A Depression Severity Prediction Model by Handwriting

Tanabe Hiroto * and Kimura Masaomi

Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan
Email: ma23117@shibaura-it.ac.jp (T.H.); masaomi@shibaura-it.ac.jp (K.M.)
*Corresponding author

***Abstract*—In this study, we propose a model for determining depression severity through handwriting analysis. Traditional diagnostic methods rely on subjective communication between doctor and patient, which can lead to varied interpretations and prolonged diagnostic timelines. It is expected that handwritten data can be used to objectively diagnose depression. We utilize Residual Neural Network (ResNet) and Gradient-weighted Class Activation Mapping++ (Grad-CAM++) to identify regions of interest in handwriting images, associating these with normalized handwriting speed, which has been shown to correlate with depressive states. Experimental results showed that the model's region of interest focuses on the slow rather than the fast rate part. This model approach facilitates early and efficient detection of depression by making the process more accessible and minimizing the need for specialized equipment.**

***Keywords*—depression, handwriting analysis, Convolutional Neural Network (CNN), Gradient-weighted Class Activation Mapping++ (Grad-CAM++), Residual Neural Network (ResNet)**

I. Introduction

According to the World Health Organization, approximately 300 million people worldwide suffer from depression [1]. Currently, diagnosis is determined from interviews based on information gathered through conversations between patients and doctors [2]. The doctor interprets this and makes a diagnosis. However, there are two main problems with this diagnostic approach: first, it is time-consuming to make a diagnosis; second, subjective judgements can lead to variations in diagnosis. Additionally, questionnaire-based assessments rely on patients' self-reported responses, which can be influenced by their subjective perception of symptoms or even deliberately altered. This reliance on patient-reported information can lead to variability and bias in diagnostic results, highlighting the need for an objective and efficient model to determine the severity of depression.

Previous studies have used a variety of data to determine depression, including voice [3, 4], facial expressions [5–8], behavior [9, 10], Electroencephalography (EEG) [11, 12] and handwriting [13, 14]. However, the above methods involve several financial costs. Microphones are needed to take voice, cameras are needed to capture facial features and behavior, expensive equipment is needed to take brain waves, and tablets and Stairs pens are needed to take time series data on handwriting. There is also time costs involved, such as setting up and staging to take the information. However, using images of handwriting is low-cost and can be carried out anywhere. Therefore, whether depression can be determined from images of handwriting should be tested in the future. It would be challenging to estimate depression from handwriting. There are two reasons for this: first, handwriting science is considered a pseudo-science, and it is unknown whether depression can be characterized from images of handwriting; second, the amount of data is small, and some ingenuity is required to create the data. Therefore, the proposed method was developed by devising how to input handwriting data into the model and how to analyze the results.

This study proposes a model to predict the severity of depression using images of handwritten text. By using a Convolutional Neural Network (CNN) based on Residual Neural Network (ResNet) with skip connections [15], handwritten data is used to infer the severity of depression.

The rest of this paper is organized as follows: Section II highlights our contributions. Section III reviews related works. Section IV explains the proposed method. Section V presents experimental results, followed by discussions in Section VI. Finally, Section VII concludes the paper with a summary and future directions.

The contributions of this paper are as follows:

- We propose a model to estimate the severity of depression based on handwriting data, demonstrating its potential for effective mental health assessment.
- We propose a method to investigate the relationship between the model's attention regions and handwriting velocity

II. Related Works

Various methods have been explored to determine depression, focusing primarily on non-verbal behaviors such as voice (speech signals), facial expressions, posture, EEG and handwriting. Approaches using speech signals have analyzed acoustic biomarkers, such as F0 and its variability, which differ between clinical and non-clinical

Manuscript received January 7, 2025; revised January 23, 2025; accepted February 13, 2025; published April 16, 2025.

subjects. However, speech signals are susceptible to external variability factors, which can affect feature reliability [3]. Other speech-related studies, including work by Williamson *et al.* [4], have examined motor coordination in speech as a potential indicator for inferring depression severity. Facial expression analysis and behavior pattern analysis approaches have also yielded promising results. For instance, in the AVEC Challenge, models using subtle changes in facial expression and voice achieved accuracies exceeding 80% in depression detection [5, 6]. Furthermore, advanced analysis of 2D and 3D facial data, as shown in studies by Zhou *et al.* [7] and Huang *et al.* [8], improved accuracy by simultaneously analyzing spatial and spatiotemporal facial dynamics. Posture and gait pattern analysis has been linked to depression signs as well. For example, Michalak *et al.* [9] found a connection between gait patterns and depression, while Canales *et al.* [10] associated poor posture with recurring depressive episodes. EEG-based approaches have also been widely studied. Research by Seal *et al.* [11] and Spyrou *et al.* [12] demonstrated that EEG patterns in depressed individuals differ significantly from healthy individuals, showing promise for objective assessments of emotional states. However, EEG measurement requires specialized equipment, limiting its accessibility in routine practice.

Previous research has demonstrated significant handwriting changes in patients with bipolar disorder during manic episodes, suggesting that handwriting features can serve as indicators of mental health states [16]. Specifically, these features have been found useful for predicting transitions into manic episodes. Building on this foundation, our study focuses on handwriting characteristics associated with depression, further supporting the potential of handwriting analysis as a tool for mental health assessment.

Two previous studies on handwriting are described. Laurence *et al.* [13] created the first publicly available handwriting and drawing database (EMOTHAW) linking emotional states to handwriting. They used random forests to analyze data from 129 participants (71 women, 58 men) performing seven tasks (five drawings, two writings). Their study revealed that people with depression took longer to complete tasks than healthy participants, based on time spent with the pen grounded versus lifted. Juan *et al.* [14] also utilized the EMOTHAW dataset to predict depression based on time-series features. By analyzing six features—total task time, pen floating time, pen placement time, stroke count, and mean writing pressure—Juan's team achieved 80.31% accuracy in depression detection. They Fourier-transformed these time-series features to account for finer temporal information, improving predictive accuracy. The above studies demonstrate successful handwriting-based approaches for predicting depression, though time-series data tends to have a larger volume.

As mentioned above, the problem with previous research is the cost of materials and the burden on the patient. The use of handwriting images can therefore minimize material costs and patient burden.

III. Proposed Method

A. Prerequisites

This study proposes a model that uses Resnet to output an estimated severity score on a scale from 0 to 25 from input handwritten images. We first describe how the dataset of images input to the model is created. In this study, we use data from a previous study [13], in which seven items are recorded in one file: x, y, timestamp, pen status, azimuth, altitude and pressure for each of the four words written. When creating the image, the x, y coordinates are used. In the original data, the time intervals between timestamps were long, so we used spline interpolation to create smooth curves between the handwriting data points. To remove points that were likely recorded in areas where the pen was not in contact with the surface, we applied a method where the central point was blacked out if there were two or fewer handwriting pixels within a surrounding 25-pixel (5×5) area. Next, time-series data for the four words were contained within a single file, necessitating the separation of this data into individual files for each word. During this process, significant shifts in the x-coordinates were observed, prompting the detection of the y-coordinates, which often exceeded a predefined threshold. This change enabled the recognition of movements corresponding to the next word once the previous work was completed, allowing for successful segmentation.

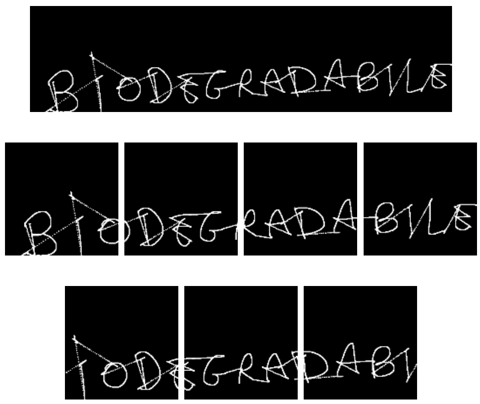

Fig. 1. Examples of processing and division.

Each word was then transformed into a 300×1200 image based on the data. The handwriting was rendered in white against a black background. Each image was divided into four 300×300 images for three reasons. First, since the amount of data is limited, segmenting the words further increases the dataset size. Second, this approach helps capture the individual handwriting tendencies of the participants rather than overall writing trends (e.g., an entire sentence slanting upward). Third, each 300×300 image is ensured to contain at least one character. Ideally, segmenting each character individually would be preferable, but due to variations in character size and individual writing habits, such division is challenging. Therefore, we opted to use 300×300 segmentation. Additionally, three intermediate images were generated to capture details that may have been lost during the initial division, resulting in a total of seven images per word (Fig. 1).

Recognizing that the dataset was unbalanced, we employed data augmentation techniques such as parallel migration and shrinking. This strategy aimed to equalize the number of data points at all severity levels from 0 to 25.

B. Our Model

Our model uses a combination of ResNet and Fully Connected (FC) layers, specifically five ResNet layers and five FC layers (Fig. 2). CNNs are usually more effective for image processing tasks such as handwriting analysis, so we initially used a CNN with convolutional layers. However, a challenge emerged in ensuring that the CNN model kept the predictions constant. This problem was attributed to the max pooling layer, which generalizes individual handwriting features and removes essential information. In our dataset, which consisted of only four words written by 129 participants in identical texts, Max Pooling reduced the sensitivity of the model to individual differences and kept it focused only on common word information. Max pooling produces an effect like thickening, as shown in the image on the right. This process can distort original straight and curved lines, potentially making it more challenging to accurately extract fine-grained features.

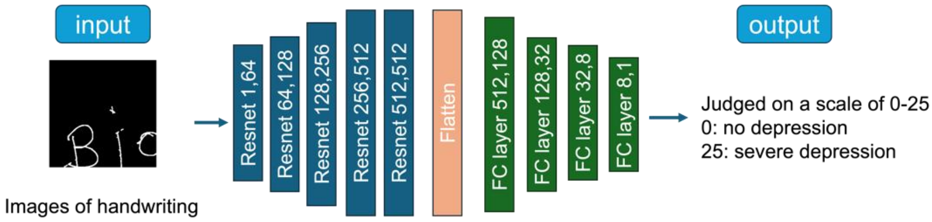

Fig. 2. Our model.

To address this issue, the removal of the Max Pooling layer reduced certain predictive value issues, but still did not achieve the desired level of accuracy. To further improve performance, the ResNet layer was made into a deeper network, which retained more nuanced handwriting features through skip connections.

C. Relationship between the Model's Region of Interest and Handwriting Speed

To validate the model, we examine the relationship between handwriting speed and the model's regions of interest. Prior research has shown that depressed patients tend to take longer to complete tasks. If it can be shown that the model focuses on areas of slow handwriting speed, it can be shown that attention is on the areas where depressed patients would cognitively require attention, i.e., at the beginning and end of writing. To show this, the following approach is taken. Handwriting speed is measured from time series data, normalized and divided by the median into fast and slow groups. The denominator represents the number of coordinators in the fast group with a Gradient-weighted Class Activation Mapping++ (Grad-CAM++) [17] value of 0.5 or higher, while the numerator represents the number of coordinators in the slow group with a Grad-CAM++ value of 0.5 or higher.

To facilitate understanding of the following analysis, we first define the key notations used in the equation:

- G: The Grad-CAM++ value for each pixel, which ranges from 0 to 1 and represents the degree of attention the model pays to that pixel.
- v_s : The group of data points with slower handwriting speeds, determined by splitting the dataset based on the median speed.
- v_f : The group of data points with faster handwriting speeds, also determined by the median speed.
- $N_{v_s}(G > 0.5)$: The total number of pixels in v_s where $G > 0.5$.
- $N_{v_f}(G > 0.5)$: The total number of pixels in v_f where $G > 0.5$.
- r: The ratio of $G > 0.5$ pixels between v_s and v_y, calculated as follows:

Using the definitions above, the equation for x is given as:

$$r = \frac{N_{v_s}(G>0.5)}{N_{v_f}(G>0.5)} \tag{1}$$

This equation represents the relative attention rate of pixels with $G > 0.5$ in the slow group (v_s) compared to the fast group (v_f). A higher value of r indicates a greater concentration of $G > 0.5$ pixels in v_s relative to v_f Therefore, a large value of r can suggest that the learning model focuses on slow areas and reflects the characteristics of depression shown in previous studies.

The ratio of these two values is then calculated. If the ratio exceeds 1, it indicates that the model focuses more on areas with slower handwriting speeds. By verifying the above method, it can be demonstrated that the model exhibits characteristics of depression as indicated in previous studies

IV. Experiment

A. Objectives

The objectives of this experiment are as follows:

- To evaluate the accuracy of depression severity estimation based on handwriting images and validate the effectiveness of the proposed model.
- To examine whether the proposed model accurately captures the features identified in previous studies by analyzing the relationship

between handwriting speed and regions highlighted by Grad-CAM++ and confirming consistency with existing findings.

B. Dataset

The EMOTHAW dataset comprises 129 participants, all of whom completed the DASS test to assess the severity of their depression. Test scores ranged from 0 to 25, with the following distribution: 95 participants scored between 0–9, 14 participants scored between 10–13, 13 participants scored between 14–20, and 7 participants scored 21 and above. Participants engaged in seven tasks—five drawing and two writing—using an iPad, while time-series data on their handwriting was collected. For each task, eight types of time-series data were recorded: x-coordinates, y-coordinates, timestamps, whether the pen was grounded to the paper, longitude, latitude, and writing pressure. To train the model as an image, we constructed images using three features: x-coordinates, y-coordinates, and timestamps.

The default dataset comprises handwriting samples from 129 participants, where each participant wrote four words. However, this dataset is considered insufficient and unbalanced due to the distribution of depression severity among the subjects, which skews toward the milder end of the spectrum. For instance, the most frequent label corresponds to 4, represented by 16 subjects, yielding 16×4×7 = 448 images. To address the imbalance, data augmentation was performed to ensure that each severity label contained 448 images. This augmentation involved techniques such as parallel movement and reduction, leading to a total of 11,200 samples (note that label 11 was not present). The complete dataset was subsequently divided into training and test subsets through three distinct approaches.

We divided 129 participants into a 9:1 ratio, using data from 116 participants for training and 13 participants for testing. The 13 individuals selected for the test set were chosen to avoid bias, with one person representing each of the 13 labels from the set {0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 21, 25}. All data related to these selected individuals were included in the test set. The rationale for implementing Dataset 1 is that it represents the most practical and realistic approach for future applications. Our goal is to train the model on historical data so that it can predict the severity of depression based on handwriting from new, unseen individuals. Therefore, this dataset reflects the partitioning method most closely aligned with this real-world use case.

C. Method

To achieve the above objectives, the following procedures are conducted:

1) Implementation of datasets

Train a model for depression severity estimation based on handwritten images. Evaluate the accuracy using MSE and validate the proposed model.

2) Feature verification

To validate a feature identified in previous research—depressed patients take longer to complete tasks—the relationship between handwriting speed and the areas highlighted by Grad-CAM++ will be analyzed.

D. Results

The training loss was less than 1. The final Mean Squared Error (MSE) value converged to approximately 68. As illustrated in the box-and-whisker diagram above (Fig. 3), the model demonstrates a discernible trend, suggesting that it has learned the underlying patterns to some extent. In previous studies, depression scores measured using the DASS scale were classified into two categories: scores of 0–9 were considered non-depressed, while scores of 10 or higher indicated depression. Since the proposed method predicts depression severity, it can be adapted for binary classification by applying the same threshold. Using the model trained on Dataset 1, we computed the binary classification score, achieving an accuracy of 72.25%.

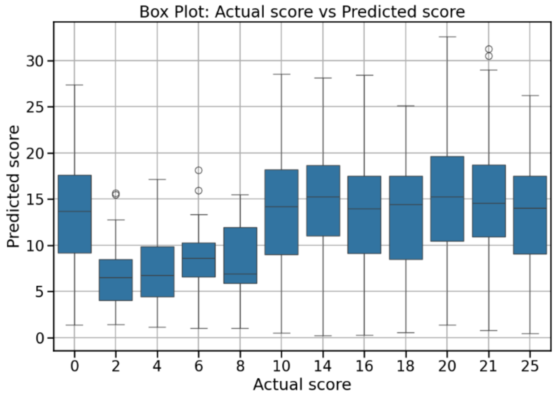

Fig. 3. Box-and-whisker plot of predicted and measured values for the first data set.

Based on the method described in the proposed method, the features of the previous studies are learnt to test whether the model learns more about the areas with slower handwriting speed. Fig. 4 shows the speed of the handwriting and the scatterplot of all Grad-CAM++ images in aggregate. The denominator represents the number of fast groups with a Grad-CAM++ value of 0.5 or more, while the numerator represents the number of coordinates in the slow groups with a Grad-CAM++ value of 0.5 or more. Let this value be denoted as r, where $r = 1.21$. This result indicates that Grad-CAM++ focuses more on regions where handwriting speed is slower.

Fig. 4. Image storing handwriting speed by adding a color element to the handwriting (upper image) and output image when using Grad-CAM++ (lower image).

V. Discussion

In this study, we successfully identified the regions of interest within handwriting images using Grad-CAM++. The results revealed that the model's attention areas were closely associated with regions where handwriting speed was slower, specifically at the beginning and end of the writing process. This finding aligns with prior research, which has shown that individuals with depression take longer to complete tasks compared to healthy individuals. The fact that the model focused on the start and end of the writing process suggests that it effectively captures cognitive changes characteristic of depression. The beginning of the handwriting process corresponds to the initial formation of letters, which is considered cognitively demanding as it involves planning and motor coordination. Similarly, the end of the writing process may impose an increased cognitive load due to the mental awareness of task completion and the need for precise motor control in finalizing character shapes. The model's strong attention to these specific regions indicates that individuals with depression may experience heightened cognitive burdens at the start and end of handwriting tasks. This suggests that our model not only recognizes these cognitive challenges but also has the potential to distinguish depression-specific handwriting patterns effectively.

In prior studies, the highest accuracy for binary classification was 71.6% in [13] and 80.31% in [14]. The proposed method achieved an accuracy of 72.25% for binary classification of depression. Although this falls short of the 80.31% achieved by the 2021 model that utilized Fourier transform on time-series data, it outperformed the 71.6% obtained by the model trained on time-series data using a random forest. These results indicate that, in this study, depression severity could be reasonably predicted using only images, without relying on temporal information.

However, this study focused only on the temporal aspects of the data without considering other factors, such as pen pressure, and trained the model using these aspects as images. Additionally, due to the limited number of data samples, further validation is needed to assess the generalizability of the results. It is also important to note that the dataset used in this study was collected in Italy. As future work, it will be essential to collect difference language datasets and validate the proposed method on data from individuals in different countries to assess its generalizability.

On the other hand, data privacy protection is also an important issue. Since handwriting data is closely related to personal information, it is necessary to implement anonymization and encryption techniques to manage the data securely. In this study, the dataset used has already been irreversibly anonymized and cannot be re-identified. From an ethical perspective, it is essential to collaborate with research institutions and medical organizations and follow appropriate procedures to utilize the data.

VI. Conclusion

We developed a model to estimate the severity of depression from handwriting and confirmed that it could make predictions to some extent. The model utilized ResNet and Grad-CAM++ to investigate the relationship between the regions of interest and handwriting speed, confirming that the model primarily learned features associated with slower speeds. This suggests that the model is capturing characteristics identified in previous research. We believe this study demonstrates the feasibility of saving time-series data as images for inference.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Tanabe conducted the data preprocessing, developed the model, performed the experiments, evaluated the results, and wrote the manuscript. Kimura was responsible for the planning. All authors approved the final version of the manuscript.

Funding

This research was funded by JSPS KAKENHI Grant Number JP23KJ1924, JP23K24935.

Acknowledgement

The authors thanks for the funds by JSPS KAKENHI Grant Number JP23KJ1924, JP23K24935. The authors also express their gratitude to the contributors of the previous study [13] for providing the dataset.

References

[1] World Health Organization. (2017). Depression and other common mental disorders: Global health estimates. [Online]. Available: https://www.who.int/publications/i/item/depression-global-health-estimates

[2] D. A. Regier *et al*., "DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses," *American Journal of Psychiatry*, vol. 170, no. 1, pp. 59–70, 2013.

[3] A. Esposito and A. M. Esposito, "On the recognition of emotional vocal expressions: Motivations for a holistic approach," *Cognitive Process*, vol. 13, no. 2, pp. 541–550, 2012.

[4] J. R. Williamson *et al.*, "Vocal biomarkers of depression based on motor incoordination," in *Proc. the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 41–48.

[5] M. Valstar *et al*., "Continuous emotion and depression recognition challenge," in *Proc. the 3rd ACM Int. Workshop Audio/Visual Emotion Challenge (AVEC 2013)*, 2013, pp. 3–10.

[6] M. Valstar *et al*., "3D dimensional affect and depression recognition challenge," in *Proc. the 4th Int. Workshop on Audio/Visual Emotion Challenge (AVEC 2014)*, 2014, pp. 3–10.

[7] X. Zhou *et al*., "Visually interpretable representation learning for depression recognition from facial images," *IEEE Trans. Affective Comput*., 2018.

[8] M. Niu *et al*., "Multimodal spatiotemporal representation for automatic depression level detection," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 294–307, 2023.

[9] J. Michalak *et al*., "Embodiment of sadness and depression—gait patterns associated with dysphoric mood," *Psychosom. Med*., vol. 71, no. 5, pp. 580–587, 2009.

[10] J. Z. Canales *et al.*, "Investigation of associations between recurrence of major depressive disorder and spinal posture alignment: A quantitative cross-sectional study," *Gait Posture*, vol. 52, pp. 258–264, 2017.
[11] Seal *et al*., "DeprNet: A deep convolution neural network framework for detecting depression using EEG," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
[12] Spyrou *et al*., "Geriatric depression symptoms coexisting with cognitive decline: A comparison of classification methodologies," *Biol. Signal Process. Control*, vol. 25, pp. 118–129, 2016.
[13] Likforman-Sulem *et al*., "EMOTHAW: A novel database for emotional state recognition from handwriting and drawing," *IEEE Trans. Hum.-Mach. Syst*., vol. 47, no. 2, pp. 273–284, 2017.
[14] Nolazco-Flores *et al*., "Emotional state recognition performance improvement on a handwriting and drawing task," *IEEE Access*, 2021.
[15] K. He *et al*., "Deep residual learning for image recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[16] N. Ayaz *et al*., "The use of handwriting changes for the follow-up of patients with bipolar disorder," *Noro Psikiyatr Ars*., vol. 59, no. 1, pp. 3–9, 2022.
[17] A. Chattopadhay *et al*., "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 839–847.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).