A Hybrid CNN–ViT Attention-Based Deep Learning Framework for Robust Fingerprint Image Classification

[1]Udit Chauhan, [2]Praveen Kumar Mohane
[1]M. Tech Scholar, Department: Computer science engineering, Millennium Institute of technology
[2]Assistant professor, Department: Computer science engineering, Millennium Institute of technology
[1]uditch94@gmail.com , [2]pkmohane9910@gmail.com

* Corresponding Author: Udit Chauhan

Abstract:

Fingerprint classification becomes central to biometrics authentication and forensic identification. Since poor-quality images, imbalanced classes, and complicated ridge patterns plague conventional classification methods, the conventional classification models' accuracy is affected. To overcome this, a hybrid deep learning framework is proposed, combining CNNs with Vision Transformers and an integrated attention mechanism to improve the fingerprint image classification performance. For local spatial feature extraction, CNNs are used, and for capturing long-range global dependencies that exist within fingerprint patterns, ViTs are used. Finally, embedding the attention module helps the model to concentrate on highly discriminative features that increase model interpretability and reduce misclassification. Data augmentation considered rotation, scaling, and shifting to instill insensitivity and diversity into the data, thus instilling robustness and broader generalities in the model. The hybrid architecture benefits from the complementary advantages of CNNs and ViTs and at the same time promotes computational efficiency. The evaluation of the proposed model using comprehensive performance measures such as accuracy, precision, recall, and ROC-AUC assures applicability in real-time biometric systems. Experimental evidence confirms that such an integrated method greatly surpasses conventional CNN-only as well as pure ViT classifiers, especially with complex, noisy fingerprint datasets. The incorporation of data augmentation techniques, including rotation, scaling, and shifting, addresses challenges such as class imbalance and data variability. With an accuracy of up to 91.15% after 30 epochs, the model proves effective across various fingerprint datasets (DB1, DB2, DB3, DB4), showing its potential for real-time biometric identification and security applications. The performance metrics, such as precision, recall, F1-score, and ROC-AUC, confirm that the model offers a reliable solution for high-accuracy classification tasks.

Keywords: *Latent Fingerprint, Minutiae Detection, Deep Learning, CNN-ViT Hybrid, Attention Mechanism, Biometric Recognition*

I. INTRODUCTION

Fingerprint recognition remains a vital authentication method with biometric application since the principle of uniqueness offers permanence applicable in many areas such as law enforcement, border security, or mobile device access. If we consider biological traits, fingerprints are highly regarded because their fine-grained features, including ridges, bifurcations, and minutiae, bear tremendous discriminatory capabilities [1]. However, intra-class variations, noisy acquisitions, partial prints, and less than clear impressions continue to be disadvantages for traditional fingerprint classification systems, particularly when latent or distorted samples are involved. Standard approaches mostly based on extracting handcrafted features and rule-based classifiers are neither adapting nor robust enough to cope with such complicated variations presented in fingerprint data [2].

Such limitations are therefore addressed with the deep learning approach and hence have been widely studied in recent years. Convolutional neural networks in particular efficiently learn localization operators, spatial hierarchies, or scale from fingerprint images. CNN architectures are capable of extracting strong features, e.g., ridge flows and minutiae points. They are, however [3], limited in extracting long-range dependency and global relationship-the very features required-to help with the holistic understanding of a complicated fingerprint class.

ViTs have been developed as powerful alternatives to bridge this chasm. Originally meant for natural image classifications, ViTs better model the global attention and spatial dependencies through self-attention mechanisms. This is a huge advantage while handling non-local context, which complements the local feature extraction operation of CNNs [4]. However, ViTs alone might still not do very well on smaller datasets or in cases where fine-grained local information is important, such as fingerprint classification.The research works on a hybrid deep learning model that combines CNN, ViT, and an attention mechanism into a single framework. The CNN component is set to generate fine-level spatial features while the ViT component is responsible for global representation. An attention module is also added to make the identification of informative regions present in the fingerprint image more dynamic, such as ridge patterns or clusters of

minutiae [5]. This hybrid setting ensures that both local details and global contexts are learned and applied for classification tasks.

In addition, it applies data augmentation techniques such as image rotation, translation, and scaling to address class imbalance and limited training sets. These methods ensure that the model improves its generalization and robustness across varying fingerprint classes and acquisition conditions [6]. Training and evaluation happened on various publicly available fingerprint datasets (DB1, DB2, DB3, and DB4), each corresponding to differentiated fingerprint classes and complexities. The proposed hybrid model shows much improved accuracy in classification, by up to 91.15%, measured after 30 epochs of training. Other metrics such as precision, recall, F1-Score, and ROC-AUC established that it performs more reliably with consistent classification.

The present study underscores the feasibility of fusing CNNs and ViTs for fingerprint classification and indicates that the proposed hybrid architecture can be utilized in real-world biometric systems. Having the ability to generalize over multiple types of fingerprints, the model can also be employed in forensic identification and mobile authentication scenarios [7]. Furthermore, this paves the way for the fruitful investigation into multi-modal biometric systems, where fingerprint data could be paired with those of other biometric traits such as iris or face, toward a higher level of security and identification precision [8]. Fig. 1 describes enhancing fingerprint matching with deep learning.

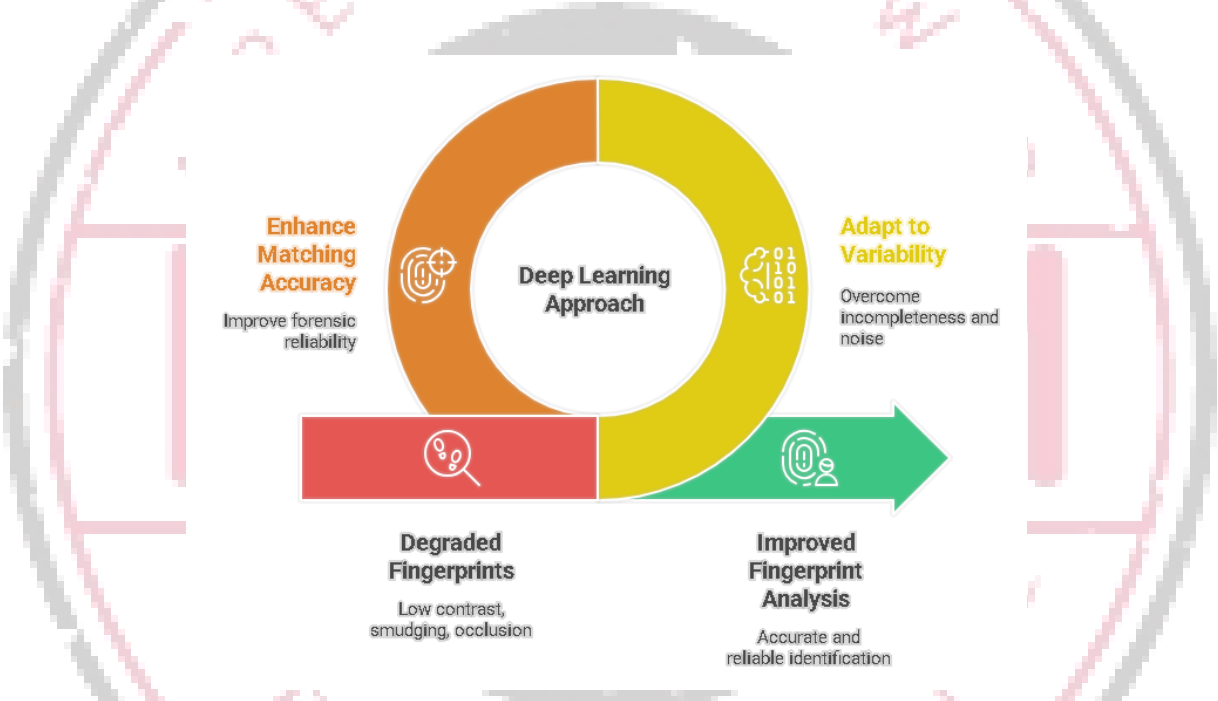

Fig. 1: Enhancing Fingerprint Matching with Deep Learning

A. Deep Learning in Fingerprint Recognition

Deep learning has revolutionized fingerprint recognition by providing powerful methods for modelling intricate fingerprint patterns and structures. Contrary to traditional algorithms wherein features are handcrafted, features here are discriminatively learned by deep neural networks from the raw input data [9]. CNNs are good at capturing local spatial information, such as ridge structures and minutiae, useful in high-resolution fingerprint classification. CNNs, however, may have limitation in grasping long-range dependencies and global contexts, especially for partial or latent prints. ViTs, conversely, build global relations using self-attention mechanisms, allowing for comprehension of the fingerprint topology in a more holistic manner. Coupled with attention modules, these models learn to weigh relevant regions dynamically, thus improving classification performance [10]. Fig .2 describes deep learning in fingerprint recognition.

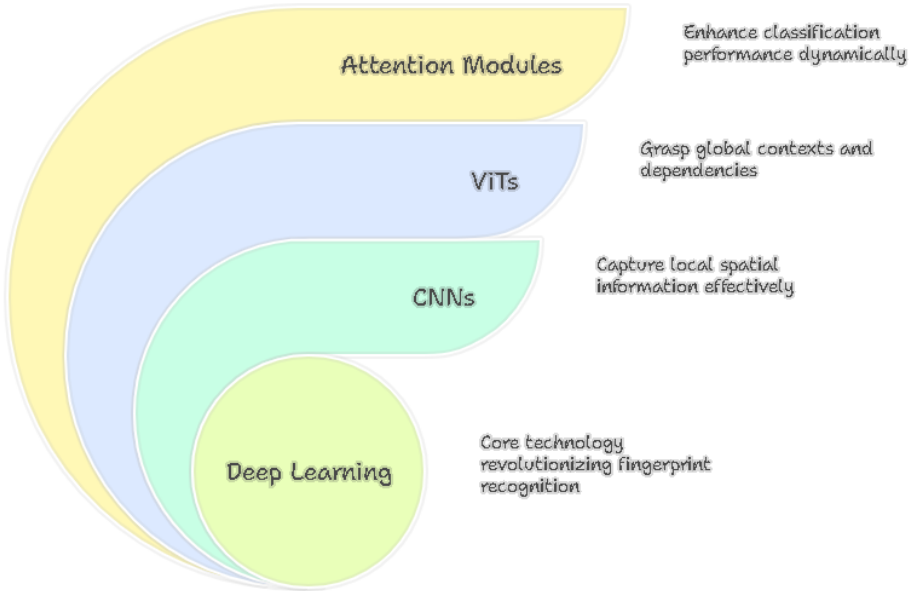

Fig .2: Deep Learning in Fingerprint Recognition

II. LITERATURE REVIEW

Abdul Wahab et al. [1] (2024) uses a GAN-based enhancement technique that integrates information about minutiae location and orientation fields to best enhance latent fingerprint clarity and ridge preservation. It is computationally expensive and very hard to balance the realism of the fingerprint with accurate feature preservation.

Temirlan Meiramkhanov et al. [2] (2024) Combined CNNs with Gabor filter enhancement techniques to improve recognition accuracy on manipulated fingerprint impressions to 94% with the Sokoto Coventry dataset. Low generalization capacity across different fingerprint kinds; heavy reliance on dataset-specific tuning.

Milind B. Bhilavade et al. [3] (2024) Compared matching scores for relatively poor fingerprint images reconstructed by conventional minutiae-based methods and deep learning, varying between 23–94% (DL) and 82–99.99% (minutiae-based). Deep learning methods performed inconsistently with different types of damage and considering the image quality.

Hongtian Zhao et al. [4] (2024) ResNet with Generalized IoU-based NMS for outlier-resistant minutiae extraction, outperforming state-of-the-art methods over the NIST SD4 and FVC2004 datasets. The performance depends on inference accuracy, with very large annotated datasets.

Sahar A. El_Rahman et al. [5] (2024) Presented CNN-based fingerprint unimodal and ECG-fingerprint multimodal systems, the sequential fusion ones yielding the highest AUC (0.99). High complexity in managing multimodal data and training large-scale fusion models.

A. A. Mulay et al. [6] (2024) Used ensemble of minutiae configurations with U-Net and ViT, gaining around 1.7% higher accuracy on challenging datasets such as NIST SD302. Slight performance gains with increased model complexity via ensemble strategies.

T. Kavitha et al. [7] (2024) Comprises an automated fingerprint recognition system for forensic crime detection with CNNs, having an accuracy exceeding 81%.Limited performance due to small data size and availability of better preprocessing techniques for noisy inputs.

P. Khare et al. [9] (2024) Introduced YOLO-based fingerprint recognition models trained on 4,000 annotated images, which raised mAP@0.5 from 93% to 97.4%. Accuracy depends heavily on annotated data quality; due to small datasets, it is difficult to generalize well.

Zexi Jia et al. [9] (2024) Finger Recovery Transformer (FingerRT) is designed to restore degraded or partial prints by harnessing the powers of Vision Transformers and enhancement networks. More computationally expensive and sensitive to segmentation errors during preprocessing.

S. Kriangkhajorn et al. [10] (2024) A frequency-domain latent print restoration framework using deep learning filter predictors was proposed to increase rank-1 identification accuracy. Complicated block partition; performance degrades strongly with the initial filter quality and level of degradation.

Z. Pan et al. [11] (2024) Dense Minutia Descriptor was developed by using deep learning concepts to encode 3D minutiae patches for precise latent matching. One limitation involved being exhaustive and computationally expensive during matching, while also having difficulties with overlapping and noisy backgrounds.

N. Bhargava et al. [12] (2024) Employed image processing followed by skeletonization in order to form bit-string-encoded fingers for easy storage and matching It is not robust to heavy noise and may discriminate against partial and distorted inputs.

R. Bano et al. [13] (2024) Approach to AFIS classification proposed by deep learning using features such as hand orientation and sweat pore patterns. The approach is far too dependent on publicly available datasets, which might not be diverse enough or representative of the true forensic variance.

Yusuf Artan et al. [14] (2024) Developed a fusion-based local matching technique that integrates handcrafted features with deep embeddings for latent print recognition. Increased complexity is brought about by the fusion process, contributing to more dependency on various stages of accurate feature extractions and, thus, more overhead for the system.

The study presented by Saket Pateriya et al. [15] (2024) propose the use of the scattering transform with the Shearlet Network (SSNet) to extract the fingerprint features with maximum robustness, and then a score-level fusion scheme is used for higher authentication accuracy. May face difficulty adapting to highly distorted or occluded fingerprints in real-time environments.

Yuhang Qiu et al. [16] (2024) proposes IFViT, a two-stage framework for accurate and interpretable fingerprint matching. It performs dense fingerprint alignment using a Siamese ViT and then extracts fixed-length, interpretable representations via retrained ViTs with a fully connected layer. Experiments on public datasets demonstrate improved matching performance and interpretability. Depends on large training data and may involve heavy computational resources during deployment.

The general analysis of fingerprint verification and forgery detection with deep learning and machine learning techniques, by M. Genel, et al. [17] (2024), is facing new security threats such as fake fingerprint attacks. Given the SOCOfing dataset, models were implemented under various configurations and hyperparameter settings for performance evaluation. Comparative results, showing the pros and cons of each method, enabled the selection of the best-performing models. Limited scope in real-time applications and relies on dataset-specific spoofing patterns.

For detecting spoofs in contactless fingerprint systems, which are increasingly being developed because of their convenient and hygienic benefits, Kanchana Rajaram et al. [18] (2024) propose CLNet, a deep learning approach. Existing methods for spoof detection often work with a limited set of features, resulting in low accuracy. Trained on the newly created S-CLAF dataset, CLNet achieved accuracy on the order of 99.07%, and it also provided high results on generalization: 98.32% on LivDet 2015 and 99.38% on the IIT Bombay dataset, bettering the results of the current state-of-the-art approaches. Performance may degrade when exposed to unseen spoofing techniques or poor lighting conditions.

H. M. Mishra and fellow analysts in 2024 [19] worked in Minutia-based mapping and Convolutional Neural Networks (CNNs), whereby deep learning means go into fingerprint matching to further its accuracy and speed. The uniqueness and permanence of fingerprints give these methods a great deal toward being implicated in modern biometric authentication and forensic investigations. Some conventional enhancement techniques may be ineffective on extremely noisy or partial prints.

A. Nóbrega et al. [20] (2024) provides an interesting possibility to produce more efficient minutiae descriptors for latent fingerprint identification, without needing private datasets. Experiments on NIST SD27 confirm an increase in hit rate of 6.59% over commercial tools, thus validating the strength of self-supervision and data augmentation methods for latent fingerprint recognition.Synthetic data may not fully capture real-world latent fingerprint distortions.

Abdulrasool Jadaan Abed – et al. [21] (2024) proposed a fingerprint identification approach using deep learning specifically geared toward low-quality fingerprint images. Experimental results demonstrate that this approach is more accurate and more robust under adverse conditions than conventional local minutia methods. This work stresses that there remains a need to innovate in biometric systems and suggests that fusion of multimodal biometrics may improve the performance

further when coming to real applications. Performance may drop without pre-enhancement or on highly distorted fingerprints.

In 2024, X. Guan et al. [22] proposed the PDRNet (Phase-aggregated Dual-branch Registration Network) to allow for enhanced dense registration of fingerprints by aligning pairs down to the pixel level. Extensive experiments across different datasets show that PDRNet truly meets the state-of-the-art accuracy and robustness paradigm, while still maintaining competitive efficiency in fingerprint registration. Complex architecture may hinder real-time deployment or mobile implementation.

The list of applications discussed by A. Juneja, et al. [23] (2024) is ever-changing. Recent developments in fingerprint recognition reviewed include contactless identification using CNNs. Identification may also follow a hygienic minutiae-based process. Indoor positioning systems rank individuals against radio signals using ML models: k-NN and SVM series. Browser fingerprinting passively recognizes users by identifying unique browser configurations.

D. Mari et al. [24] (2024) carry out the first deeper investigation into the suitability of learning-based image codecs such as JPEG-AI for storing fingerprint images, where compression artifacts might affect the extraction of biometric features. They do provide a 47.8% BD rate reduction and a +3.97 dB PSNR gain without forfeiting automatic identification accuracy and human readability. May still need optimization for edge devices with limited decoding capacity.

III. RESEARCH OBJECTIVES

- Develop a robust image classification system using a hybrid deep learning model combining CNN (Convolutional Neural Network) and ViT (Vision Transformer) with an Attention mechanism. The goal is to enhance classification accuracy by efficiently capturing both local and global features from fingerprint images.
- Address challenges such as class imbalance by applying data augmentation techniques like rotation, scaling, and shifting, which enrich the training dataset. This improves the model's ability to generalize across various image scenarios, thereby reducing misclassifications.
- Implement an integrated approach that uses CNN for spatial feature extraction, ViT for capturing long-range dependencies, and an Attention mechanism to refine focus on significant regions of the image, resulting in a more accurate and adaptive classification system.
- Evaluate the proposed model using multiple performance metrics, including accuracy, precision, recall, and ROC-AUC. These evaluations help assess the model's effectiveness in classifying fingerprint data and ensure reliable real-time decision-making support.

IV. PROPOSED METHODOLOGY

A. Dataset Description and System Configuration

Fingerprint images from public biometric databases, namely DB1, DB2, DB3, and DB4, have been used for the study. Each of these benchmark datasets is fairly popular among biometric researchers in evaluating fingerprint recognition and classification algorithms. These datasets provide a variety of fingerprint patterns, differing in image quality and acquisition conditions, so that the image classification model proposed can be trained under one condition, validated in another, and finally tested for performance under yet a third condition. The use of multiple datasets exemplifies the generalizability of the method and moves towards a more holistic evaluation against very diverse biometric conditions.
Cloud-Based Execution Environment: Google Colab
The project was implemented and executed on Google Colab, which functions as a cloud-hosted Jupyter notebook; it may be considered as a robust and scaled platform for deep learning research. Using Colab's GPU-accelerated infrastructure, the hybrid CNN + ViT could be efficiently trained and evaluated without the necessity of local computational resources.

- **Seamless Library and Dataset Integration:**

Colab supports all major deep learning and data science libraries such as TensorFlow, Keras, Scikit-learn, NumPy, OpenCV, and Matplotlib, therefore simplifying model development. Dataset access was integrated through the mounting of Google Drive, with which fingerprint image datasets can be loaded, pre-processed, and augmented rapidly; thereby implying that high-throughput experimentation was carried out entirely within the notebook.

- **Interactive Debugging and Real-Time Monitoring:**

The interface supported code execution on a step-by-step basis, an integral part of iterative debugging, hyperparameter adjustments, and pipeline refinement. In-line visualizations supported the real-time tracking of training metrics such as loss convergence and classification accuracy. Evaluation plots such as confusion matrices, ROC-AUC, and precision-recall

curves were also dynamically created, giving insightful information on the model behavior to help in decision-making through optimization during the entire development lifecycle.

B. Proposed Model

A hybrid deep learning architecture comprising Convolutional Neural Networks, Vision Transformers, and Self-Attention is thus proposed to improve the classification of fingerprints by modeling local texture features as well as global dependencies. These multi-branch setups improve the accrual of accuracy and can be scaled to fit more complex environments, making them highly capable of solving large-scale biometric problems.
The design implemented passes through a three-stage pipeline: data pre-processing, feature selection and optimization, and model training and hybridization. At first, Min-Max scaling is employed to normalize fingerprint images, and they are then resized to 80×80 to maintain consistency across the dataset. To augment data in the context of diversity generation and countering overfitting, augmentations such as rotations, flips, and affine transformations are applied. In the meanwhile, spatial filters, including Gaussian and median, are used to accentuate ridge-valley structures. With respect to the dataset, it is stratified into training, validation, and testing populations. During the feature selection phase, CNN layers extract hierarchical spatial features ranging from ridge flow to minutiae, and SMOTE handles class imbalance within fingerprint datasets (DB1–DB4) by producing synthetic samples. Next, the hybrid architecture combines the local feature extraction of CNN with ViT to capture global dependencies via Multi-Head Self-Attention on tokenized image patches. A channel-wise attention mechanism is then incorporated to refocus the attention onto the significant regions, such as cores and minutiae clusters. The Adam optimizer with adaptive learning rate serves to train the system, whereas a tailored categorical cross-entropy loss penalizes errors in classifying minority classes. Early stopping and model checkpointing are applied to encourage generalization and alleviate overfitting, thus forming a robust system for fingerprint classification.

Now, the three-stage pipeline consisting of data preprocessing, feature selection and optimization, and model training and hybridization is implemented to build the system. Fingerprint images are first normalized using Min-Max scaling and resized to 80×80 pixels for dataset consistency. Data augmentation methods, including rotations, flips, and affine transformations, are applied to improve diversity and reduce overfitting, while ridge-valley structures are enhanced with spatial filters such as Gaussian and median filters. The dataset is further stratified into three subsets: training, validation, and testing. During the feature selection phase, layers of CNN extract hierarchical spatial features ranging from ridge flow to minutiae, while SMOTE deals with class imbalances in fingerprint datasets (DB1–DB4) by generating synthetic samples. The hybrid approach, then, combines CNN-based local feature extraction with ViT for capturing global dependencies via Multi-Head Self-Attention on tokenized patches of image. Using channel-wise attention mechanisms, the focus is refined onto the important regions, such as cores and minutiae clusters. Through the Adam optimizer with an adaptive learning rate, the model is trained, and the loss function is a bespoke categorical cross-entropy penalizing mistakes in classifying minority classes. Early stopping with model check pointing ensures generalization while preventing overfitting, realizing a strong fingerprint classification system.Fig.3 shows flowchart of proposed methodology.

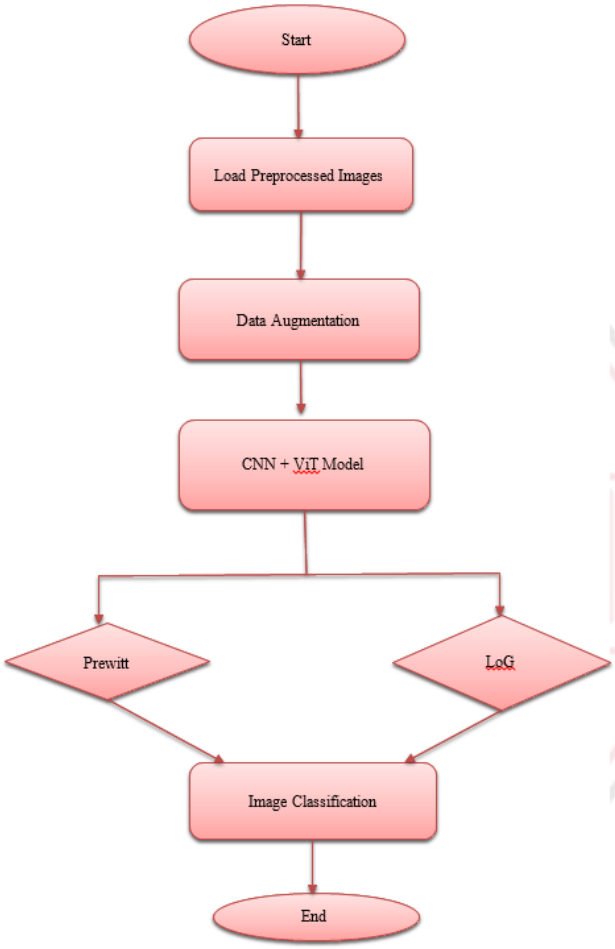

Fig. 3 Flow Chart of Proposed Model

C. Hybrid Model

- **CNN for Feature Extraction in Fingerprint Classification:**

In the first step of the proposed approach, CNNs act as the major classes for local feature extraction from the raw fingerprint images. CNNs are well-tightened bestowed to work on gridlike data such as images, and having spatial hierarchies and local dependencies are significant. Fingerprint patterns are inherently structured, comprising ridge endings, bifurcations, cores, and deltas, which must be accurately captured to differentiate among fingerprint classes (e.g., DB1–DB4).
At shallow layers, CNNs describe low-level features such as edges and gradients (e.g., Prewitt or Sobel-like filters), while deeper layers extract high-level semantic patterns such as ridge flow directions, ridge frequency, and minutiae constellations. These spatial features serve as a generous contrast to noise, illumination, and slight distortions, thus enabling high distinctiveness and robustness.
Also, to further improve fingerprint representation, convolution padding is used to maintain spatial resolution while dropout regularization prevents overfitting, especially with class-imbalance or limited-sample-diversity datasets.
This process of encoding hierarchical features serves as the foundational input for the ensuing Vision Transformer (ViT) and Attention modules with which global context modeling is performed. Extracted CNN feature maps are treated as spatial descriptors loaded with rich information on localized morphological characteristics of each fingerprint class. These descriptors serve to, on the one hand, speed up convergence in the transformer module and, on the other hand, guarantee that the hybrid model is able to take advantage both local and global fingerprint information.

- **ViT (Vision Transformer) for Global Feature Learning in Fingerprint Classification:**

Following early-stage localized feature extraction via CNN, intermediate spatial representations are fed into a ViT module to model global contextual dependencies present in the fingerprint images. In contrast with CNNs, which normally act over

local receptive fields within some window size, ViTs have a capacity to gather long-range interactions between image regions, thus being ideal for understanding global structure and holistic fingerprint patterning.
Provided the CNN feature maps, the ViT first splits them into a sequence of fixed-size image patches such as 16 × 16 (16×16 pixels), which are flattened and then linearly projected into a latent embedding space. Each of these patch embeddings is subsequently added to positional encodings to maintain spatial order, compensating for the absence of some inductive biases (locality, translation invariance) that are natively present within a CNN.

- **Hybrid Model: Merging CNN + ViT with Attention Mechanism for Fingerprint Classification:**

In the final stage, the proposed architecture performs feature fusion by combining the outputs of the CNN (which captures fine details of spatial hierarchies) and the Vision Transformer (ViT) (which encodes global contextual dependencies). The composite representation psi_g c is refined by considering channel-spatial Attention, which emphasizes the most discriminative regions and feature channels of the fingerprint image.
The CNN provides low-level features such as ridge orientations, minutiae points, and texture gradients, whereas ViT models spatial interrelationships between these features across image planes. The Attention module is key to contextually selecting features, providing adaptive weighting along spatial and channel dimensions by judging the importance of each activation map with respect to the classification goal.
This focusing ability emphasizes discriminating features such as ridge terminations, bifurcations, and rare structural features distinguishing one fingerprint class from another. The combined form of CNN–ViT–Attention pipeline is the key to building powerful representations that draw from both local details and global coherence.
The final fused feature vector is passed through a sequence of fully connected (dense) layers, culminating in a softmax output layer, which assigns the input to one of the predefined fingerprint classes (say: DB1, DB2, DB3, DB4). A custom loss function would be used-if needed-to relinquish an emphasis on an imbalanced nature of classes, such as weighted categorical cross-entropy or focal loss.

D. Evaluation Metrics

Accuracy

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{True\ positive + True\ Negetive}}{\boldsymbol{TotalNumber\ of\ Instance}} \quad (1)$$

Precision

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ positive}}{\boldsymbol{True\ Positive + False\ Positives}} \quad (2)$$

Recall (Sensitivity)

$$\mathbf{Recall} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{True\ Positives + False\ Negatives}} \quad (3)$$

F1-Score

$$\boldsymbol{F1 - Score} = 2 \times \frac{\boldsymbol{Precision \times Recall}}{\boldsymbol{Precision + Recall}} \quad (4)$$

Confusion Matrix

The confusion matrix provides detailed information on the classification accuracy by displaying true and false predictions for each fingerprint class (DB1 through DB4), thus revealing class-wise performances. It allows for the detection of misclassification trends, empowering us to improve the model in specific areas so as to increase security and accuracy in fingerprint verification.

ROC-AUC (Receiver Operating Characteristic - Area under Curve)

ROC Curve, each fingerprint class (DB1–DB4) maintains a diagonal line determined by mixing true positives with false positives for all thresholds. The higher the area under the curve, the better the discrimination power the classifier has; hence, the threshold can be selected on the basis of practical feasibility in the case of scenarios such as early alarms or missed detections.

V. RESULTS and DISCUSSION

An in-depth evaluation of the hybrid deep learning framework for fingerprint classification, identification, and verification involves CNNs employed for spatial feature extraction where local features such as ridges and bifurcations are captured, whereas ViT models are used for global context awareness through self-attention mechanisms across image patches. This canon of integration improves the classification and verification performances and is validated through taxing performance metrics like Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC to corroborate its discriminating ability on fingerprint samples from various datasets.

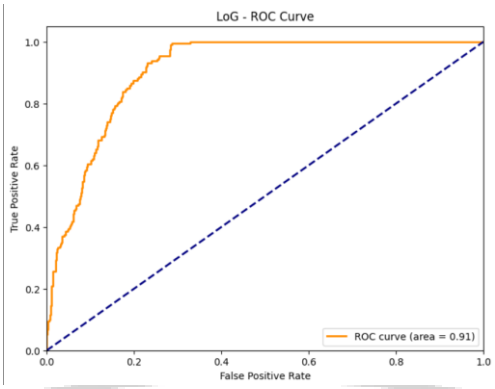

Fig. 4: LoG - ROC Curve

The model could benefit from further tuning in order to optimize its performance on the more challenging classes.The LoG ROC Curve in Fig. 4 shows an AUC of 0.91, which is indicative of good performance but not quite perfect. Given the stiff rise, the curve suggests that this model faces few challenges in distinguishing the different classes but should still ameliorate in reducing the false positives.

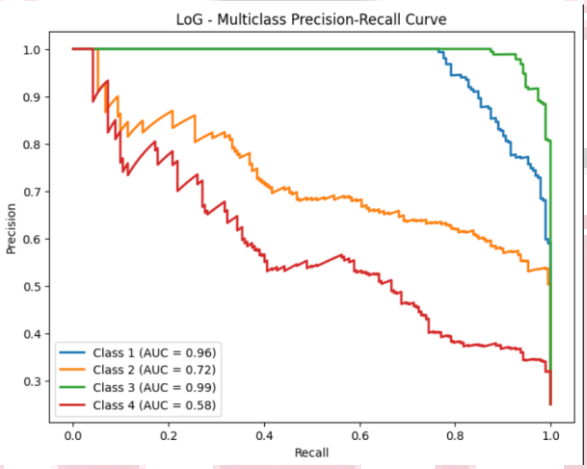

Fig. 5: LoG - Multiclass Precision-Recall Curve

The results show that while the model is great in identifying some classes, some classes, especially Class 4, may need either more data or better extraction techniques. The LoG Multiclass Precision-Recall Curve in Fig. 5 gives a comparative performance of the model for the different classes. Class 3 had a very high AUC of 0.99, meaning that the model is highly capable to discriminate when it comes to precision-recall trade-off. Class 4, on the other hand, received a poor AUC of 0.58, which means that the model really does not excel in dealing with this particular class. Classes 1 and 2 performed fairly well with AUCs of 0.96 and 0.72, respectively.

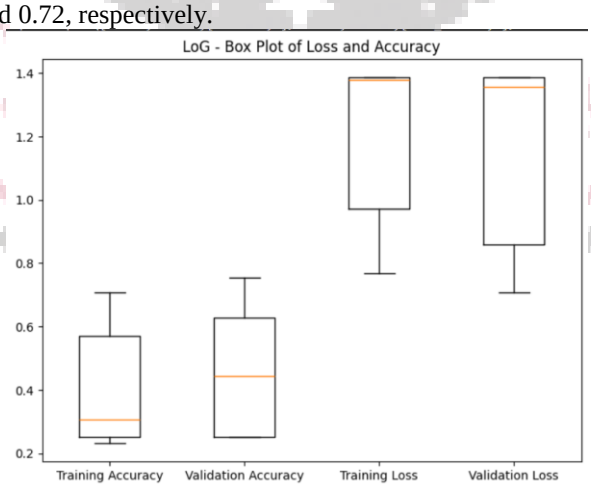

Fig. 6: LoG - Box Plot of Loss and Accuracy

A LoG Box Plot of Loss and Accuracy, as seen in Fig. 6, shows training accuracy to be slightly higher than the validation accuracy; both accuracies, however, show very little variability. And in Loss, the validation loss is still quite high as compared to the training loss. Therefore, these factors, with the rather-high validation loss and quite-good training accuracy, sometimes indicate overfitting in a model trained on training data. Additional regularization can be achieved with dropout and by providing more training data for better generalization, especially on the validation set.

Test Accuracy (LoG, 10 epochs)

The model's accuracy could likely be enhanced by tuning hyperparameters, increasing training epochs, or improving the feature extraction process. Additionally, more advanced techniques such as data augmentation, regularization, or using a deeper architecture could further improve the model's generalization ability on the test set.Hyperparameter tuning, using more training epochs, and improving the extraction of features could have been options to enhance the accuracy of the model. Advanced techniques such as augmentation, regularization, or maybe a deeper architecture could offer improvements to the model's ability to generalize to the test set. The Test Accuracy was 73.70% for the LoG model after 10 epochs.

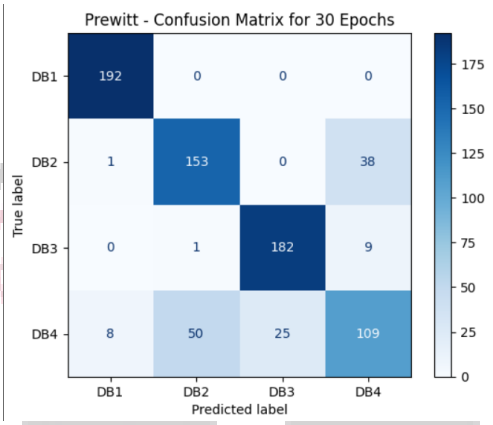

Fig. 7: Prewitt - Confusion Matrix for 30 Epochs

The Prewitt Confusion Matrix for 30 Epochs in Fig. 7 shows improved performance compared to previous epochs. The model performed well on DB1 and DB3, with 192 and 182 correct predictions, respectively. However, misclassifications remain on DB2 and DB4, with DB2 showing 38 misclassified instances and DB4 showing 109 misclassified instances. While the accuracy for most classes has improved, there are still some challenges, particularly with DB4, which could benefit from more data or further model refinement.

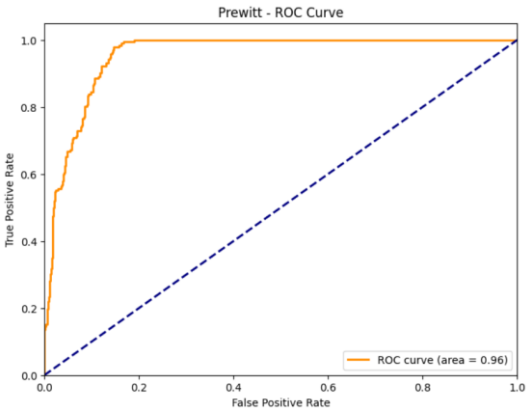

Fig. 8: Prewitt - ROC Curve

The curve shows a sharp rise in the TPR along a low FPR, showing the ability of the model to identify the classes correctly. While the AUC is high, some fine-tuning may upgrade performance, mainly to distinguish among more difficult classes.The Prewitt ROC Curve of Fig. 8 shows an AUC of 0.96, which means that the model has performed fabulously in discriminating between classes.

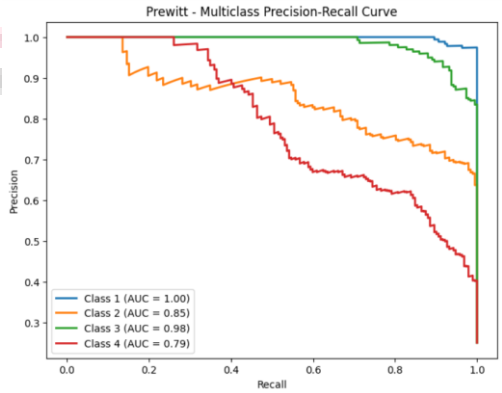

Fig. 9: Prewitt - Multiclass Precision-Recall Curve

The results show that the performance of the model differs from class to class-stellar for Class 1 and requiring optimization for Class 4. The Prewitt Multiclass Precision-Recall Curve in Fig.9 shows how the model performs for the different classes.

The AUC was perfect, at 1, for Class 1, 0.85, 0.98, and 0.79 for Classes 2, 3, and 4, respectively. While Class 1 has near-perfect precision, Class 4 continues to struggle with lower performance.

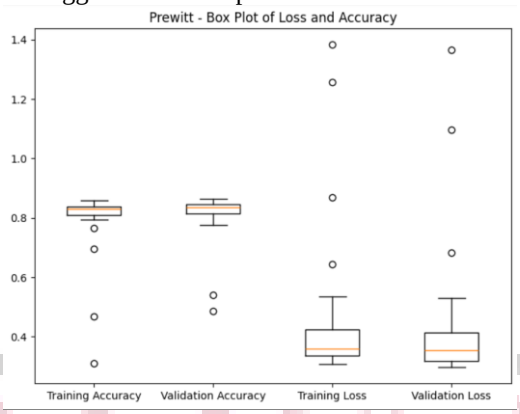

Fig. 10: Prewitt - Box Plot of Loss and Accuracy

The gaps suggest that the model might still be prone to overfitting, especially in the later epochs. Further regularization, more data, or fine-tuning might help reduce such gaps. The Prewitt Box Plot of Loss and Accuracy from Fig. 10 serves as an indication that training accuracy has always been higher than validation accuracy. From the standpoint of variability, both training accuracy and validation loss are less variable, but they do provide some avenue for improvement in terms of consistency of training and validation performance.

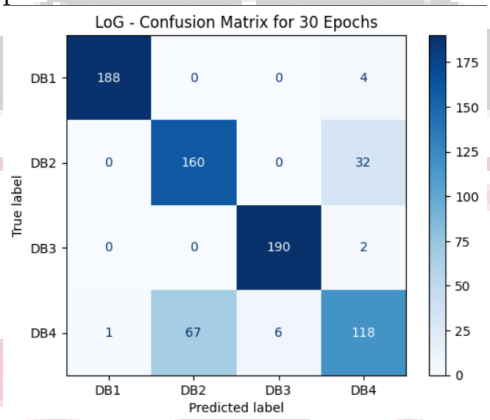

Fig. 11: LoG - Confusion Matrix for 30 Epochs

The model performed quite well on DB1, DB2, and DB3: 188, 160, and 190 correct predictions, respectively. DB4 still poses certain challenges, yielding 118 correct predictions and 67 misclassifications as DB2. There was an improvement in overall performance, but further optimization may be needed for DB4. The LoG Confusion Matrix for 30 Epochs shown in Fig. 11 shows marked improvement in classification accuracy over the earlier models.

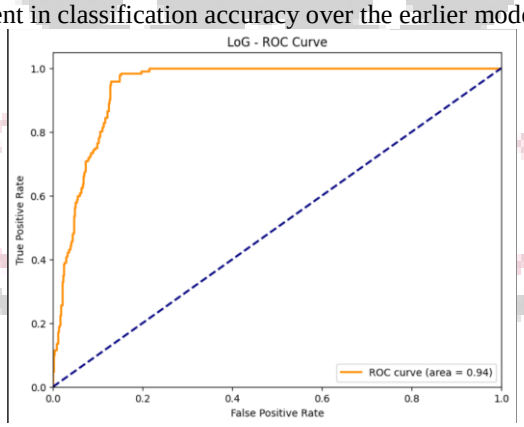

Fig. 12: LoG - ROC Curve

With an AUC of 0.94, the LoG ROC Curve in Fig. 12 is very strong, showing that the model can fairly well distinguish between the classes. Although the model performs well generally, there is still some room for improvement, particularly regarding the false positive rate, so as to allow for a higher true positive rate.

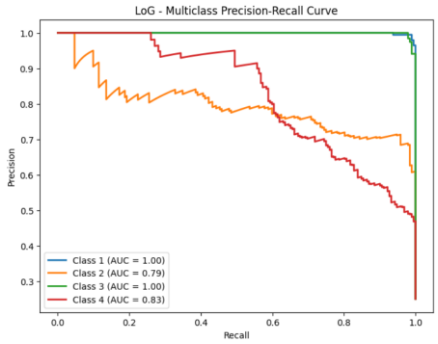

Fig. 13: LoG - Multiclass Precision-Recall Curve

The LoG Multiclass Precision-Recall Curve in Fig. 13 presents a detailed view of the precision-recall trade-off across all classes. The precision-recall curves of Class 1 and Class 3 were perfectly discriminative, with AUCs of 1.00, whereas Classes 2 and 4 had room for improvement, with AUCs of 0.79 and 0.83, respectively. These results signify a promising model for some classes but needing further optimization for others.

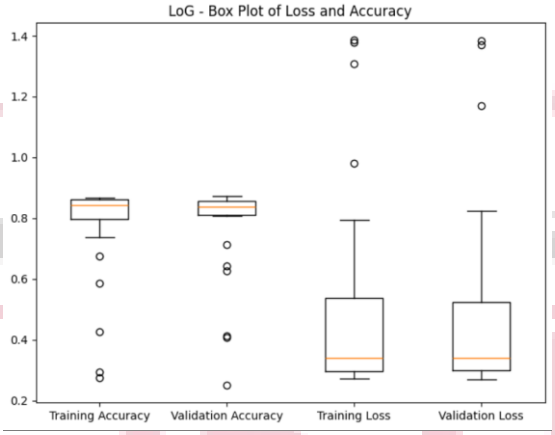

Fig. 14: LoG - Box Plot of Loss and Accuracy

The LoG boxplots of Loss and Accuracy shown in Fig.14 represent the training- and validation-based accuracies and losses data points. Looking at the plot, the evidence for overfitting is supported by the division between training accuracy and validation accuracy. Also present are variations in training loss and validation loss, which might encourage further help toward working out regularizations like dropout or data augmentation to enhance generalization.

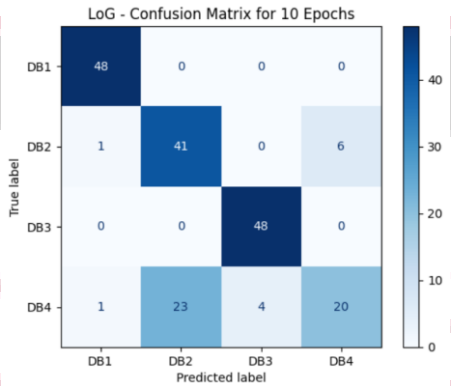

Fig. 15: LoG - Confusion Matrix for 10 Epochs

There are instances when the majority of predictions are classified appropriately, as there are true positives along the diagonal for all classes. DB1, DB2, and DB3 show best predictions whereas DB4 shows a few exceptions. This may indicate good generalization by the model; few improvements can be made for DB4 and the highest accuracy is attainable by the model.Yet, from the confusion matrix, a great performance by the model is evident.

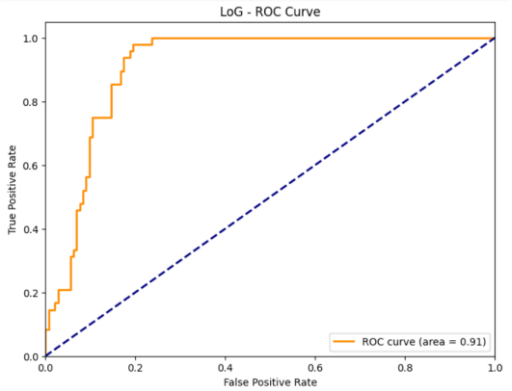

Fig. 16: LoG - ROC Curve

The value of 0.91 for area under curve (AUC) goes on to signify that the model has shown an excellent class-distinguishing ability, proving that balanced classification is being done by it. Moreover, the ROC curve (Fig. 16 epitomizes a remarkable performance with the curve suddenly shooting up to the top left corner.

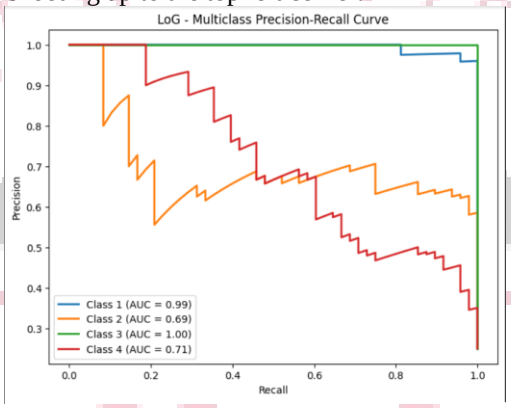

Fig. 17: LoG - Multiclass Precision-Recall Curve

Class 1 represents an impressively high AUC of 0.99. Class 3 becomes even better with an AUC score of 1.00, while Class 4 takes the third position with an AUC of 0.71. Although Class 2 should improve, the strength of the precision-recall performance for multiple classes underlines the model's great power and efficiency. The last multiclass precision-recall curve Fig. 17 guarantees the good performance of all classes.

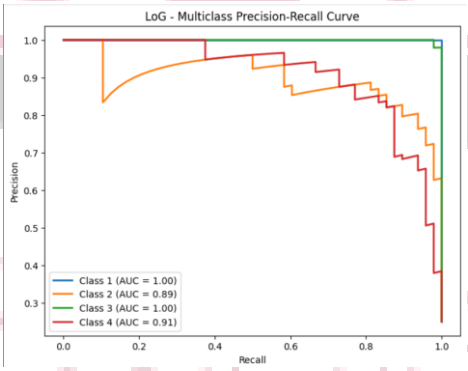

Fig. 18: LoG - Multiclass Precision-Recall Curve

The performance across classes highlights the model's versatility and its capacity to handle a range of data distributions effectively.The precision-recall curve demonstrates strong precision for all four classes, with AUC scores approaching or exceeding 0.9. This indicates that the model is not only making accurate predictions but is also performing well in terms of recall, showing its ability to detect all relevant instances.

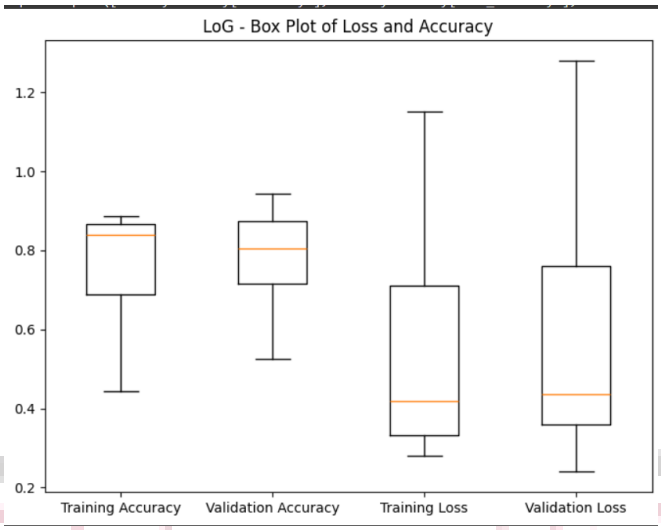

Fig. 19: LoG - Box Plot of Loss and Accuracy

The low variability and the close alignment of training and validation curves suggest the model is generalizing well and not overfitting. The training and validation losses are also notably low, reflecting the model's efficient learning process and its ability to minimize errors.The box plot of training and validation accuracy reveals an excellent consistency in model performance, with both training and validation accuracies consistently high.

VII. CONCLUSION

The hybrid deep learning model fostered by the joint architecture of CNN, ViT, and the Attention mechanism is sufficiently capable of distinguishing among fingerprint prints, increasing the accuracy of fingerprint classification. The CNN extracts local features, the ViT derives the global dependencies, and an Attention mechanism further refines the fingerprint areas paying attention to ridge patterns and minutiae points for better-based classification. After 30 epochs, this model can achieve an accuracy of up to 91.15% across several fingerprint datasets (DB1, DB2, DB3, DB4), perhaps facilitating the realization of real-time biometric identification and security applications. Performance parameters, such as precision, recall, F1-score, and ROC-AUC, establish that the model can indeed prove to be an apt solution for high-accuracy classification tasks. Several avenues can further improve the model in various ways: optimization of the model with fine-tuning of hyperparameters and advanced architectures will truly push the state-of-the-art, especially for challenging classes such as DB4. Having a more robust repertoire of data augmentation methods could help with generalization as well as tackle class imbalance. For deployment into the real-time system, it is imperative to bring down the latency of the model while maintaining its accuracy so as to serve in a biometric system in real life. Increasing the size and diversity of the dataset, including more difficult fingerprint classes or more types of biometrics, will improve robustness and adaptability. The accuracy and scalability of this model would lend itself well to integrated applications for secure access systems, law enforcement databases, and mobile authentication. Furthermore, integration with other modalities such as face or iris recognition can provide hybrid solutions to yield even more secure and accurate identification

.REFERENCES

[1] Wahab, T. M. Khan, S. Iqbal, B. AlShammari, B. Alhaqbani, and I. Razzak, "Latent fingerprint enhancement for accurate minutiae detection," Procedia Comput. Sci., vol. 246, no. C, pp. 1558–1567, Jan. 2024, doi: 10.1016/J.PROCS.2024.09.722.

[2] T. Meiramkhanov and A. Tleubayeva, "Enhancing Fingerprint Recognition Systems: Comparative Analysis of Biometric Authentication Algorithms and Techniques for Improved Accuracy and Reliability," Dec. 2024, Accessed: Aug. 07, 2025. [Online]. Available: https://arxiv.org/pdf/2412.14404

[3] M. B. Bhilavade, D. Shivaprakasha, M. R. Patil, L. S. Admuthe, R. Scholar, and A. Professor, "Fingerprint Reconstruction: Approaches to Improve Fingerprint Images," pp. 75–87, doi: 10.58346/JOWUA.2024.I1.006.

[4] H. Zhao, H. Yang, and S. Zheng, "Deep Learning-inspired Automatic Minutiae Extraction From Semi-automated Annotations," IEICE Trans. Fundam. Electron. Commun. Comput. Sci., vol. E107.A, no. 9, p. 2024EAP1043, Sep. 2024, doi: 10.1587/TRANSFUN.2024EAP1043.

[5] S. A. El-Rahman and A. S. Alluhaidan, "Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments," PLoS One, vol. 19, no. 2, p. e0291084, Feb. 2024, doi: 10.1371/JOURNAL.PONE.0291084.

[6] A. Mulay, S. A. Grosz and A. K. Jain, "Learning a Robust Minutiae Extractor via an Ensemble of Expert Models," *2024 IEEE International Joint Conference on Biometrics (IJCB)*, Buffalo, NY, USA, 2024, pp. 1-9, doi: 10.1109/IJCB62174.2024.10744473.

[7] T. Kavitha, B. Patil, S. Saraswathi, S. Rajarajeswari and A. Patil, "Recognition of Fingerprint Images using CNN for Cybercrime Detection System," *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/NMITCON62075.2024.

[8] P. Khare, S. Arora and S. Gupta, "Recognition of Fingerprint Biometric Verification System Using Deep Learning Model," *2024 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2024, pp. 01-07, doi: 10.1109/ICDSNS62112.2024.10691020.

[9] Z. Jia, C. Huang, Z. Wang, H. Fei, S. Wu and J. Feng, "Finger Recovery Transformer: Toward Better Incomplete Fingerprint Identification," in *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 8860-8874, 2024, doi: 10.1109/TIFS.2024.3419690.

[10] S. Kriangkhajorn, K. Horapong and V. Areekul, "Spectral Filter Predictor for Progressive Latent Fingerprint Restoration," in *IEEE Access*, vol. 12, pp. 66773-66800, 2024, doi: 10.1109/ACCESS.2024.3397729.

[11] Z. Pan, Y. Duan, X. Guan, J. Feng and J. Zhou, "Latent Fingerprint Matching via Dense Minutia Descriptor," *2024 IEEE International Joint Conference on Biometrics (IJCB)*, Buffalo, NY, USA, 2024, pp. 1-10, doi: 10.1109/IJCB62174.2024.10744445.

[12] N. Bhargava, P. S. Rathore, A. Goswami, K. Chauhan and S. Panda, "Performance Evaluation of Bit-String Normalisation of Fingerprint Database," *2024 Asian Conference on Intelligent Technologies (ACOIT)*, KOLAR, India, 2024, pp. 1-4, doi: 10.1109/ACOIT62457.2024.10939802.

[13] R. Bano, M. A. Zia, M. Asif, A. Noureen and A. Adnan, "Identification and Classification of Fingerprints Using Automated Deep-Learning Techniques," *2024 International Conference on IT and Industrial Technologies (ICIT)*, Chiniot, Pakistan, 2024, pp. 1-5, doi: 10.1109/ICIT63607.2024.10859518.

[14] Y. Artan and B. A. Semiz, "Fusion of Minutia Cylinder Codes and Minutia Patch Embeddings for Latent Fingerprint Recognition," Mar. 2024, Accessed: Aug. 07, 2025. [Online]. Available: https://arxiv.org/pdf/2403.16172

[15] M. Mallik, S. Chakraborty, K. Sasidhar, and C. Chowdhury, "VL-GAN: A generative classification approach for fingerprint-based indoor localization," Expert Syst. Appl., vol. 290, p. 128400, Sep. 2025, doi: 10.1016/J.ESWA.2025.128400.

[16] M. Sharafudeen and V. C. Vinod, "Dual residual learning of frequency fingerprints in detecting synthesized biomedical imagery," Appl. Soft Comput., vol. 173, p. 112930, Apr. 2025, doi: 10.1016/J.ASOC.2025.112930.

[17] Muthusamy and S. Muniyappan, "Enhancement comparison of Laplace kernelized piecewise regression-based progressive generative adversarial network for latent fingerprint," Pattern Anal. Appl., vol. 28, no. 3, pp. 1–20, Sep. 2025, doi: 10.1007/S10044-025-01491-0/METRICS.

[18] Z. Jin et al., "Channel Fingerprint Construction for Massive MIMO: A Deep Conditional Generative Approach," May 2025, Accessed: Aug. 07, 2025. [Online]. Available: https://arxiv.org/pdf/2505.07893

[19] S. R. Hosseini, O. Ahmadieh, J. Dawson, and N. Nasrabadi, "WaFusion: A Wavelet-Enhanced Diffusion Framework for Face Morph Generation," Jul. 2025, Accessed: Aug. 07, 2025. [Online]. Available: https://arxiv.org/pdf/2507.12493

[20] S. Wu *et al*., "LCVAE-CNN: Indoor Wi-Fi Fingerprinting CNN Positioning Method Based on LCVAE," in *IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2025.3575904.

[21] T. Xiang, Y. Sun and G. Shen, "MoDeFA: Multiobserver and Denoising-Enhanced Fingerprint Augmentation for Semi-Supervised Wi-Fi RSS-Based Indoor Positioning," in *IEEE Internet of Things Journal*, vol. 12, no. 15, pp. 31754-31767, 1 Aug.1, 2025, doi: 10.1109/JIOT.2025.3573967.

[22] W. Huang, Z. Yao, B. Jin, Z. Chen, and Y. Wang, "Controllable face soft-biometric privacy enhancement based on attribute disentanglement," J. Supercomput., vol. 81, no. 4, pp. 1–29, Mar. 2025, doi: 10.1007/S11227-025-07134-9/METRICS.

[23] Z. Lyu, T. T. L. Chan, G. C. M. Leung, Y. L. Chan, D. P. K. Lun and M. G. Pecht, "High-Dimensional Radio Frequency Fingerprint Synthesis for Indoor Positioning," in *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-16, 2025, Art no. 2517416, doi: 10.1109/TIM.2025.3551824

[24] D. ; Kim, J.-H. ; Park, Y.-J. Suh, D. Kim, J.-H. Park, and Y.-J. Suh, "A Wi-Fi Fingerprinting Indoor Localization Framework Using Feature-Level Augmentation via Variational Graph Auto-Encoder," Electron. 2025, Vol. 14, Page 2807, vol. 14, no. 14, p. 2807, Jul. 2025, doi: 10.3390/ELECTRONICS14142807.