International Journal of Computer Applications (0975 – 8887)
Volume 184 – No. 45, February 2023

Offline Signature Verification based on Ensemble of Features using Support Vector Machine

Sunil Kumar D.S., PhD
Department of Computer Applications
Administrative Management College,
Bangalore, India

ABSTRACT
In this paper, we propose a novel feature representation technique namely Edge Histogram and 4 Directional Histogram for offline signature verification system. Edge is a curve or point where the intensity of an image changes rapidly. Edges represent the boundary of object of an image. Edge detection is a process of detecting edges of an image. Several algorithms are available to detect edges effectively from an image. Canny, Roberts, Prewitt and Sobel are several popular available edge detectors. In our approach we used Sobel operator for edge detection. We also applied radon transform on signature samples and obtained fractal properties with the help of box-counting method. Finally fusion of these features forms a feature vector. We employed Support vector machine for classification. Experiments are conducted on bench mark dataset namely CEDAR and GPDS. The obtained experimental results exhibit the performance of the proposed method.

General Terms
Offline Signature Verification, Behavioral Biometrics.

Keywords
Signature Verification, Support Vector Machine; Classification, Edge Histogram, Edge Directional Histogram.

1. INTRODUCTION
Signatures are behavioral biometric traits of a person, used to authenticate a person. In all the legal transactions and legal documents signature is required to authenticate its legality. In such cases there are chances to forge the signature by other person to get the benefits. Therefore in order to check the genuineness of the signature, signature verification system is needed. In the state-of-the art literature there are several algorithms are proposed by different authors but still few challenges are remained to address such as the detection of skilled forgery and detection of intraclass variations. There are two types of signature verification system namely offline and online. The main difference in offline and online signature verification is in its signature acquisition method. In offline signature verification, signature samples are first written on a paper, which is then scanned using scanner and preprocessed after that fed for verification. In case of online signature verification signatures are collected using an electronic gadgets which are capable of reading and storing dynamic information like pen stroke, pen pressure, velocity, pen-up, pen-down, azimuth etc.,. The aim of signature verification is to discriminate genuine signature from forge signature. Forgeries are of three types first one is simple forgery, second one is random forgery and third one is skilled forgery. The main challenge in signature verification is to discriminate skilled forgery, because in this case the forger practices writing signature similar to genuine signature over a period and then forges. Another challenge in signature verification is intra class variation. In intra class variation the same user used to write signature with slight variations under different conditions like sickness, tiredness or with old age problems.

The order of this paper with section wise is as follows: section-2 presents a detailed literature of some of the state-of-the-art papers. The elaborated description of the proposed approach is in section-3. Section-4 illustrates the experimental results followed by the discussion. Finally, section-5 represents the conclusion part.

2. LITERATURE
Now a day's biometric applications are widely used in day to day life. Signatures are one of the behavioral biometric widely accepted as a means of authentication of a person. Biometrics has strong ability to discriminate a person's original signature from fake. This leads to development of various algorithms to recognize genuine from forge, though few challenges remain unsolved. This triggers developing new algorithms to increase the accuracy rate. We have listed few research works here. Calik et al., [1] proposed a new convolutional neural network (CNN) model called Large-Scale Signature Network (LS2Net). This approach is aimed to address the problem of small number of signature samples to train the model from large dataset. Authors introduce Class Center based Classification (C3) to classify embedded features. C3 uses class centers which are obtained by averaging in-class features. Among these class centers, 1-nearest neighbor classifier is derived as classification task. Authors also addressed Large-Scale recognition problem, by comparing ReLU and Leaky ReLU. The influence of Leaky ReLU on the performance of network is examined. Along with the addition of the C3 (Class Center for Classification) algorithm, the default network is defined as LS2Net + BN + C3 called as LS2Net_v2. Bilal et al., [2] proposes fusion of two methods one is Curvelet Transform (CT) and another is One-Class Principal Component Analysis (OC-PCA) for Open Handwritten Signature Identification System. Asyrofa et al., [3] proposes back propagation method of Neural Networks. The architecture includes input layer, hidden layer and output layer. The input layer takes the input data, hidden layer processes the data and finally output comes at output layer. If the output obtained at output layer having higher error rate then it can be propagate back to the previous layers to minimize the error by adjusting the weights of the nodes of the hidden layers, where the data processes again and gives the result at output layer. This process repeats until the desired output obtained with minimum error rate. Anwar et al, [4] proposed a Back Propagation Artificial Neural Network Matching Technique for offline signature verification. The proposed model consists preprocessing phase, codebook

generation phase and matching phase. Rohith kumar et al., [5] proposed Statistical-ANN Hybrid Technique for offline signature recognition and verification. The model involves moment invariant method and artificial neural network. The model consist two separate neural networks one is for signature recognition and another for signature verification. The dual hybrid approach comprises statistical based features extraction for signature recognition and back propagation neural network for signature verification. Luiz G Hafemann et al., [6] presents learning features formulations for offline signature verification. Learning features are used to train writer independent classifier using convolution neural networks. Grid based template matching method by Elias N Zois et al., [7] uses the geometric pattern of a signature, which is encoded by grid templates, apparently partitioned as subsets. Chibani et al., [8] presents Artificial Immune Recognition System for offline signature verification. The proposed method uses two descriptors one is gradient local binary patterns to estimate gradient features from the neighboring local binary patterns and another is longest run feature to describe the signature topology. Score level fusion of classifiers proposed by Yilmaz et al., [9] This approach extracts set of features namely, Scale Invariant Feature Transformation (SIFT), Histogram of Oriented Gradients (HoG) and Local Binary Patterns (LBP). Radhika et al., [10] proposed a combined approach of both offline and online signature verification. Author extracts features such as pen tip tracking from online signatures. Gradient features and projection profile features are extracted from offline signatures. Experiments are conducted separately. The well-known classifier Support Vector Machine is employed for classification. Results obtained are combined and verified. Yilmaz et al., [11] proposed a method where signature samples are partitioned in to different zones based on both polar and Cartesian coordinate systems. From different zones of both coordinate system histograms namely Histogram of local binary patterns as well as histogram of oriented gradients (HoG) is obtained. Classification has been done by employing Support Vector Machine (SVM).

3. PROPOSED METHOD

We propose a novel approach for signature verification based on edge histogram and 4 directional histogram using Support Vector Machine classifier. This system will use static as well as dynamic features for verification. The static features include moment features and 4direction distribution, while the dynamic features include gray distribution and stroke width distribution. At last support vector machine is used to classify the signature. The flow of the system can be seen briefly in the figure1:In the image acquisition stage we collected publically available bench mark data bases namely CEDAR and GPDS Synthetic offline Signature database. These databases consists grey scale images.

3.1 Preprocessing

After acquiring signature images we applied the following preprocessing techniques in order to get better accuracy.

3.1.1. Image Binarization:

The acquired image sample may be in different color model like RGB, gray scale or binary. The database we collected are CEDAR and GPDS, which consist grayscale signature images. We converted grey scale images in to binary images. The advantage of converting RGB color image or grayscale image in to binary scale image is to reduce the processing time as well as it will consume less storage space because the intensity of the image will be either 0 or 1. There are several algorithms are available to convert, we employed Otsu's method for binarization.

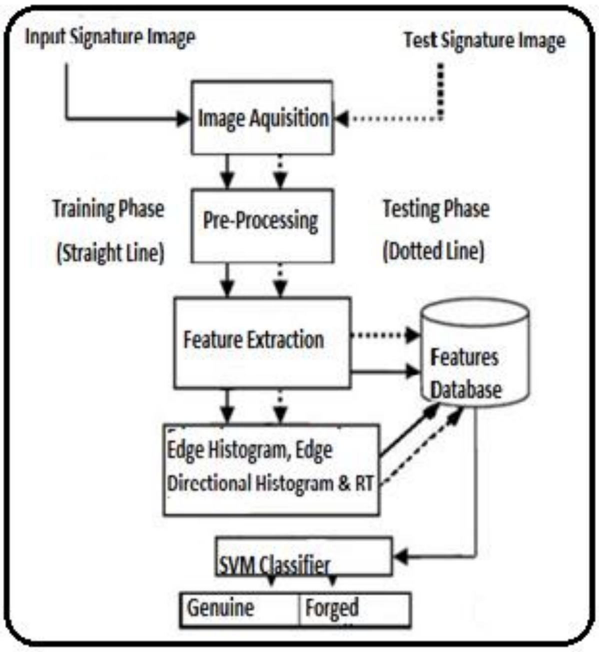

Figure 1:Offline Signature Verification Model

3.1.2. Noise Reduction

It's the process of removing noise from the image. It can be done by using median filters. The widely used noise type is salt and pepper noise. The noise present in the image is certainly degrading the image and it's difficult to extract the exact features. In order to extract the useful information from the image it's necessary to remove the noise. Gaussian median filter is applied to remove noise from the image.

3.1.3 Data Area Cropping

Image cropping is the processes where the image data area is extracted from the background area. Usually offline signature samples are acquired using a paper sheet; signatures are not spread across whole paper but on some portion of the paper. The data sample is the our region of interest, so extracting only data area from the background and processing it, will helps in getting better accuracy.

3.1.4. Width Normalization

As we know signature samples are collected from different contributors, so obviously there will be more variations in the sample signatures, during preprocessing stage its necessary to normalize the data samples not only scaling but also its width. Width is one of the local feature and varies from sample to sample. Normalized width will helps to get better accuracy.

3.1.5 Image Thinning

It's the process of keeping one pixel width information and removing redundant pixels. The collected samples are written on a paper, which having different pen width. Thinning operation makes uniform pixel width and also reducing pixel width will minimize the processing time.

After preprocessing the signature samples, we obtained edge histogram features. The edge histogram and edge directional histogram are mainly meant to extract the texture feature of the signatures. Texture is spatial intensities of an image. It is calculated by obtaining the gradient of pixels that is the maximum rate of change of co-ordinates (x, y), in the five directions with a threshold value of 100 as given in below equation (1). Where, G_x and G_y are gradient vectors in x and y direction.

$$\text{Theta} = \arctan(G_x/G_y) \quad (1)$$

Edge histogram descriptor is a well-known method to detect edges of an image. It represents frequency of occurrence of edges in five different types from an image block namely vertical edge, horizontal edge, 45^0 edge, 135^0edge and non-directional edge.

Edge Direction Histogram (EDH) is the histogram obtained from each of the images represents the frequencies of occurrences of five classes of edges in the corresponding images for texture extraction. The edge direction histogram uses the Sobel operator which helps in capturing the spatial distribution of edges in four directions (0˚,-45˚, 45˚and 90˚) with filter mask with Sobel operator in X and Y direction. Sobel operator is considered as edge descriptor which extracts an edge image from original image. Sobel operator applies 2-D spatial gradient on an image and considers high frequency on edges. Sobel operator is mainly applied to obtain absolute magnitude of intensity from an image.

We also extracted radon transform features from the images. For an image f(x, y) radon transform is the projection of image intensity along a radial line at a specific angle theta. It can be calculated using the following formula (2) which is available in MATLAB library.

R=radon (I, theta) (2)

Where, R is the omitted radon transform of the image I for an angle theta degree. The theta degree by default ranges from 0 to 179. The resulting projection is the sum of the intensities of the line in each direction. The box-counting method uses this radon transform features to compute fractal dimension. In box- counting method the pixels of an image are distributed over a grid based structure and computing the number of pixels per box.

Finally the features of edge histogram, edge directional histogram and radon transform are fused to form a feature vector. Euclidian distance metric is used to measure the similarity.

3.2 Classification:

Let’s have an insight in SVM. Support Vector Machine (SVM) is bilinear classifier used to classify two class data. But this can also be used to multiclass classification based on one versus rest paradigm, where each class is compared with rest of all other classes. SVM can be employed to classify non-linear objects using kernel trick. SVM creates a hyper plane in a high dimensional space to separate objects. A good separation is achieved when a large margin is exists, which is shown in figure 2. Signatures are typically represented by sparse vectors under the vector space model. When training classifiers on large collections of signature, both the time and memory requirements connected with these vectors may be prohibitive. This calls for the use of a feature selection method not only to reduce the number of features but also to increase the sparsity of vectors. We propose a feature selection method based on linear Support Vector Machines (SVMs). Linear SVM is used on a subset of training data to train a linear classifier which is characterized by the normal to the hyper plane dividing positive and negative instances. This is calculated using the equation 3.

$$f(x) = sign(w.X + b) \quad (3)$$

Here w → weight factor

b → threshold

X → input patterns

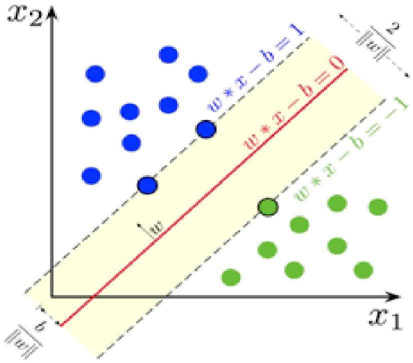

Figure 2. Block diagram of Support Vector Machine functionality [23].

Components of the normal with higher absolute values have a larger impact on data classification. Instead of predefining the number of highest scoring features to be included in a classifier we apply feature selection that aims at a predefined average sparsity level across image and classifiers for a given training set. After the feature set is determined, the model is trained on the full training data set represented within the selected feature set. The test signature is then, based on its value for the parameters from the feature set, is mapped and classified as “GENUINE” or “FORGED”.

4. EXPERIMENTAL RESULTS

The experiments are carried out in MATLAB 2016a software environment on Asus laptop with 4GB RAM, 500GB Hard Disk and i7 processor on windows 8 operating system. The following figure.3 is the snapshot of Graphical User Interface for Offline Signature Verification developed using GUID tool available in MATLAB.

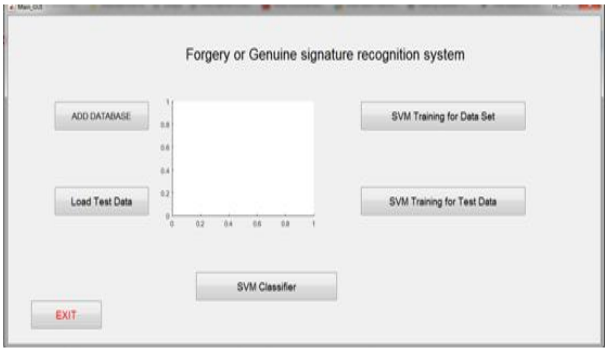

Figure. 3 Offline Signature Verification GUI

Experiments are conducted on publicly available benchmark datasets namely CEDAR (Center for Document Analysis and Recognition) which is developed by SUNNY Buffalo and GPDSSyntheticSignature database. The knowledge repository contains the Edge histogram and Edge directional histogram features extracted from every signature sample of the data set including both genuine and skilled forge signature samples. For each dataset, the signature samples are considered into two groups: training sample set and testing sample set with varying number of samples. We have carried out four sets of experiments. In Set-1, first ten genuine and first ten skilled forgeries are chosen as training samples and tested against the remaining samples of the respective datasets, where as in Set-2, we have considered first 15 samples of genuine and first 15 samples of skilled forgery for training and tested with remaining samples of the dataset. In Set-3,we have randomly chosen 10 genuine and randomly chosen 10 forge samples are considered for training, and tested with the remaining samples of the dataset, and in Set-4,there are 15 samples are chosen randomly from the respective datasets for training and remaining samples are considered for testing. In order to overcome from the effect of the randomness, Set-3 and Set-4 experimentations are repeated five times and the average result is tabulated. Signature database. The knowledge repository contains the Edge histogram and Edge directional histogram features extracted from every signature sample of the data set including both genuine and skilled forge signature samples. For each dataset, the signature samples are considered into two groups: training sample set and testing sample set with varying number of samples. We have carried out four sets of experiments. In Set-1, first ten genuine and first ten skilled forgeries are chosen as training samples and tested against the remaining samples of the respective datasets, where as in Set-2, we have considered first 15 samples of genuine and first 15 samples of skilled forgery for training and tested with remaining samples of the dataset. In Set-3,we have randomly chosen 10 genuine and randomly chosen 10 forge samples are considered for training, and tested with the remaining samples of the dataset, and in Set-4,there are 15 samples are chosen randomly from the respective datasets for training and remaining samples are considered for testing. In order to overcome from the effect of the randomness, Set-3 and Set-4 experimentations are repeated five times and the average result is tabulated.

Table1. Datasets Details

DATABASE	NO.OF USERS	NO.OF GENUINE	NO. OF FORGE	TOTAL NO.
GPDS Synthetic Signature (Offline)	4000	24	30	216000
CEDAR	55	24	24	2640

4.1 Experiments on CEDAR dataset:

CEDAR (Center for Document Analysis and Recognition) is a benchmark offline signature database, publicly available. We considered CEDAR database for our experiments. CEDAR database consists 24 genuine and 24 forgeries from 55 contributors. The total number of signature samples is 2640. The details of databases are tabulated in the table 1. Out of 2640 samples, we considered few samples for our experiments. We conducted experiments in 4 sets. In first and third set we considered 10 genuine samples and 10 skilled forgeries for training. And for testing we choose 14 genuine samples and 14 skilled forgeries for testing. In second and fourth set we considered 14 genuine and 14 skilled forgeries for training. And for testing 9 genuine samples and 9 skilled forgeries samples for testing. To avoid randomness set second and fourth are repeated 5 times and average is considered. In table 2 the results on CEDAR database is tabulated with FRR and FAR are performance evaluation metrics.

From the literature, we found the experimental results of few well known approaches on CEDAR dataset. The comparative analysis presented in table 3 shows the improvements in accuracy by the proposed approach.

Table 2. Results obtained for CEDAR

Experimental set-up	Accuracy	FRR	FAR
Set-1	96.0	4.2	5.8
Set-2	97.22	3.8	4.3
Set-3	95.24	6.0	6.2
Set-4	96.42	4.0	5.2

Table.3 Experimentation Results obtained for CEDAR Dataset - A comparison:

Proposed by	Classifier	Accuracy	FAR	FRR
Kalera et al. [12]	PDF	78.50	19.50	22.45
Chen and Shrihari [13]	DTW	83.60	16.30	16.60
Kumar et al., [14]	SVM	88.41	11.59	11.59
Shekar et al. [15]	EMD	91.06	10.63	9.4
Kumar et al. [16]	MLP	91.67	8.33	8.33
Kumar et al.	SVM	92.73	6.36	8.18
Proposed approach	SVM	97.22	3.8	4.3

B. Experiments on GPDS Synthetic Offline Signature dataset:

Digital Signal Processing Group (GPDS) of the Universidad de Las Palmas de Gran Canaria, Spain had developed a large scale signature corpus; it included with many sub corpus like GPDS-960, GPDS-300 and GPDS-160. For our experiments we considered GPDS Synthetic Offline Signature dataset, which is a well-known benchmark dataset. It can be collected by communicating with the concern developer. The GPDS Synthetic offline signature dataset consist 4000 contributors, each contributor contributed 24 genuine signature samples and 30 forge signature samples. This result a total number of signatures are 216000 which is tabulated in table 1.

The experiments are conducted on GPDS dataset on 4 sets. For set-1 and set-2, we considered 10 genuine and 10 forge samples for training. For testing we considered 14 genuine and 20 skilled forgery samples. For set-2 and set-4 we considered 15 genuine and 15 skilled forgery samples for training. We tested against 9 genuine and 15 skilled forgery samples. The obtained results are tabulated in table 4.

From the literature, we found the experimental results of few well known approaches on GPDS dataset. The comparative analysis presented in table 5 shows the improvements in accuracy by the proposed approach.

Table 4. Results obtained for GPDS

Experimental set-up	Accuracy	FRR	FAR
Set-1	95.32	3.8	5.2
Set-2	94.80	4.6	6.8
Set-3	96.52	3.2	5.0
Set-4	94.0	4.2	6.2

Table 5. Experimentation Results obtained for GPDS Dataset - A comparison:

Proposed by	Classifier	Accuracy	FAR	FRR
Ferrer et al., [18]	SVM HMM	86.65 -	13.12 12.60	15.41 14.10
Vargas et al., [19]	SVM+LBP	87.28	6.17	22.49
Ruiz-Del-Solar et al., [20]	Bayseian	84.70	14.20	16.40
Kumar et al., [21]	MLP	86.24	13.76	13.76
Shekar et al	EMD	91.06	10.63	9.4
Proposed Approach	SVM	96.52	3.2	5.0

.5. CONCLUSION

In this work we are trying to validate whether a signature sample is forged or not using support vector machine. We have acquired the signature samples of both genuine and forge from different individuals. Followed by pre-process using techniques like Binarization, complementation, add noise, remove noise, cropping, thinning. Further from these pre-processed signatures we have extracted features of edge histogram, edge directional features and radon transform. Fusion of these extracted features forms a feature vector, which is then passed to support vector machine for classification. From the experimental demonstration of results, we have improved signature recognition accuracy in terms of recognition rate is 97.22% and 96.52% for databases CEDAR and GPDS respectively. The future course of work will be applying deep learning models for classification.

6. REFERENCES

[1] Calik.N., Kurban, O.C., Yilmaz,A.R., Yild irim, T., & Ata, L.D. (2019). Large–Scale offline signature recognition via deep neural networks and feature embedding. Neurocomputing 359, 1-4.

[2] Hadjadji, B., Chibani ,Y., & Nemmour H.(2017). An Efficient Open System for Offline Handwritten Signature Identification based on Curvelet Transform and One-Class Principal Component Analysis. Neurocomputing, 265, 66-77

[3] Rahmi,A. Wijayaningrum , V.N., Mahmudy,W.F., & Parewe,A.M.(2016). Offline Signature Recognition using Back Propagation Neural Network. Indonesian Journal of Electrical Engineering and Computer Science, 4(3), 678-683

[4] Ebrahim,A.Y., & Sulong, G.(2014). Offline Handwritten Signature Verification using Back Propagation Artificial Neural Network Matching Technique. Journal of Theoretical and Applied Information Technology, 65(3).

[5] Kumar,R., Kaur,S., & Gupta,B. Signature Recognition using Statistical-ANN Hybrid Technique. International Journal of Scientific Research and Engineering Development –Volume 2 Issue 2 Mar-Apr 2019.

[6] Hafemann, L. G., Sabourin, R., & Oliveira, L. S. (2017). Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recognition*, *70*, 163-176.

[7] Zois, E. N., Alewijnse, L., & Economou, G. (2016). Offline signature verification and quality characterization using poset-oriented grid features. *Pattern Recognition*, *54*, 162-177

[8] Serdouk, Y., Nemmour, H., & Chibani, Y. (2016). New off-line handwritten signature verification method based on artificial immune recognition system. *Expert Systems with Applications*, *51*, 186-194.

[9] Yılmaz, M. B., & Yanıkoğlu, B. (2016). Score level fusion of classifiers in off-line signature verification. *Information Fusion*, *32*, 109-119.

[10] Radhika, K. S., & Gopika, S. (2015). Online and offline signature verification: a combined approach. *Procedia Computer Science*, *46*, 1593-1600.

[11] Yilmaz, M. B., Yanikoglu, B., Tirkaz, C., & Kholmatov, A. (2011, October). Offline signature verification using classifier combination of HOG and LBP features. In *2011 international joint conference on Biometrics (IJCB)* (pp. 1-7). IEEE.

[12] Kalera, M. K., Srihari, S., & Xu, A. (2004). Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, *18*(07), 1339-1360.

[13] Chen, S., Srihari, S.: Use of exterior contours and shape features in off-line signature verification. In: ICDAR, pp. 1280–1284 (2005).

[14] Kumar, R., Kundu, L., Chanda, B., Sharma, J.D.: A writer-independent off-line signature verification system based on signature morphology. In: Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia IITM ’10, pp. 261–265. ACM, New York, NY, USA (2010)

[15] Shekar, B.H., Bharathi, R.K., Pilar, B.: Local morphological pattern spectrum based approach for off-line signature verification. In: International Conference on Pattern Recognition and Machine Intelligence, pp. 335–342. Springer (2013)

[16] Kumar, R., Sharma, J., Chanda, B. (2012): Writer-independent off-line signature verification using surroundedness feature. Pattern recognition letters. **33**(3), 301–308

[17] Kumar, M.M., Puhan, N.: Inter-point envelope based distance moments for offline signature verification. In: 2014 International Conference on Signal Processing and Communications (SPCOM), pp. 1–6. IEEE (2014)

[18] Ferrer, M., Alonso, J. and Travieso, C., 2005. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27(6), pp. 993–997.

[19] Vargas, J., Ferrer, M., Travieso, C. and Alonso, J., 2011. Offline signature verification based on grey level information using texture features. Pattern Recognition 44(2), pp. 375 – 385.

[20] Ruiz-Del-Solar, J., Devia, C., Loncomilla, P. and Concha, F., 2008. Offline signature verification using local interest points and descriptors. In: Proceedings of the 13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications, CIARP '08, Springer-Verlag, Berlin, Heidelberg, pp. 22–

[21] https://en.wikipedia.org/wiki/Support-vector_machine

[22] Pilar, B., Shekar, B. H., & Sunil Kumar, D. S. (2019). Multi-scale Local Binary Patterns-A Novel Feature Extraction Technique for Offline Signature Verification. In Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part III 2 (pp. 140-148). Springer Singapore.

[23] Parameshachari, B. D., & Kumar, D. S. (2022, October). SVM Based Brain Tumor Detection and Classification System. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) (pp.1-4).IEEE.