

The Use of Attention-RNN and Dense Layer Combinations and The Performance Metrics Achieved in Palm Vein Recognition

Indriani, Yenie Syukriyah

Department of Informatics Engineering, Faculty of Engineering, Widyatama University, Bandung 40123, Indonesia

ARTICLE INFO

Article history:

Received December 30, 2024
Revised February 07, 2025
Published March 05, 2025

Keywords:

Palm vein recognition;
Deep learning;
RNN;
Attention RNN;
Dense Layer

ABSTRACT

The utilization of palm veins in vascular biometrics is widely recognized, offering significant potential and challenges for advancing individual recognition technology. Deep learning has played a crucial role in enhancing the accuracy of these recognition systems. In this study, we proposed combining Attention-RNN and Dense Layer. To validate this proposed method, three deep learning model scenarios were implemented: (1) a combined Dense Layer with RNN, (2) an Attention-RNN model, and (3) a combined Attention-RNN with a Dense Layer for palm vein recognition. Experimental results demonstrated that the Attention-RNN combined with the Dense Layer achieved the highest accuracy, outperforming the other two models. The model's performance was evaluated on two datasets, achieving 95% accuracy on the Kaggle dataset and 83% on the CASIA dataset, confirming its effectiveness in palm vein recognition.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



Corresponding Author:

Indriani, Department of Informatics Engineering, Faculty of Engineering, Widyatama University, Bandung 40123, Indonesia.
Email: indriani.st@widyatama.ac.id

1. INTRODUCTION

Biometric technology is advancing rapidly each year, presenting both opportunities and challenges for researchers and industry professionals. One of its key advantages is its ability to enhance security and operational efficiency. Biometric systems provide a reliable method for identity verification, making them essential for applications such as access control and authentication [1]. Furthermore, integrating biometrics into Internet of Things (IoT) devices enhances user experience by enabling seamless and secure interactions with digital services [2]. Innovations such as touchless biometric systems further improve accuracy and convenience. Despite these benefits, several challenges remain. Privacy concerns are a major issue due to the collection and storage of sensitive biometric data [3]. Additionally, adaptation to external factors [4], such as the COVID-19 pandemic, has exposed limitations in biometric technologies, particularly those reliant on facial and voice recognition. This has accelerated the development of alternative solutions, including contactless authentication methods. The widespread adoption of biometric technology across personal, industrial, and governmental applications continues to grow, driven by advancements in biometric data collection and processing systems.

The term "biometric" itself is derived from the Greek words *bio*, meaning "life," and *metrikos*, meaning "measurement" [5]. Biometric systems analyze physical and behavioral traits to verify identity [6], making them valuable for securing sensitive data, authenticating banking transactions, facilitating transportation and immigration procedures, and enhancing user experiences in industries like gaming and hospitality. For a biometric system to be effective, it must meet key requirements such as universality, uniqueness, permanence, technical feasibility, and user acceptability [7]. As illustrated in Fig. 1, biometric features are categorized into two types: physical biometrics, which include anatomical characteristics like the iris, fingerprint, facial

features, and vein patterns; and behavioral biometrics, which assess unique traits such as keystroke dynamics, handwriting patterns, gait, and voice recognition [8].

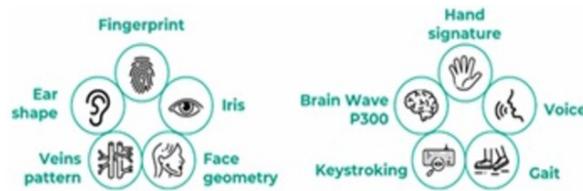


Fig. 1. The Features of Biometrics (Physical and Behaviour)

Palm vein recognition is an advanced biometric technology that utilizes the unique structure of an individual’s vascular patterns for identification. These vein patterns are highly intricate and varied making them a reliable source for biometric authentication. According to forensic professor Arrigo Tammassia from Padua University, even genetically identical individuals do not share the same vascular patterns on the back of the hand [9]. Additionally, since palm vein patterns are located beneath the skin, they are much harder to counterfeit compared to external biometric traits like fingerprint ridges or facial structures [10]. This inherent uniqueness enhances the security and accuracy of palm vein recognition in verifying identities [11], [12]. As outlined in Table 1, when compared to other biometric systems, such as fingerprint, facial, and iris recognition, palm vein recognition provides several distinct advantages, positioning it as a highly secure and effective authentication technology.

Table 1. Comparing the palm vein recognition with fingerprint, facial, and iris recognition

| Features | Palm vein recognition | Fingerprint Recognition | Facial Recognition | Iris Recognition |
|----------|---|--|---|--|
| Security | High (internal features, resistant to spoofing) | Moderate (can be spoofed with fake fingerprints) | Moderate (vulnerable face masks and photos) | High (unique and difficult to forge) |
| Accuracy | High – unique vein patterns | High – Unique fingerprint ridge (but affected by dirty, wear, or injuries) | Moderate (affected by lighting, pose, and occlusions) | High – Iris pattern remains stable over time |
| Hygiene | High (contactless) | Low (required physical touch) | High (contactless) | High (contactless) |
| Cost | High (required infrared sensor needed) | Low (widely available and inexpensive) | Low (uses standard cameras) | High (required specialized camera) |
| Adoption | Limited (mostly used in high-security applications) | High (used in phones, laptops, security) | High (used in phones, laptops, surveillance) | High (used in phones, laptops, security) |
| Speed | Fast (requires precise hand positioning) | Fast (requires fingerprint minutiae detection) | Fast (real-time facial pattern recognition) | Fast (requires accurate eye alignment) |

In deep learning models, the visual attention mechanism takes crucial role, particularly in computer vision tasks. Integrating attention mechanisms with deep learning, models can effectively identify and emphasize key features in image data using additional layers with learnable weights. During training, forward propagation and backward feedback enable networks to determine which regions of an image require greater focus [13]. This process enhances decision-making by prioritizing relevant details and addressing the complexity of visual inputs. Attention mechanisms have been successfully incorporated into various neural architectures, leading to significant advancements in multiple fields, including palm vein recognition. Several studies have investigated attention-based deep-learning techniques for palm vein recognition, each introducing unique methodologies and contributing to improved recognition performance.

Jiashu L. *et al.* [14] from Shenzhen University experimented with a Multi-Task Loss Function and Attention Layer, which effectively enhanced focus on crucial palm vein features. However, the complexity of the multi-task loss function increases the risk of overfitting, especially with smaller datasets, necessitating further validation on diverse datasets. Abdulrazzaq H. Imad [15] proposed the Residual Attention Network (RAN), which improves accuracy and computational efficiency by integrating residual learning with attention mechanisms. Despite its advantages, RAN’s complexity can result in longer training times and suboptimal performance when data is insufficient. Liao H. and Jin X. [16] introduced Gabor Attention Aggregation, which excels in capturing localized, frequency-specific vein patterns. However, its reliance on high-quality images

makes it vulnerable to noise and occlusions, limiting its robustness in real-world applications. Htet A.S. Min and Lee H. Jong [17] combined residual U-Net with an Efficient Channel Attention Residual Network (ECA-ResNet), leveraging pixel-level classification and enhanced feature focus. While this hybrid model performs well, its computational complexity poses scalability challenges, particularly when processing low-quality or blurred images. Wang P. and Qin H. [18] applied U-Net for segmentation, which, although effective in semantic segmentation tasks, struggles with distinguishing subtle vein patterns from background noise, raising concerns about generalization to large-scale datasets. Similarly, Nayar and Thomas [19] developed an authentication framework using partial palm vein patterns, which is useful in cases where full scans are unavailable. However, accurately identifying partial patterns remains a challenge, potentially increasing false acceptance and rejection rates. Sun S., Cong X., *et al.* [20] employed Neighbourhood Preserving Embedding (NPE) and Kernel Extreme Learning Machine (KELM) for dimensionality reduction and classification, effectively improving accuracy while reducing computational complexity. However, the success of this method heavily relies on high-quality feature extraction, as misalignment or poor preprocessing could degrade its performance. Across these studies, a common limitation is the dependency on high-resolution, noise-free datasets, which raises concerns about the practical implementation of these models in real-world conditions. Furthermore, more complex deep learning architecture typically requires substantial computational resources, making scalability and deployment challenging, especially in environments with limited processing power. Addressing these limitations is crucial for the efficiency of attention-based models in biometric authentication systems.

A Recurrent Neural Network (RNN) is a deep learning architecture specifically designed for processing sequential data by iteratively transforming inputs to generate meaningful outputs. When combined with the Attention mechanism, an RNN gains the ability to focus on the most important features within an image. This mechanism structures spatial focus using a top-down approach, capturing both local and global dependencies. It generates a spatial mask by diagonally scanning the image from the top corner to the opposite bottom corner, assigning attention values for specific regions. By incorporating local image context, the attention mechanism highlights significant areas, thereby improving recognition accuracy [21]. This allows the RNN to prioritize key image regions, enhancing its overall performance in feature identification. In contrast, a Dense Layer fully connects each neuron to every neuron in the previous layer, enabling the network to learn complex relationships within the data. This structure plays a crucial role in extracting meaningful representations, making it essential for deep learning tasks such as image classification and recognition. Dense layers refine the extracted features and contribute to the final decision-making process, complementing other neural network components like RNNs and attention mechanisms.

Inspired by the strengths of these models, this research proposes a deep-learning-based approach for palm-vein recognition. The key contributions of this study include:

1. Leveraging deep learning and Artificial Neural Networks (ANNs), particularly Dense Layers, RNNs, and Attention-RNNs, to enhance palm-vein recognition.
2. Conduct experiments involving three proposed algorithm variations: Dense Layer combined with RNN, Attention-RNN, and a model combined Dense Layer and Attention-RNN.
3. Evaluating the proposed models by testing them on the CASIA and Kaggle datasets under three different experimental scenarios and comparing their performances to determine the most effective approach.

2. METHODS AND MATERIALS

The following section outlines the research methodology and materials used for palm vein recognition using deep learning. It covers the proposed methodology, dataset description, deep learning model architecture, model training process, and evaluation techniques. This comprehensive approach ensures a structured understanding of how deep learning is applied to palm vein recognition, from data acquisition to performance assessment.

2.1. Methodology Research

Our research, three different approaches, each incorporating a unique algorithm to optimize palm vein recognition. As depicted in Fig. 2, it follows a structure consisting of five key stages: Data Collection, Data Preprocessing, Development or Build The Model, Model Training, and Evaluation. During the model-building phase, we refined the scenarios by integrating advanced deep-learning techniques to enhance feature extraction and recognition accuracy. Each scenario explores a different model configuration, aiming to identify the most effective approach for palm vein authentication. The subsequent sections provide a detailed explanation of each step, ensuring a comprehensive understanding of the methodology applied in this research.

Methodology Research

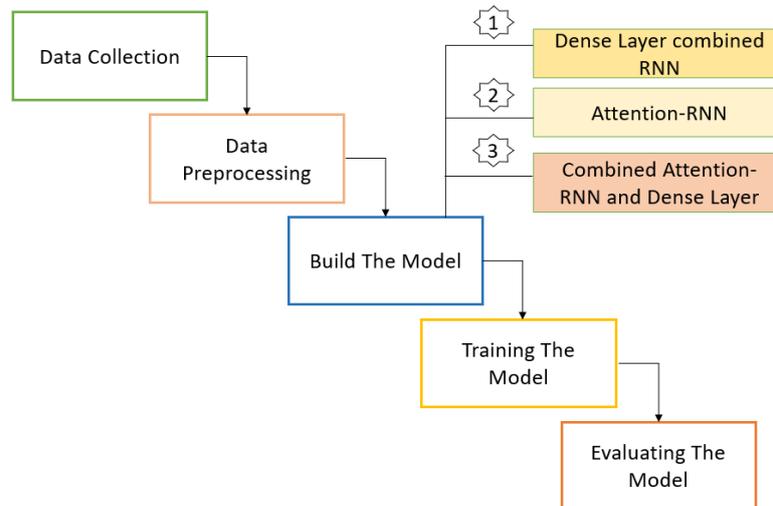


Fig. 2. Proposed methodology Palm vein recognition

2.2. Data Collection

A dataset is a collection of data that is stored and then processed so that it can produce information that meets the needs. Datasets have a vital role in research. With a dataset as a support layer, the built model will be reliable. Our research involves 2 public datasets as explained below:

1) Kaggle

Kaggle is a well-known online platform and a subsidiary of Google, serving as a hub for data scientists and machine learning professionals. It provides various tools and resources, including Datasets, where users can access, analyze, and share diverse datasets; Models, featuring machine learning, diffusion models, and Large Language Models (LLMs); Code, which enables users to explore and execute models using Kaggle Notebooks and Kaggle's public API; Competitions, which allow participants to enhance their skills through challenges; and Courses, offering educational content on data science and machine learning [36]. Additionally, Kaggle fosters an interactive community, where experts collaborate on projects, exchange insights, and participate in discussions. For this study, we utilized a contactless knuckle-palm print and vein dataset, initially shared by Michael Goh [5], [22], [23], [24]. This dataset comprises 1.958 images from 98 individuals, with each subject contributing 20 images where 10 from the right palm vein and 10 from the left palm vein. Users were instructed to position their hands naturally while slightly spreading their fingers apart, allowing movement and rotation without restrictions. The images were captured using infrared LED and visible light in the 880 nm to 920 nm range, with an optimal viewing distance of 25 cm and a focus tolerance of $25 \text{ cm} \pm 4 \text{ cm}$. Additionally, the captured images were processed to extract Regions of Interest (ROI) and stored in bitmap format, with a resolution of 640×480 pixels and 256 RGB (8 bits per channel).

2) CASIA

The Chinese Academy of Sciences Institute of Automation (CASIA) [25], established by the Center for Biometrics and Security Research (CBSR), was designed to support research and drive advancements in biometric authentication and intelligent surveillance technologies. Its primary objectives include improving the standardization of biometric databases and protocols while serving as a testing facility for biometric products. In this study, we utilized the Multi-Spectral Palmprint Image Database V1.0 (CASIA-MS-PalmprintV1), which consists of 7.200 palm vein images in 8-bit gray-level JPEG format. A sample of the CASIA dataset is illustrated in Fig. 3 and was used to support the research of Ying Hao [26], [27]. The dataset contains samples from 100 individuals, captured using a custom-designed multi-spectral imaging device. Each image was simultaneously recorded across six different electromagnetic spectrums: 460 nm, 630 nm, 700 nm, 850 nm, 940 nm, and white light. The dataset also includes images of hands in various postures, though no strict positioning constraints were imposed.

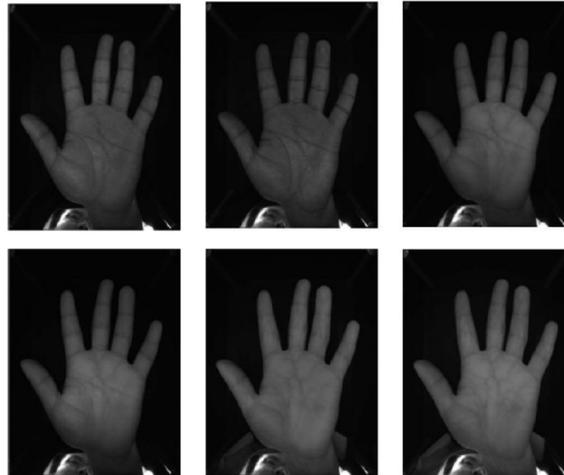


Fig. 3. Sample palm vein datasets in CASIA with multi-spectral

2.3. Data Preprocessing

The palm vein datasets were collected in the previous stage, and now, in this data processing phase, the input images undergo various transformations to ensure they can be processed effectively and yield the desired results. This step includes:

- a. Resizing Image: The dimensions of the palm vein images in Kaggle and CASIA datasets were adjusted for the image sizes to maintain consistency in processing and model compatibility.
- b. Normalization: The pixel values were scaled to a uniform range in standard intensity levels across all images, ensuring uniformity in intensity across all images.
- c. Dataset Partitioning: The dataset was divided into three subsets to facilitate:
 - Training dataset: This set is used to train or build a deep-learning model, to recognize general palm vein patterns in images.
 - Validation dataset: This data assists in optimizing the model by evaluating its ability to recognize palm vein patterns. It also helps assess the model's accuracy.
 - Testing dataset: This set evaluates the model after training is complete.
 The composition data partitioning was split into a ratio of 80:10:10, where 80% was allocated for training, 10% for validation, and 10% for testing.

2.4. Building The Model

Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a type of neural network architecture designed to process sequential data through the use of feedback connections. It includes a hidden state that updates dynamically based on information from the previous time step, allowing the network to capture temporal dependencies. This feedback mechanism enables RNNs to effectively process large sequential datasets and handle tasks that involve long-term dependencies. By linking past data to current output, RNNs are particularly well-suited for applications such as language modeling, forecasting, recognition, and other tasks that require understanding the relationship between previous and current inputs.

The development of RNNs aligns with advancements in human-computer interaction. After the concept of artificial intelligence was introduced in 1956, machine learning methods like Naïve Bayes and Decision Trees emerged in 1959. In 1986, Rumelhart *et al.* [28] introduced the backpropagation algorithm, which allowed weight updates through error propagation, revolutionizing neural network training. In 1982, Hopfield [29] developed the Hopfield Network, or Binary Hopfield Network, which was used for pattern retrieval and related to storage algorithms. Later, in 1991, Elman [30] introduced the Simple Recurrent Network (SRN), which utilized hidden units to map new inputs and previous states to outputs, creating task-specific memory representations for learning tasks. In recent years, RNN architectures have been significantly enhanced by introducing mechanisms like attention, which have greatly improved their performance and expanded their applications. RNN equations [31] can be written as:

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; W) \quad (1)$$

where h represents the network's hidden state of the network, which stores information from previous time steps and is updated at each step. x donates the input to the network. t is the time step that indicates the sequence position. W refers to the network's weight matrices, which include the input-to-hidden, hidden-to-hidden, and hidden-to-output connection weights. At each time step, the hidden state h_t is updated based on the input x_t and the previous hidden state h_{t-1} , utilizing these weight matrices to compute the new hidden state. This structure enables RNNs to capture temporal dependencies in sequential data.

Attention Mechanism

The attention mechanism has emerged as a powerful technique in deep learning, significantly improving model performance by allowing the model to focus on the most relevant parts of the input data. This selective focus enables the model to prioritize important features while disregarding irrelevant or extraneous information, which is why it is often referred to as attention. By dynamically allocating focus based on the importance of different input elements, the attention mechanism enhances the model's ability to make more informed and accurate predictions, particularly in tasks involving complex data such as images, text, and sequential information.

There are several compelling reasons why attention mechanisms have become essential in deep learning models. Firstly, models utilizing attention mechanisms consistently demonstrate superior performance across a wide range of tasks. These include medical image analysis [32] to improve diagnostic accuracy, horticulture and agriculture for detection, localization, and tracking of the fruits in the given video [33], enhanced transport networks safely by detecting road weather [34], and identifying fake news on social network platforms [35], and among many others. The success of attention mechanisms can attributed to their ability to integrate seamlessly with foundational models like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) while remaining trainable through standard backpropagation techniques [36]. Furthermore, The introduction of the Transformer model has significantly amplified the popularity of attention mechanisms, highlighting their remarkable effectiveness and versatility in modern deep-learning architectures [37].

The formula in the attention mechanism, as shown in Equation (2), involves the use of the query, key, value, and output vectors, all derived from an attention function. This function maps a query and a set of key-value pairs to an output. The output is computed as a weighted sum of the values, where the weights are determined by a compatibility function. This function measures the relationship between the query and each corresponding key, determining how much attention should be given to each value based on the query's relevance to the key [37].

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q is the representation of the feature the model is trying to identify, K is the representation of the features of the input that could be related to the query, V contains the actual data associated with the key and contributing to the output, QK^T It measures the similarity between query and keys, assists the model in determining the relevance of each input and d_k is a scaling factor. This formulation allows the attention mechanism to assign different levels of importance to different parts of the input, improving the model's ability to focus on the most relevant information.

In palm-vein recognition, incorporating the attention mechanism proves particularly useful, as it helps identify key features in palm-vein images, leading to improved recognition accuracy. By assigning different levels of importance to various elements in the input, the model can better focus on the most significant aspects, thereby refining its predictions. This adaptive technique is especially advantageous when working with data prone to noise or irrelevant information, which is a common challenge in palm-vein images. Attention-based modules can be integrated into diverse neural network designs, enhancing feature extraction and boosting model performance with minimal additional computational overhead [11].

Dense Layer

The Dense Layer, a core element in deep learning models, is a fundamental building block that functions as a fully connected layer within neural networks. The history of the Dense Layer highlights its critical role in neural network architecture, beginning with the early development of multi-layer perceptron and deep learning. The concept of the perceptron was introduced by Frank Rosenblatt, who conducted research at the Cornell Aeronautical Laboratory. In his 1957 technical report titled *The Perceptron: A Perceiving and Recognizing Automation* [38] and his 1958 journal publication titled *The Perceptron: A Probabilistic Model for Information*

Storage and Organization in the Brain [39], Rosenblatt outlined the perceptron concept. His experiments demonstrated that the perceptron system consisted of units connected in an associative structure, where each unit performed a switching function between inputs and outputs and transmitted results to one or more response units. Later, in 1986, Geoffrey Hinton [30] and collaborators identified limitations in single-layer connections and introduced the concept of multi-layer perceptron (MLPs), along with the backpropagation algorithm. This innovation enabled gradients to propagate backward through all layers of a network, facilitating the updating of weights in fully connected layers. This advancement became the foundation of the Dense Layer architecture. Today, with the advent of powerful GPUs and large datasets, deep learning has evolved to address a wide range of tasks and applications, making algorithms incorporating Dense Layers more versatile and effective such as those researched by Gupta N., *et. al.*, [40], Krasteva V, *et. al.*, [41], Helen, *et. al.*, [42] and so on.

Scenario in Research

There are 3 (three) scenarios in recognizing using palm veins that were observed, including:

a. Scenario 1: The algorithm used is a Dense Layer combined with RNN (Recurrent Neural Network)

In this scenario, the model is designed by utilizing a Dense layer, followed by a Recurrent Neural Network (RNN) to process palm vein images as illustrated in Fig. 4. Both the Dense layer and RNN play crucial roles in deep learning, contributing to feature extraction and sequential pattern recognition, respectively. The palm vein image serves as the input, reshaped into a 2D format to ensure compatibility with the Dense layer. Initially, feature extraction uses a fully connected Dense layer with 256 neurons and ReLU activation, enabling the network to learn meaningful representations. The extracted features are then passed into an RNN with 128 units, which captures sequential dependencies and patterns in the data. This sequential analysis refines the image representation into a single vector, which is then fed into the final classification layer with softmax activation to determine the class label of the palm vein image. The model is compiled using the Adam optimizer, ensuring efficient weight updates, and employs categorical cross-entropy loss, which is suitable for handling multi-class classification tasks.

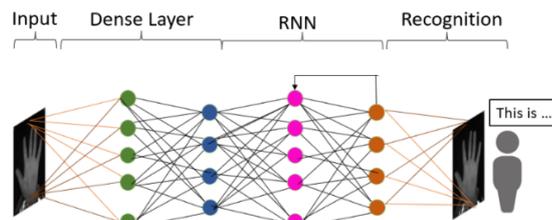


Fig. 4. Dense Layer and RNN Architecture on Palm Vein Recognition Architecture

b. Scenario 2: The algorithm used is Attention-RNN

The second scenario as illustrated in Fig. 5, this research focuses on palm vein recognition using the Attention-RNN model. In deep learning, attention mechanisms play a vital role by selectively focusing on different parts of the input sequence when predicting the output, making dataset learning more efficient. The model processes palm vein images with an input size of 1 and a hidden size of 128, incorporating the ReLU activation function to enhance feature extraction and capture meaningful representations. To classify and assign the appropriate labels for palm vein recognition, a softmax output layer is utilized. For optimization, the model is compiled using the Adam optimizer, which enables efficient weight adjustments and applies categorical cross-entropy loss, ensuring effective training and improved classification performance.

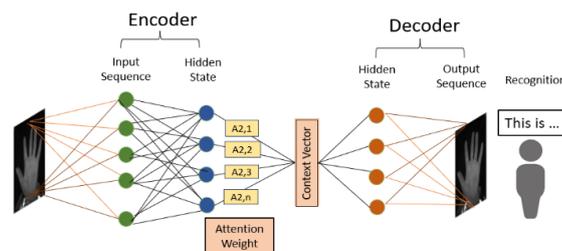


Fig. 5. Attention-RNN on Palm Vein Recognition Architecture

c. Scenario 3: The algorithm uses is combined Attention-RNN and Dense Layer.

The final scenario of this research explores an experimental model that integrates Attention-RNN with a Dense Layer for palm vein recognition, as illustrated in Fig. 6. Initially, the input palm vein image is processed using Attention-RNN with an input size of 1 and a hidden size of 128, allowing the model to capture essential features. The extracted features are then passed through a Dense Layer with 64 neurons, which further refines the learned representations, enhancing the model's ability to distinguish unique palm vein patterns. For recognition, the model employs a softmax output layer, ensuring accurate multi-class prediction. To optimize training and achieve the best results, the model is compiled using the Adam optimizer, which facilitates efficient weight adjustments, and applies categorical cross-entropy loss, ensuring effective learning and precise recognition of palm vein patterns.

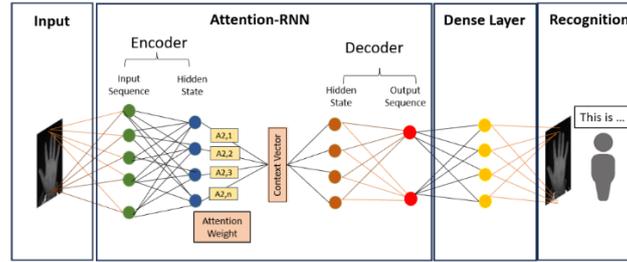


Fig. 6. Combined Attention-RNN and Dense Layer on Palm Vein Recognition Architecture

2.5. Training the Model

At this stage, the model built in each scenario undergoes training using the previously prepared training dataset. The primary objective of this training process is to enable the model to learn and recognize palm vein patterns in images, allowing it to make accurate predictions for identifying individuals. The training process is conducted over 50 epochs, meaning the model iteratively learns from the data multiple times to refine its accuracy. Additionally, training is performed in batches of size 32, ensuring efficient weight updates and optimization of the model parameters. This batch-wise training approach helps in stabilizing the learning process and improving overall performance.

2.6. Evaluating The Model

After completing model training, the next crucial stage is to evaluate the model, which focuses on assessing the performance of the trained model. This evaluation process is conducted using a separate test dataset that was not used during training to ensure an unbiased assessment of the model's generalization ability. For performance evaluation, this study employs the *Confusion Matrix* method [43]. A confusion matrix is a tool used to assess a classifier's performance on binary or multi-class test data by categorizing predictions into four key outcomes: (1) True Positives (TP) – Correctly predicted positive cases, (2) False Positives (FP) – Incorrectly predicted positive cases, (3) True Negatives (TN) – Correctly predicted negative cases, and (4) False Negatives (FN) – Incorrectly predicted negative cases. Using these metrics, several key performance measures are derived: (1) Accuracy – Measures the overall correctness of the model, (2) Precision (Positive Predictive Value) – Assesses how many of the predicted positive cases are correct, (3) Recall (Sensitivity or True Positive Rate) – Evaluate the model's ability correctly identify positive cases, and (4) F1-Score – Provides a balance between Precision and Recall, especially when dealing with imbalanced datasets. These evaluation metrics collectively help determine the effectiveness and reliability of the palm vein recognition model.

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision(Pre) = \frac{TP}{TP + FP} \quad (4)$$

$$Recall(Rec) = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = 2 \frac{Pre. Rec}{Pre + Rec} \quad (6)$$

3. RESULTS AND DISCUSSION

This section presents and analyzes the results obtained from each model variant—Combined Dense Layer and RNN, Attention-RNN, and Combined Attention-RNN with Dense Layer. The performance of these models is evaluated based on key metrics such as accuracy, precision, recall, and F1 score. Additionally, the discussion highlights the strengths and limitations of each approach, comparing their effectiveness in palm vein recognition. The results provide insights into how different deep learning architectures impact the accuracy and reliability of the recognition process.

3.1. Experiment in Scenario 1: Dense Layer Combined with RNN

The experiment recognized the palm vein image from both datasets, CASIA and Kaggle, using combined Dense Layer and RNN, which produced different results when the data were trained. The graph model loss in Kaggle and CASIA in Fig. 7, shows that both training and validation loss models indicated decreasing so we can conclude that the model combined Dense Layer and RNN is learning and generalizing new data well. The training status on the model using Kaggle datasets had a value of accuracy of 0.892 and validation accuracy of 0.783, while using CASIA datasets showed accuracy and validation accuracy higher, at 0.903 and 0.795, respectively. In addition to model accuracy and validation accuracy, we also did model training loss to see how the model calculated error on training data to discover the model's competence and validation loss to evaluate the performance which generalizes to new data. Kaggle has results of 0.479 and 0.904 while CASIA is 0.406 and 0.725.

It is meaningful to evaluate and measure the performance of the model combined Dense Layer and RNN that are built. To know the quality model by collecting, analyzing, and calculating certain assignments that were plotted before. The implementation confusion matrix as a model measures performance as outcomes as written in Table 2. We tried using 196 images of 1958 from total image palm vein images as dataset testing on Kaggle and 1000 images on CASIA. In Kaggle, the accuracy model has a value of 0.663 or 66.3%, meanwhile, CASIA is 0.619 or 61.9%. The value 0.663 in Kaggle was obtained from 196 images in dataset testing where the total correct prediction had 130 images, and 66 images were misclassified on the model, and in the CASIA accuracy model shown in point 0.619 or 61.9% from 1000 images of the dataset testing total where correct prediction had 619 images and 381 images are the total misclassified images on the model.

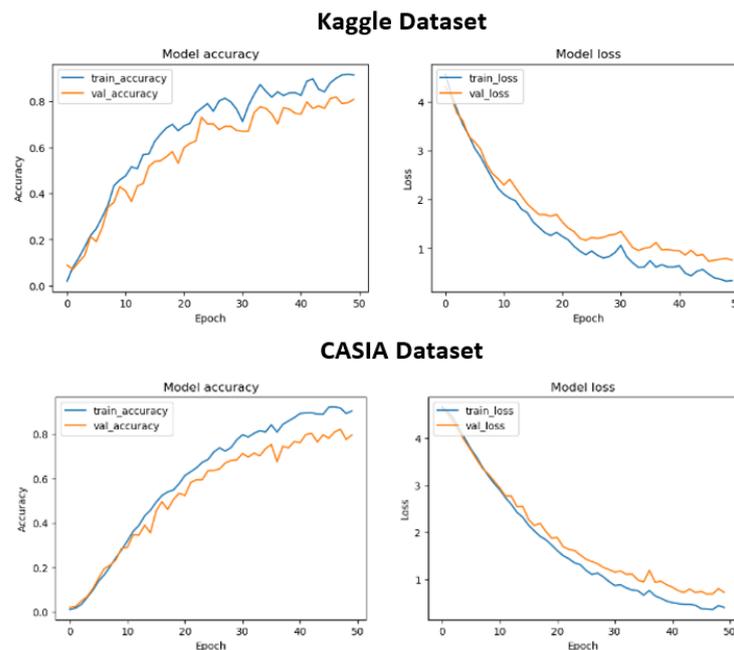


Fig. 7. Training data in Model Combined Dense Layer and RNN

Table 2. Confusion Matrix for model combined Dense Layer and RNN.

| Datasets | Accuracy | Precision | Recall | F1 Score |
|----------|----------|-----------|--------|----------|
| Kaggle | 0.663 | 0.675 | 0.663 | 0.636 |
| CASIA | 0.619 | 0.632 | 0.619 | 0.611 |

3.2. Experiment in Scenario 2: Attention-RNN

In Fig. 8, the status of training using Attention-RNN shows that the model accuracy is 1.0, with a validation accuracy of 0.936. Meanwhile, the model loss is 0.014, and the validation loss is 0.247 on the Kaggle datasets. On the other hand, on the CASIA datasets, the model accuracy has decreased to 0.786, with a validation accuracy of 0.802, while the model loss is 0.723, and the validation loss is 0.710.

After testing the model, the Attention-RNN model was evaluated to measure its performance, as shown in Table 3, using the provided test data. The performance results show an accuracy of 94% (0.948) on the Kaggle datasets and 76% (0.761) on the CASIA datasets for recognition. On Kaggle, out of 196 images, 186 were correctly predicted, and 10 were misclassified. On the CASIA datasets, from 1000 test images, 761 images were correctly predicted, and 239 images were misclassified.

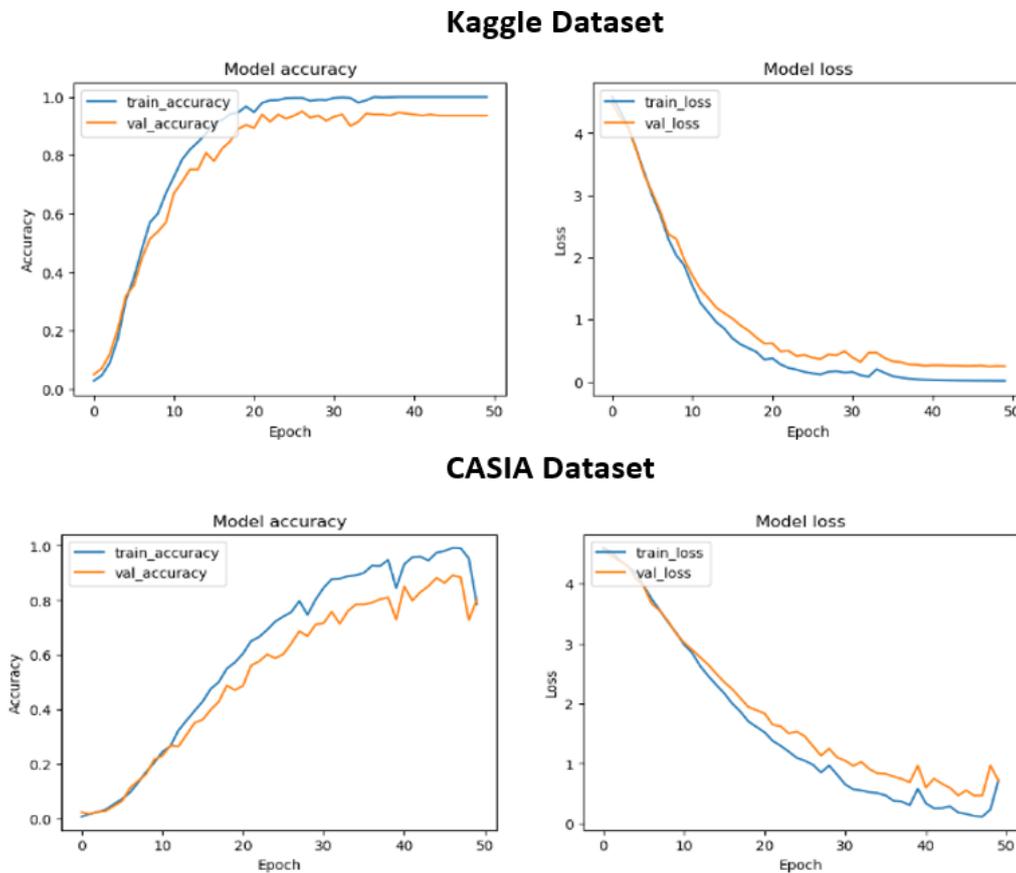


Fig. 8. Training data in Model Attention-RNN

Table 3. Confusion Matrix for model Attention-RNN

| Datasets | Accuracy | Precision | Recall | F1 Score |
|----------|----------|-----------|--------|----------|
| Kaggle | 0.948 | 0.971 | 0.948 | 0.948 |
| CASIA | 0.761 | 0.785 | 0.761 | 0.753 |

3.3. Experiment in Scenario 3: Combined Attention-RNN and Dense Layer

The model accuracy and loss for this scenario are illustrated in Fig. 9, once model and data training is completed, the combined Attention-RNN and Dense Layer model achieves a performance accuracy of 1.0 and a validation accuracy of 0.939 on the Kaggle dataset, with a model loss of 0.003 and a validation loss of 0.279.

On the CASIA dataset, the model achieves an accuracy of 1.0 and a validation accuracy of 0.967, with a model loss of 0.016 and a validation loss of 0.155.

The performance score of this model, as shown in Table 4, is better compared to the Attention-RNN model and the combined Attention-RNN and Dense Layer model when tested on both the Kaggle and CASIA datasets. The combined model achieved an accuracy of 0.954 (or 95%) on Kaggle and 0.833 (or 83%) on CASIA. The Kaggle dataset consists of 196 test images. From this, 187 images were correctly predicted, and 9 images were misclassified. For the CASIA datasets, which contain 1000 test images, 833 images were correctly predicted, and 167 images were misclassified.

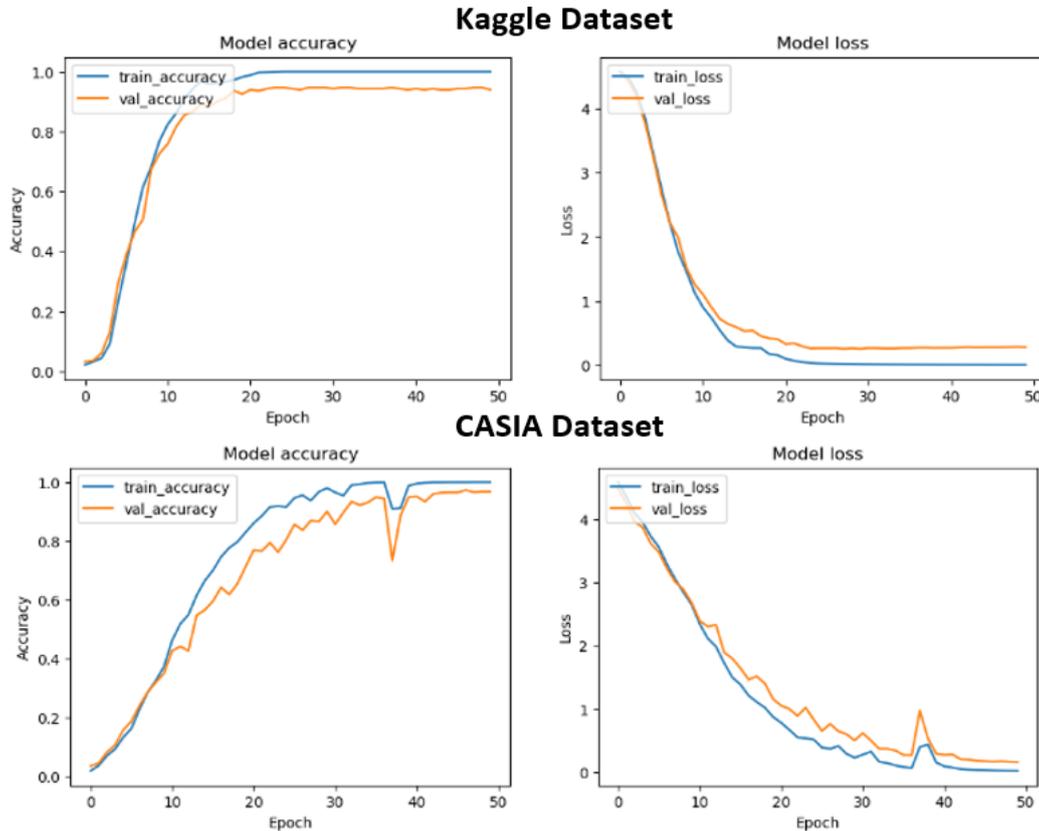


Fig. 9. Training data in Model Combined Attention-RNN and Dense Layer

Table 4. Confusion Matrix for model Combined Attention-RNN and Dense Layer

| Datasets | Accuracy | Precision | Recall | F1 Score |
|----------|----------|-----------|--------|----------|
| Kaggle | 0.954 | 0.967 | 0.954 | 0.951 |
| CASIA | 0.833 | 0.846 | 0.833 | 0.826 |

3.4. Causes of Performance Differences Across the Three Scenarios with Different Datasets

Our research, which explores various scenarios and compares different palm vein recognition methods proposed by multiple researchers, has yielded promising results in enhancing the accuracy of palm vein authentication, as shown in Table 5. The Attention-RNN model has proven to be highly effective in extracting palm vein patterns, significantly improving recognition performance. Furthermore, combining Attention-RNN with a Dense Layer has further enhanced accuracy, demonstrating the potential of the attention mechanism in refining the biometric recognition system.

The results from the three scenarios indicate that the performance of the CASIA dataset is lower than that of the Kaggle dataset. This discrepancy can be attributed to how palm vein images are captured across various electromagnetic spectra in the CASIA dataset. The CASIA dataset includes palm vein images taken under six different wavelengths: 460 nm, 630 nm, 700 nm, 850 nm, 940 nm, and white light, each offering varying visibility of palm veins. Specifically, images captured under the 460 nm, 630 nm, and 700 nm wavelengths (in Fig. 3), which fall within the visible spectrum, and show limited visibility of the palm veins or no visibility at all. These spectral bands primarily capture surface-level features like skin texture, wrinkles, and principal lines

[44] due to their limited penetration depth into the skin. The visible spectrum's inability to penetrate deeper is due to the strong absorption of light by melanin in the skin and hemoglobin in blood vessels, which hinders the capture of deeper structures. Moreover, skin conditions can further impact the consistency of the features extracted from visible spectrum images, becoming a potential weakness. In contrast, the 700 nm and 940 nm bands, which fall in the near-infrared (NIR) spectrum, have longer wavelengths that penetrate deeper into the skin. These wavelengths enhance the visibility of subcutaneous vein patterns [45], as hemoglobin absorbs NIR light, creating high-contrast images of the veins and surrounding tissue [46]. On the other hand, white light spectra, which combine a broad range of wavelengths, capture both surface and deeper subcutaneous features [47], making palm veins visible to the naked eye and allowing a more comprehensive extraction of features from the surface and underlying structures. In other words, the visible spectrum with 460 nm, 630 nm, and 700 nm fails to reveal underlying vein structures, making it less suitable for palm vein-based biometrics comparing near-infrared (850 nm, 940 nm) and white light has potentially higher accuracy with unique vein patterns in biometric recognition.

Table 5. Comparing our results with the other research.

| Authors | Proposed Method | Dataset | Accuracy |
|-------------------------------|--|--|---|
| Jianshu [14] | Multi-Task Loss Function and Attention Layer | Different Datasets | 98.89% |
| Abdulrazzaq H Imad [15] | Residual Attention Network | CASIA | 95.55% |
| Liao H, Jin X, etc [16] | Gabor Aggregation | CASIA, VERA Palm-vein dataset, The Tongji University Palm-vein dataset | CASIA 99.17%, VERA 97%, and The Tongji University 97.82% |
| Htet A.S Min, Lee H Jong [17] | Combining U-Net & ECA-ResNet | CASIA | 100% |
| Wang P and Qin H [18] | U-Net | CASIA | EER 0.47% |
| Nayar and Thomas [19] | Partial Palm Vein Authentication Framework | CASIA, VERA Palm-vein dataset, PUT | CASIA 0,0202% EER, 0,1919% EER, and PUT 0,2282% EER |
| Sun S, Cong X., et al. [20] | NPE & KELM | Hongkong Polytechnic University | 100% |
| Indriani, et al. | Our proposed method: 1. Combining Dense Layer and RNN 2. Attention-RNN 3. Combining Attention-RNN and Dense Layer | Kaggle CASIA | Kaggle 1. 66.3% 2. 94.8% 3. 95.4% CASIA 1. 61.9% 2. 76.1% 3. 83.3% |

Additionally, the three model architectures—Dense Layer combined with RNN, Attention-RNN, and the combined Attention-RNN and Dense Layer—require further improvements to enhance their performance in handling multispectral images. Advanced feature extraction methods must address the variability across different spectral bands, ensuring more effective recognition. Fusion strategies integrating information from multiple wavelengths can enhance discriminative power while developing specialized acquisition devices capable of capturing palm vein images under various spectral conditions can improve robustness. These advancements will allow the models to adapt better to real-world security and authentication applications, making them more reliable for biometric identification.

4. CONCLUSION

In this research, we explored palm vein recognition by combining a Dense Layer and Attention-RNN. To enhance our research, we implemented three different scenarios: combining a Dense Layer and RNN, using Attention-RNN, and combining Attention-RNN with a Dense Layer. We conducted an experiment using two distinct datasets, Kaggle and CASIA and compared the accuracy performance of each model. The results indicate that the combination of Attention-RNN and a Dense Layer achieved the highest accuracy for palm vein recognition, outperforming the other two models, where obtained accuracy performance in Kaggle was 0.954 or 94,5% and 0.833 or 83.3% in CASIA. For future research, we aim to improve performance in handling

multispectral palm vein images by enhancing feature extraction techniques, refining recognition algorithms, or developing specialized acquisition devices capable of adapting to various spectral conditions.

Acknowledgments

This journal article was authored by Indriani from the Faculty of Engineering, Widyatama University, based on the research report titled *The Use of Attention-RNN and Dense Layer Combinations and The Performance Metrics Achieved in Palm Vein Recognition*. The study was funded by the Bureau of Research, Community Service, and Intellectual Capital in 2024. The opinions and conclusions presented in this paper are solely those of the author and do not necessarily reflect the views of the funding institution. We extend our appreciation to the Chinese Academy of Sciences Institute of Automation (CASIA) for providing access to the CASIA-MS-PalmprintV1 dataset, which was instrumental in this study. Additionally, we acknowledge Goh Kah Ong Michael for making the Contactless Knuckle-Palm Print and Vein datasets publicly available on Kaggle, which significantly contributed to the research findings.

REFERENCES

- [1] A. De Keyser, Y. Bart, X. Gu, S. Q. Liu, S. G. Robinson, and P. K. Kannan, "Opportunities and challenges of using biometrics for business: Developing a research agenda," *J. Bus. Res.*, vol. 136, pp. 52–62, Nov. 2021, <https://doi.org/10.1016/j.jbusres.2021.07.028>.
- [2] C.-W. Lien and S. Vhaduri, "Challenges and Opportunities of Biometric User Authentication in the Age of IoT: A Survey," *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1–37, Jan. 2024, <https://doi.org/10.1145/3603705>.
- [3] Y. Zhang, B. Liao, and R. Lei, "Challenges and principled responses to privacy protection from biometric technology in China," *Acta Bioethica*, vol. 29, no. 2, pp. 249–258, Oct. 2023, <https://doi.org/10.4067/S1726-569X2023000200249>.
- [4] M. Gomez-Barrero *et al.*, "Biometrics in the Era of COVID-19: Challenges and Opportunities," *IEEE Trans. Technol. Soc.*, vol. 3, no. 4, pp. 307–322, Dec. 2022, <https://doi.org/10.1109/TTS.2022.3203571>.
- [5] S. Kodituwakku, "Biometric Authentication: A Review," *Int. J. Trend Res. Dev.*, vol. 2, pp. 113–123, Aug. 2015.
- [6] C. Wilson, *Vein pattern recognition: a privacy-enhancing biometric*. Boca Raton: Taylor & Francis, 2010. <https://doi.org/10.1201/9781439821381>.
- [7] B. Mróz-Gorgoń, W. Wodo, A. Andrych, K. Caban-Piaskowska, and C. Kozyra, "Biometrics Innovation and Payment Sector Perception," *Sustainability*, vol. 14, no. 15, p. 9424, Aug. 2022, <https://doi.org/10.3390/su14159424>.
- [8] G. K. O. Michael, T. Connie, and A. Teoh Beng Jin, "An innovative contactless palm print and knuckle print recognition system," *Pattern Recognit. Lett.*, vol. 31, no. 12, pp. 1708–1719, Sep. 2010, <https://doi.org/10.1016/j.patrec.2010.05.021>.
- [9] "The Daily News 1924," *The Daily News 1924*. [Online]. Available: <http://nla.gov.au/nla.news-article82562896>.
- [10] T. Shinzaki, "Use Case of Palm Vein Authentication," in *Advances in Computer Vision and Pattern Recognition*, pp. 145–158, 2020, https://doi.org/10.1007/978-3-030-27731-4_5.
- [11] H. Qin, M. A. El Yacoubi, "End-to-End Generative Adversarial Network for Palm-Vein Recognition," in *Lecture Notes in Computer Science*, pp. 714–724, 2020, https://doi.org/10.1007/978-3-030-59830-3_62.
- [12] W. Wu, S. J. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biom.*, vol. 9, no. 1, pp. 1–10, Jan. 2020, <https://doi.org/10.1049/iet-bmt.2019.0034>.
- [13] X. Yang, "An Overview of the Attention Mechanisms in Computer Vision," *J. Phys. Conf. Ser.*, vol. 1693, no. 1, p. 012173, Dec. 2020, <https://doi.org/10.1088/1742-6596/1693/1/012173>.
- [14] J. Lou, J. zou, and B. Wang, "Palm Vein Recognition via Multi-task Loss Function and Attention Layer," *arXiv: arXiv:2211.05970*, 2022, <https://doi.org/10.48550/arXiv.2211.05970>.
- [15] H. I. Abdulrazzaq and R. D. Al-Dabbagh, "Biometric Identification System Based on Contactless Palm-Vein Using Residual Attention Network," *Iraqi J. Sci.*, pp. 1802–1810, Apr. 2022, <https://doi.org/10.24996/ijcs.2022.63.4.37>.
- [16] H. Liao, X. Jin, H. Zhu, Y. Fu, M. A. El Yacoubi, and H. Qin, "GAN et: Gabor Attention Aggregation Network for Palmvein Identification," in *16th International Conference on Human System Interaction (HSI)*, pp. 1–6, 2024, <https://doi.org/10.1109/hsi61632.2024.10613571>.
- [17] A. S. M. Htet and H. J. Lee, "Contactless Palm Vein Recognition Based on Attention-Gated Residual U-Net and ECA-ResNet," *Appl. Sci.*, vol. 13, no. 11, p. 6363, May 2023, <https://doi.org/10.3390/app13116363>.
- [18] P. Wang and H. Qin, "Palm-vein verification based on U-Net," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 806, no. 1, p. 012043, Apr. 2020, <https://doi.org/10.1088/1757-899x/806/1/012043>.
- [19] G. R. Nayar and T. Thomas, "Partial palm vein based biometric authentication," *J. Inf. Secur. Appl.*, vol. 72, p. 103390, Feb. 2023, <https://doi.org/10.1016/j.jisa.2022.103390>.
- [20] S. Sun, X. Cong, P. Zhang, B. Sun, and X. Guo, "Palm Vein Recognition Based on NPE and KELM," *IEEE Access*, vol. 9, pp. 71778–71783, 2021, <https://doi.org/10.1109/ACCESS.2021.3079458>.
- [21] S. Khandelwal and L. Sigal, "AttentionRNN: A Structured Spatial Attention Mechanism," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3424–3433, 2019, <https://doi.org/10.1109/iccv.2019.00352>.

- [22] G. K. Ong Michael, T. Connie, and A. B. Jin Teoh, "A Contactless Biometric System Using Palm Print and Palm Vein Features," in *Advanced Biometric Technologies*, pp. 155-177, 2011. <https://doi.org/10.5772/19337>.
- [23] M. K. O. Goh, C. Tee, and A. B. J. Teoh, "Bi-Modal Palm Print And Knuckle Print Recognition System," *J. IT Asia*, vol. 3, no. 1, pp. 85–106, Apr. 2016, <https://doi.org/10.33736/jita.37.2010>.
- [24] G. K. O. Michael, T. Connie, T. C. Chin, N. H. Foon, and A. T. B. Jin, "Realizing Hand-Based Biometrics Based on Visible and Infrared Imagery," in *Neural Information Processing. Models and Applications*, pp. 606–615, 2010, https://doi.org/10.1007/978-3-642-17534-3_75.
- [25] "CASIA-MS-PalmprintV1." [Online]. Available: <http://biometrics.idealtest.org/>.
- [26] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *15th IEEE International Conference on Image Processing*, pp. 281–284, 2008, <https://doi.org/10.1109/ICIP.2008.4711746>.
- [27] Y. Hao, Z. Sun, T. Tan, "Comparative Studies on Multispectral Palm Image Fusion for Biometrics," in *Lecture Notes in Computer Science*, pp. 12–21, 2007, https://doi.org/10.1007/978-3-540-76390-1_2.
- [28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, <https://doi.org/10.1038/323533a0>.
- [29] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982, <https://doi.org/10.1073/pnas.79.8.2554>.
- [30] J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," *Mach. Learn.*, vol. 7, no. 2–3, pp. 195–225, Sep. 1991, <https://doi.org/10.1007/bf00114844>.
- [31] S. Das, A. Tariq, T. Santos, S. S. Kantareddy, I. Banerjee, "Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research," in *Neuromethods*, pp. 117–138, 2023, https://doi.org/10.1007/978-1-0716-3195-9_4.
- [32] X. Li *et al.*, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *Int. J. Netw. Dyn. Intell.*, pp. 93–116, Mar. 2023, <https://doi.org/10.53941/ijndi0201006>.
- [33] P. K. Sekharamanthy, F. Melgani, and J. Malacarne, "Deep Learning-Based Apple Detection with Attention Module and Improved Loss Function in YOLO," *Remote Sens.*, vol. 15, no. 6, p. 1516, Mar. 2023, <https://doi.org/10.3390/rs15061516>.
- [34] M. Samo, J. M. Mafeni Mase, and G. Figueredo, "Deep Learning with Attention Mechanisms for Road Weather Detection," *Sensors*, vol. 23, no. 2, p. 798, Jan. 2023, <https://doi.org/10.3390/s23020798>.
- [35] S. P. Ramya and R. Eswari, "Attention-Based Deep Learning Models for Detection of Fake News in Social Networks," *Int. J. Cogn. Inform. Nat. Intell.*, vol. 15, no. 4, pp. 1–25, Jan. 2022, <https://doi.org/10.4018/ijcini.295809>.
- [36] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv: arXiv:1409.0473*, 2016, <https://doi.org/10.48550/arXiv.1409.0473>.
- [37] A. Vaswani *et al.*, "Attention Is All You Need," *arXiv: arXiv:1706.03762*, 2023, <https://doi.org/10.48550/arXiv.1706.03762>.
- [38] F. Rosenblatt, *The Perceptron, a Perceiving and Recognizing Automaton Project Para.* in Report: Cornell Aeronautical Laboratory. Cornell Aeronautical Laboratory, 1957. [Online]. Available: https://books.google.co.id/books?id=P_XGPgAACAAJ.
- [39] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958, <https://doi.org/10.1037/h0042519>.
- [40] N. Gupta, B. Kaushik, M. Khalid Imam Rahmani, and S. Anwar Lashari, "Performance Evaluation of Deep Dense Layer Neural Network for Diabetes Prediction," *Comput. Mater. Contin.*, vol. 76, no. 1, pp. 347–366, 2023, <https://doi.org/10.32604/cmc.2023.038864>.
- [41] V. Krasteva, I. Christov, S. Naydenov, T. Stoyanov, and I. Jekova, "Application of Dense Neural Networks for Detection of Atrial Fibrillation and Ranking of Augmented ECG Feature Set," *Sensors*, vol. 21, no. 20, p. 6848, Oct. 2021, <https://doi.org/10.3390/s21206848>.
- [42] V. L. Helen Josephine, A. P. Nirmala, and V. L. Alluri, "Impact of Hidden Dense Layers in Convolutional Neural Network to enhance Performance of Classification Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1131, no. 1, p. 012007, Apr. 2021, <https://doi.org/10.1088/1757-899x/1131/1/012007>.
- [43] "Confusion Matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, pp. 209–209, 2011, https://doi.org/10.1007/978-0-387-30164-8_157.
- [44] A. Gumaei, R. Sammouda, A. Al-Salman, and A. Alsanad, "An Effective Palmprint Recognition Approach for Visible and Multispectral Sensor Images," *Sensors*, vol. 18, no. 5, p. 1575, May 2018, <https://doi.org/10.3390/s18051575>.
- [45] E. Thamri, K. Aloui, and M. S. Naceur, "Improving Palmprint based Biometric System Performance using Novel Multispectral Image Fusion Scheme," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, 2020, [Online]. Available: <https://api.semanticscholar.org/CorpusID:222130875>.
- [46] A. Kumar and K. V. Prathyusha, "Personal Authentication Using Hand Vein Triangulation and Knuckle Shape," *IEEE Trans. Image Process.*, vol. 18, pp. 2127–2136, 2009, <https://doi.org/10.1109/TIP.2009.2023153>.
- [47] A. Kumari, B. Alankar, and J. Grover, "Feature Level Fusion of Multispectral Palmprint," *Int. J. Comput. Appl.*, vol. 144, no. 3, pp. 41–46, Jun. 2016, <https://doi.org/10.5120/ijca2016910175>.

BIOGRAPHY OF AUTHORS

Indriani is a lecturer at Widyatama University in the Informatics Engineering Program, Indonesia. She completed her Bachelor's degree in Informatics Engineering at Universitas Pasundan (2006–2011) and her Master's degree in Information Science and Manufacturing Engineering at Ashikaga University, Japan (2014–2016). Her research interests include computer vision, image processing, and artificial intelligence. Email: indriani.st@widyatama.ac.id.



Yenie Syukriyah is a lecturer at Widyatama University in the Informatics Engineering Program, in Indonesia. She earned her Master's degree in 2013 and her Doctoral Degree in 2021 from the Mathematics Department at Bandung Institute of Technology, Indonesia. Her research interests include mathematical modeling. Email: yenie.syukriyah@widyatama.ac.id.