

RESEARCH ARTICLE

Enhancing Security: Infused Hybrid Vision Transformer for Signature Verification

MUHAMMAD ISHFAQ¹, AYESHA SAADIA², FAEIZ M. ALSERHANI³, AND AMMARA GUL⁴

¹Department of Cyber Security, Air University, Islamabad 44000, Pakistan

²Department of Computer Science, Air University, Islamabad 44000, Pakistan

³Department of Computer Engineering and Networks, College of Computer and Information Sciences, Jof University, Sakaka, Al-Jouf 72388, Saudi Arabia

⁴Faculty of Computing, Engineering and the Built Environment, Birmingham City University, B4 7XG Birmingham, U.K.

Corresponding authors: Ayesha Saadia (ayesha.saadia@mail.au.edu.pk) and Faeiz M. Alserhani (fmserhani@ju.edu.sa)

ABSTRACT Handwritten signature verification is challenging because there is a huge variation between the orientation thickness and appearance of handwritten signatures. A strong signature verification system is essential to refine the accuracy of confirming user authentication. This investigation introduces an inclusive framework for training and evaluating hybrid vision transformer models on diverse signature datasets, aiming to refine the accuracy in confirming user authentication. In previous studies, transformer & MobileNet were used for computer vision classification and signature verification separately. Drawing inspiration from the Convolutional Neural Network (CNN), the hybrid model is proposed as a deep-learning model (ResNet-18 & MobileNetV2) with the Vision Transformer model (proposed method 1 & proposed method 2). To bring originality to this study, we excluded the final layer of the feature extractor and smoothly integrated it with the initial layer of the vision transformer. In the scope of this research, we introduced a unique hybrid vision transformer model. Furthermore, we incorporated swish and tangent hyperbolic (tanh) activation functions into the validation model to enhance its performance. Experimental results showcase the effectiveness of the proposed hybrid model, achieving notable accuracies on various datasets, including 92.33% accuracy on Bhsig-Bengali, 99.89% accuracy on Bhsig-Hindi, 99.96% accuracy on Cedar, and 74.09% accuracy on UTsig-Persian datasets, respectively. The practical implications of this research extend to real-time signature verification for secure and efficient user authentication, particularly in mobile applications. This advancement in signature verification technology presents new possibilities for practical use in diverse scenarios beyond academia.

INDEX TERMS Vision transformer ResNet-18, MobileNetV2, handwritten character verification, signature verification, hybrid vision transformer, handwritten signature verification, UTsig-Persian.

I. INTRODUCTION

In the era of rapid technology progression, the need for a robust identity verification system is more pronounced than ever. As we continue to navigate in the digitally interconnected world, the authentication of personal information has become a critical aspect of various sectors, particularly in the realms of finance, government, and legal documents. Among the myriad methods employed the significance of handwritten signature persists, encapsulating unique bio-metric traits [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil¹.









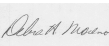
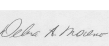
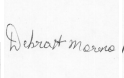
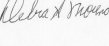




As we look at the history of mankind signatures have been considered the best source of identification [2].

The early man used signatures to form a truce, build empires, sell laborers, sell people into slavery, and take over others' property. All the mischief's related to documents done in history were part of signature theft. In today's developed era still, we are facing identity theft, illegal document ownership, fraud, and illegal property loans [3]. These problems are caused by skillful people who can forge the signatures of individuals with such accuracy that it is difficult for the naked eye to identify the fake from the real signature. The signature owner only finds out after he/she loses his/her financial aid from their digital bank accounts [4].

Signature verification can be carried out using either **i)** online or **ii)** offline methods, depending on how information is gathered [5]. If the information is obtained in real-time from devices like stylus and their authenticity remains intact then we can use it for an online verification system. For the offline, we can get data from public repositories or websites that provide datasets for the research that are authentic for the offline signature verification systems [6] or we can create our dataset with the permission of authors. The information provided by the online signature includes angle, speed of strokes, and emotions whereas the offline signature will have less information as it will be the image of the original signature [7]. As there is less information, offline authentication is somewhat difficult although many scientists have proved high accuracy results still there is much potential in the practical and designing of applications for signature verification.

The forgeries of signatures are of different levels depending upon similarities [8], the more similar are considered more skilled, a real signature image and a skilled fake signature image are shown in Table 1.

TABLE 1. Samples of genuine and forged signatures.

| Dataset | Genuine | Genuine | Forged | Forged |
|--------------------|---|---|---|---|
| BHSig-Bengali [9] |  |  |  |  |
| BHSig-Hindi [9] |  |  |  |  |
| CEDAR [9] |  |  |  |  |
| UTSig-Persian [10] |  |  |  |  |

In this study, we tackle the critical task of signature authentication by devising a machine learning-driven methodology aimed at discerning between genuine and forged signatures. Through rigorous experimentation utilizing a comprehensive dataset comprising genuine and forged signature images, we apply advanced feature extraction techniques detailed in Section III-D to capture salient characteristics. Given the sensitive nature of signature data, utmost care is taken to ensure privacy and confidentiality following ethical considerations; details regarding the dataset handling are elaborated in Section III-D. Given the pivotal role of signatures as individuals' primary form of identity verification, stringent measures are implemented to safeguard data integrity and privacy. Our results highlight the effectiveness of our proposed models, In distinguishing between real and fake signatures. This proves to be the best solution for diverse domains that require signature verification protocols for security management.

The image classification systems can be classified into three groups depending on the model [11]. Notably they can be identified as 1. data acquisition, 2. Pre-processing and 3. last Validation. The Development of artificial neural networks has resulted in the creation of convolutional neural network models specifically for verifying handwritten signatures [12]. Deep neural networks are currently viral and researched subject [13]. This has led to the use of the robust technique of transformer in the validation of handwritten signatures [14], [15], [16].

Handwritten signature recognition is challenging because there is a huge variation in the orientation thickness and appearance of signatures written by individuals at the same time [17], [18]. A new and robust approach is required for signature verification to enhance user authentication accuracy.This paper explores the complexities of signature verification systems, tracing their development from traditional techniques to modern innovations. By addressing the issues of fake signatures and the rising necessity for secure identity validation, this research seeks to further the field of signature verification technology. Drawing inspiration from the convergence of artificial intelligence and deep learning methodologies, this study aims to propose an innovative approach for signature verification, paving the way for enhanced accuracy and efficiency in identity authentication.

In this research, we will be building a robust signature verification system. This research can be used in mobile devices to verify an individual's identity using signatures by building Android applications that can verify original signatures from fake, by this process we will be able to create a local database of individual users and verify their signatures from scanned documents and predict. We proposed a hybrid model by using pre-trained MobileNetV2 [19], and ResNet-18 [20] combined individually with Vision Transformers. Traditional models used a single model (Conventional Neural Network, MobileNetV2 [19], or swim-transformer (Machine learning model named by authors) [21]) for signature verification, our proposed model outperformed these traditional models and obtained high accuracy. For novelty we also added two layers in the feature extractor module other than the originally available layers, we also added tangent-hyperbolic and sigmoid activation functions with the dropout layer. These models reached high accuracy results with minimum false acceptance rate, false rejection rate and compilation time for each Experiment will also be recorded by our system.

A. ATTACKS ON SIGNATURES

In Cyber-security, businesses suffer from various challenges regarding the signature verification system. One significant threat is forgery, replicating legitimate signatures, and compromising the integrity of the document and transactions. Signature manipulation heightens the risk of cyber-security, with Cyber adversaries employing digital tools to counterfeit authentic signatures. In the tech industry, the process of signature tracing is challenging, requiring sophisticated software to monitor and replicate authentic signatures. The rise in

bio-metric spoofing complicates matters further, as attackers strive to trick bio-metric signature recognition systems using fraudulent data. To tackle these challenges, organizations must develop and implement robust cyber security strategies to enhance their signature-based authentication systems, effectively defending against the ever-evolving tactics of cyber criminals. A proactive approach is Vital for preserving the integrity and security of authentication processes in the ever-changing and complex field of cyber security. Our research aims to build an understanding of new techniques for the developers to build a system that can easily verify the signatures easily and accurately.

B. CHALLENGES

The main challenges that come up while verifying signatures are detecting skilled forgery (professional tries to copy signature or uses instruments to copy signatures), intra-writer variability (individuals can write their signature in different form depending upon their moods and requirements), random forgery (randomly any one tries to copy individuals signature) and simple Forgery.

The proposed model tackles the challenges of handwritten signature verification by integrating ResNet-18 with Vision Transformer and MobileNetV2 [19] with Vision Transformer. This technique results in capturing local as well as global features of signatures, improving its ability to distinguish between real and fake signatures. Advanced activation functions like Swish and Tanh are used to enhance its generalization capabilities, making it more adaptable to different signature features. Additionally, the model's real-time processing ensures fast and accurate verification, lowering the chances of forgery and bio-metric spoofing.

1) INTRA-PERSON VARIABILITY

Individual signature can change each time a He/she signs because of a lot of factors like mood, health, or the speed at which he/she signs.

2) INTER-PERSON VARIABILITY

Everyone has a unique signature, and distinguishing these different styles accurately can be challenging.

3) FORGERY DETECTION SKILLED FORGERIES

Some people practice replicating another person's signature, making it look very similar to the original and difficult to detect.

4) RANDOM FORGERIES

These signatures are created without any knowledge of the genuine signature. While they are generally easier to spot, their unpredictable nature can still be challenging.

5) SIMPLE FORGERIES

When someone copies or traces a genuine signature, it can sometimes be very accurate and hard to distinguish from the real one.

The benchmark dataset used in this research are Bhsig-Bengali [9], Bhsig-Hindi [9], Cedar [9] and Utsig-Persian [10] dataset. The proposed hybrid vision transformer model demonstrates high performance on These datasets. The main difference between these dataset is the language barrier and thus they have different features for machine models.

The following sections of this paper are organized in the following manner.

- Section II provides related work of the relevant studies in the dynamic handwritten signature verification system.
- Section III provides a detailed overview of the proposed methodology for this research.
- Section IV gives details about the results of the discussion and comparison of the proposed systems.
- Section V provides an overview of the conclusion and future work.

II. LITERATURE REVIEW

To understand the historical context of this research, it is essential to delve into basic concepts of artificial intelligence initially. Artificial intelligence encompasses the capacity of a computer system to perform tasks traditionally carried out by human intelligence [20]. Within the broader realm of artificial intelligence, machine learning (ML) stands out as a subset of algorithms proficient in handling intricate tasks by acquiring knowledge from data, departing from the conventional dependence on hard-coded rules or heuristics [22].

The technical innovation lies in the seamless integration of ResNet-18 [20], MobileNetV2 [19], and Vision Transformer architectures. ResNet-18 [20] and MobileNetV2 [19] are used for their strong feature extraction capabilities, capturing fine details of the signatures. The Vision Transformer adds the ability to model long-range dependencies and global context through self-attention mechanisms. By excluding the final layer of the feature extractors and connecting them with the initial layer of the Vision Transformer, the model benefits from a holistic representation of signature features, leading to improved accuracy and robustness.

A. BACKGROUND OF MACHINE LEARNING

Machine learning represents a shift away from the conventional approach based on manually devised rules, allowing computer systems to learn and improve performance through data-driven processes [23]. This change in paradigm places a strong emphasis on data, enabling algorithms to recognize patterns, make predictions, and continually optimize performance [24]. Positioned as a fundamental component in the wider field of artificial intelligence, machine learning holds the potential for intelligent systems that can autonomously learn and adapt [25]. This research delves into the integration of Computer Vision (CV) and Machine Learning (ML) as its primary focus.

Artificial neural networks, often called neural networks, are computational structures inspired by the organization and functioning of biological neural networks present in the

brains of living organisms [26]. These networks comprise multiple layers of interconnected neurons, where each layer's neurons are linked in a specific manner to the neurons in the subsequent layer. Serving as potent representation learning tools, neural networks possess the capability to autonomously learn and derive significant representations or functions from input data. Their application extends to diverse learning tasks, encompassing classification and prediction, among others [26].

The proposed neural network designed for the binary classification images is represented in Figure 1. The initial layer serves as the input layer, accommodating various data forms such as images, binary data, information from preceding layers, or properties from data rows in tabular data—tailored to the specific problem and data format. Situated between the input and output layers, the hidden layers preserve internal representations. Akin to the way humans interpret the complexities of images in our surroundings. Ensuring the model's predictive accuracy requires these hidden layers to encapsulate abstract information.

Initially, two hidden layers are represented in the above Figure, but in practice neural networks can have any number of hidden layers depending on the complexity of the problem each model has a different number of standard layers. The last two layers are the layers of the Vision transformer used in this research. We can always add and remove or even select any layer according to our problem requirement.

The concluding layer is the output layer, responsible for producing predictions. In instances of binary classification problems, like distinguishing the authenticity of a queried signature, this layer is comprised of a solitary neuron. When hyperbolic tangent function (\tanh) is applied, it transforms the input in-between 1 and -1 , this helps neurons to model complex relations and capture non-linear information in data. The classification is usually considered “True” if the output value exceeds 0, and “False” otherwise. This facilitates the categorization of the signature class based on the output of the \tanh function [27].

Figure 1 presents the architecture diagram of an artificial neural network (ANN) meticulously designed for image classification tasks. The ANN structure comprises interconnected layers, each meticulously crafted to fulfill a specific role in processing input data and making accurate predictions. The neural network's process begins with the Input Layer, which receives image data. This data then flows through Convolutional Layers, where convolutional operations are used to extract relevant features. These layers are often paired with activation functions such as the Rectified Linear Unit (ReLU) to introduce non-linearity. To control the spatial dimensions and computational load, Pooling Layers like max pooling or average pooling are applied after convolutional operations. Following this, Fully Connected Layers, or dense layers, come into play. These layers are responsible for complex feature learning and decision-making by creating connections between every neuron in one layer and every

neuron in the next. The network culminates in the Output Layer, which produces predictions. In image classification tasks, this layer includes neurons for each class and uses soft-max activation to estimate probabilities. To prevent over-fitting, Dropout Layers are employed. These layers randomly deactivate a portion of neurons during training, which helps improve the model's generalization capabilities. Additionally, Batch Normalization is used to standardize inputs across layers, which enhances training stability and speeds up convergence towards optimal solutions.

The researchers introduced a U-NET model enhanced with Multi attention-gated modules and residual blocks specifically for medical image segmentation [28]. This approach effectively minimized noise, which is crucial for the precise identification of diseases in X-ray images. They utilized unique activation functions, including Mish and ReLU, [29] and applied AdamW and Adam optimization strategies to improve performance. Their findings demonstrated that this model surpassed traditional U-Net and Res-UNet models on benchmark datasets, such as Brain Tumor MRI and Skin Lesion data.

LSTMs are best choice in capturing the temporal dependencies, Thus making them suitable for predicting the patterns of Water flow. For this reason The researchers used it to discover the water-cycle time-series data for basin of Gilgit river. Their results indicated that LSTM model significantly improve prediction accuracy for predicting the flow of water at Gilgit Basin compared to traditional statistical methods [30].

In some other research, The authors used Grad-CAM (Gradient Weighted Class Activation Mapping) For the purpose of identifying brain tumors from the MRI images. Grad-CAM uses forward and backward pass for image classification, Where at initial stage images are passed to gather feature maps from Deep CNN model and then they are backward passed to gradient the score for the classes are calculated from the feature maps. Thus making it more relevant to making decision accurately and interprets the results, which are crucial in making diagnostic decisions [31].

In an earlier research [21], a deep-learning approach was introduced and applied to signature detection, featuring three significant improvements. Firstly, a novel Signature network was established based on a feature pyramid architecture to glean more comprehensive signature information [32]. Secondly, an image pre-processing procedure, encompassing image splicing [33], [34] and pixel value transformation [35], was developed to enhance the contrast of original images. Initially conducted on random images, these tasks were subsequently adapted for signature images [35], [36]. The study sought to explore the extent to which transfer learning from deep convolutional neural networks (CNNs), pre-trained on signature images, could be employed for automated signature verification systems in banks and other official establishments [37]. Notably, the CNN model was trained on official documents storing signatures, addressing

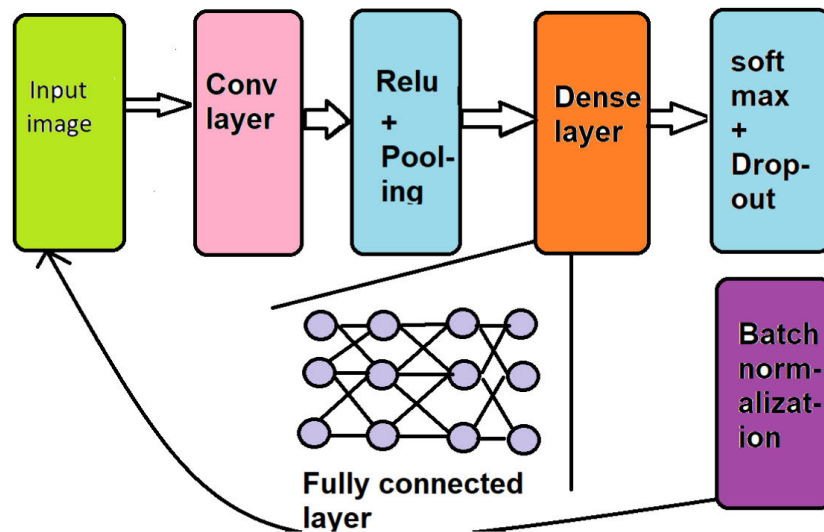


FIGURE 1. Architecture diagram of artificial neural network.

the challenge of disparate storage systems in individual banks, which had previously posed a significant impediment.

B. COMBINING CNN AND TRANSFORMERS FOR SIGNATURE CLASSIFICATION

The transformer approach was used by researchers in which an original image is received as input to the system and the cropped original image is then made input for the central system. The up-sampling enhancement module directly steers the model to focus on useful information and results in classifying the received image as fake or real. The researchers have used shallow techniques on CNN in such a way that they have developed fewer layers for the classification which in return resulted in fewer features for classification [38]. They have also missed out on quite a lot of features for sampling. They used a dataset with minimum images for testing has caused the over-fitting, a result they were able to generate high-value results. The researchers proposed a technique to overcome this over-fitting problem by using a benchmark data set like GDPS synthetic although it outperformed many techniques still it was not good enough because they applied the technique on only 1 dataset with the highest accuracy results and the dataset contained round about 3952 images.

Again researchers proposed a model in which real and fake signature images were used on the benchmark data They used the 1:5 for real and fake images in the training and testing but this ratio should have been equal so that Artificial Intelligence must be trained properly without any loss of data shortage [37]. Other Chinese researchers have proposed an encoder and decoder technique to classify signatures [38]. According to this technique, they first encoded the real signatures and then they used a decoder to decode the real signature and match it with the encoded signature image. It proved quite a technique but it lacked one thing it did not use a local dataset rather it used a benchmark dataset for

training and local for testing the results that they received were also extraordinary. The researcher tried to identify signatures by forming a hybrid method by combining CNN & Bidirectional Long Short Term Memory network [33]. Firstly they use CNN for pre-processing and feature selection of the signature images and then that data is passed to a Bidirectional Long Short Term Memory network for classification. The hybrid Deep learning network classifies the input signature as skilled forgery or genuine with high accuracy.

Prove that CNN is still the hot topic of research for signature verification systems among researchers around the world [9], [33], [39]. It is not wrong to say that CNN with no doubt is the father of image classification, face detection, AI creation, and computer vision tasks. CNN paved way for the researchers to perform computer tasks. CNN models have achieved remarkable accuracy and made the tasks of forensic experts much easier with minimum human error. CNN excels at automatic learning of hierarchical and spatial features without manually performing those tasks. CNN has proven itself in performing identification classification tasks.

The transformer integration in signature classification is perhaps the first of this kind of architecture [33]. They used the approach NLP (natural language processing) technique in the signature verification system. According to them, they used two streams for signature verification one stream is used for up-sampling, and the second stream is used for the prediction of the results. In Up-sampling the input image is broken down into batches and only the most acceptable batch information is selected and results are generated.

It is the latest technique which is composed of CNN and a vision transformer [36]. Proposed a new technique in machine learning for automatic detection of diseases and to prevent adversarial attacks to misdiagnoses and disastrous consequences in the safety realm. We have adopted our

research approach from them and implemented an artificial neural network with a vision transformer to achieve the best results for the signature classification.

This study aims to address several limitations and gaps in existing research, including:

Accuracy Across Diverse Datasets: The hybrid approach enhances accuracy by combining CNN and transformer architectures.

Inconsistency in Real Signatures: The models extract global and local features at same time thus improving it's ability to predict accurately across different writing styles.

Fabrication and Hoax detection: The unique feature extraction technique and attention mechanism increase the model's ability to identify fake signatures inconsistently.

Live Conformation: The model's high performance proves it can detect fake signatures in real life (like application can be applied in mobile devices).

Elasticity: The model's performance on Benchmark dataset's Shows its potential for flexible deployment in reality scenarios.

In short, various techniques have been explored for signature verification, each showing it's own features and disadvantages. This study explores itself by using a hybrid model by joining features from CNN, and a two-channel, two-stream transformer-based method. Building our understanding on recent developments in NLP techniques and the successful integration of artificial neural networks combined with vision transformers. This remarkable integration places our approach addressing limitations identified in previous research. This extra ordinary approach positions our investigation at the edge of signature verification, presenting potential advancements in accuracy and reliability within the fields of artificial intelligence and machine learning.

III. PROPOSED METHODOLOGY

The proposed hybrid visual transformer model is meticulously designed to efficiently capture and process features from input images for handwritten signature verification. This methodology seamlessly integrates components from established CNN architectures, specifically ResNet-18 [20] and MobileNetV2 [19], with a custom-designed vision transformer architecture. Initially, MobileNetV2 [19] serves as the feature extractor, adept at processing input images and extracting relevant visual features crucial for signature verification tasks. Subsequently, adaptive pooling and flattening techniques are applied to obtain a fixed-size representation of the extracted features, facilitating further processing by subsequent layers. The vision transformer architecture, comprising linear layers with relu activation functions, dropout layers, and batch normalization layers, is then employed for higher-level feature processing and classification. The model's output layer generates predictions for signature verification, with softmax activation converting raw output scores into probability distributions over signature classes. The training process involves iterative passes through the dataset, utilizing the Cross-EntropyLoss criterion, adam

optimizer [40], and StepLR scheduler for efficient parameter optimization and learning rate adjustment. Validation is conducted using a separate dataset to assess the model's generalization capability and identify potential issues such as over-fitting or under-fitting, thereby ensuring robust performance in real-world scenarios.

The choice and quality of dataset are the key points to success or failure of any machine learning model.

- Section III-D delves into the features of selected benchmark data sets, Highlighting its importance, relevancy and diversity to ensure a detailed evaluation of proposed methodologies.
- Section III-E has detail about evaluation metrics used in our experiments.
- And finally section III-F store the details about experimental setup, parameters and configurations and any extra considerations providing rigorous evaluation for validation of our proposed method.

A. PREPROCESSING

Preprocessing is a Vital step for development of any Machine Learning model, particularly tasks involving image classification. Pipelining our Preprocessing, we tackle the discrepancies of class labeling in benchmark datasets, ensuring uniform images of real and fake signatures and choosing the standard size of images. Notably, the benchmark datasets such as CEDAR [9] and BHSIG-BENGALI [9] utilize a class labeling scheme ranging from 1 to 100, while the BHSIG-HINDI [9] benchmark dataset extends this range to 160 classes. In contrast, the Utsig-Persian [10] dataset adopts a distinct approach, where the main class denotes real & forgery, and sub-classes are labeled from 1 to 115. Despite these variations, all images undergo resizing to a standardized dimension of 224×224 pixels. This resizing process ensures uniformity in image dimensions, facilitating efficient processing and model training. Additionally, standardization techniques are applied to mitigate variations in image attributes, such as brightness and contrast, further enhancing the consistency and reliability of the dataset. By harmonizing class labeling and image dimensions through preprocessing, we create a cohesive dataset conducive to robust model development and accurate signature verification.

In Figure 2 The model is given a single input of signature as an image The first ViT [41] module performs preprocessing step like converts the size of image to the one same size for this purpose we have used the tensor function of vision transformer. The tensor function performs the preprocessing by removing noise and extracting features from the input image and then converts image into binary using tensor and sends the tensor values to the MobileNetV2 [19] and ResNet-18 [20] model. Where these two models extract the local features of input image. After matching they send the values to the ViT [41] model again which extracts global features and verify the signatures by matching all the image data and

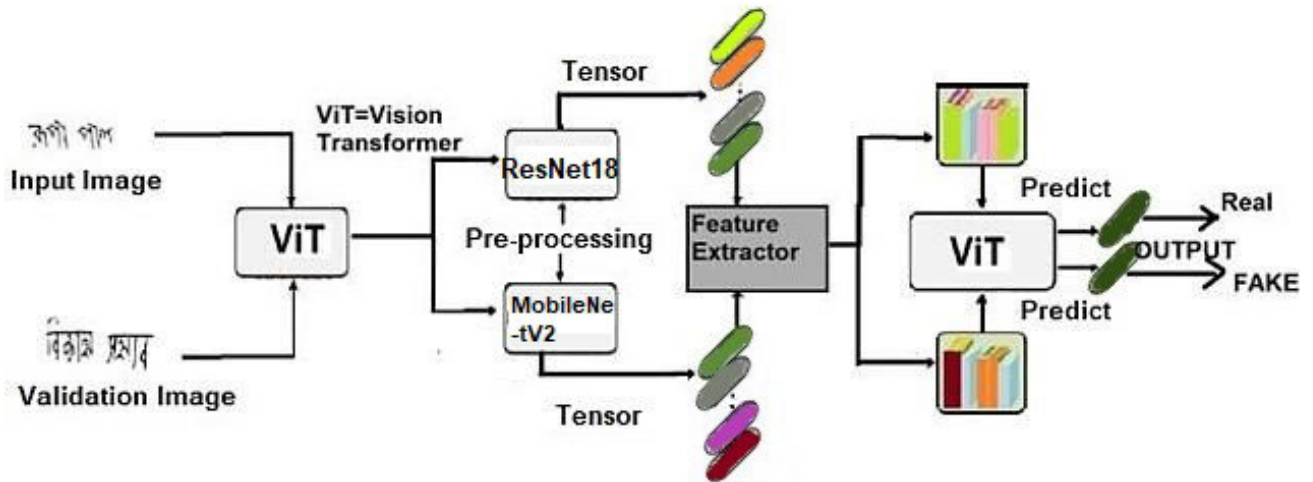


FIGURE 2. The proposed methodology.

predicts input image as a real or fake signature. So we have used ViT [41] two times for two different tasks.

B. IMAGE TRANSFORMATION

Data transformation refers to the modification of the initial input data to align it with a particular task or model. In the field of machine learning and computer vision, this procedure typically encompasses a set of operations applied to input data, such as images or text, to ready it for training or evaluation. These operations may involve resizing, cropping, normalization, and various pre-processing steps. The central objective of data transformation is to improve the quality, relevance, and adaptability of the data for a specific machine learning purpose. In the process of preparing input data for the proposed deep learning model, we have implemented a sequence of transformations aimed at improving the quality and appropriateness of the dataset—initially, the `Transforms.Color-Jitter` with `brightness = 0.2`, `contrast = 0.2`, `saturation = 0.2`, `hue = 0.2` operation is employed. After these spatial and color adjustments, the `Transforms.ToTensor()` transformation is implemented to convert images into PyTorch [42] tensors, ensuring compatibility with the deep learning model—finally, the `Transforms.Normalize` with `mean = [0.485, 0.456, 0.406]`, `std = [0.229, 0.224, 0.225]` operation is utilized to normalize tensors using specified mean and standard deviation values, thereby standardizing the input for the neural network. These sequential transformations collectively contribute to the resilience and adaptability of the model by introducing diverse perspectives and normalizing the input data.

C. PROPOSED METHOD

The proposed methodology presents a hybrid approach for authenticating individuals' signatures, accommodating diverse writing styles by merging MobileNetV2 [19] and ResNet-18 [20], both pre-trained deep learning models, with a Vision Transformer.

This fusion capitalizes on the diverse strengths of these architectures to enhance signature verification. Initial data preparation involves curating a dataset encompassing genuine and forged signatures sourced from diverse datasets such as BHSIG-BENGALI [9], BHSIG-HINDI [9], CEDAR [9], and UTSIG-Persian [10]. Each image is processed to convert it to a standard size of 224×224 from the original size of 5331×331 (the size of image in the benchmark dataset), resulting in maintaining uniformity through the database. To further classify the data techniques such as rotation, scaling, and flipping are applied to benchmark datasets for enhancing model generalization.

The model is trained using the Adam optimizer [40], with a learning rate scheduler to adjust the learning rate and minimize the cross-entropy loss function, which will help us align the predicted signatures with the actual ones. Dropout layers are incorporated to prevent over-fitting, and model parameters are updated over multiple epochs based on gradients from the training data. After training, we evaluate the model's performance on a separate validation dataset to test its ability to generalize. We use metrics such as accuracy, precision, recall, and F1-score to measure how effectively the model distinguishes between genuine and forged signatures. Any observed discrepancies or limitations during validation are meticulously analyzed, prompting potential adjustments to the model architecture or training process to enhance performance.

The overarching objective of this methodology is to elevate the precision and efficacy of signature verification by strategically integrating principles from deep learning and harnessing the complementary capabilities of MobileNetV2 [19], ResNet-18 [20], and Vision Transformer architectures.

D. DATASET

The dataset required for this research was downloaded from the publicly available dataset site Kaggle [9]. The handwritten signature data consists of three datasets. The first dataset in the folder was Bhsig-Bengali, the second

Algorithm 1 Signature Verification Algorithm

Input: Signature dataset

Output: FAR, FRR, ACCURACY, COMPILATION TIME, EXECUTION TIME

Model Architecture:

- 1) Load pre-trained ResNet-18 [20]/MobileNetV2 [19] model
- 2) Define vision transformer architecture
- 3) Combine ResNet-18 [20] and vision transformer to create a hybrid model

Training Loop:

- 1) Define optimizer and loss function
- 2) Initialize empty lists for training metrics
- 3) Set the model to training mode
- 4) Initialize total loss, correct predictions, total samples
- 5) for each batch in the training loader do
 - a) Forward pass through the model
 - b) Compute loss and perform a backward pass
 - c) Update optimizer's parameters
 - d) Update total loss and correct predictions
- 6) Calculate accuracy and average loss
- 7) Append metrics to the respective lists

Validation:

- 1) Set model to evaluation mode
- 2) Initialize validation loss, correct predictions, and total samples
- 3) for each batch in the validation loader do
 - a) Forward pass through the model
 - b) Compute validation loss
 - c) Update correct predictions
- 4) Calculate validation accuracy and average loss
- 5) Append metrics to the respective lists

FAR, FRR Calculation:

- 1) Set threshold for signature verification
- 2) Initialize FAR, FRR, total authentic, total impostor
- 3) for the batch in the validation loader do
 - a) Forward pass through the model
 - b) Calculate probabilities and predicted labels
 - c) for each sample in the batch do
 - i) if the actual label is authentic then
 - A) Update total authentic and FAR
 - ii) else
 - A) Update total impostor and FRR
- 4) Calculate FAR and FRR percentages
- 5) Append FAR and FRR to their respective lists

one was Bhsig-Hindi and the third one was the Benchmark dataset, Cedar. Bhsig-Bengali contains the signatures of 100 authors, each author has 24 original signatures and 30 forged signatures. The Bhsig-Hindi dataset contains signatures from 160 people each person has 24 real signatures

and 30 forged signatures. The Cedar dataset contains signatures by 100 people each person has 24 real signatures and 24 forged signatures. Apart from these, we have a Utsig-Persian dataset that contains signatures by 115 people each person has 27 real signatures and 42 forgeries. The images in the benchmark dataset are of three different formats including—PNG, TIFF, and jpg format [43]. The Utsig-Persian dataset is obtained from the original site of the University of Tehran [10]. Table 2 shows the number of authors, genuine, and forged signatures, and the writing style of those signatures. For the visuals of the signature visit Table 2.

TABLE 2. Dataset authors, genuine, forged, style.

| Name | Style | Authors | Genuine | Forged |
|--------------------|----------------|---------|---------|--------|
| BHSig-Bengali [9] | Bengali Script | 100 | 24 | 30 |
| Cedar [9] | Latin Script | 55 | 24 | 24 |
| BHSig-Hindi [9] | Hindi Script | 160 | 24 | 30 |
| UTsig-Persian [10] | Persian Script | 114 | 27 | 42 |

E. EVALUATION METRICS

The analysis of our data has revealed several noteworthy results, which are discussed in this portion. To measure performance we use training and validation accuracy and loss are calculated for each hybrid model. In our model evaluation, a set of comprehensive metrics is utilized to assess its performance across traditional benchmarks, providing key insights into its learning dynamics and generalization capabilities. These metrics include the average training and validation losses (Average Training Loss, Average Validation Loss), training and validation accuracies (Training Accuracy, Validation Accuracy), as well as specialized metrics like False Rejection Rate (FRR) and False Acceptance Rate (FAR) for the sole purpose of making our system more precise and Advanced decision maker.

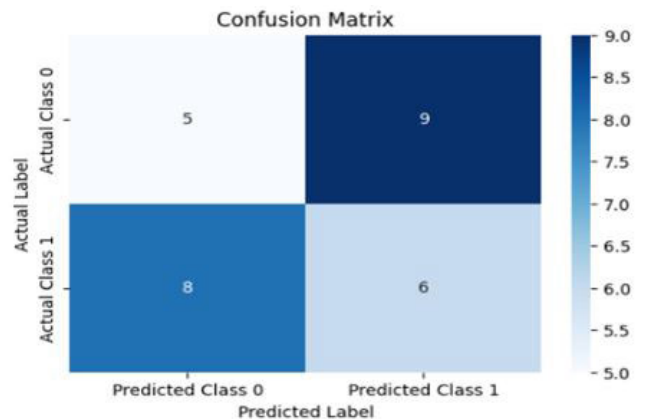


FIGURE 3. Confusion matrix.

Figure 3 points toward the key aspects of the confusion matrix for the signature verification system. The label Actual 0 corresponds to Tensor 0 which represents Binary image for validation. similarly, Actual 1 corresponds to Tensor 1 which represents the Binary input image. Same-wise Predicted class 0 represents a Real signature and Predicted class 1 represents a fake signature, thus helping in identifying real from fake signatures.

The training loss for each epoch is calculated as the average of batch losses during training:

$$\text{AverageTrainingLoss} = \left(\frac{1}{N}\right) \sum_{i=1}^n \cdot \text{batchloss} \quad (1)$$

where N is the number of batches.

The training accuracy is calculated as the percentage of correctly classified samples over the total training samples:

$$\text{TrainingAccuracy} = \left(\frac{\text{TotalTrainingSamples}}{\text{CorrectPredictions}}\right) \times 100 \quad (2)$$

The validation accuracy is determined in the same manner on the separate validation dataset:

$$\text{ValidationAccuracy} = \left(\frac{\text{TotalValidationSamples}}{\text{CorrectPredictions}}\right) \times 100 \quad (3)$$

The learning rate is implemented to adjust the learning process during training based on a step schedule. The learning rate update equation is:

$$\text{NewLearningRate} = \text{OldLearningRate} \times \Gamma \quad (4)$$

where T is the reduction factor specified in the StepLR scheduler. NEXT, we record the False Rejection Rate (FRR) and False Acceptance Rate (FAR), for assessing the model's performance in recognizing authentic signatures and rejecting false ones:

$$\text{FRR} = \left(\frac{\text{Total Genuine Samples}}{\text{False Rejections}}\right) \times 100 \quad (5)$$

The FAR (False Acceptance Rate) is formulated:

$$\text{FAR} = \left(\frac{\text{Total Forged Samples}}{\text{False Acceptances}}\right) \times 100 \quad (6)$$

These equations are crucial for evaluating our model, providing a clear understanding of its learning dynamics, accuracy, and performance metrics. Figure 3 visually represents the trends of these metrics for the training epochs.

Figure 3 presents the confusion matrix, which provides a detailed breakdown of the model's performance. It shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class. "Actual Class 0" and "Actual Class 1" refer to the true class labels, while "Predicted Class 0" and "Predicted Class 1" represent the model's predictions. For example, the matrix indicates that 5 instances of Actual Class 0 were correctly

predicted as Class 0 (True Negatives) and 6 instances of Actual Class 1 were accurately predicted as Class 1 (True Positives). Additionally, it shows 8 instances of Actual Class 0 that were incorrectly predicted as Class 1 (False Positives) and 9 instances of Actual Class 1 that were wrongly predicted as Class 0 (False Negatives). The color intensity in the matrix reflects the frequency of instances, with darker shades signifying higher counts. This matrix is crucial for evaluating the model's performance and calculating metrics such as accuracy, precision, and recall, which support effective decision-making in classification tasks.

F. EXPERIMENTAL SETUP

During the experimental phase, we rigorously studied two hybrid models: ResNet-18 [20] with Vision Transformer (Proposed Method 1) and MobileNetV2 [19] with Vision Transformer (Proposed Method 2) for a classification task. To enhance the performance of our proposed models, we improvised old models and made some adjustments, such as integrating activation functions other than Relu like Tanh and Swish. To enhance integration and improve efficiency, we used the Keras [44] and TensorFlow libraries [45]. As the original images in our dataset were of 531×331 pixels, we enhanced them to a consistent 224×224 pixels using TensorFlow and a transformer model. This conversion ensured a consistent input size for our model and remove and kind of disbalance in our benchmark datasets, which operates with a fixed number of trainable parameters, contributing to the stability of the evaluation process. The experimental coding and output were executed using the Google Collaborator notebook, providing a collaborative and efficient coding environment.

Benchmark Dataset organization involved classes labeled from 1 to 100 in CEDAR [9] and BHSIG-BENGALI [9], with the class range extending to 160 in BHSIG-HINDI [9]. In Utsig-Persian [10], the primary class pertained to forgery, with additional original and sub-classes labeled from 1 to 115. There was also the chance of intra writer variability, to overcome this issue we gained 20 real signatures of authors and similarly 20 fake signatures. Thus preventing any kind of dicriminacy. This duplication of signatures from real authors resulted in our solution for intra-writer variability as all individuals tend to change a small line in signature. This results in a lot of features for Machine Learning models. Our results proved that we have successfully achieved our goals as high accuracies were recorded except Utsig-Persian [10] benchmark dataset, however its accuracy is also not very low.

Our proposed model was developed using the Google Collaborator platform, leveraging Python 3 on a Google compute engine backend for coding. The system's computational resources included a RAM capacity of 12.7GB, and the collaborative notebook utilized the standard Google GPU for accelerated processing during model training. Access to datasets was facilitated through Google Drive, providing a centralized and accessible repository for training the model. This meticulously configured experimental setup aims to

provide a robust foundation for evaluating the performance of our hybrid models in the context of signature verification.

We have trained our model based on Google Collaborator Notebook with Python 3 Google Compute Engine backend. The system RAM used in building this model is 12.7 GB and the disk is 107 GB. The initial learning rate for the model is 0.0005 and weight decay is $1e-5$. The batch size is set to 6-32 owing to the model requirements. The hybrid models are composed of a ResNet-18 [20] pre-trained model with a Vision transformer model (proposed 1) and a MobileNetV2 [19] pre-trained model with a Vision transformer model (Proposed 2).

The proposed model for signature verification employs two distinct architectures ResNet-18 [20] and MobileNetV2 [19] from pre-trained PyTorch [42] model infused with the Vision transformer model to improve performance and accuracy. In Proposed Method 1, ResNet-18 [20] undergoes 100 epochs using Adam optimizer [40] and Cross Entropy Loss function, With a learning rate of 0.005 (we can use small values for learning rate). The results are recorded using Loss, Training accuracy, and validation Accuracy metrics.

In Proposed Method 2; MobileNetV2 [19] pre-trained model from PyTorch [42] is used to train with the Vision transformer model. This hybrid model is trained over 150 epochs using Adam optimizer [40], and Cross Loss Entropy function, With a learning rate of 0.05. The metrics used for measuring results are Loss, Training accuracy, and Validation accuracy.

Notably, changes were necessary for integrating the ResNet-18 [20] and MobileNetV2 [19] model with The vision transformer model. We begin the process of integrating by first understanding the Vision Transformer mode. This model begins with Patch embedding (The image is divided into fixed-size patches and each patch is flattened and projected into lower dimensional space using a linear layer). Then comes Positional Encoding. In positional encoding information is added to the Patch Embeddings to retain spatial information, This is implemented using Positional encoding. then we have Transformer Encoder Layers, which are the main component of the Vision Transformer. This layer has a multi-head self-attention mechanism followed by a feed-forward neural network. To stabilize the training of the model layer normalization and residual connection are used. For the classification task classification token is used as an output corresponder, which is prepended to the sequence of patch embeddings.

The integration of ResNet-18 [20] and MobileNetV2 [19] with the Vision transformer is carried out in the following manner. ResNet and MobileNet are used for Feature extraction from images, The output of these convolutional layers is fed into the Vision transformer. This requires adapting the final layers of both models to generate feature maps for patch embedding. Feature maps of ResNet-18 [20] and MobileNetV2 [19] are split into patches, Each patch is flattened and then processed through a linear layer to produce

patch embedding's. The integration points for creating a hybrid model are as follows, The ViT [41] module is integrated after the feature extraction phase from ResNet-18 [20] or MobileNetV2 [19]. This arrangement allows the ViT [41] to efficiently process and analyze the spatial information extracted by these models.

IV. RESULT

In this section, we provide a detailed analysis of the results from our extensive testing and training of the hybrid models for signature verification. We review essential performance metrics, including training and validation losses, accuracy, False Rejection Rate (FRR), and False Acceptance Rate (FAR). Additionally, we assess how effectively the models learn, their ability to differentiate between signatures, and their success in generalizing to new data. We also examine computational efficiency, training times, and resource usage to offer a complete understanding of the models' performance. This thorough evaluation aims to provide insightful conclusions on the practical effectiveness and feasibility of our hybrid model approach for signature verification.

Achieving 99.96% accuracy on the BHSIG-Hindi [9] dataset is a remarkable feat, demonstrating the model's exceptional proficiency in signature verification. This level of accuracy guarantees dependable user authentication, significantly mitigating the risks of unauthorized access and fraud. In practical applications such as mobile banking and document verification, this performance enhances security, boosts user confidence, and improves operational efficiency. It also shows the model's ability to manage diverse signature styles and conditions, making it well-suited for critical applications.

A. APPROACH

Before exploring the results, it's crucial to explain the approach used in testing and evaluating our hybrid models. We combined ResNet-18 [20] architectures with custom Vision Transformers (ViT [41]) to make proposed method 1 MobileNetV2 [19] architectures with custom Vision Transformers (ViT [41]) in our models to produce Proposed Method 2. These architectures were selected for their effectiveness in feature extraction and representation learning. Our training process integrates these components to fine-tune the model for signature verification. Important factors such as preprocessing steps, hyperparameter tuning, and model configurations play a significant role in determining performance. This section provides a clear overview of our methodology, laying the groundwork for interpreting the results.

1) PROPOSED METHOD 1

These results were achieved by using Proposed Method 1 (comprising of ResNet-18 [20] + Vision Transformer). To further breakdown ResNet-18 [20] was utilized for extracting

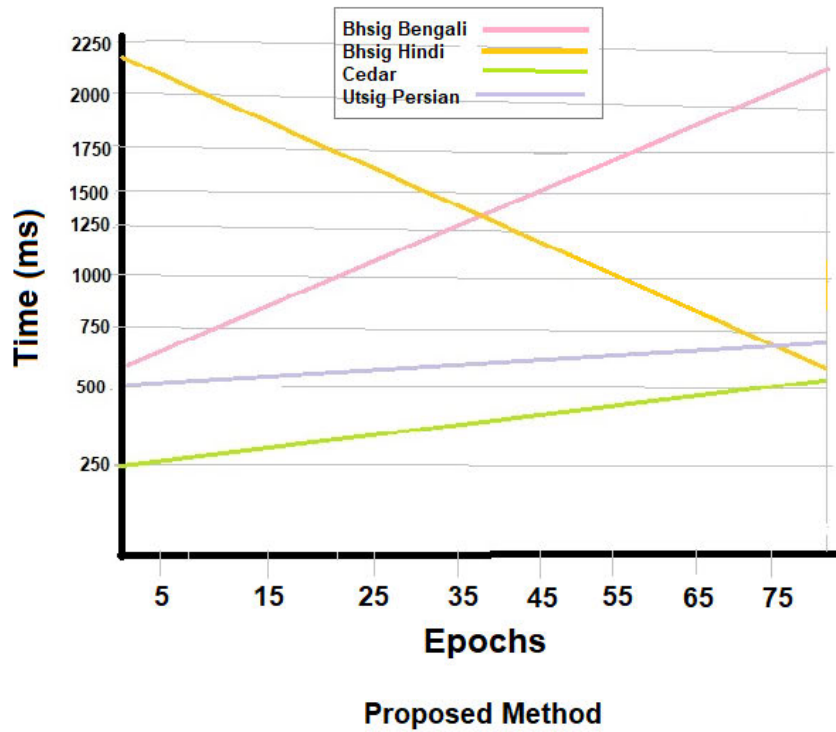


FIGURE 4. Proposed compilation time of the experiment.

global features., and ViT [41] was employed for classification and preprocessing.

The highest accuracy recorded by each experiment was as follows:

- Experiment 1: 92.33%
- Experiment 2: 99.36%
- Experiment 3: 99.9%
- Experiment 4: 74.69%

These accuracies were achieved across all the datasets mentioned in Table 3.

2) PROPOSED METHOD 2

The model used for obtaining these results was the MobileNetV2 [19] + Vision Transformer hybrid model (proposed method 2). MobileNetV2 [19] was utilized for global feature extraction, while Vision Transformer was employed for the validation of outputs from the feature extractor.

The highest accuracy recorded by each experiment was as follows:

- Experiment 1: 81.02%
- Experiment 2: 99.89%
- Experiment 3: 99.96%
- Experiment 4: 74.06%

These accuracies were achieved across all the datasets mentioned in Table 3. The results of the extensive experiments have been recorded in Table 3 and CT(time in table stands for Compilation-Time).

B. FAR, FRR, COMPILATION TIME

The FRR, FAR, COMPILATION TIME, and EXECUTION TIME are also recorded for each experiment. The far records are recorded by this formula, The False Acceptance Rate (FAR) is calculated using the formula:

$$FAR = \left(\frac{\text{Number of False Acceptances}}{\text{Total Number of Impostor Samples}} \right) \times 100 \quad (7)$$

Similarly, the False Rejection Rate (FRR) is calculated using the formula:

$$FRR = \left(\frac{\text{Number of False Rejections}}{\text{Total Number of Impostor Samples}} \right) \times 100 \quad (8)$$

The graph in Figure 5 represents the compilation time, execution time, FRR and FAR combined in one graph.

TABLE 3. Experimental results of proposed method.

| Dataset | Model | Accuracy (%) | CT (seconds) |
|--------------------|-------------------|--------------|--------------|
| BHSig-Bengali [9] | Proposed method 1 | 92.33 | 723.23 |
| | Proposed method 2 | 81.02 | 1600.032 |
| BHSig-Hindi [9] | Proposed method 1 | 99.36 | 2200 |
| | Proposed method 2 | 99.89 | 2300 |
| CEDAR [9] | Proposed method 1 | 99.90 | 300 |
| | Proposed method 2 | 99.96 | 500 |
| Utsig-Persian [10] | Proposed method 1 | 75.69 | 400 |
| | Proposed method 2 | 74.06 | 500 |

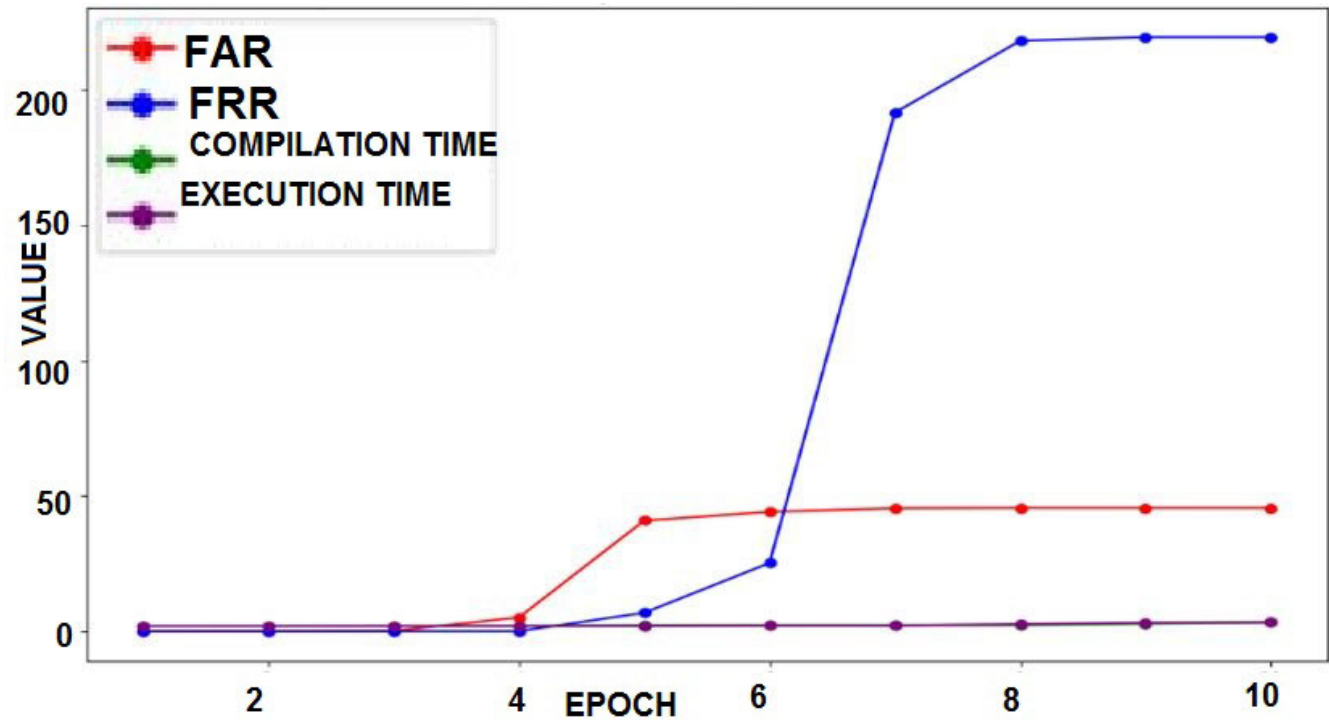


FIGURE 5. FRR, FAR, compilation time, and execution time.

Different experiments have recorded different compilation times on all four given datasets. The accuracies and the compilation time of our trained model are recorded in Table 3.

We examined the impact of relu [29], Swish, and tanh activation functions on the performance of our hybrid vision transformer model across four signature datasets. The results show that Swish significantly improves accuracy over relu [29] and tanh. Swish consistently delivered the highest accuracies across all datasets, highlighting its effectiveness in enhancing model performance. Tanh also outperformed relu [29], though its improvements were less pronounced compared to Swish.

TABLE 4. Performance metrics for different activation functions.

| Activation Function | Bhsig-Bengali [9](%) | Bhsig-Hindi [9](%) | Cedar [9](%) | UTsig-Persian [10](%) |
|---------------------|----------------------|--------------------|--------------|-----------------------|
| relu [29] | 90.45 | 98.75 | 99.50 | 70.30 |
| Swish [46] | 92.33 | 99.89 | 99.96 | 74.09 |
| Tanh [47] | 91.70 | 99.50 | 99.85 | 73.45 |
| Sigmoid [48] | 35 | 45.98 | 47.54 | 23.78 |

The results in Table 4 demonstrate that the Swish and Tanh activation functions are highly effective in feature extraction, leading to the highest accuracy in signature verification. While ReLU also performs well, the Sigmoid function exhibits the lowest accuracy and poorest feature extraction

capabilities. Consequently, Swish and Tanh were selected as the activation functions for our model.

These results recorded using 32 and 16-batch sizes are shown in the table 5.

TABLE 5. Recorded FAR and FRR results.

| FAR /FRR | Model | BHSig-Bengali [9] | BHSig-Hindi [9] | Cedar [9] | UTSig-Persian [10] |
|----------|------------|-------------------|-----------------|-----------|--------------------|
| FAR | Proposed 1 | 0.85 | 0.35 | 1.25 | 0.92 |
| | Proposed 2 | 9.6 | 51.37 | 11.5 | 0.98 |
| FRR | Proposed 1 | 0.90 | 1.35 | 0.125 | 22 |
| | Proposed 2 | 0.97 | 4.15 | 8.63 | 11 |

Figure 5 shows the compilation time of the best experiments. the graph proves our proposed method has experimented with the minimum required to compile results.

The analysis of our data has revealed several noteworthy results. To measure the performance of **Proposed method 1** we use training and validation accuracy and loss are calculated for each hybrid model.

Figure 6(a) shows the results of the ResNet-18 [20] and ViT [41] hybrid model. The training accuracy is 87% and the validation accuracy is 86% for 45 experiments. These results are recorded for the Bhsig-Bengali [9] dataset. Figure 6(b) shows the results of the ResNet-18 [20] and ViT [41] hybrid model on Bhsig-Hindi [9].

Proposed method 2 is also evaluated upon performing different numbers of experiments. Figure 7 shows the

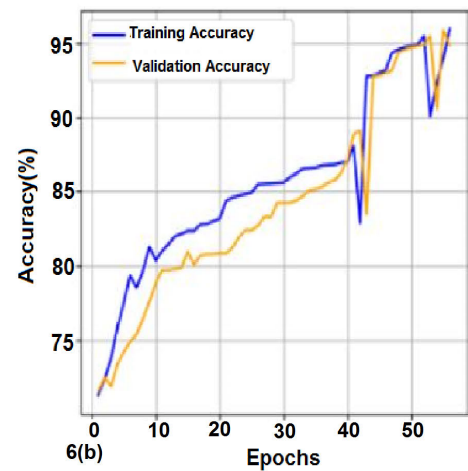
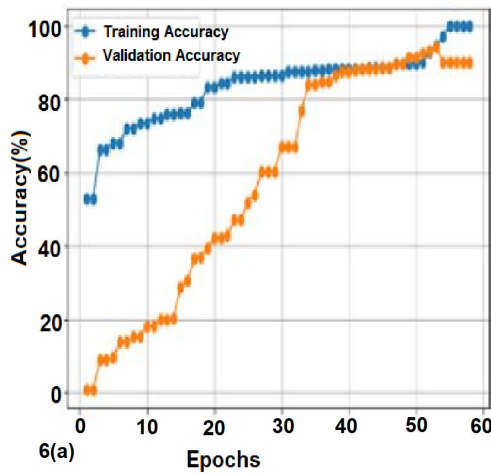


FIGURE 6. Training accuracy vs validation accuracy BHSIG-Bengali 6(a), BHSIG-Hindi 6(b).

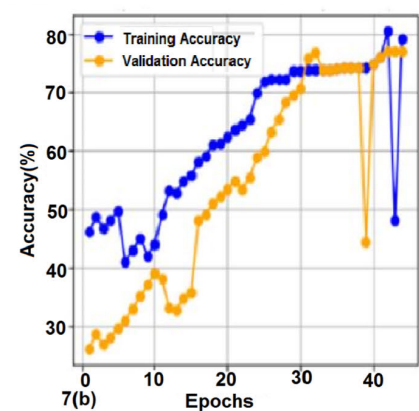
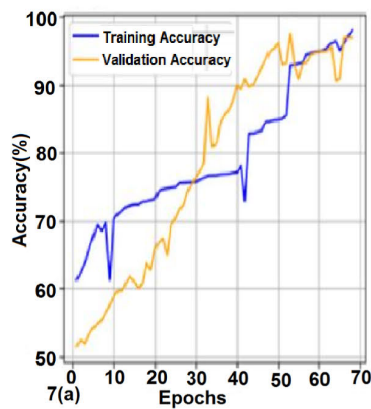


FIGURE 7. Training accuracy vs validation accuracy BHSIG- Bengali 7(a), Utsig-Persian 7(b).

results of different experiments performed using hybrid MobileNetV2 [19] and transformer machine learning models.

In Figure 7(a) the results are for Bhsig-Bengali [9]. In Figure 7(b) the results are for Utsig-Persian [10].

In Figure 8(a) the results are for Bhsig-Hindi [9]. In Figure 8(b) the results are for Cedar [9]. The training accuracy is 97% and the validation accuracy is 96% for 55 experiments.

Figure 9(a) shows the results of the ResNet-18 [20] and ViT [41] hybrid model on Utsig-Persian [10]. The training accuracy is 88% and the validation accuracy is 87% for 55 experiments.

Figure 9(b) shows the results of the ResNet-18 [20] and ViT [41] hybrid model on Cedar [9]. The training accuracy is 89.09% and the validation accuracy is 89% for 70 experiments.

With the elaborative analysis done in the research, we can prove that the proposed models have outperformed state-of-the-art models. Proposed models have achieved higher accuracy lower FRR(False Rejection Rate) and low FAR (false

Acceptance Rate) values than traditional models. Table 7 records the results of experiments on the Utsig-Persian.

TABLE 6. Results of compilation time (seconds) of proposed methods.

| No | Model | BHSig-Bengali [9] | BHSig-Hindi [9] | Cedar [9] | UTSig-Persian [10] |
|----|-------------------|-------------------|-----------------|-----------|--------------------|
| 1 | Proposed Method 1 | 723.232 | 2200 | 300 | 400 |
| 2 | Proposed Method 2 | 1800 | 1200 | 600 | 800 |

TABLE 7. Results on Utsig-Persian dataset.

| No | Model | UTSig-Persian [10] (%) | FRR | FAR | CT (seconds) |
|----|-------------------|------------------------|-------|-----|--------------|
| 1 | Proposed Method 1 | 75.69 | 0.925 | 22 | 400 |
| 2 | Proposed Method 2 | 74.06 | 0.987 | 11 | 800 |

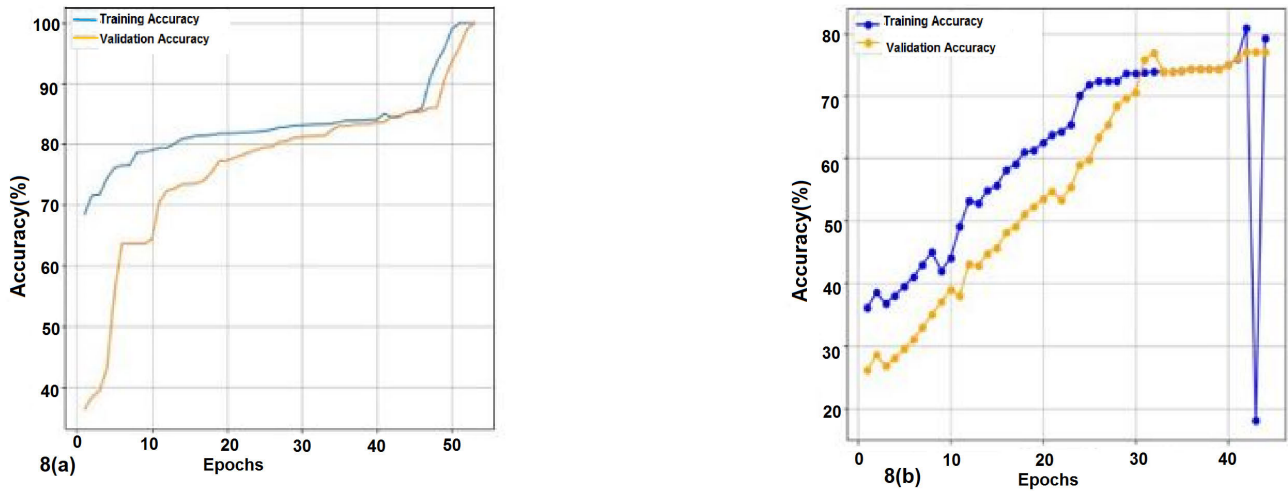


FIGURE 8. Training accuracy vs validation accuracy Bhsig-Hindi 8(a), Cedar 8(b).

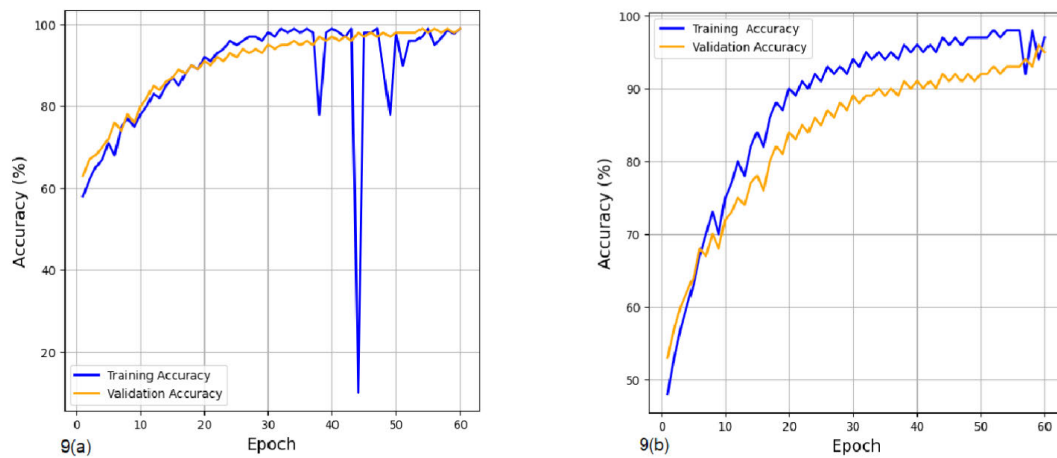


FIGURE 9. Training accuracy vs validation accuracy on Utsig-Persian 9(a), Cedar 9(b).

During the validation phase, the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are computed, offering additional perspectives on the model's efficacy in terms of false positives and signature verification.

C. DISCUSSION

The discussion section serves as the intellectual nexus of our exploration, providing a platform to interpret and contextualize the findings unearthed during our research journey. In this segment, we delve into the intricacies of the results, elucidating their significance, implications, and potential 97% and the validation accuracy is 96% for 55 experiments. Figure 9(a) shows the results of the ResNet-18 [20] and ViT [41] applications. It is within these paragraphs that we unravel the nuances of our methodologies, critically assess the limitations encountered, and articulate the broader implications of our research within the realms of artificial intelligence, signature verification,

and hybrid model integration. As we navigate through the outcomes, we aim to offer a holistic understanding of our contributions to the existing body of knowledge, paving the way for further inquiry, refinement, and practical implementations in the dynamic landscape of signature verification systems.

The experimental results indicate the model's robustness and ability to generalize across diverse datasets. The high accuracy rates on the Bhsig-Bengali, Bhsig-Hindi, and Cedar datasets suggest that the model can handle various signature styles and variations effectively. The slightly lower accuracy on the Utsig-Persian dataset suggests features extracted from this benchmark dataset were more difficult as compare to other datasets. The results prove that the model can maintain high performance even with different signature of different writing styles having different characteristics, which is crucial for real-world applications where the system needs to verify signatures from a diverse user base.

The proposed methodology introduces a novel hybrid model for signature verification, representing a significant departure from traditional approaches. Our approach merges Convolutional Neural Networks (CNNs) with Vision Transformers (ViT [41]s) to leverage the strengths of both technologies, enhancing performance and adaptability. Central to this strategy is the combination of ResNet-18 [20], MobileNetV2 [19], and a custom vision transformer model.

This hybrid method offers significant improvements for secure user authentication systems. The high accuracy result proves to be a reliable solution for signature verification, which is crucial for secure access and online in mobile applications, thereby increasing user security and self reliance. For document identification, the model's high accuracy helps in identifying fake signatures and ensuring document realness. Its high efficiency and best scalability make it applicable for use in various real-world scenarios.

Our hybrid model effectively combines the feature extraction capabilities of ResNet-18 [20] and MobileNetV2 [19] with the processing power of vision transformers. This integration allows the model to automatically identify and refine complex patterns in signature images, capturing subtle details essential for accurate verification.

Compared to traditional methods that may struggle with diverse datasets and signature styles, our approach benefits from the strengths of both CNNs and transformers, leading to better performance in various scenarios. Modern techniques such as transfer learning and data augmentation are also employed, allowing the model to continuously improve with minimal manual adjustment, making it resilient and adaptable to evolving challenges in signature verification.

The methodology includes a thorough framework for training and evaluating the hybrid Vision Transformer model on signature image datasets. It begins with detailed data preparation, involving resizing, cropping, flipping, and normalization to standardize the input. The dataset is then divided into training and validation subsets, ensuring a well-rounded evaluation of the model's effectiveness. The hybrid model architecture combines the feature extraction capabilities of pre-trained CNN backbones with a custom vision transformer, creating a versatile model that blends established features with the adaptability of a transformer.

We have computed and compared precision, recall, and F1-score metrics across the Bhsig-Bengali, Bhsig-Hindi, Cedar, and Utsig-Persian benchmark datasets to understand the model's performance better. High precision and recall across all datasets would indicate robust performance, and the F1-score would help balance any trade-offs between precision and recall. Table 8 records the experimental results for all activation functions.

Table 8 points towards the performance of different activation functions on different benchmark datasets. The table demonstrates that Tanh and Swish activation functions collectively perform best on Bhsig-Bengali [9] and Cedar

TABLE 8. Comparison of activation functions across different datasets.

| Activation Function | Evaluation Metric | Bhsig-Hindi [9] (%) | Bhsig-Bengali [9] (%) | Cedar [9] (%) | Utsig-Persian [10] (%) |
|------------------------|-------------------|---------------------|-----------------------|---------------|------------------------|
| relu [29] | Precision | 0.5069 | 0.4105 | 0.3986 | 0.3103 |
| | Recall | 0.5571 | 0.6679 | 0.4987 | 0.5571 |
| | F1 Score | 0.5291 | 0.4034 | 0.2998 | 0.3986 |
| Tanh [47] | Precision | 0.3986 | 0.2230 | 0.2590 | 0.2032 |
| | Recall | 0.4412 | 0.3412 | 0.2981 | 0.2978 |
| | F1 Score | 0.3987 | 0.3542 | 0.2450 | 0.3001 |
| Swish [46] | Precision | 0.3103 | 0.5076 | 0.6512 | 0.4108 |
| | Recall | 0.5571 | 0.7676 | 0.7342 | 0.5821 |
| | F1 Score | 0.3986 | 0.5059 | 0.4989 | 0.6781 |
| Tanh [47] + Swish [46] | Precision | 0.3103 | 0.6089 | 0.6091 | 0.7812 |
| | Recall | 0.5571 | 0.6542 | 0.4008 | 0.5643 |
| | F1 Score | 0.3986 | 0.8912 | 0.7654 | 0.5589 |

[9] datasets whereas relu activation function performs moderately on benchmark datasets. Swish activation function alone performs well and gives us the best accuracies.

Hyper-parameters such as learning rate, weight decay, and epochs are meticulously configured to optimize the training regimen. We use the Adam optimizer [40] and cross-entropy loss function to increase the learning and classification process. During training, the model is iteratively improved over multiple epochs, with losses and accuracies regularly computed and evaluated against a validation set. To measure the model's performance comprehensively, we measure metrics such as the False Acceptance Rate (FAR), False Rejection Rate (FRR), and Overall Error Rate (ERR), which help us understand the model's decision-making precision and effectiveness.

To present our findings clearly, we utilize Matplotlib to generate visualizations that illustrate the model's learning progress, because we used python programming language to train our model and plot graphs for our results. These graphs display losses, accuracies, and other relevant metrics throughout the training epochs, offering a detailed view of the model's performance. In conclusion, our approach introduces an advanced method for signature verification, combining the strengths of hybrid modeling to achieve high accuracy and adaptability in real-world applications.

The proposed model tackles major cyber-security issues such as forgery and bio-metric spoofing using several advanced techniques:

Enhanced Feature Extraction: Our hybrid architecture captures both detailed local features and broad global patterns, making it challenging for attackers to duplicate authentic signatures accurately.

Attention Mechanisms: Vision Transformers focus on critical areas of the signature, which enhances the model's ability to detect subtle inconsistencies.

Real-Time Verification: The model's efficient processing capabilities enable the swift identification and rejection of fake or altered signatures, ensuring strong security in real-time scenarios.

Together, these strategies strengthen the security and reliability of signature verification systems, reducing the likelihood of unauthorized access and fraudulent activities.

D. COMPARISON

Table 9 points towards the comparison between our proposed model and conventional models accuracies on Bhsig-Bengali [9], Bhsig-Hindi [9], and Cedar [9] datasets.

TABLE 9. Comparison of proposed accuracy with conventional models accuracies.

| No | Model | Bhsig-Bengali [9](%) | Bhsig-Hindi [9](%) | Cedar [9](%) |
|----|-----------------------|----------------------|--------------------|--------------|
| 1 | 2-Chanel-2-Logit [21] | 88.08 | 86.66 | 100 |
| 2 | MobileNetV2 [49] | 85.63 | 75.00 | 96.00 |
| 3 | 2C2S [33] | 93.25 | 90.68 | 100 |
| 4 | SigNet [50] | 86.11 | 84.64 | 100 |
| 5 | SURDS [51] | 90.36 | 0 | 89.50 |
| 6 | MSN [52] | 91.56 | 88.88 | 98.40 |
| 7 | Proposed Method 1 | 92.33 | 99.9 | 92.50 |
| 8 | Proposed method 2 | 81.02 | 99.89 | 99.96 |

Table 10 elaborates on the comparison of the False Acceptance Rate of conventional models and proposed models. We can prove from Table 10 that the proposed models have outperformed conventional models.

TABLE 10. Comparison of proposed FAR with conventional models FAR.

| No | Model | Bhsig-Bengali [9](%) | Bhsig-Hindi [9](%) | Cedar [9](%) |
|----|-----------------------|----------------------|--------------------|--------------|
| 1 | 2-Chanel-2-Logit [21] | 10.44 | 0 | 0 |
| 2 | MobileNetV2 [49] | 0 | 0 | 0 |
| 3 | 2C2S [33] | 5.37 | 8.66 | 0 |
| 4 | SigNet [50] | 13.89 | 15.36 | 0 |
| 5 | SURDS [51] | 19.89 | 12.01 | 0 |
| 6 | MSN [52] | 10.42 | 17.06 | 3.18 |
| 7 | Proposed Method 1 | 0.90 | 1.35 | 1.22 |
| 8 | Proposed method 2 | 0.97 | 4.15 | 8.63 |

Table 11 elaborates on the comparison of the False Rejection Rate of conventional models and proposed models. We can prove from Table 11 that the proposed models have outperformed conventional models.

The findings of this study significantly impact the development of secure user authentication systems by providing a highly accurate and robust solution for signature verification. In mobile applications, this ensures secure access and transactions, enhancing user trust and satisfaction. For document verification, the model's high accuracy reduces the risk of forgery and fraud, ensuring the authenticity of signed documents. The efficiency and scalability of the model make it suitable for widespread deployment, offering reliable security solutions across various real-world applications.

TABLE 11. Comparison of proposed FRR with conventional models FRR.

| No | Model | Bhsig-Bengali [9](%) | Bhsig-Hindi [9](%) | Cedar [9](%) |
|----|-----------------------|----------------------|--------------------|--------------|
| 1 | 2-Chanel-2-Logit [21] | 9.37 | 0 | 0 |
| 2 | MobileNetV2 [49] | 0 | 0 | 0 |
| 3 | 2C2S [33] | 8.11 | 8.66 | 0 |
| 4 | SigNet [50] | 13.89 | 9.98 | 0 |
| 5 | SURDS [51] | 5.42 | 8.98 | 0 |
| 6 | MSN [52] | 6.44 | 5.16 | 0 |
| 7 | Proposed Method 1 | 0.85 | 0.35 | 1.25 |
| 8 | Proposed method 2 | 9.60 | 51.73 | 11.5 |

V. CONCLUSION AND FUTURE WORK

In this study, we conducted an extensive examination of hybrid models integrating ResNet-18 [20] or MobileNetV2 [19] with the Vision Transformer (ViT [41]) architecture across diverse datasets, including BHSIG-BENGALI, CEDAR, BHSIG-HINDI, and UTSIG-PERSIAN. Analysis across various metrics, including training and validation accuracy, loss, compilation time, execution time, and error rates (FRR and FAR), consistently demonstrated promising results. In particular, the hybrid combinations exhibited average training accuracies ranging from 68% to 80% on the BHSIG-BENGALI dataset, while the Res-Net and ViT [41] hybrid models stood out with an average accuracy of 92.25% on the CEDAR benchmark dataset. Notably, CEDAR showcased an outstanding 100% accuracy, and UTSIG-PERSIAN displayed accuracies ranging from 78% to 90% across diverse hybrid models. These findings underscore the adaptability and effectiveness of hybrid models across datasets and enrich our understanding of their performance nuances in varied scenarios, contributing to the dynamic field of artificial neural networks.

As we wrap up this research, it encourages us to consider future directions. Identifying the key causes and challenges in these areas will be essential for creating effective solutions. Looking forward, the integration of hybrid techniques, such as Shuffle-Net + ViT [41] or ViT + Nasnet, emerges as potential avenues for further refining signature validation and verification standards. As Shuffle-Net provides more efficiency in mobile and low compute environments and Nas-Net has auto-learn architecture which can adapt to dynamic benchmark datasets. This avenue of research opens exciting possibilities to challenge existing methodologies and introduce innovative approaches.

Furthermore, the realm of dataset creation and secure storage of authentic signatures invites creative exploration. As technology evolves, the future holds potential for refining methodologies and unearthing novel approaches to fortify the resilience and efficiency of signature validation systems. In conclusion, this study provides a snapshot of the current landscape, emphasizing the imperative for ongoing exploration and innovation in the dynamic fields of hybrid models and signature validation. The journey continues, promising

continuous strides in understanding, improving, and applying artificial neural networks in real-world scenarios.

AUTHOR CONTRIBUTIONS

Conceptualization: Muhammad Ishfaq and Ayesha Saadia; methodology: Muhammad Ishfaq and Faeiz M. Alserhani; software: Muhammad Ishfaq; validation: Muhammad Ishfaq, Ayesha Saadia, and Ammara Gul; formal analysis: Muhammad Ishfaq; investigation: Faeiz M. Alserhani; resources: Muhammad Ishfaq; data curation: Muhammad Ishfaq; writing—original draft preparation: Muhammad Ishfaq and Ammara Gul; writing—review and editing: Muhammad Ishfaq; visualization: Muhammad Ishfaq; supervision: Ayesha Saadia; project administration: Muhammad Ishfaq and Ayesha Saadia. All authors have read and agreed to the published version of the manuscript.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable

DATA AVAILABILITY STATEMENT

Not applicable

ACKNOWLEDGEMENTS

Not applicable

CONFLICT OF INTEREST

The authors declared no conflict of interest.

REFERENCES

- [1] N. Cavus and N. Sancar, "The importance of digital signature in sustainable businesses: A scale development study," *Sustainability*, vol. 15, no. 6, p. 5008, Mar. 2023.
- [2] M. Mambo, K. Usuda, and E. Okamoto, "Proxy signatures for delegating signing operation," in *Proc. 3rd ACM Conf. Comput. Commun. Secur. (CCS)*, 1996, pp. 48–57.
- [3] K. Zhang, "Threshold proxy signature schemes," in *Proc. Int. Workshop Inf. Secur. Berlin, Germany: Springer*, 1997, pp. 1–10.
- [4] C. Mwangi, "An algorithm for identity theft mitigation: Keypoint signature verification," Ph.D. dissertation, School Comput. Inform., Univ. Nairobi, Kenya, 2016.
- [5] Y. M. Al-Omari, S. N. H. S. Abdullah, and K. Omar, "State-of-the-art in offline signature verification system," in *Proc. Int. Conf. Pattern Anal. Intell. Robot.*, vol. 1, Jun. 2011, pp. 59–64.
- [6] S. Mushtaq and A. H. Mir, "Signature verification: A study," in *Proc. 4th Int. Conf. Comput. Commun. Technol. (ICCCCT)*, Sep. 2013, pp. 258–263.
- [7] E. Argones Rua and J. L. Alba Castro, "Online signature verification based on generative models," *IEEE Trans. Syst. Man, Cybern., B, Cybern.*, vol. 42, no. 4, pp. 1231–1242, Aug. 2012.
- [8] K. Franke, "Analysis of authentic signatures and forgeries," in *Proc. Int. Workshop Comput. Forensics*, Hague, The Netherlands. Berlin, Germany: Springer, Aug. 2009, pp. 150–164.
- [9] D. Rai. *Handwritten Signatures Dataset*. Accessed: Jul. 17, 2023. [Online]. Available: <https://www.kaggle.com/datasets/divyanshrai/handwritten-signatures>
- [10] OpenML. *Openml Dataset: ID 23411*. Accessed: Sep. 24, 2024. [Online]. Available: <https://www.openml.org/search?type=data&sort=runs&id=23411&status=active>
- [11] M. Adamski and K. Saeed, "Online signature classification and its verification system," in *Proc. 7th Comput. Inf. Syst. Ind. Manage. Appl.*, Jun. 2008, pp. 189–194.
- [12] A. Rehman, S. Naz, M. I. Razzak, and I. A. Hameed, "Automatic visual features for writer identification: A deep learning approach," *IEEE Access*, vol. 7, pp. 17149–17157, 2019.
- [13] S. Carter and M. Nielsen, "Using artificial intelligence to augment human intelligence," *Distill*, vol. 2, no. 12, p. 9, Dec. 2017.
- [14] B. Zohuri and F. M. Rahmani, "Artificial intelligence versus human intelligence: A new technological race," *Acta Sci. Pharmaceutical Sci.*, vol. 4, no. 5, pp. 50–58, Apr. 2020.
- [15] N. Swarnkar, A. Rawal, and G. Patel, "A paradigm shift for computational excellence from traditional machine learning to modern deep learning-based image steganalysis," in *Data Science and Innovations for Intelligent Systems*. Boca Raton, FL, USA: CRC Press, 2021, pp. 209–240.
- [16] N. Kühl, M. Goutier, R. Hirt, and G. Satzger, "Machine learning in artificial intelligence: Towards a common understanding," 2020, *arXiv:2004.04686*.
- [17] G. R. Yang and X.-J. Wang, "Artificial neural networks for neuroscientists: A primer," *Neuron*, vol. 109, no. 4, p. 739, Feb. 2021.
- [18] F. Xiao, Y. Honma, and T. Kono, "A simple algebraic interface capturing scheme using hyperbolic tangent function," *Int. J. Numer. Methods Fluids*, vol. 48, no. 9, pp. 1023–1040, Jul. 2005.
- [19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [21] J.-X. Ren, Y.-J. Xiong, H. Zhan, and B. Huang, "2C2S: A two-channel and two-stream transformer based framework for offline signature verification," *Eng. Appl. Artif. Intell.*, vol. 118, Feb. 2023, Art. no. 105639.
- [22] M. H. Siddiqi, K. Asghar, U. Draz, A. Ali, M. Alruwaili, Y. Alhwaiti, S. Alanazi, and M. M. Kamruzzaman, "Image splicing-based forgery detection using discrete wavelet transform and edge weighted local binary patterns," *Secur. Commun. Netw.*, vol. 2021, pp. 1–10, Sep. 2021.
- [23] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1257–1260.
- [24] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Pattern Recognit.*, vol. 45, no. 12, pp. 4292–4299, Dec. 2012.
- [25] M. T. Rahman, M. A. Al-Amin, J. B. Bakkre, A. R. Chowdhury, and M. A.-A. Bhuiyan, "A novel approach of image morphing based on pixel transformation," in *Proc. 10th Int. Conf. Comput. Inf. Technol.*, Dec. 2007, pp. 1–5.
- [26] S. Naz and G. S. Kashyap, "Enhancing the predictive capability of a mathematical model for pseudomonas aeruginosa through artificial neural networks," *Int. J. Inf. Technol.*, vol. 16, no. 4, pp. 2025–2034, Apr. 2024.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [28] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, and A. Acosta, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8099502>
- [29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [30] D. Hussain, T. Hussain, A. A. Khan, S. A. A. Naqvi, and A. Jamil, "A deep learning approach for hydrological time-series prediction: A case study of Gilgit river basin," *Earth Sci. Informat.*, vol. 13, no. 3, pp. 915–927, Sep. 2020.
- [31] T. Hussain and H. Shouno, "Explainable deep learning approach for multi-class brain magnetic resonance imaging tumor classification and localization using gradient-weighted class activation mapping," *Information*, vol. 14, no. 12, p. 642, Nov. 2023.
- [32] M. M. Yapıcı, A. Tekerek, and N. Topaloğlu, "Deep learning-based data augmentation method and signature verification system for offline handwritten signature," *Pattern Anal. Appl.*, vol. 24, no. 1, pp. 165–179, Feb. 2021.
- [33] T. Longjam, D. R. Kisku, and P. Gupta, "Writer independent handwritten signature verification on multi-scripted signatures using hybrid CNN-BiLSTM: A novel approach," *Expert Syst. Appl.*, vol. 214, Mar. 2023, Art. no. 119111.
- [34] A. Jain, S. K. Singh, and K. P. Singh, "Handwritten signature verification using shallow convolutional neural network," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 19993–20018, Jul. 2020.

- [35] H. Li, P. Wei, Z. Ma, C. Li, and N. Zheng, "Offline signature verification with transformers," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2022, pp. 1–6.
- [36] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, "MedViT: A robust vision transformer for generalized medical image classification," *Comput. Biol. Med.*, vol. 157, May 2023, Art. no. 106791.
- [37] J. McHatton and K. Ghazinour, "Mitigating social media privacy concerns—A comprehensive study," in *Proc. 9th ACM Int. Workshop Secur. Privacy Anal.*, vol. 33, Apr. 2023, pp. 27–32.
- [38] C. Li, F. Lin, Z. Wang, G. Yu, L. Yuan, and H. Wang, "DeepHSV: User-independent offline signature verification using two-channel CNN," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 166–171.
- [39] F. M. Alsuhat and F. S. Mohamad, "A hybrid method of feature extraction for signatures verification using CNN and HOG a multi-classification approach," *IEEE Access*, vol. 11, pp. 21873–21882, 2023.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [41] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [42] A. Paszke, "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*.
- [43] R. H. Wiggins, H. C. Davidson, H. R. Harnsberger, J. R. Lauman, and P. A. Goede, "Image file formats: Past, present, and future," *RadioGraphics*, vol. 21, no. 3, pp. 789–798, May 2001.
- [44] F. Chollet. (2015). *Keras*. [Online]. Available: <https://keras.io>
- [45] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–283.
- [46] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.
- [47] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Jan. 1998.
- [48] K. S. Sharvari, "A comparative study of transfer learning models for offline signature verification and forgery detection," *J. Univ. Shanghai Sci. Technol.*, vol. 23, no. 7, pp. 1129–1139, Jul. 2021.
- [49] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," *Pattern Recognit.*, vol. 70, pp. 163–176, Oct. 2017.
- [50] S. Chattopadhyay, S. Manna, S. Bhattacharya, and U. Pal, "SURDS: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 1600–1606.
- [51] Y.-J. Xiong and S.-Y. Cheng, "Attention based multiple Siamese network for offline signature verification," in *Proc. 16th Int. Conf. Document Anal. Recognit.*, vol. 16, Lausanne, Switzerland. Cham, Switzerland: Springer, Sep. 2021, pp. 337–349.



MUHAMMAD ISHFAQ is currently pursuing the master's degree in cyber security with Air University, Islamabad, Pakistan. He is a Cyber-Security Enthusiast. He is passionate about cyber-security, particularly penetration testing, and aims to deepen his understanding of various security concepts and methodologies. He actively engages in hands-on learning experiences and practical cyber-security projects to enhance his skills and expertise.

AYESHA SAADIA received the Ph.D. degree from National University of Sciences and Technology (NUST), Pakistan. She is currently an Assistant Professor with the Faculty of Computing and Artificial Intelligence, Air University, Islamabad, Pakistan. Her research interests include medical image processing, image restoration and enhancement, image-to-image translation, and machine learning.



FAEIZ M. ALSERHANI received the bachelor's degree in computer engineering from King Saud University, Riyadh, Saudi Arabia, the M.S. degree in computer and information networks from the University of Essex, U.K., and the Ph.D. degree in network and information security from the University of Bradford, U.K. He is currently with the Department of Computer Engineering and Networks, Jouf University, Saudi Arabia. His research interests include network security, cyber-security, intrusion detection systems, and the application of AI in cyber-security.



AMMARA GUL received the Ph.D. degree in information security from the Royal Holloway, University of London, U.K. She is currently a Lecturer in cyber security with the Faculty of Computing, Engineering and Built Environment, Birmingham City University, Birmingham, U.K. Her research interests include power grid security, security auditing and evaluations, cyber-security, intrusion detection systems, and application of AI in cyber-security.

...