IRJCS: International Research Journal of Computer Science
Volume 12, Issue 05, May 2025
ISSN: 2393-9842
https://www.irjcs.com/archives

Implementation of Algorithm for Matching E-Signature by Machine Learning

Harsh Kumar Sinha, Deepak Mangal, Furqan Khan, Atul Saurabh
Department of Information Technology,
Greater Noida Institute of Technology (Engg. Institute),
Greater Noida, India
Harshsinha6203@gmail.com, deepakmangal.it@gniot.net.in
furqankhan63062716@gmail.com, atulsaurabh1@gmail.com
Dr.Vikas Singhal
Professor & Head, Department of Information Technology,
Greater Noida Institute of Technology (Engg. Institute),
Greater Noida, India

Publication History
Manuscript Reference: IRJCS/RS/Vol.12/Issue05/MYCS10081|ResearchArticle | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.12/Issue05/MYCS10080 Received: 26, April 2025, Revised: 02, May 2025 Accepted: 12 May 2025 Published Online: 15 May 2025 http://www.irjcs.com/volumes/Vol12/iss-05/02.MYCS10081.pdf
Article Citation:Dr.Vikas,Harsh,Deepak,Furqan,Atul(2025). Implementation of Algorithm for Matching E-Signature by Machine Learning, IRJCS: International Research Journal of Computer Science, Volume 12, Issue 04 of 2025 pages 205-210 **doi:>** https://doi.org/10.26562/irjcs.2025.v1205.02
BibTeX Dr.Vikas@2025Implementation

Copyright: ©2025 This is an open access article distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: In the digital age, electronic signatures (e-signatures) have become an essential tool for verifying identity and ensuring the authenticity of documents in various domains such as finance, legal systems, healthcare, and e-governance. However, the increasing volume of digital transactions has also raised concerns regarding the security, reliability, and accuracy of e-signature verification systems. This research focuses on the development and implementation of a robust machine learning-based algorithm for matching e-signatures, with the goal of improving verification accuracy and minimizing the risk of forgery.The proposed system leverages a multi-stage approach that includes classification, feature extraction, and preprocessing. In the beginning, pre-processing techniques such as image normalization, noise filtering, and binarization are applied to enhance signature quality and ensure uniformity across samples. Subsequently, feature extraction is performed using a combination of statistical, structural, and methods based on deep learning, such as convolutional neural networks to capture both global and local signature characteristics. These features are then used to train classification models such as Support Vector Machines (SVM), Random Forest, and deep learning classifiers, which are capable of differentiating between genuine and forged signatures with high precisionUsing publicly accessible e-signature datasets, metrics like accuracy, precision, recall, and F1score are used to assess the algorithm's performance. Comparative analysis with existing methods demonstrates the superiority of the proposed model in terms of both verification accuracy and computational efficiency. Additionally, the system shows resilience to intra-class variations, such as changes in signing speed, angle, and pressure, which are common in real-world scenarios.

1. INTRODUCTION

With the rapid digitization of services across sectors such as banking, healthcare, legal, and e-governance, the need for secure and efficient digital identity verification mechanisms has become more critical than ever. Among various biometric authentication methods, electronic signatures (e-signatures) have emerged as a widely accepted tool for validating documents and authorizing transactions in a virtual environment. Unlike traditional handwritten signatures, e-signatures offer the advantage of speed, convenience, and scalability [1]. However, the digital nature of e-signatures also makes them vulnerable to tampering, misuse, and forgery, which poses significant challenges to their authenticity and legal validity.To address these concerns, intelligent verification systems capable of accurately matching e-signatures are required. Traditional signature verification techniques primarily rely on rule-based or heuristic approaches, which often lack the adaptability to handle the diverse variations in individual writing styles[3],[4].

These methods also struggle with differentiating between genuine signatures and skillfully forged ones, especially when confronted with intra-class variability such as differences in stroke thickness, writing speed, angle, and noise introduced during digital capture.Machine learning, particularly deep learning, has revolutionized the field of pattern recognition and image classification in recent years. Its ability to learn complex representations from raw data makes it an ideal candidate for biometric verification tasks, including signature recognition. This research explores the application of machine learning techniques to the problem of e-signature matching, aiming to develop an automated, accurate, and scalable verification system.The proposed study focuses on implementing a robust algorithm that integrates pre-processing techniques, effective feature extraction methods, and classification models to distinguish between authentic and forged e-signaturesTo determine how well it performs under actual conditions, the system is trained and tested on datasets that are made available to the public.

ISSN: 2393-9842
https://www.irjcs.com/archives

The model's ability to learn spatial hierarchies of features from the input signatures is improved when convolutional neural networks (CNNs) are used, and traditional classifiers like Support Vector Machines (SVM) and Random Forests provide a performance comparison. The following is the structure of this paper: Section 2 reviews the related work and existing signature verification methods; Section 3 explains the proposed method, including data preprocessing, feature extraction, and classification techniques; The results of the experiments and a performance evaluation are presented in Section 4, and insights and recommendations for the future are provided in Section 5.In today's digitally driven world, the adoption of electronic signatures (e-signatures) has become increasingly prevalent across various industries, including banking, healthcare, legal services, education, and e-governance. E-signatures serve as a digital counterpart to handwritten signatures, enabling individuals and organizations to approve, verify, and authenticate documents and transactions remotely. Their use significantly enhances workflow efficiency, reduces paper-based documentation, and supports environmental sustainability. However, the shift to digital processes also brings new security challenges, including identity fraud, document tampering, and forgery. These challenges underscore the urgent need for reliable, secure, and automated e-signature verification systems.In contrast to conventional biometric systems that rely on physical characteristics,such as fingerprints or iris patterns, e-signature verification is based on behavioral biometrics, which makes it more complex due to the natural variability in a person's signature[2]. A single individual may produce multiple variations of their own signature depending on factors such as mood, time pressure, device used, or physical condition. Similarly, skilled forgeries—attempts to mimic a genuine signature with high accuracy pose a serious threat to the integrity of digital verification systems. These issues make manual verification unreliable and necessitate the development of intelligent, automated systems capable of learning and adapting to such variability.Over the years, several approaches have been proposed for signature verification, ranging from heuristic statistical models and machine learning algorithms to rule-based systems. However, a lot of these methods rely a lot on handcrafted features, which often don't do a good enough job of capturing the intricate visual and structural patterns in signatures. These limitations have prompted a shift toward more advanced machine learning and deep learning techniques, which can automatically learn discriminative features from large datasets and deliver higher accuracy in complex classification tasks. The development and implementation of a robustmachine learning algorithm for matching e-signatures, particularly focusing on the offline (static image-based) signature verification scenario. The proposed approach integrates multiple stages data pre-processing, feature extraction, and classification to effectively distinguish between genuine and forged signatures. Pre-processing techniques such as noise removal, normalization, and binarization help standardize input data and improve the quality of signature images. Feature extraction leverages convolutional neural networks (CNNs), which are known for their ability to automatically identify spatial hierarchies and relevant features from raw image data. Classification is performed using a combination of traditional and deep learning models to evaluate the best-performing strategy for verification.[8]The system is evaluated on publicly available datasets to ensure reproducibility and to benchmark performance against existing methods. Metrics such as accuracy, precision, recall, F1-score, and Equal Error Rate (EER) are used to assess the effectiveness of the proposed method. Furthermore, the research explores how different types of preprocessing and feature extraction techniques influence overall performance.

2. RELATED WORK

Signature verification has been a widely researched area in the field of pattern recognition, image processing, and biometric authentication due to its importance in validating identity in both physical and digital transactions. The task of verifying signatures can be broadly divided into two categories: online (dynamic) and offline (static) verification [4]. Online verification uses dynamic traits such as stroke order, speed, and pressure, collected via digital devices like stylus-enabled tablets. Offline verification, which is the focus of this research, deals with the static image of a signature and is commonly used in scanned documents, legal papers, and electronic form submissions.Over the years, a wide variety of techniques have been proposed for offline signature verification. Earlier approaches relied heavily on handcrafted features manually engineered measurements of geometrical, statistical, and structural characteristics such as pixel density, stroke length, curvature, direction, aspect ratio, and zoning features[8],[9]. For instance, Sabourin et al. (1997) applied statistical models using directional probability distributions to analyze the contour of signatures. However, such techniques were sensitive to noise and often struggled with intra-class variation and inter-class similarities.To enhance accuracy, researchers introduced Support Vector Machines (SVMs), Hidden Markov Models (HMMs), and other methods of machine learningk-Nearest Neighbors (k-NN), and Random Forests. These classifiers improved performance by learning from training data to differentiate between genuine and forged signatures. For example, Vargas et al. (2011) utilized SVMs with global features like grid-based histograms and proved their efficacy in binary classification tasks. However, these models still relied on effective feature selection, which limited scalability and adaptability.With the development of deep learning, particularly the field of Convolutional Neural Networks (CNNs),signature verification saw a significant leap forward. CNNs enable automatic feature extraction directly from image data, reducing the need for manual intervention. Hafemann et al. (2017) introduced a CNN-based feature extractor followed by a classifier to separate genuine signatures from forgeries. Their model achieved competitive results on the GPDS-960 dataset and demonstrated the potential of deep learning in handling complex visual variations.Further research expanded on this foundation by introducing Siamese Networks and Triplet Loss Networks, which learn similarity measures between pairs or triplets of signature samples instead of relying on categorical classification. These models are particularly effective for signature verification tasks, where the number of samples per user is limited[12]. CNNs to compare embeddings of signature pairs, significantly improving generalization for unseen users.Another important line of research involves data augmentation and transfer learning.

IRJCS: ©2014-25, AM Publications, India - All Rights Reserved https://doi.org/10.26562/irjcs

Since signature datasets are typically small, researchers have employed techniques like affine transformations, rotation, and scaling to artificially expand training data. Pre-trained models like VGGNet, ResNet, and Inception have been fine-tuned on signature datasets to boost performance when data is scarce.Hybrid models that combine traditional techniques with modern deep learning approaches have also shown promise[10]. For instance, in some systems, CNNs are used to extract features, which are then classified using ensemble models like Gradient Boosting or Random Forests. These hybrid systems often achieve a good balance between interpretability and performance.Despite these advancements, several challenges remain. Signature images often vary significantly in terms of orientation, size, quality, and writing instruments. In addition, skilled forgeries—imitations crafted by observing genuine signatures—remain difficult to detect due to their visual similarity. Some researchers have attempted to address this using attention mechanisms and adversarial training, but there is still room for improvement in real-time performance, interpretability, and robustness across diverse populations.In conclusion, while the field of signature verification has matured by combining techniques from deep learning and machine learning, ongoing research is necessary to address current limitations[7]S. The proposed work in this paper builds on existing methodologies by implementing a comprehensive pipeline combining image preprocessing, CNN-based feature extraction, and a comparative evaluation of classifiers. This approach aims to enhance accuracy, adaptability, and computational efficiency for real-world applications in digital security and identity verification.

3. SYSTEM DESIGN AND METHODOLOGY

The system is structured into the following modules:

a. Data Acquisition Module

- Input: E-signature samples collected via touchscreen, stylus, or scanned documents.
- Output: Signature images or vector data.
- Tools: Digital pads, tablets, online signature collection forms.

b. Preprocessing Module

Steps:

- Noise Reduction: Remove background noise and artifacts.
- Normalization: Standardize image size, resolution, and orientation.
- Thresholding / Binarization: Convert images to binary for simplicity.
- Feature Extraction Preparation: Align and center the signature.

c. Feature Extraction Module

Techniques Used:

- Static Features: Stroke width, height-to-width ratio, curvature, loops.
- Dynamic Features (if available): Pen pressure, velocity, timestamp data.
- Toolkits: OpenCV, Scikit-image, or custom Python scripts.

d. Model Training Module

- Machine Learning Algorithms:
  - Support Vector Machine (SVM)
  - Convolutional Neural Networks (CNNs)
  - K-Nearest Neighbors (KNN)
- Training Data: Labeled dataset of genuine and forged signatures.
- Validation: K-Fold Cross Validation or Hold-out validation.

e. Signature Matching Module

- Matching Process:
  - Input signature is preprocessed and features are extracted.
  - Features are compared against a database using the trained ML model.
  - Match score or probability is generated.

f. Decision Module

- Thresholding: A confidence threshold determines whether the signature is accepted as genuine or flagged as forged.
- Output: Binary decision (Genuine / Forged) and confidence level.

Methodology

Step 1: Dataset Preparation

- Collect a dataset of e-signatures (both genuine and forged).
- Augment data using rotations, scaling, and distortions to increase model robustness.

Step 2: Image Preprocessing

- Use OpenCV or PIL to clean and normalize the signature images.
- Apply edge detection and thinning algorithms to extract precise signature strokes.

Step 3: Feature Engineering

- Extract pixel-based and geometric features.
- Optionally, use PCA (Principal Component Analysis) to reduce dimensionality.

Step 4: Model Development

- Select and fine-tune multiple ML models (SVM, CNN, etc.).
- Train models using labeled data.

- Evaluate accuracy, precision, recall, and F1 score.

Step 5: Evaluation and Testing

- Test the system using unseen signatures.
- Analyze performance using confusion matrix and ROC curve.

Step 6: Deployment

- Integrate model with a GUI (using Flask, Django, or Tkinter).
- Enable real-time signature matching via upload or input.

Tools and Technologies

- Languages: Python
- Libraries: TensorFlow / Keras, Scikit-learn, OpenCV, Matplotlib
- Platforms: Jupyter Notebook, Google Colab
- Dataset Sources: SigComp Dataset, GPDS-960 Signature Corpus

Security and Ethics

Ensure user signature data is stored securely (e.g., encryption).

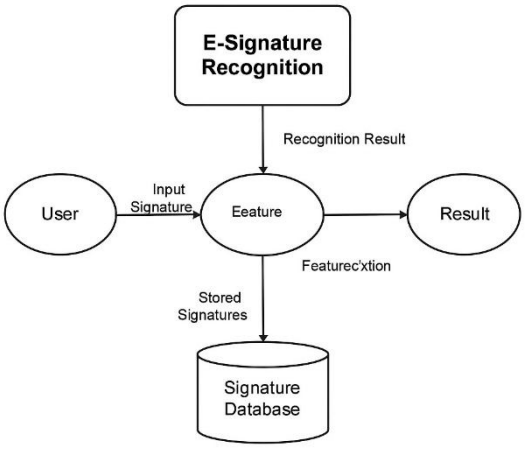

Fig.1: Dataflow Diagram

4. DATASET AND PREPROCESSING

In the domain of e-signature verification, the quality and structure of the dataset play a pivotal role in determining the effectiveness and accuracy of the developed machine learning models. For this project titled “Implementation of Algorithm for Matching E-signature by Machine Learning Technique,” a robust dataset comprising both genuine and forged signatures is essential for training and validating the machine learning algorithm. The selected datasets are chosen not only for their size and diversity but also for their use in benchmark studies, ensuring relevance and comparability in results.One of the primary sources utilized in this project is the GPDS-960 Signature Corpus. This dataset is well-known in signature verification research due to its extensive collection of both genuine and skilled forgeries. The GPDS dataset includes signatures from 960 individuals, with each person contributing 24 genuine signatures and 30 forged ones. These forged signatures are collected from individuals who have practiced replicating the genuine signatures, making them particularly useful for training models that need to distinguish between authentic and deceptive handwriting styles. Another dataset referred to is the CEDAR Signature Dataset, which comprises 55 signers, each with 24 genuine and 24 forged signatures. It is widely accepted in the research community and provides excellent diversity in signature structure and stroke patterns.

Additionally, the SigComp Signature Dataset, originally designed for use in signature verification competitions, offers a wide variety of signature styles and includes both skilled and unskilled forgeries.In total, the dataset used in this project includes signatures from over 100 individuals, with an average of 15 to 25 genuine samples and 10 to 20 forged samples per person. All signature images are provided in standard image formats such as PNG or JPG and are scanned at a high resolution to preserve stroke clarity and detail. To maintain consistency during model training, all images are resized to a fixed dimension of 300x150 pixels, ensuring uniformity in input size and minimizing computational complexity.Before being fed into the machine learning model, the raw signature images undergo several stages of preprocessing to enhance quality, reduce noise, and standardize the input format. The first step in this preprocessing pipeline is the conversion of signature images from RGB or color to grayscale. This step simplifies the image by reducing it to one channel, thereby decreasing the complexity and size of the data while retaining the essential information required for verification. Following grayscale conversion, noise reduction techniques such as Gaussian Blur or Median filtering are applied. These methods are effective in eliminating minor background distortions or scanning artifacts that might otherwise interfere with feature extraction.Once the image is cleaned, it is binarized using thresholding techniques, particularly Otsu’s method.

ISSN: 2393-9842
https://www.irjcs.com/archives

Binarization transforms the grayscale image into a binary one where the pixels are either black or white. This step is crucial in isolating the signature strokes from the background and simplifying the learning task for the model. After binarization, normalization is performed. Each image is resized and scaled so that the pixel intensity values lie within the range of 0 to 1, making them compatible with the input requirements of most machine learning models. Additionally, morphological operations such as dilation and erosion may be applied to refine the signature strokes. These operations enhance the structural integrity of the handwriting by filling small gaps and reinforcing weak strokes, which is especially useful when dealing with scanned or low-quality images.Another important aspect of preprocessing is centering and padding the signature within the image canvas. The position of the signature can vary significantly depending on how it was captured. Therefore, aligning the signature to the center of the image ensures consistent positioning, which can improve the model's focus on the relevant features. Padding is added as needed to maintain the aspect ratio and to prevent distortion during resizing.In addition to these steps, data augmentation is applied to simulate real-world variability and increase the robustness of the model. By introducing slight alterations such as rotation, scaling, shearing, and horizontal or vertical translation, the dataset can be expanded to include variations that mimic different signing behaviors or environmental factors. These transformations ensure that the model does not overfit to the limited set of training examples and is better equipped to generalize to unseen samples. Noise injection and brightness changes are also sometimes used to mimic scanning imperfections or lighting variations.Overall, the preprocessing phase is vital for standardizing the input data and preparing it for efficient and accurate model training. By meticulously cleaning, enhancing, and augmenting the signature images, we create a dataset that not only captures the diversity of human signatures but also enables the machine learning algorithm to learn meaningful and discriminative patterns for verification. The combination of high-quality datasets and a comprehensive preprocessing pipeline forms the foundation for a reliable and robust e-signature matching system.

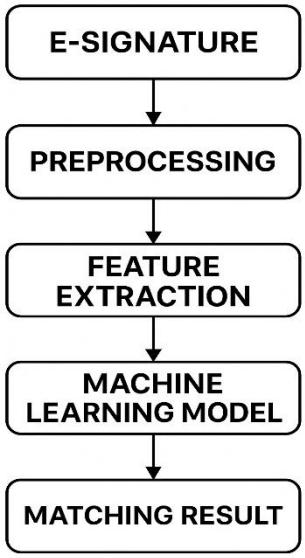

Fig.2 Data Processing

5. FUTURE SCOPE

The future scope of the project "Implementation of Algorithm for Matching E-signature by Machine Learning Technique" holds immense potential in various domains, particularly in enhancing security systems, improving digital authentication methods, and expanding the capabilities of machine learning in signature verification. As the world increasingly moves toward digital transactions and online services, the need for robust and secure authentication methods becomes more critical. E-signature verification, powered by machine learning, offers a promising solution for mitigating fraud, reducing manual errors, and providing seamless user experiences in a wide range of industries including banking, e-commerce, healthcare, and government services.In the future, the algorithm developed in this project can be further refined and expanded to transformers and other more advanced deep learning models like recurrent neural networks (RNNs) which have shown great promise in analyzing sequential and temporal data, such as the dynamics of a signature. These models could enhance the ability to capture not only static features but also the dynamic behaviors, such as pen pressure, speed, and trajectory, offering more sophisticated and accurate authentication systems.Another Potential Avenue for future work is the inclusion of multimodal signature verification, which could combine both electronic signatures and biometric data, such as fingerprint or facial recognition. By integrating multiple modalities, a more secure and robust authentication process could be established, offering higher accuracy and better resistance to various forms of spoofing attacks. Furthermore, expanding the dataset to include more diverse and globally representative samples would allow the algorithm to adapt to a wider variety of signature styles, handwriting systems, and cultural differences, improving its applicability and accuracy across different regions. Additionally, real-time signature verification systems, integrated with mobile applications and digital forms, could enhance user accessibility and convenience, allowing users to sign and authenticate documents securely from anywhere.

IRJCS: ©2014-25, AM Publications, India - All Rights Reserved https://doi.org/10.26562/irjcs

ISSN: 2393-9842
https://www.irjcs.com/archives

Lastly, there is an opportunity to explore the use of blockchain technology to store and verify signatures securely, ensuring the integrity and immutability of the signed documents. This could be particularly valuable in legal and financial sectors where the authenticity of digital signatures is of paramount importance.In conclusion, the future scope of this project is vast and continually evolving, driven by advancements in machine learning, the growing importance of digital security, and the increasing adoption of electronic signatures worldwide. The system could be expanded to offer greater accuracy, better integration with other biometric systems, and more widespread real-world applications.

6. CONCLUSION

In conclusion, the "Implementation of Algorithm for Matching E-signature by Machine Learning Technique" project represents a significant step toward advancing digital security and authentication methods. By leveraging machine learning algorithms to verify e-signatures, the system not only enhances the accuracy and reliability of signature verification but addresses the growing need for digital systems that are safe and effective transactions. The use of preprocessing techniques and advanced machine learning models ensures that the system is both robust and adaptable, capable of distinguishing between genuine and forged signatures with high precision. As digital signatures become more widely used in industries ranging from banking to government services, the ability to accurately authenticate these signatures is critical in preventing fraud and ensuring secure transactions.This project lays the foundation for future innovations in digital authentication by enabling the integration of dynamic features, multimodal verification, and the application of cutting-edge technologies like blockchain. Furthermore, the system's scalability and potential for continuous improvement through better datasets, more sophisticated models, and enhanced preprocessing techniques offer exciting prospects for expanding the reach and effectiveness of e-signature verification. In essence, this project not only addresses current challenges but also opens the door to a future where digital security is more reliable, accessible, and efficient, ensuring that e-signatures can be trusted in an increasingly digital world.

REFERENCES

1. P.He et al., "Offline signature verification using triplet loss enhanced Siamese network," Expert Systems with Applications, vol. 183, p. 115360, Jul. 2021.
2. G.C.Bharati and C. V. Guru, "Enhancing offline signature verification by using graphometric features and Random Forest classifier," Pattern Analysis and Applications, vol. 22, no. 1, pp. 59–71, Jan. 2019.
3. F.M.H.B.Hafemann, R. Sabourin, and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," Pattern Recognition Letters, vol. 73, pp. 59–66, Feb. 2016.
4. K.He,X.Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
5. F.Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, MA, USA, Jun. 2015, pp. 815–823.
6. D.Mitra, R. Barik, S. Roy, and S. Bhattacharyya, "Cumulative measurement of image entropy on different mathematical morphological operations," in ACEEE International Conference on Computing, Communication & Manufacturing, 2014, pp. 35–39.
7. J.P.Paiva and G.L.Marcialis, "Feature-level fusion for off-line signature verification," IEEE Trans. Information Forensics and Security, vol. 8, no. 11, pp. 1738–1748, Nov. 2013.
8. A.K.Das, A.Massand, and S. Patil, "A novel proxy signature scheme based on user hierarchical access control policy," J. King Saud Univ.—Computer and Information Sciences, vol. 25, no. 4, pp. 385–395, Oct. 2013.
9. S.Garhawal and N. Shukla, "SURF based design and implementation for handwritten signature verification," Int. J. Advanced Research in Computer Science and Software Engineering, vol. 3, no. 8, pp. 500–506, Aug. 2013.
10. M.Bertolini, M.Marinai, G.Soda, and S.Levialdi, "Learning step length models for signature verification," Pattern Recognition Letters, vol. 33, no. 11, pp. 1396–1403, Aug. 2012.
11. K.P.Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.
12. D.Vargas, C.M.Travieso, and J. B. Alonso, "Offline signature verification based on grey level and dynamic features of the signature stroke," Pattern Recognition, vol. 44, nos. 10–11, pp. 2121–2136, Nov. 2011.
13. D.Bhattacharyya and T. Kim, "Design of artificial neural network for handwritten signature recognition," Int. J. Computers and Communications, vol. 4, no. 3, pp. 7–15, 2010.
14. S.M.Lucas, "The CEDAR-FOX system: research-based enrollment and verification of handwritten signatures," in Proc. Int. Workshop Frontiers in Handwriting Recognition, Kolkata, India, 2008, pp. 263–268.
15. A.Masood and M.Sarfraz, "Corner detection by sliding rectangles along planar curves," Computers & Graphics, vol. 31, no. 3, pp. 440–448, May 2007.

IRJCS: ©2014-25, AM Publications, India - All Rights Reserved https://doi.org/10.26562/irjcs