GMR
The Original

SECURE MULTIMODAL BIOMETRIC AUTHENTICATION USING DARTS-OPTIMIZED SWIN TRANSFORMERS AND DIFFUSION-ASSISTED CANCELABLE TEMPLATE PROTECTION

S. Jebapriya[1] and M. Ganaga Durga[2]

[1]Research Scholar, Department of Computer Science, Madurai Kamaraj University, Madurai, Tamilnadu, India, E-mail: jebapriyas.research@gmail.com
[2]Associate Professor, Department of Computer Applications, Sri Meenakshi Govt. Arts College for Women Autonomous), Madurai, Tamilnadu, India, E-mail: sivamgdurga@gmail.com

ABSTRACT:
Multimodal biometric authentication is offering an effective method to identity verification by combining matching qualities of face, iris and fingerprint modalities, in real-world scenarios, but still, reaching trustworthy authentication while protecting stored biometric templates remains a noticeable hurdle; this paper indicates a secure multimodal biometric authentication framework blending Swin Transformer-based feature extraction, Differentiable Architecture Search (DARTS), reliability-aware feature fusion, and diffusion-assisted latent representation learning, in many respects. Modality-specific Swin Transformer encoders are, as a practical matter, employed to pull discriminative features from biometric inputs; indeed, what's more, until then, the network architecture is improved, within a predefined search space by dARTS. Authentication is ran using Hamming-distance-based matching between enrolled and probe templates, and diffusion-assisted learning is worked in to support latent representation robustness, followed by a secret-dependent cancelable transformation to make binary biometric templates. The pulled representations are, broadly speaking, combined through the reliability-aware fusion to create a integrated feature representation, and the proposed framework is found using precision, recall, F-Measure, and precision and compared with BiLSTM, FOA and CNN, BiLSTM, FOA approaches. The experimental results show that the proposed framework hits 94.852% precision, 98.640% recall, 96.709% F-Measure, and 98.914% accuracy, outperforming the checked strategies across all four metrics; these findings are emphasizing the potential of combining transformer-based feature extraction, architecture optimization, and multimodal feature fusion to sharpen biometric authentication performance.

KEYWORDS: Multimodal biometric authentication, Swin Transformer, Differentiable Architecture Search, DARTS, diffusion-assisted learning, cancelable biometrics, biometric template protection, Hamming distance, in actual application.

1. INTRODUCTION

Biometric authentication has become an significant method to identity verification in applications that are needing trustworthy access control and personal identification, to a significant degree. As, well, side by side to standard password-based frameworks, biometric recognition depends on physiological or behavioural attributes such as facial traits, iris patterns, and fingerprints, broadly speaking. These qualities usefully are informing and decrease the dependence on credentials that can, to some extent, be forgotten, shared, or stolen, on the whole. However, biometric systems also introduce specific security and privacy challenges because biometric characteristics are inherently associated with individuals and cannot be replaced as easily as conventional passwords following a security breach. When the dust settles, the development of biometric authentication systems needs attention to recognition accuracy and to the protection of stored biometric information [1], broadly speaking.

Multimodal biometric authentication has emerged as an approach to raise the reliability of standard unimodal systems by combining information from two or more biometric traits, in many respects; matching identity-related attributes can be employed from separate sources by a system based on face, iris, and fingerprint, to a significant degree. But mixed feature representations are being generated by these modes and are impacted by different acquisition conditions; indeed, as, a finding, combining them needs an proper fusion strategy that preserves useful information while lowering the affect of unreliable samples. Ross and Jain [2] explored the function of information fusion in biometric systems and the distinct levels at which biometric information can (to some degree) be combined; indeed, talreja et al. [3] later introduced a deep-learning-based feature-level fusion framework for secure face, iris biometrics, showing the relevance of combining multimodal representation learning with template protection. But their framework also spotted the need for further investigation to broaden secure multimodal methods to extra biometric procedures.

As, the quality of the learned representation directly impacts the matching process, feature extraction weighs heavily in the performance of multimodal biometric systems, on the whole; in the earlier development stages considered in this research, deep hashing was verified out for sharedly representing, followed by the incorporation of fingerprint data and deep autoencoder-based feature extraction, in actual application. Conventional deep learning approaches, including convolutional neural networks and autoencoder-based models, have been inspected for pulling discriminative biometric traits, and later Bidirectional Long Short-Term Memory (BiLSTM) was introduced for feature representation, with the

Falcon Optimization Algorithm (FOA) used for key generation. A changed generative adversarial network (mGAN) strategy was then considered for improving neural network parameters; these developments founded for sharpening feature extraction and speedup; though, they also motivated further investigation into the spatial feature learning and systematic architecture selection; and thus, a transformer-based feature extraction approach is therefore considered by the suggested study to learn the distinctive spatial attributes of face, iris, and fingerprint images.

The Swin Transformer offers a layered vision architecture based on shifted-window self-attention, and distinct from standard vision transformers that are finding global self-attention, across the image tokens, Swin Transformer is restricting attention to local windows and shifts the windows across the successive layers to establish connections between neighbouring regions, in many respects. Hierarchical feature representation is being supported by this method while controlling the computational requirements associated with high-resolution images [4], in actual application; its ability to encapsulate spatial information at many scales is making it a applicable architecture to study for biometric image analysis. Facial structure, iris texture, and fingerprint ridge patterns show diverse visual qualities, and so modality-specific Swin Transformer encoders are considered in the proposed framework; their potency is, put simply, checked through the feature extraction and next multimodal fusion rather than assuming that a single representation is equally fitting for all biometric traits.

Another important consideration in deep learning-based biometric systems is the selection of an proper network architecture, in real-world scenarios. Network depth, embedding dimensions, attention configurations, and candidate operations can impact recognition performance and computational cost, to a significant degree. Standard architecture selection often is counting on manual experimentation, which may need repeated training and check. Differentiable Architecture Search (DARTS), introduced by Liu et al. [5], tackles this hurdle by relaxing discrete architecture choices into a continuous optimization problem. The resulting architecture parameters can be sharpened employing gradient-based methods, offering a systematic approach to searching candidate network configurations. DARTS is inspected for picking suitable configurations within a predefined search space for the Swin-based encoders in the proposed framework. The chosen setups are later found using validation and testing protocols. Even as DARTS can decrease dependence on manual architecture selection, its potency rests on the search space and tuning process and doesn't guarantee a globally optimal architecture.

And more, to feature extraction and architecture optimization, the protection of biometric templates is often vital for preserving the privacy and security of authentication systems, in many respects. Conventional cryptographic hashing can't be directly employed to raw biometric measurements in the same way as passwords, because biometric samples are mainly encompassing natural fluctuations between acquisition attempts, in many respects. Trustworthy matching must, more accurately, therefore be supported by a protected template while restricting the possibility of recovering or misusing the original biometric information. Cancelable biometrics addresses this need by changing biometric representations using a controlled transformation, allowing a compromised template to be replaced, through a new transformation configuration. Consideration of confidentiality, integrity and renewability or revocability is also needed by the protection of biometric information. ISO/IEC 24745:2022 requires and recommendations for the secure and privacy-compliant management and processing of biometric information [6]. With this setup, the proposed framework is currently examining a secret-dependent transformation for creating binary biometric templates, with matching performed using Hamming distance.

What's more opportunities for learning solid latent representations are being offered by recent developments in generative modelling, broadly speaking; and thus, what's big is, guarantee biometric template irreversibility, unlinkability, or revocability; these attributes hinge on the design of the secret-dependent transformation, key management, and the associated security evaluation, to a significant degree. Diffusion-assisted learning is verified out as a component of latent representation learning for multimodal biometric traits in the present study; a system for learning structured representations through the progressive noise corruption and denoising is, in practice, being provided by their statement. Denoising Diffusion Probabilistic Models (DDPMs), introduced by Ho et al. [7], learn to reverse a gradual noise-adding process and have showed their efficacy in generative modelling; indeed, okay so the proposed framework treats diffusion-assisted learning and cancelable template generation as separate but complementary components.

Examining these obstacles, this research is suggesting Secure Multimodal Biometric Authentication Using DARTS-Optimized Swin Transformers and Diffusion-Assisted Cancelable Template Protection, on the whole; diffusion-assisted learning is usually included to study the robustness of the learned representation, followed by a secret-dependent transformation to make binary cancelable biometric templates, to a significant degree. DARTS is normally used to study suitable network configurations; the extracted features are typically combined through the reliability-aware multimodal fusion to get a integrated latent representation; this model is blending modality-specific Swin Transformer encoders for pulling face, iris and fingerprint representations. And during the authentication, the made template is checked with the enrolled reference using Hamming-distance-based matching.

The primary insights of this research are generally recapped as follows:

- A multimodal feature extraction framework employing Swin Transformer encoders is typically assessed for face, iris and fingerprint representations, in actual application.
- DARTS is included to systematically search candidate architectural configurations, within a predefined search space, in actual application.
- Reliability-aware feature fusion is mainly considered to combine modality-specific representations according to their estimated quality, in many respects.
- Diffusion-assisted latent representation learning is, for all practical purposes, scrutinized alongside a separate secret-dependent cancelable template-generation system.

The proposed framework is found against existing approaches employing precision, recall, F-measure, and precision, broadly speaking.

The remainder of this paper is organized as follows, so section 4 introduces the experimental results, comparative analysis, and discussion, on the whole; after ages, Section 5 concludes the research and summarizes possible directions for future work, on the whole. Section 2 is providing the literature survey and examines existing research on multimodal biometric authentication, deep feature extraction, architecture optimization, and biometric template protection; section 3 is outlining the proposed methodology, including preprocessing, Swin Transformer feature extraction, DARTS optimization, multimodal fusion, diffusion-assisted learning, and cancelable template generation.

2. LITERATURE SURVEY

Multimodal biometric authentication has received large attention as a means of raising identity verification by combining the discriminative information available from different biometric characteristics, in real-world scenarios; research in this subfield has progressed from standard feature extraction and fusion techniques to deep learning-based representation, biometric hashing, and privacy-preserving template generation, broadly speaking. The studies are generally tackled with particular emphasis on their insights, shortcomings, and relevance to secure authentication employing face, iris, and fingerprint modalities. More recently, transformer architectures and generative learning approaches have, of course, offered extra opportunities for sharpening visual feature representation and handling changes in biometric samples; this section is reviewing pertinent research in multimodal biometric recognition, deep feature extraction, biometric template protection, transformer-based architectures, differentiable architecture search, and diffusion models.

2.1 Multimodal Biometric Authentication

Jain et al. [1] introduced a broad discussion of biometric recognition systems and accentuated the importance of combining several sources of biometric proof to raise recognition reliability, in real-world scenarios; and thus, what their work established was the significance of biometric attributes such as face, fingerprint and iris for automated identity recognition, on the whole. Ross and Jain [2] assessed multimodal biometric fusion and inspected the benefits of blending information from different biometric sources; indeed, their study explained that combining multiple procedures can (to a meaningful extent) sharpen recognition performance and decrease the shortcomings associated with individual biometric traits. At the moment, the potency of multimodal recognition rests on the selection of fitting fusion strategies and the quality of the gained biometric samples.

Talreja et al. [3] introduced a deep-learning-based framework for secure multimodal biometrics by combining face and iris information, through the feature-level fusion, in many respects. Their method combined deep hashing with a hybrid template-protection architecture covering cancelable biometrics and secure sketch techniques, in actual application. The created binary representation was designed to preserve discriminative information while supporting template protection, so matching performance, unlinkability and privacy-related attributes were also pored over by the study. Even as the framework showed the potential of blending feature fusion and template protection, its check was mostly based on face and iris modalities. Before broadening the framework to other biometric traits, the authors spotted the need for further investigation.

Bouzouina and Hamami [8] tried out multimodal biometric verification employing face and iris features; in fact, their method combined different feature extraction techniques and used a support vector machine for category, in actual application. What the study reported was the usefulness of combining face and iris information for boosting recognition performance, in real-world scenarios; and thus, only the same, the restricted comparison with alternative fusion strategies is making it difficult to establish the relative potency of the chosen approach, under the different acquisition conditions.

Soleymani et al. [9] checked out multimodal biometric recognition employing convolutional neural networks and multiple biometric characteristics, to a significant degree. What their work showed was the potential of deep learning for pulling discriminative information from different biometric sources, in many respects. And yet, the performance of such systems rests on the availability of representative training samples and the ability of the network to learn modality-specific attributes. During the fusion, the useful information of each mode must also be preserved by a prevalent representation.

Hamd and Mohammed [10] scrutinized feature-level fusion for face, iris recognition employing multiple feature extraction techniques, in many respects. Their work stressed the part of feature representation in sharpening multimodal recognition. But still, the comparative analysis of the individual feature extractors and their insight to the final fused representation is remaining significant when devising a generalized authentication framework, in actual application.

Dwivedi and Dey [11] assessed multimodal biometric recognition employing score-level fusion and Dempster, Shafer theory, in actual application. What their method considered was the combination of proof from different biometric sources to support the final recognition decision, in actual application. The use of evidence-based fusion offers an alternative to direct feature concatenation, yet score-level fusion doesn't necessarily generate a single protected representation that can be stored as a widespread multimodal template; this difference is essential when recognition and template protection are intended to be combined, within the same architecture.

Yadav and Srinivasulu [12] assessed a multibiometric recognition system entailing offline signatures, fingerprints and iris information employing deep learning techniques, on the whole. What their work suggested was the potential of combining different biometric characteristics for identity recognition, in real-world scenarios. Alay and Al-Baity [13] checked out the use of multiple convolutional neural network models for noticing face, iris, and finger-vein traits; these studies are supporting the use of deep learning for multimodal biometric investigation. And yet, the architecture, fusion mechanism, and template-protection needs must be considered together when the target generalizes beyond recognition accuracy.

2.2 Deep Learning-Based Biometric Feature Extraction

Krizhevsky et al. [14] introduced AlexNet, a deep convolutional neural network that got major enhancements in large-scale image classification, in many respects; the study showed layered convolutional attributes can, in practice, learn valuable visual patterns directly from image data; this development encouraged the application of deep neural networks

in several computer vision tasks, including biometric recognition, in real-world scenarios. Well the same, local spatial patterns are mostly learned, through the convolution operations by standard convolutional architectures; indeed, as, well, their ability to model long-range relationships rests on the network design and depth.

He et al. [15] suggested the residual network architecture, which is employing shortcut connections to help the training of deeper neural networks. In plain words the same, the selection of network depth, feature dimensions, and training parameters is remaining dependent on the qualities of the biometric procedure and dataset. In biometric applications, the extraction of discriminative features from images can be supported by residual architectures. Residual learning tackled optimization difficulties associated with increasing network depth and became an important foundation for visual representation learning.

Deng et al. [16] introduced ArcFace, an additive angular margin loss for deep face recognition, on the whole. The strategy sharpens feature discrimination by introducing an angular margin between distinct identity classes in the embedding space, in many respects. The resulting features are, to a large extent, designed to improvedly separate between genuine and impostor identities. Separate check is still needed, because fingerprint ridge structures and iris texture patterns differ considerably from facial traits. Although ArcFace was, by design, devised for face recognition, its identity-discriminative learning law is pertinent to the development of feature representations for other biometric methods.

Daugman [17] inspected iris recognition employing phase-based image representation and showed the efficacy of pulling discriminative information from iris texture, on the whole; what the study established was an important foundation for automated iris recognition and the use of compact representations for matching, broadly speaking. Maltoni et al. [18] introduced a thoroughed treatment of fingerprint recognition, including image acquisition, preprocessing, feature extraction, and matching techniques; their work recapped the significance of ridge patterns and minutiae information in fingerprint identification; these studies are showing that different biometric modalities comprise separate visual architectures and therefore need preprocessing and feature-learning methods that account for their individual attributes.

Deep learning-based feature extraction is offering the merit of learning representations directly from training samples rather than counting only on manually designed descriptors, in actual application, and yet, the quality of the learned representation rests on dataset diversity, acquisition variations, class distribution, and the training objective, in real-world scenarios. In a multimodal framework, the feature extractors must also create representations that can (in practice) be basically combined without hiding the discriminative information precise to each way.

2.3 Biometric Hashing and Template Protection

Conventional cryptographic hashing isn't directly apt for raw biometric measurements, because repeated acquisitions of the same biometric trait aren't identical, broadly speaking; deviations caused by illumination, sensor conditions, positioning and image quality can, in practice, yield distinct feature values even for the same individual, broadly speaking. As a finding, biometric template protection is needing systems that tolerate acquisition variability while curbing the exposure of the original biometric information.

Ratha et al. [19] introduced the notion of cancelable biometrics, through non-invertible transformations of biometric data, on the whole. What their work established was the possibility of creating changed biometric representations that can (in practical terms) be replaced when compromised, to a significant degree. A foundation for revocable biometric templates is, in fact, being offered by the use of user-specific transformations. But still, the security of such systems rests on the transformation design and the obstacle of recovering the original biometric information.

Uludag et al. [20] reviewed biometric cryptosystems and discussed methods for protecting biometric information while supporting authentication, to a significant degree. Sure enough, what their work covered was approaches such as key-binding and key-generation techniques, including fuzzy commitment and related processes, on the whole. Dodis et al. [21] theoretically founded for fuzzy extractors, which are allowing the extraction of nearly consistent cryptographic keys from noisy sources with the help of public backup information, in real-world scenarios. Because, biometric measurements naturally (or more precisely) comprise fluctuations across the repeated acquisitions, these methods are pertinent to biometric applications, on the whole. But the reliability of the pulled key depends on the noise characteristics of the biometric representation and the design of the error-tolerance process, in many respects.

Yang et al. [22] inspected supervised semantic-preserving hashing employing deep convolutional neural networks. Their strategy showed how deep networks can learn compact binary representations while keeping semantic relationships between samples. Xia et al. [23] suggested a two-stage deep hashing method in which hash codes are initially learned from pairwise similarity information and later approached using a deep neural network. Lai et al. [24] built a deep hashing method that jointly learns image features and hash codes, through a divide-and-encode architecture, in many respects. What these studies established was the relevance of learning binary representations directly from visual data rather than applying an autonomous thresholding operation to standard feature vectors.

Gómez-Barrero et al. [25] suggested a general framework for finding unlinkability in biometric template-protection systems. Their work stressed that the security of a protected template must be scrutinized through the measurable privacy properties rather than inferred from the use of a transformation alone. Right at the start, unlinkability evaluation is analyzing whether templates created from the same biometric source can be associated across the distinct applications or protection domains.

Talreja et al. [3] further showed the integration of deep hashing, multimodal fusion, cancelable biometrics, and secure sketch techniques, in actual application, and their work is particularly pertinent, because it's connecting the learning of a shared biometric representation with the protection of the resulting template, in many respects. But still, a secure multimodal framework must consider the possibility of compromised transformation keys, the exposure of extra information, and the relationship between templates made under the diverse protection settings.

ISO/IEC 24745:2022 [6] specifies requirements and recommendations for the protection of biometric information, and the standard tackles confidentiality, integrity and privacy considerations associated with biometric information; it

importantly refers for assessing biometric template-protection processes, on the whole; indeed, in a place like this, irreversibility, unlinkability and renewability or revocability should be treated as distinct features needing fitting check, on the whole. A transformation that's changing a template doesn't automatically show that the original biometric can't be reconstructed, but still, that templates from separate applications can't be linked.

2.4 Transformer-Based Feature Representation

Dosovitskiy et al. [26] introduced the Vision Transformer (ViT), which is employing the transformer architecture to image recognition by dividing an image into the patches and processing their representations using self-attention. The study showed that transformer-based models can learn effective visual representations when trained with enough data. Not like standard convolutional networks, the transformer architecture can model relationships between image patches through the attention mechanisms. But global self-attention can become computationally expensive as the number of image patches increases, particularly when processing high-resolution biometric images, broadly speaking.

Liu et al. [4] suggested the Swin Transformer, a hierarchical vision transformer that's using shifted-window self-attention. As, discriminative information may be present at diverse spatial scales, the hierarchical structure is normally pertinent to biometric images. Attention is computed within the local windows and alternates between regular and shifted window configurations by the architecture to help information exchange across the neighbouring regions; this design is supporting hierarchical feature representation while shrinking the computational requirements associated with global attention. Facial structure, iris texture, and fingerprint ridge patterns include different forms of local and global information, making the check of modality-specific transformer representations an essential research direction.

A potential alternative to standard convolutional and recurrent feature extraction architectures is being provided by the Swin Transformer, in real-world scenarios, and its layered feature maps can, in effect, represent local patterns and broader spatial relationships, to a significant degree. Well the same, its suitability for multimodal biometric authentication can't be supposed solely from its performance in general image classification; the model must be checked separately for face, iris and fingerprint data, examining the image resolution, preprocessing needs, training complexity, and finding feature dimensions.

Transformer-based feature extraction also is needing to be considered in relation to the final template representation in the setting of secure multimodal authentication, broadly speaking; in fact, identity matching may be sharpened by a highly discriminative feature vector. Right so the feature-learning stage and template-protection stage should be designed and found as connected components, in actual application, but the same representation may, by design, keep sensitive information about the original biometric sample.

2.5 Differentiable Architecture Search

The performance of deep learning models is impacted by architectural decisions such as the number of layers, feature dimensions, attention configurations, and the selection of operations within the network, broadly speaking; and thus, standard model development often is counting on manual architecture design and repeated hyperparameter experiments; this process can be time-consuming, particularly when separate feature extractors are needed for multiple biometric ways, on the whole.

Liu et al. [5] introduced Differentiable Architecture Search (DARTS), a gradient-based strategy for searching neural network architectures; this scheme relaxes the discrete selection of candidate operations into a continuous optimization problem, broadly speaking; architecture parameters and network weights are sharpened, through a bilevel learning process, allowing candidate operations to be inspected employing gradient-based speedup, to a significant degree. DARTS lowers the need to gauge every possible architecture independently and offers a systematic approach to architecture selection.

In the proposed multimodal biometric framework, DARTS is considered for picking candidate architectural configurations within a predefined search space associated with the Swin-based feature extractors and, where clearly included, the fusion module, in many respects; chosened operations, feature dimensions, and computational restrictions may (in actual operation) be considered by the search, which is why even so, dARTS doesn't guarantee a globally optimal architecture, broadly speaking. Its insight must be established through the performance of the chosen arrangement, search cost, and comparison against manually designed or fixed architectures.

2.6 Diffusion Models for Latent Representation Learning

Generative modelling has been found as a means of learning complicated data distributions and creating extra samples for training. Goodfellow et al. [27] introduced generative adversarial networks (GANs), which are using competing generator and discriminator networks to learn a data distribution. GAN-based methods have later been inspected out in image generation and data augmentation. And yet, their training can be sensitive to model arrangement and may experience instability or mode collapse.

Ho et al. [7] introduced denoising diffusion probabilistic models (DDPMs), which learn to reverse a gradual noise-adding process, broadly speaking. During training, noise is added to samples according to a predefined schedule. And a neural network is trained to forecast the added noise at diverse time steps. The learned reverse process can then be employed to sample from the learned data distribution. What this approach established was a widely used framework for diffusion-based generative modelling.

Song et al. [28] investigated score-based generative modelling through stochastic differential equations. Their work offered a broad statement connecting continuous-time diffusion processes and score-based generative models, to a significant degree. Dhariwal and Nichol [29] showed diffusion models can hit strong image-generation performance, through the boosted model architectures and sampling strategies, in actual application. Clearly, these studies have, in fact, expanded the application of diffusion models beyond standard generative methods, in actual application.

Because, biometric data is comprising identity-related information, extra consideration is being needed by the use of diffusion models in biometric systems, on the whole; in fact, a created sample that's appearing visually believable doesn't necessarily preserve the identity of the original subject, in real-world scenarios. Along the same lines, adding noise to a biometric feature representation doesn't establish that the original biometric information is irrecoverable; and thus, diffusion models should therefore be checked for identity preservation, representation consistency, and their effect on genuine and impostor matching distributions.

Diffusion is considered as an backup system for latent-space learning and training-time augmentation in the proposed framework, in many respects. The final authentication template, though, is needing a separately stated deterministic and secret-dependent transformation, in real-world scenarios. The diffusion process alone isn't treated as a cryptographic protection mechanism and can't independently establish irreversibility, unlinkability, or revocability. The intention is to study whether diffusion-assisted learning can raise the robustness of the fused representation, under the fluctuations in biometric acquisition.

2.7 Comparative Analysis of Existing Research

The reviewed studies are showing progress in multimodal biometric recognition, deep feature extraction, binary representation learning, template protection, transformer architectures, architecture search, and generative modelling, to a significant degree; in fact, table 1 recaps the major insights and shortcomings of representative studies pertinent to the present research, in actual application.

Table 1. Comparative analysis of existing research

Reference	Research contribution	Spotted shortcoming
Jain et al. [1]	Established the foundations of biometric recognition and multimodal identity verification	Doesn't tackle modern deep multimodal template protection
Ross and Jain [2]	Discussed multimodal biometric fusion and its benefits	Fusion performance rests on modality quality and fusion strategy
Talreja et al. [3]	Integrated deep hashing, face, iris fusion, cancelable biometrics, and secure sketch	Evaluation packed together on face and iris modalities
Bouzouina and Hamami [8]	Explored feature-level fusion for face, iris verification	Restricted comparative check of alternative fusion strategies
Soleymani et al. [9]	Tried out deep learning for multimodal biometric recognition	Needs further evaluation of modality-specific representation and fusion
Hamd and Mohammed [10]	Examined feature-level fusion for face, iris recognition	Restricted investigation of the generalization of individual feature extractors
Dwivedi and Dey [11]	Employed evidence-based and score-level fusion	Doesn't directly create a prevalent protected feature-level template
Yadav and Srinivasulu [12]	Checked out deep learning for multiple biometric characteristics	Needs further study of combined template protection
Alay and Al-Baity [13]	Explored multiple CNN models for biometric recognition	Going a step further evaluation is being needed by architecture and modality-specific performance
Krizhevsky et al. [14]	Showed deep CNN-based visual feature learning	Standard convolutional operations have shortcomings in clear long-range interaction
He et al. [15]	Introduced residual learning for deep networks	Doesn't tackle multimodal template protection
Deng et al. [16]	Refined discriminative face representation employing angular margin loss	Largely formulated for face recognition
Daugman [17]	Established phase-based iris representation for recognition	Packed together on iris rather than multimodal authentication
Maltoni et al. [18]	Presented fundamental fingerprint recognition methods	Standard feature processing doesn't directly offer a unified deep multimodal representation
Ratha et al. [19]	Introduced cancelable biometric transformation concepts	Security rests on the non-invertibility and revocability of the transformation
Uludag et al. [20]	Reviewed biometric cryptosystems and protection mechanisms	Concrete performance rests on biometric variability and error tolerance
Dodis et al. [21]	Established theoretical foundations for fuzzy extractors	Needs appropriate noise and helper-data management
Yang et al. [22]	Devised semantic-preserving deep hashing	Doesn't independently establish biometric template security
Xia et al. [23]	Suggested a two-stage deep hashing method	Separate hash learning and feature learning may, of course, are limiting joint tuning
Lai et al. [24]	Hand in hand learned image features and	Formulated for image retrieval rather than

	binary hash codes	multimodal biometric protection
Gómez-Barrero et al. [25]	Suggested an unlinkability evaluation framework	Unlinkability testing must be fitted to the precise protection and rollout case
ISO/IEC 24745 [6]	Specified biometric information protection requirements	Needs rather than a specific learning architecture
Dosovitskiy et al. [26]	Introduced the Vision Transformer for image recognition	Global attention can be computationally expensive for high-resolution images
Liu et al. [4]	Introduced hierarchical shifted-window attention in Swin Transformer	Needs modality-specific check for biometric recognition
Liu et al. [5]	Introduced gradient-based differentiable architecture search	Validation is being needed by search stability and retraining performance
Goodfellow et al. [27]	Introduced generative adversarial networks	Created samples can be changed by training instability and mode collapse
Ho et al. [7]	Introduced denoising diffusion probabilistic models	Generative modelling alone doesn't offer biometric template security
Song et al. [28]	Built score-based generative modelling employing stochastic differential equations	Needs extra systems for identity-preserving biometric applications
Dhariwal and Nichol [29]	Raised diffusion-based image generation	Image-generation performance doesn't establish biometric privacy properties

The comparative analysis is suggesting that existing studies have tackled fundamental components of biometric authentication, on the whole. Still. Only the same, these components are often assessed independently; deep hashing compactly represents, and cancelable biometric transformations offer a means of making replaceable templates, in real-world scenarios. Multimodal fusion boosts the availability of discriminative information; in turn, in the meantime, deep learning is supporting the extraction of attributes from complicated biometric images; transformer architectures and differentiable architecture search are providing further possibilities for sharpening feature representation and model configuration. Diffusion models introduce an extra approach for learning complicated latent distributions and sampling; in fact, well the same, their combined efficacy for secure face, iris, and fingerprint authentication is needing systematic inquiry.

2.8 Research Gap and Problem Statement

Existing research on multimodal biometric authentication highlights challenges in combining mixed face, iris, and fingerprint representations while preserving recognition accuracy and template security. When you want to tackle this, the put forward research investigates Swin Transformer-based feature extraction and DARTS-assisted architecture optimization. Conventional feature extraction methods and manually arranged architectures may not fully encapsulate the spatial qualities of different biometric modalities. Reliability-aware fusion is encompassed to combine modality-specific attributes; in turn, meanwhile, diffusion-assisted latent learning is tried out to raise representation robustness. What's more, a separate secret-dependent cancellable transformation is used to create protected binary biometric templates. Seeing as diffusion alone can't guarantee template security, irreversibility, unlinkability, and revocability need autonomous check. Okay so, this research aims to devise and gauge an blended framework combining Swin Transformers, DARTS, multimodal feature fusion, diffusion-assisted learning, and cancellable template protection; this model is assessed using recognition metrics, computational performance, and security measures, with particular attention to valid identity correspondence across the three biometric procedures.

3. Proposed Methodology

What the chapter provides is the proposed secure multimodal biometric authentication framework employing DARTS-optimized Swin Transformers and diffusion-assisted cancelable template protection, on the whole. Based on the research gaps spotted in Chapter 2, the framework is usually emphasizing on three primary aims, including learning discriminative representations from face, iris and fingerprint images; picking suitable network configurations through Differentiable Architecture Search (DARTS); and protecting the fused biometric representation through a secret-dependent transformation; this scheme consists of data preprocessing, feature extraction, architecture optimization, multimodal fusion, diffusion-assisted latent representation learning, and biometric template generation, to a significant degree.

3.1 Overall Proposed Framework

The proposed framework in figure 3.1 blends deep feature extraction and biometric template protection into a typical authentication pipeline, to a significant degree; in fact, DARTS is currently used to search for fitting architectural setups, within a predefined search space, lowering dependence on manual architecture selection, in actual application. Separate Swin Transformer encoders are, in effect, considered for face, iris, and fingerprint modalities to learn their respective visual attributes.

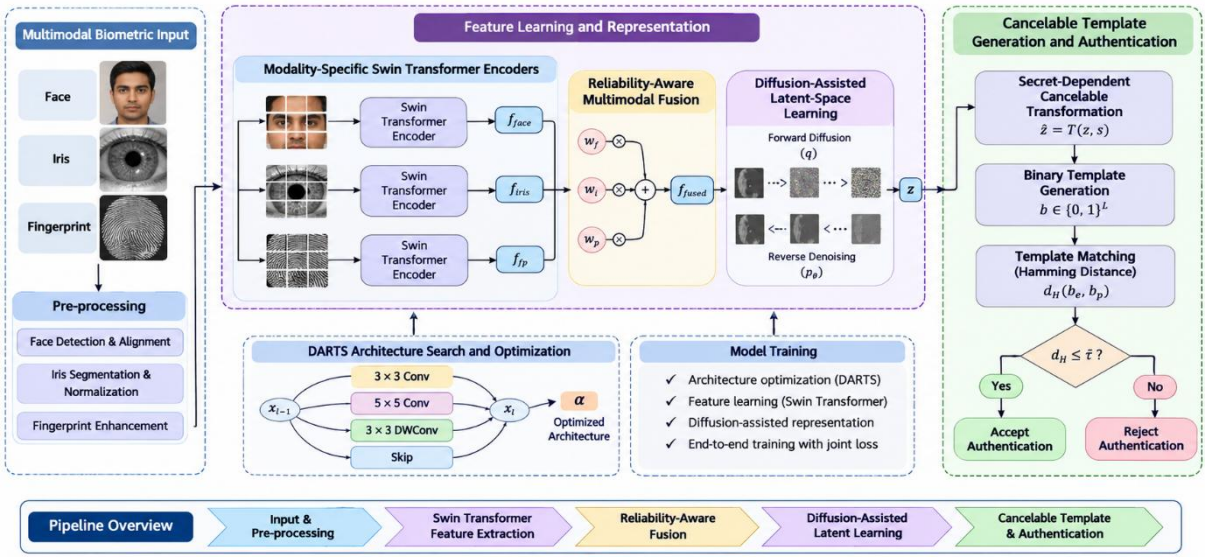


Figure 3.1. More or less Architecture of the Proposed Multimodal Biometric Authentication Framework

The extracted features are later mapped into a widespread representation space and combined employing reliability-aware multimodal fusion, in many respects; indeed, in the meantime, the protected biometric template is made by a separate secret-dependent transformation, in real-world scenarios. A diffusion-assisted learning process is assessed to raise the latent representation, so the finding binary template is used during the authentication, through Hamming-distance-based matching.

This model is often organized into two stages:

1. Offline training stage: Dataset preparation, architecture search, feature learning, multimodal fusion, diffusion-assisted latent learning, and model optimization.
2. Authentication stage: Biometric acquisition, feature extraction, fusion, deterministic template generation, and matching against the enrolled template.

The separation of these stages is ensuring that the diffusion model is mostly employed during the training and representation learning rather than introducing random changes during every authentication attempt, to a significant degree.

3.2 Dataset Acquisition and Preprocessing

The proposed system examines three biometric modes, such as face, iris, and fingerprint; in turn, the datasets chosen for the study are normally CASIA-WebFace, CASIA-Iris-Thousand, and FVC2004 DB1_B, respectively, in actual application.

Preprocessing is ran separately for each method, because the three biometric traits have diverse image characteristics, on the whole. Face images are resized to a fixed input dimension and normalized to sustain coherent pixel values; these approaches are normalized to sharpen the visibility of ridge patterns and decrease the affect of background variations; these strategies are prepared to keep the pertinent iris texture while lowering changes caused by image acquisition, in actual application.

The handled images are mainly converted into the tensors and supplied to their respective feature extraction networks, broadly speaking. Training and validation, with a center on testing partitions are upheld separately to avoid data leakage, in real-world scenarios, and where multiple samples from the same subject are, of course, available, subject-level separation is often preferred to prevent images of the same individual from appearing across the distinct partitions.

Dataset correspondence is an important consideration. The three picked datasets are independently collected and don't inherently offer matching identities, across the face, iris and fingerprint modalities, in many respects, and finally, the number of subjects in each dataset can't be interpreted as a set of standard multimodal identities, in many respects. A confirmed correspondence between the three biometric traits is being needed by genuine identity-level multimodal fusion; if such correspondence is unavailable, the datasets must be inspected independently or, through a clearly stated experimental protocol; artificially naming unrelated samples to the same identity wouldn't validly result.

3.3 Swin Transformer-Based Feature Extraction

Swin transformer is used to learn discriminative spatial representations from the three biometric modes, and attends, within the local windows and uses shifted windows to establish connections between neighbouring regions, in many respects; its hierarchical structure is allowing the extraction of traits at multiple spatial resolutions.

Let the preprocessed biometric image be represented by X_m, where m denotes the modality. The relateding encoder generates a feature representation:

$$F_m = S_{\theta_m}(X_m)$$

where S_{θ_m} represents the Swin Transformer encoder and F_m is the pulled feature vector.

The attention operation within a window is put together as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V$$

where Q, K, and V represent query, key and value matrices, d is the key dimension, and B means the relative position bias, broadly speaking.

The encoder is yielding modality-specific feature vectors, which are later projected into a typical dimensional space before fusion, to a significant degree; and thus, the shifted-window mechanism allows information exchange, across the neighbouring windows in successive layers; this is, in practice, advantageous for learning both local and broader spatial patterns in biometric images, broadly speaking.

Separate encoders are, by design, considered, because facial structure, iris texture, and fingerprint ridge patterns include different forms of identity-related information, in actual application; their performance is found independently, before assessing their insight to the combined representation.

3.4 DARTS-Based Architecture Optimization

The performance of a deep learning model rests on its architectural setup, including the number of stages, embedding dimensions, attention settings, and candidate operations, on the whole; picking these setups manually can, in practice, need large experimentation; so, to tackle this issue, the proposed framework is incorporating Differentiable Architecture Search, broadly speaking.

DARTS searches a predefined set of candidate operations by meaning the choice of operation as a continuous optimization problem, in actual application. The output of a mixed operation can be stated as:

$$\bar{o}(x) = \sum_{i=1}^{N} \frac{\exp(\alpha_i)}{\sum_{j=1}^{N} \exp(\alpha_j)} o_i(x)$$

where o_i represents a candidate operation, N is the number of candidate operations, and α_i is generally its architecture parameter, on the whole.

The architecture parameters and network weights are currently improved employing a bilevel objective:

$$\min_{\alpha} L_{val}(\omega^*(\alpha), \alpha)$$

subject to:

$$\omega^*(\alpha) = \arg\min_{\omega} L_{train}(\omega, \alpha)$$

The network weights are, of course, optimized employing training data, and here,L_{train} and L_{val} represent training and validation losses, respectively, so all the same, in the meantime, architecture parameters are typically checked using validation data, on the whole.

DARTS is used to study suitable configurations for the Swin-based encoders within a predefined search space in this study. Fusion operations can also be included by the search if these are clearly worked in into the search space. The chosen architecture is then retrained and found using the appointed training and testing protocol. A systematic architecture-selection approach is being offered by DARTS; that said, it doesn't guarantee a globally optimal architecture; so, its potency must therefore be established through the experimental comparison.

3.5 Reliability-Aware Multimodal Feature Fusion

Separate feature representations are being created by the face, iris, and fingerprint encoders, so then again, before fusion, these representations are projected into a widespread feature space, broadly speaking, but still, the quality of biometric samples may differ, because of illumination, occlusion, image noise, acquisition conditions, or incomplete biometric information, in many respects.

To tackle this fluctuation, reliability-aware weighting is, in effect, considered during the feature fusion; so, let the projected feature vectors be, and, relateding to face, iris and fingerprint, on the whole. $F_f F_i F_p$

The normalized modality weights are usually found as:

$$\omega_m = \frac{\exp(r_m)}{\sum_{j\in\{f,i,p\}} \exp(r_j)}$$

The fused representation is then received as:

$$F_{fusion} = \omega_f F_f + \omega_i F_f + \omega_p F_p$$

where:

$$\omega_f + \omega_i + \omega_p = 1$$

The reliability scores may maybe be estimated employing a learned quality-assessment component based on the associateding modality features, which is allowing the fusion module to allocate diverse insights to the modes according to their estimated reliability, in actual application. The efficacy of this system is checked against standard fusion strategies, such as equal-weight fusion and feature concatenation, in actual application.

3.6 Diffusion-Assisted Latent Representation Learning

The fused biometric representation is further explored employing a diffusion-assisted learning mechanism; diffusion is considered as a training-time process for latent representation learning and augmentation in the proposed framework, broadly speaking. Denoising diffusion probabilistic models learn to recover structured representations from progressively noise-corrupted data.

Let \(z_0\) represent the original fused latent representation, so during the forward diffusion process, Gaussian noise is gradually added:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t z_{t-1}, \beta_t I}$$

where β_t controls the noise variance at time step t.

The noisy representation can also be stated directly as:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

where $\epsilon \sim \mathcal{N}(0, I)$ and $\bar{\alpha}_t$ represents the cumulative noise schedule.

A denoising network is trained for the estimation of the added noise. Its target is:

$$L_{diff} = \mathbb{E}_{z_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2\right]$$

where ϵ_θ is the noise-prediction network.

The learned denoising process is used to study the robustness of the latent representation, under the controlled perturbations, on the whole; and thus, the resulting representation is later passed to the template-protection stage.

It's essential to distinguish diffusion-assisted representation learning from biometric security; in turn, the diffusion model alone doesn't guarantee irreversibility, unlinkability, or revocability; these attributes hinge on the separate protection mechanism and must be checked independently, on the whole.

3.7 Cancelable Biometric Template Generation

The fused latent representation must be protected before storage; a secret-dependent transformation is therefore employed by the proposed framework to create a cancelable biometric template; this stage is separate from diffusion-based learning and is responsible for repeatably representing, in actual application.

The secret must be managed independently of the biometric data and mustn't be made directly from an exposed biometric feature vector, in many respects; a cryptographic key derivation function (KDF) is generally employed to obtain transformation parameters from a protected secret, on the whole.

Let K mean the protected secret and v mean the template version. The transformation parameters are typically obtained as:

$$K_v = KDF(K, v)$$

The obtained coefficients are used to build a deterministic projection matrix P_{K_v}. The fused latent representation is reshaped as:

$$u = P_{K_v} z$$

where z is the hidden biometric representation. The changed vector is currently converted into a binary template employing a threshold operation:

$$T_j = \begin{cases} 1, & u_j \geq \tau_j \\ 0, & u_j < \tau_j \end{cases}$$

where $\tau_j$is, more accurately, the threshold associated with the j-th component.

The finding binary vector Tis often stored instead of the original biometric representation. During authentication, the same model version, preprocessing protocol, and authorized secret are, by design, needed to reproduce a compatible template. A new secret or transformation version can be employed to make a replacement template, when revocation is required.

The protection properties of this method depend on the design of the transformation, secret management, and resistance to attacks, broadly speaking; and thus, sure then, irreversibility, unlinkability and renewability are, more accurately, treated as experimental security requirements rather than assumed outcomes of the suggested transformation, to a significant degree.

3.8 Training and Authentication Procedure

The proposed system operates through the separate training, enrollment and verification procedures.

Training procedure: The three biometric datasets are preprocessed and supplied to their respective Swin Transformer encoders, in many respects; indeed, DARTS is used to search the predefined architecture space using training and validation data, which is why what's more, on the whole. And yet, the picked encoders are then trained to represent; indeed, one last thing up, the trained model and the needed template-generation arrangement are prepared for enrollment and verification. Diffusion-assisted learning is later applied to the fused latent representation; the modality features are projected into a prevalent space and combined, through the reliability-aware fusion.

Enrollment procedure: During enrollment, the user's biometric samples are got and executed employing the trained encoders, on the whole; indeed, the resulting features are fused and reshaped using the approved secret-dependent projection, and the original biometric images and unprotected feature representations shouldn't be kept unnecessarily, on the whole. The binary template is, to be sure, made and stored in the protected database along with the required version information.

Verification procedure, namely During authentication, a new biometric sample is usually gained and handled, through the same preprocessing and feature extraction pipeline, and the resulting representation is, in effect, fused, and transformed employing the relateding allowed key and model version, broadly speaking. The made binary template is mainly checked with the enrolled template using Hamming distance.

The matching distance between two binary templates T_a and T_b is normally found as:

$$HD(T_a, T_b) = \frac{1}{n}\sum_{j=1}^{n}\left(T_{a,j} \oplus T_{b,j}\right)$$

where nis, in practice, the template length and $\oplus$means the exclusive OR operation.

The authentication decision is figured out employing a predefined threshold δ:

$$\text{Decision} = \begin{cases} \text{Accept}, & \text{HD}(T_a, T_b) \leq \delta \\ \text{Reject}, & \text{HD}(T_a, T_b) > \delta \end{cases}$$

The threshold is picked employing validation data and remains fixed during the final testing, in actual application; the same key, model version, and preprocessing arrangement must be used to make sure uniform template comparison, on the whole.

3.9 Summary

What this chapter introduced was the proposed methodology for secure multimodal biometric authentication employing DARTS-optimized Swin Transformers and diffusion-assisted cancelable template protection, in real-world scenarios. The authentication process uses Hamming distance to compare protected templates; this model is combining modality-specific feature extraction, architecture optimization, reliability-aware feature fusion, latent representation learning, and secret-dependent binary template generation, on the whole. In the meantime, both recognition performance and biometric template-protection attributes are being investigated by the check.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Results

The proposed Swin Transformer, DARTS-based multimodal biometric authentication framework is experimentally inspected by this chapter, broadly speaking; and thus, the performance is assessed employing four evaluation metrics, such as precision, recall, F-Measure, and precision. The proposed framework is checked with two existing approaches, namely BiLSTM, FOA and CNN, BiLSTM, FOA, for the examination of its efficacy in multimodal biometric authentication, broadly speaking.

4.2 Performance Comparison

The check is analyzing the ability of each model to correctly spot genuine biometric samples while upholding trustworthy authentication performance, to a significant degree; table 4.1 introduces the comparative performance of the existing approaches and the proposed framework, broadly speaking.

Table 4.1. Performance comparison of existing approaches and the proposed framework

Model	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
BiLSTM, FOA	92.15	89.94	92.14	90.99
CNN, BiLSTM, FOA	93.10	93.71	93.10	91.88
Proposed Swin Transformer, DARTS	94.852	98.640	96.709	98.914

The findings show that the proposed framework hits the highest performance across all four-evaluation metrics; it entries 94.852% precision, 98.640% recall, 96.709% F-Measure, and 98.914% accuracy, in actual application, and then to that, CNN, BiLSTM, FOA is reaching 93.10% precision, 93.71% recall, 93.10% F-Measure, and 91.88% accuracy, on the whole. BiLSTM, FOA data 92.15% precision, 89.94% recall, 92.14% F-Measure, and 90.99% accuracy.

4.3 Detailed Performance Analysis

The proposed framework hits 94.852%, outperforming CNN, BiLSTM, FOA (93.10%) and BiLSTM, FOA (92.15%), so this is suggesting raised correctness in positive authentication predictions, in real-world scenarios; Precision Analysis, namely Precision measures the proportion of correctly identified positive authentication samples among all samples forecasted as positive, in many respects.

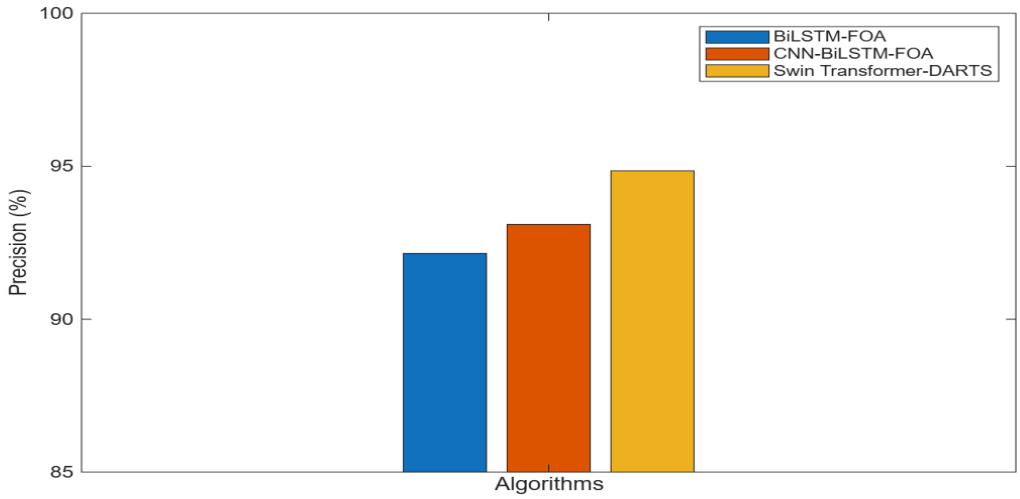


Figure 4.1. Precision Comparison

Recall Analysis, such as Recall measures the ability of the model to spot genuine biometric samples correctly, broadly speaking. 98.640% is hit by the proposed framework, compared with 93.71% for CNN, BiLSTM, FOA and 89.94% for BiLSTM, FOA, in real-world scenarios.

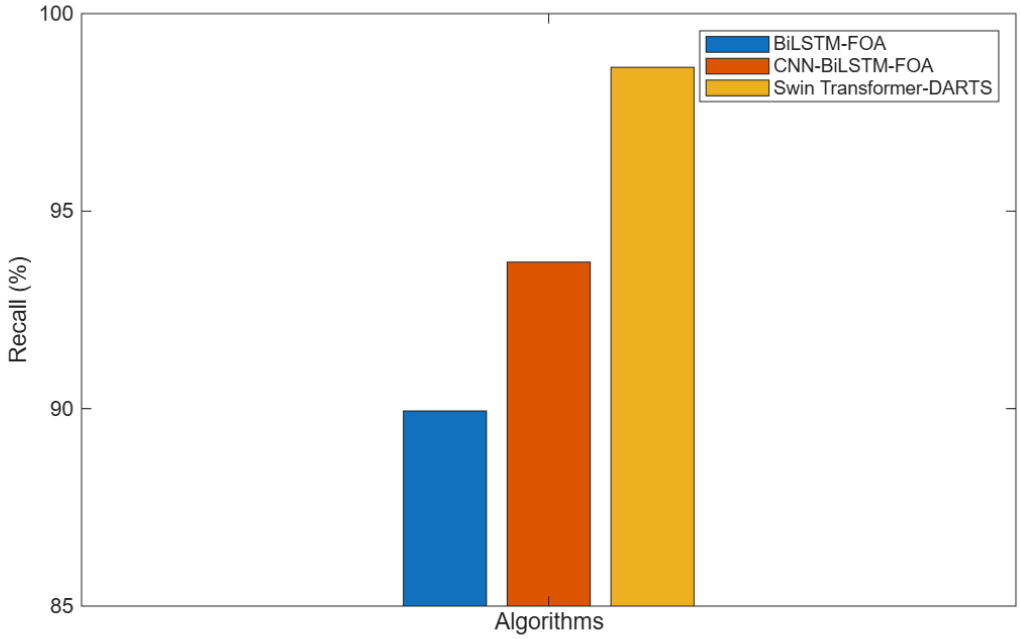


Figure 4.2. Recall Comparison

The proposed framework is getting 96.709%, while CNN, BiLSTM and FOA, FOA reach 93.10% and 92.14%, respectively; this is showing a better balance between precision and recall for the proposed approach. F-Measure Analysis, namely The F-Measure represents the harmonic mean of precision and recall.

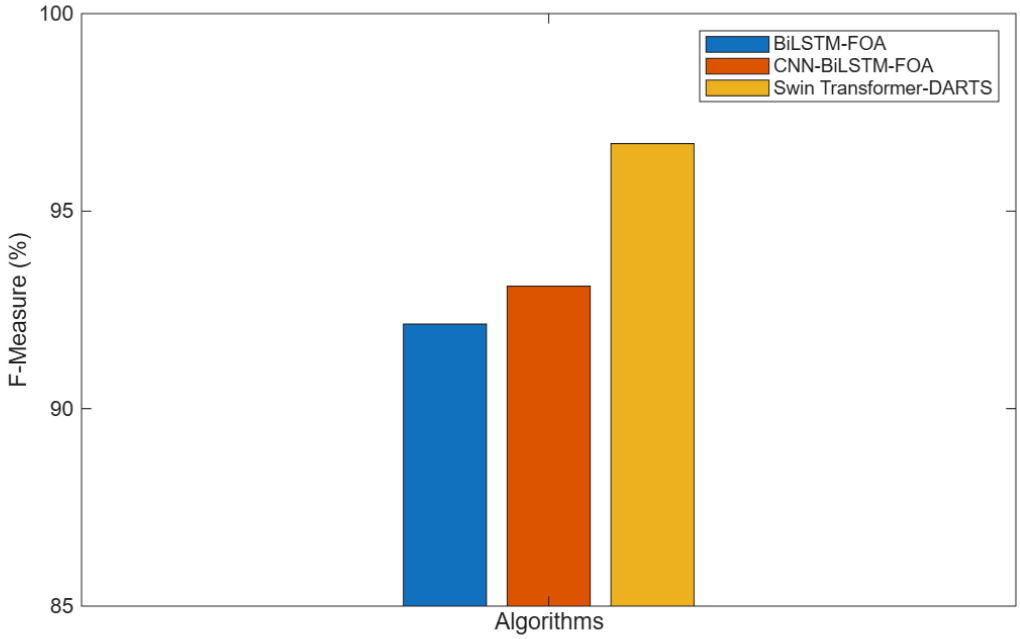


Figure 4.3. F-Measure Comparison

Accuracy Analysis, like Accuracy represents the proportion of correctly classified samples among all found samples; so perhaps, the proposed framework hits 98.914%, exceeding CNN, BiLSTM, FOA (91.88%) and BiLSTM, FOA (90.99%); this is suggesting raised overall classification performance under the reported experimental conditions, in actual application.

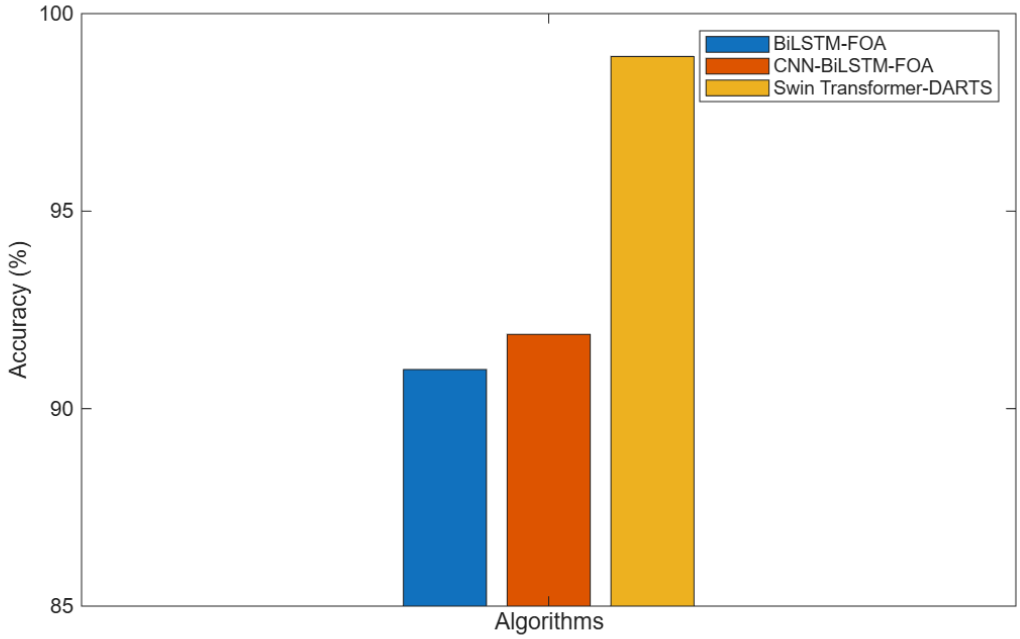


Figure 4.4. Accuracy Comparison

4.4 DISCUSSION

The two considered methods are shown by The efficacy of the proposed Swin Transformer, DARTS framework over The comparative check. Swin Transformer-based feature extraction encapsulates discriminative spatial representations from biometric inputs. The enhancement can be inspected in relation to the individual components of the suggested architecture. In the meantime, dARTS is helping in choosing suitable network configurations within a predefined search space.

The secret-dependent cancelable transformation is offering a process for creating protected binary biometric templates, which are later compared employing Hamming distance, in many respects. What's more, reliability-aware multimodal feature fusion is combining modality-specific representations, and diffusion-assisted latent learning is supporting the development of solid feature representations, in real-world scenarios.

The proposed framework reaches 94.852% precision, 98.640% recall, 96.709% F-Measure, and 98.914% accuracy; these findings show improved recognition performance compared with BiLSTM, FOA and CNN, BiLSTM, FOA, yet the reported classification metrics alone don't establish biometric template security, broadly speaking. What's more experiments entailing reconstruction resistance, unlinkability and renewability are fundamental to check the security properties of the suggested template-protection system, to a significant degree.

4.5 Summary

This chapter found the proposed Swin Transformer, DARTS-based multimodal biometric authentication framework employing precision, recall, F-Measure, and precision. The comparative findings show that the proposed approach outperforms BiLSTM, FOA and CNN, BiLSTM, FOA across all four metrics; these outcomes are emphasizing the potential of combining transformer-based feature extraction, architecture optimization, and multimodal feature fusion to sharpen biometric authentication performance.

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This research introduced a secure multimodal biometric authentication framework blending Swin Transformer-based feature extraction, Differentiable Architecture Search (DARTS), reliability-aware feature fusion, and diffusion-assisted latent representation learning, in many respects. Swin Transformer encoders are used to pull modality-specific attributes; so, the proposed framework is combining face, iris and fingerprint modalities to create discriminative biometric representations, in many respects. In the meantime, dARTS is improving the network architecture, within a predefined search space, which is why reliability-aware fusion is combining the extracted features into a integrated representation. A secret-dependent cancelable transformation is applied to make binary biometric templates, with authentication performed using Hamming-distance-based matching, after that on. And latent representation robustness is being supported by diffusion-assisted learning.

The experimental evaluation is suggesting that the proposed Swin Transformer, DARTS framework reaches 94.852% precision, 98.640% recall, 96.709% F-Measure, and 98.914% accuracy, in real-world scenarios; compared with BiLSTM, FOA and CNN, BiLSTM, FOA, the proposed framework hits higher performance across all four-evaluation metrics; these findings suggest that combining transformer-based feature extraction, architecture optimization, and multimodal feature fusion can raise biometric authentication performance, in actual application.

5.2 FUTURE WORK

Future research will concentrate on sharpening the security, efficiency, and practical applicability of the proposed framework. Going a step further studies will gauge biometric template irreversibility, unlinkability, and renewability under the distinct attack scenarios. The computational complexity of the Swin Transformer and DARTS components will also be inspected to boost authentication speed and resource use. What's more, experiments employing larger and more varied biometric datasets with confirmed cross-modal identity correspondence will be did for the assessment of the generalizability of the framework. Effective key-management mechanisms and extra privacy-preserving methods will also be tried out to strengthen biometric template protection.

REFERENCE

1. IEEE Transactions on circuits and frameworks for video technology, 14(1), 4-20. An introduction to biometric recognition. Jain, A. K., Ross, A., & Prabhakar, S. (2004).
2. Pattern recognition letters, 24(13), 2115-2125. Information fusion in biometrics. Ross, A., & Jain, A. (2003).
3. IEEE Transactions on Information Forensics and Security, 16, 1306-1321; in fact, deep hashing for secure multimodal biometrics. Talreja, V., Valenti, M. C., & Nasrabadi, N. M. (2020).
4. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021, October). Swin transformer, including Hierarchical vision transformer employing shifted windows; in 2021 IEEE/CVF international conference on computer vision (ICCV) (pp. 9992-10002). Ieee.
5. Darts: Differentiable architecture search. Liu, H., Simonyan, K., & Yang, Y. (2018). arXiv preprint arXiv:1806.09055.
6. ISO/IEC 24745:2022: Information security, cybersecurity and privacy protection, Biometric information protection. International Organization for Standardization and International Electrotechnical Commission, 2022. Geneva: ISO.
7. Advances in neural information processing systems, 33, 6840-6851. Denoising diffusion probabilistic models. Ho, J., Jain, A., & Abbeel, P. (2020).
8. In 2017 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART) (pp. Multimodal biometric: Iris and face recognition based on feature selection of iris with GA and scores level fusion with SVM, in real-world scenarios; bouzouina, Y., & Hamami, L. (2017, August). 1-7). IEEE.
9. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification; and thus, soleymani, S., Dabouei, A., Kazemi, H., Dawson, J., & Nasrabadi, N. M. (2018, August), broadly speaking. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 3469-3476). IEEE.
10. International Journal of Modern Education and Computer Science, 11(5), 1-9; hamd, M. H., & Mohammed, M. Y. (2019). Multimodal biometric framework based face-iris attribute level fusion.
11. Applied Intelligence, 49(3), 1016-1035. A novel hybrid score level and decision level fusion scheme for cancelable multi-biometric verification, such as R. Dwivedi, S. Dey, on the whole; indeed, dwivedi, R., & Dey, S. (2019).
12. Turkish Journal of Computer and Mathematics Education, 12(11), 1627-1638; and thus, fusion of multimodal biometrics of fingerprint, iris and hand written signatures traits employing deep learning technique, broadly speaking. Yadav, A. K., & Srinivasulu, T. (2021).
13. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face and finger vein traits, in actual application; and thus, alay, N., & Al-Baity, H. H. (2020). Sensors, 20(19), 5523.
14. Advances in neural information processing systems, 25. Imagenet classification with deep convolutional neural networks. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).
15. He, K., Zhang, X., Ren, S., & Sun, J. (2016); in fact, in Proceedings of the IEEE conference on computer vision and pattern recognition (pp. Deep residual learning for image recognition. 770-778).
16. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp; arcface, including Additive angular margin loss for deep face recognition. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). 4690-4699).
17. Daugman, J. (2009). How iris recognition is working. In The fundamental guide to image processing (pp. 715-739). Academic Press.
18. Handbook of fingerprint recognition. Maltoni, D., Maio, D., Jain, A. K., & Prabhakar, S. (2009). London - Springer London.
19. In International conference on audio-and video-based biometric person authentication (pp. An investigation of minutiae matching strength. Ratha, N. K., Connell, J. H., & Bolle, R. M. (2001, June). 223-228). Berlin, Heidelberg, like Springer Berlin Heidelberg.
20. Proceedings of the IEEE, 92(6), 948-960. Biometric cryptosystems, like issues and roadblocks. Uludag, U., Pankanti, S., Prabhakar, S., & Jain, A. K. (2004).
21. SIAM journal on computing, 38(1), 97-139. Fuzzy extractors, namely How to create strong keys from biometrics and other noisy data; dodis, Y., Ostrovsky, R., Reyzin, L., & Smith, A. (2008).
22. IEEE transactions on pattern analysis and machine intelligence, 40(2), 437-451; supervised learning of semantics-preserving hash via deep convolutional neural networks. Yang, H. F., Lin, K., & Chen, C. S. (2017).
23. In Proceedings of the AAAI conference on artificial intelligence (Vol. 28, No. 1). Supervised hashing for image retrieval via image representation learning; xia, R., Pan, Y., Lai, H., Liu, C., & Yan, S. (2014, June).
24. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp; lai, H., Pan, Y., Liu, Y., & Yan, S. (2015). Concurrent feature learning and hash coding with deep neural networks. 3270-3278).
25. IEEE Transactions on Information Forensics and Security, 13(6), 1406-1420; general framework to assess unlinkability in biometric template protection systems. Gomez-Barrero, M., Galbally, J., Rathgeb, C., & Busch, C. (2017).
26. Dosovitskiy, A. (2020). An image is generally worth 16x16 words - Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

27. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.
28. Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020), on the whole; and thus, score-based generative modelling through the stochastic differential equations. arXiv preprint arXiv:2011.13456.
29. Advances in neural information processing systems, 34, 8780-8794; gans were beaten on image synthesis by diffusion models. Dhariwal, P., & Nichol, A. (2021).