Application of Texture Based Features for Text Non-text Classification in Printed Document Images with Novel feature Selection Algorithm

Soulib Ghosh1, Khalid Hassan Sk1, Ali Hussain Khan1, Ankur Manna1, Showmik Bhowmik2,\*, Ram Sarkar 1

1Department of Computer Science and Engineering, Jadavpur University, India.

2Department of Computer Science and Engineering, Ghani Khan Choudhury Institute of Engineering and Technology, Malda, West Bengal, India

E-mail: [ghoshsoulib@gmail.com](mailto:ghoshsoulib@gmail.com), [skkhalidhassan7@gmail.com](mailto:skkhalidhassan7@gmail.com), [ahk4815@gmail.com](mailto:ahk4815@gmail.com), [ankurcasino108@gmail.com](mailto:ankurcasino108@gmail.com), showmik@gkciet.ac.in, [ramjucse@gmail.com](mailto:ramjucse@gmail.com)

\*Corresponding author

**Abstract**: ~~The amount of document images are rapidly increasing due to the emerging multimedia technology and the acceptance of printed documents among the people. In case of any processing human intervention is not possible due to the ample amount of documents. So, it needs to be converted in editable format which requires~~ **~~Optical~~****~~character~~****~~recognition (OCR)~~**~~.~~ The text non-text separation is the initial step of any OCR system to convert a document image into editable format. Hence, an efficient text non-text separation module is a pressing need. For that purpose, we have proposed a texture based feature for region based text non-text classification method followed by a novel feature selection technique. The entire work is divided into two stages – feature extraction and feature selection. In the feature extraction stage, we have incorporated rotation invariant property with local ternary pattern to form a new texture based feature namely, rotation invariant local ternary pattern (RILTP). Whereas, in the second stage, a novel feature selection technique is proposed which is a modified version of binary particle swarm optimization (**BPSO**). For the evaluation purpose, we have constructed a dataset from an august competition namely, Recognition of Documents with Complex Layouts (RDCL) organized by International Conference on Document Analysis and Recognition (ICDAR) community. We have extracted total 690 images from 70 pages of RDCL 2015 and 75 pages of RDCL 2017, where each class contain 345 data samples. The proposed texture based feature provides an accuracy of 97.09%. Whereas, after applying feature selection the feature dimension is reduced by approximately 55% and at the same time the accuracy increases to 97.5%. The comparative study reveals the strength of the proposed system. It outnumbers most of the widely used state-of-the art texture based features. On the other hand, the proposed feature selection method is also compared with various extensively used wrapper feature selection methods. The performance of the proposed feature selection methodology is very impressive and outperforms some widely used wrapper methods.

1. **Introduction**

Due to recent growth of multimedia technologies, the capability of capturing document content using portable devices have amplified research opportunities in the domain of automatic textual and graphical content acquisition from documents and scene images [1]. The contents of any offline document can be broadly classified into two categories – text and non-text. In the textual part, we mainly deal with the text portion. Whereas, non-text part may contain graphs, pie-chart, tables, pictures, drawings etc. The aim of this work is to successfully identify a segmented region as text or non-text. In the literature, this type of work is referred in different ways namely, Text and Non-text Separation [2], Text localization [3], Text detection and extraction [4] and Suppression of non-text components [5] etc. The text non-text separation task can be formally defined as follows. Let be an image that represents non-empty set of elements. is also non-empty set of output class that contains two elements - represents the text class and stands for non-text class. Where,. Our main task is to devise a function, **which** maps into. It can be represented as:

Text non-text separation plays a key role in OCR related works. Any typical OCR machine deals with the textual content of an input document image. Therefore, the presence of non-texts coupled with the texts may cause an OCR system to produce error. Due to this, researchers have begun to develop various approaches in order to separate out the non-text portion from text portion more efficiently. There are many challenges faced in this domain. Firstly, non-text detection becomes difficult as it can occur in many forms namely, graph, images, pie-charts etc. In some cases, text and non-text occur together as example in graphs. A graph contains figures as well as some labeling (Axis label for example) that occur as a text. Besides, due to various reasons like printing error, external noise due to casual handling, image acquisition error or incompetent preprocessing text components may get attached with non-texts.

Due to emerging multimedia technology, the modern documents are getting more expressive but the coupling of texts and non-texts in those documents are getting stronger as well. This makes the text non-text separation task more challenging. Hereby, an efficient algorithm for the said purpose is a pressing need. Thus in this paper, we have proposed a region based text non-text classification method. For that purpose we have introduced a modified version of local ternary pattern (LTP) based texture feature and a modified version of binary Particle Swarm Optimization (BPSO) based feature selection algorithm to decontaminate the less informative features. For the final classification, we have used Random Forest classifier.

LTP is a texture based feature that encodes any component in three address pattern.In this paper, we have proposed rotation invariant LTP (RILTP). In RILTP, we rotate each pattern obtained from LTP and take the pattern that has minimum equivalent decimal value. The idea behind using RILTP is that it reduces the feature dimension, eliminating the redundant features. It is also parallel towards slight image rotation. Besides, rotation invariance captures the edges properly. Here we have also adopted a feature selection method to remove the redundant features, if any, and to improve the recognition accuracy. The feature selection methodology, applied here, is based on a modified version of BPSO. PCC **(Pearson correlation coefficient**) is incorporated with the BPSO for estimating the fitness value instead of classifier accuracy. Key modules of the proposed approach is shown in figure 1.

|  |
| --- |
|  |
| Figure 1: Pictorial representation of the proposed algorithm for text non-text classification. |

1. **Related Work**

As, in the present work, we have dealt with a feature selection method in text non-text classification, so, description of the related work is divided into two sub-sections. First sub-section deals with the study of text non-text separation, whereas, some existing feature selection methods are briefed in the later sub section.

**2.1. Text non-text separation**

The fundamental task in non-text classification is to recognize the text and non-text parts in an image, which can broadly be carried out using three major approaches namely, region based, connected component (CC) based and pixel based. In the first category, the entire image is segmented into regions following some methods and then each region is classified either as text or as non-text. In this approach, researchers have mostly relied on texture based features for text non-text classification. For example, Oyedotun et al. [6] have employed a feed forward neural network (NN) that trains on first-order statistical features like median and modal pixel intensity values and grey level co-occurrence matrix (GLCM) based second order statistical features like entropy, contrast and energy. Antonacopoulos et al. [7] have used white tiles based features to distinguish between text and graphics. In the paper [8], Shih et al. have used block segmentation algorithm using run length smoothing algorithm (RLSA) and then applied a rule based block recognition algorithm to label a block as: text, horizontal/vertical line, and graphics. In the paper [9] Safonov et al. have proposed a simple and fast mechanism to classify a block of a magazine or newspaper into text, picture, or background. They have performed the classification with the help of textual features and AdaBoost classifier with dual leave-group-of-sources-out cross-validation.

Region based methods, however, fail in case of dense documents where both text and non-text can be present in a region together, thus leading to misclassification. Hence, CC based classification is better in this case where each component is handled individually to classify it as text or non-text. For example, Ghosh et al. [10] have come up with a threshold-based approach, which considers various shape-based features for different categories of commonly used non-texts to classify the components. In another work, Sah et al. [11] have used a modified version of histogram of oriented gradients (HOG) feature descriptor followed by multi-layer perceptron (MLP) classifier to perform the same. In [12], Bhowmik et al. have used a texture-based feature named as rotation invariant local binary pattern (RILBP) followed by MLP based classification. Khan and Mollah [13] [14] have also proposed CC based text and non-text classification techniques. In [13], the authors have followed two different approaches i.e., handcrafted features with shallow learner and automatic feature extraction and classification with convolutional neural network (CNN) for setting the benchmark for results on a self-prepared database. In this work, the authors have used a histogram of distance values that are calculated from the medial skeleton map and distance transform map of the component images as hand-crafted features and the designed CNN model is similar to the LeNet-5 model [15]. In [14], the authors have designed an area occupancy profile based feature for the said purpose for hand-held camera captured scene and document images. Area occupancy is calculated using the equidistant pixels in distance transform map. In another work, Ghosh et al. [16] have first provided an empirical study on different variants of local binary pattern (LBP) based texture features like basic LBP, improved LBP (IBLP) and rotation invariant LBP (RILBP) and then designed robust uniform LBP (RULBP) for text non-text classification. In their work, they have shown that RULBP outperforms the other existing LBP variants with the help of five different classical classifiers namely, MLP, Naïve Bayes (NB), sequential minimal optimization (SMO), K-nearest neighbors (k-NN) and random forest (RF).

CC based classification, however, fails in cases where the components are faded out or broken due to poor document quality. Hence, for such documents, few researchers have considered pixel based technique. In this context, Garz et al. [17] have used scale invariant feature transform (SIFT) and Difference of Gaussian (DoG) based features to perform text non-text classification. In another work, Gobbi et al. [18] have proposed a pixel level text non-text classification technique for old map images. In their work, they have used three features dubbed as fractal dimension, square ratio and circle ratio and, a majority voting based classification. k-NN, NB, support vector machine (SVM), MLP and decision tree (DT) classifiers are used for majority voting. Also, Kosaraju et al. [19] have designed a layout analysis method for complex document images. In their work, they have designed a CNN model that extracts texture based features for classifying pixels as either text or non-text. However, the drawback of these methods is the time required for processing. A more detailed study for text and non-text classification can be found in the survey paper by Bhowmik et al. [20]. A brief summary of all the described methods is presented in Table 1. The table contains the features and classifiers used in each work along with their individual pros and cons.

Table 1. Overview of the state-of-the-art methods for text and non-text separation

|  |  |  |
| --- | --- | --- |
| **Method** | **Features and Classifier** | **Remark** |
| *Region Based Classification* | | |
| Oyedotun and Khashman [6] | **Feature**: First-order statistical features, namely median and modal pixel intensity values and grey level co-occurrence matrix (GLCM) based second order statistical features like entropy, contrast, energy and homogeneity.  **Classifier**: Feed forward neural network (NN). | **Advantage**: Many efficient texture based features are used that captures the texture property appropriately. The method performs well in the cases where text and non-text are situated in different regions.  **Disadvantage**: The method requires ample amount of training samples. Besides, the technique may fail in situations where text and non-text overlap with each other within region under consideration. |
| Antonacopoulos et al. [7] | **Feature**: white tiles based features.  **Classifier**: Rule based. | **Advantage**: skewed and complex-shaped regions are handled efficiently.  **Disadvantage**: Main drawback of the method is that it does not yield satisfactory results in case of complex situations like texture similarity between text and non-text and overlap between regions. It is highly affected by poor binarization. |
| Shih et al. [8] | **Feature**: RLSA based features.  **Classifier**: Rule based | **Advantage**: The method is independent of character font and size and the scanning resolution.  **Disadvantage**: The method assumes that the text and non-text regions will be divided into blocks in any document. This may not hold in case of complex documents. |
| Safonov et al. [9] | **Feature**: Texture based features like mean brightness, standard deviation and gradient based features.  **Classifier**: AdaBoost | **Advantage**: The method is simple and fast.  **Disadvantage**: The method is sensitive towards noise. It falters in case of very complex background surfaced image. |
| *Connected Component based classification* | | |
| Ghosh et al. [10] | **Feature**: Various shape-based features like aspect ratio, Euler number, number of cross-over and stroke width.  **Classifier**: Rule based | **Advantage**: As the method is rule based, the execution time is very less. Obtained accuracy is quiet impressive considering the time complexity and simplicity of the method.  **Disadvantage**: Certain category of components are dealt with. It is not suitable for a common framework. |
| Sah et al. [11] | **Feature**: Basic HOG  **Classifier**: MLP. | **Advantage**: The proposed technique is very simple and fast. The accuracy obtained is quiet impressive. The method is insensitive towards small noise.  **Disadvantage**: The method is evaluated in a very small dataset containing very limited challenges. |
| Bhowmik et al. [12] | **Feature**: RILBP  **Classifier**: MLP | **Advantage**: The proposed technique is invariant towards image rotation. Apart from this, due to the rotation invariant property, the number of features are reduced thus, eliminating the redundant information.  **Disadvantage**: The method is sensitive towards noise. |
| Khan and Mollah [13] | **Feature**: CNN based feature.  **Classifier**: Deep learning. | **Advantage**: The method is script independent. Deep convolution neural network (D-CNN) based automated feature extraction and classification framework is developed.  **Disadvantage**: The method is sensitive towards noise. It needs large training set. |
| Khan and Mollah [14] | **Feature**: Features based on area occupancy profile of equidistant pixels.  **Classifier**: Accuracies reported for five classifiers – NB, MLP, SVM, RF and AdaBoost. | **Advantage**: The method is script invariant and effective for scene images also.  **Disadvantage**: The method extract features from binary image. It is highly sensitive towards binarization procedure. |
| Ghosh et al. [16] | **Feature**: RULBP.  **Classifier**: k-NN, NB, MLP, SMO and RF. | **Advantage**: The method is script invariant. It performs better than the other variants of LBP based texture features.  **Disadvantage**: It is highly sensitive towards binarization procedure. Performance on document with complex layout and background is not ensured. |
| *Pixel based classification* | | |
| Garz et al. [17] | **Feature**: SIFT and DoG based features.  **Classifier**: SVM with RBF as kernel function. | **Advantage**: The method does not reply on any binarization technique. The method is robust towards noise and background clutter.  **Disadvantage**: The method is pixel based contains two stages. As a result, the time complexity is very high. |
| Gobbi et al. [18] | **Feature**: Fractal dimension, square ratio and circle ratio.  **Classifier**: Majority voting of k-NN, NB, SVM, MLP and DT. | **Advantage**: The method is applicable for maps of various kinds and challenges. It is capable of detecting texts, symbols, lines etc.  **Disadvantage**: The method is relied on a segmentation based method. Poor segmentation will lead to a poor outcome. |
| Kosaraju et al. [19] | **Feature**: CNN model based texture features.  **Classifier**: DoT-Net | **Advantage**: The proposed approach is capable of identifying text, image, table, mathematical expression, and line-diagram etc. DoT-Net can capture textural variations among the multiclass regions of documents.  **Disadvantage**: The proposed technique need ample amount of training samples. As it is a pixel based classification, the time complexity is very high. |

**2.2. Feature selection**

Feature selection methods can be divided into three categories: filter, wrapper and embedded. Filter based feature selection has been applied in various domains since a long time [21][22][23]. In [24], the author has proposed a Mutual Information based Feature Selection (MIFS) technique which has applied a heuristic modification to **Mutual Information (MI).** In [25], the authors have proposed a greedy method based on MI**,** mentioning the drawbacks of MIFS. In [26], the authors have proposed a Conditional MIbased feature selection method where the MI of a feature with respect to the class is taken conditionally to already selected features. This method avoids redundancy by considering already selected features. The authors of [27] have proposed a quadratic MI and pruned Parzen window estimator used to estimate MI for Feature selection. In [28], authors have used Chi-square feature selection technique for intrusion detection by using Support Vector Machine (SVM) classifier. In [29], the authors have used Pearson Correlation Coefficient (PCC) for feature selection and used these selected features as input to different classification algorithms - SVM, K-Nearest Neighbour (KNN), C4.5, and compared the results with standard feature selection algorithms.

The rest of the paper is organized as follows. Introduction and literature survey are already discussed in the first two sections consecutively. Section 3 deals with motivation and contribution of the proposed work. Section 4 describes the proposed method. In that section, the feature extraction and the feature selection methodology are thoroughly discussed. Result and comparison are put forward in the section 5. Finally section 6 contains the conclusion and the future scope.

1. **Motivation and Contribution**

As the literature survey reveals, the deep learning based approaches provide a satisfactory outcomes. But, there are few drawbacks of any deep learning based approach. First of all, deep learning based approaches require ample amount of training to train the model properly [30]. The adequate availability of labeled is a difficult task. Moreover, deep learning based approaches are computationally expensive and need high training time. Finally, deep learning based techniques demand high performing GPU machine. Due to these shortcomings, we have followed a feature based approach for text non-text separation. On the other hand feature based approaches suffers from the problem of redundant features [31]. The presence of redundant features hamper the training process and effects the final accuracy. The solution of this problem is to apply a proper feature selection method. Based on the above facts, we have followed a feature based approach facilitated with an efficient feature selection method. The key contribution of the proposed work is listed as follows.

1. We have proposed a modification local ternary pattern (LTP) namely, rotation invariant local ternary patter (RILTP).
2. In any wrapper based feature selection method the use of classifier in each step is necessary. We have proposed novel Pearson co-relation co-efficient (PCC) based BPSO that use PCC instead of classifier, which essentially reduce the execution time.
3. The proposed method is compared with many state-of-the-art texture based features. It has been found that the proposed modification RILTP outperforms all the methods. This essentially make the proposed method empirically sound.
4. **Proposed method**

In this section, the detailed study of the feature extraction and the feature selection methods are covered. Firstly, feature extraction is thoroughly discussed. We have touched the concept of local binary pattern followed by the uniform and the rotation invariant property of LBP. Then, we have discussed LTP and the proposed modified version of LTP. Finally, the feature selection methodology is discussed.

* 1. **Feature Extraction**

As mentioned earlier, in some recent text non-text classification models, researchers have mainly depended on texture based features. A typical shape based feature descriptor estimates its shape information which is highly affected by image dimension. Components present in document image, either text or non-text, may vary in terms of shape and dimension. For example, non-text may be tables, figures, graphs of different sizes. So, shape based features may lead us to wrong estimation. Whereas, texture based features study the texture of the component and performs statistical evaluation based on the neighboring pixels. As a result, texture based features are quiet invariant to the component dimension, rather it is sensitive towards to the edge and contour information. To build an efficient text non-text separation model, in the present work, we have applied texture based features which are capable of capturing the edge and corner information of any component followed by a novel feature selection method.

* + 1. LBP

Ojala et al. in their paper [32] first introduced the texture based operator, LBP, a very effective and computationally simple texture descriptor for monochromatic image.

In the original definition given in [32], there are gray scale pixels surrounding a center pixel at having a radius of unit. So, the position of pixel can be calculated as given in the equation 1.

Where

Let the gray scale intensities of the center and neighboring pixels are denoted as

Respectively.

Let, be the binary feature for the center pixel, calculated using equation 2.

Here, is a -bit binary pattern that can also be encoded into bit unsigned integer. denotes the center pixel. The function in the equation 2 is defined as.

…………. (3)

Where, are two variables that can be replaced with a pixel intensity.

Let is the final feature encoding the relation between the center pixel and the surrounding pixels which is stated as equation 4. The input in the equation 4 will be the center pixel along with all the surrounding pixels. The output will be the corresponding LBP value.

……… (4)

As it can be observed that, for, the value of becomes 8. Whereas, the value of F varies from 0 to 255 which is basically the range defined using 8 unsigned bits.

The basic LBP is further modified by researchers. Two useful modifications of LBP are rotation invariant LBP (RILBP) and uniform LBP (ULBP) which are discussed below.

* + - 1. Rotation Invariant LBP

RILBP [32] is a LBP variant, in which is bit-wise rotated to achieve the least decimal value. So, RILBP function can be defined as equation 5.

Here rotates the binary pattern by bits.

* + - 1. Uniform LBP:

For ULBP [32], a binary pattern is said to be uniform if there is less than or equal to two, 1 to 0 or 0 to 1 transitions. So, we label all uniform patterns distinctly and all non-uniform patterns are labeled with one distinct value. This feature is created because, often, in normal LBP features, a certain type of patterns constitutes majority of the feature sets, so we can label the rest as only one single value without losing too much information. ULBP uses different values to label the patterns.

* + 1. Local ternary pattern (LTP):

LBP [32] has proven to be very effective texture based feature descriptor and resistant lighting effects also. But it is very much sensitive to noise. To overcome this limitation Tan et al. [33], proposed another texture based feature descriptor called LTP. In LBP, the intensity difference between each neighboring pixel and the center pixel is considered even when this difference is too small.. This makes the LBP feature very sensitive to minor intensity changes and thereby sensitive to noise. On the other hand, LTP decides a threshold to detect the inequality. Only if the change is greater than the threshold, it captures the variation. In this regard, LTP is much invariant to noise. LTP is a three-valued code whereas, LBP is a two-valued. So, LTP is more informative. In case of any binary patterns, we mainly target to get the difference between the center pixel and the surrounding pixels to attain the edge information. In this regard, the cases where center pixel is greater and less than the surrounding pixel, both are informative. In case of LBP, we can only obtain the information where center pixel is greater than the surrounding pixel. Whereas, in case of LTP we can acquire both of these information. Therefore, LTP becomes more informative and less sensitive towards noise. LTP method is explained below.

To calculate LTP as explained in [33], without loss of generality we maintain a threshold value and thereafter we consider the pixel values in the range of to be 0 and the ones greater than as 1 and ones less than as -1. So we use 3 values (i.e. -1, 0, 1) to encode the texture information in LTP as explained in equation 7.

can be used to generate two binary patterns and , with the help of equations 8 and 9. Equation 10 and 11 are used to change the individual bits of the two obtained pattern.

Here we apply functions and on individual bits from

4.1.2.1. Uniform LTP (ULTP):

We have incorporated uniform variant to the two binary patterns. Section 3.1.1.2 explains uniform pattern.

4.1.2.2. Rotation Invariant LTP (RILTP):

In this section, we have proposed a modification of LTP, which is capable of capturing information about the data better than LBP (or its variants) and basic LTP. In case of rotation invariant, the feature dimension becomes reduced. As a result, the redundant information get ignored and the important ones are kept. In case of rotation invariant, the string is rotated bitwise and the string with minimum decimal value is considered. So, the edge and the corner information are better acknowledged. After receiving the two binary strings from LTP, the rotation invariant property is applied on the strings following the similar fashion as explained in section 3.1.1.2. Then we take the histogram of the equivalent decimal value.

An example for each feature descriptor methods are given in Figure 2.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| (a) | (b) | (c) | (d) |
|  |  |  |  |
| (e) | (f) | (g) | (h) |

Figure 2: Pictorial representation of the edge and corner information after applying various methods. (a) Input color image (b) Corresponding gray scale image (c) LBP (d) ULBP (e) RILBP (f) LTP (g) ULTP (h) RILTP.

As we can observe from the Figure 2 that RILTP provides better edge information. In other methods such as LBP and its variation, the edge information are not preserved properly. Besides, LTP is also more invariant towards noise which is also evident from Figure 2(h). An example of calculating the RILTP feature is shown below.

Suppose the intensity of center pixel and its surrounding eight pixels are as follows:

|  |  |  |
| --- | --- | --- |
| 116 | 122 | 74 |
| 98 | 95 | 23 |
| 145 | 205 | 92 |

Let us consider the threshold is 5. Following the equation 7 of LTP, the encoded matrix will be as follows.

|  |  |  |
| --- | --- | --- |
| 1 | 1 | -1 |
| 0 |  | -1 |
| 1 | 1 | 0 |

The three valued encoded string will be. After segregating the 1 and -1 according to equation 10 and 11, we get two strings as follows: . Then we apply the rotation invariant property in both string as explained in 3.1.1.1. Finally, the equivalent decimal value is obtained from the rotation invariant strings. The similar approach is applied to all pixels of an image and we take the histogram of those values as our feature. Feature dimension for different methods are given in the Table 2. Please note that in LBP from one code or binary string we get one decimal value whereas in LTP from one binary string we get two decimal values by applying equation 10 and 11. As a result the number of feature in LTP is just twice the number of feature in LBP.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature descriptor** | **Basic** | **Rotation Invariant** | **Uniform** |
| LBP |  |  |  |
| LTP |  |  |  |

Table 2: Feature lengths of different feature descriptors considered here.

Feature Selection

Particle Swarm optimization shares many similarities with Evolutionary Computation (EC) techniques in general and GAs in particular. All these techniques begin with a group of a randomly generated population and utilize a fitness value to evaluate the population. They all update the population and search for the optimum with random techniques. The main difference between the PSO approach compared to EC and GA is that PSO does not have genetic operators such as crossover and mutation. Particles update themselves with the internal velocity; they also have a memory important to the algorithm. Also, in PSO only the ‘best’ particle gives out the information to others. It is a one-way information sharing mechanism, the evolution only looks for the best solution. Compared to GAs, the advantages of PSO are that PSO is easy to implement and there are few parameters to adjust.

In any machine learning model, feature vector which represents the pattern/image under consideration plays a crucial role. A strong discriminant feature vector leads to better classification model. In order to design a suitable feature vector for a particular classification problem researchers sometimes add more and more features without understanding the importance of the same in the classification process. Even features extracted from certain portion of an image may not be informative for classification. But when the feature dimension becomes high exhaustive searching for optimal feature subset will not be a viable option due to high computation complexity and storage requirements. Here comes the importance of introducing some intelligent ways to feature selection. In this paper, to remove the redundant and irrelevant features from LTP variants we have proposed a novel feature selection algorithm based on binary PSO (BPSO) in combination with . Previously, feature selection has been used in many fields of document image processing like – digit recognition [30], word recognition [34], character recognition [35]etc. But to the best of our knowledge, feature selection has not been used in the domain of text non classification.

Firstly, is described followed by the BPSO algorithm. Is a correlation measure which is used in statistics to find the strength of linear dependence between two random variables x and y. It is denoted by the equation 12:

Where, is the value, and represents sample in x and y respectively and n is the total number of samples.

In present context, we have considered different features as random variables. From the equation, we know that the gives a real value between -1 to +1 for any two features. A value lesser than 0 means both the features are inversely dependent on each other i.e. if one variable increases other will decrease. A value equal to 0 means both the features are independent to each other. A value greater than 0 signifies they both are directly dependent on each other i.e. if one feature increases then the other will also increase.

**BPSO:**   
We define our search space as n dimensional space where each dimension refers to an individual feature in the dataset present. We start with p particles where each particle is a dimensional position vector representing features present in that particle i.e. which represents the position vector of particle. The velocity of the particle is which represents velocity of particle. There is a personal best position for each particle. We define as the personal best solution for ith particle. There is also a global best solution for all the particles which is represented as

We perform a finite number of iterations until convergence is reached. At each iteration position vector for each particle is updated. Velocity for each feature (dimension) in a particle in an iteration is calculated according to the following formula:

Here and are random numbers taken in the interval of [0,1]. and are positive constants whose values are taken randomly as well. W is the inertia of weight whose value decreases from 1 to 0 over the course of iterations .

Now the actual velocity for each feature in a particle is BPSO is taken as:

Each individual particle’s position is updated in each iteration according to the following equation 15:

Here denotes position of dimension of particle in the search space in the iteration. The value of 1 signifies that the feature j is included in the particle and value of 0 signifies that feature j is excluded.

After a particle has completed updating its position, its personal best solution is checked with the current solution. Here we introduce the usage of values. It is to be noted that in the basic BPSO, after each iteration personal best and global best get updated. But in the present work, instead of following the conventional approach, we have used values to take the decision when personal best and global best would be updated. An advantage of a correlation co-efficient such as Pearson r is that it provides effect size information (in unit free terms). Many other statistics, such as the independent samples t test, can be converted to r. Effect size indexes such as r can be combined across studies in a meta-analysis. If the current feature vector is having a lower value than that of the personal best then personal best is updated to the current position. Besides personal best position, the global best solution is also checked with that of current position. If the value is lower than that of then global best position is changed to the current position.

The value of each particle will signify the nature of that feature vector. A positive value will mean that the features are having direct dependence with each other. Hence the possibility of having redundant features in that particle is high. But the main goal of feature selection is to reduce the number of features. It can be attained by removing those redundant features. If we obtain a negative value of of a particle, it surely signifies that the features present in that particle are independent with each other or the number of independent features are more than the number of dependent features in it. Hence by not considering those features in the next iteration (but we cannot take decision for single feature, rather we take decision for feature vector itself.) redundancy can be minimized. So a lesser value of will represent a better solution.

The steps are performed for a finite number of iterations or until convergence is reached.

PSO is based on the concept of information sharing through social interaction among the particles which form the swarm (population). Each particle which is a feature vector is shifted towards its local best position as well as the global best position in the swarm at the same time as evident from equation 13. The algorithm for the proposed based BPSO is explained in Algorithm 1.

**Algorithm 1:**

Input: Original feature vector containing n features.

Number of iterations

: Number of particles present in the population

Output: A reduced vector of r features where

1. Create a random matrix (swarm) of size where is number of features and is number of particles (feature vectors). Populate each particle with 0s and 1s randomly.
2. Velocity matrix of size is created with all 0s.
3. is a vector of local best feature vectors achieved for the particle of size is defined which is initialized to the feature vector for the particle obtained in step 1.
4. is the global best feature vector of size 1xn is defined storing the particle with lowest mean value. is set to the value of
5. for = 1 to do:
6. for = 1 to do:
7. for = 1 to do:
8. Calculate the velocity of the feature of particle using

equation 13.

1. end for
2. end for
3. for = 1 to do:
4. for = 1 to do:
5. Calculate the sigmoid of the velocity of the feature of particle

using equation 14

1. Update the position of the feature of particle using equation 15
2. If of particle is lesser than mean PCC of local best

solution ():

1. particle is made the local best solution ().
2. end if
3. If of particle is lesser than of global best

solution ():

1. particle is made the global best solution ()
2. is set to the value of
3. end If
4. end for
5. end for
6. End for
7. When convergence is reached then the global best particle gives the optimal feature vector selected with number of features as .
8. **Experimental Results and Comparison**

This section contains the dataset description, classifier selection, results obtained using proposed method and comparison with some state-of-the art methods. In the comparison section, first stage deals with the comparison of the proposed RILTP with some other texture based features. Whereas, the second stage contains comparative study of feature selection methods. Finally, few error cases are analyzed briefly.

* 1. Preparation of database:

The text non-text classification database is created from an eminent competition named, Recognition of Documents with Complex Layouts (RDCL) organized by International Conference on Document Analysis and Recognition (ICDAR) community. In our present work, 70 pages from RDCL 2015 [36] and 75 pages from RDCL 2017 [37] are taken into consideration. These datasets are generated in PRImA research lab, University of Salford, UK. One of the reasons for considering this dataset is that it contains diverse challenges and complexities in terms of non-texts like equations, tables, images, graphic separator etc. Text and non-text components are manually cropped from these pages to generate two sets – texts and non-texts. We have considered total 690 images, where 345 are texts and other 345 are non-texts. All the cropped text non-text images are converted into gray scale image from their RGB format for feature extraction. Some of the examples of text and non-text data are shown in Figure 3.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 3: sample images from text non-text RDCL database used in the present work. (a) – (c) represent text data, whereas, (d) – (f) represent non-text data..

* 1. Classifier Selection:

We have used random forest (RF) for the classification purpose. RF follows a bootstrapping algorithm [38] depending on a decision tree model which is very helpful in solving the over fitting problem. The less number of hyper parameters and a good prediction result also make it simple and useful. RF deals with the outliers by essentially binning them and also impartial towards non-linear features. RF also breakdown the entire process into several threads which make it computationally more efficient. Due to the above mentioned facts, RF is selected as the classifier. We have availed the facility of WEKA [39], which is a common machine learning tool for the classification task. Initially, the input of the classifier is the features extracted from the training samples. The classifier provides a trained model obtained from these training features. The trained model is used to determine the labeled class (text or non-text) of the test samples and the test accuracy is obtained. This test accuracy is reported as final accuracy of the proposed model.

* 1. Results obtained on the proposed method:

In the dataset, there are total 690 images comprising of 345 text and 345 non-text. The entire dataset is divided into train and test set of ration 75:25. As a result, total 516 images are used as training and the rest is used for testing. The classification accuracy obtained using the proposed RILTP is 97.09%. Table 3 contains train test ratio versus the accuracy values. As we can see, even in a very low train set the method yields good accuracies. Table 3 confirms the strength of the proposed method empirically.

Table 3: Train test ratio versus the accuracy values obtained using RILTP

|  |  |
| --- | --- |
| Train test ratio | Accuracy (in %) |
| 10-90 | 81.96 |
| 20-80 | 89.67 |
| 25-75 | 91.49 |
| 50-50 | 93.62 |
| 66-34 | 94.47 |
| **75-25** | **97.09** |
| 80-20 | 96.38 |

* 1. Testing to check classifier overfitting

In any machine learning algorithm, over fitting is a major issue. Over fitting happens when a model adapts the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize. To establish a generalized feature descriptor, it is always necessary to avoid these issues. To show that the result obtained by our method is not due to classifier overfitting, we have prepared a completely separate dataset for testing. The new dataset is constructed from media team document pages [40], which is a publicly available and widely used dataset. This dataset contains total 526 old printed document pages consisting of various noises due to aging. We have cropped 100 text and 100 non-text region images from these page images to prepare our new dataset.

For this experiment, the random forest classifier is trained based on the RILTP feature vector using the previous training set (as mentioned in section 5.1) and tested on the newly prepared dataset.. As previously discussed, if there is over fitting, then the method will eventually fail to yield satisfactory result in different dataset containing data with unknown nature (in our case noisy data) . However in this experiment, the proposed method have achieved 92.01% test accuracy, which is quite satisfactory if we consider the image quality. Therefore the suspection of classifier overfitting can be eliminated..

As previously mentioned, in many feature extraction techniques the main concern is to handle the redundant features. The unwanted features hampers the trained model, hence, decreasing the classification accuracy. So we have proposed a BPSO based feature selection technique to get rid of the redundant features. Initially, the number of features are 72, whereas, the feature dimension have decreased to 32 after applying feature selection method. At the same time the accuracy has also increased to 97.5%. The proposed feature selection method decreased the feature dimension by 55% and increased the classification accuracy simultaneously. Table 4 shows a detailed result obtained from the feature selection method.

Table 4: outcome of the feature selection method:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature Dimension before Feature selection | Accuracy before feature selection | Feature Dimension After Feature selection | Accuracy after feature selection |
| 72 | 97.09% | 32 | 97.5% |

The *population size* vs. *accuracy* graph is shown in Figure 4(a). The BPSO algorithm is applied for 30 iterations over the population to calculate results for this graph. The population size denotes the number of particles present in the swarm in the BPSO. The classification accuracy starts off with 89.89% for 5 number of particles and increases linearly to a maximum of 95.7% at 20 population size. Further increase in population size does not increase accuracy any higher.

The *number of iteration* vs *accuracy* graph is shown in Figure 4(b). The population size consisted of 20 particles. The Accuracy starts with 89.85% at 5 iterations and increases linearly with iterations. It reaches a global maximum of 97.5% at 25 iterations.

|  |  |
| --- | --- |
|  |  |
| 4(a) | 4(b) |

Figure 4: Result analysis of BPSO (a) Population size vs. Accuracy (b) Iteration vs. Accuracy

* 1. Comparison with state-of-the-art methods:

In this subsection, the proposed feature extraction method is compared with some popular texture based features. Whereas, the feature selection is also compared with few state-of-the-art feature selection methods. The proposed RILTP is compared with some widely used text non-text separation methods GLCM [41], GLRLM [42] and HOG [43]. Besides, the comparative study also includes three types of LBP [32] namely, basic LBP, rotation invariant LBP, Uniform LBP and basic LTP [33], uniform LTP. In the paper [44], Murala et al. have proposed Local Ternary Pattern (LTrP) based texture feature for content based image retrieval. We have also shown empirically that our proposed RILTP is superior to LTrP in the domain of text non-text classification. For more detailed comparison, we have also compared our result with three versions of LTrP namely, basic LTrP, rotation invariant LTrP and uniform LTrP. Table 5 contains the thorough comparison of the proposed RILTP with all the existing methods. Table 5 also contains the precision, recall, F-measure for all the methods along with their corresponding accuracy. It can be noticed from the Table 4 that, RILTP outnumbers all the state-of-the methods.

Table 5: Texture based features and their corresponding precision, recall, F-measure and accuracy. In all of the comparisons, the classifier used in RF and the train test split is 75-25.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Accuracy (in %) | Precision | Recall | F-Measure |
| Basic LBP [32] | 96.51 | 0.965 | 0.965 | 0.965 |
| Uniform LBP [32] | 95.34 | 0.954 | 0.953 | 0.953 |
| Rotation invariant LBP [32] | 94.18 | 0.943 | 0.942 | 0.942 |
| Basic LTP [33] | 95.93 | 0.959 | 0.959 | 0.959 |
| Uniform LTP [33] | 92.44 | 0.925 | 0.924 | 0.924 |
| Basic LTrP [44] | 96.25 | 0.963 | 0.963 | 0.962 |
| Rotation Invariant LtrP [44] | 95.34 | 0.954 | 0.953 | 0.953 |
| Uniform LTrP [44] | 95.93 | 0.959 | 0.959 | 0.959 |
| HOG [43] | 85.46 | 0.856 | 0.855 | 0.854 |
| GLCM [41] | 92.44 | 0.924 | 0.924 | 0.924 |
| GLRLM [42] | 91.86 | 0.922 | 0.919 | 0.919 |
| **RILTP (Proposed)** | **97.09** | **0.971** | **0.971** | **0.971** |

From the above comparison, it is evident that the proposed RILTP outperforms all the methods and stands best. Though the reported accuracy is 1-to-3% more than the other comparing methods, it has a significant implications. First of all, the text non-text separation in document image is one of the most challenging task due to the presence of text and non-text with varied shape, size and orientations. Moreover, many non-text regions containing charts, tables, graphs etc also contain text in the form of labels and data makes it even more complex. The regions containing signature which are generally considered as non-text but poses very close characteristics to text. These signifies the inherent complexity of this task. Therefore increasing the accuracy by 1-3% margin is also a crucial task. On the other hand, text non-text separation does not exist as a discrete domain. Rather, it is used as a preprocessing step for any OCR system. Therefore, many tasks like – handwritten printed separation, character recognition are performed after this. If there is any error in text non-text separation, then that error will affect the later steps. The errors in each step will get accumulated. As a result, the entire system will be error prone. Thus, 1-3% increase in accuracy is quite convincing in this domain.

For comparing the feature selection method, we have considered few state-of-the-art wrapper based feature selection methods. In each case, the feature selection technique is applied on the RILTP feature set. The comparison set contains methods like – GA [45], GSA [46], PSO [47], DGA [48], BGSO [49] and HMOGA [50]. Table 6 contains the number of feature and their corresponding accuracy obtained from each feature selection methods. It also contains the precision, recall and F-measure for all the methods. The input for each feature selection method is RILTP feature set that has a feature dimension of 72.

Table 6: Detailed comparative study of the feature selection methods. The second and the third column in the table contain reduced feature dimension and the final accuracy. The other columns contain precision, recall, F-measure of the compared methods.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | Number of reduced feature | Accuracy (in %) | Precision | Recall | F-measure |
| GA [45] | 46 | 86.53 | 0.865 | 0.864 | 0.864 |
| GSA [46] | 22 | 88.94 | 0.888 | 0.889 | 0.889 |
| PSO [47] | 39 | 83.17 | 0.830 | 0.830 | 0.831 |
| DGA [48] | 26 | 88.94 | 0.890 | 0.889 | 0.889 |
| BGSO [49] | 31 | 94.23 | 0.941 | 0.943 | 0.942 |
| HMOGA [50] | 32 | 89.9 | 0.899 | 0.899 | 0.899 |
| **BPSO (Proposed)** | **32** | **97.5** | **0.975** | **0.975** | **0.975** |

It is evident from the Table 6 that the proposed BPSO outperforms all the basic wrapper based feature selection methods. In any feature based approach, the main drawback is redundant features. These redundant features also hamper the training phase and misguide the train model. So, removal of these redundant information is a pressing need. That’s why, a feature selection technique is adopted. The main purpose of the feature selection module is to eliminate redundant information by reducing the feature dimension. From table 6, we can notice that the feature dimension is reduced to 32 from 72, which is almost 55% reduction. Though, the increase in accuracy is marginal, the reduction is feature dimension is very impressive. On the other hand, unlike other wrapper methods, the proposed BPSO does not use classifier in each steps. It uses PCC instead of classifier. Hence, the execution time is also reduced. This is a major advantage of the proposed BPSO over other wrapper methods. Besides, the performance of the proposed BPSO method is also impressive. Based on the table 6, the proposed BPSO outperforms other algorithms in terms of reduction in feature dimension and increase in accuracy.

* 1. Error Case Analysis:

Though the proposed RILTP and BPSO perform well and yields desirable classification accuracy with the help of very little number of features, it has certain limitations. Due to the immense challenges in the non-text detection, the method falters in certain cases. In many cases, the non-texts occur as a table. But at the same time, the content of the tables are mostly texts. So there is a combination of texts and non-texts as shown in Figure 4(a). In this cases, the features obtained are misleading for the model. Similarly, there exists many graphs that contains texts and non-texts both as shown in Figure 4(b). We have considered signature as non-text. Whereas, most of the signatures contain prominent texts as shown in Figure 4(c). So, signatures are also point of concerns. On the hand, the main problem faced in text detection is the color of background. In the case of inverted text or revers video where the background is black or colored as shown in Figure **3(d),** whereas, in other cases the color of the background is white as shown in Figure 4(e). In this work, we have considered texture based feature that computes features based on some statistical measures of the pixel intensities. So, pixel intensity values play a key role in feature extraction method. If the background color varies, the method will also face difficulties in handling the issues.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
| (a) | (b) | | (c) |
|  | |  | |
| (d) | | (e) | |

Figure 4: Error case analysis. (a) – Table that contain texts and non-texts, (b) – Graph containing texts and non-text, (c) – signature, (d) –text region with black background (or inverted text), (e) –text region with white background.

1. **Conclusion**

In recent days, OCR plays a very important role due to the emerging multimedia technology. As a result text non-text separation has begun to gain courtesy. A proper text non-text separation algorithm efficiently separate non-text components from textual part, which increase the performance of the OCR in further steps. In this paper, we have proposed a region based text non-text separation method for the said purpose. We have proposed a modified version (rotation invariant) of LTP. The main idea of using rotation invariant is that it removes the redundant features and decreases the feature dimension. The proposed method is compared with various state-of-the-art methods and various versions of LBP, LTP and LTrP. Our proposed method, rotation invariant LTP yields an outstanding outcome providing recognition accuracy of 97.09%. The proposed method outperforms other variations of LBP, LTP and LtrP as well as some commonly used methods like – GLCM, GLRLM and HOG on the used database. Besides, we have also proposed a novel feature selection method which is a modified version of BPSO. The proposed feature selection method successfully removes all the redundant features that decrease the feature dimension as well as increase the classification accuracy. Though the increase in classification accuracy is very marginal (accuracy increases to 97.5% from 97.09%), we have noticed a feature dimension reduction of approximately 55%. Besides, we have compared the feature selection method with other state-of-the-art wrapper methods namely, GA, GSA, PSO, DGA, BGSO and HMOGA. The comparison result has also empirically revealed that the proposed BPSO method outperforms all the methods mentioned before. Though the current system provides a promising result with the help of very little feature dimension, but the method falters in some of the cases like – pie chart, situations where text embeds with non-text. We have also discussed with several error cases in the section 4.5. As a future scope, we can improve the method to deal with those error cases and extend our experimental domain to some other complex datasets.

**Reference:**

[1] X. C. Yin, Z. Y. Zuo, S. Tian, and C. L. Liu, “Text Detection, Tracking and Recognition in Video: A Comprehensive Survey,” *IEEE Transactions on Image Processing*, vol. 25, no. 6. Institute of Electrical and Electronics Engineers Inc., pp. 2752–2773, Jun-2016.

[2] T. A. Tran, I. S. Na, and S. H. Kim, “Separation of text and non-text in document layout analysis using a recursive filter,” *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 10, pp. 4072–4091, 2015.

[3] C. Yu, Y. Song, and Y. Zhang, “Scene text localization using edge analysis and feature pool,” *Neurocomputing*, vol. 175, no. PartA, pp. 652–661, 2016.

[4] M. D. A. Asif, U. U. Tariq, M. N. Baig, and W. Ahmad, “A novel hybrid method for text detection and extraction from news videos,” *Middle - East J. Sci. Res.*, vol. 19, no. 5, pp. 716–722, 2014.

[5] R. Sarkar, S. Moulik, N. Das, S. Basu, M. Nasipuri, and M. Kundu, “Suppression of non-text components in handwritten document images,” in *ICIIP 2011 - Proceedings: 2011 International Conference on Image Information Processing*, 2011.

[6] O. K. Oyedotun and A. Khashman, “Document segmentation using textural features summarization and feedforward neural network,” *Appl. Intell.*, vol. 45, no. 1, pp. 198–212, Jul. 2016.

[7] A. Antonacopoulos and R. T. Ritchings, “Representation and classification of complex-shaped printed regions using white tiles,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2, no. May, pp. 1132–1135, 1995.

[8] F. Y. Shih and S.-S. Chen, “Adaptive document block segmentation and classification,” *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 26, no. 5, pp. 797–802, 1996.

[9] I. V Safonov, I. V Kurilin, M. N. Rychagov, and E. V Tolstaya, “Segmentation of Scanned Images of Newspapers and Magazines,” in *Document Image Processing for Scanning and Printing*, Springer, 2019, pp. 107–122.

[10] S. Ghosh, R. Bhattacharya, S. Majhi, S. Bhowmik, S. Malakar, and R. Sarkar, “Textual Content Retrieval from Filled-in Form Images,” in *Workshop on Document Analysis and Recognition*, 2018, pp. 27–37.

[11] A. K. Sah, S. Bhowmik, S. Malakar, R. Sarkar, E. Kavallieratou, and N. Vasilopoulos, “Text and non-Text recognition using modified HOG descriptor,” *2017 IEEE Calcutta Conf. CALCON 2017 - Proc.*, vol. 2018-Janua, pp. 64–68, 2018.

[12] S. Bhowmik, R. Sarkar, and M. Nasipuri, “Text and non-text separation in handwritten document images using local binary pattern operator,” in *Advances in Intelligent Systems and Computing*, 2017, vol. 458, pp. 507–515.

[13] T. Khan and A. F. Mollah, “AUTNT-A component level dataset for text non-text classification and benchmarking with novel script invariant feature descriptors and D-CNN,” *Multimed. Tools Appl.*, vol. 78, no. 22, pp. 32159–32186, 2019.

[14] T. Khan and A. F. Mollah, “Text non-text classification based on area occupancy of equidistant pixels,” *Procedia Comput. Sci.*, vol. 167, pp. 1889–1900, 2020.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] S. Ghosh, D. Lahiri, S. Bhowmik, E. Kavallieratou, and R. Sarkar, “Text/non-text separation from handwritten document images using LBP based features: An empirical study,” *J. Imaging*, vol. 4, no. 4, 2018.

[17] A. Garz, R. Sablatnig, and M. Diem, “Using local features for efficient layout analysis of ancient manuscripts,” in *2011 19th European Signal Processing Conference*, 2011, pp. 1259–1263.

[18] S. Gobbi, M. Ciolli, N. La Porta, D. Rocchini, C. Tattoni, and P. Zatelli, “New Tools for the Classification and Filtering of Historical Maps,” *ISPRS Int. J. Geo-Information*, vol. 8, no. 10, p. 455, 2019.

[19] S. C. Kosaraju *et al.*, “DoT-Net: Document Layout Classification Using Texture-Based CNN,” in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 1029–1034.

[20] S. Bhowmik, R. Sarkar, M. Nasipuri, and D. Doermann, “Text and non-text separation in offline document images: a survey,” *Int. J. Doc. Anal. Recognit.*, vol. 21, no. 1–2, Jun. 2018.

[21] P. K. Singh, I. Chatterjee, and R. Sarkar, “Page-level handwritten script identification using modified log-Gabor filter based features,” in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems, ReTIS 2015 - Proceedings*, 2015, pp. 225–230.

[22] X. He, D. Cai, and P. Niyogi, “Laplacian score for feature selection,” in *Advances in neural information processing systems*, 2006, pp. 507–514.

[23] K. K. Bharti and P. K. Singh, “A survey on filter techniques for feature selection in text mining,” in *Advances in Intelligent Systems and Computing*, 2014, vol. 236, pp. 1545–1559.

[24] R. Battiti, “Using Mutual Information for Selecting Features in Supervised Neural Net Learning,” *IEEE Trans. NEURAL NETWORKS*, vol. 5, no. 4, p. 531, 1994.

[25] N. Kwak and C.-H. Choi, “Input Feature Selection for Classification Problems,” *IEEE Trans. NEURAL NETWORKS*, vol. 13, no. 1, p. 143, 2002.

[26] F. Fleuret, “Fast binary feature selection with conditional mutual information,” *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, 2004.

[27] T. W. S. Chow and D. Huang, “Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information,” *IEEE Trans. Neural Networks*, vol. 16, no. 1, pp. 213–224, 2005.

[28] I. Sumaiya Thaseen and C. Aswani Kumar, “Intrusion detection model using fusion of chi-square feature selection and multi class SVM,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.

[29] J. Biesiada and W. Duch, “Feature selection for high-dimensional data - A pearson redundancy based filter,” *Adv. Soft Comput.*, vol. 45, pp. 242–249, 2007.

[30] S. Ghosh, S. Bhowmik, R. Sarkar, K. K. Ghosh, and S. Chakraborty, “A filter ensemble feature selection method for handwritten numeral recognition,” in *International Conference on Emerging Technologies for Sustainable Development (ICETSD ’19)*, p. EMR (2016): 007213.

[31] S. Saha *et al.*, “Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm,” *Appl. Sci.*, vol. 10, no. 8, p. 2816, Apr. 2020.

[32] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[33] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[34] S. Malakar, M. Ghosh, S. Bhowmik, R. Sarkar, and M. Nasipuri, “A GA based hierarchical feature selection approach for handwritten word recognition,” *Neural Comput. Appl.*, pp. 1–20, Jan. 2019.

[35] G. Kim, G. Kim, S. Kim, T. Tek, and S. Kyungki, “Feature Selection Using Genetic Algorithms for Handwritten Character Recognition,” 2000.

[36] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher, *ICDAR2015 Competition on Recognition of Documents with Complex Layouts-RDCL2015 †*. .

[37] C. Clausner, A. Antonacopoulos, and S. Pletschacher, “ICDAR2017 Competition on Recognition of Documents with Complex Layouts - RDCL2017,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2017, vol. 1, pp. 1404–1410.

[38] K. K. Ghosh, S. Ghosh, S. Sen, R. Sarkar, and U. Maulik, “A two-stage approach towards protein secondary structure classification,” *Med. Biol. Eng. Comput.*, vol. 58, no. 8, pp. 1723–1737, Aug. 2020.

[39] M. Hall, G. Holmes, B. Pfahringer, P. Reutemann, E. Frank, and I. H. Witten, “The WEKA data mining software: An update EThOS for EAP View project Commercial data mining projects using the ADAMS platform View project The WEKA Data Mining Software: An Update,” 2014.

[40] “The MediaTeam document database.” [Online]. Available: http://www.mediateam.oulu.fi/downloads/MTDB/download2.html. [Accessed: 29-Jul-2020].

[41] F. Albregtsen, “Statistical Texture Measures Computed from Gray Level Coocurrence Matrices,” 2008.

[42] F. Albregtsen, “Statistical Texture Measures Computed from Gray Level Coocurrence Matrices,” *… Lab. Dep. Informatics, Univ. …*, pp. 1–14, 2008.

[43] N. Dalal, B. T. Histograms, and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” pp. 886–893, 2005.

[44] R. P. Maheshwari, B. Raman, S. Murala, and R. Balasubramanian, “Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval,” *IEEE Trans. IMAGE Process.*, vol. 21, no. 5, 2012.

[45] K. F. Man, K. S. Tang, and S. Kwong, “Genetic algorithms: Concepts and applications,” *IEEE Trans. Ind. Electron.*, vol. 43, no. 5, pp. 519–534, 1996.

[46] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, “GSA: A Gravitational Search Algorithm,” *Inf. Sci. (Ny).*, vol. 179, no. 13, pp. 2232–2248, Jun. 2009.

[47] Y. Z. Yao and Y. R. Xu, “Parameter analysis of particle swarm optimization algorithm,” *Harbin Gongcheng Daxue Xuebao/Journal Harbin Eng. Univ.*, vol. 28, no. 11, pp. 1242–1246, 2007.

[48] F. F. Easton and N. Mansour, “Distributed genetic algorithm for deterministic and stochastic labor scheduling problems,” *Eur. J. Oper. Res.*, vol. 118, no. 3, pp. 505–523, Nov. 1999.

[49] H. Teimourzadeh, B. Mohammadi-Ivatloo, and S. Asadi, “Binary Group Search Optimization for Distribution Network Reconfiguration,” 2020, pp. 103–126.

[50] O. B. Augusto, S. Rabeau, P. Dépincé, and F. Bennis, “Multi-objective genetic algorithms: A way to improve the convergence rate,” *Eng. Appl. Artif. Intell.*, vol. 19, no. 5, pp. 501–510, Aug. 2006.