International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 02 | Feb 2021 www.irjet.net p-ISSN: 2395-0072

DIMENSION REDUCTION OF HIGH DIMENSIONAL IMAGE DATA USING HYBRID COMPRESSION TECHNIQUE

Priyanka Kujur

SUIIT, Sambalpur University, India

---***---

Abstract - *A large amount of storage capacity is required to store the digital image. To store a digital image using less storage capacity, the actual size of the image has to be reduced without losing its quality. The reduction of the size of an image can also be done by reducing the dimension of the image. Previously various feature selection and feature extraction methods like Principal component analysis (PCA) have been used to reduce the dimension. This paper proposes a hybrid compression technique which uses the combination of two lossy compression techniques, Discrete cosine transform (DCT) and Discrete wavelet transform (DWT). This hybrid compression technique gives a better result when compared to standalone DCT and DWT.*

Key Words: *Discrete cosine transform (DCT), Principal component analysis (PCA), Discrete wavelet transform (DWT).*

1. INTRODUCTION

Due to the rapid development of computer technology, a large number of high dimensional data is generated with much useful information. Such data includes high-resolution images, text documents, gene expression and so, on. High dimensional data includes a large number of irrelevant and redundant information [11]. So many feature selections, feature extraction, and compression techniques are used to eliminate irrelevant and redundant information. Image compression is used to reduce the redundancies and randomness to increase the storage capacity and efficiency level of the image. Therefore, it is essential to compress the image and store only the required information needed to reconstruct the image. The high-resolution image is referred to as high dimensional data space as each image data is organized into two-dimensional pixel values in which each pixel consists of its respective RGB bit values [1]. The digital image plays an important role in digital communication. The original image contains a large amount of data that requires large storage space and large bandwidth, so due to low storage space and limited bandwidth, it becomes difficult to provide quality communication with the original image so it is required to reduce the size of an image data for better quality compression. Sometimes images having a large area of the same color will have large redundancies and similar images that have frequent and large changes in color will be less redundant and hard to compress [10].

The objective of this paper is to reduce the size of the digital image data using a combination of compression techniques. Image compression can be done in two ways Lossy and lossless compression. In this paper, we present a performance evaluation of a hybrid model which is the combination of several compression techniques like DWT and DCT lossy compression techniques and Huffman encoding technique which is lossless. At last, a lossless technique Huffman encoding is used so that our PSNR and MSE will go better, and due to DWT and DCT, we will get a good level of compression without losing any image. When compared to standalone DCT and DWT algorithms, the proposed hybrid algorithm gives better results in terms of a higher compression ratio. The compression ratio of this hybrid technique is then compared with the compression ratio of PCA and it is found that hybrid technique gives a higher compression ratio hence it goes better than PCA in terms of compression ratio.

This paper contains following more sections, section 2 describes the related works done by different authors. Section 3 describes the basic compression and dimension reduction techniques used along with the proposed work. Section 4 shows the result and analysis of the proposed method and section 5 winds up the paper with a conclusion.

2. BACKGROUND STUDY

Image compression has many practical applications because of huge data storage, transmission, and retrieval for medical imaging, documents, and videoconferencing [10]. So, ZHAO Zhongwen Beijing, China and GUO Huanghuang, Equipment Academy Beijing, China used PCA, a statistical method that conducts feature dimension reduction and data set simplification. By utilizing this idea with dimension reduction, many indexes can be transferred into a few uncorrelated indexes without information corruption of original data, which leads

to reduce computing and helps to improve calculation efficiency together with accuracy [11].
S. C. Ng in his paper aimed to evaluate the application of PCA on digital image feature reduction and compare the quality of the feature reduced images with different variance values. Experimental results showed that the PCA technique effectively reduces the dimension of image data while still maintaining the principal properties of the original image [1]. Shereena V. B.and Julie M. David in their work presented a comparative study of two linear dimension reduction methods namely PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). Experimental results show that PCA gives better performance in terms of higher precision and recall values with lesser computational complexity than LDA [2]. In their research, Sarabjot Singh Grewal, Er. Beant Kaur Department of Electronics and Communication Punjabi University Patiala presented image compression and error correction control using deep neural networks. First, Deep Neural Network (DNN) is implemented for image compression. They applied the algorithm on four different images including 512*512, 256*256, 128*128, and 64*64. The result of DNN implementation shows the better quality of the decompressed images along with less computational capacity and the proposed algorithm is effective in accomplishing better error correction and reducing the storage requirements [3]. In their paper, A.H.M. Jaffar Iqbal Barbhuiya, Tahera Akhtar Laskar, K. Hemachandran3 Department of Computer Science Assam University, modified and efficient image compression scheme based on DWT which results in a good compression ratio without degrading the quality of the image. The proposed algorithm has been analyzed and compared with some other existing methods. Several quality measurement techniques like PSNR and MSE have been considered to determine the image compression concerning the reference image [5]. There are several methods available to compress an image file. In this paper Uday Bhade, Sanjeet Kumar, Prashant Dwivedi, Shahbaz Soofi, Avinash Ray, Department of Electrical Electronics Engineering use different types of image compression techniques SVD, BTC DWT DCT to compress image file. They find various parameters i.e. compression ratio, MSE, BPP, PSNR from each image compression technique, and then compare each technique's parameters from one another [7]. In their paper, R.Praisline Jasmi Department of ECE, Kalasalingam University, Mr.B.Perumal Department of ECE Dr.M.Pallikonda Rajasekaran Head of the Department of ECE proposed image compression using simple coding techniques called Huffman; Discrete Wavelet Transform (DWT) coding and a fractal algorithm are done. Huffman coding technique was involved in reducing the redundant data in input images. DWT enabled to improve the quality of the compressed image. The fractal algorithm involved the encoding process and gave a better compression ratio. By using the above algorithms Peak signal to noise ratio (PSNR), Mean Square Error (MSE) and compression ratio(CR) and Bits per pixel(BPP) of the compressed image by giving 512512 input images were calculated and also the comparison of performance analysis of the parameters with that above algorithms was done. The result clearly explained that the Fractal algorithm provides a better Compression ratio (CR) and Peak Signal to noise ratio (PSNR) [9]. The DCT-based image compression quality was studied by experiments in the subjective and the objective methods in this paper. At first, XiHongZHOU, Department of Computer and Information Technology Northeast Petroleum University, illustrated the option method of DCT coefficient as the DCT coefficients option is important to affect the static image compression quality. Next, according to the experiments data and simulation images, in the objective evaluation method, was evaluated by computing the MSE and PSNR, that affect the compression of image quality is further analysis, the result indicates: the bit rate of less than 0.75bbp is the most applicable for the good image compression quality [10].

3. PROPOSED WORK

Digital image and video images require a large amount of capacity to store. Image compression means reducing the size of the image, without losing its quality. Depending on the constructed image, to be the same as the original, two techniques for compression exist. Two techniques are lossy techniques and lossless techniques. Digital imaging plays an important role in image processing therefore it is necessary to develop a system that produces a high degree of compression while preserving the critical image. In this paper, we present a hybrid model which is the combination of several compression techniques. This paper presents the implementation of DWT and DCT these are the lossy techniques and also introduce the Huffman encoding technique which is lossless. At lossless technique is also used so that the PSNR and MSE will go better, and due to DWT and DCT, a good level of compression takes place without losing any information. When compared to standalone DCT and DWT algorithms, the proposed hybrid algorithm gives better results in terms of a higher compression ratio. The compression ratio of this hybrid technique is then compared with the compression ratio of PCA and it is found that hybrid technique gives a higher compression ratio hence it goes better than PCA and standalone DCT and DWT.

3.1 Hybridization of compression techniques

The objective of the hybrid algorithm is to utilize the properties of both DWT and DCT. Here the original image of size 256x256 or any resolution provided divisible by 32, is first divided into blocks of NxN. Each block is decomposed using 2-D DWT. Then the coefficients with low frequency are passed to the next stage whereas the coefficients with high frequency are discarded. Then the passed LL components are further decomposed using another 2-D DWT. Then the 8-point DCT is applied to the DWT Coefficients. To achieve higher compression, the majority of high coefficients are discarded. To achieve more compression quantization is performed. In this stage, many of the high-frequency components are rounded to zero. The quantized coefficients are then scaled using the scaling factor. Then the image is reconstructed using the inverse procedure. During inverse DWT, zero values are been added in place of detailed coefficients.

Hybrid compression techniques use advantages of all three techniques, like compression time for DCT is lowest, DWT is moderate and Huffman coding. Compression takes the highest time. But at the same time quality of the image after decompression in DCT and DWT is not good as compared to Huffman coding. In the Huffman algorithm, we get a decompressed image the same as the original image.

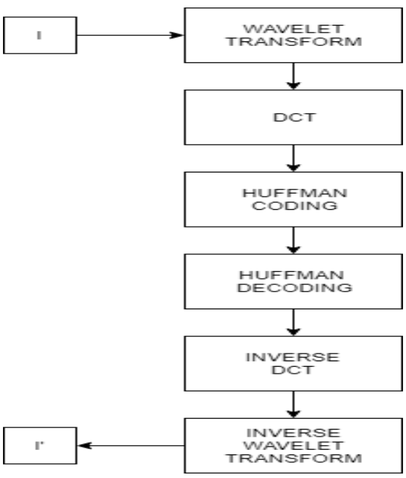


Fig 1: Hybridization of Compression Techniques

3.1.1 Discrete wavelet transform (DWT)-

DWT compresses the given image without the loss of any information in that particular image, which comes under the lossless type of image compression [7].DWT is used to transform a discrete-time signal into Discrete Wavelet Representation. Wavelets are functions that are defined over a finite interval and

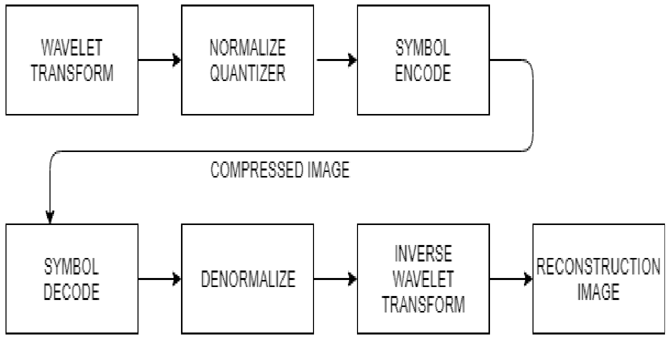


Fig 2: Discrete wavelet transform

have an average value of zero. The wavelet transform is the most significant and practical computational tool for the multiplication of signal and image processing applications. Unwanted noise and blurring in a digital image can be removed by using a wavelet transform [7]. Wavelet transform represents an arbitrary function as a superposition of a set of such wavelets or basis functions. These basic functions are also called baby wavelets which are obtained from a single prototype wavelet called the mother wavelet, by scaling and shifting. The Discrete Wavelet Transform (DWT), can display the image at a different resolution so it has emerged as an efficient tool for medical image compression. In wavelet transform, the decomposition of a particular image consists of two parts, first is lower frequency or approximation of an image and second is a higher frequency or detailed part of an image [7].

The discrete wavelet transform in two dimensions of functions f(x,y) of size MXN is

$$W\emptyset(j_0,m,n) = \frac{1}{\sqrt{MN}}\sum_{X=0}^{M-1}\sum_{Y=0}^{N-1} f(x,y)\emptyset(j_0,m,n)(X,Y)$$

Here, f(X), $\emptyset_{j0,m,n}(X)$ and $\emptyset_{j0,m,n}(X)$ are functions of the discrete variable X=0,1,2. . . . . . M-1 [5].

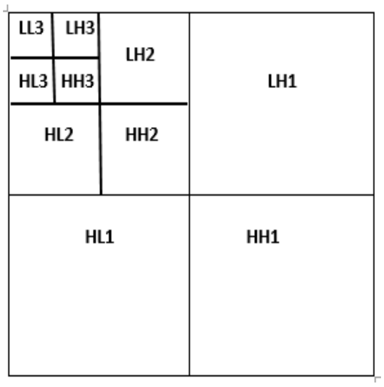


Fig 3: Wavelet Transform

The wavelet filter decomposition of an image is explained in the figure below where four different sub-images are obtained, the approximation (LL), the vertical detail (LH), the horizontal detail (HL), and the diagonal detail (HH). The LL subband contains the maximum description of the image. LH contains mostly the vertical information, which corresponds to horizontal edges, HL represents the horizontal detail information from the vertical edges and HH contains the diagonal detail information [7]. As the LL sub-band is the replica of the original image, taking only the approximate image reduces the image size without any distortion.

- we create a multi-resolution pyramid of images
- At each level, we just store the differences between the image
- We can reconstruct the image by just adding up all the residuals

3.1.2 Discrete cosine transform (DCT)-

Discrete cosine transform (DCT) is widely used in compression and image processing. In DCT most of the visually significant information about the image is concentrated in just a few coefficients of the DCT. For example, the DCT is at the heart of the international standard lossy image compression algorithm known as JPEG. DCT separates the images into parts of different frequencies. During quantization, the actual compression occurs, and the less important frequencies are discarded. Only the important frequencies that remain are been retrieved for further processing. DCT is applied to every non-overlapping block of the image. The type-II DCT and its inverse in one dimension are given by the following equation [8]:

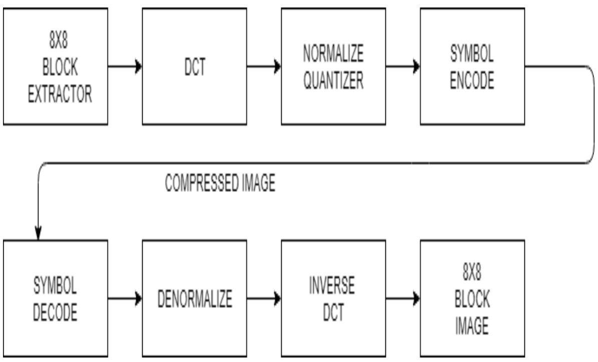


Fig 4: Discrete cosine transform

STEPS

- The original image is divided into blocks of 8 x 8.
- The pixel values within each block range from [-128 to 127] but pixel values of a black and white image range from [0-255] so, each block is shifted from [0-255] to [-128 to 127].
- The DCT works from left to right, top to bottom thereby it is applied to each block.
- Each block is compressed through quantization.
- The array of compressed blocks that constitute the image is stored in a drastically reduced amount of space.
- Compressed image is reconstructed through the reverse process.
- This process uses the inverse Discrete Cosine Transform (IDCT).

3.1.3 HUFFMAN CODING-

Huffman coding is a form of statistical coding which reduces the number of bits required to represent a string of symbols. Huffman code is widely used to increase coding efficiency. The goal of this algorithm is to allow symbols to vary in length. It uses the Huffman source-coding algorithm to generate the uniquely decipherable Huffman code with a minimum expected codeword length when the probability distribution of a data source is known to the encoder [9]. Shorter codes are assigned to the most frequently used symbols, and longer codes to the symbols which appear less frequently in the string. Huffman coding is based on the frequency of occurrence of a data item in pixel. Its principle is to use a lower number of bits to encode the data that occurs more frequently. Codes are stored in a Code Book which may be constructed for each image or a set of images. In all cases, the codebook plus encoded data must be transmitted to enable decoding. By using this Huffman algorithm, we can be able to design the most efficient compression method. Huffman Coding comes under lossless technique here in lossless compression no information is lost during Image Compression.

3.2 Principal component analysis (PCA)

Principal Component Analysis is a statistical technique that uses mathematical principles to transform a large possibly correlated variable into a small variable called principal components. The advantage of PCA is the patterns in the data and compressing data by reducing the number of dimensions without losing any information. The image dimension reduction technique by PCA is consolidated into four major steps:

- image normalization
- finding the covariance matrix of the image data
- calculate the eigen vectors and eigen values of the covariance matrix and
- lastly transforming image data into new basis

4 RESULT AND ANALYSIS

The result of different compression techniques used in a digital image is shown below. The compression ratio of different techniques is found out using the given formula as

$$Compression\ Ratio = \frac{size\ of\ actual\ image}{size\ of\ compressed\ image}$$

Mean square error (MSE) is one of the parameters to evaluate the quality of the compressed image. If the value of MSE is less, then the quality of the compressed image will increase [7].

$$MSE = \frac{1}{M}\sum_{X=1}^{M}\sum_{Y=1}^{N}[f(X,Y) - f'(X,Y)]^2$$

Where f(x, y) is the original input image, f'(x, y) is compressed image and M, N are the dimensions of the images. PSNR is ratio between sizes of the input image to the square of Mean Square Error (MSE). If PSNR is high then the quality of compressed image is also increased [7].

$$PSNR = 10log_{10}\left[\frac{M*N}{MSE}\right]$$

Where, MxN is the size of an input image.

In PCA component1 we extract large information thus the resulting image has data that can be further used, in PCA component 2 the rest of the required information are been extracted, which gives a bit blurry result, and PCA component3 extract the rest of the remaining information. the result of wavelet transform whose compression ratio is less than the compression ratio of PCA thus we can say that PCA gives better compression than Wavelet transform. The compression ratio of DCT is less than WT and PCA thus we can say WT and PCA are better than DCT. The compression ratio of the hybrid compression technique is greater than the other above shown compression ratios. The compression ratio of this hybrid technique is then compared with the compression ratio of PCA and it is found that hybrid technique gives a higher compression ratio hence it goes better than PCA and the other compression techniques.

TECHNIQUES	ORIGINAL VALUE	COMPRESSED VALUE	COMPRESSION RATIO	MEAN SQUARE ERROR(MSE)	PEAK SIGNAL-TO-NOISE RATIO(PSNR)
PCA	47609	6856	6.944	10756.452	11.5698
DWT	47609	11619	4.097	8389.065	12.649
DCT	47609	15346	3.102	6741.544	13.599
HYBRID	47609	6492	7.333	10949.461	11.493

Table 1: Result and Analysis

Table 1, shows the value of an image in its original form, compressed form, its compression ratio, MSE, and PSNR value. Figure 5 shows the comparison between the CR and PSNR values that are been mentioned in the above table. The figure clearly shows that the compression ratio of Hybrid is greater than that of PCA, DCT, and DWT. Although the PSNR value of hybrid is less than that of DCT and DWT it is almost equal to the value of PCA. So, it can be said that hybrid technique gives better compression ratio than the other technique mentioned above.

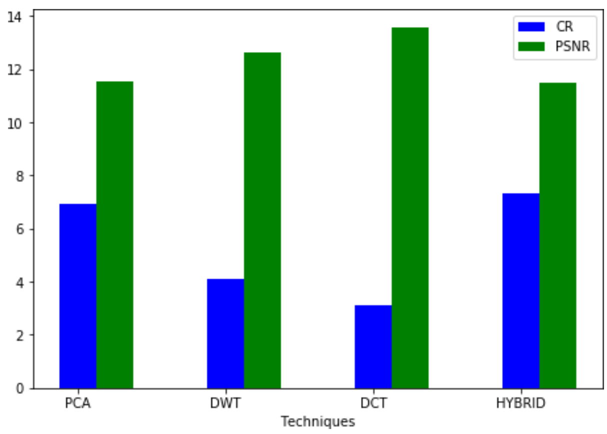


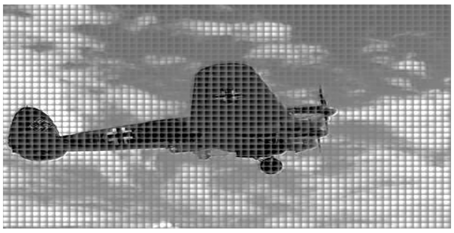
Fig 5: Comparison of hybrid technique with other techniques

a) Original

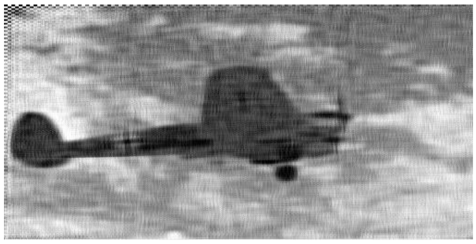
b) Grayscale

c) Dimension reduction using PCA

d) Compressed Image using Discrete wavelet transform



e) Compressed Image using Discrete cosine transform



f) Compressed Image using Hybrid Technique

Fig6: (a)Original, (b)Grayscale, (c)Dimension reduction using PCA, (d)Compressed image using DWT, (e)Compressesd image using DCT, (f)Compressed image using Hybrid technique

Figure 6 shows the result of the PCA technique, DWT, DCT, and Hybrid technique. (a), (b) of Figure 6 shows the original digital image which is given as input and the grayscale image of the original image. In (c), (d), and (e) the result of PCA, DWT, DCT are been shown, (f) shows the result of the hybrid technique used in an image where the grayscale image is given as an input to perform DWT and DCT compression. Huffman coding is applied to the image after DCT and DWT, which gives a resultant image as shown in (f) which is the final result of the hybrid compression technique.

5 CONCLUSION

In this work, the problem of the storage capacity of digital images and the ways to reduce it, by reducing the actual size of the image without losing its actual quality has been discussed. The study and reviews of the existing approaches for dimension reduction of high dimensional image data have also been done. The two-dimension reduction methods are feature selection and feature extraction. It is explored that the most used techniques for dimension reduction are principal component analysis (PCA). Finally, in our work, a hybrid compression method is been shown, which is a combination of three compression schemes DWT, DCT, and Huffman compression, and its compression ratio is compared with standalone DWT, DCT, and PCA.It shows DWT and DCT are very good to cope with compression ratio, but as they are lossy techniques so our quality measurement is decreased so, further to enhance the compression ratio a Huffman compression method is used because of its lossless compression nature. After applying lossy techniques it's better to use lossless. So, from the above experimental result, we can say the hybrid technique gives a better compression ratio than PCA, WT, and DCT. We can also say that PSNR of Hybrid is greater than that of DWT and DCT, and almost equal to the PSNR of PCA. PSNR and CR are traditional methods to estimate the quality of the image after any modification, but the further improvement over these traditional methods can be done by using a perception-based model SSIM.

REFERENCES

[1] S. C. Ng, 2016, Principal component analysis to reduce dimension on digital image 8th International Conference on Advances in Information Technology, IAIT2016, 19-22.

[2] Shereena V.B. and Julie M.David, 2015, SIGNIFICANCE OF DIMENSIONALITY REDUCTION IN IMAGE PROCESSING, Signal & Image Processing : An International Journal (SIPIJ) Vol.6, No.3.

[3] Sarabjot Singh Grewal, Er.Beant Kaur, 2017, Full Resolution Image Compression with Deep Neural Net- works, International Journal for Research in Applied Science & Engineering Technology (IJRASET)ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887.

[4] Jigna J.Patel, Dr.S & S.S.Ghandhy GEC, 2017, Importance of Dimensionality Reduction in Image Pro- cessing, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 8 21 – 24.

[5] A.H.M.Jaffar Iqbal Barbhuiya, Tahera Akhtar Laskar, K. Hemachandran, 2014, An Approach for Color Image Compression of JPEG and PNG Images using DCT And DWT, Sixth International Conference on Computational Intelligence and Communication Networks.

[6] Monica Benito, Daniel Peña, 2005, A fast approach for dimensionality reduction with image data, www.elsevier.com/locate/patcog, Pattern Recognition 38 (2005) 2400 – 2408.

[7] Uday Bhade, Sanjeet Kumar, Prashant Dwivedy, Shahbaz Soofi, Avinash Ray, 2017, Comparative study of DWT, DCT, BTC and SVD Techniques for Image Compression, Proceeding International Conference on Recent Innovations is Signal Processing and Embedded Systems (RISE-2017) 27-29.

[8] Lei Wang, Jiaji Wu, Licheng Jiao, Li Zhang and Guangming Shi, 2008, LOSSY TO LOSSLESS IMAGE COMPRESSION BASED ON REVERSIBLE INTEGER DCT, 978-1-4244-1764-3/08.

[9] R.Praisline Jasmi, Mr.B.Perumal, Dr.M.Pallikonda Rajasekaran, 2015, COMPARISON OF IMAGE COMPRESSION TECHNIQUES USING HUFFMAN CODING, DWT AND FRACTAL ALGORITHM, 978-1-4799-6805-3/15/ IEEE, 2015 International Conference on Computer Communication and Informat- ics (ICCCI -2015), Jan. 08 – 10, 2015.

[10] XiHongZHOU, 2011, Research on DCT based Image Compression Quality, 2011 Cross-Strait Quad-Regional Radio Science and Wireless Technology Conference.

[11] ZHAO Zhongwen, GUO Huanghuang, 2017, Visualization Study of High-dimensional Data Classification Based on PCA-SVM, IEEE Second International Conference on Data Science in Cyberspace.