

Handwriting-Based Alzheimer's Disease Detection Using VGG-19

Nicholas Wilyanto ^{1*}, Derry Alamsyah ^{2**}

* Department of Informatics, Universitas Multi Data Palembang

** Faculty of Computer Science and Engineering, Universitas Multi Data Palembang

nicholas.wilyanto@mhs.mdp.ac.id¹, derry@mdp.ac.id²

Article Info

Article history:

Received 2026-02-04

Revised 2026-02-21

Accepted 2026-04-08

Keyword:

*Alzheimer's Disease,
Convolutional Neural Network,
Early Detection,
Handwriting,
Image Preprocessing,
VGG-19.*

ABSTRACT

Alzheimer's disease is a progressive neurological disorder that impairs cognitive function and daily activities. Conventional diagnostic methods are often invasive, costly, and reliant on specialized clinical facilities, limiting their suitability for large-scale early screening. This study proposes a handwriting-based early detection approach using a Convolutional Neural Network built on the VGG-19 architecture. Experiments were conducted on the DARWIN-I dataset comprising 174 participants (89 Alzheimer's patients and 85 healthy controls) across two handwriting tasks (TASK_04 and TASK_05), considering both raw and cleansed image subsets. Preprocessing strategies included standard VGG-19 preprocessing as well as enhanced techniques combining Gaussian Blur, normalization, CLAHE, and Otsu thresholding. Hyperparameter optimization was performed using Hyperband, followed by 5-fold stratified cross-validation to ensure stable performance. The best configuration was retrained and evaluated on a hold-out test set. The highest experimental accuracies reached 87% for TASK_04 and 90% for TASK_05. For deployment, the final model was selected based on stability and robustness, achieving 81% and 90% accuracy for TASK_04 and TASK_05, respectively. The selected model was deployed in a cloud-based environment and integrated into a mobile application for real-time prediction. A usability evaluation involving 21 participants showed positive responses in usefulness, ease of use, ease of learning, and satisfaction. These findings demonstrate that the proposed approach is effective and suitable for supporting practical early-stage Alzheimer's disease screening.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Alzheimer's disease is a long-term neurodegenerative disorder and is widely recognized as the primary cause of dementia, accounting for approximately 60–80% of dementia cases worldwide [1], [2]. The number of individuals affected by dementia continues to increase globally, with approximately 55.2 million cases reported in 2019 and projections estimating a rise to 139 million by 2050 [3]. A similar trend is observed in Indonesia, where the number of people living with dementia is expected to increase significantly from about 1.2 million in 2016 to nearly 4 million by 2050 [4]. Despite extensive research, there is currently no cure. Medications such as donepezil, rivastigmine, and galantamine which are categorized as

acetylcholinesterase inhibitors only help slow the rate of cognitive decline rather than stop the disease entirely [5], [6], [7]. Because of these limitations, early detection has grown increasingly important to support timely intervention, improve treatment planning, and help preserve patient quality of life [6], [8], [9], [10], [11].

However, many diagnostic procedures currently depend on expensive neuroimaging, detailed clinical assessments, or invasive procedures such as lumbar puncture, which require specialized expertise and equipment, making early detection inaccessible for many communities [10], [12], [13], [14]. These challenges underscore the need for screening approaches that are cost-effective, non-invasive, and simple to deploy. Handwriting analysis has emerged as one such alternative, based on the understanding that Alzheimer's

disease affects cognitive processing and fine motor function, which naturally manifest in handwriting performance [6], [15], [16], [17]. This creates opportunities for practical and efficient early screening using machine learning models trained on handwriting data.

Recent research has reinforced this direction. Cilia et al. [18] evaluated transfer learning models such as VGG-19, ResNet50, InceptionV3, and InceptionResNetV2 using the DARWIN dataset, one of the largest public handwriting datasets for Alzheimer's research. Their findings showed that VGG-19 and ResNet50 achieved the strongest results with 72.26 percent accuracy on the first tracing task and 75.17 percent on the second. The findings indicate that deep learning-based models are capable of identifying subtle variations in handwriting patterns that distinguish individuals with Alzheimer's disease from healthy control subjects. Erdogmus and Kabakus [19] expanded this work by introducing a lightweight twelve-layer CNN designed specifically for handwriting-based screening, trained on all twenty-five DARWIN tasks. Their model reached an accuracy of 90.4 percent, showing that tailored CNN architectures can perform well on handwriting-based cognitive assessment. Complementary findings were reported by Sweidan et al. [20], who introduced a real-time framework that employs one-dimensional CNNs to identify early-stage Alzheimer's disease using dynamic handwriting signal data. Their subject based soft voting method achieved an accuracy of 94.44 percent, demonstrating that both static and dynamic handwriting cues hold diagnostic potential.

Collectively, these studies highlight handwriting analysis as a non-invasive approach with growing reliability for early Alzheimer's disease screening. They also point to the need for refinement in CNN design and preprocessing strategies, especially when working with varied handwriting tasks and digital input sources, which collectively motivates the present work.

To better understand the methodological foundation of this study, it is important to highlight the role of deep learning approaches in handwriting-based analysis.

Handwriting-based Alzheimer's disease detection requires analytical methods capable of capturing subtle visual and motor-related variations in handwriting patterns. Handwriting images reflect fine motor skills that tend to deteriorate in individuals with Alzheimer's disease, making them suitable for image-based cognitive assessment [20]. Convolutional Neural Networks (CNNs) are particularly suitable for handwriting-based analysis because of their strong capability in image classification and their robustness to common variations such as translation, rotation, and scale [21]. In contrast to conventional methods that depend on handcrafted feature extraction, CNNs learn discriminative representations directly from input images, thereby reducing subjectivity while improving computational efficiency and classification reliability [22] – [24].

Among various CNN architectures, deep transfer learning models such as VGG-19, ResNet50, and Inception-based networks have demonstrated strong performance in medical

image classification tasks. Prior comparative studies on the DARWIN dataset have shown that VGG-19 performs competitively alongside other deep architectures, particularly ResNet50, in handwriting-based Alzheimer's detection [18].

Considering its stable performance, architectural simplicity, and proven transfer learning capability, VGG-19 was selected as the backbone model in this study. Rather than benchmarking multiple CNN architectures, this research focuses on systematically analyzing the impact of data cleansing and preprocessing strategies under a consistent deep learning framework. Maintaining a fixed architecture enables controlled evaluation of how input data quality and preprocessing variations influence classification stability and real-world deployment performance.

The depth of VGG-19 allows the model to capture hierarchical visual representations with increasing levels of abstraction, which is beneficial for analyzing subtle structural variations in handwriting patterns [25] – [28]. Furthermore, its compatibility with pre-trained ImageNet weights facilitates effective knowledge transfer when working with limited medical datasets [29]. These characteristics make VGG-19 a reliable and methodologically appropriate choice for the objectives of this study.

Although earlier studies have demonstrated that deep learning models can recognize Alzheimer's disease through handwriting, several challenges still remain. Some existing works rely on a limited range of handwriting tasks or employ architectures that are computationally heavy, making real-world deployment difficult. In addition, many studies do not examine how well their models perform beyond controlled experimental environments, such as in mobile or web-based applications where handwriting data may vary in quality because of lighting, noise, and paper characteristics. At the same time, preprocessing techniques such as noise reduction, contrast enhancement, and thresholding have not been consistently evaluated, even though these steps can substantially influence prediction quality. In particular, there is limited analysis on how different preprocessing strategies affect CNN-based handwriting classification when applied to real-world mobile data. These gaps suggest the need for solutions that are accurate, consistent, lightweight, and adaptable to digital handwriting collected from everyday devices.

With these challenges in mind, this research focuses on developing and evaluating a VGG-19 based convolutional neural network for early Alzheimer's detection using handwriting images. Multiple preprocessing strategies are systematically evaluated to assess their impact on classification performance while maintaining a consistent model architecture. To ensure practical relevance, the best-performing configuration is deployed as a mobile screening application that integrates cloud-based prediction and automated reporting. The overall goal is to develop a non-invasive, efficient, and accessible handwriting-based detection system that supports early screening and enables

earlier intervention in a practical, affordable, and widely deployable way.

II. METHODS

This study adopts a quantitative experimental approach to develop and evaluate a handwriting-based Alzheimer's disease detection model based on the VGG-19 convolutional neural network architecture. The main objective is to systematically investigate how variations in data quality and image preprocessing techniques affect classification performance. All experiments utilize the DARWIN-I dataset, a publicly available handwriting dataset commonly used for cognitive assessment related to Alzheimer's disease. The dataset is divided into two subsets: original handwriting images and manually cleansed images that have been refined to minimize noise, background artifacts, and irrelevant visual elements. This dataset organization enables a structured evaluation of the impact of image quality on model robustness and prediction accuracy.

To investigate the impact of image preprocessing on classification performance, multiple preprocessing strategies are evaluated. These strategies include the standard preprocessing procedure required by the VGG-19 architecture, as well as several combinations with additional image enhancement techniques, namely Gaussian Blur, Normalization, Contrast Limited Adaptive Histogram Equalization (CLAHE), and Otsu Thresholding. Each preprocessing configuration is applied consistently across both raw and cleansed datasets. This experimental design enables a systematic and controlled comparison of preprocessing effects while maintaining a fixed model architecture and learning framework.

Following preprocessing, the dataset is divided into training-validation and test subsets to support an objective performance evaluation. Model development is carried out through hyperparameter optimization using the Hyperband algorithm, which efficiently explores the search space to identify promising configurations. The selected configurations are further evaluated using 5-fold Stratified Cross Validation to assess model stability and generalization performance across different data splits. The most promising hyperparameter configurations are retrained using the full training dataset and then assessed on a separate hold-out test set. The final model is determined based on its performance on this unseen data.

Finally, the selected model is deployed to a cloud-based environment hosted on Microsoft Azure using Docker containerization. The deployed model is exposed through a web-based API, enabling seamless integration with the Android-based MindTrace application for real-time handwriting-based Alzheimer's disease prediction.

The overall research workflow is illustrated in Figure 1, which summarizes the main stages of the proposed methodology, including data preparation, preprocessing scenarios, model training and validation, performance evaluation, and system deployment.

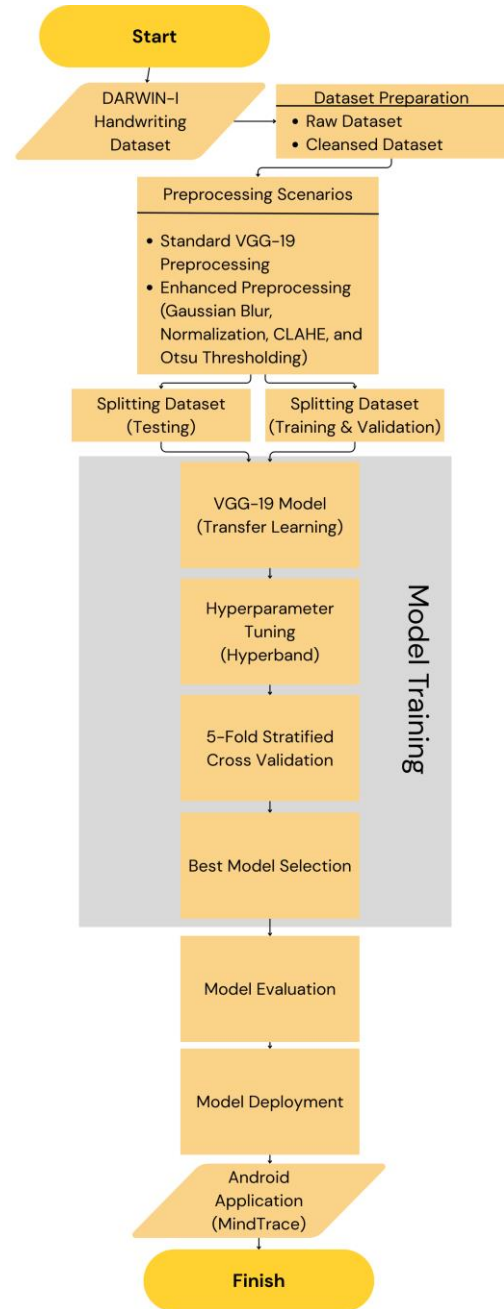


Fig. 1. Flowchart of the proposed Alzheimer's disease classification framework based on handwriting images.

A. Dataset Description

The DARWIN-I dataset is employed in this study as the primary source of handwriting data. This publicly available dataset was developed specifically for Alzheimer's disease research and was first introduced by Cilia et al. [18]. It contains handwriting samples from 174 participants, comprising 89 individuals diagnosed with Alzheimer's disease and 85 healthy control subjects. Participant ages range from 44 years and above. The Alzheimer's disease group has an average age of 71.5 years with a standard deviation of 9.5, whereas the healthy control group has an

average age of 68.9 years with a standard deviation of 12 [16].

DARWIN-I provides offline handwriting images obtained from paper-based drawing tasks, where participants performed predefined graphical activities using conventional writing tools. In this research, only the offline image modality is used, as it better represents natural handwriting characteristics and is more suitable for practical screening scenarios outside controlled laboratory environments. From the available handwriting tasks in the dataset, this study focuses on TASK_04 and TASK_05, which involve repeatedly drawing circular shapes with different diameters. The selection of these two tasks was based on both methodological and clinical considerations. In addition to balanced class representation and consistent image quality, the choice was supported by expert consultation with a neurologist specializing in cognitive disorders. According to the expert recommendation, TASK_04 and TASK_05 are considered more sensitive for detecting fine motor deterioration associated with neurodegenerative conditions.

Both tasks are designed to evaluate movement automation, regularity, and coordination in sequential hand movements. In TASK_04, participants draw circles with a diameter of 6 cm, while in TASK_05 the diameter is reduced to 3 cm. Each circle is drawn four consecutive times. The smaller diameter in TASK_05 requires higher movement precision and visuospatial control, which may increase sensitivity to subtle motor impairment.

This approach is consistent with findings reported by Cilia et al. [18], which indicate that writing-based tasks may be influenced by individual writing habits and educational background, potentially masking cognitive impairment. In contrast, structured graphical tasks such as repetitive circle drawing provide a more objective representation of motor and cognitive decline.

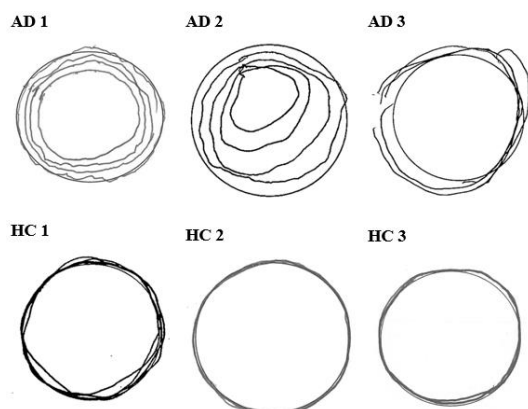


Fig. 2. Example handwriting samples from the DARWIN-I dataset for TASK_04 (circle drawing with a diameter of 6 cm). Samples are shown for Alzheimer's disease (AD) and healthy control (HC) participants.

Other available tasks in the dataset were not included in order to maintain experimental consistency and reduce task heterogeneity, thereby enabling a more controlled evaluation of preprocessing strategies and model performance. Examples of handwriting samples for TASK_04 and

TASK_05 from both Alzheimer's disease (AD) and healthy control (HC) participants are illustrated in Figure 2 and Figure 3, respectively.

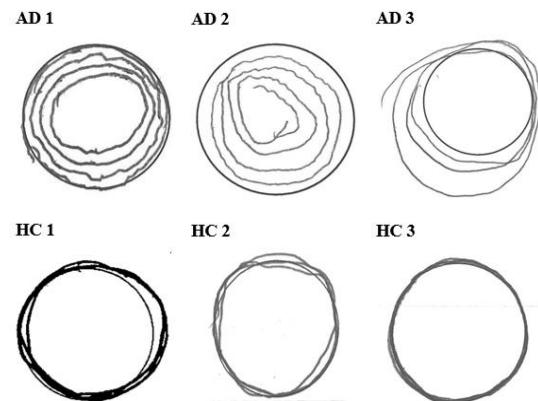


Fig. 3. Example handwriting samples from the DARWIN-I dataset for TASK_05 (circle drawing with a diameter of 3 cm). Samples are shown for Alzheimer's disease (AD) and healthy control (HC) participants.

To examine the influence of data quality on model performance, the dataset is organized into two subsets: raw images and cleansed images. The cleansed images have been manually refined to reduce noise and artifacts caused by scanning conditions and paper variations, while the raw images are retained in their original form. This separation allows a systematic evaluation of data cleansing when combined with different preprocessing strategies.

B. Data Preprocessing

Prior to model training, all handwriting images are preprocessed to ensure compatibility with the VGG-19 architecture and to enhance discriminative visual features. All handwriting images are resized to 224×224 pixels to match the input requirements of the VGG-19 architecture. The color channels are rearranged from RGB to BGR, and pixel intensities are standardized by subtracting the ImageNet mean values of [123.68, 116.779, 103.939] for each channel, in accordance with the conventional VGG-19 preprocessing pipeline [30] – [31].

In this study, two preprocessing approaches are evaluated. The first approach applies only the standard VGG-19 preprocessing to the input images. The second approach introduces additional image enhancement techniques prior to the VGG-19 preprocessing step to improve the quality and highlight the handwritten patterns from the images. These techniques include Gaussian Blur for noise reduction, Normalization to balance local intensity variations, Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance local contrast, and Otsu Thresholding to separate handwriting strokes from the background [32] – [36].

All preprocessing configurations are applied consistently across both raw and cleansed subsets of the DARWIN-I dataset. This experimental design allows a structured comparison of different preprocessing strategies while examining the influence of data quality and image enhancement techniques on classification performance.

C. Experimental Scenarios

This study defines four experimental scenarios to examine how data quality and preprocessing strategies influence classification performance. These scenarios are constructed by combining two dataset conditions (raw and cleansed) with two preprocessing approaches (standard and enhanced).

The first scenario utilizes raw handwriting images from the DARWIN-I dataset and applies only the standard VGG-19 preprocessing. This scenario serves as a baseline to evaluate model performance without data cleansing or additional image enhancement.

The second scenario applies the standard VGG-19 preprocessing to the cleansed dataset, allowing an assessment of the impact of data cleansing while maintaining a conventional preprocessing pipeline.

The third scenario uses raw images combined with the enhanced preprocessing approach, which incorporates additional image enhancement techniques prior to the standard VGG-19 preprocessing. This scenario evaluates the effectiveness of preprocessing enhancements when applied to unrefined data.

The fourth scenario combines the cleansed dataset with the enhanced preprocessing approach, representing the most optimized experimental condition. This scenario is designed to evaluate the cumulative effect of data cleansing and advanced preprocessing on classification performance.

All experimental scenarios follow the same model architecture, training strategy, and evaluation protocol to ensure fair comparison. The differences in model performance across scenarios are therefore attributable solely to variations in data condition and preprocessing strategy.

D. Model Architecture

This study employs a Convolutional Neural Network (CNN) based on the VGG-19 architecture for handwriting-based Alzheimer's disease classification. A pre-trained VGG-19 model initialized with ImageNet weights is used as the backbone to leverage transfer learning and improve convergence on a limited dataset [37]. The overall architecture of the proposed model is illustrated in Figure 4.

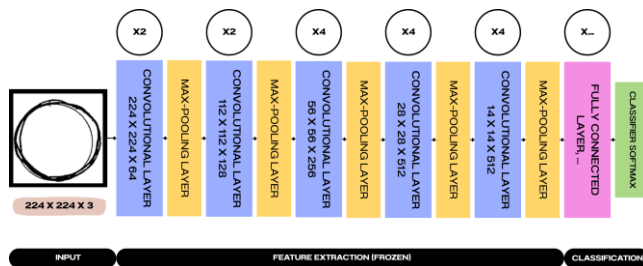


Fig. 4. Architecture of the modified VGG-19 model used for handwriting-based Alzheimer's disease classification.

The convolutional layers of VGG-19 are retained and frozen to preserve generic visual feature representations learned from large-scale data, while the original fully connected layers are replaced with custom classification layers. The modified architecture consists of a pooling layer

followed by one or two fully connected layers with ReLU activation, depending on the task configuration. A softmax activation function is used in the output layer to perform binary classification between Alzheimer's disease and healthy control classes.

Dropout regularization is incorporated into the fully connected layers to mitigate overfitting, with dropout rates varying between 0.1 and 0.3 across different experimental settings. Model training is performed using the Adam optimization algorithm, while categorical cross-entropy is employed as the loss function, which aligns with the softmax-based output used for both multi-class and binary classification tasks in this study.

E. Hyperband Tuning and Validation

Hyperparameter optimization is performed to identify the most effective model configuration for each handwriting task. This study employs the Hyperband algorithm, implemented using Keras Tuner, to efficiently search the hyperparameter space by allocating more training resources to promising configurations while early-stopping underperforming ones [38].

The hyperparameters explored include the number of fully connected layers, number of neurons, dropout rate, and learning rate, with learning rates evaluated within the range of 0.00001 to 0.001. All models are trained using the Adam optimizer with a fixed batch size of 16. Training is conducted for a maximum of 20 epochs, with early stopping applied using a patience value of 5 to prevent overfitting.

Following the Hyperband optimization, the five best-performing hyperparameter configurations are selected for further evaluation. Each configuration is evaluated using 5-fold stratified cross-validation, with class proportions preserved across all folds. In each fold, four subsets are used for training and one subset for validation, and the process is repeated until all subsets have served as validation data [39].

The performance of each candidate configuration is summarized by computing the mean validation accuracy and standard deviation across the five folds. Based on these results, the most stable and well-performing configurations are identified and selected for final training and evaluation.

F. Model Evaluation

The classification performance of the selected model is evaluated using unseen data. Once the optimal hyperparameter setting is identified, the model is retrained on the complete training dataset and tested on an independent hold-out set comprising 20% of the overall data, which is not involved in any training or validation process.

The evaluation is performed using a confusion matrix, from which standard performance metrics are derived, including accuracy, recall, precision, and F1-score. The mathematical formulations of these evaluation metrics are presented in Equations (1)-(4) [40].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

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

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. These evaluation metrics are used to measure how effectively the model differentiates between individuals with Alzheimer's disease and healthy control subjects.

Evaluation is carried out separately for TASK_04 and TASK_05, as well as for each experimental scenario defined in this study. This approach enables a detailed comparison of classification performance across different preprocessing strategies and data conditions.

G. Deployment and Implementation

The final selected model is deployed in a cloud-based environment hosted on Microsoft Azure using Docker containerization to ensure portability, scalability, and reproducible execution. The trained VGG-19-based classification model is packaged within a Docker container and deployed as a FastAPI-based inference service, which exposes a RESTful endpoint for handwriting image classification. The inference service receives handwriting image inputs and returns classification results in JSON format.

A RESTful API gateway is implemented using the Laravel framework and deployed on Vercel to support communication between the mobile application and the inference server. This API gateway is responsible for handling client requests, performing authentication and request validation, forwarding image data to the cloud-based inference service, and managing prediction results returned by the model.

An Android-based mobile application, MindTrace, is developed as the client-side interface of the proposed system. The application allows users to capture or upload handwriting images, which are then transmitted to the Laravel-based API gateway for real-time prediction. The API forwards the image data to the deployed inference server on Azure, receives the prediction results, and returns the final classification output to the mobile application.

In addition, Supabase, which is built on PostgreSQL, is utilized as the backend data storage service to manage user information, prediction history, and generated report metadata. This separation between inference processing and data management enables centralized maintenance, persistent storage, and efficient system scalability.

Figure 5 illustrates the overall deployment architecture of the proposed system, highlighting the interaction between the MindTrace Android application, the Laravel-based API gateway, the cloud-based FastAPI inference server, and the Supabase database. This deployment framework follows a remote inference approach, enabling centralized model updates and practical integration for real-world early Alzheimer's disease screening applications.

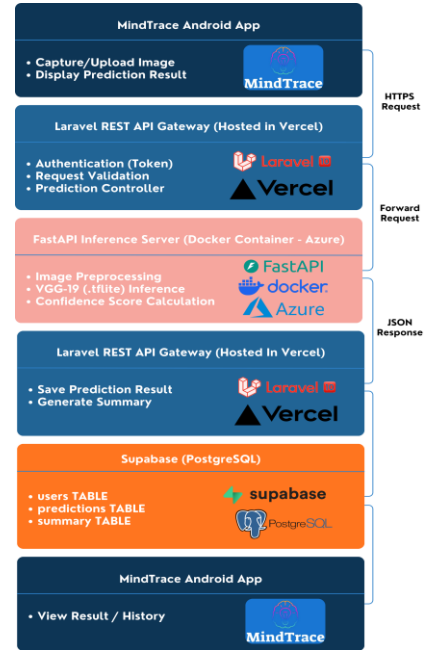


Fig. 5. Deployment architecture of the proposed MindTrace system, illustrating the interaction between the Android application, Laravel-based API gateway, cloud-based FastAPI inference server, and Supabase database.

H. System Usability Evaluation

A system usability evaluation was conducted to assess user interaction and acceptance of the developed mobile application. The evaluation was carried out after the system deployment using the USE Questionnaire, which measures key usability aspects, including ease of learning, ease of use, usefulness, and user satisfaction [41]. The collected responses were quantitatively analyzed, and the detailed results are presented and further discussed in relation to system deployment and practical usage.

III. RESULTS AND DISCUSSION

This section reports the experimental findings and provides a discussion of the proposed Alzheimer's disease detection system, with emphasis on classification performance across four experimental approaches applied to the TASK_04 and TASK_05 datasets. The results are discussed per approach to highlight the impact of preprocessing strategies and hyperparameter optimization on model performance.

In this study, recall for the Alzheimer's disease (AD) class is interpreted as sensitivity, while recall for the healthy control (HC) class corresponds to specificity. This interpretation is clinically relevant in early screening contexts, where minimizing false negatives (i.e., maximizing sensitivity) is particularly important.

A. Scenario 1: Raw Dataset with Standard VGG-19 Preprocessing

Scenario 1 represents the baseline experimental condition in this study. In this scenario, the model is trained using raw

handwriting images from the DARWIN-I dataset without any data cleansing or additional image enhancement techniques. Only the standard VGG-19 preprocessing is applied. The objective of this scenario is to establish a baseline performance that serves as a reference for evaluating the impact of data cleansing and enhanced preprocessing strategies in subsequent scenarios.

Table 1 summarizes the classification performance of the final selected model for Scenario 1 on TASK_04 and TASK_05, evaluated using accuracy, precision, recall, and F1-score.

TABLE I
CLASSIFICATION PERFORMANCE OF SCENARIO 1 ON TASK_04 AND 05

TASK	Accuracy	Precision	Recall	F1-Score
TASK_04	79%	79%	79%	79%
TASK_05	76%	77%	77%	76%

As shown in Table 1, the model achieves an accuracy of 79% on TASK_04, with balanced precision, recall, and F1-score values. This result indicates that, even when trained on raw handwriting images without prior data cleansing or enhancement, the model is capable of distinguishing between Alzheimer's disease patients and healthy control subjects with reasonable effectiveness.

A more detailed analysis of the classification behavior for TASK_04 is illustrated by the confusion matrix shown in Figure 6. Based on Figure 6, the model correctly classifies 14 out of 17 Alzheimer's disease patients and 12 out of 16 healthy individuals. In terms of recall, the Alzheimer's disease class reaches a value of 82%, while the healthy control class records a recall of 75%, indicating that the model is slightly more sensitive in detecting Alzheimer's disease cases. Regarding precision, the Alzheimer's disease class achieves a value of 78%, whereas the healthy control class attains a precision of 80%. These results suggest that the model's predictions for both classes are relatively reliable and that no strong bias toward a particular class is observed under baseline conditions.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	14 42.42%	3 9.09%	17 82.35% 17.65%
Healthy	4 12.12%	12 36.36%	16 75.00% 25.00%
SUM	18 77.78% 22.22%	15 80.00% 20.00%	26 / 33 78.79% 21.21%

Fig. 6. Confusion matrix of Scenario 1 on TASK_04.

For TASK_05, as shown in Table 1, the model achieves an accuracy of 76%, with precision, recall, and F1-score values that remain relatively balanced, although slightly lower compared to TASK_04. The corresponding confusion matrix is presented in Figure 7, which shows that the model correctly identifies 13 out of 18 Alzheimer's disease patients and 13 out of 16 healthy individuals. The recall values for the Alzheimer's disease and healthy control classes are 72% and 81%, respectively, indicating a more balanced sensitivity between the two classes compared to TASK_04. In terms of precision, the model records a value of 81% for the Alzheimer's disease class and 72% for the healthy control class, suggesting that predictions related to Alzheimer's disease are made with slightly higher confidence.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	13 38.24%	5 14.71%	18 72.22% 27.78%
Healthy	3 8.82%	13 38.24%	16 81.25% 18.75%
SUM	16 81.25% 18.75%	18 72.22% 27.78%	26 / 34 76.47% 23.53%

Fig. 7. Confusion matrix of Scenario 1 on TASK_05.

The reduced performance observed in TASK_05 can be attributed to the smaller circle diameter (3 cm), which imposes higher precision requirements on repetitive motor movements. This condition increases stroke density, overlap, and sensitivity to background noise and scanning artifacts when no additional preprocessing is applied.

Overall, the results of Scenario 1 demonstrate that standard VGG-19 preprocessing alone is sufficient to achieve moderate and stable classification performance on handwriting-based Alzheimer's disease detection. The observed performance limitations, particularly for TASK_05, indicate that data quality plays a critical role and justify the use of data cleansing and image enhancement techniques in the subsequent experimental scenarios.

B. Scenario 2: Cleansed Dataset with Standard VGG-19 Preprocessing

Scenario 2 evaluates the impact of data cleansing on classification performance while maintaining the standard VGG-19 preprocessing pipeline. In this scenario, handwriting images from the DARWIN-I dataset are manually cleansed to reduce noise and scanning artifacts before being processed using standard VGG-19 preprocessing. The objective of this scenario is to isolate the

effect of improved data quality without introducing additional image enhancement techniques.

Table 2 summarizes the classification performance of the final selected model for Scenario 2 on TASK_04 and TASK_05, evaluated using accuracy, precision, recall, and F1-score.

TABLE II
CLASSIFICATION PERFORMANCE OF SCENARIO 2 ON TASK_04 AND 05

TASK	Accuracy	Precision	Recall	F1-Score
TASK_04	87%	87%	87%	87%
TASK_05	86%	87%	85%	86%

As shown in Table 2, Scenario 2 demonstrates a substantial improvement in classification performance compared to the baseline scenario. For TASK_04, the model achieves an accuracy of 87%, with precision, recall, and F1-score values that are consistently high and well balanced across both classes. This improvement indicates that data cleansing significantly enhances the model’s ability to extract more consistent and discriminative handwriting features.

A more detailed analysis of the classification behavior for TASK_04 is illustrated by the confusion matrix shown in Figure 8. Based on Figure 8, the model correctly classifies 15 out of 17 Alzheimer’s disease patients and 12 out of 14 healthy control subjects. In terms of recall, both the Alzheimer’s disease and healthy control classes achieve a value of 88% and 86%, respectively, indicating balanced sensitivity across classes. Regarding precision, the Alzheimer’s disease class attains a value of 88%, while the healthy control class records 86%. These results confirm that the model does not exhibit bias toward either class and demonstrates stable and reliable classification behavior under improved data quality conditions.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	15 48.39%	2 6.45%	17 88.24% 11.76%
Healthy	2 6.45%	12 38.71%	14 85.71% 14.29%
SUM	17 88.24% 11.76%	14 85.71% 14.29%	27 / 31 87.10% 12.90%

Fig. 8. Confusion matrix of Scenario 2 on TASK_04.

For TASK_05, as shown in Table 2, the model achieves an accuracy of 86%, with high and well-balanced precision, recall, and F1-score values, indicating strong and consistent classification performance. The corresponding confusion

matrix is presented in Figure 9, which shows that the model correctly identifies 15 out of 16 Alzheimer’s disease patients and 10 out of 13 healthy control subjects. In terms of recall, the Alzheimer’s disease class reaches a value of 94%, demonstrating that the model is highly sensitive in detecting affected individuals. Conversely, the recall for the healthy control class is slightly lower at 77%, indicating a modest increase in false positive predictions. Regarding precision, the Alzheimer’s disease class records a value of 83%, while the healthy control class achieves a higher precision of 91%. This behavior suggests that the model tends to prioritize the detection of Alzheimer’s disease cases while maintaining a high level of confidence when predicting healthy individuals.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	15 51.72%	1 3.45%	16 93.75% 6.25%
Healthy	3 10.34%	10 34.48%	13 76.92% 23.08%
SUM	18 83.33% 16.67%	11 90.91% 9.09%	25 / 29 86.21% 13.79%

Fig. 9. Confusion matrix of Scenario 2 on TASK_05.

The performance characteristics observed in Scenario 2 indicate a clear improvement in model robustness across both handwriting tasks. In particular, the high recall achieved for Alzheimer’s disease in TASK_05 suggests that data cleansing enhances the model’s sensitivity in identifying affected individuals, albeit with a slight increase in false positive predictions. This behavior reflects a trade-off that is often acceptable in early screening applications, where prioritizing sensitivity is more critical than maximizing specificity. These findings highlight that improved input data quality can substantially stabilize model behavior across different handwriting tasks without the need for complex preprocessing pipelines.

Overall, the results of Scenario 2 indicate that data cleansing alone yields notable gains in classification performance across both handwriting tasks. Compared to Scenario 1, the increase in accuracy, improved class balance, and higher sensitivity toward Alzheimer’s disease detection demonstrate that input data quality plays a crucial role in handwriting-based Alzheimer’s disease detection systems. These findings highlight that even without additional image enhancement techniques, careful data preparation can substantially enhance model reliability and robustness, making this scenario particularly suitable for real-world and early screening applications.

C. Scenario 3: Raw Dataset with Enhanced Image Preprocessing

Scenario 3 evaluates the effectiveness of enhanced image preprocessing techniques when applied directly to raw handwriting images without prior data cleansing. In this scenario, additional image enhancement methods are incorporated before the standard VGG-19 preprocessing stage. The objective of this scenario is to assess whether preprocessing enhancements alone are sufficient to improve classification performance when dealing with unrefined data.

Table 3 summarizes the classification performance of the final selected model for Scenario 3 on TASK_04 and TASK_05, evaluated using accuracy, precision, recall, and F1-score.

TABLE III
CLASSIFICATION PERFORMANCE OF SCENARIO 3 ON TASK_04 AND 05

TASK	Accuracy	Precision	Recall	F1-Score
TASK_04	79%	79%	79%	79%
TASK_05	74%	74%	74%	74%

As shown in Table 3, the classification performance for TASK_04 reaches an accuracy of 79%, with balanced precision, recall, and F1-score values. This result indicates that the application of enhanced preprocessing on raw data is able to preserve classification performance at a level comparable to the baseline scenario.

A more detailed analysis of the classification behavior for TASK_04 is illustrated by the confusion matrix shown in Figure 10.

Based on Figure 10, the model correctly classifies 14 out of 17 Alzheimer’s disease patients and 12 out of 16 healthy control subjects. In terms of recall, the Alzheimer’s disease class achieves a value of 82%, while the healthy control class records a recall of 75%, indicating a slightly higher sensitivity toward detecting Alzheimer’s disease cases. Regarding precision, the Alzheimer’s disease class attains a value of 78%, whereas the healthy control class reaches 80%. These results suggest that the model maintains balanced predictive behavior across both classes despite the presence of noise in the raw dataset.

For TASK_05, as reported in Table 3, the model achieves an accuracy of 74%, with precision, recall, and F1-score values that are consistently moderate and slightly lower than those observed in TASK_04. The corresponding confusion matrix is presented in Figure 11, which shows that the model correctly identifies 13 out of 18 Alzheimer’s disease patients and 12 out of 16 healthy control subjects.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	14 42.42%	3 9.09%	17 82.35% 17.65%
Healthy	4 12.12%	12 36.36%	16 75.00% 25.00%
SUM	18 77.78% 22.22%	15 80.00% 20.00%	26 / 33 78.79% 21.21%

Fig. 10. Confusion matrix of Scenario 3 on TASK_04.

The recall values for the Alzheimer’s disease and healthy control classes are 72% and 75%, respectively, indicating a relatively balanced sensitivity between the two classes. In terms of precision, the Alzheimer’s disease class records a value of 76%, while the healthy control class achieves 71%. This behavior suggests a slight tendency toward improved recognition of healthy individuals, although overall performance remains moderate.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	13 38.24%	5 14.71%	18 72.22% 27.78%
Healthy	4 11.76%	12 35.29%	16 75.00% 25.00%
SUM	17 76.47% 23.53%	17 70.59% 29.41%	25 / 34 73.53% 26.47%

Fig. 11. Confusion matrix of Scenario 3 on TASK_05.

The results obtained in Scenario 3 indicate that enhanced preprocessing alone is not sufficient to substantially improve classification performance when applied to raw handwriting images. Compared to Scenario 1, the performance gains are minimal, and for TASK_05, the accuracy is slightly reduced. This finding suggests that preprocessing enhancements are more effective when combined with prior data cleansing, as noise and scanning artifacts in raw images may still limit the extraction of discriminative features.

Overall, the results of Scenario 3 demonstrate that while enhanced preprocessing can stabilize model performance on raw data, its impact remains limited without improved input data quality. The findings indicate that data cleansing serves as a necessary step before the application of more advanced

preprocessing strategies in handwriting-based Alzheimer’s disease detection systems.

D. Scenario 4: Cleansed Dataset with Enhanced Image Preprocessing

Scenario 4 represents the most comprehensive experimental setting in this study. In this scenario, handwriting images from the DARWIN-I dataset undergo both data cleansing and enhanced image preprocessing prior to the standard VGG-19 preprocessing stage. The objective of this scenario is to evaluate the combined impact of improved data quality and advanced preprocessing techniques on the classification performance of the proposed Alzheimer’s disease detection system.

Table 4 summarizes the classification performance of the final selected model for Scenario 4 on TASK_04 and TASK_05, evaluated using accuracy, precision, recall, and F1-score.

TABLE IV
CLASSIFICATION PERFORMANCE OF SCENARIO 4 ON TASK_04 AND TASK_05

TASK	Accuracy	Precision	Recall	F1-Score
TASK_04	81%	80%	80%	80%
TASK_05	90%	90%	90%	90%

As shown in Table 4, Scenario 4 yields competitive and stable classification performance for both handwriting tasks. For TASK_04, the model achieves an accuracy of 81%, with macro-average precision, recall, and F1-score values of 80%. This result indicates that the integration of data cleansing and enhanced preprocessing contributes to maintaining balanced performance across both classes, although the improvement over previous scenarios is moderate for this task.

A more detailed analysis of the classification behavior for TASK_04 is illustrated by the confusion matrix shown in Figure 12. Based on Figure 12, the model correctly classifies 14 out of 17 Alzheimer’s disease patients and 11 out of 14 healthy control subjects. In terms of recall, both the Alzheimer’s disease and healthy control classes achieve values of 82% and 79%, respectively, indicating comparable sensitivity toward both groups. Regarding precision, the Alzheimer’s disease class records a value of 82%, while the healthy control class achieves 79%. These results demonstrate a balanced predictive behavior, with no significant bias toward either class.

For TASK_05, as shown in Table 4, Scenario 4 achieves the highest classification performance among all evaluated scenarios, with an accuracy of 90% and equally high precision, recall, and F1-score values. The corresponding confusion matrix is presented in Figure 13, which shows that the model correctly identifies 14 out of 16 Alzheimer’s disease patients and 12 out of 13 healthy control subjects.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	14 45.16%	3 9.68%	17 82.35% 17.65%
Healthy	3 9.68%	11 35.48%	14 78.57% 21.43%
SUM	17 82.35% 17.65%	14 78.57% 21.43%	25 / 31 80.65% 19.35%

Fig. 12. Confusion matrix of Scenario 4 on TASK_04.

In terms of recall, the Alzheimer’s disease class attains a value of 88%, while the healthy control class records a higher recall of 92%, indicating excellent sensitivity in recognizing both affected and non-affected individuals. Regarding precision, the Alzheimer’s disease class achieves a value of 93%, whereas the healthy control class records 86%, reflecting high confidence in Alzheimer’s disease predictions with minimal misclassification.

TARGET \ OUTPUT	Alzheimer	Healthy	SUM
Alzheimer	14 48.28%	2 6.90%	16 87.50% 12.50%
Healthy	1 3.45%	12 41.38%	13 92.31% 7.69%
SUM	15 93.33% 6.67%	14 85.71% 14.29%	26 / 29 89.66% 10.34%

Fig. 13. Confusion matrix of Scenario 4 on TASK_05.

The strong performance observed in TASK_05 indicates that the combination of data cleansing and enhanced preprocessing is particularly effective for circular drawing tasks with smaller diameters, where higher stroke density and fine-grained motor variations make the images more sensitive to noise and contrast variations. The improved preprocessing pipeline enables the model to better capture relevant structural features, resulting in significantly reduced misclassification rates.

Overall, the results of Scenario 4 confirm that combining data cleansing with enhanced image preprocessing yields the most optimal performance for TASK_05 and stable, balanced performance for TASK_04. Compared to previous scenarios, this approach demonstrates improved robustness

and reliability, particularly for tasks that are highly affected by image quality variations. The results indicate that combining data quality improvement with preprocessing enhancements contributes to more effective handwriting-based Alzheimer's disease detection systems.

E. Comparative Discussion of All Scenarios

To provide a comprehensive comparison across all experimental scenarios, Table 5 summarizes the classification performance of the four scenarios for TASK_04 and TASK_05 using accuracy, precision, recall, and F1-score. This comparative analysis highlights the impact of data cleansing and enhanced preprocessing strategies on handwriting-based Alzheimer's disease classification.

TABLE V
SUMMARY OF CLASSIFICATION PERFORMANCE FOR ALL SCENARIOS

Scenario	TASK	Accuracy	Precision	Recall	F1-Score
1	04	79%	79%	79%	79%
	05	76%	77%	77%	76%
2	04	87%	87%	87%	87%
	05	86%	87%	85%	86%
3	04	79%	79%	79%	79%
	05	74%	74%	74%	74%
4	04	81%	80%	80%	80%
	05	90%	90%	90%	90%

For TASK_04, Scenario 2 demonstrates the best overall performance, achieving the highest accuracy of 87% with well-balanced precision and recall across both classes. This indicates that the application of data cleansing alone significantly improves input quality and enables the model to learn discriminative handwriting patterns more effectively. Scenario 4 also shows stable performance with an accuracy of 81%, suggesting that the combination of data cleansing and enhanced preprocessing remains beneficial, albeit with slightly lower gains compared to Scenario 2. In contrast, Scenarios 1 and 3, which rely on raw data without cleansing, exhibit lower accuracy and less balanced class-wise performance, particularly in identifying healthy subjects.

In TASK_05, Scenario 4 achieves the highest classification performance, reaching an accuracy of 90%. From a clinical screening perspective, this corresponds to a sensitivity of 88% (14/16 AD correctly identified) and a specificity of 92% (12/13 HC correctly identified). This result demonstrates the effectiveness of combining data cleansing with enhanced preprocessing techniques when handling more complex handwriting tasks. Scenario 2 also yields strong performance with an accuracy of 86% and notably high sensitivity of 94% for Alzheimer's disease cases. However, the specificity for the healthy class is lower at 77%, reflecting a trade-off that is commonly acceptable in early screening applications where prioritizing sensitivity is critical. Meanwhile, Scenarios 1 and 3 show comparable but lower accuracy, suggesting that enhanced preprocessing alone is insufficient when applied to uncleaned data.

Although TASK_04 and TASK_05 are both circular drawing tasks, the smaller diameter used in TASK_05 (3 cm) introduces higher motor precision demands and visual complexity compared to TASK_04 (6 cm). This difference increases stroke density and sensitivity to noise and minor distortions, which helps explain the observed variations in classification performance across scenarios, particularly the greater impact of enhanced preprocessing in TASK_05.

Overall, the comparative results across both tasks confirm that data cleansing plays a critical role in improving classification accuracy and model stability. Enhanced visual preprocessing further contributes to performance gains, particularly when combined with cleansed data. Among all evaluated approaches, Scenario 4 consistently delivers robust and superior results across TASK_04 and TASK_05, making it the most reliable approach for practical deployment. Consequently, this scenario is selected as the final model configuration to be implemented in the MindTrace mobile application for handwriting-based Alzheimer's disease screening.

Moreover, the robustness of Scenario 4 against noise and image quality variations supports its suitability for real-world mobile deployment, where handwriting data are collected under less controlled conditions.

Despite the promising performance observed across the evaluated scenarios, several potential sources of data bias and limitations should be acknowledged. The DARWIN-I dataset represents a specific population group with particular demographic characteristics, including age distribution and clinical diagnosis context. Variations in educational background, cultural factors, and individual handwriting habits across broader populations may influence handwriting patterns and consequently affect model generalization.

Furthermore, although the deployment experiments demonstrate stable performance, real-world handwriting acquisition may involve diverse lighting conditions, camera qualities, and environmental factors that differ from the controlled dataset setting. These variations may partially explain performance differences observed between experimental evaluation and deployment results.

Therefore, further validation using larger and more diverse multi-center datasets is necessary to comprehensively assess model robustness and generalizability across broader clinical and demographic contexts.

F. System Deployment

The best-performing model selected from Scenario 4 was successfully deployed in a cloud-based environment and integrated into the MindTrace Android application to support real-time handwriting-based Alzheimer's disease classification. The deployed system enables an end-to-end prediction workflow, starting from handwriting image acquisition on the mobile device, followed by cloud-based inference, and concluding with result delivery to the user interface.

The integration results demonstrate that handwriting images submitted through the MindTrace application can be

effectively processed by the deployed inference service. The classification results for TASK_04 and TASK_05 are returned in real time and presented to users in the form of diagnostic information, task-level predictions, and an overall classification outcome. Figure 14 shows the prediction result interface displayed on the MindTrace Android application after successful inference.

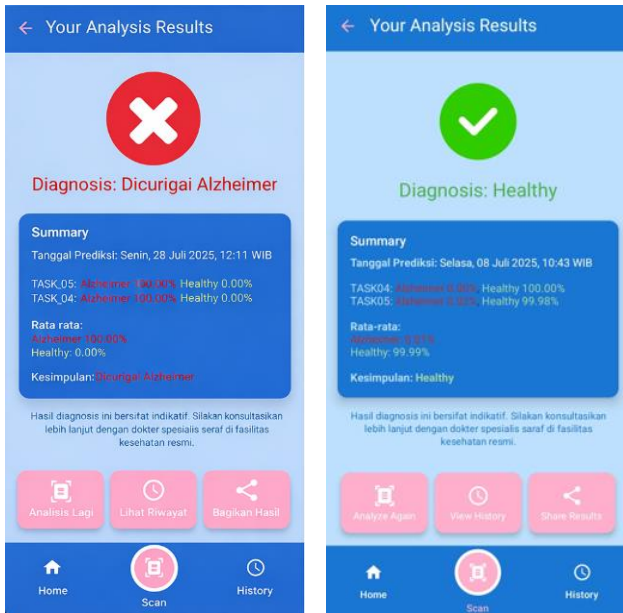


Fig. 14. Prediction result interface of the MindTrace Android application showing the classification outcome for Alzheimer’s disease based on handwriting analysis.

In addition to real-time prediction display, the system supports persistent data storage and report generation. Prediction outcomes, along with associated user information and handwriting images, are stored in the Supabase database to enable result tracking and access to prediction history. Furthermore, a structured prediction report in PDF format is automatically generated, summarizing the prediction results and visualizing the input handwriting samples, as illustrated in Figure 15.

These results confirm that the proposed deployment strategy operates reliably in a real-world application setting and is consistent with the remote inference approach evaluated in the experimental scenarios.

Regarding data privacy and ethical considerations, handwriting images submitted through the mobile application are transmitted securely via authenticated API communication and stored in a controlled database environment. The system is designed solely for early screening support and does not provide a definitive medical diagnosis. Users are informed that clinical confirmation from healthcare professionals remains essential. These measures aim to ensure responsible AI deployment in real-world health-related applications.

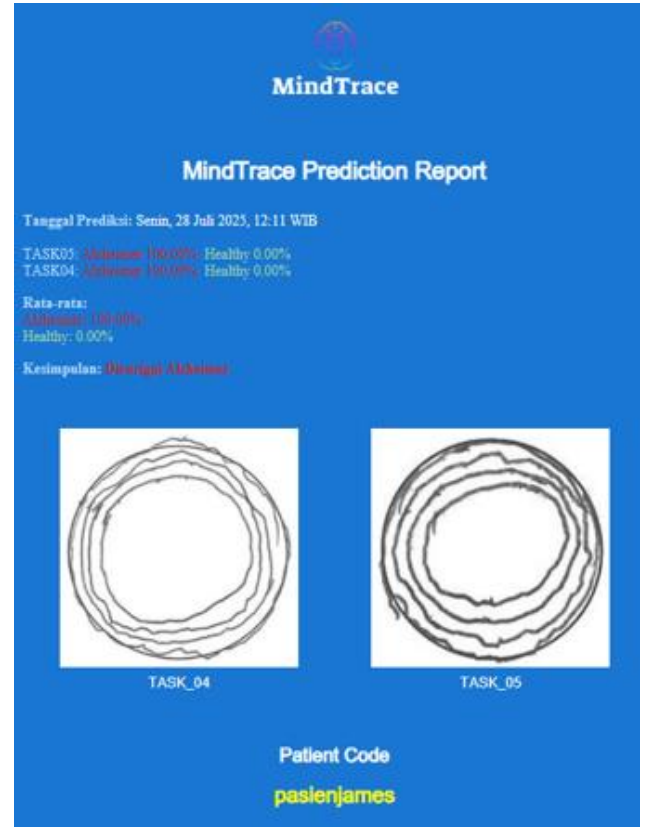


Fig. 15. Automatically generated prediction report in PDF format summarizing task-level results and handwriting image inputs.

G. System Usability Evaluation

The usability of the deployed MindTrace application is assessed using the USE Questionnaire, which measures four dimensions: usefulness, ease of use, ease of learning, and user satisfaction. The evaluation consists of 12 statements rated on a five-point Likert scale, with response options ranging from strongly disagree (1) to strongly agree (5). In total, 21 participants take part in the usability assessment.

Each response was converted into a numerical score, and an index percentage was calculated by comparing the obtained score with the maximum possible score for each item. The average index percentage for each usability aspect was then computed to obtain an overall usability assessment. The detailed questionnaire items are presented in Table 6, while the summary of the usability evaluation results is shown in Table 7.

TABLE VI
USE QUESTIONNAIRE ITEMS FOR MINDTRACE USABILITY EVALUATION

Aspect	Item Code	Statement
Usefulness	U1	The application facilitates early detection of Alzheimer’s disease through handwriting analysis.
	U2	The application provides early detection results that can be used as initial information for further medical consideration.

	U3	The application saves time in the early detection process of Alzheimer's disease.
Ease of Use	EOU1	I can use the application without requiring special training or assistance.
	EOU2	The application has a user-friendly and easy-to-use interface.
	EOU3	The error messages displayed by the application are easy to understand.
Ease of Learning	EOL1	I can quickly understand how the application works.
	EOL2	I can easily remember the steps to use the application.
	EOL3	I quickly become familiar with and can use the application without difficulty.
Satisfaction	S1	I am satisfied with the features and performance of the application.
	S2	The application's design (colors, icons, layout) supports ease of use.
	S3	I feel comfortable and satisfied while using the application.

The results show that all usability aspects achieved an average index above 80%, indicating a very positive user response. The highest score was observed in the ease of learning aspect (89.52%), suggesting that users were able to quickly understand and adapt to the application without difficulty. High scores in satisfaction (89.21%) and usefulness (88.25%) further indicate that users perceived MindTrace as beneficial and comfortable to use in supporting early Alzheimer's disease screening. Although the ease of use aspect obtained a slightly lower score (86.99%), it still falls within the "very good" category, demonstrating that the system interface and interaction flow are generally intuitive and user-friendly.

TABLE VII
SUMMARY OF USABILITY EVALUATION RESULTS

Aspect	Number of Items	Average Index
Usefulness	3	88.25%
Ease of Use	3	86.99%
Ease of Learning	3	89.52%
Satisfaction	3	89.21%

Overall, these findings confirm that the deployed MindTrace system not only performs well in terms of classification accuracy but also provides a positive user experience, supporting its feasibility for real-world usage as a mobile-based early screening tool.

IV. CONCLUSION

This study presented the development and evaluation of a VGG-19-based deep learning approach for handwriting-based early Alzheimer's disease detection. The proposed approach utilizes a deep learning model based on the VGG-19 architecture to provide a non-invasive, efficient, and accessible alternative for preliminary cognitive assessment. The results show that the developed system achieves

consistent classification performance across both handwriting tasks, namely TASK_04 and TASK_05.

Among the evaluated scenarios, the fourth scenario, which combines cleansed data with additional visual preprocessing techniques, produced the most consistent and robust results. This configuration achieved accuracies of 81% for TASK_04 and 90% for TASK_05 when deployed within the mobile application. The deployed configuration achieved a sensitivity of 88% and specificity of 92% for TASK_05, indicating strong capability in identifying both affected and non-affected individuals in early screening contexts. Furthermore, the application of preprocessing techniques such as Gaussian Blur, adaptive normalization, CLAHE, Otsu thresholding, and standard VGG-19 preprocessing proved effective in enhancing the quality of handwriting features, especially under real-world imaging conditions involving noise, uneven illumination, or low contrast.

The selected model was successfully deployed on a cloud-based server and integrated into an Android mobile application, enabling real-time handwriting-based prediction. Functional implementation confirms that the system supports essential features such as user authentication, image submission, prediction processing via an application programming interface, and result presentation in the form of downloadable reports. In addition, the usability evaluation indicates positive user responses across all assessed aspects, suggesting that the application is easy to learn, useful, and satisfactory for end users.

Overall, the results suggest that the developed system can support early-stage Alzheimer's disease screening in practical mobile-based applications. Nevertheless, several limitations should be acknowledged. First, the study relies on a single publicly available dataset with a relatively limited sample size, which may restrict the diversity of handwriting characteristics represented in the training data. Second, only two graphical tasks were evaluated, which may not fully capture the range of cognitive and motor variations associated with Alzheimer's disease. Third, variations in real-world image acquisition conditions, such as lighting, camera resolution, and background noise, may influence deployment performance.

Future research should therefore incorporate larger and more diverse multi-center datasets, evaluate additional handwriting tasks, and investigate cross-dataset validation to further assess generalization capability. In addition, the integration of explainable artificial intelligence techniques, such as Grad-CAM-based visual attention mapping, may provide greater interpretability by highlighting salient handwriting regions that influence classification decisions. Enhancing model transparency would strengthen clinical trust and support broader adoption in medical screening contexts.

REFERENCES

- [1] G. Lepa, "Kajian pustaka efektivitas dan efek samping memantine pada pasien Alzheimer," Universitas Katolik Widya Mandala

- Surabaya, Surabaya, 2023.
- [2] Alzheimer's Association, "Alzheimer's Association 2024 Alzheimer's disease facts and figures," 2024.
 - [3] WHO, "Global status report on the public health response to dementia," 2021.
 - [4] Alzheimer's Indonesia, "Statistik tentang Demensia," Apr. 22, 2019. <https://alzi.or.id/statistik-tentang-demensia/>.
 - [5] A. Arfina, "Pengaruh edukasi terhadap pengetahuan masyarakat tentang deteksi dini Alzheimer di Kelurahan Labuh Baru Pekanbaru," *Heal. Care J. Kesehat.*, vol. 10, no. 01, pp. 257–258, 2021.
 - [6] R. Saha, A. Mukherjee, A. Sadhukhan, A. Roy, M. De, and Anirban Mukherjee, *Handwriting analysis for early detection of Alzheimer's disease*, 1st ed. 2020.
 - [7] R. Singh and N. M. Sadiq, *Cholinesterase inhibitors*. StatPearls Publishing, 2024.
 - [8] D. A. Arafa, H. E. D. Moustafa, A. M. T. Ali-Eldin, and H. A. Ali, "Early detection of Alzheimer's disease based on the state-of-the-art deep learning approach: A comprehensive survey," *Multimed. Tools Appl.*, vol. 81, no. 17, pp. 23736–23737, 2022, doi: 10.1007/s11042-022-11925-0.
 - [9] H. Qi *et al.*, "A study of auxiliary screening for Alzheimer's disease based on handwriting characteristics," *Frontiers (Boulder)*, pp. 1–11, 2023.
 - [10] A. Juganavar, A. Joshi, and T. Sheekar, "Navigating early Alzheimer's diagnosis: A comprehensive review of diagnostic innovations," *Cureus*, vol. 15, no. 9, pp. 1–8, 2023, doi: 10.7759/cureus.44937.
 - [11] A. Rosyida and T. B. Sasongko, "Deteksi dini penyakit Alzheimer dengan algoritma C4.5 berbasis BPSO (binary particle swarm optimization)," *J. Sist. Inf. dan Komput.*, pp. 341–342, 2023.
 - [12] M. L. Vicchietti, F. M. Ramos, L. E. Betting, and A. S. L. O. Campanharo, "Computational methods of EEG signals analysis for Alzheimer's disease classification," *Sci. Rep.*, vol. 13, no. 1, pp. 1–2, 2023, doi: 10.1038/s41598-023-32664-8.
 - [13] A. Samad and E. Samet Aydu, "Rapid Alzheimer's disease diagnosis using advanced artificial intelligence algorithms," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 6, p. 1760, 2024, doi: 10.38124/ijisrt/ijisrt24jun1915.
 - [14] I. Aberathne, D. Kulasiri, and S. Samarasinghe, "Detection of Alzheimer's disease onset using MRI and PET neuroimaging: Longitudinal data analysis and machine learning," *Neural Regen. Res.*, vol. 18, no. 10, pp. 2134–2149, 2023, doi: 10.4103/1673-5374.367840.
 - [15] A. Bensalah, A. Parziale, G. De Gregorio, A. Marcelli, A. Fornes, and J. Lladós, "I can't believe it's not better: In-air movement for Alzheimer handwriting synthetic generation," in *IGS*, 2023, pp. 1–2, doi: 10.1007/978-3-031-45461-5_10.
 - [16] U. Mitra and S. U. Rehman, "ML-powered handwriting analysis for early detection of Alzheimer's disease," *IEEE Access*, vol. 11, p. 6, 2023, doi: 10.1109/ACCESS.2024.3401104.
 - [17] A. G. Vrahatis, K. Skolariki, M. G. Krokidis, K. Lazaros, T. P. Exarchos, and P. Vlamos, "Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: The role of artificial intelligence and deep learning," *Sensors*, vol. 23, no. 9, pp. 1–3, 2023, doi: 10.3390/s23094184.
 - [18] N. D. Cilia, C. de Stefano, C. Marrocco, F. Fontanella, M. Molinara, and A. S. di Freca, "Deep transfer learning for Alzheimer's disease detection," in *International Conference on Pattern Recognition*, 2020, pp. 9904–9911, doi: 10.1109/ICPR48806.2021.9412603.
 - [19] P. Erdogmus and A. T. Kabakus, "The promise of convolutional neural networks for the early diagnosis of Alzheimer's disease," *Eng. Appl. Artif. Intell.*, vol. 123, pp. 1–13, 2023, doi: 10.1016/j.engappai.2023.106254.
 - [20] J. Sweidan, M. A. El-Yacoubi, and A.-S. Rigaud, "Explainability of CNN-based Alzheimer's disease detection from online handwriting," *ResearchSquare*, pp. 1–15, 2024, [Online]. Available: <https://doi.org/10.21203/rs.3.rs-4215056/v1>.
 - [21] S. A. Maulana, S. H. Batubara, T. A. Amelia, and Y. P. P. Pasaribu, "Penerapan metode CNN (convolutional neural network) dalam mengklasifikasi jenis ubur-ubur," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, p. 122, 2023, doi: 10.55606/juprit.v2i4.3084.
 - [22] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep learning approach for early detection of Alzheimer's disease," *Cognit. Comput.*, vol. 14, no. 5, pp. 1711–1712, 2022, doi: 10.1007/s12559-021-09946-2.
 - [23] K. Azmi, S. Defit, and S. Sumijan, "Implementasi convolutional neural network (CNN) untuk klasifikasi batik Tanah Liat Sumatera Barat," *J. Unitek*, vol. 16, no. 1, p. 29, 2023, doi: 10.52072/unitek.v16i1.504.
 - [24] K. R. Wardani and L. Leonardi, "Klasifikasi penyakit pada daun menggunakan metode convolutional neural network," *J. Tekno Insentif*, vol. 17, no. 2, pp. 112–114, 2023, doi: 10.36787/jti.v17i2.1130.
 - [25] C. Yulian, "Penerapan deep learning style transfer untuk menciptakan kombinasi gambar digital dengan batik," Universitas Multimedia Nusantara, 2021.
 - [26] A. Lutfiyah, "Implementasi visual geometry group (VGG19) untuk mengenali pola citra dalam mendeteksi penyakit mata katarak," 2023.
 - [27] A. Z. Noorizki and W. I. Kusumawati, "Perbandingan performa algoritma VGG16 dan VGG19 melalui metode CNN untuk klasifikasi varietas beras," *J. Comput. Electron. Telecommun.*, vol. 4, no. 2, p. 7, 2023, doi: 10.52435/complete.v4i2.387.
 - [28] L. A. Susanto, A. Nilogiri, and L. Handayani, "Klasifikasi citra lesi kulit serupa virus Monkeypox menggunakan VGG-19 convolutional neural network," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 8, no. 1, p. 2, 2023, doi: 10.32528/justindo.v8i1.168.
 - [29] M. M. Siregar, R. Hizria, and D. Pardede, "Perbandingan kinerja kernel SVM dalam klasifikasi kategori kanker kulit menggunakan transfer learning," vol. 4, no. 1, p. 84, 2024.
 - [30] Keras.io, "VGG16 and VGG19," 2024. https://keras.io/2/api/applications/vgg/?utm_source=chatgpt.com.
 - [31] OpenVino, "Use Case and High-Level Description," 2023. https://docs.openvino.ai/2023.3/omz_models_model_vgg19.html.
 - [32] T. G. Devi, N. Patil, S. Rai, and C. S. Philipose, "Gaussian blurring technique for detecting and classifying acute lymphoblastic leukemia cancer cells from microscopic biopsy images," *Life*, vol. 13, no. 2, p. 4, 2023, doi: 10.3390/life13020348.
 - [33] E. Fauziah, F. Bagus Sadewo, H. M. Putra, P. Rosyani, and F. I. Komputer, "Penerapan image processing menggunakan OpenCV dan Python untuk memperhalus gambar melalui smoothing image dengan metode Gaussian blur," *Indones. J. Netw. Secur.*, vol. 13, no. 2, p. 6, 2024.
 - [34] M. Cossio, "Augmenting medical imaging: A comprehensive catalogue of 65 techniques for enhanced data analysis," p. 7, 2023, [Online]. Available: <http://arxiv.org/abs/2303.01178>.
 - [35] I. M. Mohammed and A. N. A. M. Isa, "Contrast limited adaptive local histogram equalization method for poor contrast image enhancement," *IEEE Access*, vol. 13, p. 62606, 2025, doi: 10.1109/ACCESS.2025.3558506.
 - [36] Q. Cao, L. Qingge, and P. Yang, "Performance analysis of Otsu-based thresholding algorithms: A comparative study," *J. Sensors*, pp. 1–2, 2024, doi: 10.1155/2021/4896853.
 - [37] D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, and D. Li, "An efficient and lightweight convolutional neural network for remote sensing image scene classification," *MDPI*, p. 2, 2020.
 - [38] N. Asyrofiyyah and E. Sugiharti, "Hyperparameter optimization using Hyperband in convolutional neural network for image classification of Indonesian snacks," vol. 2, no. 1, p. 48, 2024, doi: 10.15294/rji.v2i1.72720.
 - [39] Wijiyanto, A. I. Pradana, Sopingi, and V. Atina, "Teknik K-fold cross validation untuk mengevaluasi kinerja mahasiswa," pp. 241–242, 2024, doi: 10.33364/algoritma/v.21-1.1618.
 - [40] A. M. Siregar, S. Faisal, Y. Cahyana, and B. Priyatna, "Perbandingan algoritme klasifikasi untuk prediksi cuaca," *Univ. Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 16–19, 2020.
 - [41] A. Ningtiyas, S. N. Faizah, M. Mustikasari, and I. Bastian, "Pengukuran usability sistem menggunakan USE questionnaire pada aplikasi OVO," *Ilm. KOMPUTASI*, vol. 20, pp. 102–103, 2021.