International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Investigating Digital Illiterate Classification Techniques Based on DeepFace Technology

Dae-Kyeon Shin [a], Tae-Ho Kim [b], Dong-Jin Seo [b], Howard Kim [a], Jin-Gil Cha [a], Seong-Kweon Kim [a,*]

[a] *Department of Information Technology and Media Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Seoul, 01811, Republic of Korea*

[b] *Department of Smart ICT Convergence Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Seoul, 01811, Republic of Korea*

Corresponding author: [*]kim12632@ seoultech.ac.kr

***Abstract*— This paper presents an algorithm to identify digital illiterates by analyzing age and emotions from facial recognition. In this paper, digital illiterates refer to people who struggle in using digital devices. The study assumed that older individuals who display surprised or angry facial expressions while using digital devices are more likely to be identified as digitally illiterate. For age detection, the study used MTCNN (Multi-task Cascaded Convolutional Networks), and for emotion detection, it employed the VGG-Face model. MTCNN detects facial features and landmarks to preprocess images and distinguish facial characteristics. The VGG-Face model uses convolution operations to analyze facial images and classify emotional states. The dataset used in the study consisted of 3,000 facial images collected from the internet. The research team categorized the images into faces of individuals aged over 50, angry expressions, and surprised expressions. The dataset included 411 individuals (13.7%) aged over 50, 163 individuals (5.4%) with angry expressions, and 145 individuals (4.8%) with surprised expressions. Accuracy was estimated by comparing the results from the DeepFace algorithm with those from the research team's classifications. The DeepFace algorithm achieved 95.77% accuracy in detecting individuals aged over 50, 83.45% accuracy for surprise, and 76.07% for anger. The results demonstrate that it is possible to identify digital illiterates based on their age and emotional expressions and could enable the development of personalized services to directly or indirectly support digital illiterates, potentially improving digital accessibility.**

***Keywords*— Deep learning; DeepFace; digital illiterate; face recognition; emotion recognition.**

Manuscript received 15 Mar. 2024; revised 10 Jun. 2024; accepted 12 Sep. 2024. Date of publication 31 Oct. 2024.

I. INTRODUCTION

In the current digital epoch, rapid digitization permeates various sectors, predominantly driven by the extensive use of computer technology. This progression has facilitated the ubiquitous dissemination of information via remote, digital means. Nonetheless, this paradigm shift presents substantial hurdles for individuals with limited digital proficiency, who frequently need help utilizing these technological modalities effectively.

This investigation seeks to ascertain the utility of facial and emotional recognition technologies in identifying individuals facing challenges with digital literacy. The hypothesis conjectures that the exhibition of surprise or anger, particularly in older demographics, may signify a need for digital acumen. Under this premise, the study aims to delineate such expressions as indicative markers for identifying users with limited digital literacy. Moreover, the research contemplates the design of bespoke service offerings, including potential direct interactions, to cater to the specific needs of this demographic, thereby augmenting their engagement with digital platforms. By concentrating on users' palpable emotional and facial cues, this research aspires to substantially contribute to enhancing inclusive and accessible digital services. Such advancements aim to alleviate the obstacles encountered by individuals with limited digital literacy, promoting a more inclusive digital environment in an era characterized by pervasive digitization.

II. MATERIALS AND METHOD

Recent advancements in deep learning paradigms have significantly transformed the landscape of facial detection and recognition technologies. The shift towards deep learning-based methodologies has markedly outperformed traditional algorithmic approaches, primarily attributed to the proliferation of royalty-free facial databases and the evolution of deep learning architectures adept at extracting insights from these voluminous datasets [1].

This technological advancement has not been confined to facial recognition alone but has extended its tentacles into ancillary domains such as age, gender, and emotion recognition. Convolutional neural networks (CNNs) have emerged as a robust framework, exhibiting exceptional proficiency in deciphering complex facial recognition tasks [2].

Fig. 1 delineates the established methodology for facial recognition employing CNNs. The initial phase involves preprocessing, which is imperative for isolating and extracting the frontal facial features from the given image or video input. This stage encompasses facial detection, which aims to pinpoint the facial region within the entire input, and facial alignment, which meticulously outlines critical facial features, including the eyes, nose, and mouth. Following this, face embedding is undertaken, where feature vectors are derived. These vectors compare against a pre-trained database, facilitating the verification and eventual execution of facial recognition.

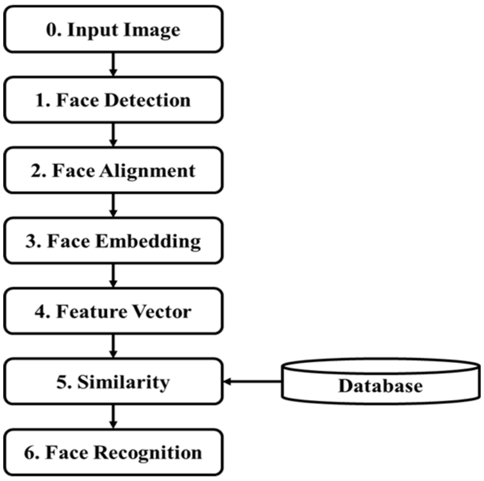

Fig. 1 Facial recognition process utilizing CNNs

The pivotal elements in this procedure are the accurate detection of the facial region and the subsequent recognition of the face from the identified area. This research, which focuses on identifying individuals with limited digital literacy, integrates the Multi-task Convolutional Neural Network (MTCNN) technique during the preprocessing phase.

The MTCNN framework employs a cascading sequence of P-Net, R-Net, and O-Net, each progressively refining the detection and alignment of the facial features. This sequence culminates in producing the final facial candidate and bounding box regression and landmark localization values. These elements undergo training through a multi-tiered process, which is intrinsic to the MTCNN's methodology, thus enhancing the precision and reliability of the facial recognition process. Fig. 2 presents a detailed schematic of the MTCNN deployment.

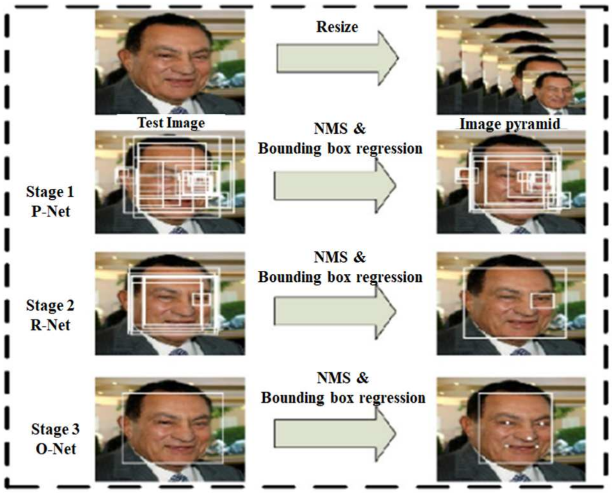

Fig. 2 Operational schema of MTCNN [3]

The MTCNN's initial phase involves the dynamic resizing of the original image across multiple scales, a critical step to ensure the comprehensive detection of faces of varying dimensions. Following this preparatory resizing, the image is subjected to the p-net, the first of the MTCNN's three sequential networks. The p-net serves a dual purpose: firstly, to ascertain the presence of facial regions within the image through a classification process, and secondly, to propose preliminary bounding boxes around these potential facial regions [3], [4], [5].

Subsequently, the process transitions to the r-net, the second stage in the MTCNN architecture. The r-net is tasked with refining the outcomes of the p-net. It achieves this by identifying the true facial regions within the proposed bounding boxes more accurately and fine-tuning them for enhanced precision. The final stage is the o-net, which builds upon the refined detections from the r-net. The o-net's primary function is the detailed demarcation of facial landmarks on the detected faces, such as the eyes, nose, and mouth. This stage is crucial for detailed facial feature recognition, providing the granular data necessary for sophisticated facial analysis tasks. Collectively, these three networks operate in a cascading manner, each contributing to a progressively more accurate and detailed facial detection and recognition capability, as illustrated in Fig. 2 [6], [7], [8], [9], [10], [11].

Fig. 3 delineates the intricate architecture and operational sequence of the MTCNN constituent algorithms: p-net, r-net, and o-net. The initial phase of the MTCNN's operation commences with the p-net, which is entrusted with the task of processing the resized images inputted at $12 \times 12 \times 3$ dimensions. This network undertakes a convolutional operation to categorize facial regions, outputting not only the four bounding box regression values, which encapsulate the dimensions and location of the detected facial region, but also ten landmark localization values that meticulously plot the coordinates of significant facial landmarks such as the eyes, nose, and mouth [3].

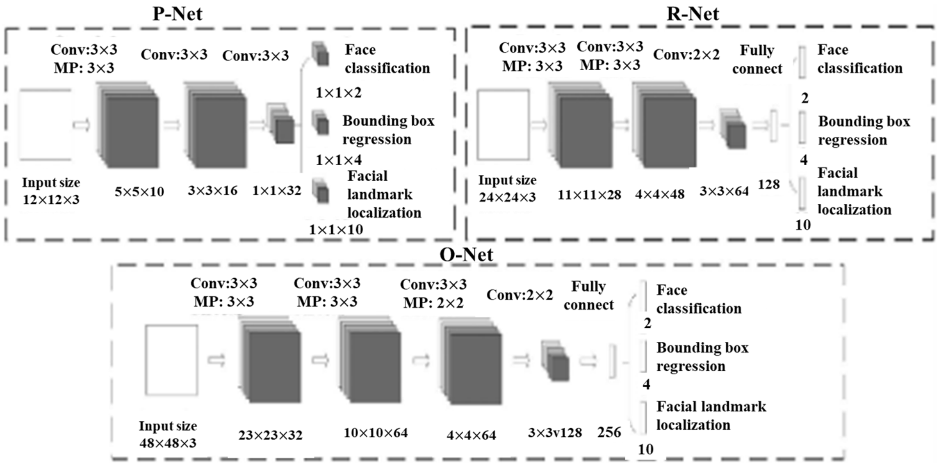

Fig. 3 Architectural configuration of P-Net, R-Net, and O-Net within MTCNN [3]

Following the p-net stage, these detected facial regions are recalibrated to their original dimensions, and procedures like on-maximum Suppression and bounding box regression are employed to enhance the accuracy and reduce overlaps in the detected areas. The processed regions are then relayed to the r-net, the second phase of the MTCNN, specifically designed to refine the detection by sieving out the genuine facial areas from the preliminary detections and applying a more nuanced bounding box regression for heightened precision.

The o-net, the third and final phase, concludes the MTCNN's sequence. The net's primary focus is the meticulous localization and application of facial landmarks on the candidate facial areas ascertained by the r-net. This stage is pivotal for the detailed recognition of facial features, contributing significantly to the MTCNN's comprehensive facial analysis capability [12], [13], [14], [15], [16].

Fig. 4 delineates a systematic flowchart elucidating the sequential steps involved in emotion recognition. The initial stage commences with image data input through a meticulously curated dataset. This input is subjected to a series of image processing techniques, where images are resized and refined utilizing the advanced capabilities of the MTCNN algorithm. This critical phase is instrumental in differentiating between facial and non-facial regions, accurately identifying the face, and detecting quintessential facial features and landmarks, such as the eyes, nose, and mouth. Upon completion of these preprocessing steps, the refined image is channeled into the subsequent model.

The core model, integral to this study, is the VGG-face model, which undertakes the task of emotion recognition. This model reinstates convolutional processes to analyze the patterns inherent in the facial features, classifying them pertinent for emotion analysis. Following the classification, a detailed examination of the features is conducted to extract and interpret attributes that are emblematic of specific emotions, including but not limited to anger, disgust, fear, sadness, happiness, and surprise. The culmination of this intricate process is annotating the detected emotion onto the image, thereby rendering a definitive representation of the emotional state [17].

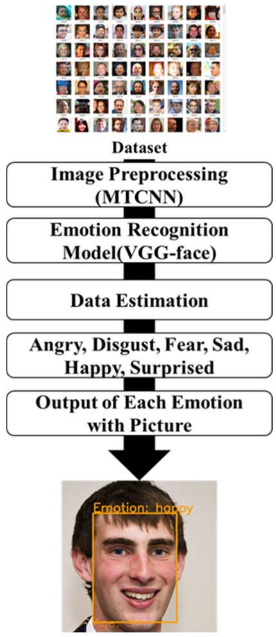

Fig. 4 Schematic representation of emotion recognition process

Fig. 5 delineates the block diagram of the VGG Face Model, a profound CNN architecture designed for intricate facial recognition tasks. The model's architecture is methodically structured, commencing with an input layer followed by a series of convolutional layers, max pool layers, and fully connected layers, culminating in a softmax layer for classification [18].

Commencing with the conv3-64 layer, the model employs 64 filters of 3×3 dimension to perform the initial convolution operation, extracting fundamental features from the input image. A subsequent layer of the same specification delves deeper, enhancing the detail of low-level visual patterns. Maxpool layers intersperse these convolutional layers, distilling the feature representation by downsampling, thus retaining critical information while reducing the overall data footprint. Progressing deeper, conv3-128 layers augment the filter count, enabling the capture of more intricate features, and are further refined by additional conv3-256 and conv3-512 layers. Each increment in filter count corresponds to an increased ability to recognize more complex and abstract features within the image.

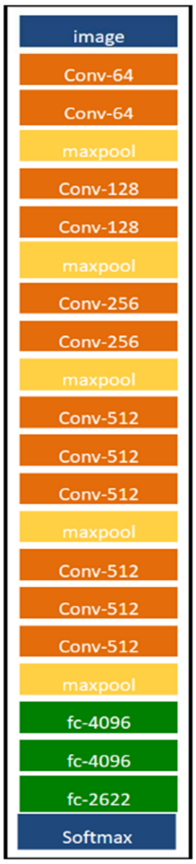

Fig. 5 Architectural overview of VGG face model [18]

The architecture's depth is significant, with multiple layers dedicated to progressively extracting and refining features. The final stages involve fully connected layers, specifically FC-4,096 layers that integrate all preceding feature information into a high-dimensional vector, setting the stage for the final classification. The FC-2,622 layer, an extension of this fully connected network, is particularly adept at distinguishing between nuanced variations across facial features. The concluding Softmax layer plays a pivotal role in probabilistically determining the most likely classification from the possible set of outcomes, effectively pinpointing the most accurate facial recognition and emotional inference [19], [20], [21], [22], [23].

III. Results and Discussion

A. Deepface Model

Fig. 6 provides a schematic representation of the image alignment methodology conceptualized within the DeepFace framework. The DeepFace alignment protocol is a multi-stage process, each meticulously designed to enhance the precision of facial recognition. Initially, the protocol commences with facial detection, employing six critical reference points derived from the original image; specifically, two points from the eyes, one from the nasal apex, and three from the contour of the lips. This foundational step ensures the accurate localization of the face within the image.

Progressing to the second stage, the protocol extracts a 2D cropped image of the detected facial region from the original image. This 2D representation serves as the basis for further alignment refinements. The subsequent third stage introduces the application of Delaunay triangulation on the 2D-aligned image. This technique is pivotal for rectifying any off-plane rotations, thus enhancing the dimensional consistency of the facial image. Concurrently, this phase also involves the construction of a transformational model that facilitates the transition from 2D to 3D representation, effectively interlinking 67 distinct facial reference points.

The culmination of this multi-phased process is the final stage, where the 2D and 3D structural elements are cohesively aligned. This crucial stage focuses on minimizing alignment errors and standardizing the facial representation to a frontal orientation. This systematic alignment process not only improves the accuracy of subsequent facial recognition tasks but also standardizes the facial data for consistent processing and analysis [24], [25], [26].

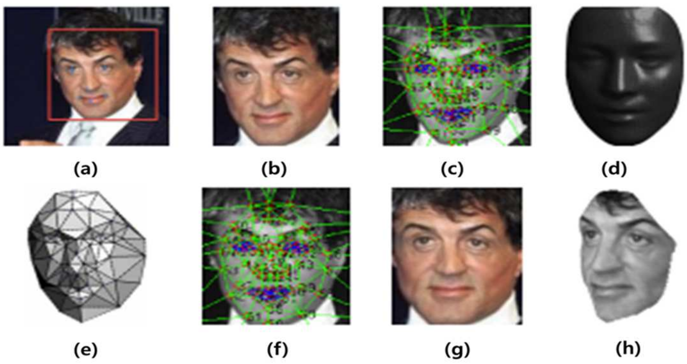

Fig. 6 DeepFace image alignment procedure [24]

Fig. 7 provides an analytical representation of the DeepFace architecture's structural composition. The architecture is initiated with a 152×152 dimensioned 3D aligned RGB image as the input. Commencing with the first layer, C1, this layer is comprised of a convolution operation utilizing 32 filters, each of 11×11 dimension, coupled with the Rectified Linear Unit (ReLU) activation function. This initial layer is pivotal for the primary feature extraction from the input image.

Proceeding to the second layer, M2, it incorporates a max pooling operation over a 3×3 pixel area with a stride of 2, effectively reducing the spatial dimensions while preserving the most salient features. The subsequent third layer, C3, continues the convolutional process with 16 filters of 9×9 dimension, also utilizing the ReLU activation function for non-linear transformation.

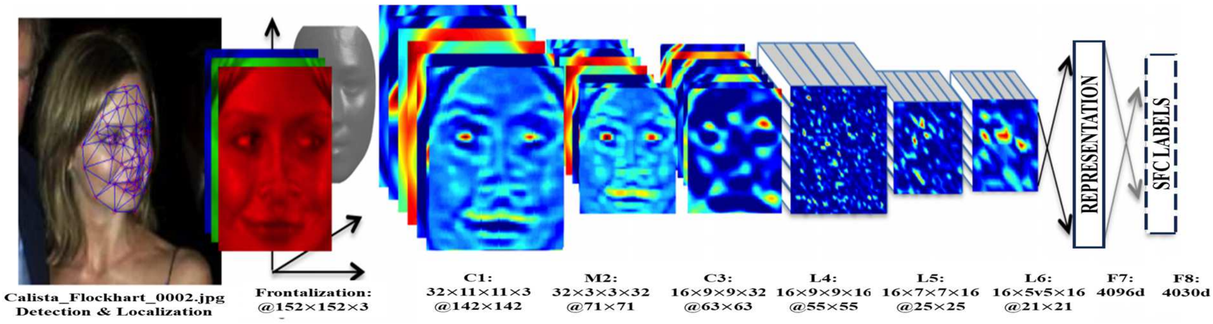

Fig. 7 Structural composition of DeepFace architecture [24]

Further layers, L4, L5, and L6, are specialized layers of local convolution, each employing 16 filters with dimensions of 9×9, 7×7, and 5×5 respectively, and all are activated by the ReLU function. These layers are intricately designed for localized feature extraction, progressively capturing more nuanced and detailed aspects of the facial features.

Advancing into the deeper layers, F7 is a fully connected layer with a substantial 4096 units, followed by ReLU activation. This layer is instrumental in integrating the features extracted by previous layers, transforming them into a high-dimensional feature space. Following F7, L8 is another fully connected layer with 4030 units, culminating with a softmax function. The softmax function is crucial for classification purposes, assigning probabilistic values to each possible class and facilitating the final decision on the facial recognition outcome.

The DeepFace architecture's meticulously designed layers and functions are a testament to the complex and sequential processing required for effective and accurate facial feature analysis and recognition. The initial segment, labeled C1, within the DeepFace architecture, undertakes a convolution operation on the 152×152×3 input image utilizing 32 filters of an 11×11 dimension. This stage employs the ReLU as its activation function, facilitating the extraction of primary features from the image and yielding an output dimension of 32 × 142 × 142. Subsequently, the second segment, M2, implements a downsampling mechanism on the output from C1. This is achieved using a 3×3 window with a stride of 2, selectively retaining the maximum value within each 3×3 region for the next layer. This process effectively reduces spatial dimensions while preserving critical feature information.

The third stage involves a convolution operation on M2's output using 16 filters of 9×9 dimension continues to employ ReLU. This layer is instrumental in extracting more intricate shapes and patterns, refining the output dimension to 16×63×63. Progressing to the localized convolution layers, L4, L5, and L6, each utilizes 16 filters of dimensions 9×9, 7×7, and 5×5 respectively, with ReLU activation. These layers are specifically designed for extracting and refining regional features of the image, with L4 adjusting the output size to 16×55×55, L5 to 16×25×25, and L6 to 16×21×21.

Advancing deeper into the network, the F7 layer, the architecture's first fully connected layer, employs 4,096 neurons to conduct a comprehensive analysis of the image. It amalgamates all preceding features using ReLU activation, synthesizing the image's features to discern complex patterns and relationships. Following this, the L8 layer, another fully connected layer, utilizes 4,030 neurons and a Softmax activation function. This layer is crucial for determining probabilistic outcomes, effectively facilitating face recognition and classification.

Each layer within this architecture plays a pivotal role in extracting and processing various levels of abstract features from the image, culminating in synthesizing complex information essential for accurately recognizing and classifying faces. The final output dimension is refined to 32×71×71. This multi-layered process not only reduces the data's dimensions but also preserves essential features while minimizing noise, thereby enhancing the overall efficiency of the model [14], [15]. Consequently, images processed through the DeepFace architecture yield critical feature information, facilitating nuanced individual classification and analysis [27], [28].

B. Dataset Employed in the Study

Fig. 8 visually depicts the categorization framework for human emotional expressions. It demonstrates the application of emotion labeling from the FER-2013 dataset, a widely recognized resource in emotion recognition. Facial and emotion recognition technologies, such as those analyzed in this study, are pervasive in diverse sectors and can delineate seven fundamental human emotions - happiness, anger, fear, surprise, disgust, sadness, and contempt.

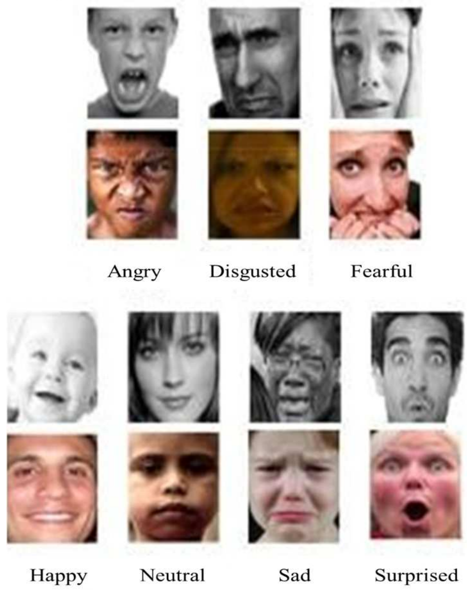

Fig. 8 Categorization of human emotional expressions [29]

These classifications are deduced from the analysis of facial muscle movements and expressions. In the context of this research, 3,000 instances from the dataset were methodically classified into these seven emotional states: anger, disgust, fear, happy, neutral, sad, and surprised. These categorizations were subsequently utilized to identify the corresponding emotions on human faces via the DeepFace algorithm, demonstrating the potential of deep learning in understanding and interpreting human affective states [29].

Fig. 9 presents a bifurcated visualization of the human facial image dataset and its corresponding classified dataset, both integral to implementing the DeepFace algorithm. These data models are formulated by applying the C++ facial recognition library, dlib, employed on Flickr, a prominent photo-sharing platform under the aegis of an American corporation, Yahoo. This procedure autonomously discerns and systematizes photographs containing human facial features, extracting and editing them to conform to copyright stipulations, thereby constructing the dataset. This dataset is an extensive collection of 70,000 PNG images, each with a resolution of 1024×1024. For this research, a subset comprising 3,000 of these images was meticulously selected and subjected to a classification process utilizing the capabilities of the library above [30].

Fig. 9 Compilation of human facial data(left) and its classified counterpart(Right)

Fig. 10 delineates the emotional labeling data spectrum as established in this study. From the corpus of 3,000 images, the emotion 'happy' emerged as the most predominant, constituting 44.7% with 1,342 images. This was followed by the 'neutral' classification, accounting for 36.8% with 1,142 images. The distribution of subsequent emotions is as follows: 'angry' with 163 images (5.4%), 'surprise' with 145 images (4.8%), 'disgust' with 92 images (3.1%), 'sad' with 86 images (2.9%), and 'fear' with 67 images (2.2%). This distribution serves as a foundational reference for subsequent comparative analysis with the outcomes derived from the DeepFace algorithm, thereby evaluating its proficiency in accurate emotion recognition.

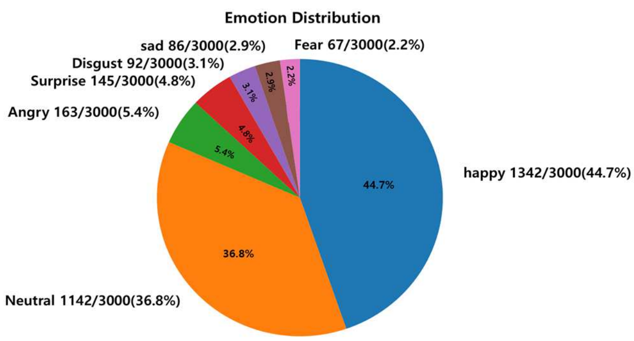

Fig. 10 Emotional labeling data spectrum in the current study

C. Analysis of Digital Illiterate Detection via the Deep Face Algorithm

Fig. 11 elucidates a graphical representation that contrasts the classification outcomes as discerned by human evaluators and the DeepFace algorithm within the contextually labeled dataset of this study.

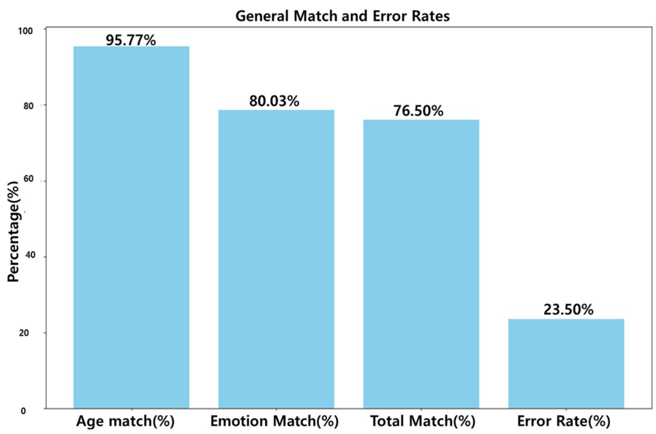

Fig. 11 Comparative analysis of age and emotion concordance and discrepancy rates via DeepFace

The graph quantifies the alignment between the human-labeled 'old' classification and the algorithm's designation of individuals as over 50, termed as 'age match.' Additionally, it defines the 'emotion match' as the overall concordance rate for the seven identified emotions. The 'total match' represents the mean concordance rate for age and emotional classifications. Conversely, the 'Error Rate' is computed by subtracting the total match from 100%. The empirical results demonstrate an age match of 95.77%, an emotion match of 80.03%, a total match of 76.5%, and an error rate of 23.5%.

Within the scope of this research, individuals above the age of 50 are hypothesized to be digital illiterates; accordingly, images depicting such individuals are labeled as 'old.' The notably high age match rate of 95.77% underscores the algorithm's precision in age categorization. Consequently, the efficacy of emotion recognition is posited as a pivotal metric for ascertaining digital illiterates, providing a nuanced understanding of the relationship between demographic variables and digital engagement.

Fig. 12 presents a graphical analysis delineating the concordance rates between emotional classifications conducted by human evaluators and those ascertained through the DeepFace algorithm, focusing on surprise and angry emotions. These emotions are hypothesized to be prevalent among individuals who experience difficulty in navigating digital interfaces. The graph quantitatively illustrates a concordance rate of 83.45% for the surprise emotion and 76.07% for the angry emotion, as determined by human assessment and the DeepFace algorithm's interpretation. The underlying assumption of this analysis is that the manifestation of surprise and angry emotions, particularly in conjunction with advanced age, may indicate an individual's digital illiteracy. It presupposes that individuals identified under this criterion will likely benefit from tailored assistance in enhancing their proficiency with digital devices.

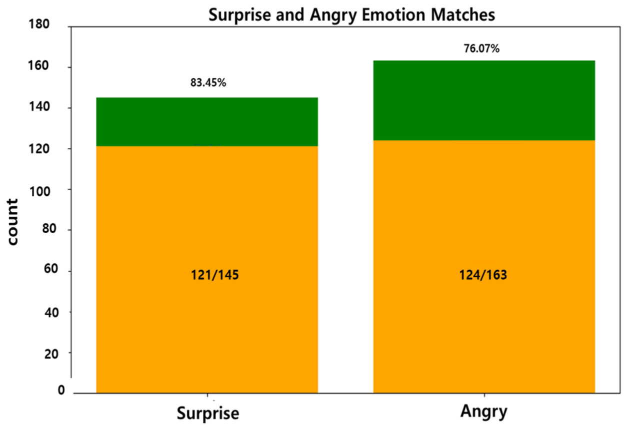

Fig. 12 Analysis of concordance rates for surprise and angry emotions among digital illiterates

IV. Conclusion

This study has systematically investigated an algorithmic approach to detecting digital illiteracy, using the DeepFace algorithm to analyze facial expressions and age. Utilizing a meticulously labeled dataset of 5,000 images, the research examined the concordance between human-labeled emotions and those identified by DeepFace, focusing mainly on the surprise and angry emotions prevalent among individuals struggling with digital devices. The results demonstrated a concordance rate of 83.45% for surprise and 76.07% for anger, highlighting the potential of facial analysis in identifying digital illiterates.

The implications of this research are multifold. Firstly, the high concordance rates, especially for angry emotions, indicate that emotional recognition can be a significant indicator of the challenge faced by digital illiteracy. Secondly, the study confirms the efficacy of the DeepFace algorithm in interpreting complex emotional and age-related facial cues, suggesting its applicability in real-world scenarios where digital illiterates need to be assessed quickly and non-invasively. Lastly, the findings open avenues for developing targeted services and support systems for digital illiterates, potentially enhancing their interaction with digital technologies and reducing the digital divide.

Acknowledgment

This study is supported by the Research Program funded by the SeoulTech (Seoul National University of Science and Technology).

References

[1] Achmad, Balza, and Kartika Firdausy. “Neural Network-Based Face Pose Tracking for Interactive Face Recognition System”. *International Journal on Advanced Science, Engineering and Information Technology*, vol. 2, no. 1, pp. 105-108, Feb. 2012, doi:10.18517/ijaseit.2.1.164.

[2] Y. Kim, W. Park, M. Roh and J. Shin, "GroupFace: Learning Latent Groups and Constructing Group-Based Representations for Face Recognition", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 5620-5629, 2020.

[3] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[4] S. Tariyal, R. Chauhan, Y. Bijalwan, R. Rawat and R. Gupta, "A comparitive study of MTCNN, Viola-Jones, SSD and YOLO face detection algorithms," *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bangalore, India, pp. 1-7, 2024, doi:10.1109/iitcee59897.2024.10467445.

[5] S. Qi, X. Zuo, W. Feng and I. G. Naveen, "Face Recognition Model Based On MTCNN And Facenet," *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, Karnataka, India, pp. 1-5, 2022, doi:10.1109/icmnwc56175.2022.10031806.

[6] R. M. Prakash, N. Thenmoezhi and M. Gayathri, "Face Recognition with Convolutional Neural Network and Transfer Learning," *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, pp. 861-864, 2019.

[7] Parkhi, O., Vedaldi, A., & Zisserman, A. “Deep face recognition,” *BMVC*, pp. 1–12, 2015.

[8] Simonyan, K., & Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *CoRR*, pp.1409-1556, 2014.

[9] J. Johnson, A. Karpathy and L. Fei-Fei, "DenseCap: Fully Convolutional Localization Networks for Dense Captioning," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 4565-4574, 2016.

[10] A. Sajjanhar, Z. Wu and Q. Wen, "Deep Learning Models for Facial Expression Recognition," *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, ACT, Australia, pp. 1-6, 2018, doi: 10.1109/dicta.2018.8615843.

[11] A. Sharma, V. Bajaj and J. Arora, "Machine Learning Techniques for Real-Time Emotion Detection from Facial Expressions," *2023 2nd Edition of IEEE Delhi Section Flagship Conference*, Rajpura, India, pp. 1-6, 2023, doi: 10.1109/delcon57910.2023.10127369.

[12] D. Gyawali, P. Pokharel, A. Chauhan and S. C. Shakya, "Age Range Estimation Using MTCNN and VGG-Face Model," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1-6, 2020, doi: 10.1109/icccnt49239.2020.9225443.

[13] Honey and S. S. Oberoi, "Revolutionizing Face Detection: Exploring the Potential of MTCNN Algorithm for Human Face Recognition," *2023 Seventh International Conference on Image Information Processing (ICIIP)*, Solan, India, pp. 300-306, 2023, doi:10.1109/iciip61524.2023.10537691.

[14] Kasim, Shahreen, et al. “A Study on Facial Expression Recognition Using Local Binary Pattern,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 5, pp. 1621-6, Oct. 2017, doi:10.18517/ijaseit.7.5.3390.

[15] N. Zhang, J. Luo and W. Gao, "Research on Face Detection Technology Based on MTCNN," *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, Xi'an, China, pp. 154-158, 2020, doi: 10.1109/iccnea50255.2020.00040.

[16] O. Khajuria, R. Kumar and M. Gupta, "Facial Emotion Recognition using CNN and VGG-16," *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, pp. 472-477, 2023, doi: 10.1109/icict57646.2023.10133972.

[17] A. Atanassov and D. Pilev, "Pre-trained Deep Learning Models for Facial Emotions Recognition," *2020 International Conference Automatics and Informatics*, Varna, Bulgaria, pp. 1-6, 2020, doi:10.1109/icai50593.2020.9311334.

[18] D. Sil, A. Dutta and A. Chandra, "Convolutional Neural Networks for Noise Classification and Denoising of Images," *TENCON 2019 - 2019 IEEE Region 10 Conference*, Kochi, India, pp. 447-451, 2019, doi:10.1109/tencon.2019.8929277.

[19] Y. Tao, "Image Style Transfer Based on VGG Neural Network Model," *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, pp. 1475-1482, 2022, doi: 10.1109/aeeca55500.2022.9918891.

[20] D. Gyawali, P. Pokharel, A. Chauhan and S. C. Shakya, "Age Range Estimation Using MTCNN and VGG-Face Model," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1-6, 2020, doi: 10.1109/icccnt49239.2020.9225443.

[21] A. R. S and V. R. V, "Age & Gender Recognition Using Deep Learning," *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, Kannur, India, pp. 386-390, 2022, doi:10.1109/icicict54557.2022.9917573.

[22] B. R. Ilyas, B. Mohammed, M. Khaled and K. Miloud, "Enhanced Face Recognition System Based on Deep CNN," *2019 6th International Conference on Image and Signal Processing and their Applications (ISPA)*, Mostaganem, Algeria, pp. 1-6, 2019, doi: 10.1109/ispa48434.2019.8966797.

[23] Z. Lu, X. Jiang and A. Kot, "Enhance deep learning performance in face recognition," *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, China, pp. 244-248, 2017, doi: 10.1109/icivc.2017.7984554.

[24] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1701-1708, 2014, doi:10.1109/cvpr.2014.220.

[25] A. Awana, S. V. Singh, A. Mishra, V. Bhutani, S. R. Kumar and P. Shrivastava, "Live Emotion Detection Using Deepface," *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, pp. 581-584, 2023, doi: 10.1109/ic3i59117.2023.10397747.

[26] A. Firmansyah, T. F. Kusumasari and E. N. Alam, "Comparison of Face Recognition Accuracy of ArcFace, Facenet and Facenet512 Models on Deepface Framework," *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Jakarta, Indonesia, pp. 535-539, 2023, doi:10.1109/iccosite57641.2023.10127799.

[27] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1701-1708, 2014, doi:10.1109/cvpr.2014.220.

[28] Cate, Hardie & Dalvi, Fahim & Hussain, Zeshan, “DeepFace: Face Generation using Deep Learning”, 2017.

[29] L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim and S. Bahri Musa, "The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi," *2020 Fifth International Conference on Informatics and Computing (ICIC)*, Gorontalo, Indonesia, pp. 1-9, 2020, doi:10.1109/icic50835.2020.9288560.

[30] H. Arabian, V. Wagner-Hartl, J. Geoffrey Chase and K. Möller, "Facial Emotion Recognition Focused on Descriptive Region Segmentation," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico, pp. 3415-3418, 2021, doi: 10.1109/embc46164.2021.9629742.