HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 4 (2024)

Enhancing Quality of Life for Dementia Patients through Intelligent Memory Aids in Social Communication

Yoonwoo Chyung[1], Yunhoo Kang[1] and Jae-Yeon Sim[#]

[1]St.Paul Preparatory Seoul, Republic of Korea
[#]Advisor

ABSTRACT

The prevalence of dementia in South Korea has shown a marked increase, with the number of affected individuals rising from approximately 630,000 in 2015 to nearly 1 million in 2023. This growth aligns with the aging population, as the proportion of elderly individuals has expanded, driving the prevalence rate from 9.5% in 2015 to 10.3% in 2023. This trend is not isolated to South Korea; in the United States, the number of dementia patients has also risen sharply, from around 4.7 million in 2010 to an estimated 13.8 million by 2050, representing a nearly threefold increase over 40 years. As time progresses, the number of dementia patients is expected to continue climbing, with a particularly significant rise among individuals aged 85 and older. Globally, the number of dementia patients is projected to reach 131.5 million by 2050, nearly doubling every 20 years, highlighting the significant impact of population aging as a major risk factor for dementia onset. Memory aids like notes and reminder calendars are commonly used by dementia patients to support daily activities, but they have limitations. While helpful for routine tasks, these simple tools cannot assist patients in remembering faces or recognizing familiar people. To address this issue, we introduce a machine learning-based memory aid system designed to support social communication for dementia patients. The proposed system processes face images to perform facial identification which provides relevant information about the identified person, such as their relationship to the dementia patient and any scheduled activities with them. To improve the system's accuracy, we proposed a metric-based loss function. Experimental results demonstrated that this approach enhanced accuracy, achieving a 2.96% improvement over the previous method.

Introduction

Dementia, including Alzheimer's disease, is characterized by a set of symptoms that impact cognitive functions such as memory, thinking, and reasoning, significantly interfering with a person’s daily life and activities. The number of these patients is increasing. According to the Centers for Disease Control and Prevention (CDC), the prevalence of dementia is rising due to an aging population. It is estimated that by 2050, nearly 14 million Americans will be living with dementia disease.

Generally, many patients rely on the use of memory aids due to the nature of the disease. Memory aids for dementia patients are tools and techniques designed to assist dementia patients in compensating for memory loss. These aids include a variety of items such as photographs, notebooks, and electronic devices, which facilitate the retention and retrieval of essential information for daily activities. For instance, photo albums or dairies can help patients remember important individuals, events, or locations that might otherwise be forgotten. Despite the considerable benefits these aids provide in enhancing memory recall, they exhibit significant limitations in improving social communication. Patients often continue to face challenges in engaging in meaningful dialogues, interpreting social cues, and sustaining interpersonal relationships, as memory aids primarily address information retention rather than complex dynamics of social interaction. Nevertheless, even with the utilization of memory aids, dementia patients

may still encounter difficulties in recognizing the faces of their family, friends, and acquaintances. This inability to identify loved ones can result in significant confusion and distress. Furthermore, the frustration and sadness associated with inability to remember important aspects of their lives can exacerbate feelings of depression, thereby further compromising their overall well-being.

To address the aforementioned issue, we propose a machine learning-based social communication aid system with face recognition techniques to enhance the quality of life for dementia patients. The proposed system is composed of three modules: an image-obtaining camera, a face recognition module, and an output information speaker. The image-obtaining camera is a wearable device attached to glasses which makes it capture images of what the patient sees. The face recognition module identifies the faces of people the patient communicates. Finally, the output information speaker provides the patient with relevant information about the recognized face through audio feedback.

Background Knowledge

Social Isolation Dementia

Dementia is a progressive neurological disorder that significantly impacts the daily lives of those afflicted. Patients experience a range of symptoms that deteriorate their cognitive functions which makes everyday tasks increasingly difficult. As the disease advances, individuals struggle with memory loss, confusion, and impaired judgment, which interfere with their ability to perform routine activities.

A particularly challenging symptom of Alzheimer's disease is the difficulty in recognizing and remembering people's faces. This condition is common among Alzheimer's patients and can severely impact their social interactions. Patients may struggle to engage in spontaneous conversation, missing nonverbal cues, or experiencing difficulty formulating responses. Social isolation exacerbates the progression of dementia disease by reducing mental stimulation and opportunities for social engagement. It also contributes to feelings of loneliness and depression further diminishing the quality of life for patients. Therefore, it is essential to consider incorporating communication-focused strategies alongside memory aids to ensure a more holistic approach to supporting dementia patients.

Face Recognition

Face recognition can be divided into two main components: face verification and face identification. Face verification is the process of confirming whether a given face matches a specific known face. The algorithm compares a presented face with a stored memory of another face to verify their similarity.

Face identification, on the other hand, recognizes a face as belonging to a particular individual. This process requires the ability to retrieve specific information about a person from memory such as their name and relationship to the observer.

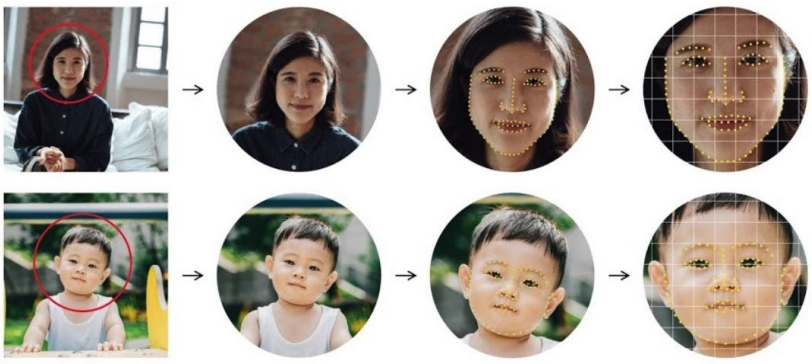

Figure 1. Example of face area crop and alignment (Wiki Docs 2021)

Accurate face recognition relies heavily on effective preprocessing techniques. The initial step involves face detection, which locates the presence and position of faces within an image. This ensures the system focuses on relevant regions and avoids processing irrelevant background information. Following successful detection, preprocessing continues with cropping and rotation. Cropping removes unnecessary background clutter and focuses solely on the facial region of interest. Rotation, often referred to as canonical form transformation in this context, corrects for head pose variation. This transformation aims to align all faces into a standard orientation, typically an upright frontal view.

In the proposed approach, we utilize face identification to recognize the friends and acquaintances of individuals with dementia. The system outputs social log information related to the recognized persons to assist the individual's memory. A detailed explanation of the proposed system is elaborated in Chapter 3.

Proposed Facial Identification System

This chapter provides a detailed explanation of the proposed system which consists of three modules. Figure 2 illustrates the technical flowchart of the system. First, a Facial Verification Convolutional Neural Network (FV-CNN) processes images to detect facial areas. The detected face is then cropped and rotated into a canonical form for further processing by the Facial Identification Convolutional Neural Network (FI-CNN). The FI-CNN performs the identification process on these face images. Once the individual is identified, the system informs the dementia patient via an audio output. The text-to-speech (TTS) module then provides relevant information, such as the person's relationship to the patient or any scheduled activities with them.

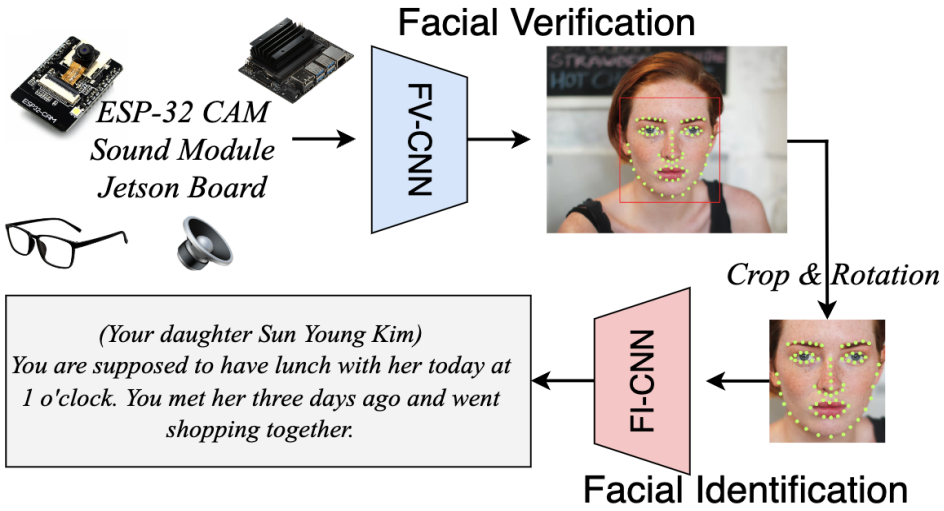

Figure 2. Overall architecture of the proposed intelligent memory aids system

Training Strategy Using Triplet Loss

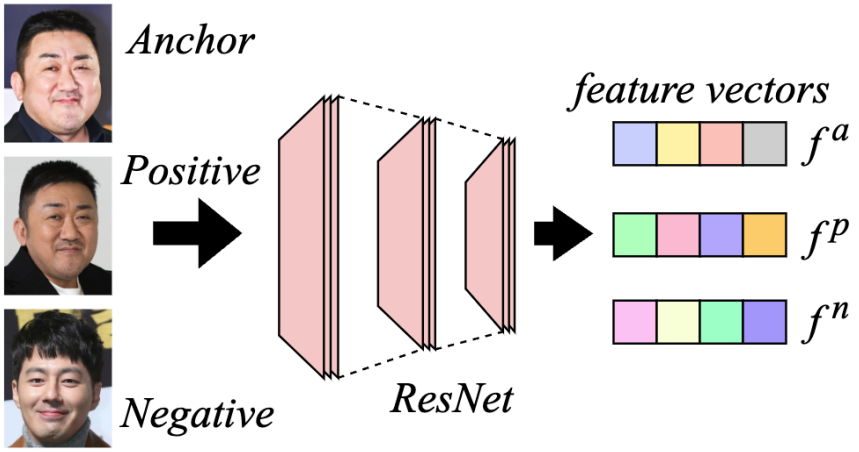

Figure 3. Architecture of the proposed training approach using triplet loss function

In facial identification systems, particularly those utilizing deep learning, three key image types play a crucial role: anchor, positive and negative images. These terms are tightly linked to the loss function that helps the system learn how to distinguish between different faces. An anchor image serves as a reference point, randomly selected from a dataset of facial images. A positive image corresponds to the same individual depicted in the anchor image, signifying a positive match. Conversely, a negative image represents a different person, establishing a negative match in relation to the anchor image.

Equation 1. Distance between anchor and positive features

$$dist^{a,p} = \sqrt{(f_1^a - f_1^p)^2 + (f_2^a - f_2^p)^2 + \cdots + (f_L^a - f_L^p)^2}$$
$$dist^{a,n} = \sqrt{(f_1^a - f_1^n)^2 + (f_2^a - f_2^n)^2 + \cdots + (f_L^a - f_L^n)^2}$$

Equation 1 quantifies the distance between two feature vectors, where f_k^a , f_k^p and f_k^n represent the $\boldsymbol{k}$-th elements of feature vectors f^a, f^p, and f^n, respectively. The variable $\boldsymbol{L}$ denotes the dimension of the feature vectors, which we set to 1024 in this research. $dist^{a,t}$ represents the calculated distance between the anchor feature vector and the target feature vector, which can be either positive or negative. These distances are then used to compute the loss value as described in Equation 2.

Equation 2. Triplet loss function

$$L_T = \sum_i \max\left(0, dist_i^{a,p} - dist_i^{a,n} + T\right)$$

The triplet loss function pushes away the negative (different picture) far away from the anchor meaning the negative image is not the right one. This is a very useful function for the trained model to understand what image to output and distinguish each of them. During the training process, the model learns to extract consistent features for samples from the same individual. The effectiveness of this approach will be demonstrated through extensive experiments in Chapter 4.

Transfer Learning (Facial Identification)

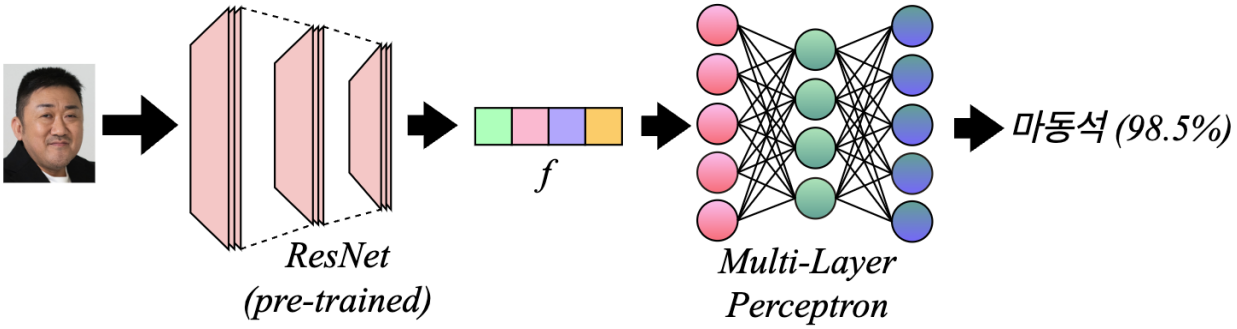

Figure 4. Architecture of the proposed transfer learning

Figure 4 illustrates the architecture of the proposed transfer learning approach for facial identification. The facial identification network takes facial images as input and produces an individual's identity. The proposed network consists of two main components: a convolutional neural network (CNN) which is pre-trained as detailed in Chapter 3.1, extracts related features from the inputted individual facial images, transforming them into a compact feature vector. This feature vector is then fed into the multi-layer perceptron (MLP). The MLP's task is to further analyze the feature vector and make a final prediction about the individual's identity. To train the proposed facial identification network, we utilized the cross entropy loss function which frequently used to train many classification models.

Equation 3: Cross entropy loss function

$$L_c = -\ln(prob)$$

The loss function measures the difference between the predicted probability distribution and the actual distribution of class labels. The process involves calculating the negative logarithm of the predicted probability for the correct class for each data point. These values are then averaged across all data points to obtain the final cross entropy loss. By minimizing the loss function during training, the model is optimized to output higher probability values for the correct class, resulting in improved classification accuracy.

System Implementation

To develop the system, we used a Jetson Orin Nano (Scalcon et al. 2024) embedded board as the computing platform and an ESP-32 CAM (Venu 2022) as the input camera. The camera module communicates with the board wirelessly via Wi-Fi. For the TTS module's audio output, any type of speaker can be used. In our setup, we utilized standard audio earphones connected to the board via a sound jack.

Experimental Results

Facial Identification Dataset

To train and evaluate the proposed system, we used the VGG Face Dataset (Cao et al. 2018), which includes 2,622 individuals and 26 million samples. We trained the model by randomly selecting 10 individuals from the dataset and testing performance across various convolutional neural network architectures. This process was repeated 10 times, and we calculated the average performance and standard deviation to ensure statistical significance.

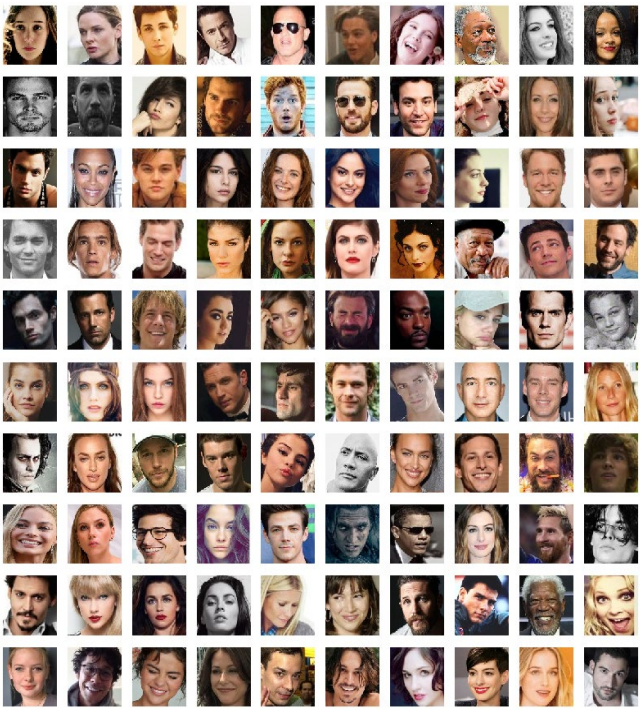

Figure 5. Snapshot of the dataset used in this paper (Cao et al. 2018)

Evaluation Comparison

For evaluation, we selected four convolutional neural network architectures known for their comparable performance in various computer vision tasks: AlexNet (Krizhevsky et al. 2012), VGG (Simonyan et al. 2014), MobileNet-V2 (Sandler et al. 2018), and ResNet (He et al. 2016) . These architectures were chosen for their different layer depths, allowing us to examine how performance varies with network complexity.

The evaluation metrics used include accuracy, recall, precision, and F1-score, which are commonly employed to assess the performance of classification models.

Table 1. Evaluation comparison with state-of-the-art convolutional neural networks

	Accuracy	**Recall**	**Precision**	**F1-Score**
AlexNet (Krizhevsky et al. 2012)	0.7761	0.7772	0.7778	0.7775
VGG-19 (Simonyan et al. 2014)	0.7998	0.7989	0.7985	0.7987
MobileNet-V2 (Sandler et al. 2018)	0.8035	0.8017	0.8014	0.8015
ResNet-50 (He et al. 2016)	0.8191	0.8191	0.8193	0.8192

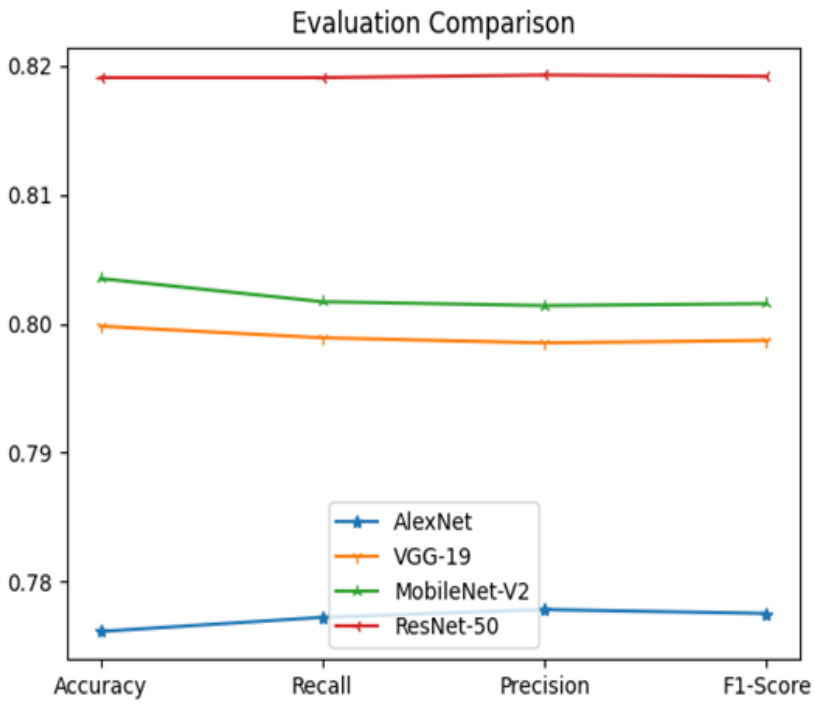

Figure 6. Evaluation comparison (line graph)

AlexNet has the lowest accuracy among the 3 other convolutional neural network architectures by the accuracy of 0.7761, recall 0.7772, precision 0.7778, and f1-score of 0.7775. The ResNet-50 has the highest scores among the 3 other CNN architectures. Additionally, the accuracy of the convolutional neural network architectures are ordered from the highest accuracy to the lowest accuracy depending on the amount of depth.

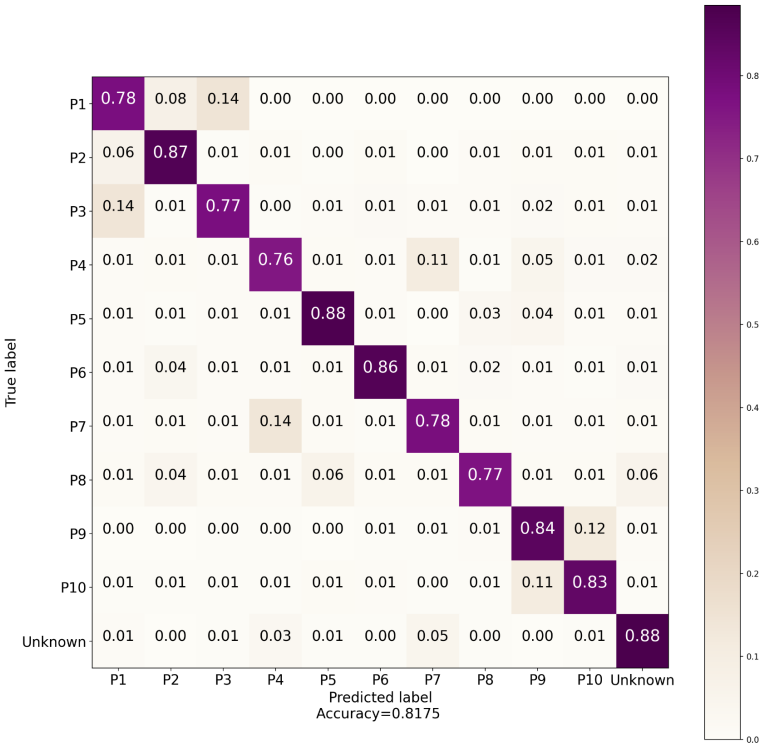

Figure 7. Evaluation comparison (confusion matrix)

Additionally, we generate the confusion matrix to examine the robustness of the proposed approach. The confusion matrix depicted in Figure 7 provides an insightful evaluation of ResNet-50's ability to accurately identify specific individuals within the face datasets based on their facial snapshots. The color gradient, ranging from white to dark blue, visually represents the model's accuracy across different classifications. On average, ResNet-50 achieved an accuracy of 0.8175, with is performance ranging from a maximum of approximately 0.88 to a minimum of 0.77.

Several factors contributed to instances of reduced accuracy. One notable limitation was the dataset's demographic imbalance; in an experiment where only two individuals with darker skin tones were included, the recognition rate declined, likely due to the model's bias toward skin color. Furthermore, the model's performance was adversely affected when identifying individuals with similar facial features, as the resemblance among these faces led to a decrease in recognition accuracy.

Ablation Study

Table 2. Ablation study result (triplet loss evaluation)

	Accuracy (proposed method)	**Accuracy (previous method)**
AlexNet	0.7761	0.7564
VGG-19	0.7998	0.7654
MobileNet-V2	0.8035	0.7721
ResNet-50	0.8191	0.7863

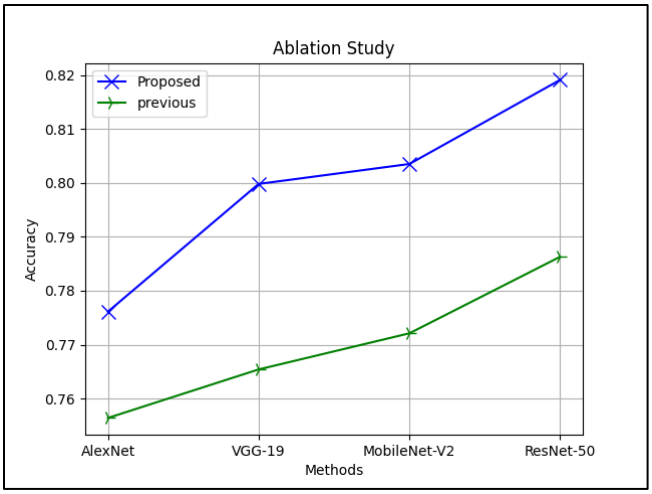

Figure 8. Ablation study result (triplet loss evaluation)

To assess the efficacy of the proposed methodology, an ablation study was conducted by comparing the performance of four distinct CNN architectures–AlexNet, VGG-19, MobileNet-V2, and ResNet-50–both with and without the integration of the proposed method. The results of this comparative analysis are visualized in Figure 8 and tabulated in Table 2.

Significantly, the implementation of the proposed approach led to a consistent improvement in accuracy across all architectural models, with gains of 2.10% for AlexNet, 3.45% for VGG-19, 3.23% for MobileNet-V2, and 3.28% for ResNet-50. These findings demonstrate that the proposed method substantially enhances model accuracy, thereby validating its effectiveness in improving performance across diverse CNN architectures.

Conclusion

In this research, we proposed a machine learning-based face identification system to enhance the quality of life for dementia patients. The system is developed with intelligent memory aids designed to improve social communication. The proposed system was evaluated with four different convolutional neural network architectures, with ResNet-50 achieving the highest accuracy at 81%. Additionally, the implementation of the triplet loss function further enhanced performance, increasing the average accuracy by 2.96%. Looking forward, we plan to integrate the proposed system into wearable technology, such as smart glasses equipped with a camera, to assist dementia patients in recognizing and identifying individuals in real-time. This work demonstrates the potential of machine learning-driven solutions to significantly support the cognitive and social needs of individuals living with dementia.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018, May). Vggface2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 67-74). IEEE.

DFROBOT. (2024, Jul 22). “ESP32-CAM Development Board”: DFROBOT. https://www.dfrobot.com/product-1879.html

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). https://doi.org/10.48550/arXiv.1801.04381

Scalcon, F. P., Tahal, R., Ahrabi, M., Huangfu, Y., Ahmed, R., Nahid-Mobarakeh, B., ... & Emadi, A. (2024, June). AI-Powered Video Monitoring: Assessing the NVIDIA Jetson Orin Devices for Edge Computing Applications. In 2024 IEEE Transportation Electrification Conference and Expo (ITEC) (pp. 1-6). IEEE.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Venu, D. N. (2022). IOT Surveillance Robot Using ESP-32 Wi-Fi CAM & Arduino. IJFANS International Journal of Food and Nutritional Sciences, 11(5), 198-205.

Wiki Docs. (2021, Jun 1). “*Face recognition*”: Wiki Docs. https://wikidocs.net/151311