Scale-aware Economy-Accuracy Trade-off in Contactless Palmprint Recognition Using 14-Tap Q-Shift DTCWT and Layer-Wise Attention

Divyang S. Jadav
IIT Bombay, and SVKM's NMIMS, Navi-Mumbai
divyang.jadav@gmail.com

Siddhant Dongare
IIT Bombay
siddhantdongare718@gmail.com

Inderjeet
IIT Bombay
inderjeet30642@gmail.com

Saket Pateria
IIT Bombay
saket12may85@gmail.com

Deekshant Kumar
IIT Bombay
deekshantsaxena@gmail.com

Hemant Goklani
IIIT Surat
hsgoklani@gmail.com

Vikram Gadre
IIT Bombay
vmgadre@ee.iitb.ac.in

Abstract— **Contactless palmprint recognition has recently emerged as a popular alternative to conventional biometrics among researchers. The exploratory analysis of a data-driven economy for contactless palmprint recognition systems is relatively less explored in the literature. This work investigates the economy-accuracy trade-off for contactless palmprint recognition through a Dual-Tree Complex Wavelet Transform (DTCWT)-driven attention framework. A multiscale multiresolution representation of palmprint characteristics is obtained from the high-frequency subbands of the DTCWT and is learned by a compact convolutional neural network (CNN) integrated with configurable Convolutional Block Attention Module (CBAM). The study examines the joint effects of the depth of progressive decomposition and the selective placement of CBAM within progressive CNN layers on recognition performance in a moderate data scenario. The results indicate that CBAM applied at a deeper convolutional layer in a shallow or moderately deep CNN at an intermediate decomposition depth of the wavelet transform provides an optimal framework in terms of recognition accuracy and economy.**

Index Terms—**Contactless Palmprint Recognition, Dual-Tree Complex Wavelet Transform, Convolutional Block Attention Module, Multiscale Multiresolution Feature Representation**

I. Introduction

Identity management systems and computer vision-based security applications widely use biometrics to verify users, with fingerprints as a widely adopted biometric type, due to their uniqueness, permanence, and ease of deployment. The paradigm has shifted towards contactless fingerprint recognition following the COVID-19 pandemic, away from conventional touch-based systems. However, contactless fingerprint images often suffer from quality degradation caused by illumination variation, scale inconsistency, motion blur, and reduced ridge clarity, which adversely affect the extraction of fine discriminative features, such as minutiae points [13]. In contrast, contactless palm acquisition captures both fingerprint and palmprint characteristics over a larger surface area. Hence, it provides richer texture, principal lines, wrinkles, and ridge structures, allowing more reliable discriminative feature extraction even when the image quality is suboptimal [1].

Palmprint recognition process generally involves four fundamental stages: image acquisition, pre-processing, feature extraction, and matching. Existing palmprint recognition approaches include line-based, subspace learning, local directional coding, texture-based, and deep learning (DL)-based methods, which differ primarily in how structural and textural palmprint cues are encoded [1]. Extracting local features such as minutiae points from contactless palm images is challenging due to acquisition-related degradations, similar to those of fingerprints, and texture-based representations are highly sensitive to noise and imaging inconsistencies [1].

Recent DL models such as Fast CNN [2], EE-PRNet [3], PalmNet [5], [4], DDR [6], and RFN [7] have demonstrated strong recognition performance in palmprint biometrics. More broadly, DL has become the dominant paradigm in biometric recognition, object detection, image classification, and restoration tasks due to its ability to learn hierarchical feature representations directly from the data. Despite their high accuracy, deep hierarchical networks often require extensive hyperparameter tuning, large-scale annotated datasets, and substantial computational resources, while also exhibiting reduced interpretability of learned representations [11], [12], [14]. DL frameworks often lead to under-utilization of resources in biometric applications, where public datasets are available, but very large-scale data collection, annotation, and deployment-oriented optimization remain challenging. This underscores the need for compact, computationally efficient architectures that retain recognition capability without relying on excessive model complexity.

Conventional signal-processing approaches such as the Fourier Transform, Histogram of Oriented Gradients (HoG), Scale-Invariant Feature Transform (SIFT), and Local Binary Patterns (LBP) provide handcrafted and interpretable descriptors, but often lack robust joint spatial-frequency localization. The Discrete Wavelet Transform (DWT) partially addresses this limitation through multiscale decomposition; however, it remains shift-variant, translation-sensitive, and offers limited directional selectivity [8], [9], [10]. Advanced multiresolution frameworks belonging to the X-let family, particularly the Dual-Tree Complex Wavelet Transform (DTCWT), offer approximate shift invariance, enhanced directional selectivity, and stable multiscale analysis, making it especially suitable for contactless palmprint feature representation [8]. While transfer learning, few-shot learning, and transformer-based architectures have recently been explored to address a data-efficient biometric recognition system, they often introduce substantial computational overhead, reduced interpretability, and sensitivity to patch design [11], [12]. Consequently, integrating signal-processing priors within compact learning architectures offers a promising direction to achieve computationally economical and structurally interpretable biometric recognition under moderate-data conditions.

However, prior palmprint studies have not systematically examined how transform depth and attention placement jointly affect recognition accuracy and computational efficiency in compact wavelet-driven CNN architectures. To address this gap, the present study proposes extracting discriminative palmprint representations emphasizing principal lines and wrinkles using DTCWT, followed by feature refinement through the Convolutional Block Attention Module (CBAM).

The main contributions of the research are as follows:

- A compact DTCWT-CBAM framework is proposed for line-dominant contactless palmprint recognition using high-frequency transform features, for systematic evaluation of transform depth and attention placement.
- The joint analysis of progressive DTCWT decomposition and layer-wise attention in progressively complex CNN architectures is performed to identify the optimal architecture under limited-data settings. This prime contribution presents data-driven design-space analysis of contactless palmprint recognition.
- Experimental analysis shows that deeper-stage attention in a moderately compact CNN, combined with an intermediate transform depth, provides an effective balance between recognition accuracy and computational efficiency.

II. Proposed Methodology

The proposed framework, evident in Fig. 1, starts with pre-acquired palm images, followed by region-of-interest (RoI) extraction and pre-processing, multiscale multiresolution (MMR) feature extraction, MMR representation learning via a compact CNN combined with CBAM, and classification.

A. RoI Extraction and Pre-processing

The palm region of interest (RoI) is extracted from hand images by locating landmarks using the Mediapipe library, followed by geometric normalization to compensate for variations in palm positioning, orientation, and scale during contactless acquisition, thereby ensuring consistent alignment across all samples. Normalized RoI images are resized to square images sized (512×512) and converted to grayscale, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the principal line patterns and fine wrinkle structures.

B. DTCWT-Based Multiscale Feature Representation

The DTCWT with a 14-Tap Q-shift filter bank extracts wavelet coefficients from the normalized and enhanced palm images. The choice of the 14-Tap Q-shift filter bank is due

Diagram.png

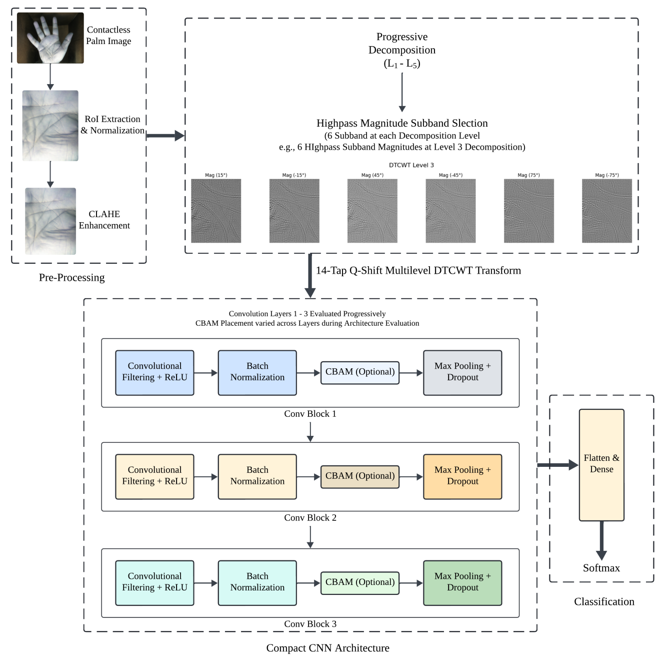

Fig. 1. Overall Framework of the Proposed System

to its ability to extract better directional features while maintaining moderate computational overhead compared to smaller length filter banks [8], [9], [10]. The transform employs two parallel real filter bank trees to generate complex coefficients, expressed as

$$d_c(l, n) = d_r(l, n) + j\, d_i(l, n), \tag{1}$$

where $d_r(l, n)$ and $d_i(l, n)$ denote the real and imaginary coefficients at decomposition level l and spatial location $n = (x, y)$. For each level, six directional complex subbands are produced corresponding approximately to $\pm 15°$, $\pm 45°$, and $\pm 75°$. Magnitude and phase representations are computed as

$$d_c(l, n) = \sqrt{d_r^2(l, n) + d_i^2(l, n)} \quad \angle \tan^{-1}\left(\frac{d_i(l, n)}{d_r(l, n)}\right). \tag{2}$$

Preliminary experiments indicated a limited contribution from the phase subbands, whereas magnitude subbands effectively capture discriminative palmprint features: principal lines and wrinkles (as evident in Fig. 2). Therefore, only magnitude subbands are retained for further experimentation. Low-frequency approximation coefficients are excluded from the experiments. After each decomposition scale, the spatial resolution of the feature maps is reduced by a factor of 2, but resized to (128×128) using bilinear interpolation to meet the fixed spatial resolution requirement of CNN.

C. Model Architecture and Configuration

The feature maps obtained from the magnitude subbands by DTCWT decomposition are concatenated along the channel dimension for each image, resulting in an n-channel representation, where n is given by $6 \times S$ with S as the number of decomposition scales (e.g., Level 3 yields $n =$

18). Consequently, the CNN input has a spatial resolution of $(128{\times}128{\times}n)$, in contrast to the original grayscale input of size $(128{\times}128{\times}1)$. These representations are evaluated using CNN architectures of increasing depth: one-layer, two-layer, and three-layer variants.

Each convolution block comprises a $(3{\times}3)$ convolution kernel (same padding, unit stride), followed by rectified linear unit (ReLU) activation, batch normalization (momentum = 0.99, ϵ = 0.001), L1–L2 regularization (L1 = $1{\times}10^{-5}$, L2 = $1{\times}10^{-4}$), max pooling with stride 2, and progressively increasing dropout. The number of filters increases across the hidden layers: 16, 32, and 64 for the first, second, and third layers, respectively. The CBAM is inserted after batch normalization and before max pooling in selected layers.

For an intermediate feature map $F \in \mathbb{R}^{H\times W\times C}$, CBAM sequentially applies channel attention and spatial attention as

$$F' = M_C(F) \odot F, \qquad F'' = M_S(F') \odot F', \tag{3}$$

where $M_C(\cdot)$ and $M_S(\cdot)$ represent channel and spatial attention maps, respectively, and $\odot$ denotes element-wise multiplication. Channel attention emphasizes discriminative directional subbands, and spatial attention localizes discriminative palm regions.

A flatten layer with dropout is used in the classification head, followed by a fully connected softmax layer with output neurons equal to the number of classes.

D. Training Setup and Protocol

Experiments begin with a Level 1 decomposition, followed by progressive exploration of CNN depth and selective placement of CBAM. In particular, the following configurations are evaluated: single-layer CNN with and without CBAM; CBAM applied at the first layer, the second layer, or omitted entirely for the two-layer CNN; CBAM applied at individual layers, pairwise combinations, and no attention for the three-layer CNN. Similar approaches are followed for higher decomposition scales up to Level 5.

Training uses the Adam optimizer (initial learning rate of 10^{-3}) with categorical cross-entropy loss and accuracy as the primary metric. A ReduceLROnPlateau learning rate scheduler (factor = 0.5, patience = 7, minimum learning rate 10^{-6}) is employed to improve convergence. Models are trained for 50, 100, and 150 epochs.

All configurations are evaluated using validation and test accuracy, classification metrics, trainable parameter count, and model size. In this study, economy refers to the computational efficiency of the proposed approach while maintaining competitive recognition performance. It is quantified through trainable parameter count and model size, which reflect model complexity and memory requirements. The objective of the study is not merely to achieve the highest recognition performance, but to determine the minimum structural complexity required to achieve stable classification under moderate-data conditions.

III. Results and Discussion

All variants of the proposed framework are evaluated separately for the left and right palm samples using test accuracy (Test Acc), precision (Prec), recall (Rec), F1-score (F1), area under the receiver operating characteristic curve (AUROC), area under the precision–recall curve (AUPR), and equal error rate (EER), approximate number of trainable parameters (Params, in millions) and model size (Size, in MB). All experiments are repeated with different random splits, and consistent trends are observed. Tables I and II report the performance results of the top three performing configurations for each decomposition scale (Level) for the left and right palms, respectively.

A. Dataset Organization

Experiments in the study use the IITD contactless palmprint dataset [15], which comprises palm images from 230 individuals. Each person has provided mostly 5 samples of their left and right hands under a contactless acquisition setup. The dataset contains 2601 RGB hand images in total, each with $(1600{\times}1200{\times}3)$ size, stored at 180 dots per inch (dpi) resolution [15], [16].

To simulate practical constraints, a single random RoI sample per subject (separately for the left and right) is used for training, augmented with controlled perturbations (scaling, Gaussian blurring, brightness variation, and random noise). Each subject contributes one original and two augmented images for training, while the remaining augmented samples are split into validation and test sets, resulting in a 60:20:20 split. The chosen augmentations are to represent real-time deployment of a biometric system, where enrollment samples are collected in a controlled environment, while query samples may have acquisition perturbations.

B. Performance Across Progressive DTCWT Decomposition Levels

Relatively weak performance is observed at Level 1 decomposition across all configurations due to limited directional detail and mostly coarse structural information. The performance improves significantly at Level 2, while it nearly saturates at Level 3, achieving about 99% accuracy and very low ERR on palm samples from both left and right hands. Decomposition at Levels 4 and 5 shows only slight improvements or no reliable gains. This trend suggests that intermediate decomposition depth captures sufficient discriminative directional information, making deeper decomposition unnecessary under practical constraints. Representative DTCWT feature maps at different decomposition scales for a particular image are shown in Fig.2.

C. Effect of CNN Depth and CBAM Placement

Single-layer CNN variants are the least effective models, indicating their inefficiency at learning the MMR representation. Two-layer CNN variants show their efficacy over single-layer models, with significant gains in recognition performance. Three-layer CNN models outperform the other two, which can be attributed to their larger receptive fields and stronger ability to integrate local directional responses into more discriminative feature representations.

TABLE I

TOP THREE CONFIGURATIONS FOR LEFT PALM SAMPLES ACROSS DECOMPOSITION LEVELS

Dec Level	Configuration	Test Acc (%)	Prec	Rec	F1	AUROC	AUPR	EER (%)	Params (M)	Size (MB)
L1	2L-CNN + CBAM(2)	73.91	0.66	0.74	0.69	0.85	0.79	13.04	7.54	28.8
	3L-CNN + CBAM(1,3)	74.78	0.68	0.75	0.71	0.87	0.80	8.70	3.79	14.5
	3L-CNN + CBAM(2,3)	75.22	0.68	0.75	0.70	0.88	0.80	8.19	3.79	14.5
L2	3L-CNN + CBAM(2,3)	93.04	0.90	0.93	0.91	1.00	0.998	1.74	3.80	14.5
	3L-CNN + CBAM(3)	94.78	0.93	0.95	0.94	1.00	0.998	0.43	3.80	14.5
	3L-CNN + CBAM(1,3)	95.22	0.94	0.95	0.94	1.00	0.998	0.56	3.80	14.5
L3	3L-CNN + CBAM(2,3)	99.13	0.99	0.99	0.99	1.00	1.00	0.061	3.80	14.5
	3L-CNN + CBAM(1,3)	99.13	0.99	0.99	0.99	1.00	1.00	0.065	3.80	14.5
	3L-CNN (No CBAM)	99.13	0.99	0.99	0.99	1.00	1.00	0.019	3.79	14.5
L4	3L-CNN + CBAM(2)	100.00	1.00	1.00	1.00	1.00	1.00	0.00	3.80	14.5
	3L-CNN + CBAM(2,3)	100.00	1.00	1.00	1.00	1.00	1.00	0.00	3.80	14.5
	2L-CNN + CBAM(2)	100.00	1.00	1.00	1.00	1.00	1.00	0.00	7.55	28.8
L5	3L-CNN + CBAM(1,3)	98.26	0.98	0.98	0.98	1.00	1.00	0.030	3.80	14.5
	2L-CNN + CBAM(2)	99.57	0.99	1.00	0.99	1.00	1.00	0.013	7.55	28.8
	3L-CNN + CBAM(2,3)	100.00	1.00	1.00	1.00	1.00	1.00	0.004	3.80	14.5

TABLE II

TOP THREE CONFIGURATIONS FOR RIGHT PALM SAMPLES ACROSS DECOMPOSITION LEVELS

Dec Level	Configuration	Test Acc (%)	Prec	Rec	F1	AUROC	AUPR	EER (%)	Params (M)	Size (MB)
L1	3L-CNN + CBAM(1,3)	65.22	0.55	0.65	0.59	0.81	0.72	11.27	3.79	14.5
	3L-CNN + CBAM(2,3)	65.22	0.54	0.65	0.58	0.80	0.71	13.55	3.79	14.5
	3L-CNN + CBAM(3)	69.13	0.60	0.69	0.64	0.86	0.75	6.52	3.79	14.5
L2	3L-CNN + CBAM(1,3)	92.61	0.90	0.93	0.91	1.00	0.994	1.26	3.80	14.5
	3L-CNN + CBAM(3)	94.35	0.92	0.94	0.93	1.00	0.998	0.81	3.80	14.5
	2L-CNN + CBAM(2)	94.35	0.93	0.94	0.93	1.00	0.998	0.70	7.54	28.8
L3	3L-CNN + CBAM(1,3)	99.57	0.99	1.00	0.99	1.00	1.00	0.006	3.80	14.5
	3L-CNN + CBAM(3)	99.57	0.99	1.00	0.99	1.00	1.00	0.010	3.80	14.5
	3L-CNN (No CBAM)	100.00	1.00	1.00	1.00	1.00	1.00	0.00	3.79	14.5
L4	3L-CNN + CBAM(2,3)	99.57	0.99	1.00	0.99	1.00	1.00	0.006	3.80	14.5
	3L-CNN (No CBAM)	99.57	0.99	1.00	0.99	1.00	1.00	0.361	3.80	14.5
	3L-CNN + CBAM(2)	100.00	1.00	1.00	1.00	1.00	1.00	0.00	3.80	14.5
L5	2L-CNN (No CBAM)	98.70	0.98	0.99	0.98	1.00	1.00	0.053	7.55	28.8
	2L-CNN + CBAM(1)	99.13	0.99	0.99	0.99	1.00	1.00	0.344	7.55	28.8
	2L-CNN + CBAM(2)	99.13	0.99	0.99	0.99	1.00	1.00	0.283	7.55	28.8

TABLE III

BASELINE CONFIGURATION (WITHOUT DTCWT) COMPARISON USING GRAYSCALE IMAGES AS INPUT ($128\times128\times1$)

Palm Side	Configuration	Test Acc (%)	Prec	Rec	F1	AUROC	AUPR	EER (%)	Params (M)	Size (MB)
Left	3L-CNN (No CBAM)	95.65	0.94	0.96	0.94	0.99	0.98	1.30	3.80	14.5
	3L-CNN + CBAM(2,3)	98.26	0.97	0.98	0.98	0.99	0.98	0.87	3.80	14.5
Right	3L-CNN (No CBAM)	95.65	0.94	0.96	0.94	0.99	0.98	1.70	3.79	14.5
	3L-CNN + CBAM(2,3)	95.22	0.93	0.95	0.94	0.99	0.98	0.87	3.80	14.5

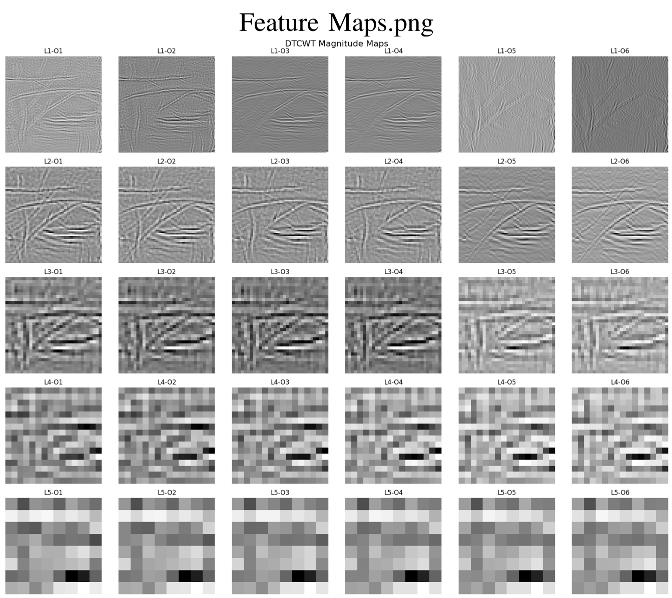

Fig. 2. DTCWT Feature Map Visualization at Each Scale, Level 1: Row 1, Level 2: Row 2, Level 3: Row 3, Level 4: Row 4, Level 5: Row 5

CBAM applied to deeper convolutional blocks consistently outperforms shallow-stage attention, suggesting that attention becomes more effective after sufficient feature abstraction. For left palm samples, CBAM placed at the second and third hidden layers provides strong performance from Level 3 onward. For right palms, CBAM is not always necessary at intermediate depth (Level 3), but contributes to improved stability at higher decomposition levels, indicating a dependency on both feature abstraction and transform depth. Representative CNN feature maps at different hidden layers are shown in Fig. 3.

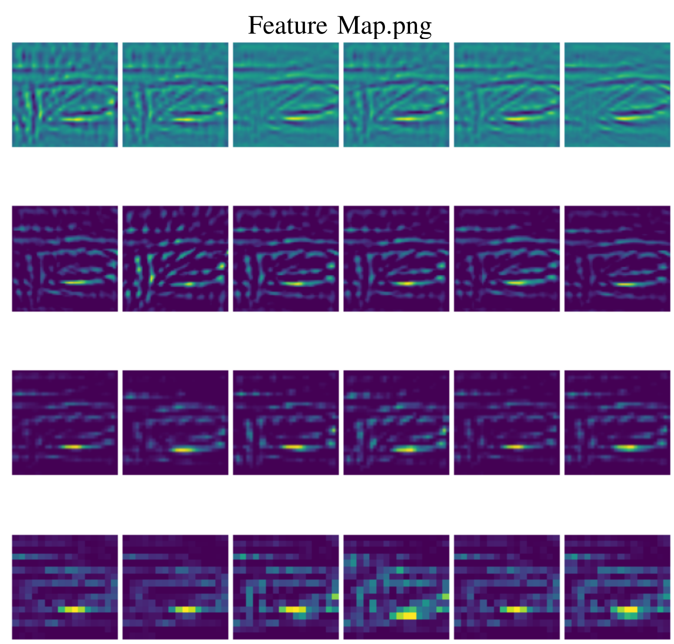

Fig. 3. Feature Maps Hidden Layers of CNN, Row 1: Convolution Layer 1 Output, Row 2: CBAM Output at Hidden Layer 2, Row 3: CBAM Output at Hidden Layer 3, Row 4: Output Layer

D. Baseline Configurations

To study the role of an MMR representation of the input, experiments are conducted using a baseline three-layer CNN with identical complexity without attention and with attention applied at deeper layers. Baseline three-layer CNNs trained on CLAHE-enhanced grayscale inputs show lower performance despite identical complexity, as evident from Table III. Adding the CBAM block provides a marginal improvement in the spatial domain. This confirms that frequency domain feature representation learning improves performance and plays a vital role in achieving an optimal economy-accuracy trade-off under moderate-data conditions. This trend is consistent across both left and right palm samples.

E. Recognition Saturation and Economy Tradeoff

At intermediate decomposition scales (Level 3), recognition performance saturates and provides only marginal gains beyond that point under controlled perturbations, indicating dataset-specific saturation rather than improved generalization. Consequently, a compact three-layer CNN with CBAM applied at the deeper layers configuration represents the most practical operating point, as it achieves the best accuracy-economy trade-off among the evaluated models. Compared with existing methods on the IITD contactless palmprint dataset [15], as presented in Table IV, the proposed framework achieves an EER of 0.06% with 3.8 million parameters, demonstrating a noticeable reduction in model complexity while maintaining competitive recognition performance.

TABLE IV
COMPARISON WITH EXISTING PALMPRINT RECOGNITION METHODS ON IITD CONTACTLESS PALMPRINT DATASET

Method	**Approach**	**EER (%)**	**Params (M)**
SAC-Net [18]	DL	0.25	8–10
CO3Net [17]	DL	0.47	7–9
CompNet [19]	DL	0.54	5–7
EE-PRNet [3]	DL	1.02	25
RFN [7]	DL	2.10	10–20
DDR [6]	DL	2.37	12–14
PalmNet [5]	DL	4.20	1–2
Baseline	DL	0.87	3.80
Proposed Method	**MMR+DL**	**0.06**	**3.80**

IV. CONCLUSION

The study demonstrates that the MMR feature representation effectively reduces the CNN's learning burden by capturing discriminative information in the frequency domain. Consequently, compact shallow architectures achieve a favorable trade-off between recognition performance and computational economy, particularly under data-scarce conditions. The results further establish that the wavelet transform decomposition depth and selective attention placement jointly govern model optimality. An intermediate decomposition depth combined with a compact CNN and a deeper-stage attention is identified as the most effective configuration under the curated experimental setup, providing a practical design guideline for efficient biometric recognition. While the proposed framework achieves promising results, the experiments are limited to a moderate-sized contactless palmprint dataset. Although repeated evaluations across multiple random splits yielded consistent performance trends, the possibility of dataset-specific learning cannot be completely ruled out. Future work will extend the framework to larger datasets and additional biometric modalities to further assess its robustness, scalability, and generalization capability.

ACKNOWLEDGMENT

The authors acknowledge the support of the Anusandhan National Research Foundation (ANRF) through the Partnerships for Accelerated Innovation and Research (PAIR) Project at IIT Bombay (RD/0125-ANRFPVG-001). The authors also thank the Department of Electrical Engineering, IIT Bombay, and SVKM's NMIMS School of Technology Management and Engineering, Navi Mumbai, for providing institutional support and research facilities.

REFERENCES

[1] N. Amrouni, A. Benzaoui, and A. Zeroual, "Palmprint Recognition: Extensive Exploration of Databases, Methodologies, Comparative Assessment, and Future Directions," *Applied Sciences*, vol. 14, no. 1, p. 153, 2023.

[2] M. Izadpanahkakhk, S. M. Razavi, M. Taghipour-Gorjikolaie, S. H. Zahiri, and A. Uncini, "Deep Region of Interest and Feature Extraction Models for Palmprint Verification using Convolutional Neural Networks Transfer Learning," *Applied Sciences*, vol. 8, no. 7, p. 1210, 2018.

[3] W. M. Matkowski, T. Chai, and A. W. K. Kong, "Palmprint Recognition in Uncontrolled and Uncooperative Environment," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1601–1615, 2019.

[4] T. Chai, S. Prasad, and S. Wang, "Boosting Palmprint Identification with Gender Information using DeepNet," *Future Gener. Comput. Syst.*, vol. 99, pp. 41–53, 2019.

[5] A. Genovese, V. Piuri, K. N. Plataniotis, and F. Scotti, "PalmNet: Gabor-PCA Convolutional Networks for Touchless Palmprint Recognition," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3160–3174, 2019.

[6] S. Zhao and B. Zhang, "Deep Discriminative Representation for Generic Palmprint Recognition," *Pattern Recognit.*, vol. 98, p. 107071, 2020.

[7] Y. Liu and A. Kumar, "Contactless Palmprint Identification using Deeply Learned Residual Features," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 2, pp. 172–181, 2020.

[8] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, 2005.

[9] N. G. Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, 2001.

[10] F. Yeganli, "A Design of Q-shift Filter for Dual-Tree Complex Wavelet Transforms," Ph.D. dissertation, Eastern Mediterranean Univ., 2010.

[11] J. Ren, C. Hu, Z. Shang, Y. Li, Z. Zhao, and R. Yan, "WavFormer: An Interpretable Wavelet-constrained Transformer for Industrial Acoustics Diagnosis," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, 2024, pp. 1–6.

[12] J. Ren, C. Hu, Z. Shang, Y. Li, Z. Zhao, and R. Yan, "Learning Interpretable and Transferable Representations via Wavelet-constrained Transformer for Industrial Acoustic Diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, p. 3512312, 2025.

[13] D. Kumar, G. Parasnis, R. Bhope, A. Chokshi, V. Jain, D. Jadav, V. Gadre, B. Banerjee, and A. Chokhani, "Economy and Explainability with Non-separable Wavelet-based CNN Experiments," in *Proc. Int. Conf. Comput. Vis. Image Process.*, 2024, pp. 266–280.

[14] S. Pateriya, V. Ghule, D. S. Jadav, D. Kumar, and V. M. Gadre, "SSNet: Novel Approach for Fingerprint Recognition in Data-scarce Scenarios," *Authorea Preprints*, 2024.

[15] A. Kumar, "IIT Delhi Palmprint Image Database Version 1.0," [Online]. Available: IITD Palmprint Database

[16] A. Kumar, "Incorporating Cohort Information for Reliable Palmprint Authentication," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, 2008, pp. 583–590.

[17] Z. Yang, W. Xia, Y. Qiao, Z. Lu, B. Zhang, L. Leng, and Y. Zhang, "CO3Net: Coordinate-aware Contrastive Competitive Neural Network for Palmprint Recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.

[18] C. Gao, Z. Yang, M. Zhu, and A. B. J. Teoh, "Scale-aware Competition Network for Palmprint Recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 4580–4584.

[19] X. Liang, J. Yang, G. Lu, and D. Zhang, "CompNet: Competitive Neural Network for Palmprint Recognition using Learnable Gabor Kernels," *IEEE Signal Process. Lett.*, vol. 28, pp. 1739–1743, 2021.