Grenze International Journal of Engineering and Technology, June Issue

Machine Learning-based Signature Verification using OCR and Line Sweep Technique

Vidhya S G[1], Anusha T S[2], Arpitha M R[3], Geetha P S[4] and Nisarga M[5]
[1]Asst Prof., Dept. of Information Science and Engineering, ACU, BGSIT, India
Email: vidhyasg@bgsit.ac.in
[2-5]UG Student, Dept. of Information Science and Engineering, ACU, BGSIT, India

***Abstract*— A prevalent method for personal identification and validation, handwritten signatures find extensive use in various settings. Documents such as bank checks and legal agreements necessitate signature authentication, presenting a formidable challenge given the sheer volume. This underscores the need for a robust automated signature verification tool to combat fraud across financial transactions. Presently, manual verification relies heavily on the verifier's subjective judgment, which is prone to inefficiencies and errors. Expert evaluators struggle to discern the subtle differences in line and angle ratios between genuine and fraudulent signatures. To address these challenges, we propose an automated signature verification system leveraging advancements in image processing and machine learning. Upon capturing a customer's handwritten signature, preprocessing steps will enhance its clarity and extract relevant features. Subsequently, a verification process will compare these features against stored data in the database. Our approach employs OCR-based techniques for signature localization, combined with Connected Components and geometric feature analysis. Classification using Support Vector Machine (SVM) achieves a commendable 91% accuracy in signature authenticity verification.**

I. INTRODUCTION

Machine Learning (ML) has become a cornerstone in various fields, and one intriguing application is the authentication of handwritten signatures. This process involves verifying individuals based on their handwritten marks. Recent advancements in ML, along with innovative techniques like Line sweep algorithms and Optical Character Recognition (OCR), have significantly bolstered the accuracy and efficiency of signature verification systems. Line sweep algorithms, when combined, synergize to enhance the signature verification process. Line sweep's geometric insights complement algorithmic decision-making, resulting in a comprehensive and precise analysis of handwritten signatures. OCR techniques further enrich signature verification systems by converting handwritten text into machine-readable data. OCR facilitates the extraction of textual information from signatures, enabling additional analysis and comparison. Algorithms serve as the foundation of ML-based signature verification systems, providing the structure for learning and decision-making. Supervised learning algorithms, such as Support Vector Machines (SVM) and Neural Networks, are commonly utilized. SVMs excel in distinguishing authentic signatures from forgeries by generating optimal decision boundaries based on training data

II. MOTIVATION

The development of a Machine Learning (ML) based signature verification process utilizing Optical Character

Grenze ID: 01.GIJET.10.2.742

Recognition (OCR) and Line Sweep techniques is driven by various factors. The primary objective of the project is to enhance security in industries where the authenticity of signatures holds significant importance, such as finance, legal, and government sectors. With the rise of digital transactions and the increasing sophistication of signature-related fraud, there is a pressing need for a robust system to safeguard against unauthorized activities. Integrating ML and OCR aims to automate manual processes, thereby improving efficiency, minimizing human error, and offering a scalable solution capable of managing expanding transaction volumes and document loads. The versatility of ML models in accommodating diverse signature styles ensures a thorough and dependable verification process. Additionally, incorporating OCR enables a comprehensive analysis of both visual and contextual elements within documents. Ultimately, the project endeavors to deliver a technologically advanced, compliant, and user-friendly solution. Such a solution not only strengthens security measures but also drives innovation in document verification processes, contributing to overall advancements in the field.

III. Literature Survey

The aim of Shashidhar Sanda and Sravya Amaretti's paper is to develop a method capable of distinguishing between authentic and forged signatures. This method utilizes a Support Vector Machine (SVM) and represents signatures as a boundary of x-y coordinates stored as a text file in a sign database. SVMs are commonly employed in pattern recognition and regression tasks. The method involves several steps:
1. Initial Step: Input signatures are read in x-y coordinates to determine the position of the pen, such as its direction.
2. Data Acquisition: Real-time signatures from specialized pens and digital discs are processed and stored in a signature database.
3. Preprocessing: This involves normalization, sampling time adjustment, and resampling distance calibration.
4. Feature Extraction: Various features are extracted to enhance the ability to distinguish between categories. The generated features are then added to the database and matched against input image features.
The authors use two methods to validate the correctness of the program: Shashikumar KB Raja proposes an offline signature verification method using a neural network (SVFGNN) that integrates global and network features. S. Priya, A. Hruday Kumar, and K. Somesh propose a system for signature verification based on pixels and strokes, utilizing methods like the Harris algorithm and surf.
Furthermore, M.B. Sudhan, Abhilasha Sarkar; Manjula Subramaniam, Arpith Mathew, and Teja E; k. Thitthilakaraj, T. Santha Perumal; Chandra Sekhar V, Prerana Mukherjee, Devanur S Guru, Viswanath Pulabaigari; Antonio Parziale, Moises Diaz, Miguel A. Ferrer, Angelo Marcelli; Tasrina Parvin, and Alimul Rajee; and Luiz G. Hafemann, Robert Sabourin all propose various methods and techniques for signature verification and detection, using approaches such as deep learning, image processing, and dynamic time warping, among others. These methods contribute to the advancement of secure authentication systems and document verification processes.

IV. Proposed System

The proposed machine learning-based signature verification system integrates Optical Character Recognition (OCR) techniques and the Line Sweep algorithm to enhance the accuracy and efficiency of signature authentication across various domains. Here's a breakdown of the key components and approaches:
1. Optical Character Recognition (OCR) Integration: The system starts with acquiring high-quality signature images using scanning devices or cameras. OCR techniques preprocess these images, extracting both textual information and relevant signature features to standardize representation and enhance clarity.
2. Feature Extraction and Machine Learning Model: Machine learning algorithms, such as Support Vector Machines (SVM) or Convolutional Neural Networks (CNNs), are employed for signature verification. Feature extraction methods capture essential characteristics like stroke patterns and curvature. The model is trained on a diverse dataset of genuine and forged signatures to refine parameters iteratively.
3. Line Sweep Algorithm for Spatial Analysis: The Line Sweep algorithm facilitates spatial analysis of signature features by systematically sweeping through segments of the signature space. This approach identifies patterns efficiently, reducing computational complexity and enhancing scalability.
4. Multi-Mode Fusion: To enhance robustness, the system integrates information from OCR, machine learning models, and the Line Sweep algorithm. This fusion approach enables comprehensive analysis and more accurate decision-making.
5. Dynamic Learning and Adaptability: The system incorporates dynamic learning mechanisms to adapt to evolving signature patterns. It continuously updates its model based on new data, ensuring adaptability to changes in writing styles and forgery techniques.

6. Real-time Processing and Authentication: With a focus on real-time processing, the system delivers swift and accurate signature authentication. The combination of OCR, machine learning, and the Line Sweep algorithm ensures reliability and meets speed requirements for various applications.
Regarding the Signature Verification Approaches:

- *OCR Approach*: Identifies and crops signature images.
- *Connected Components Algorithm*: Extracts signatures from images.
- *Line Sweep Algorithm*: Fits rectangles across signatures, utilizing OCR result images.

The functionalities to be implemented include dataset loading, model training, output classification, performance measurement, and result generation. These approaches collectively form a comprehensive and advanced signature verification system capable of enhancing security and efficiency across domains relying on signature authentication processes.

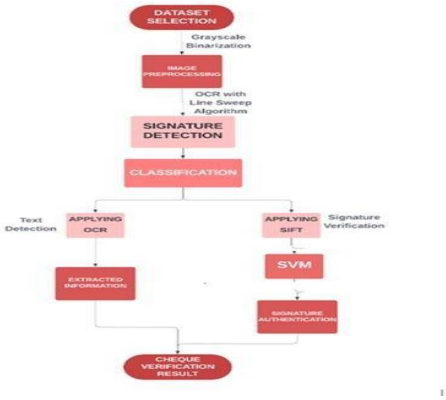

Fig 1: System Architecture

The Machine Learning Cognitive Verification Process, integrating OCR and line sweep techniques, is a robust system designed to bolster document authentication and verification. This advanced architecture seamlessly incorporates various components to ensure precision and efficiency in cognitive verification tasks. Let's break down the key elements:
1.OCR Module: The system's foundation lies in its OCR module, employing advanced machine learning algorithms to extract text information from documents. Trained on diverse datasets, it recognizes various fonts, languages, and document formats, ensuring adaptability to different scenarios.
2. Line Sweep Technique: This technique systematically scans the document's textual content to identify patterns, relationships, and anomalies. It aids in assessing document structure and detecting irregularities indicating potential fraud or forgery.
3. Feature Extraction: Characteristics like text alignment, spacing, and orientation are extracted, enhancing the system's ability to differentiate genuine documents from fraudulent ones. This step prepares the data for the subsequent machine learning-based verification phase.
4. Machine Learning Model: The core involves a machine learning model trained on authentic and forged documents. Capable of learning complex patterns, it makes informed decisions regarding document authenticity. Continuous refinement through iterative training ensures adaptability to evolving forgery techniques.
5. Real-time Processing and Scalability: Parallel processing capabilities enable efficient handling of large document volumes without compromising verification accuracy. The system seamlessly integrates with existing verification workflows in various industries.
6. Feedback Loop: User feedback and verification outcomes inform the training process, allowing the model to adapt and refine its decision-making over time. Security and privacy are ensured through encryption protocols.
The UML Diagram depicts the signature verification process, illustrating user authentication triggering OCR, Line Sweep, and Feature Extraction modules. These modules collaborate with the ML Model for verification, with components, interactions, and server deployment succinctly represented.
In conclusion, the Machine Learning Cognitive Verification Process leveraging OCR and line sweep techniques offers a sophisticated solution for document authentication. By integrating OCR, spatial analysis, feature extraction, and machine learning, it creates an adaptive, secure, and efficient system enhancing cognitive verification across diverse applications.

V. Conclusion

With the rapid evolution of technology, traditional protocols are becoming outdated, prompting the development of new tools and technologies to combat fraud and crime. Systems are continually updated to address vulnerabilities as they are uncovered. One such system has been devised to combat forged and altered signatures, ensuring precise disclosure of individuals' identities and thwarting identity-related crimes.

The integration of machine learning, OCR, and the line sweep technique represents a significant advancement in signature verification. This integration holds the promise of enhanced accuracy and adaptability. However, further refinement and research are imperative to optimize its performance in real-world scenarios. Continued efforts in research and development will be crucial to ensure the effectiveness and reliability of these advanced verification systems.

References

[1] Sounak Deya, Anjan Dutta, J. Ignacio Toledo, Suman K.Ghosha, Josep Llados´a, Umapada Pal ”SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification” Issue 30, September,2017.

[2] Julita A., Fauziyah S., Azlina O., Mardiana B., Hazura H., Zachariah A.M.”Online Signature Verification System” Issued April 2009.

[3] Shashi Kumar D R, K B Raja, R. K Chhotaray, Sabyasachi Pattanaik ”Off-line Signature Verification Based on Fusion of Grid and Global Features Using Neural Networks” Vol. 2(12), 2010.

[4] Poonam Chaudhary and Vijay Kumar Singh,” Online Signature Verification: A Review”, Volume3, Issue-2, Feb 2016.

[5] Liyang Xie, Zhongcheng Wu, Xian Zhang, Yong Li, Xinkuang Wang “Writer-Independent Online Signature Verification based on 2-D Representation of Time Series Data using Triplet Supervised Network” June 2022.

[6] S. Priya, A.K.R.N.Supreeth, K. Somesh, A. Hruday Kumar ” Signature Verification System using Different Algorithms”, Volume-8, Issue6S3, April 2019.

[7] Tushara D, Shridevi Raddy, Shreya KM, Spoorthy Y,” Signature Verification System using Neural Networks”, NCCDS - 2021 Conference Proceedings.

[8] Yuchen Zheng, Brian Kenji Iwana, Muhammad Imran Malik, Sheraz Ahmed, Wataru Ohyama, Seiichi Uchida “Learning the MicroDeformation by Max-Pooling for Offline Signature Verification” October 2021.

[9] Chandra Sekhar Vorungunti, Viswanath Pulabaigari, Rama Krishna Sai Subrahmanyam, Gorthi, Prerana Mukherjee “OSVFuseNet: Online Signature Verification by Feature Fusion and DepthWise Separable Convolution Based Deep Learning” October 2020.

[10] Antonio Parziale, Moises Diaz, Miguel A. Ferrer, Angelo Marcelli “SM-DTW: Stability Modulated Dynamic Time Warping for Signature Verification” April 2019.

[11] Xia Xingua, Song Xiaoyu, Luan Fangun, Zheng Jungang, Chen Zhili, Ma Xiaofu “Discriminative Feature Selection for Online Signature Verification” February 2018.