testing.

Table 1: Data set descriptions. (f -feature, c -class, s -sample).

	# f	# c	# s	# train s	# test s
GESTURE	18	2	448	150	298
SEEDS	7	2	140	46	94
TELESCOPE	10	2	400	134	266
WINE	13	2	130	44	86
WISCONSIN	33	2	198	66	132

5.2 Experimental Results

In this study, the new LDC classifier in Lorentzian space is suggested. This algorithm uses the optimal feature pairs which selected by FSLS method based on Lorentzian space metric. To evaluate the proposed classifier performance some public data sets as GESTURE, SEEDS, TELESCOPE, WINE and WISCONSIN were used in experiments. As clearly seen from Table 2. the number of features in these data sets are different. Hence, in experiments the number of feature subsets obtained from these data sets are also different. As we see from this statement the large number of features in data set is considerably increased the subsets number. Hence, the FSLS method in term of classification is important. Moreover, as mentioned before, the best outputs of LDC method is linked to number of selected optimal feature pair subsets. Therefore, in experiments, the meaning of k was defined as 20. Subsequently, from all feature pair subsets only 20 feature pairs were selected according to FSLS method. On the one hand, the new LDC classifier with value $k = 20$ in terms of computational complexity does not produce the perceivable difference in comparison with classic Bayes, kNN and SVM classifiers. For example, for feature pair from GESTURE data set case the classic classifiers Bayes, kNN and SVM are produced the work times as 0.0078, 0.0349 and 0.0596 second respectively. The work time of our method for the same case was produced as 0.0677 second. The computational time of our method as seen from results is little more than SVM output which is the biggest among the others. However, it can be explained by use of pre-processing step which is reported in section 3.2.

Despite the fact that the number of feature pairs for data sets are dissimilar as it has been seen from experimental results definition of k as 20 was sufficient to get the best success rate with LDC classifier. Also, it was found out that the meaning of $k_{opt} = (1, k)$ which produces the best success rate in LDC method can be less than k. The last statement enhances the proposed methods validity in terms of computational complexity and effectiveness. The numerical information about the features and feature pair subsets obtained from data sets take place in Table 2. Also, the differences between k and k_{opt} which produce the best classification outputs with proposed LDC method is given.

Table 2: feature (f), feature combination (fc), k- selected subsets, k_{opt}- optimal subsets that produce best result.

	# f	#fc	# k	# k_{opt}
GESTURE	18	153	20	20
SEEDS	7	21	20	12
TELESCOPE	10	45	20	8
WINE	13	78	20	14
WISCONSIN	33	528	20	15

As mentioned in section III the meaning of w, q parameters are important in terms of transforming the data and making them usable in Lorentzian space. In this regard, the optimal values of these parameters were found out for all data set. The meanings of parameters changes according to distribution of points in data set. The whole list of optimal parameter values obtained for data sets that produce the best classification results with proposed LDC method are took place in Table 3. below.

Table 3: The optimal parameters of compression matrix for data sets.

	w_{opt}, q_{opt}
GESTURE	0.9, 1.8
SEEDS	2, 1.4
TELESCOPE	1.9, 1.8
WINE	0.9, 1.9
WISCONSIN	2, 1.8

The performance of new LDC classifier over all data set was evaluated by comparing the classification results with Bayes, kNN and SVM classifiers outputs. For classic classifiers the Euclidean analogue of proposed feature selection method was used. It means that except the compression of data set which is explained in the section 3.2. and special for Lorentzian space the other steps of proposed algorithm are common for classic classifiers. It was made with the aim of to keep the experiment path similar and meaningful in term of comparison the classification results. Also, in experiments the classic classifiers result for data sets with all features were investigated and compared with the results of new proposed method. For example, for GESTURE data set the results of classic Bayes, kNN and SVM classifiers were recorded as 84.56%, 80.20% and 53.69% respectively. It was made to define the superiority of presented method.

Thus for GESTURE data set, the best classification rate for SVM is obtained as 67.45%. The best results for kNN is obtained as 82.21% and for Bayes as 93.29%. Under these circumstances, the proposed LDC classifier produced the best finding as 96.64%. Despite of the kNN method result which is sufficiently high almost 4% superiority was provided by our method in GESTURE data set. For GESTURE data set case new proposed classifier produced the best classification rate in $k_{opt} = 20$ which is equal to threshold meaning. It means that, the new LDC classifier using the FSLS method selects only 20 optimal feature pairs from 153 subsets and obtains the best result. This statement can be used as a considerable measure in proving the validity and usability of the proposed LDC classifier. Further, in $k = 1$ case, namely, only with two feature our method produces success rate as 71.48% and in this wise left behind the classic classifiers and this superiority continues in all feature pair subsets cases. The illustration of the classification results of classic method and the outputs recorded by proposed classifier for GESTURE data set in varies meaning of k is imaged in Figure 2.

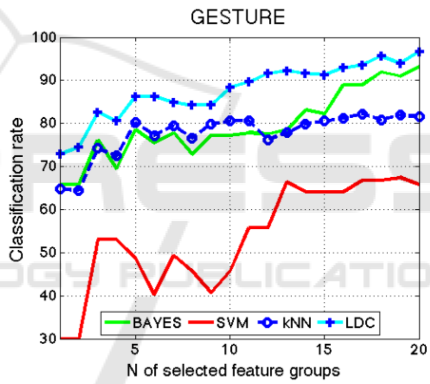

Figure 2: Classification results for GESTURE data set.

Totally 21 feature pair subsets were extracted by (4) from SEEDS data set. The number of selected feature pair subsets by FSLS method was 20 and the best classification result was produced by new LDC classifier as 97.87%. The worst success rate was recorded by kNN as 95.74%. For SEEDS data set Bayes and SVM classifiers have produced the same classification rate as 96.81%. As a result of experiments, an optimal meaning of k_{opt} which produces the best classification rate with the proposed new LDC classifier was found out as 12. As clearly visible from Figure 3. in $k = 12$ case the best result for SEEDS which produced by both of Bayes and SVM was increased almost for 5%. Moreover, in comparison to outputs that were recorded by classic methods the findings of suggested classifier for SEEDS data set in most of means k are the best ones. Additionally, despite of the high success rate obtained by classic classifiers our method is able to produce better outputs. The visual comparison of classic classifiers and the proposed methods outputs for SEEDS data set are illustrated in Figure 3.

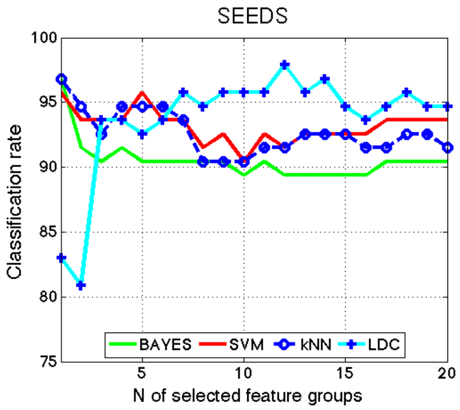

Figure 3: Classification results for SEEDS data set.

For TELESCOPE data set having 45 feature pair subsets in total which were extracted from 10 features both of Bayes and SVM method produced the same success rates as 53.01% and it is the worst one among others. The same situation was observed in PI DIABTES data set case between kNN and SVM. In TELESCOPE case the best result was obtained by the proposed LDC classifier in eighth iteration ($k_{opt} = 8$) as 68.42%. The closest classification result to LDC classifier output is 66.17% that recorded by kNN. As we clearly can see from Figure 4. in all selected feature pair subsets, except four of them, the new suggested classifier produces better results than other methods. The variations of the new proposed classifier results throughout all means k are imaged in Figure 4. Also, in TELESCOPE case our algorithm with only two feature ($k = 1$) obtains better results than SVM and Bayes in all iterations.

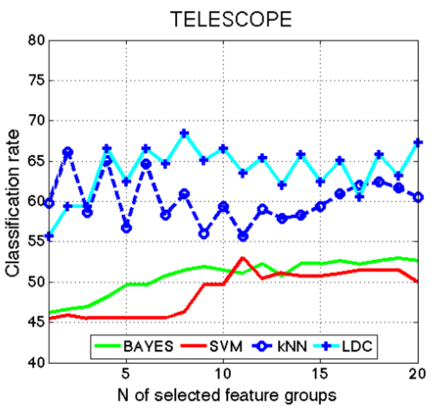

Figure 4: Classification results for TELESCOPE data set.

The similar course of action as in SEEDS case was exhibited by LDC classifier for WINE data set. Namely, in first iterations the proposed method produces the worst success rate than other classifiers and from $k = 6$ to end only the best ones. For WINE data set the worst one among the best classification results was produced by Bayes as 89.53%. Also, the best results of SVM and kNN classifiers were recorded as 91.86% and 94.19% respectively. The proposed LCMF classifier in SEEDS case produces the best classification output as 98.84% and for it only 14 optimal feature pairs of selected 20 subsets has been enough. As mentioned above, the suggested LDC classifier in most of the selected subsets that were extracted from WINE and SEEDS data sets produces better classification outputs. Even in GESTURE case the supremacy was observed in whole iterations. Essentially, this fact describes that the new classifier is not useful only on specific feature pair groups and also available in all subsets. The classic classifiers outputs and the results of LDC classifier for selected feature pairs from WINE data set were visualized in Figure 5.

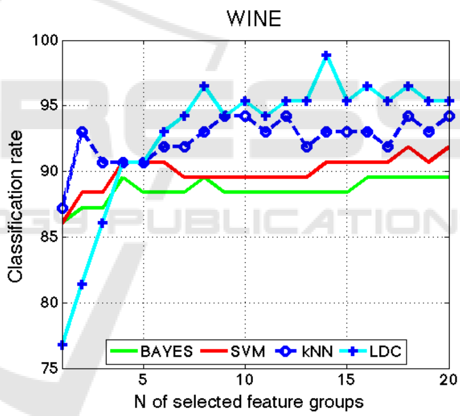

Figure 5: Classification results for WINE data set.

WISCONSIN is the last data set which was used in this study to validate the LDC classifier. The worst classification results in entire the selected subsets from WISCONSIN data set were produced by SVM and the best of them was recorded as 61.36%. And, 75.00% and 78.03% are the best results of kNN and Bayes classifiers for WISCONSIN data set respectively. For the same case the new LDC classifier with 15 optimal feature pairs produces 80.30% classification rate. In this study, from WISCONSIN data set were occurred in total 528 feature pair subsets by (4) and only 15 of them that selected according to FSLS method was sufficient to produce the best classification result. Moreover, in more than half of the selected feature pairs the results obtained by proposed classifier are better than others. The comparison of classification results for WISCONSIN data set are illustrated in Figure 6.

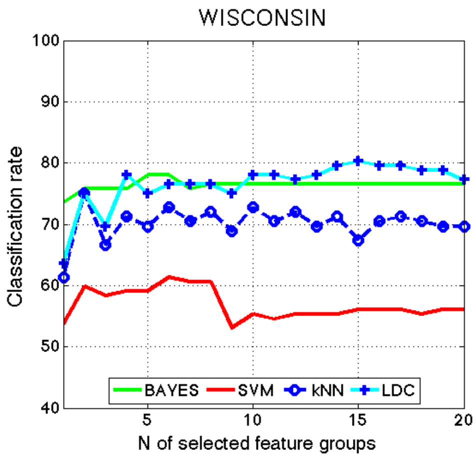

Figure 6: Classification results for WISCONSIN data set.

Generally, as result of experiments in this study, the classification rates obtained from GESTURE, SEEDS, TELESCOPE, WINE and WISCONSIN data sets by new LDC classifier are better than other classic methods outputs. In terms of classification the proposed classifier is superior to kNN, Bayes and SVM methods. This situation and the best classification results obtained by classic classifier methods can be seen in comparison from Table 4.

Table 4: The comparison of the best classification results.

	Bayes	SVM	kNN	LCMF
GESTURE	93.29	67.45	82.21	**96.64**
SEEDS	96.81	95.74	96.81	**97.87**
TELESCOPE	53.01	53.01	66.17	**68.42**
WINE	89.53	91.86	94.19	**98.84**
WISCONSIN	78.03	61.36	75.00	**80.30**

6 CONCLUSIONS

In this study, the novel Lorentzian Distance Classifier for Multiple Feature (LCDMF) method is developed. The proposed classifier uses the improved Feature Selection in Lorentzian Space (FSLS) method. The FSLS method was restructured according to Lorentzian metric and based on *LJ* discriminative criterion. It selects optimal feature subsets from data set with the aim of to reduce the dimension. Thus, by selecting most important feature subsets from original data set according to Lorentzian space metric the best classification results can be produced by proposed LDC classifier. Also, in this study, the pre-processing step is proposed. This pre-processing step is important in terms of transforming the data and making them suitable in Lorentzian space. Further, the covariance matrix calculation in Lorentzian space was described. The validity and correctness of the proposed classifier were tested over GESTURE, SEEDS, TELESCOPE, WINE and WISCONSIN data sets. The performance of new proposed LDC classifier over all data set was evaluated by comparing the classification results with Bayes, kNN and SVM classifiers outputs. In experiments besides the results of the classical classifiers for selected feature pairs, also the results for all features were investigated and compared with the results of new proposed method. As result of experiments, the superiority of proposed LDC classifier to other classic methods is clearly seen.

In future studies, Lorentzian metric may be used for Principal Component Analysis by reconstruction of its internal calculations. Furthermore, the structure of the SVM method may also be reorganized according to properties of the Lorentzian space. These modifications could improve the success rate of the classification.

ACKNOWLEDGEMENTS

This study was supported by TUBITAK (Turkish Scientific and Technical Research Council). The project number is 115E181.

REFERENCES

Louridas P., Ebert C., 2016. Machine Learning, in *IEEE Software*, 33 (5), pp. 110-115.

Wang et al., 2016. Nonlinearity Mitigation Using a Machine Learning Detector Based on k -Nearest Neighbours, in *IEEE Photonics Technology Letters*, 28 (19), pp. 2102-2105.

Bkassiny, M., Li, Y., Jayaweera, S. K., 2013. A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys & Tutorials*, 15(3), 1136-1159.

Theodoridis S., Koutroumbas K., 2009. *Pattern Recognition*, Elsevier, 4th ed.

Kerimbekov, Y., Bilge, H. Ş., Uğurlu, H. H., 2016. The use of Lorentzian distance metric in classification problems. *Pattern Recognition Letters*, 84, 170-176.

Bilge, H. Ş., Keřimbekov, Y., 2015, May. Classification with Lorentzian distance metric. In 2015 *23nd Signal*

Processing and Communications Applications Conference (SIU), pp. 2106-2109. IEEE.

Bilge, H. Ş., Kerimbekov, Y., Uğurlu, H. H., 2015, September. A new classification method by using Lorentzian distance metric. *In Innovations in Intelligent SysTems and Applications (INISTA), 2015 International Symposium* on, pp. 1-6. IEEE.

Gündogan, H., & Kecilioglu, O., 2006. Lorentzian matrix multiplication and the motions on Lorentzian plane. *Glasnik matematički*, 41(2), 329-334.

R. Brualdi, 2010., *Introductory Combinatorics*, Pearson Prentice Hall, 5th ed.

Marcus, M., Minc, H., 1992. *A survey of matrix theory and matrix inequalities,* Courier Corporation.

Lichman, M., 2013. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.