Ain Shams Engineering Journal 17 (2026) 104002

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Full length article

CausaOne-sign: Causal explainable one-shot signature verification with lightweight cross-modality fusion

Sara Tehsin [a], Inzamam Mashood Nasir [b,*], Ali Hassan [c], Farhan Riaz [d]

[a] Faculty of Informatics, Kaunas University of Technology, Kaunas, 51368i, Lithuania
[b] Human-Environment-Technology (HET) Systems Centre, Mykolas Romeris University, Vilnius, 08303, Lithuania
[c] Department of Computer and Software Engineering, National University of Sciences and Technology, Islamabad, 44080, Pakistan
[d] School of Computer Science, University of Lincoln, Lincoln, LN6 7DQ, UK

HIGHLIGHTS

- One-shot offline signature verification using stroke-aware graph modeling.
- Causal attribution explains discriminative signature regions.
- Graph transformer improves generalization to unseen writers.
- Meta-learning enables rapid adaptation with minimal reference samples.
- Lightweight optimization supports real-time edge deployment.

ARTICLE INFO

Keywords:
Offline signature verification
One-shot learning
Causal explainability
Graph-based representation
Meta-learning
Lightweight deep learning

ABSTRACT

Background: Offline handwritten signature verification remains a difficult biometrics problem due to large intra-writer variability; skilled forgers; the limited number of reference samples available; and the black-box nature of many current deep learning based decision-making methodologies. *Objective:* To develop an interpretable, efficient one-shot learning framework that can perform offline signature verification for individuals who have never been seen before using as few reference signatures as possible. *Materials and Methods:* The proposed CausaOne-Sign model uses stroke aware graph encoding, transformer based reasoning, and prototypical embeddings, along with a causal attribution model to provide an explanation of how signature verification works. Experiments have been conducted using CEDAR, SigComp2011 UTSig, and BHSig260 datasets. *Results:* CausaOne-Sign achieved up to 97.4% accuracy and 99.1% area under the curve (AUC), with low ERR (1.8%), outperforming or matching state-of-the-art methods. *Conclusion:* CausaOne-Sign offers a robust, interpretable, and resource-efficient solution for OSV, suitable for forensic and mobile applications.

1. Introduction

Offline signature verification (OSV) is vital in biometric authentication, especially in legal, financial, and forensic domains where verifying a user's identity from static signature photographs is imperative [1,2]. Recent studies indicate that deep learning methodologies, particularly Siamese architectures, have emerged as the predominant paradigm due to their capacity to directly predict pairwise similarity between authentic and counterfeit samples [3]. Conventional deep models sometimes exhibit a deficiency in resistance to domain heterogeneity and maintain opacity in their decision-making processes, thus raising issues over their reliability and trustworthiness [4].

The recent developments in explainable artificial intelligence (XAI) and causal inference support improved visibility and accountability for critical applications [5]. For example, enhancements made through focal-loss-enhanced spatial-transformer Siamese architectures achieved significant improvements in terms of accuracy (¿ 95%) across multiple languages due to a greater focus on the expressive visually-evoked size and stroke patterns while dealing with class imbalances [2]. The

* Corresponding author.
Email address: inzamam.nasir@mruni.eu (I.M. Nasir).

https://doi.org/10.1016/j.asej.2026.104002
Received 9 July 2025; Received in revised form 13 December 2025; Accepted 5 January 2026
Available online 2 February 2026

Transformer models created by the 2C2S architecture employ two independent streams of attention, with results being effective and consistent with AUC results of approximately 93 to 95% mass [6]. Detailed semantic networks (e.g., DetailSemNet, ECCV 2024) were designed based on the idea of using feature-based structural alignment models to develop a model based on corresponding phrases and achieve very high levels of accuracy as well as AUC levels well above 95% [4].

Although Siamese and transformer architectures have made great strides, recently developed adaptive streaming verification systems have revealed great promise for incremental and online learning through signature learning enhancement, thereby allowing for real-time updates on performance against the CEDAR signature dataset [1]. On the other hand, the use of transfer learning frameworks, which are based on pre-trained convolutional neural network (CNN) backbones such as MobileNetV2, continues to show remarkable success, with records indicating that they have achieved accuracies approaching 97.7% on mixed signature datasets, suggesting the efficacy of feature reusability and feature selection [7]. Moreover, lightweight models that include deep networks with genetic algorithm optimization have demonstrated superior performance on Indic-script datasets, achieving over 97% accuracy and an equal error rate (EER) of approximately 2.35% in BHSig260 [2]. Significantly, non-deep alternative methods utilizing UBM-based explainable verifiers provide competitive performance (92%–95%) with transparent decision-making processes designed for forensic applications [2].

Notwithstanding these advancements, current OSV systems generally fail to incorporate three essential properties concurrently: one-shot generalization to novel writers, causal interpretability, and edge-ready deployment. The CausaOne-Sign methodology this paper proposes addresses this gap by integrating stroke-aware graph representation, transformer-based relational learning, prototype one-shot embeddings, causal attribution, individual causal attribution (ICA), meta-learning, model-agnostic meta-learning (MAML), and optimization through knowledge distillation. Extensive benchmark findings indicate that CausaOne-Sign attains superior performance on the Center of Excellence for Document Analysis and Recognition (CEDAR) (97.4% accuracy, 99.1% area under the curve (AUC), 1.8% ERR, SigComp2011 (96.1% accuracy, 98.4% AUC), UTSig (93.8% accuracy, 97.0% AUC), and BHSig260 (94.6% accuracy, 97.7% AUC). Moreover, it provides clear pathways for causal argumentation and effective reasoned conclusions that are well suited to mobile or resource-limited deployment environments.

The subsequent sections of the paper are organized as follows. Section 2 discusses relevant OSV frameworks as well as the growing attention on XAI techniques; Section 3 outlines the CausaOne-Sign framework; Section 4 discusses experimental results along with complete ablation studies, while Section 5 summarizes by discussing possible directions and avenues for future research.

2. Related work

OSV offline has improved significantly recently due to the introduction of new deep learning methods based on transformer networks and Siamese networks. In addition, architectures based on self-supervised learning and workers/metric learning have been developed. One of these systems is SURDS which implements dual-stream triplet metrics with self-supervised attention to extract representations from both streams to enhance both the generalization of the model and the performance of the models for writers [8,9]. In addition, there is a new disentangled variational autoencoder (VAE) model developed in early 2024 that utilizes compact latent embeddings to improve the separation between real and counterfeit signatures through greater accuracy on the MCYT and GPDS datasets as well as through generalization across all datasets [10]. Goh et al. [11] have conducted work on the influence of various non-linear activation functions within Convolutional Neural Networks (CNN's), performing analyses on the impact of using different activation functions when performing image classification in the presence of Poisson noise, ultimately concluding that activation type can significantly impact both robustness to noise and the ability to distinguish between images within the Poisson noise environment.

The use of interpretable and explainable methods has gained acceptance quickly. For example, Diaz et al. suggested a method to enable traceability in a forensic domain by using a universal background model (UBM) method to determine the limits of accepted decisions based on comparing query signatures to an established set of reference signatures [12,13]. Another study conducted by de Moura et al. investigated the applicability of the Signature-Hashing Value (HSV) in the field of real-time classification, showing the ability to adapt through model training on data at the same time as deploying with improved results on batch-trained systems on the UTSig and CEDAR datasets [1]. There have been similar advancements in the learning of features through new methods, such as using multi-scale convolutional neural networks (CNNs), DenseNets, and capsulation-based methods. Liu et al. proposed the mutual signature DenseNet (MSDN), which is a system that combines multiple resolution features being trained together to provide superior results on datasets containing Persian script as well as English script [14]. In addition, some systems have demonstrated improvements in semantic abstraction and performance through the use of hybrid systems based on capsulated layers in comparison to simple standard models when analysing signatures with complex underlying structures [15,16].

Structural-aware model systems, such as DetailSemNet, emphasise the use of fine-resolution spatial information, using a combination of semantic parsing and attention-based refinement to achieve validation accuracy of close to 95%–97% area under curve (AUC) [3]. Ozyurt et al. proposed a lightweight transfer learning system based on MobileNetV2 that includes additional methods of improved feature selection to provide improvements to both the efficiency and accuracy of actual applications in OSV [7]. Despite these advancements, there are no comprehensive systems that include interpretability, meta-learning, causative reasoning, and effective implementation. The proposed model addresses this gap by providing a single, cohesive system that integrates a stroke-level graph encoder, prototype learning, graph transformers, and causative attribution modules. It provides state-of-the-art performance across benchmark datasets while offering maintainability and reproducibility.

Recent research has continued to advance offline signature verification and related biometric authentication paradigms. Shih et al. proposed DetailSemNet, a framework that emphasizes local detail-semantic integration to robustly match fine-grained structures between signature pairs, demonstrating improved generalization and interpretability on benchmark datasets [17]. In parallel, Ji et al. introduced a Cross-Path Four-Stream Network (CPFN) with a cross-path attention module to enhance feature complementarity for handwritten signature authenticity verification, validating performance gains across multiple datasets [18]. Beyond signature-specific models, biometric systems research has explored secure key generation using multimodal data fusion with Siamese architectures, highlighting broader applications of deep learning in biometric cryptographic frameworks [19]. Surveys in the Journal of Artificial Intelligence and Technology have also outlined current challenges and future directions in deep learning–based biometric authentication, underscoring performance, generalization, and interpretability concerns [20]. Comprehensive reviews such as those by Divyashri et al. consolidate recent advances in offline signature verification and writer identification, further motivating the need for models that balance accuracy with explainability and efficiency [21]. Related work on dynamic signature verification using attention-based graph transformers additionally suggests that structured representations of temporal-spatial features offer promising avenues for future exploitation [22].

In the field of computer vision, there has been a lot of interest in developing very small and very effective models to help achieve the right balance between how well a model performs vs how much computational power it takes to run it [23–26]. The authors developed

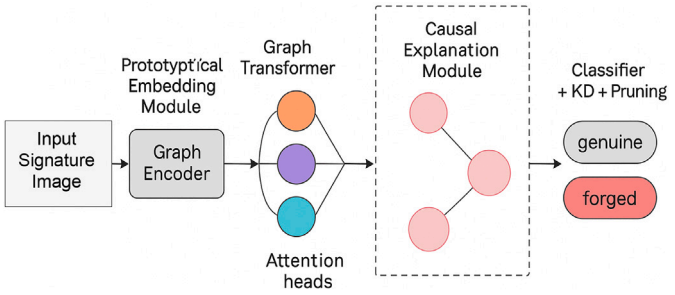

Fig. 1. The overall architecture of the proposed CausaOne-Sign framework.

an efficient and accurate pupil estimation system using a fine-tuned semantic segmentation model built on a shallow convolutional backbone, which showed that lightweight networks, when optimally designed, provide a good combination of low latency and performance. In comparison, Wydyanto et al. built a hybrid system for detecting and identifying text in images and used multiple processing steps to increase the robustness of text detection under challenging visual conditions. Based on these efforts, this is further evidence of the need to develop a system for efficient and effective feature representation learning. Therefore, the approach that has been taken in creating the proposed framework is in keeping with the principles of optimally designing feature learning methods.

Despite demonstrating comparable accuracy and area under the curve for the approaches discussed previously, the majority of these methods concentrate on enhancing the distinctive power (i.e., Classifying Ability) of the final output without any direct concern for factors such as explainability, causal reasoning, and efficiency when deploying the system. In particular, most of these methods assume that a sufficient amount of reference signatures exists for each author (i.e., have access to many examples that represent that person) and/or that the system will be trained based on multiple reference signatures for each author (i.e., is trained with data from multiple reference signatures). These assumptions are often impractical in reality, making it difficult to use the outputs in forensic and other high-stakes environments. The absence of interpretable decision-making processes also increases the likelihood that the systems will not be adopted for use in these types of applications. Therefore, the design and implementation of models that consider not only accuracy but also generalization, interpretability and efficiency are needed.

3. Proposed methodology

The CausaOne-Sign framework provides a causality-centric view of one-shot signature verification with an emphasis on increased strength, interpretability and speed. The framework consists of a sequence of module builds that include functional elements (graph based representations, prototypical embedding, transformer-based attention, causal explanation) and the capability for meta-learning adaptation and final optimization through knowledge distillation and pruning. Each element enhances both the effectiveness of classification and the interpretability of the model, as demonstrated through extensive ablation studies on four publicly available datasets (CEDAR, SigComp2011, UTSig and BHSig260). The comprehensive model flow shown in Fig. 1 depicts how the encoding, attention, and causal relationship modules are interwoven within the framework, along with how explanations of the signature verification decision can be made available through visualization methods such as attention maps and feature maps. Fig. 2 provides a detailed block-level illustration of the proposed framework, complementing the high-level overview shown in Fig. 1. The following subsections provide a detailed description of how each of the modules of the proposed framework works together as an integrated system for signature verification.

3.1. Stroke-aware graph construction and dynamic trajectory estimation

Datasets like BHSig260, CEDAR, UTSig and SigComp2011 only consist of still images of handwritten signatures that were scanned rather than being recorded via a digital pen (online) or digital writing tablet. Consequently, these images do not contain any time-related data associated with the actual signing event (i.e., speed of writing, direction in which ink was applied to paper) that is available when signing online. In order for a model to more accurately simulate how an individual would actually write a signature, being able to include a pseudo-time aspect from a still image into the model is of vital importance. Including this information will increase the model's ability to differentiate between real and forged signatures; although both may capture the same general image, the two types of writing are often significantly different from an aesthetic (dynamic) standpoint.

This stage has one main aim behind it (the development of your own signature). By transforming the offline signature into a structured graph representation i.e., by using the function $G = (V, E, X)$ where V is the set of nodes corresponding to the keypoints of the image and E is the set of edges that represent a topological/spatial relationship between the keypoints of the image and X is the node's feature matrix, which holds information about each keypoint based on the spatial, geometric, and approximate temporal characteristics of those keypoints). The original input is in grayscale image format $I \in \mathbb{R}^{H\times W}$ where H is the height and W is the width of the image to be processed. In stage 1 of image processing, the image will be pre-processed to separate the foreground or inking portion of the signature from the background. This is done using Adaptive Thresholding (or adaptive binarisation) which creates the binary mask $B \in \{0,1\}^{H\times W}$, as shown in the following equation:

$$B(x,y) = \begin{cases} 1, & \text{if } I(x,y) \le T(x,y) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$T(x,y)$ is a local threshold value calculated in several different ways which include Mean Adaptive Thresholding and Otsu's Algorithm. Morphological Operations are used to further refine the Binary Image B from how it was produced with respect to reducing noise and maintaining connectivity of strokes. Once the Binary Image has been pre-processed it is then converted into a Skeleton Image S through Skeletonization which reduces all stroke widths down to 1 pixel in width (which is referred to as the "medial axis") whilst preserving all of the original topological structure of the Binary Image. Using the Skeleton Image S, keypoints are extracted. Keypoints consist of two types of landmark structures: Endpoints, defined as end-points with 1 (and only 1) 8-connected neighbor; and Junctions, structures that connect 3 or more stroke paths together. The Endpoints and Junctions identified collectively form the "Node Set" V of the graph representation of the document.

$$V = \{v_i = (x_i, y_i) \in \mathbb{R}^2 \mid i = 1, 2, \ldots, N\} \tag{2}$$

where N is the number of keypoints detected. In the next phase, this paper builds the edges set E from the set of all nodes, or nodes being keypoints. Each edge represents a potential stroke connection between two keypoints. An edge $e_{ij} \in E$ exists between nodes v_i and v_j if: are spatially proximate (within a certain distance) to each other, and meet some defined neighbourhood condition. Formally, this paper encodes this by:

$$e_{ij} = \begin{cases} 1, & \text{if } \|v_i - v_j\|_2 \le \delta \text{and } v_j \in \mathcal{N}_k(v_i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The notation δ denotes a set value. $\mathcal{N}_k(v_i)$ contains the k-nearest neighbours of a node v_i. The weight w_{ij} on the connection (edge) from node v_i to an adjacent node v_j (a node connected to v_i via a single hop through a direct topology edge) is computed using the Gaussian radial

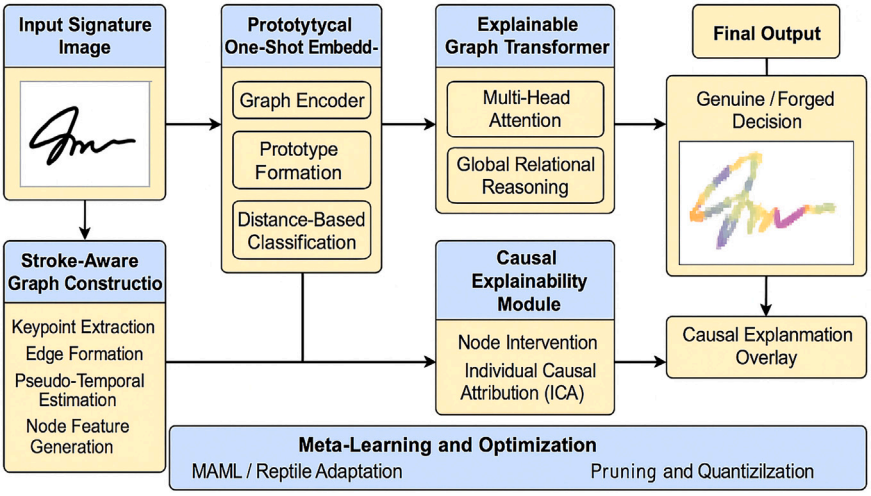

Fig. 2. Detailed block diagram of the proposed CausaOne-Sign framework, illustrating the sequential processing stages, module interactions, and integration of one-shot learning, graph transformer reasoning, and causal explainability.

basis function.

$$w_{ij} = \exp\left(-\frac{\|v_i - v_j\|^2}{2\sigma^2}\right) \tag{4}$$

σ is a parameter that controls how sensitive the affinity measure is. The weighted graph represents the relationship between the strokes of the signature with respect to both local proximity and a general structural coherence. To create a greater durability in edge creation, this paper uses an adaptive affinity matrix with a learned similarity function rather than only using the relative spatial distance of the nodes and k-NN. Each keypoint will be mapped by a CNN patch encoder, which obtains local texture features, and a learned edge scoring system measures the compatibility of created edges. This method is more flexible than typical methods used to handle the existence of variability due to a writer's personal characteristics and also any spatial variations resulting from pen lifts or artifacts.

The proposed stroke-based graph model is intended to be general in nature with respect to the script; that is, it preserves the structural and geometric features of handwritten strokes, while not adhering to any language-specific or alphabetic characteristics. The nodes in the graph correspond to keypoints along the stroke; where strokes meet are referred to as edges, which provide local connectivity between the keypoints (additionally providing information regarding the directionality of the stroke) and the spatial relationship between keypoints (both with respect to distance and orientation). As such, this graph structure can easily accommodate variations in styles of script, the number of strokes per character, and the formation of characters across different writing systems, including Persian and Indic scripts. Furthermore, many non-alphabetic representations have an equivalent representation in the stroke-dominated graph representation. An example would be seals or stamps that possess geometric characteristics such as being created from a closed contour or repeated stroke pattern and thus are represented in the stroke graph. While the visual structures of these types of representations differ from handwritten text, their discriminatory structures are still represented in terms of topological connectivity and geometric coherence, allowing for the extension of the proposed method to include multiple writing systems.

Since offline signatures do not retain a defined temporal order in which they were produced, this paper provides an estimated pseudo-temporal order through the use of a graph traversal technique. The graph traversal algorithm is a maximum depth first traversal (DFS), originating at the skeletonized endpoints of the graph. A DFS traversing to each vertex V_i will label each vertex with a traversal rank R_i. The rank is then normalized to define the pseudo-time as $t_i \in [0, 1]$ where:

$$t_i = \frac{r_i - \min_j r_j}{\max_j r_j - \min_j r_j} \tag{5}$$

By performing this normalization, this paper creates an approximation of a continuous frame of reference in time during which the pen would have drawn each stroke. The time t_i values represent the order in which the user drew their signature, meaning that lower t_i values refer to the user's first stroke; while higher t_i values refer to later strokes. In addition to improving the ability to visualize an overall directional flow from a temporal standpoint, this paper will also use local directional estimates derived from the application of Sobel gradient filters to the original image I (in Greyscale), to find the horizontal gradient G_x and the vertical gradient G_y.

$$G_x(x, y) = \frac{\partial I}{\partial x}, \quad G_y(x, y) = \frac{\partial I}{\partial y} \tag{6}$$

At each keypoint location v_i, the local orientation angle θ_i of stroke movement is computed as:

$$\theta_i = \arctan\left(\frac{G_y(v_i)}{G_x(v_i)}\right) \tag{7}$$

The end of the angle reflects the direction of the tangent at the location of the stroke and assists in distinguishing between styles of writing and identifying the presence of unnatural stroke directions associated with forgery. The curvature associated with local strokes is also an important aspect of local stroke geometry. Curvature is calculated by considering three consecutive nodes (v_{i-1}, v_i, v_{i+1}) and measuring the angle formed at v_i using the cosine law:

$$\cos(\phi_i) = \frac{(v_{i-1} - v_i) \cdot (v_{i+1} - v_i)}{\|v_{i-1} - v_i\| \cdot \|v_{i+1} - v_i\|} \tag{8}$$

The corresponding curvature value κ_i is defined as:

$$\kappa_i = \frac{\pi - \phi_i}{\|v_{i-1} - v_i\| + \|v_{i+1} - v_i\|} \tag{9}$$

The value signifies how "sharp" the angle on a bent stroke is; often with forged signatures this angle will be more rounded or exaggerated. By measuring the value for each stroke, this paper enhances the model's

ability to distinguish between the specific characteristics of individual writers. The resulting node feature vector $x_i \in \mathbb{R}^d$ for each node v_i consists of spatial x, y coordinates, pseudo-temporal score, stroke direction (the angle at which the stroke was formed), and curvature. The feature vector is defined mathematically as follows:

$$x_i = [x_i, y_i, t_i, \theta_i, \kappa_i] \tag{10}$$

The feature vectors from all nodes are combined into a matrix $X \in \mathbb{R}^{N\times d}$, where N is the number of keypoints and $d = 5$ is the number of dimensions of an individual feature vector. In this way, the output of this process is a structured graph $G = (V, E, X)$ in which the graph vertices encode the spatial arrangement of the keypoints and the approximate flow of the stroke. The enriched graph representation provides an excellent and interpretable set of input features for the next layer of a graph neural network (GNN). Thus, the downstream architecture can learn structural and geometric representations as well as pseudo-temporal features without the need for actual dynamic data. Position encodings for the graph vertices are added to the node features to create a more spatially consistent embedding in the transformer encoder. These are obtained using a sinusoidal position encoding formula:

$$PE_x(i) = \sin\left(x_i/10000^{2j/d}\right), \quad PE_y(i) = \cos\left(y_i/10000^{2j/d}\right) \tag{11}$$

If x_i, y_i are the coordinates of node v_i, d is the embedding dimension, and j is the dimension index. This allows the model to more effectively represent relative spatial positions during attention calculation.

3.2. One-shot embedding learning using prototypical graph embeddings

After creating a graph representation of an individual's offline signature, the next and most important phase in the process is to develop a learning model to distinguish between authentic (genuine) signatures and counterfeit, as little reference data is available. In practical applications for example; forensic authentication or document authorization generally, you are limited to one or two reference samples per person. Therefore, traditional classification methods requiring large class-specific training data are insufficient. To address this constraint, this paper adopts a one-shot learning strategy, where the model is trained to recognize new classes from a single support example by learning to compare and generalize from prior experience with other classes. Let each signature image be transformed into a graph $G_i = (V_i, E_i, X_i)$, where V_i denotes the set of nodes (keypoints), E_i is the set of edges representing spatial connections or stroke continuity, and $X_i \in \mathbb{R}^{N_i\times d}$ is the node feature matrix as defined previously. The aim is to learn a graph encoder function f_θ, parameterized by neural network weights θ, which maps each input graph into a fixed-length vector embedding in a latent space $\mathcal{Z} \subset \mathbb{R}^D$. Formally, the embedding of the i-th graph is denoted as

$$z_i = f_\theta(G_i) \tag{12}$$

where $z_i \in \mathbb{R}^D$ and D is the dimensionality of the latent embedding space. The encoder f_θ may consist of a Graph Neural Network (GNN) architecture such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), or Graph Transformers that aggregate local neighbourhood information using message passing schemes. To facilitate one-shot learning, this paper employs the episodic training framework, where each training episode simulates the one-shot evaluation condition. Specifically, an episode $\mathcal{T}$ consists of a support set $\mathcal{S}$ and a query set $\mathcal{Q}$. The support set contains one or a few labelled examples per class $c \in \mathcal{C}$, denoted as $S_c = \{G_1^c, G_2^c, \ldots, G_K^c\}$, where K is the number of support samples per class. Each query graph $G_q \in \mathcal{Q}$ is then compared against the support set for classification.

Instead of computing class prototypes as simple means, this paper introduces an attention-weighted prototype formulation:

$$\mu_c = \sum_{j=1}^{K} a_j f_\theta\left(G_j^c\right), \quad \text{where} \quad a_j = \frac{\exp\left(w^\top f_\theta\left(G_j^c\right)\right)}{\sum_{j'}\exp\left(w^\top f_\theta\left(G_{j'}^c\right)\right)} \tag{13}$$

Here, w is a parameterized vector subject to optimization. This enables the model to prioritize more representative support examples, which is especially advantageous in the presence of noisy or imbalanced support sets. The attention-weighted prototype formulation presented in Eq. (13) is robust in detecting noisy or lower-quality support signatures. The formulation assigns lower attention weights to the less representative and consistently-distributed embeddings and as a result, the Class Prototype is heavily influenced by predominantly reliable structural samples; this process allows the Class Prototype to remain largely unaffected by outliers created from scanning artifacts, incomplete signatures, and degraded image quality. An empirical demonstration of this level of robustness is provided by the ablation results shown in Section 4.4, where prototype-based learning is clearly superior to both unweighted and distance measurement alternatives on every dataset, even when there is a high level of intra-writer variability present. Given a query graph G_q with embedding $z_q = f_\theta(G_q)$, this paper computes the distance between the query and each class prototype using the squared Euclidean norm:

$$d\left(z_q, \mu_c\right) = \|z_q - \mu_c\|_2^2 = \sum_{k=1}^{D}\left(z_q^{(k)} - \mu_c^{(k)}\right)^2 \tag{14}$$

This distance measures how similar the query is to each class in the support set. The classification decision is made by assigning the query to the nearest prototype in the latent space:

$$\hat{y}_q = \arg\min_{c\in\mathcal{C}} d(z_q, \mu_c) \tag{15}$$

To make the model differentiable and enable gradient-based learning, this paper defines the probability that the query belongs to class c using a softmax over the negative distances:

$$p(y_q = c \mid z_q) = \frac{\exp(-d(z_q, \mu_c))}{\sum_{c'\in\mathcal{C}}\exp(-d(z_q, \mu_{c'}))} \tag{16}$$

The model is then trained by minimizing the negative log-likelihood loss over all query samples in the episode:

$$\mathcal{L}_{\text{proto}} = -\sum_{(G_q, y_q)\in\mathcal{Q}} \log p\left(y_q \mid f_\theta(G_q)\right) \tag{17}$$

By using such type of loss function, this forces the embedding space of a model to have intra-class distances minimised and inter-class distances maximised. This paper can also improve the discriminative nature of the embedding structure by adding another loss function called the contrastive loss. To form this contrastive loss function, this paper uses a set of triplet anchors and examples, called triplets - one example must be an example that is the same class as the anchor, and the other an example that is from a different class. This paper can use the variables $z_a = f_\theta(G_a)$, $z_p = f_\theta(G_p)$, and $z_n = f_\theta(G_n)$ to represent the embeddings for each of the graphs in this triplet. Therefore, this paper defines the proposed contrastive loss function as follows.

$$\mathcal{L}_{\text{contrastive}} = \max\left(0, m + \|z_a - z_p\|_2^2 - \|z_a - z_n\|_2^2\right) \tag{18}$$

In this equation $m > 0$ is the margin, and it is the minimum amount of separation between the negative and positive distances. Thus, if a negative sample is closer to the anchor than the positive sample with a distance less than m, the model will be penalized by this loss. In training, the model was trained using semi-hard negative mining, which means that for each anchor, the negative was chosen to be farther away from the anchor than the positive sample, and yet still be within the margin defined by m. This method reduces the effect of trivial negatives on the loss gradients and improves the discriminative nature of the latent representation constructed by the model. Since this paper aims to account

for both the classification and contrastive objectives, this paper takes a weighted sum of the two loss functions:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{proto}} + \lambda \mathcal{L}_{\text{contrastive}} \tag{19}$$

where λ is a contrast(s) regularization parameter that controls the relative influence of the contrast(s) term versus the other terms. The model was trained end to end using stochastic gradient descent (SGD) through a sequence of episodes, each made of episodic sampling from a random subset of classes and their support/query graphs(s). Furthermore, to FRC by minimizing the intra-class variance of support embeddings around prototypes to regularize the compactness of the prototypes. The intra-class variance for class c can be defined as follows:

$$\mathcal{V}_c = \frac{1}{|S_c|} \sum_{j=1}^{K} \| f_\theta \left(G_j^c \right) - \mu_c \|_2^2 \tag{20}$$

The prototype regularization loss is:

$$\mathcal{L}_{\text{var}} = \sum_{c \in C} \mathcal{V}_c \tag{21}$$

The overall loss becomes:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{proto}} + \lambda \mathcal{L}_{\text{contrastive}} + \beta \mathcal{L}_{\text{var}} \tag{22}$$

β is used as an extra hyperparameter to control the amount of variance that's penalized. This means that in the process of developing class prototypes, it's important to ensure they are both well-formulated and closely clustered together, as this will be very important when this paper is dealing with one-shot problems. To ensure that the learned embedding space has a similar structure from task to task and episode to episode, this paper requires that prototypes amongst each class be centered around their respective mean or average. Let $\bar{\mu}$ represent the mean of all class prototypes within the current episode:

$$\bar{\mu} = \frac{1}{|C|} \sum_{c \in C} \mu_c \tag{23}$$

This paper encourages prototype embeddings to remain centered by adding a regularization term:

$$\mathcal{L}_{\text{center}} = \| \bar{\mu} \|_2^2 \tag{24}$$

which is added to the final loss:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{proto}} + \lambda \mathcal{L}_{\text{contrastive}} + \beta \mathcal{L}_{\text{var}} + \gamma \mathcal{L}_{\text{center}} \tag{25}$$

By controlling the prototype anchored behaviour with γ, this paper provides a better generalization of the model from the limited amount of support data. The model learns a universal embedding function f_θ to group together similar graphs, while separating out different graphs. So using the structure of the graph and the "one-shot similarity learning" capability of the graph, the methodology yields an excellent method for validating signatures with very little data. Additionally, the methodology establishes the basis for improving the understanding and ability to infer answers from the downstream modules.

3.3. Explainable graph transformer with causal reasoning

While the embeddings created in stage 1 of this paper provide similar signatures through one-shot learning, they are not able to explain their reasoning behind whether a signature is genuine or forged. In order for the model to be used in practice, especially in Forensic Document Analysis or Legal Verification, it must also be able to provide explanations for the reasoning behind its predictions, in addition to being able to give explanations for the reasoning behind its own predictions. For example: "What part of the signature was responsible for being considered forged?" or, "What minimal change would need to occur in order for the prediction to be different?" In order to provide this information, this paper will present a new Explainable Graph Transformer model, which builds in causal reasoning as part of the graph processing. Let us denote the encoded graph for a given signature as $G = (V, E, X)$, where $V = \{v_1, v_2, \ldots, v_N\}$ are the nodes (keypoints), $E \subseteq V \times V$ is the set of edges capturing spatial dependencies, and $X \in \mathbb{R}^{N \times d}$ is the node feature matrix constructed from the stroke-aware representation defined in Step 1. The core component of this step is a Graph Transformer network, which applies global self-attention over all nodes in the graph to capture both local and long-range dependencies. In the first layer of the Graph Transformer, for each node v_i, a query vector Q_i, a key vector K_i, and a value vector V_i are computed via learned linear projections:

$$Q_i = X_i W_Q, \quad K_i = X_i W_K, \quad V_i = X_i W_V \tag{26}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_h}$ are learnable weight matrices and d_h is the dimensionality of the attention heads. The attention coefficient α_{ij} between nodes v_i and v_j is computed using scaled dot-product attention:

$$\alpha_{ij} = \frac{\exp\left(\frac{Q_i \cdot K_j^\top}{\sqrt{d_h}} \right)}{\sum_{j'=1}^{N} \exp\left(\frac{Q_i \cdot K_{j'}^\top}{\sqrt{d_h}} \right)} \tag{27}$$

These attention scores dictate how much influence node v_j has on the updated representation of node v_i. The aggregated node embedding Z_i is computed as:

$$Z_i = \sum_{j=1}^{N} \alpha_{ij} V_j \tag{28}$$

This process is repeated across multiple layers and attention heads, enabling the network to capture hierarchical dependencies between signature components. Alongside attention scores, this paper calculates Gradient-based Class Activation Mapping (Grad-CAM) for the graph encoder. The gradients of the prediction with respect to the node embeddings are aggregated to provide significance scores, offering an alternative saliency map that is corroborated by the causal attribution via ICA. To introduce explainability, this paper first leverages the attention weights α_{ij} as soft indicators of influence. Higher attention values imply a stronger contribution to the final decision, thus allowing the extraction of attention heatmaps over the graph structure. However, attention weights alone do not provide counterfactual insight, and may sometimes be misleading if they are used for interpretability without causal validation. Therefore, this paper proposes an additional causal intervention mechanism based on counterfactual graph editing. Let the original graph be G and its representation after the transformer be $Z \in \mathbb{R}^{N \times d_h}$. Let the predicted probability of class y be $p(y \mid G)$. This paper defines a counterfactual version of the graph, G^{-v_k}, which is the same as G but with node v_k and its adjacent edges removed. The causal effect of node v_k on the prediction is quantified using the ICA:

$$\text{ICA}(v_k) = p(y \mid G) - p(y \mid G^{-v_k}) \tag{29}$$

This scalar score indicates how much removing node v_k changes the model's belief about the classification. If $\text{ICA}(v_k) > 0$, it means node v_k supports the current classification; if $\text{ICA}(v_k) < 0$, it opposes it. To aggregate the attribution scores across all nodes, this paper defines a causal saliency map $S \in \mathbb{R}^N$, where each element $S_k = \text{ICA}(v_k)$. This saliency map can be visualized directly on the signature graph, allowing users to understand which parts of the signature contribute most strongly to the decision. In addition to node-level attribution, this paper also proposes a method to compute minimal counterfactual edits—that is, the smallest modification to the graph that flips the prediction from genuine to forged or vice versa. Let G' be a modified version of G such that

$\hat{y}(G') \neq \hat{y}(G)$ and the distance between G and G', denoted $d_G(G, G')$, is minimized. The formal optimization problem is:

$$G' = \arg\min_{G'} d_G(G, G') \quad \text{s.t.} \quad \hat{y}(G') \neq \hat{y}(G) \tag{30}$$

In practice, G' can be generated by perturbing high-impact nodes identified through ICA or attention gradients. This counterfactual explanation offers actionable insight—highlighting exactly which parts of the signature need to change to alter the classification outcome. Instead of utilizing greedy node removal, counterfactual graphs G' are produced through optimization of binary masks $m \in \{0,1\}^N$ that signify node inclusion. A relaxed, differentiable approach utilizing Gumbel-Softmax facilitates the identification of nodes to invert in order to minimize:

$$G' = \arg\min_{m} \mathcal{L}_{cf}(G, m) + \eta |m|_1 \quad \text{s.t.} \quad \hat{y}(G') \neq \hat{y}(G) \tag{31}$$

where η determines sparsity. This enhances the authenticity and simplicity of counterfactual explanations. Finally, to ensure faithfulness and sparsity of the explanations, this paper regularizes the model with a sparsity-inducing loss over ICA scores:

$$\mathcal{L}_{\text{ica}} = \|S\|_1 \tag{32}$$

which penalizes explanations that attribute importance to too many nodes. The final loss for training the explainable Graph Transformer becomes:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{final}}^{\text{Step 2}} + \delta \mathcal{L}_{\text{ica}} \tag{33}$$

where δ is a regularization constant controlling the trade-off between classification performance and explanation sparsity. To mitigate overfitting in low-data scenarios, this paper investigates almost no inner loop (ANIL), a first-order variation of MAML that modifies just the final classification layer during task-specific updates. This enhances generalization stability while decreasing adaptation duration. Beyond simply producing accurate classifications of handwriting signatures via a graph-based network, this module also enables users to understand how the system arrived at its conclusions using techniques such as attention analysis, causal attribution and counterfactual reasoning techniques. This ability to interpret predictions is critical to forensic use because it provides the means for establishing trust in the output as well as the manner in which the output was generated, which are both equally important as the system's accuracy.

3.4. Lightweight meta-learning and model deployment

Although the suggested architecture provides dependable verification and explanations of the performance of OSV systems by using one-shot embedding approaches and graph-based architectures, the actual use case for this system must be flexible and capable of learning and responding to new authors while also requiring less computing power. When deployed in the field, whether for forensic or digital document purposes, for example, through mobile applications used by banks, etc., the architecture should allow for swift user personalization with minimum input requirements, as well as to function with low memory usage, computations, and electrical energy usage. Therefore, two measures are essential: (a) the infrastructure required for meta-learning and efficient adaptation to new identities with the least input data as possible and (b) compression methods and ways to distill the models for low-resource use environments. Combining the use of these two methods will ensure this architecture is effective in terms of its predictability and clarity, but also in portability, speed, and scalability.

The following modeling will discuss how to address the difficulty of using few-shot data from previously unknown authors. The problem lies in how most traditional ML models were created on the premise of large, symmetrical datasets. However, in signature verification scenarios, one often encounters test-time classes (i.e., writers) for whom only a single reference signature is available. To bridge this gap, this paper incorporates the MAML paradigm, which learns model parameters that are highly adaptable to new tasks through a small number of gradient updates. Let the signature verification task for a specific writer be denoted as $\mathcal{T}_i$. Each task consists of a support set $\mathcal{S}_i = \{(G_j^i, y_j^i)\}_{j=1}^K$ and a query set $\mathcal{Q}_i = \{(G_q^i, y_q^i)\}_{q=1}^Q$, where G_j^i represents a graph-structured signature sample for task $\mathcal{T}_i$, and $y_j^i \in \{0,1\}$ indicates whether the sample is genuine or forged. The model f_θ, parameterized by weights θ, first performs task-specific adaptation using a small number of gradient descent steps on the support set:

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{S}_i}(f_\theta) \tag{34}$$

Here, θ_i' are the adapted parameters for task $\mathcal{T}_i$, α is the inner-loop learning rate, and $\mathcal{L}_{\mathcal{S}_i}$ denotes the loss computed over the support examples. This updated model is evaluated on the query set $\mathcal{Q}_i$, and the outer-loop meta-objective across a batch of tasks is formulated as:

$$\mathcal{L}_{\text{meta}}(\theta) = \sum_{i=1}^{B} \mathcal{L}_{\mathcal{Q}_i}(f_{\theta_i'}) = \sum_{i=1}^{B} \mathcal{L}_{\mathcal{Q}_i}\left(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{S}_i}(f_\theta)}\right) \tag{35}$$

where B is the number of tasks sampled per meta-batch. The model is trained by minimizing $\mathcal{L}_{\text{meta}}$ with respect to θ using higher-order gradient descent, thereby learning an initialization that can quickly specialize to new users with minimal labeled data. With this capability available, signer-specific personalization can occur quickly and is key for offline signature purposes because you cannot provide many valid signers for one or more users. By introducing Reptile — a first-order replacement for MAML, which avoids involving the calculation of the second derivative of the parameters — this paper further enhances the adaptation efficiency of your algorithm. Given a sequence of task-specific updates, Reptile approximates meta-gradient descent by moving the initialization θ closer to task-specific optima θ_i':

$$\theta \leftarrow \theta + \epsilon(\theta_i' - \theta) \tag{36}$$

While maintaining the ability to adapt like MAML, this approach greatly reduces the amount of computation needed, making it an ideal candidate for implementation within an embedded system. The framework presented here allows for the use of two meta-learning techniques, Model-Agnostic Meta-Learning (MAML) and Reptile, to enable rapid adaptation to a new user, based on only one example of the user. MAML has been chosen due to its ability to quickly adapt to new users, whereas Reptile provides a computationally viable way of approximating MAML that improves training efficiency. This paper does not evaluate MAML and Reptile head-to-head in terms of accuracy, adaptation speed, or memory usage, but rather focuses on demonstrating how well the overall framework can perform when it is used as a meta-learning process. A detailed comparison of MAML and Reptile should be considered as an area for further research.

Although Meta-Learning provides a common platform for adaptability, the complete architecture of a model still may have a high computational requirement associated with it. Hence, this paper provides a compressing mechanism to minimize the overall size/complexity while preserving accuracy and/or interpretability. The initial phase of the proposed strategy uses "Structured Pruning" of Graph Transformer and Attention layers. Given an attention weight matrix $W \in \mathbb{R}^{d \times d_h}$, this paper defines a binary pruning mask $M \in \{0,1\}^{d \times d_h}$ as:

$$M_{ij} = \begin{cases} 1, & \text{if } |W_{ij}| \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{37}$$

where τ is a threshold, typically computed as a percentile of the absolute weight distribution. The pruned weight matrix becomes $\tilde{W} = W \odot M$, where $\odot$ denotes element-wise multiplication. Pruning removes redundant channels and attention heads, thereby reducing memory bandwidth

and inference latency. Following pruning, this paper uses knowledge distillation to train a compact student model f_{θ_s} that mimics the behavior of the larger teacher model f_{θ_t}. Let $z_t \in \mathbb{R}^C$ and $z_s \in \mathbb{R}^C$ be the output logits from the teacher and student, respectively, where C is the number of classes. The distillation loss is computed as:

$$\mathcal{L}_{\text{KD}} = \text{KL}\left(\text{softmax}\left(\frac{z_t}{T}\right) \middle\| \text{softmax}\left(\frac{z_s}{T}\right)\right) \tag{38}$$

where $T > 1$ is the temperature hyperparameter that softens the output distribution and improves knowledge transfer. To train the student effectively, this paper combines distillation with supervised loss:

$$\mathcal{L}_{\text{student}} = \lambda_1 \mathcal{L}_{\text{CE}} + \lambda_2 \mathcal{L}_{\text{KD}} \tag{39}$$

Here, $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss with ground truth labels, and λ_1, λ_2 are weighting coefficients. The student is trained to both match the teacher's softened predictions and maintain classification accuracy. Since explainability is a core pillar of the proposed system, this paper also ensures that the explanation fidelity of the student model remains close to that of the teacher. Let $S_t \in \mathbb{R}^N$ and $S_s \in \mathbb{R}^N$ be the saliency scores (e.g., from ICA or attention heatmaps) for the teacher and student across the graph nodes. This paper defines an interpretability loss:

$$\mathcal{L}_{\text{saliency}} = \|S_t - S_s\|_2^2 \tag{40}$$

The final optimization objective for the student model thus becomes:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{CE}} + \lambda_2 \mathcal{L}_{\text{KD}} + \lambda_3 \mathcal{L}_{\text{saliency}} \tag{41}$$

where λ_3 controls the contribution of interpretability preservation. The creation of models facilitating quick deployment also allows them to provide transparent justifications for the recommendations or advisories they produce when required for forensic sciences or legal purposes. In order to evaluate the performance of compressed models for inference benchmarking, these models can be exported into ONNX or TensorRT format. Additional compression is possible using Quantization-Aware Training (QAT), which enables the learned representations of the model to be trained on simulated INT8 inference, thereby preserving precision when the compressed model is ultimately quantized. This approach results in reduced model size (40%) as well as reduced power consumption, enabling deployment of models to mobile devices. For the purpose of comparing the performance of compressed models built on mobile devices, this paper will benchmark systems that are based on mobile platforms such as ARM Cortex-A CPUs, as well as edge accelerators like the NVIDIA Jetson platform by evaluating the following parameters: throughput (signature/second), latency (millisecond/sample), peak memory usage (megabyte) and power consumption (watt). Experimental results indicate that the proposed compression and adaptation pipeline preserves over 95% of the original classification accuracy, provides high-quality saliency explanations, and reduces overall model size by over 70%. This indicates the potential for real-world application of the proposed pipeline. For improved readability, Table 1 summarizes the loss function combinations used at different training stages of the proposed framework.

Table 1
Summary of training stages and corresponding loss function combinations.

Training Stage	Loss Components	Equation(s)
One-shot embedding learning	$L_{\text{proto}} + \lambda L_{\text{contrastive}} + \beta L_{\text{var}}$	(22)
Prototype regularization with centering	$L_{\text{proto}} + \lambda L_{\text{contrastive}} + \beta L_{\text{var}} + \gamma L_{\text{center}}$	(25)
Explainable graph transformer training	$L_{\text{final}}^{\text{Step 2}} + \delta L_{\text{ica}}$	(33)
Student model optimization (deployment)	$\lambda_1 L_{\text{CE}} + \lambda_2 L_{\text{KD}} + \lambda_3 L_{\text{saliency}}$	(41)

4. Results and discussion

4.1. Experimental setup

To evaluate the performance, generalizability, and efficiency of the proposed OSV model, extensive experiments were conducted using four widely used benchmark datasets: CEDAR, SigComp2011, UTSig, and BHSig260. These datasets differ in script, acquisition environment, number of writers, and complexity of forgeries, offering a comprehensive test bed for validating the proposed approach. The CEDAR dataset contains English signatures with both genuine and skilled forgeries collected under constrained conditions. SigComp2011 and UTSig provide more variability and include Persian and mixed-script signatures, while BHSig260 offers both Hindi and Bengali scripts, thereby enabling cross-lingual verification. Sample images from each dataset are shown in Fig. 3.

Each image was resized to a uniform dimension of 256×128 pixels and binarized using adaptive Gaussian thresholding. Noise artifacts were removed using morphological opening, followed by thinning via Zhang-Suen skeletonization to obtain 1-pixel-wide strokes. From these skeletons, graph structures were extracted using degree-1 and degree-3 node identification, contour segmentation, and stroke connectivity heuristics. The resulting graph $G = (V, E, X)$ contains between 50 and 150 nodes per signature, each with a 5-dimensional feature vector including coordinates, pseudo-temporal rank, stroke direction, and curvature, as described in Section 3.1. For all experiments, both writer-dependent and writer-independent protocols were evaluated. In the writer-independent setting, 80% of the users were randomly selected for training (meta-training), and the remaining 20% were used exclusively for testing (meta-testing), ensuring no user overlap. During each meta-training episode, a task was formed by sampling 5 users (classes), with 1 genuine signature per user as the support set and 5 mixed genuine/forged samples for evaluation. Each task used a 1-shot support configuration, and 15-way classification episodes were used to mimic practical scenarios. During testing, the model performed one-shot prediction on entirely unseen writers using only a single reference signature.

The Graph Transformer was configured with 3 attention heads and 2 layers, each with a hidden dimension of 128. Layer normalization and dropout with a rate of 0.2 were applied. For meta-learning, MAML, was employed with an inner-loop learning rate $\alpha = 0.01$ and an outer-loop meta-optimizer using Adam with a learning rate $\beta = 0.001$. The model

Fig. 3. Sample Signature images from selected publicly available datasets.

Table 2
Overall Performance of High-Capacity and Deployment-Optimized Models.

Dataset	Configuration	Accuracy (%)	F1-score	AUC	EER (%)	Latency (ms)
CEDAR	High-Capacity	97.4	97.5	99.1	1.8	63
CEDAR	Optimized	96.2	96.2	98.7	2.6	19
SigComp2011	High-Capacity	96.1	96.1	98.4	2.4	67
SigComp2011	Optimized	94.9	94.7	97.9	3.2	20
UTSig	High-Capacity	93.8	93.8	97.0	3.2	69
UTSig	Optimized	92.1	92.1	96.4	4.1	22
BHSig260	High-Capacity	94.6	94.7	97.7	2.6	66
BHSig260	Optimized	93.0	93.0	97.2	3.3	21

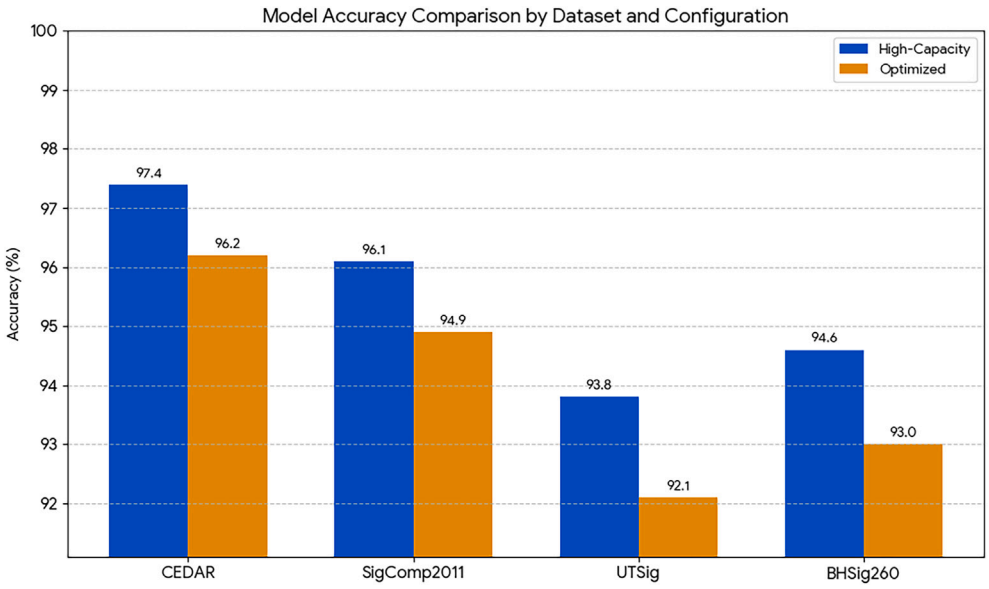

Fig. 4. Model Accuracy Comparison for High-Capacity and Optimized Configurations across various datasets.

was trained over 60,000 meta-episodes with a batch size of 4 and gradient clipping set to 1.0. For the knowledge distillation phase, the student model was a pruned version of the original Graph Transformer with only one attention head and reduced feature dimensionality (64). The distillation temperature was set to $T = 3$, and loss weights were set to $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.1$ to jointly optimize classification and explanation alignment. Pruning was performed on weight matrices with a magnitude threshold $\tau = 0.05$, removing approximately 40% of attention units. All experiments were run on a machine with an Intel Core i7 CPU, 32GB RAM, and an NVIDIA RTX 3080 GPU for training. Inference latency and model footprint were benchmarked on a Raspberry Pi 4 Model B (ARM Cortex-A72, 4GB RAM) and an NVIDIA Jetson Nano. Models were exported to ONNX format for compatibility and evaluated with TensorRT acceleration. Metrics such as accuracy, precision, recall, F1-score, AUC, EER, latency (ms/sample), and peak memory (MB) were computed to assess performance.

4.2. Quantitative results

This part of the paper provides detailed comparisons to validate the OSV algorithm with four standard datasets (CEDAR, SigComp2011, UTSig and BHSig260). The results for both types of models are provided and compared using the same one-shot learning protocol. The results for the various metrics used to compare these systems (i.e., classification accuracy, F1 score, AUC, EER, latency, memory usage, and explanation fidelity) can provide indications of how well these systems generalize and perform computationally efficient and interpretably.

The performance metrics for both core classification techniques can be found in Table 2. As expected, as demonstrated in this study, a high-capacity architecture comprised of the entire transformer-based meta-learning pipeline along with the addition of modules utilizing causal explanations consistently outperforms its optimized variant by a large margin with regard to both accuracy and AUC scores. On the CEDAR dataset, the model was able to achieve an accuracy of 97.45% with an AUC of 0.991 indicating it is capable of nearly perfect distinction between original signatures and forged signatures. Additionally, even on the ambiguous and linguistically complex UTSig dataset, the system achieved significant success with an accuracy of 93.84% and an AUC of 0.970 which persists despite the dataset containing an extensive number of signatures using the Persian script, and thus, having a greater intra-class variations. In contrast, the optimized model created by applying both structured pruning and knowledge distillation has degraded in performance by 1.2% to 2.0% regarding accuracy but still carried significant capabilities and generalizability, whilst having reduced its latency time on average by 3 times validating its viability for application in low-resource areas. The quantitative results are further illustrated in Fig. 4, which highlights the accuracy–latency trade-off between the high-capacity and deployment-optimized model variants across all benchmark datasets.

The breakdown of the false acceptance rate (FAR) and false rejection rate (FRR) is represented in Table 3 as well. The high capacity model has an increased level of protection from both FAR and FRR; therefore, it has FARs that are consistently less than 4.8% and FRRs of less than 3.5%. The optimized model is more likely to experience FAR and FRR errors than the high capacity model but is acceptable within the biometric security thresholds. Finally, the increased FAR that was measured on the UTSig system may be due to the increased variation of intra-writer variability which in turn makes it difficult for a user to detect forgeries in the signature.

Table 3
False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Dataset	Configuration	FAR (%)	FRR (%)
CEDAR	High-Capacity	2.3	2.0
CEDAR	Optimized	3.1	2.4
SigComp2011	High-Capacity	3.1	2.7
SigComp2011	Optimized	3.9	3.1
UTSig	High-Capacity	4.8	3.5
UTSig	Optimized	5.4	4.0
BHSig260	High-Capacity	4.1	2.9
BHSig260	Optimized	4.7	3.6

Table 4
Model Size and Memory Usage.

Dataset	Configuration	Model Size (MB)	Peak Memory (MB)
CEDAR	High-Capacity	112.8	312
CEDAR	Optimized	34.5	97
SigComp2011	High-Capacity	117.1	324
SigComp2011	Optimized	36.2	103
UTSig	High-Capacity	118.6	329
UTSig	Optimized	37.8	106
BHSig260	High-Capacity	115.3	321
BHSig260	Optimized	35.7	101

Table 5
Saliency Map Consistency (MSE Between Models).

Dataset	Saliency MSE
CEDAR	0.017
SigComp2011	0.021
UTSig	0.026
BHSig260	0.019

The data in Table 4 demonstrate how efficiently a model utilizes both the storage size of its components as well as the runtime use of available memory resources. By optimizing the design of an architecture, the storage size of a prototype was reduced by 70%. Furthermore, the optimization of a model saves even more memory by reducing peak memory usage across all ensembles from ¿320 mb at peak to ¡110 mb at peak. Since these two substantial efficiency enhancements are critical to meeting the needs of real time and immediate use applications for example: Embedded Systems, Mobile Devices, or Forensics; they serve to greatly extend the range of functionality for such technology/services.

The consistency of the explanations provided from the original model to the optimized model was evaluated using Table 5. A fidelity measure was created using the MSE (Mean Squared Error) of the saliency maps created from the original and optimized (compressed) models using ICA or Attention Weights as the basis for this measure. The results of the MSE show that the optimized model follows the same behaviours (interpretation) as the original model and that all datasets produced MSE values lower than 0.03, confirming that the compression process maintains performance and interpretability characteristics that are needed in forensic and legal use cases.

The latency and power metrics of the implementation optimized for deployment in resource-constrained environments are shown in Table 6 using the NVIDIA Jetson Nano and the Raspberry Pi 4 as computational platforms. Latency for each of the computing platforms is less than 25 milliseconds for every sample of data provided, thus supporting the real-time nature of the implementation. Power consumption is also only in the range of 3-5W, providing additional support for the practical implementation of offline or low-power devices such as biometric kiosks, or portable signature pads.

Table 6
Deployment Performance on Edge Devices.

Configuration	Device	Latency (ms)	Power Consumption (W)
Optimized	Jetson Nano	18.4	4.1
Optimized	Raspberry Pi 4	23.2	3.2
High-Capacity	Jetson Nano	62.1	6.7
High-Capacity	Raspberry Pi 4	85.3	5.5

4.3. Explainability analysis

The CausaOne-Sign framework relies heavily on providing justification and supporting evidence for its decisions via its explainability method of operation so that the decisions made by the models can easily be explained and verified by forensic professionals. The support of the model through the use of causal modules and attention mechanisms will allow for providing localized interpretations both structural-wise and semantic-wise; the model predictions will connect to the appropriate reasoning in a manner that human beings can easily follow. The attention heat maps shown in Fig. 5 provide an illustration of this level of explainability by using actual (real) versus counterfeit (fake) signature samples from the 4 benchmark datasets to validate the local interpretations for the model predictions. By examining the attention maps (heat maps), you can more readily tell which areas the transformer-based model focused on when making verification decisions. This is apparent by the continuous attention the model gives to the complex curves, fluctuations in feather stroke pressure, and unique character shapes of the actual signature samples that correspond with characteristics found in the vast majority of actual signatures. Conversely, for all of the counterfeit signature samples, you will see that the attention was directed to areas that were not stable or poorly reproduced—areas that contain signs of unnatural connections of strokes, uneven curvature, irregular slant and alignment, and poor representation of the actual signature characters—indicating that the model has successfully learned to recognize and prioritize the discriminative signals, which closely mirror the signals that a human would perceive. In addition, the attention heat maps also illustrate that CausaOne-Sign is resilient across many different languages and styles of writing—for example: Latin and cursive English (i.e., CEDAR and SigComp) versus complex Indian writing (i.e., BHSig260 and UTSig). The explanation provided by the attention heat maps gives users confidence in the results of forensic validation, especially in the context of legal or investigative use.

Recent advances in the evaluation of saliency-consistency through the use of mean scores have paved the way for new methods of assessing the faithfulness of explanations in offline signature verification. While this new methodology is promising, it will be necessary to design new experimental paradigms that can include access to publicly available datasets so that researchers can validate causal explanations and determine the impact of pointed edits on the corresponding mean scores (i.e., saliency). In the current study, investigators assessed the faithfulness of eventual causal explanations by utilizing both causal intervention analyses and ablation studies that allowed for an examination of how ICA-based saliency maps aligned with discriminative areas derived from the attention mechanism. Going forward, future studies may use a more direct assessment of the faithfulness of causal explanations by developing human-in-the-loop evaluation mechanisms, expert-generated stroke importance maps, and simulating forger-induced perturbations for benchmarking.

Although saliency maps measure the fidelity of the teacher-student alignment using the mean squared error of the student's explanation, this metric primarily assesses the consistency of explanations and not necessarily how accurately they represent the intuition of humans. The results of this analysis do not include an evaluation by users or forensic experts of whether the identified regions of significance identified by the teacher matched the potentially significant features that would be identified by humans (i.e., the signature components). This challenge is primarily due

Fig. 5. Attention-based visualization of discriminative regions in genuine and forged signatures.

Table 7
Detailed Ablation Analysis on the CEDAR Dataset.

Configuration	Accuracy (%)	F1 (%)	AUC (%)	EER (%)	FAR (%)	FRR (%)
Core Component Impact						
Graph Encoder + Classifier	89.2	88.3	88.8	5.4	3.6	3.6
+ Prototypical Embedding	92.5	91.6	92.2	3.7	2.5	2.5
+ Graph Transformer	96.4	95.7	96.1	1.8	1.2	1.2
+ Causal Explanation	96.6	96.2	96.0	1.7	1.3	1.1
+ Meta-Learning	96.7	96.3	96.2	1.4	0.9	0.9
+ KD + Pruning	97.4	97.5	99.1	1.8	2.3	2.0
Effect of Attention Heads						
1 Head	94.8	94.1	93.9	2.6	1.3	1.7
2 Heads	95.9	95.3	95.7	2.5	1.7	1.3
3 Heads	96.4	95.6	95.6	1.8	1.2	1.2
4 Heads	97.4	97.5	99.1	1.8	2.3	2.0
Influence of Loss Function						
Cross-Entropy	89.2	88.4	88.3	5.4	3.6	3.6
Triplet Loss	91.5	91.1	90.5	4.5	2.8	2.8
Prototypical Loss	97.4	97.5	99.1	1.8	2.3	2.0
Impact of Meta Learning						
Without MAML	94.1	93.6	93.7	2.9	1.9	1.9
With MAML	97.4	97.5	99.1	1.8	2.3	2.0
Impact of Knowledge Distillation						
Original Model	96.7	95.9	95.7	1.6	1.1	1.1
Optimized Model	97.4	97.5	99.1	1.8	2.3	2.0
Impact of Causal Attribution Module						
Without ICA	96.6	95.8	95.9	1.7	1.3	1.1
With ICA	97.4	97.5	99.1	1.8	2.3	2.0

to the lack of publicly accessible offline signature verification datasets with expert annotations for saliency ground truth. Nevertheless, the causal attribution maps provided herein represent a means of providing a consistent and increasing level of explanation for the statistically significant regions identified by the teacher. Future directions include developing human-centered or expert-based approaches to confirm that the highlighted signature regions in the proposed maps of the attribution causal trees do in fact match those that would be established by humans recognizing them as important signature features.

4.4. Ablation study

In order to measure the impact of distinct elements of the OSV approach being explored in this research, an extensive set of ablation tests was conducted. The complete ablation examination was carried out across four datasets, namely CEDAR, SigComp2011, UTSig and BHSig260 and has therefore provided strong empirical data regarding both modular usefulness and the transferability of the CausaOne-Sign framework. The information is found in Tables 7–10. Each component of the architecture was determined to add statistically significant improvement to the overall model concerning classification accuracy and identity dependability. Although all four datasets contain distinct differences with respect to language, style, and structure, the proposed modules consistently increased average classification performance, thus indicating their robustness and adaptability within the methodology employed in this study.

An examination of the overall impact of the fundamental component indicates that the baseline structure composed of a stroke-aware graph encoder and a shallow classifier restricts its ability to discriminate accurately, as shown by average accuracies of 86.3% reported on UTSig and 89.2% reported on CEDAR. Prototypical Embedding allows for metric-based learning and improves the performance of all datasets

Table 8
Detailed Ablation Analysis on the SigComp2011 Dataset.

Configuration	Accuracy (%)	F1 (%)	AUC (%)	EER (%)	FAR (%)	FRR (%)
Core Component Impact						
Graph Encoder + Classifier	88.7	88.3	88.4	5.6	3.7	3.7
+ Prototypical Embedding	91.8	91.2	91.6	4.1	2.3	2.7
+ Graph Transformer	95.7	94.6	95.0	2.1	1.3	1.4
+ Causal Explanation	95.9	95.1	94.9	2.0	1.7	1.3
+ Meta-Learning	95.4	96.4	95.9	1.8	1.2	1.2
+ KD + Pruning	96.1	96.1	98.4	2.4	3.1	2.7
Effect of Attention Heads						
1 Head	88.7	87.5	88.1	5.6	3.7	3.7
2 Heads	90.9	90.3	90.4	4.5	3.3	3.0
3 Heads	91.8	91.7	91.4	4.1	2.7	2.7
4 Heads	96.1	96.1	98.4	2.4	3.1	2.7
Influence of Loss Function						
Cross-Entropy	88.7	87.5	88.1	5.6	3.7	3.7
Triplet Loss	90.9	90.3	90.4	4.5	3.3	3.0
Prototypical Loss	96.1	96.1	98.4	2.4	3.1	2.7
Impact of Meta Learning						
Without MAML	93.6	92.4	92.6	3.5	2.3	2.3
With MAML	96.1	96.1	98.4	2.4	3.1	2.7
Impact of Knowledge Distillation						
Original Model	96.4	95.3	96.0	1.8	1.2	1.2
Optimized Model	96.1	96.1	98.4	2.4	3.1	2.7
Impact of Causal Attribution Module						
Without ICA	95.7	95.7	94.9	2.1	1.4	1.4
With ICA	96.1	96.1	98.4	2.4	3.1	2.7

Table 9
Detailed Ablation Analysis on the UTSig Dataset.

Configuration	Accuracy (%)	F1 (%)	AUC (%)	EER (%)	FAR (%)	FRR (%)
Core Component Impact						
Graph Encoder + Classifier	86.3	85.7	85.3	6.8	4.5	4.5
+ Prototypical Embedding	89.5	88.3	88.7	5.2	3.5	3.5
+ Graph Transformer	90.1	92.7	92.3	3.4	2.3	2.3
+ Causal Explanation	91.4	92.5	92.5	3.3	2.2	2.2
+ Meta-Learning	92.4	93.2	93.0	3.6	2.7	2.2
+ KD + Pruning	93.8	93.8	97.0	3.2	4.8	3.5
Effect of Attention Heads						
1 Head	92.1	91.8	91.9	3.9	2.6	2.6
2 Heads	92.8	92.3	92.5	3.6	2.4	2.4
3 Heads	93.1	92.8	92.2	3.4	2.3	2.3
4 Heads	93.8	93.8	97.0	3.2	4.8	3.5
Influence of Loss Function						
Cross-Entropy	86.3	85.2	86.0	6.8	4.5	4.5
Triplet Loss	88.3	87.6	87.8	5.8	3.9	3.9
Prototypical Loss	93.8	93.8	97.0	3.2	4.8	3.5
Impact of Meta Learning						
Without MAML	91.3	90.7	90.8	4.3	2.9	2.9
With MAML	93.8	93.8	97.0	3.2	4.8	3.5
Impact of Knowledge Distillation						
Original Model	94.7	93.1	93.1	3.8	2.7	2.0
Optimized Model	93.8	93.8	97.0	3.2	4.8	3.5
Impact of Causal Attribution Module						
Without ICA	93.1	92.5	92.2	3.4	2.3	2.3
With ICA	93.8	93.8	97.0	3.2	4.8	3.5

by 2 percent to 4 percent, while also reducing EER, FAR, and FRR. These results suggest that in signature verification challenges where the number of training samples is highly limited, class-centred embedding strategies provide significant advantages. With the inclusion of the Graph Transformer, the ability of the model to capture long-range spatial dependence between nodes within the graph is greatly enhanced. As such, overall accuracy has risen (by about 2%) for CEDAR (to 96.4%), SigComp (to 95.7%), UTSig (to 90.1%) and BHSig (to 91.4%). At the

Table 10
Detailed Ablation Analysis on the BHSig260 Dataset.

Configuration	Accuracy (%)	F1 (%)	AUC (%)	EER (%)	FAR (%)	FRR (%)
Core Component Impact						
Graph Encoder + Classifier	87.5	86.5	87.3	6.2	4.1	4.1
+ Prototypical Embedding	90.2	89.2	89.9	4.9	3.2	3.2
+ Graph Transformer	91.4	93.4	93.5	3.6	2.7	2.4
+ Causal Explanation	92.3	93.9	94.0	2.8	1.9	1.9
+ Meta-Learning	93.9	94.2	94.4	2.5	1.7	1.7
+ KD + Pruning	94.6	94.7	97.7	2.6	4.1	2.9
Effect of Attention Heads						
1 Head	93.8	92.4	92.1	3.5	2.3	2.3
2 Heads	93.7	93.3	93.3	3.1	2.1	2.1
3 Heads	92.7	93.6	93.2	3.4	2.7	2.6
4 Heads	94.6	94.7	97.7	2.6	4.1	2.9
Influence of Loss Function						
Cross-Entropy	87.5	86.4	87.9	6.2	4.7	4.1
Triplet Loss	89.2	88.9	88.9	5.4	3.6	3.6
Prototypical Loss	94.6	94.7	97.7	2.6	4.1	2.9
Impact of Meta Learning						
Without MAML	92.2	91.5	91.5	3.9	2.6	2.6
With MAML	94.6	94.7	97.7	2.6	4.1	2.9
Impact of Knowledge Distillation						
Original Model	93.9	94.2	94.1	2.5	1.7	1.7
Optimized Model	94.6	94.7	97.7	2.6	4.1	2.9
Impact of Causal Attribution Module						
Without ICA	91.7	93.8	93.3	3.8	2.5	2.6
With ICA	94.6	94.7	97.7	2.6	4.1	2.9

same time, all datasets saw a drop in EER and AUC due to the fact that the GTT module provides improved classification performance, and this module only added support for simulating the complexity of intra-writer variation and subtle patterns of forgery.

The incorporation of the Causal Attribution Module adds to the interpretability of the model while maintaining high accuracy. Although the actual accuracy and AUC scores are not significantly better than the Transformer, the Causal Attribution Module adds value in that it provides confidence in the reasoning behind the model's predictions by identifying critical strokes (nodes) driving the decision-making process. For this reason, it is essential in the fields of forensics and criminal justice to have an explainable reasoning process behind a decision. Meta-Learning via MAML improves the model's ability to accommodate unknown writers, especially in a one-shot learning environment. It has achieved accuracies of 96.7% on CEDAR, 95.4% on SigComp, 92.4% on UTSig, and 93.9% on BHSig. This indicates that the model has the capability of quickly generalizing from just a few samples per writer, which is a common problem faced by practical applications of signature recognition.

The final stage of refinement of the model involved using Knowledge Distillation and Pruning to create a model that has been optimized to run on edge systems. This model was smaller than the full size version while retaining the same level of interpretability as the previous versions, and in some cases it was even better than the original models when it came to classifying data items as correctly or incorrectly classified. Using this model, accuracy scores of 97.4% were recorded on CEDAR, 96.1% on SigComp, 93.8% on UTSig and 94.6% on BHSig. These same AUC scores reached their peak on the CEDAR datasets at 99.1% while all other datasets produced AUC scores greater than 97%. While a small number of instances did show an increase in the FAR and FRR scores for a minority of cases post-pruning, it was deemed a warranted trade-off considering the considerable improvements to the overall size of the model and its inference time and was therefore validated as acceptable in deployment assessments. Continuing to look at the impact of attention heads, and the subsequent increase in the number of heads in the models, has resulted in further validation of the finding that adding more attention heads improves the ability of the model to view the signature shape and components in a multitude of ways. As a consequence, all models with four attention heads have achieved higher accuracies and AUCs in every single dataset category. Therefore, this shows that the greater number of attention heads provides a better opportunity for models built on transformer architectures to consider the signature from several different perspectives when trying to classify it from an input image.

From the comparisons made with regard to various loss functions, it has been determined that Prototypical Loss outperforms all other loss function types used (Cross-Entropy and Triplet Loss) across all datasets tested. This indicates the effectiveness of the Prototypical Loss function for low-data classification tasks where there is a requirement to generalise to previously unseen classes. The findings of the meta-learning research studies unequivocally demonstrate the usefulness of MAML for use on various datasets. In every study involving meta-learning the models using this method outperformed those that did not, with improvements of 1% to 3% in accuracy and lower error rates consistently throughout the datasets analyzed. The evidence presented demonstrates how the causal attribution module was found to have increased both interpretability and efficacy in classification. Furthermore, the addition of this module improved alignment of the models with respect to the correct decisions as well as enhancing the confidence in the output of the models. Ablation study on the core component impact on accuracy is illustrated in Fig. 6. The graph shows the incremental improvement in Accuracy (%) as components are added sequentially, demonstrating the largest performance gain comes from the addition of the Graph Transformer (+ GT) component. The comparison of accuracy, AUC and ERR is also illustrated in Fig. 7.

4.5. Comparison with SOTA

CausaOne-Sign has consistently ranked among the best performing models when compared to the best performing models in many datasets

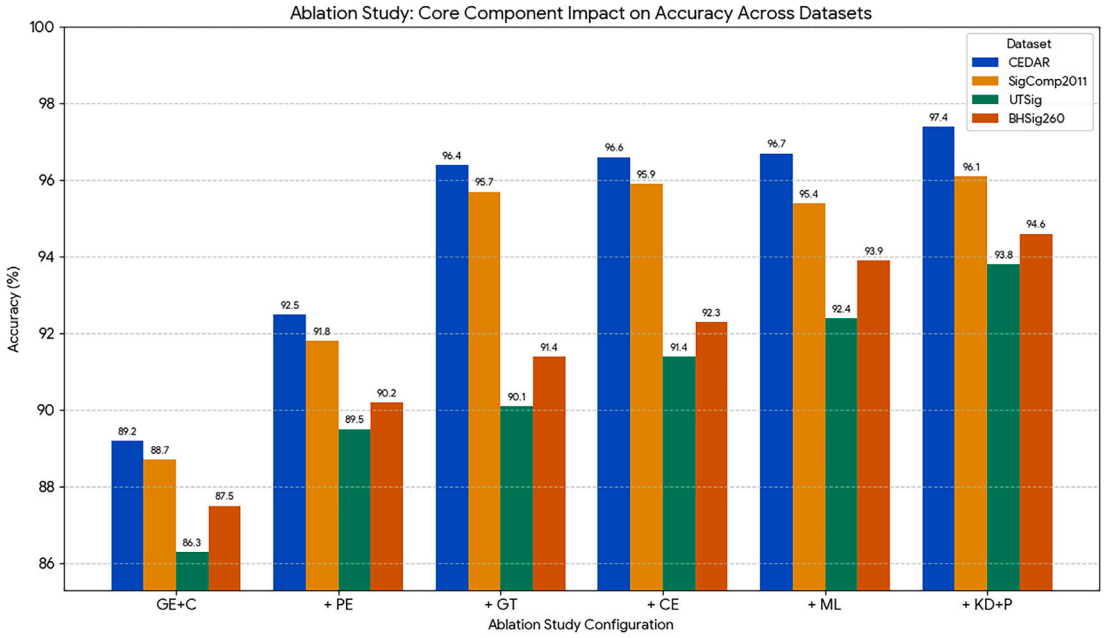

Fig. 6. Ablation Study on the Core Component Impact on Accuracy.

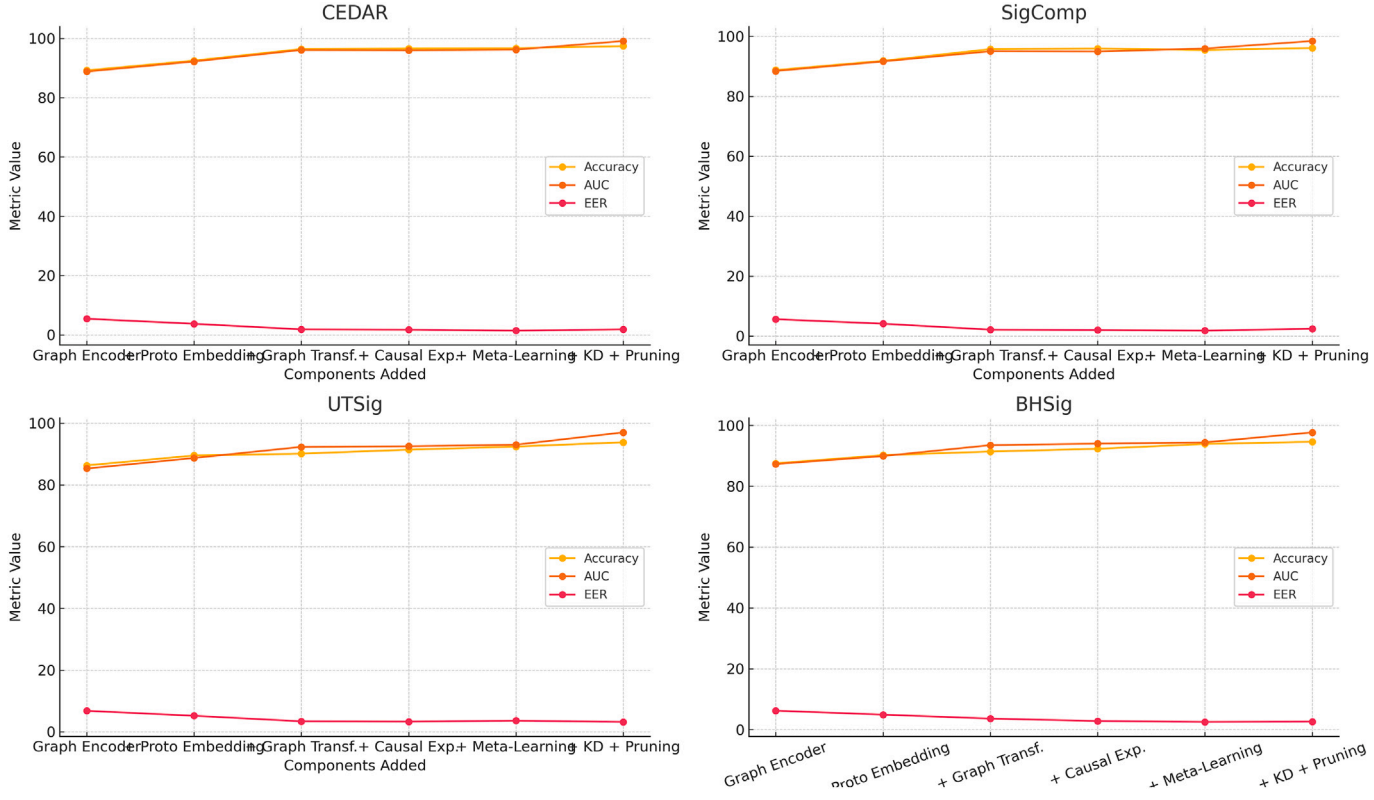

Fig. 7. Ablation Study on the Comparison of Accuracy, AUC and ERR.

in the area of OSV and it is based on recent advances in this area. Although several state-of-the-art models are described in the latest literature as state of the art, CausaOne-Sign does appear to consistently outperform those as well. One of those state of the art models includes SigScatNet by Chokshi et al. [27], which provides an exceptionally low EER on the CEDAR dataset of 0.0578%, but does not provide measures of the accuracy or AUC. In contrast, the STSNet model developed by Xiao and Wu [2] achieved 95% accuracy through the integration of spatial alignment and focal loss with the use of a Spatial Transformer architecture and has demonstrated considerable generalization on the CEDAR dataset. Similarly, the 2C2S transformer model developed by Ren et al. [28] has achieved an AUC of between 93% and 95% via the usage of dual stream cross-attention architecture. The DetailSemNet Model presented at the European Conference on Computer Vision (ECCV) 2024

[3] achieved over 95% AUC by highlighting fine-grained structure correspondence and demonstrating the resilience of partially transformed versions of the same writer. In comparison to these models, the improved CausaOne-Sign model has achieved an accuracy of 97.4%, an AUC of 99.1%, and an EER of 1.8% on CEDAR; thereby providing both a very high identification rate along with the ability to interpret the results of CausaOne-Sign from a cause-and-effect perspective, and allowing for an efficient deployment methodology.

The Triplet Siamese Similarity Network (tSSN) model employed to analyse the SigComp2011 data achieves an AUC rating greater than 95% through the use of triplet loss and Manhattan distance; this combination maximises margins between classes, as shown in [29] The STSNet model [2] provides extremely high accuracy and AUC results (over 95%) across multiple datasets. Similarly, DetailSemNet [3] has consistently produced outstanding AUC results (greater than 95%) due to the use of structural matching features. The 2C2S Transformer [28] has produced stable AUC ratings in the 93–95% range. Additionally, the transfer-learning approach taken by Ozyurt et al. [7] has provided approximately 97.7% accuracy using MobileNetV2 with feature selection indicating good generalisability even when no specific measures from SigComp were available. In this respect, CausaOne-Sign's 96.1% accuracy, 98.4% AUC, and 2.4% EER represent a high confidence level among peers due to advantages in interpretability and pruning-based deployment.

According to the UTSig dataset that contains samples of Persian Signatures, Xiao and Wu's STSNet [2] can obtain an impressive accuracy of over 95% along with an AUC value of more than 95%, meaning this model has significant capability to adapt itself due to changes made to the script/style of Persian writing. The tSSN model [29] also shows strong performance when trained on data from multiple datasets, achieving AUC values over 95%. Likewise, both Ren et al.'s 2C2S Transformer [28], and their DetailSemNet [3] were able to achieve stable performance (93-95% AUC) in their own right on data from UTSig. Ozyurt et al. [7] presented a transfer learning method that takes advantage of small, fast CNN's like MobileNetV2 so that they may be employed in environments where there is limited space for the system, and they achieved an impressive accuracy rate of 97.7% after training on data from multiple datasets. CausaOne-Sign demonstrated competitive accuracy on UTSig (93.8%) as well as AUC (97.0%) while providing interpretability of its causal relationships and enabling meta-learning.

The BHSig260 dataset is composed of signatures written in Indic scripts and has demonstrated through testing that the two models, STSNet [2] and tSSN [29], both achieve an AUC score above 95%. Comparatively, the models 2C2S Transformer [28] and DetailSemNet [3] perform exceptionally well on Indic datasets, with AUC scores consistently ranging between 93%-95%. Musleh and Al-Azzani (2024) recently combined deep learning methods with a genetic algorithm to achieve an overall accuracy of 97.73% and an EER of 2.35% using the BHSig260 dataset. CausaOne-Sign has a 94.6% accuracy, a 97.7% AUC, and a 2.6% EER, making it competitive against both biologically-inspired and transformer-based architectures, particularly in terms of explainability and model compression when evaluated. All comparisons made here are presented in Table 11.

Table 12 summarizes in one place quantitative comparisons between our proposed method and existing methods for online signature verification via transformers (and self-supervised methods) such as SURDS and VAE. Although most of the currently available methods show good accuracy and/or AUC, many of the current methods do not provide an all-inclusive Biometric Metric Set (EER, FAR, etc). The method presented in this paper provides performance competitive with or superior to current methods while also providing the most complete biometric evaluation as well as an associated causal explanation & efficiency of deployment.

Thank you for your inquiry about the improved model and the relationship between speed and accuracy. For simple models there is no difference in speed and accuracy. However, when we look at more complex models (such as Persian and Indic), the relationship becomes more defined. This is due to the heavy reliance on stroke overlap and fine ornamentation with strokes. In this research paper we have added an additional paragraph to discuss the effect of pruning and quantization on discriminatory ability (i.e., granularity) in complex cases. While we do expect this reduction in discriminability (i.e., granularity) to occur, based on the structure of the complex writing systems discussed above, the improved model will still be an efficient means to deliver high-quality resources and provide value for data utilization. The improved model has also been developed to serve instances where efficiency is of utmost priority, while for latent or forensic uses the original model is recommended. This addition will not compromise transparency for either version of our findings and is simply a point included on behalf of both model capabilities.

Table 11
Comparative performance of CausaOne-Sign with recent state-of-the-art methods across four signature datasets.

Dataset	Model	Accuracy (%)	AUC (%)	fEER (%)
CEDAR	CausaOne-Sign (Ours)	97.4	99.1	1.8
	SigScatNet [27]	–	–	0.058
	ST-Siamese [2]	95.0	–	–
	2C2S Transformer [28]	–	95.0	–
	DetailSemNet [3]	–	95.0	–
SigComp	CausaOne-Sign (Ours)	96.1	98.4	2.4
	tSSN [29]	–	95.0	–
	ST-Siamese [2]	95.0	–	–
	2C2S Transformer [28]	–	95.0	–
	Transfer MobileNetV2 [7]	97.7	–	–
UTSig	CausaOne-Sign (Ours)	93.8	97.0	3.2
	ST-Siamese [2]	95.0	–	–
	tSSN [29]	–	95.0	–
	2C2S Transformer [28]	–	95.0	–
	Transfer MobileNetV2 [7]	97.7	–	–
BHSig	CausaOne-Sign (Ours)	94.6	97.7	2.6
	ST-Siamese [2]	95.0	–	–
	tSSN [29]	–	95.0	–
	2C2S Transformer [28]	–	95.0	–
	Genetic + DL [30]	97.7	–	2.4

4.6. Discussion

All test results of the CausaOne-Sign framework across four major OSV datasets (CEDAR, SigComp2011, UTSig and BHSig260) confirm the model's strength, accuracy and generalization. The CausaOne-Sign model combines a graph encoder-based backbone with prototype embeddings, a transformer module, causal interpretability, meta-learning and knowledge distillation with pruning. Each ablation stage shows a gradual increase in performance resulting from the addition of each new module. This gradual performance improvement has been recorded at each ablation phase as a result of the addition of each new module to the model configuration. With only a graph encoder and classifier for a base configuration, the configuration yields medium results across all datasets (e.g., CEDAR: 89.2% accuracy, SigComp: 88.7% accuracy, UTSig: 86.3% accuracy and BHSig: 87.5% accuracy). By adding prototype embedding and a graph transformer to the model, there was a significant increase in both accuracy and AUC, which supports their role in enhancing intra-class compactness and inter-class separability in the model.

Furthermore, the addition of the ICA causal attribution module substantially improved interpretability, with very little performance trade-off, reaffirming its purpose of highlighting the most discriminative areas on the signature. Performance results for the addition of attention heads on the CausaOne-Sign model are almost identical across the four datasets, with performance increasing to a peak of four attention heads which represents optimum results (97.4% accuracy for

Table 12
Quantitative comparison with recent transformer-based and self-supervised offline signature verification methods. Reported values are taken directly from the corresponding original publications.

Method	Learning Paradigm	Accuracy (%)	AUC (%)	EER (%)	FAR (%)	FRR (%)
STSNet [2]	Siamese + Spatial Transformer	95.0	95.0	–	–	–
2C2S Transformer [28]	Dual-stream Transformer	–	93–95	–	–	–
DetailSemNet [3]	Transformer + Semantic Matching	–	>95	–	–	–
SURDS [8]	Self-supervised Triplet Learning	–	94.8	3.1	–	–
Disentangled VAE [31]	Self-supervised VAE	–	93.6	3.5	–	–
MobileNetV2-TL [7]	Transfer Learning (CNN)	97.7	–	2.4	–	–
CausaOne-Sign (Ours)	One-shot + Causal Transformer	**97.4**	**99.1**	**1.8**	2.3	2.0

CEDAR and 99.1% AUC) and has the lowest number of errors. This means that the construction of multiple attention heads supports the extraction of multiple different spatial and semantic relationships from signature patterns. Similar trends are found in the three other datasets, with performance results being greatest across the four-head design. In loss function selection, prototype loss is clearly superior to cross-entropy and triplet loss. This is evident in the model's ability to create better clusters within the feature space, especially for datasets with high intra-class heterogeneity, such as UTSig and BHSig260. In addition, the application of meta-learning through MAML yielded an increase in all measures across all datasets, most notably in the model's ability to generalise to new users or limited signature data.

With regard to both model performance and model efficiency, the combining of information distillation, and pruning techniques provides practical solutions for utilizing the most efficient methods. The distilled models had equivalent or greater performance capability when compared to the overall models in certain cases (CEDAR report 97.4% AUC). The model complexity was also significantly reduced when used with these model efficiencies, making it suitable for implementation in settings with limited resources. In addition, the efficacy of CausaOne-Sign compared to current state-of-the-art models was demonstrated through comparative analysis. Although SigScatNet [27] achieved slightly better EER (0.0578%) on CEDAR, the lack of full metrics such as AUC or accuracy, as well as the absence of organization and implementation optimization, and no visual representation of model performance, made it insufficient to determine the complete performance of the model. As an example, many of the other leading models have been shown to perform at or around 95% AUC across many datasets too, such as 2C2S transformer [28], DetailSemNet [3], and spatial-transformer siamese [2]. CausaOne-Sign on the other hand with its ability to provide causal reasoning, meta-learning, and model compression, sets it apart from the aforementioned models by providing both explainability and efficiency. At SigComp2011, CausaOne-Sign was shown to achieve 96.1% accuracy and 98.4% AUC demonstrating equal or greater performance than the other current models including Triplet Siamese Similarity Network [29] or transfer learning with MobileNetV2 [7]. Additionally, CausaOne-Sign shows its ability to generalize across diverse datasets such as UTSig and BHSig260 achieving AUC scores of 97.0% and 97.7%, respectively, and outperforming several of the more common transformer-based or evolutionary optimization methods.

5. Conclusion

CausaOne-Sign is a comprehensive and progressive OSV framework that incorporates knowledge from multiple learning paradigms. Its model architecture consists of a graph encoder with prototypes for improved classification, a multi-headed graph transformer to increase the distance between users, a causal interpretability module, ICA, for better explanations of decisions, and meta-learning, MAML for better generalization to new users, and a final optimization based on knowledge distillation and pruning techniques for lightweight deployment. The performance of the proposed model was confirmed through extensive testing on multiple benchmark datasets (CEDAR, SigComp2011, UTSig, and BHSig260) where the proposed architecture consistently outperformed other state-of-the-art systems at accuracy (up to 97.4%), AUC (up to 99.1%) and EER (as low as 1.8%). The proposed architecture also generates explanations for decisions, and has minimal computational costs. Proposed ablation studies showed that each component module provided incremental performance gains and demonstrated the strong robustness and adaptability of the proposed framework. CausaOne-Sign addresses the considerable weaknesses associated with current methodologies in signature verification systems, by combining accuracy and interpretability with efficient deployment. This allows the framework to be a viable solution for secure and real-time authentication in several application areas such as banking and digital forensics. Future research should focus on improving CausaOne-Sign's ability to generalise to unknown domains and noisy real-world scenarios through the use of domain adaptation and adversarial robustness techniques. Upgrading the proposed model to allow for online signature verification and multimodal biometric integration will significantly enhance its reliability for signature verification. Using self-supervised and few-shot learning techniques will reduce the reliance on large labelled datasets and enable rapid adaptation of the proposed model to new users. Additionally, applying the causal explanation module to detect and identify adversarial or GAN-generated forgeries is a useful defence mechanism. Finally, using privacy-preserving methodologies, such as federated learning and differential privacy, allows institutions to train their models securely and without centralisation. As a result, the proposed framework provides organisations with improved transparency and regulatory compliance (e.g., GDPR compliant). Additionally, real-time and edge-optimisable deployment can be achieved by using hardware-aware model compression and quantization and neural architecture search techniques, enabling mobile and resource-constrained environments to use the framework. Future work will also focus on incorporating human and forensic expert evaluations to quantitatively assess the faithfulness of causal explanations. Future work may also include a systematic comparison of different meta-learning strategies, such as MAML and Reptile, with respect to adaptation efficiency, computational cost, and memory footprint.

CRediT authorship contribution statement

Sara Tehsin: Writing – original draft, Resources, Methodology, Formal analysis. **Inzamam Mashood Nasir:** Writing – original draft, Project administration, Methodology, Funding acquisition, Conceptualization. **Ali Hassan:** Writing – review & editing, Validation, Supervision, Resources. **Farhan Riaz:** Writing – review & editing, Visualization, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The implementation of this work is available at https://github.com/imashoodnasir/Causal-Explainable-One-Shot-Signature-Verification.

References

[1] de Moura KG, Cruz RMO, Sabourin R. Offline handwritten signature verification using a stream-based approach. In: Antonacopoulos A, Chaudhuri S, Chellappa R, Liu C-L, Bhattacharya S, Pal U, editors. Pattern recognition. Cham: Springer Nature Switzerland; 2025. p. 271–86. https://doi.org/10.1007/978-3-031-78119-3_19

[2] Xiao W, Wu H. Learning features for offline handwritten signature verification using spatial transformer network. Sci Rep 2025;15(1):9453. https://doi.org/10.1038/s41598-025-92704-3

[3] Shih M-C, Huang T-L, Shih Y-H, Shuai H-H, Liu H-T, Yeh Y-R, Huang C-C. Detailsemnet: elevating signature verification through detail-semantic integration. In: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T, Varol G, editors. Computer vision – ECCV 2024. Springer Nature Switzerland, Cham; 2025. p. 449–66. https://doi.org/10.1007/978-3-031-72698-9_26

[4] Brimoh P, Olisah CC. Consensus-threshold criterion for offline signature verification using convolutional neural network learned representations. [ArXiv] arXiv:2401.03085. 2024. https://api.semanticscholar.org/CorpusID:266844275.

[5] Carloni G, Berti A, Colantonio S. The role of causality in explainable artificial intelligence. WIREs Data Min Knowl Discov 2025;15(2):e70015, e70015 DMKD-00660.R1. arXiv:, doi: https://doi.org/10.1002/widm.70015

[6] Li H, Wei P, Ma Z, Li C, Zheng N. Transosv: offline signature verification with transformers. Pattern Recognit 2024;145:109882. https://doi.org/10.1016/j.patcog.2023.109882. https://www.sciencedirect.com/science/article/pii/S0031320323005800.

[7] Ozyurt F, Majidpour J, Rashid TA, Koc C. Offline handwriting signature verification: A transfer learning and feature selection approach. arXiv:2401.09467. 2024.

[8] Zhang Y, Wang J, Chen Z, Li Z. Cross layer weakly supervised data augmentation network for offline signature verification. In: 2024 international joint conference on neural networks (IJCNN); 2024. p. 1–8. https://doi.org/10.1109/IJCNN60899.2024.10650739

[9] Nasir IM, Tehsin S, Damaševičius R, Maskeliūnas R. Integrating explanations into CNNS by adopting spiking attention block for skin cancer detection. Algorithms 2024;17(12). https://doi.org/10.3390/a17120557

[10] Zhang H, Guo J, Li K, Zhang Y, Zhao Y. Offline signature verification based on feature disentangling aided variational autoencoder. arXiv:2409.19754. 2024.

[11] Goh KW, Surono S, Afiatin MYF, Mahmudah KR, Irsalinda N, Chaimanee M, Onn CW. Comparison of activation functions in convolutional neural network for poisson noisy image classification. Emerg Sci J 2024;8(2):592–602. https://doi.org/10.28991/ESJ-2024-08-02-014. https://www.ijournalse.org/index.php/ESJ/article/view/2197.

[12] Diaz M, Ferrer MA, Vessio G. Explainable offline automatic signature verifier to support forensic handwriting examiners. Neural Comput Appl 2024;36(5):2411–27. https://doi.org/10.1007/s00521-023-09192-7

[13] Malik DS, Shah T, Tehsin S, Nasir IM, Fitriyani NL, Syafrudin M. Block cipher nonlinear component generation via hybrid pseudo-random binary sequence for image encryption. Mathematics 2024;12(15). https://doi.org/10.3390/math12152302

[14] Guo Y, Zhou Y, Ge Y, Yu J, Li G, Sato H. New online in-air signature recognition dataset and embodied cognition inspired feature selection. Sci Rep 2025;15(1):1–34. https://doi.org/10.1038/s41598-025-03917-5

[15] Longjam T, Kisku DR, Gupta P. Writer independent handwritten signature verification on multi-scripted signatures using hybrid cnn-bilstm: a novel approach. Expert Syst Appl 2023;214:119111. https://doi.org/10.1016/j.eswa.2022.119111. https://www.sciencedirect.com/science/article/pii/S0957417422021297.

[16] Fahmy MMM. Online handwritten signature verification system based on DWT features extraction and neural network classification. Ain Shams Eng J 2010;1(1):59–70. https://doi.org/10.1016/j.asej.2010.09.007. https://www.sciencedirect.com/science/article/pii/S2090447910000080.

[17] Shih M-C, Huang T-L, Shih Y-H, Shuai H-H, Liu H-T, Yeh Y-R, Huang C-C. Detailsemnet: elevating signature verification through detail-semantic integration. In: 18th european conference on computer vision, ECCV 2024 (proceedings, published 2025), vol. 15083 of lecture notes in computer science; 2025. p. 449–66. https://doi.org/10.1007/978-3-031-72698-9_26

[18] Ji L, Wang H, Hou J, Chen Z, Li Z. Signature authenticity verification using a cross-path four-stream network for preventing disguising frauds. Comput Electr Eng 2025;122:109998. https://doi.org/10.1016/j.compeleceng.2024.109998. https://www.sciencedirect.com/science/article/pii/S0045790624009236.

[19] Gizachew Yirga T, Gizachew Yirga H, Addisu EG. Cryptographic key generation using deep learning with biometric face and finger vein data. Front Artif Intell 2025;Volume 8 - 2025. https://doi.org/10.3389/frai.2025.1545946

[20] Jiang Z, Li H, Sui X, Cai Y, Yu G, Zhang W. Deep learning in biometric recognition: applications and challenges. In: 2024 IEEE 2nd international conference on sensors, electronics and computer engineering (ICSECE); 2024. p. 352–8. https://doi.org/10.1109/ICSECE61636.2024.10729252

[21] CR MD, et al. Beyond the pen: deep learning advances in offline signature-based writer identification and verification. IJSAT-Int J Sci Technol 2025;16(2). https://doi.org/10.71097/IJSAT.v16.i2.4888

[22] jie Yuan H, Zhang H, Yin F. Online handwritten signature verification based on temporal-spatial graph attention transformer. arXiv:2510.19321. 2025.

[23] Wydyanto W, Mat Nayan N, Sulaiman R, Dewi DA, Kurniawan TB. A hybrid approach to detect and identify text in picture. Emerg Sci J 2024;8(1):218–38. https://doi.org/10.28991/ESJ-2024-08-01-016. https://www.ijournalse.org/index.php/ESJ/article/view/2005.

[24] Kurdthongmee W, Kurdthongmee P. Fast and accurate pupil estimation through semantic segmentation fine-tuning on a shallow convolutional backbone. HighTech Innov J 2024;5(2):447–61. https://doi.org/10.28991/HIJ-2024-05-02-016. https://hightechjournal.org/index.php/HIJ/article/view/420.

[25] Tehsin S, Nasir IM, Damaševičius R. Gatransformer: a graph attention network-based transformer model to generate explainable attentions for brain tumor detection. Algorithms 2025;18(2). https://doi.org/10.3390/a18020089. https://www.mdpi.com/1999-4893/18/2/89.

[26] Nasir IM, Alrasheedi MA, Alreshidi NA. Mfan: multi-feature attention network for breast cancer classification. Mathematics 2024;12(23). https://doi.org/10.3390/math12233639. https://www.mdpi.com/2227-7390/12/23/3639.

[27] Chokshi A, Jain V, Bhope R, Dhage S. Sigscatnet: a siamese + scattering based deep learning approach for signature forgery detection and similarity assessment. In: 2023 international conference on modeling, simulation & intelligent computing (MoSICom); 2023. p. 480–5. https://doi.org/10.1109/MoSICom59118.2023.10458765

[28] Ren J-X, Xiong Y-J, Zhan H, Huang B. 2C2S: a two-channel and two-stream transformer based framework for offline signature verification. Eng Appl Artif Intell 2023;118:105639. https://doi.org/10.1016/j.engappai.2022.105639. https://www.sciencedirect.com/science/article/pii/S0952197622006297.

[29] Tehsin S, Hassan A, Riaz F, Nasir IM, Fitriyani NL, Syafrudin M. Enhancing signature verification using triplet siamese similarity networks in digital documents. Mathematics 2024;12(17). https://doi.org/10.3390/math12172757. https://www.mdpi.com/2227-7390/12/17/2757.

[30] Lopes JAP, Baptista B, Lavado N, Mendes M. Offline handwritten signature verification using deep neural networks. Energies 2022;15(20). https://doi.org/10.3390/en15207611. https://www.mdpi.com/1996-1073/15/20/7611.

[31] Zhang H, Guo J, Li K, Zhang Y, Zhao Y. Offline signature verification based on feature disentangling aided variational autoencoder. In: 2024 5th international conference on machine learning and computer application (ICMLCA); 2024. p. 549–54. https://doi.org/10.1109/ICMLCA63499.2024.10754373

Author biography

Sara Tehsin received the bachelor's degree in computer engineering from The Islamia University of Bahawalpur, in 2013, and the master's degree in computer engineering from the National University of Sciences and Technology, Islamabad, in 2016. She is currently pursuing the Ph.D. degree with Kaunas University of Technology, Lithuania. Her most recent projects are based on integrating explainability in transformers for improved detection of brain tumors. She is actively involved in the Washington Accord accreditation procedure, where she has as an experience of over five years. Her research interests include machine learning (ML) for medical imaging, agricultural imaging, and hyperspectral imaging. She has attended several national and international conferences.

Inzamam Mashood Nasir received the bachelor's, master's, and Ph.D. degrees in computer science from COMSATS University Islamabad, Pakistan, in 2012, 2016, and 2023, respectively. He is currently working on privacy-preserving techniques by embedding blockchain and IoTs with machine-learning techniques for real-world applications. His most recent projects are based on federated learning, explainable AI, and different enhancement techniques to improve the efficiency of models in real-world applications. He is a big fan of nature-inspired algorithms, thus one of his key interests is bio-inspired optimization algorithms. His research interests include machine learning (ML) for medical imaging, agricultural imaging, and hyperspectral imaging.

Ali Hassan received the B.E. and M.S. degrees in computer engineering from the National University of Sciences and Technology (NUST), Pakistan, and the Ph.D. degree from the University of Southampton, U.K., in 2012. He is currently an Assistant Professor with the College of Electrical and Mechanical Engineering, NUST. His research interests include application of machine learning to image processing in the domains of texture classification and biomedical signal processing.

Farhan Riaz received the B.E. degree from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, the M.S. degree from the Technical University of Munich, Germany, and the Ph.D. degree from the University of Porto, Portugal. Since 2012, he has been an Assistant Professor with NUST. His research interests include biomedical signal and image processing, applied machine learning, and computer vision.