BMC Medicine

RESEARCH

Open Access

Major depressive disorder recognition based on electronic handwriting recorded in psychological tasks

Chong Li[1], Kunxue Zhang[3], Qunxing Lin[2], Shan Huang[2], Wanying Cheng[2], Yueshiyuan Lei[2], Xinyu Zhao[2] and Jiubo Zhao[1,2,4*]

Abstract

Background This study aimed to determine whether handwriting patterns are altered in individuals experiencing depressive episodes. Additionally, we developed a model for the recognition of major depressive disorder (MDD) based on electronic handwriting in psychological tasks.

Methods A total of 130 patients and 117 healthy controls completed 21 psychology-related handwriting tasks. The electronic tablet recorded several handwriting characteristics, including horizontal and vertical coordinates, nib pressure and speed, and inclination angle. The statistical indicators for each handwriting characteristic were calculated. Statistical analyses, including differential analysis, were performed to identify predictors of depression. Furthermore, logistic regression and machine learning models were developed to discriminate MDD.

Results The study included 130 patients with onset depression (mean (standard deviation (SD)) age, 20.42 (5.21)) and 117 healthy controls (mean (SD) age, 20.54 (2.60)). The *t*-test and logistics analysis results indicated that depressed patients exhibited a higher minimum of handwriting pressure, an elevated median of handwriting speed, and greater pen tip jitter. The LightGBM machine learning model exhibited satisfactory performance, with a cross-validated area under the receiver operating curve of mean 0.90 (SD, 0.01). The analysis of variance revealed that the negative question–answer task model exhibited superior performance compared to the neutral and positive task models.

Conclusions The present study indicates that depressed patients exhibit modal handwriting changes and developed a cost-effective, rapid, and valid model for identifying MDD. This finding established a strong foundation for developing multimodal recognition models in the future.

Keywords Major depressive disorder, Handwriting, Psychology, Machine learning, Mental health recognition

*Correspondence:
Jiubo Zhao
jiubozhao@126.com
[1] Department of Psychiatry, Zhujiang Hospital, Southern Medical University, No. 253, Industrial Avenue Zhong, Guangzhou, Guangdong, China
[2] Department of Psychology, School of Public Health, Southern Medical University, No. 1838 Guangzhou Dadao Road North, Guangzhou, Guangdong, China
[3] Department of Neurology, Nanfang Hospital, Southern Medical University, No. 1838 Guangzhou Dadao Road North, Guangzhou, Guangdong, China
[4] Guangdong-Hong Kong Joint Laboratory for Psychiatric Disorders, Guangzhou, China

Background

Major depressive disorder (MDD) is a prevalent mental disease that is associated with significant physical and emotional harm [1], negatively affecting the lives of many people worldwide. The 2017 global estimate from the World Health Organization indicates that the number of individuals with MDD worldwide exceeds 300 million and has been on the rise for the past decade [2]. The morbidity rate of MDD in China is higher than the global average, which results in frequent suicides among adolescents [3]. Given that earlier diagnosis, prompt psychosocial interventions, and medical care can efficiently alleviate and even cure MDD, it is imperative to enhance the efficiency of identifying adolescents with MDD [4].

In clinical practice, it may be challenging to establish a successful diagnosis and the corresponding effective treatment due to the significant heterogeneity of MDD. A significant factor contributing to the delayed or underdiagnosis is the absence of accepted objective criteria for evaluating MDD [5]. Therefore, multiple studies started to investigate the efficacy of objective biomarkers for depression diagnosis in adolescents. Magnetic resonance imaging of the brain [6–8], monoaminergic [9–11], and immune and endocrine systems [12–14] have all been associated with depression in adolescents. However, meta-analyses suggest these candidate biomarkers have limited clinical predictive value [15]. In addition to scale scores and biomarkers, objective measures, including voice [16–19], heart rate variability [20–22], and facial expressions [23–25], have been the focus of research in recent years. These measures could serve as impartial criteria to facilitate MDD identification. The present study employs a novel approach to identify MDD through handwriting features, which presents a quantitative index of psychomotor symptoms of MDD, including slowed thought processes, physical restlessness, and severe retardation in gross and fine body movements [26].

A significant correlation has been observed between physical activity and MDD [27]. Recent studies have suggested that individuals experiencing fatigue are more likely to suffer from depression, with fatigue itself being a strong predictor of depression [28, 29]. In biological studies, T.M. Gao et al. reported reduced ATP levels in a mouse model of depression [30, 31]. As handwriting can be described as a fine body movement, ATP reduction can affect handwriting activities. Furthermore, Multiple Code Theory [32], a general theory of emotional information processing, can explain the strong interrelatedness among bodily activation, symptoms, and how individuals communicate their emotional experiences [32–34]. A recent study established a correlation between the speed of speech and writing in patients with depression [35]. These findings suggest that depressed people could have altered handwriting patterns during handwriting activities.

In 2002, Oliver Tucha found that before and after taking medication, patients with depression exhibited changes in handwriting behavior, including handwriting speed and acceleration of the pen movement [36]. A 2004 study by R. Mergl and colleagues compared handwriting speed in patients with depression and healthy controls and found that patients with depression wrote significantly slower [37]. In 2010, Sara Rosenblum and colleagues found that older patients with mild depression had significantly altered handwriting patterns. The depression and control groups exhibited significant differences in the speed at which they wrote, the duration of time they left the pen unused, and the strength of the pen [38]. Additionally, Jeremia Heinik and colleagues reported that individuals with mild depression exhibited distinct patterns in areas such as pen pressure during handwriting compared to normal controls [39]. These studies indicated that people with depression could have differences in handwriting speed, strength, and font size compared to healthy people. However, previous studies primarily focused on middle-aged [36, 37] or elderly populations [38] [39], with small sample sizes ($n < 50$) and a lack of comprehensive handwriting features. This study focused on the adolescent population, with a large sample size ($n > 100$) and more comprehensive indicators, including writing coordinates, writing pressure, speed, and inclination in several psychological tasks.

In this study, we aimed to determine whether distinctive alterations in handwriting characteristics were present in adolescent individuals with depression and develop a rapid, cost-effective, and valid model of depression identification based on handwriting characteristics.

Methods

Subjects

The study included 130 outpatients with onset MDD recruited from the Southern Medical University, the Zhujiang Hospital, and the Baiyun Mental Health Hospital in Guangzhou. All participants provided written informed consent before inclusion. The study was registered at the Chinese Clinical Trial Registry with the number ChiCTR2400083328. Our study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.

The authors asserted that all procedures contributing to this work complied with the ethical standards of the relevant national and institutional committees on human experimentation and the Helsinki Declaration of 1975, as revised in 2013. All procedures involving human subjects/patients were approved by the Health Research

Ethics Committee of the Southern Medical University (NFYKDX003).

The inclusion criteria included patients who had a score higher than 7 on the 17-item Hamilton Depression Rating Scale (HAMD-17), which indicates a mild to severe depressive episode [40]. Patients were screened with the Mini-International Neuropsychiatric Interview [41]. All patients met the criteria for MDD according to the *International Statistical Classification of Diseases and Related Health Problems; a specialist in psychiatry confirmed the Tenth Revision code and the MDD diagnosis*. None of the participants received any sedative or hypnotic medication. The exclusion criteria included acute suicidal ideation or psychosis, other primary Axis I psychiatric disorder, alcohol or substance use disorder or dependence, severe somatic illness, history of severe brain injury, no fluency in Chinese, failure to complete all tasks, and current or planned pregnancy or breastfeeding. A total of 152 patients were enrolled in the trial; however, 22 were excluded due to non-compliance.

Furthermore, 131 healthy controls with no Axis I psychiatric disorder were recruited from the Southern Medical University, Zhujiang Hospital, or an online recruitment site. The healthy control group was age- and sex-matched as closely as possible to the patient group following the inclusion and exclusion criteria. However, the healthy individuals did not have a current diagnosis or a history of mental illness. Besides, 14 participants were excluded from the study due to failure to complete the tasks.

Record of handwriting features

In real time, we extracted and recorded the handwriting characteristics from a digital tablet (Hanvon-100040774427, Beijing, China) and saved them as a file in CSV format. There were six handwriting characteristics, including the horizontal and vertical coordinates of the handwriting nib position (sampling frequency 200 Hz, hereinafter referred to as x and y coordinate), instantaneous handwriting speed (sampling frequency 200 Hz), instantaneous handwriting pressure ($1-8192$ degrees, sampling frequency 200 Hz), instantaneous tilt of the nib X-axis ($0-90$ degrees, sampling frequency 200 Hz, hereinafter referred to as angle 1), and instantaneous tilt of the nib Y-axis ($0-90$ degrees, sampling frequency 200 Hz, hereinafter referred to as angle 2). The experiment consisted of 21 handwriting tasks (Table 1) in which subjects wrote their answers on a digital pad (Additional File 1: Fig. S1). To accurately represent the subjects' emotional state, there was no predetermined completion time. Most subjects could complete all the tasks within 20–30 min.

Statistics

A total of 13 basic statistical values were calculated for the six handwriting characteristics in each task, including the sum value of all data points (Sum), the number of data points (Len), mean, median, upper quartile, lower quartile, maximum (Max), minimum (Min), extreme deviation (Range), interquartile range, standard deviation (SD), variance (Var), and coefficient of discretion being recorded (Table 1). This resulted in 78 variables for each task. In each task, the following statistics strategy was used. Numeric variables are presented as mean ± SD, while categorical variables are presented as counts and frequencies. The chi-square (χ^2) test was used to compare categorical data between the two groups. For continuous variables, an independent sample t-test was used as appropriate. To identify MDD predictors, variables passing Bonferroni correction ($p_{\text{adjusted}} < 0.05/78$) were included in univariate logistic regression. Furthermore, variables from the univariate analysis passing Bonferroni

Table 1 Description of handwriting tasks and features

Tasks/variables	Description
Task 1	Name writing task
Task 2 – 5	Number-related tasks
Task 6 – 8	Vocabulary transcription tasks
Task 9 – 11	Negative question and answer tasks
Task 12 – 14	Neutral question and answer tasks
Task 15 – 17	Positive question and answer tasks
Task 18 – 21	Picture description tasks
Handwriting features	*x* coordinate, *y* coordinate, speed, pressure, tilt of the *x*-axis, and tilt of the *y*-axis
Basic statistical values	The sum value of all data points, the number of data points, mean, median, upper quartile, lower quartile, maximum, minimum, extreme deviation, interquartile range, standard deviation, variance, and coefficient of discretion

Six handwriting features were recorded for each task, and 13 basic statistical values were computed for each handwriting feature. This resulted in $6 \times 13 = 78$ parameters for each task. Details of the 21 tasks are provided in Additional File 1 correction $p_{adjusted} < 0.05$ and $|\beta| > 0.01$ were included in the multivariate logistic regression to identify independent determinants of MDD. Significant variables ($p < 0.05$) in multivariate logistics analysis were used to construct a new logistic regression model for each task. The receiver operating characteristics curve was used to indicate the sensitivity and specificity of the model. If there were no significant variables in the multivariate analysis, all predictors in the univariate analysis were included in the new logistics model to predict MDD. Moreover, positive predictive value (PPV) and negative predictive value (NPV) were calculated to evaluate models' performance. Data processing and analysis were performed using R (version 4.3.0; 2023–04–21), GraphPad Prism (version 9.0.0), and Storm Statistical Platform (www.medsta.cn/software).

Machine learning model construction

The lightweight variant of the gradient boosting decision tree algorithm (LightGBM), which Microsoft developed, is an iterative, decision tree-based algorithm. The LightGBM algorithm employs a histogram-based traversal method (Histogram) to navigate through each segmentation point, thereby reducing memory consumption and the complexity of data separation, facilitating the acceleration of the training process. Furthermore, LightGBM uses an exclusive feature bundling algorithm, consolidating multiple mutually exclusive features into a single feature. Additionally, it utilized a leaf-wise strategy to grow trees and identified the leaf with the greatest gain in variance for a split. Consequently, LightGBM exhibits superior training speeds, reduced memory usage, and enhanced accuracy. Therefore, the LightGBM model was selected as the primary method to establish a MDD recognition model.

Hyperparameters were adjustable parameters that regulated the process of model training and exerted a significant impact on the performance of the model. A threefold cross-validation strategy was employed to identify the optimal set of parameters. This involved an iterative training process whereby the model was trained using multiple sets of hyperparameters. The optimal parameter set was selected based on evaluating the area under the receiver operating curve (AUC) metric. The hyperparameter searches and the specific values are presented in Additional File 1: Table 1.

We constructed a LightGBM model for each task based on the 78 parameters and demographics: age and gender. A threefold cross-validation scheme was employed. The original data sets were randomly divided into three equal-sized subsets. Of the three subsets, a single one was retained as the validation data to test the model's predictive ability. The remaining two subsets were used as training samples to estimate model parameters. The results were presented across the folds by calculating the mean, SD, and 95% confidence interval. Shapley additive explanation (SHAP) values were calculated to demonstrate which variables were more important. This analysis was performed on Python 3.7.

To compare the model performance of different stimuli tasks, one-way analysis of variance (ANOVA) tests were used to measure continuous variables between three or more groups, with Tukey's multiple comparisons test as the post hoc analysis.

Results

There were no significant differences between MDD and control groups with regard to age, gender, and years of education. The mean (SD) of HAMD-17 scores in the MDD group was 21.07 (6.96) (Additional File 1: Table 2). Additionally, tests of variance, univariate logistic regression, stratification analyses, multi-factor logistic regression model construction, and LightGBM models were performed for each handwriting task. The metrics, including AUC, accuracy, sensitivity, specificity, PPV, and NPV, were used to evaluate the models. We described task 7 in the main text in detail because it is where the LightGBM model performed better (Additional File 1: Table 3). The detailed results on the other tasks are provided in Additional File 1.

In task 7, 34 variables were significantly different between MDD and control groups after Bonferroni correction ($p < 0.05/78$) (Additional File 2). In the univariate logistics analysis, all 34 features in task 7 were predictors of MDD (Additional File 3). Significant variables (Bonferroni correction $p < 0.05$ and $|\beta| > 0.01$) were selected for multivariate analysis, resulting in six features, including task 7 *X* coordinate quantile lower, task 7 *X* coordinate max, task 7 *X* coordinate inter-quantile range, task 7 angle *X* std, task 7 angle *Y* inter-quantile range, and task 7 speed median as the MDD predictors (Table 2). In gender-stratified multivariate analysis, all the variables in task 7 were MDD predictors (Additional File 1: Table 4). The logistic regression model with the six features acquired a validation AUC of mean 0.82 (SD, 0.1), the highest in the three-fold cross-validation (Additional File 1: Table 5 and Fig. S2). In the MDD group, task 7 *X* coordinate quantile lower (mean (SD), 142.75 (34.56) versus 158.04 (28.08), $p_{adjusted} < 0.05$), task 7 *X* coordinate max (mean (SD), 189.82 (49.34) versus 215.73(37.78), $p_{adjusted} < 0.01$), and task 7 *X* coordinate inter-quantile range (mean (SD), 76.28 (28.53) versus 94.09 (22.40), $p_{adjusted} < 0.01$) were lower compared to the control group. However, task 7 angle *X* std (mean (SD), 5.03 (1.55) versus 4.01(1.15), $p_{adjusted} < 0.01$), task 7 angle *Y* inter-quantile range (mean (SD), 4.82 (1.96) versus 3.78 (1.22),

Table 2 Multivariable logistic analysis of various variables in task 7 with age as covariate

Variables	β	S.E	Z	P	OR (95% CI)
Intercept	−9.58	3.60	−2.66	0.008	0.01 (0.00 ~ 0.08)
Gender					
female					1.00 (Reference)
male	−0.16	0.41	−0.38	0.705	0.86 (0.38 ~ 1.92)
Age	0.06	0.04	1.45	0.146	1.07 (0.98 ~ 1.16)
Task 7 *X* coordinate quantile lower	0.13	0.04	3.17	0.002	1.14 (1.05 ~ 1.23)
Task 7 *X* coordinate Max	−0.08	0.03	−2.48	0.013	0.92 (0.86 ~ 0.98)
Task 7 *X* coordinate range	0.03	0.03	1.03	0.302	1.03 (0.97 ~ 1.10)
Task 7 *X* coordinate inter-quantile range	0.19	0.06	3.11	0.002	1.20 (1.07 ~ 1.36)
Task 7 *X* coordinate Std	0.08	0.15	0.55	0.584	1.09 (0.80 ~ 1.47)
Task 7 *X* coordinate discrete coefficient	0.14	0.07	1.91	0.056	1.15 (1.00 ~ 1.33)
Task 7 *Y* coordinate discrete coefficient	−0.03	0.02	−1.94	0.052	0.97 (0.94 ~ 1.00)
Task 7 angle *X* inter-quantile range	0.34	0.18	1.84	0.066	1.40 (0.98 ~ 2.01)
Task 7 angle *X* Std	2.17	0.79	2.74	0.006	8.80 (1.86 ~ 41.69)
Task 7 angle *X* Var	−0.12	0.07	−1.70	0.090	0.89 (0.77 ~ 1.02)
Task 7 angle *Y* inter-quantile range	−0.63	0.24	−2.61	0.009	0.53 (0.33 ~ 0.86)
Task 7 angle *Y* Std	−1.78	1.18	−1.51	0.131	0.17 (0.02 ~ 1.70)
Task 7 angle *Y* Var	0.13	0.15	0.84	0.404	1.13 (0.84 ~ 1.53)
Task 7 speed median	0.07	0.02	3.51	< 0.001	1.08 (1.03 ~ 1.12)

OR odds ratio, *CI* confidence interval

$p_{adjusted} < 0.01$), and task 7 speed median (mean (SD), 56.02 (16.47) versus 47.75 (10.06), $p_{adjusted} < 0.01$) were higher (Additional File 2).

Performance of LightGBM model

The LightGBM model was used to construct a more effective depression recognition model. The hyperparameters were adjusted using a grid search, and the final specific model parameters are provided in Additional File 1: Table 1. Recognition models were initially constructed using all the handwriting features from each task. The threefold cross-validation results revealed that the model based on task 7 exhibited an AUC value of mean 0.90 (SD, 0.01) and an accuracy of mean 0.82 (SD, 0.01) (Additional File 1: Table 3). The SHAP analysis revealed that the three most influential features were task 7 speed discrete coefficient, task 7 pressure min, and task 7 angle *X* std (Fig. 1) . The performance of the LightGBM model surpasses that of the logistic regression model (Additional File 1: Table 5). Furthermore, the handwriting features of the 21 tasks were incorporated into the model. The results of the three-fold cross-validation exhibited an AUC value of mean 0.90 (SD, 0.01) and an accuracy value of mean 0.83 (SD, 0.02) (Additional File 1: Table 3). SHAP values indicated that the three most influential features were task 7 speed discrete coefficient, task 19 angle *X* sum, and task 8 pressure min (Fig. 2).

Comparison of model results across different stimuli tasks

Among the 21 tasks, we assigned positive, neutral, and negative tasks. Tasks 6, 7, and 8 were all copying tasks in negative, neutral, and positive; tasks 9–17 were question and answer, 9–11 were negative, 12–14 were neutral, and 15–17 were positive. We performed an ANOVA on model AUC values to determine whether a particular task model performed better. The results demonstrated a statistically significant difference among the three stimuli groups in the copying task ($F = 25.69$, $p = 0.001$, Additional File 1: Table 6–7). Subsequent post hoc analyses revealed that the neutral task exhibited superior performance compared to the negative task ($p < 0.01$); however, there was no significant difference between the neutral and positive tasks ($p > 0.05$) (Fig. 3A). In the question and answer tasks, a significant difference was observed among the three stimuli groups ($F = 31.45$, $p < 0.0001$, Additional File 1: Table 8–9). The negative task was found to be superior to the positive task ($p < 0.0001$) and the neutral task ($p < 0.01$), while the neutral task was demonstrated to be more effective than the positive task ($p < 0.001$) (Fig. 3B).

Review of the main findings in all tasks

In multiple tasks, logistic regression analysis indicated that angle *X* std and angle *Y* std were the predictors of MDD (Additional File 1: Figs. S3–4). The two variables were greater in the patient group (Additional File 2). This

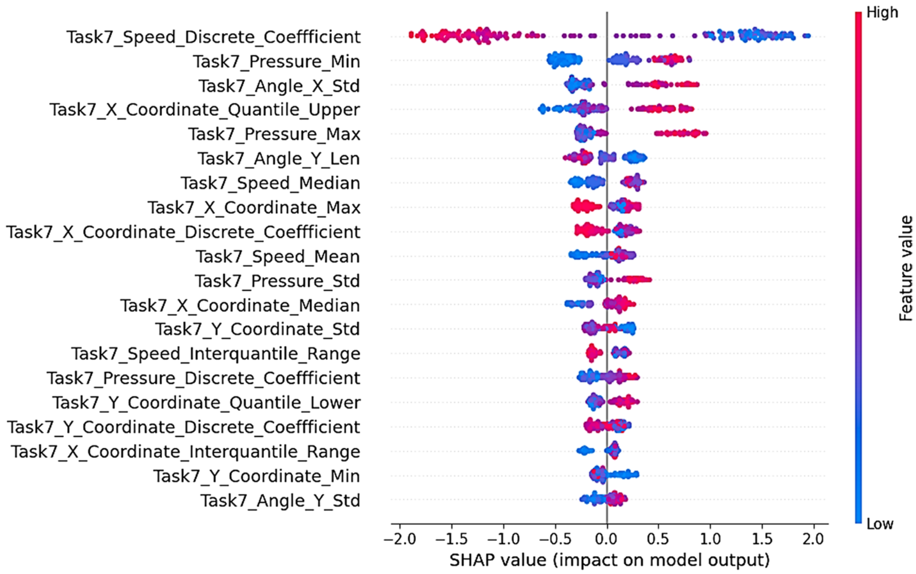

Fig. 1 The LightGBM model was constructed using the 78 handwriting variables from task 7. Variable importance was subsequently calculated using the SHAP method, with class 0 representing the control group and class 1 representing the patient group. The results demonstrated that the three most influential indicators on the model outcomes were task 7 speed discrete coefficient, task 7 pressure min, and task 7 angle *X* Std

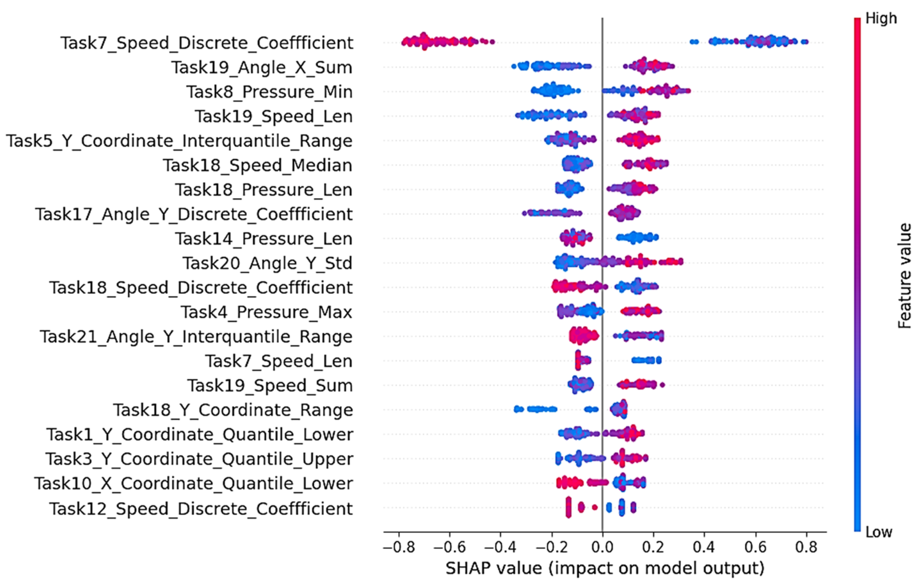

Fig. 2 The LightGBM model was constructed using the 78 × 21 = 1638 handwriting variables from all tasks. Variable importance was subsequently calculated using the SHAP method, with class 0 representing the control group and class 1 representing the patient group. The results demonstrated that the three most influential indicators on the model outcomes were task 7 speed discrete coefficient, task 19 angle *X* sum, and task 8 pressure min

suggested that individuals with MDD exhibited more jitteriness in handwriting. Regarding handwriting pressure, the multivariate logistic regression results of various tasks indicated that the pressure min was the MDD predictor (Additional File 1: Figs. S3–4), with a greater value observed in the patient group (Additional File 2). This suggested that patients with MDD experienced elevated levels of handwriting pressure. Regarding handwriting speed, the multivariate logistic regression results of various tasks indicated that the speed median was an MDD predictor (Additional File 1: Figs. S3–4). The patient group's median handwriting speed was faster than the

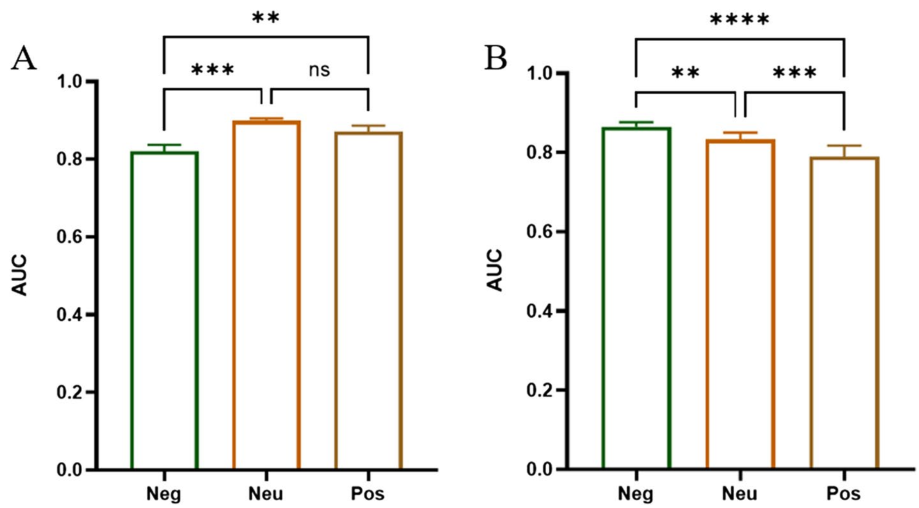

Fig. 3 Results of ANOVA analysis. **A** Bar plot of AUC comparison between the three stimuli groups of the copying tasks. **B** Bar plot of AUC comparison between the three stimuli groups of question and answer tasks. Neg, negative; Neu, neutral; Pos, Positive. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$

control group's (Additional File 2). Furthermore, SHAP scores indicated that task 7 speed discrete coefficient was the most important variable in the LightGBM model for discriminating MDD (Figs. 1 and 2).

Discussion

This study used a novel methodology to quantify human handwriting activity and constructed an MDD recognition model based on handwriting characteristics. Our findings confirmed that patients with MDD exhibited altered handwriting patterns and validated the hypothesis that handwriting characteristics can be used to identify these patients. This finding further validates the Multiple Coding Theory that a significant association exists between physical activity, behavior, and a person's emotional state. Besides, the biological mechanism behind this difference can be associated with ATP levels. However, the experimental results indicate that the handwriting pressure and speed of patients with MDD were higher than those of healthy individuals. Conversely, biological studies suggest that ATP levels are decreased in patients with MDD. However, these results are inconsistent. It can be hypothesized that handwriting is not analogous to the movement of large joints, which require a significant amount of energy from the body. Consequently, it can be plausible that the reduction in ATP levels does not directly result in a decline in handwriting pressure and speed. Furthermore, handwriting is subject to regulation by several different brain regions and is an activity that is inherently complex from an intellectual perspective. A potential avenue for elucidating the mechanism by which brain networks regulate handwriting activity can provide insights into the increased handwriting speed and stress observed in patients with MDD.

Furthermore, ANOVA of the AUC values of logistic regression models constructed for the positive, neutral, and negative tasks in question-and-answer tasks revealed that the negative tasks models exhibited the highest degree of discrimination, followed by the neutral and the positive tasks, indicating the lowest degree of discrimination. These findings are in accordance with the results of previous psychological experiments. However, no statistically significant difference was observed in the transcription task, which can be attributed to the fact that the transcription task is less emotionally stimulating for the subjects than the question-and-answer task. Previous studies suggest that depression is associated with an increased focus on negative information [42, 43]. According to the hypothesis [44], individuals with depression interpret neutral and ambiguous stimuli through themes of loss, failure, worthlessness, and rejection. Stress can activate dysfunctional schemas, resulting in negative automatic thoughts centered on the cognitive triad: pessimistic views of the self, world, and future. This sensitivity to negative stimuli is particularly increased in those with elevated depression [45–48] and suicide risk [42, 49–52]. Multiple studies identify online language patterns that are indicative of depression-related distorted thinking [53–55]. Negative materials thus effectively simulate high-stress situations, facilitating accurate assessments of emotional and cognitive states. This kind of emotional reactivity is associated with the development and maintenance of psychopathological conditions [56]. Therefore, using negative materials should enhance the realism of simulated stress scenarios, improving data quality and the model's accuracy in identifying depression.

A preliminary review of the existing studies suggests that handwriting assays have been analyzed for various other disorders, including mild cognitive impairment and Parkinson's disease [57–59]. Future research could be promising if it can determine whether the electronic handwriting parameters identified are unique to MDD. The handwriting variables tested exhibit sensitivity; however, their specificity is limited compared to healthy controls. A more informative approach would be identifying pathognomonic handwriting features in MDD and other pathologies.

In clinical practice, the application of electronic handwriting models in this study can facilitate the identification of depression, thereby assisting psychiatrists in the diagnosis process. In Chinese community hospitals in particular, the lack of specialist psychiatrists often prevents depression from being diagnosed. This method can be utilized initially for preliminary screening during a patient's visit to the clinic, thereby obviating the doctor's requirement to employ a structured interview. In summary, this method facilitates rapid and accurate diagnosis, ensuring effective referral and treatment.

Limitations

This study has certain limitations. The study population is relatively fixed and predominantly comprised of adolescents, and the extent to which the findings can be generalized to other age groups requires further investigation. Using convenience sampling methods constitutes a bias to external validity. In this experiment, the group of patients with MDD are all in depressive episodes, and some of the subjects are on medication. However, further investigation is required to determine whether the medication affects handwriting. Considering reports in the studies indicating that antidepressant drugs can precipitate tremors in patients, we meticulously observed the hands of the subjects before the experiment. We requested detailed information regarding the presence of tremors in their daily lives. Only two subjects reported occasional tremors; however, they did not affect their handwriting activities. The patient group was not followed up further, and it would be advantageous to investigate whether the handwriting characteristics of the patients would change following improvements in their clinical symptoms as a result of medication or psychotherapy. In this study, the handwriting content is limited to Chinese characters and Arabic numerals. However, it would be beneficial to investigate whether the same change was observed in English handwriting to enable further generalization of the model.

Conclusions

This study demonstrates that depressed patients exhibit characteristic alterations in their handwriting and proposes a cost-effective, rapid, and valid model for identifying MDD. This finding offers further insight into the objective indicators of MDD and provides a robust foundation for the future development of multimodal recognition models.

Abbreviations

AUC	Area under the receiver operating characteristic curve
GBDT	Gradient boosting decision tree
LEN	Length value
MDD	Major depressive disorder
MEAN	Average value
MEDIAN	Median value
MAX	Maximum value
MIN	Minimum value
NPV	Negative predictive value
PPV	Positive predictive value
SD	Standard deviation
SHAP	Shapley additive explanation
STD	Standard deviation
SUM	Summation value
VAR	Variance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04101-2.

Additional File 1: Table 1-119. Table 1: Hyperparameter space search for LightGBM. Table 2: Demographic Characteristics data for Patients with MDD vs Healthy Controls. Table3: Performance of LightGBM model based on each task. Table 4: Gender-Stratified Univariable Logistic analysis of predictors in Task7 with age as covariate. Table 5: The best performance in the 3-fold cross-validation of Logistics model based on Task 7. Table 6-9: ANOVA analysis and multiple comparison of models' AUC in task 6-17. Table 10-119: Multivariable Logistic analysis results, Gender-Stratified Univariable Logistic analysis results, and 3-fold cross-validation model metrics of each task. Figures S1-S64. Figure S1: A handwriting example of one subject in Task 7. Figure S2: ROC curve of the 3-fold cross-validation of logistics model in task 7. Figure S3: Count of each significant variable in multi-variable logistics regression across all tasks. Figure S4: Count of each significant variable in univariable logistics regression after Bonferroni correction across all tasks. Figure S5-S64: ROC curve of the 3-fold cross-validation of logistics model in each task except task 7.

Additional file 2: T-test results of handwriting parameters in all tasks.

Additional file 3: Univariable logistics regression results in all tasks.

Acknowledgements

We appreciate the help from the Baiyun Mental Health Hospital, Guangzhou, as some MDD patients were recruited from the hospital.

Authors' contributions

CL had full access to all data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: CL, KZ, and JZ. Acquisition, analysis, or interpretation of data: CL, QL, SH, WC, YL and XZ. Drafting of the manuscript: CL, QL, and KZ. Critical review of the manuscript for important intellectual content: All authors. Statistical analysis: CL, QL. Obtained funding: JZ. Administrative, technical, or material support: CL, QL, and KZ. Supervision: JZ. All authors read and approved the final manuscript.

Funding
This study was supported by the National Natural Science Foundation of China (Grant No. 72174082 and Grant No. 82373695), the Natural Science Foundation of Guangdong Province (Grant No. 2023 A1515011825), and the Guangdong-Hong Kong Joint Laboratory for Psychiatric Disorders (2023B1212120004).

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Health Research Ethics Committee of the Southern Medical University (NFYKDX003).

Consent for publication
All authors agreed with the publication of this manuscript.

Competing interests
The authors declare no competing interests.

Received: 21 December 2024 Accepted: 24 April 2025
Published online: 13 May 2025

References

1. Malhi GS, Mann JJ. Depression Lancet. 2018;392(10161):2299–312.
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1789-858.https://doi.org/10.1016/S0140-6736(18)32279-7. Erratum in: Lancet. 2019;393(10190):e44. https://doi.org/10.1016/S0140-6736(18)32279-7.
3. Chen R, An J, Ou J. Suicidal behaviour among children and adolescents in China. Lancet Child Adolesc Health. 2018;2(8):551–3.
4. Hawton K, Saunders KEA, O'Connor RC. Self-harm and suicide in adolescents. Lancet. 2012;379(9834):2373–82.
5. Gururajan A, Clarke G, Dinan TG, Cryan JF. Molecular biomarkers of depression. Neurosci Biobehav Rev. 2016;64:101–33.
6. Klooster D, Voetterl H, Baeken C, Arns M. Evaluating Robustness of Brain Stimulation Biomarkers for Depression: A Systematic Review of Magnetic Resonance Imaging and Electroencephalography Studies. Biol Psychiatry. 2024;95(6):553–63.
7. Binnewies J, Nawijn L, van Tol M-J, van der Wee NJA, Veltman DJ, Penninx BWJH. Associations between depression, lifestyle and brain structure: A longitudinal MRI study. Neuroimage. 2021;231: 117834.
8. Xu M, Li X, Teng T, Huang Y, Liu M, Long Y, et al. Reconfiguration of Structural and Functional Connectivity Coupling in Patient Subgroups With Adolescent Depression. JAMA Netw Open. 2024;7(3): e241933.
9. Cavaleri D, Bartoli F, Capogrosso CA, Guzzi P, Moretti F, Riboldi I, et al. Blood concentrations of neopterin and biopterin in subjects with depression: A systematic review and meta-analysis. Prog Neuropsychopharmacol Biol Psychiatry. 2023;120: 110633.
10. Carotenuto A, Valsasina P, Preziosa P, Mistri D, Filippi M, Rocca MA. Monoaminergic network abnormalities: a marker for multiple sclerosis-related fatigue and depression. J Neurol Neurosurg Psychiatry. 2023;94(2):94–101. https://doi.org/10.1136/jnnp-2022-330109.
11. Yan Z, Rein B. Mechanisms of synaptic transmission dysregulation in the prefrontal cortex: pathophysiological implications. Mol Psychiatry. 2022;27(1):445–65.
12. Jespersen A, Yilmaz Z, Vilhjálmsson BJ. Harnessing the Power of Population Cohorts to Study the Relationship Between Endocrine-Metabolic Disorders and Depression. Am J Psychiatry. 2022;179(11):788–90.
13. Snigdha S, Ha K, Tsai P, Dinan TG, Bartos JD, Shahid M. Probiotics: Potential novel therapeutics for microbiota-gut-brain axis dysfunction across gender and lifespan. Pharmacol Ther. 2022;231: 107978.
14. Rasmussen M-LH, Poulsen GJ, Videbech P, Wohlfahrt J, Melbye M. Endocrine disease history and the risk of postpartum depression. Br J Psychiatry. 2023;222(3):119–24.
15. Winter NR, Blanke J, Leenings R, Ernsting J, Fisch L, Sarink K, et al. A Systematic Evaluation of Machine Learning-Based Biomarkers for Major Depressive Disorder. JAMA Psychiat. 2024;81(4):386–95.
16. Wasserzug Y, Degani Y, Bar-Shaked M, Binyamin M, Klein A, Hershko S, et al. Development and validation of a machine learning-based vocal predictive model for major depressive disorder. J Affect Disord. 2023;325:627–32.
17. Schultebraucks K, Yadav V, Shalev AY, Bonanno GA, Galatzer-Levy IR. Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. Psychol Med. 2022;52(5):957–67.
18. Kim AY, Jang EH, Lee S-H, Choi K-Y, Park JG, Shin H-C. Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach. J Med Internet Res. 2023;25: e34474.
19. König A, Tröger J, Mallick E, Mina M, Linz N, Wagnon C, et al. Detecting subtle signs of depression with automated speech analysis in a non-clinical sample. BMC Psychiatry. 2022;22(1):830.
20. Ettore E, Müller P, Hinze J, Riemenschneider M, Benoit M, Giordana B, et al. Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review. JMIR Ment Health. 2023;10: e37225.
21. Xie F, Zhou L, Hu Q, Zeng L, Wei Y, Tang X, et al. Cardiovascular variations in patients with major depressive disorder versus bipolar disorder. J Affect Disord. 2023;341:219–27.
22. Li C, Yang Y, Wang W, Li H, Mai Y, Zhao J. Measurement of differential activation by heart-rate-variability for youth MDD discrimination. J Affect Disord. 2025;376:169–76.
23. Li X, Zhang X, Zhu J, Mao W, Sun S, Wang Z, et al. Depression recognition using machine learning methods with different feature generation strategies. Artif Intell Med. 2019;99: 101696.
24. Zhou Y, Han W, Yao X, Xue J, Li Z, Li Y. Developing a machine learning model for detecting depression, anxiety, and apathy in older adults with mild cognitive impairment using speech and facial expressions: A cross-sectional observational study. Int J Nurs Stud. 2023;146: 104562.
25. Yasin S, Othmani A, Raza I, Hussain SA. Machine learning based approaches for clinical and non-clinical depression recognition and depression relapse prediction using audiovisual and EEG modalities: A comprehensive review. Comput Biol Med. 2023;159: 106741.
26. Liberg B, Rahm C. The functional anatomy of psychomotor disturbances in major depressive disorder. Front Psychiatry. 2015;6:34.
27. Choi KW, Chen C-Y, Stein MB, Klimentidis YC, Wang M-J, Koenen KC, et al. Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. JAMA Psychiat. 2019;76(4):399–408.
28. Corfield EC, Martin NG, Nyholt DR. Co-occurrence and symptomatology of fatigue and depression. Compr Psychiatry. 2016;71:1–10. https://doi.org/10.1016/j.comppsych.2016.08.004.
29. Deng M-G, Liu F, Liang Y, Wang K, Nie J-Q, Liu J. Association between frailty and depression: a bidirectional Mendelian randomization study. Sci Adv. 2023;9(38):3902.
30. Lin S, Huang L, Luo Z-C, Li X, Jin S-Y, Du Z-J, et al. The ATP Level in the Medial Prefrontal Cortex Regulates Depressive-like Behavior via the Medial Prefrontal Cortex-Lateral Habenula Pathway. Biol Psychiatry. 2022;92(3):179–92.
31. Cao X, Li L-P, Wang Q, Wu Q, Hu H-H, Zhang M, et al. Astrocyte-derived ATP modulates depressive-like behaviors. Nat Med. 2013;19(6):773–7.
32. Bucci W. Psychoanalysis and cognitive science: A multiple code theory. New York, NY, US: Guilford Press; 1997. xiv, 362-xiv, p.
33. Pennebaker JW, Francis ME. Cognitive, emotional, and language processes in disclosure. Cognition and Emotion. 1996;10(6):601–26. https://doi.org/10.1080/026999396380079.
34. Damasio AR. Emotion and the Human Brain. Ann N Y Acad Sci. 2001;935(1):101–6.
35. Mariani R, Di Trani M, Negri A, Tambelli R. Linguistic analysis of autobiographical narratives in unipolar and bipolar mood disorders in light of multiple code theory. J Affect Disord. 2020;273:24–31.
36. Tucha O, Aschenbrenner S, Eichhammer P, Putzhammer A, Sartor H, Klein HE, et al. The impact of tricyclic antidepressants and selective serotonin

re-uptake inhibitors on handwriting movements of patients with depression. Psychopharmacology. 2002;159(2):211–5.
37. Mergl R, Juckel G, Rihl J, Henkel V, Karner M, Tigges P, et al. Kinematical analysis of handwriting movements in depressed patients. Acta Psychiatr Scand. 2004;109(5):383–91.
38. Rosenblum S, Werner P, Dekel T, Gurevitz I, Heinik J. Handwriting process variables among elderly people with mild Major Depressive Disorder: a preliminary study. Aging Clin Exp Res. 2010;22(2):141–7.
39. Heinik J, Werner P, Dekel T, Gurevitz I, Rosenblum S. Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study. Int Psychogeriatr. 2010;22(3):479–88.
40. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6(4):278–96.
41. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59(Suppl 20):22-57.
42. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol. 2010;6:285–312.
43. Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. Depress Anxiety. 2010;27(12):1135–42.
44. Beck AT. A 60-Year Evolution of Cognitive Theory and Therapy. Perspect Psychol Sci. 2019;14(1):16–20.
45. Joormann J, Quinn ME. Cognitive processes and emotion regulation in depression. Depress Anxiety. 2014;31(4):308–15.
46. Joormann J, Stanton CH. Examining emotion regulation in depression: A review and future directions. Behav Res Ther. 2016;86:35–49.
47. Iqbal N, Dar KA. Negative affectivity, depression, and anxiety: Does rumination mediate the links? J Affect Disord. 2015;181:18–23.
48. Bailen NH, Wu H, Thompson RJ. Meta-emotions in daily life: Associations with emotional awareness and depression. Emotion. 2019;19(5):776–87.
49. Hatkevich C, Penner F, Sharp C. Difficulties in emotion regulation and suicide ideation and attempt in adolescent inpatients. Psychiatry Res. 2019;271:230–8.
50. Mirkovic B, Delvenne V, Robin M, Pham-Scottez A, Corcos M, Speranza M. Borderline personality disorder and adolescent suicide attempt: the mediating role of emotional dysregulation. BMC Psychiatry. 2021;21(1):393.
51. Naifeh JA, Ursano RJ, Stein MB, Mash HBH, Aliaga PA, Fullerton CS, et al. Prospective associations of emotion reactivity and risk behaviors with suicide attempts in US Army soldiers. Psychol Med. 2023;53(13):6124–31.
52. Douglas VJ, Kwan MY, Gordon K. The roles of weight stigma, emotion dysregulation, and eating pathology in suicide risk. Body Image. 2021;38:162–70.
53. Bathina KC, Ten Thij M, Lorenzo-Luaces L, Rutter LA, Bollen J. Individuals with depression express more distorted thinking on social media. Nat Hum Behav. 2021;5(4):458–66.
54. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoțiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A. 2018;115(44):11203–8.
55. Dozois DJA. Depressive cognition on Twitter. Nat Hum Behav. 2021;5(4):414–5.
56. Nock MK, Wedig MM, Holmberg EB, Hooley JM. The emotion reactivity scale: development, evaluation, and relation to self-injurious thoughts and behaviors. Behav Ther. 2008;39(2):107–16.
57. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med. 2016;67:39–46.
58. Kawa J, Bednorz A, Stępień P, Derejczyk J, Bugdol M. Spatial and dynamical handwriting analysis in mild cognitive impairment. Comput Biol Med. 2017;82:21–8.
59. Rosenblum S, Meyer S, Richardson A, Hassin-Baer S. Patients' Self-Report and Handwriting Performance Features as Indicators for Suspected Mild Cognitive Impairment in Parkinson's Disease. Sensors (Basel). 2022;22(2):569. https://doi.org/10.3390/s22020569.

Publisher's Note