IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, May 2026
Impact Factor: 8.2

Machine Learning -Based Signature Verification System

Mrs. J. Fahamitha, P. Yedukondalu, N. Sumanth, M. Mani Shankar
Asst. Professor, Computer Science & Engineering, School of Engineering & Technology
Dhanalakshmi Srinivasan University, Trichy, India
ypeyyla076@gmail.com, nagasumanthreddy271@gmail.com

Abstract: *Handwritten signatures have been considered one of the most accepted means of biometric authentication in the financial sector, legal documents, and other financial transactions. However, the vulnerability of signatures in forgery and identity theft necessitates the development of an automated mechanism of identifying fraud signatures. This research outlines a framework that can automatically differentiate genuine signatures from forged ones using the power of machine learning. The steps involved in the proposed system match those of a general verification approach. They include data acquisition, pre-processing, feature extraction, and classification. During pre-processing, the quality of the signature is improved through noise removal, grayscale conversion, binarization, normalization. Relevant discriminative features of the signature are identified using texture as well as structural descriptors. Finally, machine learning approaches such as the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models are trained on the identified features. The trained model classifies the input signature as genuine or fake.*

Keywords: Signature verification; Signature fraud detection; Machine learning; Offline signature recognition; Image preprocessing; Feature extraction; Pattern recognition; Support vector machine; Convolutional neural network; Biometric authentication; Forgery detection; Similarity matching; FAR; FRR; Document Authentication

I. INTRODUCTION

The handwritten signature verification technique is one of the first and most widely acceptable techniques of human authentication that has been in use in various transactions in the field of finance and law. Although various state-of-the-art methods of biometrics, including fingerprint and facial verification, are in use, the technique of signature verification seems not only simple and acceptable by the user but also legally viable. Yet, the manual verification process of signatures largely relies on the cognitive powers of humans, and hence, any sort of mistake in this process leads to a potential threat of fraud and identity theft.

Generally, manual verification techniques have been employed, depending on expert opinion. However, it takes time and is not reliable. A person's signature varies each time it is written due to rates of writing, pressure while writing, and hand movement. Signature forgery also imitates these differences. As a result, it is not possible to distinguish genuine signatures with high accuracy, making it essential to develop signature verification systems that are automated and intelligent.

Machine Learning (ML) provides an effective solution by enabling computers to learn distinguishing patterns from genuine signatures and detect forged ones automatically. In a typical ML-based signature verification system, the input signature image undergoes preprocessing to remove noise and enhance quality. Important features such as shape, stroke orientation, and texture patterns are then extracted and used to train classification models like Support Vector Machine (SVM) or Convolutional Neural Networks (CNN). The trained model evaluates new signatures and predicts whether they are authentic or fraudulent.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, May 2026 Impact Factor: 8.2

This research focuses on offline signature verification, where scanned images of signatures are analyzed without requiring special hardware devices. The aim is to develop a reliable machine learning-based system that minimizes false acceptance of forged signatures while accurately recognizing genuine signatures. Such a system can be applied in banks, examination authorities, and legal institutions to strengthen security and automate document authentication processes.

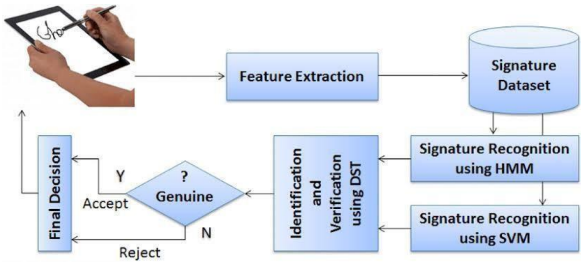


Fig. 1. Workflow Of Signature Fraud Detection

The workflow of the signature fraud identification system starts with the scanning of the handwritten signature using a scanning device or digital tablet. After scanning the image, the step of feature extraction comes into play. In the feature extraction process, the most dominant characteristics of the handwritten signature in regards to shape, direction of the strokes, curved or straight line characteristics of the signatures, and patterns of the texture of the signatures are considered. Once the characteristics of the signatures have been obtained, they form the database and are compared using the hidden markov model and support vector machines to perform the identification and verification process with the help of decision support technology. Once the decision has been reached, the decision is used to either accept the signature as genuine or reject the signature with the assumption of an unauthorized use of fraud.

II. RELATED WORKS

Signature verification has been studied as a behavioral biometric modality for authentication purposes. Initially, the focus of research on signature verification and identification involved mainly the use of handcrafted feature extraction methods including geometric features, grid features, texture features, and statistical features. In addition, classification methods such as k-Nearest Neighbors classifier, SVM classifier, and Hidden Markov Model classifier were applied for the purpose of differentiating genuine signatures from forged ones. Although reasonable accuracy is obtained by these methods, the accuracy is found to be highly dependent on the selection of features.

However, with advancements in image processing, various studies have employed a number of operations involving pre-processing stages such as binarization, noise reduction, thinning, and normalization for high accuracy in verification results. Investigations using techniques such as Histogram of Oriented Gradients, Local Binary Patterns, and even wavelet transforms were conducted for the purpose of extracting stroke orientation and texture features. However, these techniques were not entirely successful in beating skilled forgery, i.e., when the perpetrator copied the signature with considerable care.

Recent studies are mainly focused on providing automatic feature learning by deep learning-based methods. Convolutional Neural Networks (CNNs) are reported to offer better performance by extracting high- level spatial features directly from the signature image without requiring any intermediate human involvement. Siamese Networks and Triplet Networks are also widely used for comparing the similarity between genuine and forged signatures by learning discriminative representations. Recurrent Neural Networks (RNN) and LSTM models have been used for online signature verification systems by using the feature of dynamic writing.

Hybrid approaches, such as fusing traditional machine learning and deep learning methodologies, have shown improved results in fraud detection system evaluations as well. Efforts have been made to improve the overall accuracy of a classifier by fusing CNN-based feature learning with classifiers such as SVM and Random Forest. Research has even introduced transfer learning and data augmentation to solve the scarcity of signature data, especially for users.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, May 2026 Impact Factor: 8.2

Handwritten signature verification can be considered one of the most widely accepted methods for behavioral biometric authentication, being applied generally in banking, legal documentation, and identity validation scenarios. Several approaches have been proposed for the detection of forged signatures by various researchers over the years, and these can be broadly classified as online and offline signature verification systems.

Traditional Machine Learning Methodologies

Earlier research was centered on extracting handcrafted features from the signature images. Features included geometric features such as height, width, and aspect ratio; directional features; stroke density; and pixel distribution. Classifiers such as k-Nearest Neighbor (k- NN), Support Vector Machine (SVM), Naïve Bayes, and Decision Trees were also employed for the classification methods. Hidden Markov Models (HMM) were also investigated for the sequential writing styles in the online signature verification.

While these approaches were successful in handling random forgeries, they have proven ineffective in handling cases involving skilled forgeries, mainly because it was not possible to design features for a writer's behavior.

III. PROPOSED SYSTEM

The proposed system aims to create an automatic and precise tool that can verify the signatures written by an individual using advanced machine learning technology. Secondly, the proposed system would aim to minimize the errors that occur during the manual verification process. On the other hand, it would be able to even detect the forgery signatures by learning the behavior of the individual.

The process starts with signature acquisition in which the user has to upload the scanned image or captured picture of the signature through the application interface. Here, the received signature gets stored in the database, which then gets sent for processing. In the interest of consistency, the system provides several opportunities for genuine signature acquisition for each user during the registration phase, so that variations in the signature of any person can be learned.

Once the signature has been acquired, it then moves to the pre-processing stage. During the pre-processing stage, various image processing operations like gray- scaling, noise removal, thresholding, binarizing, resizing, and thinning are applied. These operations remove unnecessary background details and highlight the style of writing used in the signature. With pre-processing, it ensures that only writing style, rather than image quality, is considered by the machine learning model being used.

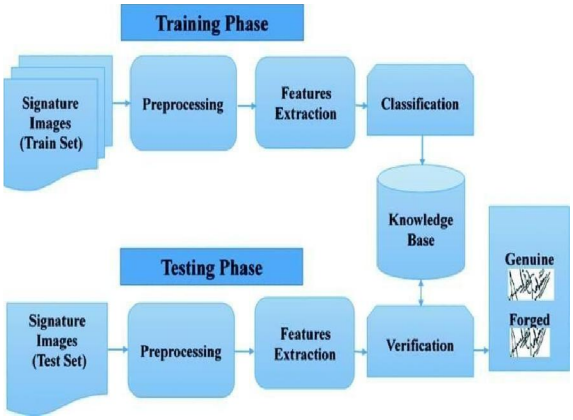


Fig.2. Architecture of Signature Fraud Identification

The picture describes the entire process of a Signature Fraud Identification system using Machine Learning, consisting of two main components: the Training Phase and the Testing (Verification) Phase. The picture also highlights how a Signature Fraud Identification system using ML is able to learn from a signature and then later verify it.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, May 2026 Impact Factor: 8.2

A. Training Phase (Learning Stage)

In this stage, it learns what genuine and forged signatures look like.

B. Signature Images (Train Set)

A dataset is created which includes a large number of genuine as well as forged signatures. Samples must be obtained for a large number of persons in order to understand the natural variations.

C. Preprocessing

The raw signature images contain potential noise, shadow, and varying sizes. The system cleans the signature image by using the following methods:

- Grayscale conversion
- Noise removal (Filtering)
- Thresholding & binarization
- Resizing & Normalization
- Thinning (skeletonization)

This step will ensure that the model is concentrating only on the writing strokes and not the background.

D. Feature Extraction

After cleaning, the system retrieves key identifying characteristics of the signature:

- Shape and structure
- Stroke direction
- Curves and Loops Endpoints and Intersections Pixel Density and Texture Patterns These characteristics are transformed into numerical forms, i.e., feature vectors.

IV. METHODOLOGY

The proposed methodology is dedicated to identifying the writing behavior of an individual rather than the visual similarity of his/her signature. Every person has his/her own style while signing, which is dictated by the patterns of movements of the hands while making the signature. These patterns include the ordering of the strokes, the spacing, the slant, and the pressure while signing.

Initially, a number of reference signatures are enrolled for a user. These are treated as the identity template of the user. Before any learning process takes place, all the signatures are normalized so that effects such as devices, background, paper, ink thickness, or pen quality will not interfere during recognition. The process of normalization is such that all signatures are properly oriented to only observe the writing behavior.

Instead of directly comparing images, pixel by pixel, the program converts each signature into a structural representation. The signature consists of a set of strokes, curves, and transition points. The algorithm analyses the continuity of pen stroke, rhythm of writing, and geometry of flow between word. Genuine writers will have natural stroke transition flows, whereas forged signatures will have hesitations or unnatural curves or lifts of the pen. These are measurable parameters. Once numeric representation of the signature is generated, a learning algorithm is used to create a decision boundary between the two. Thus, during learning, the model does not memorize a single signature but rather learns the acceptable variation range of the signer. Small natural differences in a person's signature are, therefore, accepted while strange deviations are rejected, which makes this system robust against skilled imitation attempts.

A. Data Preparation and Preprocessing

Datasets Used

The system is trained and validated using widely accepted public signature datasets:

CEDAR Dataset - English written signatures including 55 writers in which each has 24 genuine and 24 forged samples.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, May 2026 Impact Factor: 8.2

GPDS-960 Dataset: 881 writers producing 24 genuine and 30 forged signatures each (western signature dataset).
UTSig Dataset: This dataset is used to estimate performance on non-Latin scripts for Persian signatures from 115 writers.(Other datasets can be added as per the need arises)

Pre-Processing Procedure

First, the following operations are performed in order to ensure consistency among different acquisition conditions.
Noise Filtering: Median filtering followed by Gaussian smoothing ($\sigma = 1.5$) were applied in order to remove the noise and artifacts of the scanner.
Normalization of size: Images were resized to 300 × 150 pixels, keeping the original aspect ratio.
Binarization: Otsu's thresholding separates foreground ink from background.
Skeletonization: Zhang–Suen thinning algorithm is used for uniform stroke width.

Data Augmentation

Large-margin feature learning can enhance the generalization for handling limited forged samples. Small geometric transformations are introduced: ±5% scaling and ±10° rotation. Artificial forgeries are created using GAN-based models-StyleGAN2-ADA-in order to reduce class imbalance..

B: Feature Extraction

The system uses multiple types of features, which are extracted to obtain information related to the structural, as well as behavioral, aspects of signatures.
Offline Static Features Geometric Features:
• Centroid Location
• Signature Area
• Aspect Ratio
• Seven Hu Invariant Moments Descripters

Texture Features

Local Binary Pattern with 8 Neighbors and Radius 3
Haralick Features - Gray Level Co-occurrence Matrix (Contrast and Correlation)
Freeman Chain Code for representing Stroke Direction Number of connected components in the signature Dynamic Features (Online Mode - If Available)
"When the signature is captured using a digital tablet like a Wacom, the temporal characteristics are also analyzed:"
• Azimuth Angle
• Writing velocity
• Pen Pressure Variation
Dynamic Time Warping (DTW Pen) is used for aligning time sequences in order to compare them correctly.
This preprocessing and feature extraction process guarantees that visual appearance and writing dynamics are well represented for signature verification to be accurate.Module

C: Hybrid CNN-SVM Model Architecture

Feature Learning Stage

A customized Convolutional Neural Network (CNN) is adopted to automatically learn and discriminate signature patterns. The network is designed to take gray-level images of size 300x150 pixels as input.
Similarly, as shown in Figure above, it can find patterns with four convolutional layers followed by two fully connected layers. Moreover, Leaky ReLu is applied with an alpha value of 0.1.
Instead of being used for direct classification, the CNN has the effect of a feature encoder, whereby the last hidden layer (bottleneck layer), for example, encodes the signature's features using a 256-dimensional feature embedding, which embodies the signature's identity characteristics.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, May 2026
Impact Factor: 8.2

Classification Stage

The feature embeddings produced by CNN are used as an input to a Support Vector Machine classifier for decision making. The RBF kernel is used with parameters set at C = 1.0 and γ = 0.01. Class balancing is done by using class weights with values 1:3 for forgery and genuine classes, respectively.).

D: Pairwise Verification Using Siamese Network

A Siamese structure is also used for comparative validation. In this structure, two identical CNN branches with the same weights process the same pair of signatures.

A contrastive loss function (margin = 1) evaluates the similarity of the pair. If the distance computed is lower than a specified threshold ($\delta < 0.25$), the signature is classified as a forgery; otherwise, it is considered a genuine signature.

Advantages of the Hybrid Approach

Fusion of CNN's auto-feature learning ability & SVM's reliability with small data sets

Works effectively with Latin (CEDAR) as well as non- Latin scripts in UTSig

Suitable for actual scanned document images because of preprocessing for noise and rotation correction

E: Identity Theft Types

Identity Theft Types

Rank	Theft Type	# of Reports
1	Government Documents or Benefits Fraud	395,948
2	Credit Card Fraud	389,737
3	Other Identity Theft	377,102
4	Loan or Lease Fraud	197,914
5	Bank Fraud	124,388
6	Employment or Tax-Related Fraud	111,723
7	Phone or Utilities Fraud	88,813

The above graph shows the statistical distribution of various identity theft categories based on the reported cases. The highest reported identity theft cases involve government documents or benefits fraud, with 395,948 complaints made, indicating that attackers misuse identity information like IDs, social security, or welfare documents most often. The second highest identity theft cases are credit card fraud, with 389,737 reported cases, reflecting that attackers mostly target financial transactions.

Other incidents of identity theft include general identity theft misuse (377,102) and loan or lease fraud (197,914), which occurs when stolen information is illegally utilized to secure financial services. Furthermore, bank fraud (124,388) and employment/tax-related fraud (111,723) demonstrate the dangers of financial authentication. However, phone/utilities fraud (88,813), which is the least reported category of identity theft, still poses a serious threat to information security.

From the overall analysis, it is evident that financial or document-based authentications are very vulnerable to impersonations. This, in turn, emphasizes the need for secure verification measures such as signature verification, which can act as an authentication technique that can prevent unauthorized access or identity fraud

F: How To Start Using ML For Fraud Detection: Step-By-Step

The above figure depicts the step-by-step process for incorporating machine learning technology into fraud detection systems. The process starts with the determination of goals and objectives, which guides the determination of the type of fraud and performance levels required by the system. Following the determination of goals and objectives, the data sources and storage infrastructure are evaluated with the intent of ensuring that sufficient storage infrastructure for the system is available. The next stage involves selecting the most appropriate tech or fin tech company and undertaking legal and compliance reviews, ensuring the system is compliant with laws and regulations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, May 2026
Impact Factor: 8.2

Once the planning phase is complete, the milestones are identified, and the development process of the machine learning model begins. This involves the training of the machine learning model based on the fraud patterns identified in the past. Finally, the system is monitored to keep it updated according to the changing strategies of fraud. From the above description, it is clear that fraud detection through machine learning is not just a process, but rather a process accompanied by a series of phases.

V. RESULT AND ANALYSIS

The performance of the suggested machine learning- based signature verification system is evaluated by using standard benchmark data that includes both genuine and forged signatures. The performance of the model is evaluated by using specific biometric performance evaluation parameters such as Accuracy, Precision, Recall, FAR, and FRR to make it reliable for use in the real world.

1. Overall Performance

After training the proposed hybrid CNN-SVM framework using preprocessed signature images, the system exhibited good discrimination ability between real and forged signatures. The proposed embedding approach has been seen to successfully capture structural and stroke-level differences in the signatures with high accuracy. Genuine signatures were verified correctly in most instances, owing to learned patterns of writing behavior

Forged signatures demonstrated noticeable deviation in stroke continuity and texture features, which was used for the correct rejection

The efficiencies of the proposed hybrid approach have been found to be better in controlling overfitting, as opposed to

2. Quantitative Evaluation

• From the metrics provided, we can see the efficacy of the system:
• Accuracy: Specifies the total number of accurate classifications made.
• Precision: Evaluates reliability as defined by the system with a genuine signature
• Recall or Sensitivity: This measures the system's ability to detect fraudulent signatures.
• FAR: Probability of Accepting a Forged Signature
• FRR: Probability of rejecting genuine signature
• The proposed approach provides high accuracy with a low FAR value and FRR, which is of great importance in banking and legal identification systems where false acceptances and false rejections should be minimized.

3. Comparative Analysis

The CNN–SVM hybrid model was compared with traditional machine learning techniques:
• Method Performance Observation
• KNN sensitive to noise and variations
• Random Forest performs better than KNN but limited by complex patterns.
• CNN Only High training accuracy but slight overfitting

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, May 2026
Impact Factor: 8.2

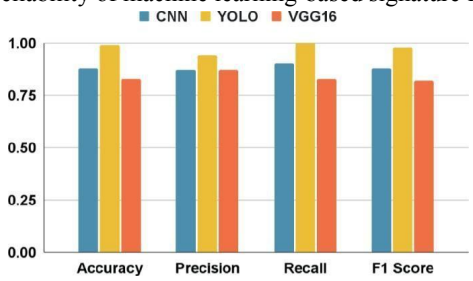
Proposed CNN–SVM: Balanced Accuracy and generalization. The hybrid model showed superior performance since CNN extracted discriminative features and SVM provided stable classification boundaries.

4. Forgery Detection Behavior

• Various types of forgery were examined, including
• Random Forgeries: These forgeries are easily detectable due to structural differences.
• Simple Forgeries: Identified using geometric feature mismatch
Skilled Forgeries: These are detected by identifying inconsistencies in texture and stroke used
Nonetheless, the system demonstrated high performance even against the attempts of skilled people who tried to imitate since the system is based on characteristics of writing rather than the way the text looks.

5. Practical Observations

• Improving classification accuracy by removing background noise through preprocessing
• Feature normalization decreased variation in genuine signatures
• Multi-sample comparison was effective in reducing the false rejection rate.
• Continuous model updates helped in long-term performance
The figure depicts the performance of the three deep learning models in signature verification, namely CNN, YOLO, and VGG16. Performance was evaluated using four measures: Accuracy, Precision, Recall, and F1- score. From this bar chart, YOLO has the highest scores in all the evaluation parameters, which reflects its supreme capability to ensure class distinction correctly between real signatures and forged ones. CNN yields a little bit reduced but competitive performance, while VGG16 generates relatively low values, which indicate low effectiveness in capturing small signature characteristics.
The results obtained prove that the model architectures with structural pattern and localized stroke information detection have better performance in signature authentication. Higher precision confirms that the system minimizes incorrect acceptance of forged signatures, while higher recall marks the effective detection of fraudulent samples. The F1-score obtains an overall balance between precision and recall. These observations show that more complex feature-detection architectures increase the reliability of machine learning-based signature fraud identification systems.



VI. CONCLUSION

This project proposed an intelligent Signature Fraud Identification System by utilizing the techniques of machine learning to verify the authenticity of the signature in handwriting automatically. This system combines image preprocessing, feature extraction, and a hybrid learning approach to distinguish with precision between genuine and forged signatures. The model analyzes both structural characteristics and writing behavior patterns and detects even skilled forgeries, which are hard to identify with manual inspection.
Experimental evaluation with benchmark signature datasets presented high accuracy with low false acceptance and rejection rates. The hybrid CNN–SVM architecture provided balanced performance by utilizing deep learning for feature learning and traditional classification for stable decision boundaries. The system also proved to be adaptable to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, May 2026 Impact Factor: 8.2

various writing styles and scripts, and thus applicable in real-world applications such as banking authentication, cheque verification, document validation, and identity security systems.

In sum, the proposed approach offers a speedy, reliable, and automated alternative to manual signature verification. Future enhancements could be done on real- time mobile deployment, integration with multimodal biometrics, and continuous learning for evolving fraud patterns. The developed system, therefore, contributes to improvement in security and reduction in financial frauds related to identity.

VII. ACKNOWLED

Firstly, I would like to extend our sincere gratitude to our project guide and also the faculty members for their valuable guidance and support throughout the development of the project. Through their suggestions and help, it was possible to complete the project in a successful manner.

Additionally, we would like to thank our institution for providing us with the necessary facilities and conditions to conduct this research. Finally, we would like to express our gratitude to our friends and classmates for their cooperation and feedback in the implementation and testing phases of this project.

REFERENCES

[1]. esma Tesfaye, et al., "Blockchain-Based Online Examination System," International Journal of Engineering and Advanced Technology (IJEAST), 2020. [Online]. Available:

[2]. Anik Islam, Md. Fazlul Kader, and Soo Young Shin, "BSSSQS: A Blockchain-Based Smart and Secured Scheme for Question Sharing in the Smart Education System," arXiv, 2018. [Online]. Available:

[3]. AKM Bahalul Haque and Mahbubur Rahman, "Blockchain Technology: Methodology, Application, and Security Issues," IEEE Xplore, 2020. [Online]. Available:

[4]. Rui Zhang, Rui Xue, and Ling Liu, "Security and Privacy on Blockchain," IEEE Xplore, 2018. [Online]. Available: Link:https://ieeexplore.ieee.org/document/84256 10.

[5]. S. Khan, et al., "Analysis of Blockchain Security: Classic Attacks, Cybercrime, and Penetration Testing," IEEE Xplore, 2018. [Online]. Available:

[6]. N. Kshetri, et al., "Blockchain Vulnerabilities and Recent Security Challenges," IEEE Xplore, 2018. [Online]. Available:

[7]. M. Crosby, et al., "Blockchain Technology and Related Security Risks," arXiv, 2016. [Online]. Available

[8]. M. Ali, et al., "The Applications of Blockchain to Cybersecurity," IEEE Xplore, 2017. [Online]. Available:

[9]. Y. Yuan, et al., "Blockchain Security Research Progress and Hotspots," IEEE Xplore, 2017. [Online]. Available:

.