

Research Article

Depression Predictive Model Using In-Context Learning Based on HRV with PPG Derived Validity Label

Hang Li¹ , Jing Li² , Xin Zhong³, Guo Chen³, Ruiqi Peng⁵ , Liang Zhang¹ ,
Juqiang Han⁴ , Xiaomin Luo^{2,*} 

¹Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen, China

²Jintz Biomedical Research Laboratory, Foshan, China

³Psychological Rehabilitation Department, Dongguan Rehabilitation Hospital, Dongguan, China

⁴Outpatient Department, the 7th Medical Center of PLA General Hospital, Beijing, China

⁵School of Artificial Intelligence, South China Normal University, Guangzhou, China

Abstract

Background: Traditional diagnostic approaches for major depression disorder (MDD) or clinical depression rely on subjective assessment of clinical symptoms while heart rate variability (HRV) metrics provide an objective alternative to support clinical assessments and facilitate early depression detection. However, the imperceptibility of non-stationarity and unpredictability in noticing a factor for its HRV outcome highlight the challenges in modelling of predictive AI. **Methods:** In this study, totally 139 participants were recruited including 40 patients and 99 healthy controls. Only 28 of the 40 depression patients and 34 of the 99 healthy controls were enrolled for HRV data collection according to inclusion criteria. Our experiment provided evidence for evaluation of the validation method using a photoplethysmography (PPG) derived parameter representing beat-to-beat stress-induced vascular response in terms of labelling performance and applicability. **Results:** The results demonstrated the link between depression and the autonomic nervous system (ANS) measured using HRV both in statistical analysis and AI-driven classification, as seen in the GPT-4-based LLM outperformed baseline models across multiple data sets. The validity labeling contributed significantly to model performance and robustness, especially in small-sample scenarios. Although small sample size was used in HRV-based depression prediction training via a large language model (LLM) with in-context learning (ICL), the performance was definitely improved with validity labeling activated compared to labeling disabled. **Conclusions:** Through comparison of observational accuracy in predictive models, the reliability of HRV recordings is crucial for improving AI-driven depression prediction and aligning AI analysis with the expectations on physiological and psychological effects. Among factors that could cause HRV value to change in unexpected ways, stationarity is a prerequisite for short-term HRV (ST-HRV), thus validation strategy, a labeling method capable of identifying and rejecting recordings of false signals, is necessarily needed.

Keywords

Heart Rate Variability (HRV), Photoplethysmography (PPG), Stress-induced Vascular Response Index (sVRI), In-context Learning (ICL), Large Language Model (LLM), AI-driven Depression Prediction

*Corresponding author: luo.xiaomin@139.com (Xiaomin Luo)

Received: 25 March 2025; **Accepted:** 14 April 2025; **Published:** 11 June 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Depression, or depressive disorder, is a globally prevalent mental disorder affecting individuals across all age groups, recognized as a significant public health concern [1]. Depression is a serious medical illness with great burdens for the affected individuals and public health care systems. Major depression disorder (MDD), also called clinical depression, is a more severe form of depression, diagnosed with specific criteria from the Diagnostic and Statistical Manual of Mental Disorders, 5th ed., criteria (DSM-5) [2], with symptoms like persistent low mood, anhedonia, and significant impairment in daily functioning. The onset and progression of depression are influenced by both biological and environmental factors. Genetic predisposition and neurochemical imbalances contribute to its pathophysiology [3], while external stressors such as prolonged familial conflicts or traumatic life events further elevate susceptibility [4]. The complexity of these contributing factors complicates both the diagnosis and treatment of depression.

Traditional diagnostic approaches for MDD rely on subjective assessment of clinical symptoms, such as interviews and self-reported questionnaires like the Patient Health Questionnaire (PHQ) [5, 6]. However, these assessments are inherently subjective, as their accuracy depends on both the patient's ability to articulate symptoms and the clinician's interpretation [7, 8]. Reliance upon clinical assessments and patient interviews for diagnosing MDD is frequently associated with misdiagnosis and suboptimal treatment outcomes. As such, there is increasing interest in objective methods to determine depressive symptoms. Heart rate variability (HRV) metrics, as a noninvasive and objective physiological indicator of autonomic nervous system function, are technically sophisticated to diagnostic tests in psychiatry.

HRV, as a physiological monitoring of autonomic nervous system, provides an objective alternative to support clinical assessments and facilitate early depression detection. Many studies have identified a link between depression and the autonomic nervous system (ANS) measured using HRV. A compelling body of research indicates a reduced HRV indexed by lower values of SDNN, RMSSD and HF power and increased values of LF power for patients with depression in comparison to healthy controls [8]. Thus, as simplified method, some HRV parameters have been used to separate patients with MDD and healthy controls [9-11].

According to the neurovisceral integration model, lower HRV reflects impaired regulatory control, linking autonomic dysregulation to symptoms of anxiety and depression [13]. Given its affordability and ease of measurement, HRV has been widely explored as a potential biomarker for depression assessment [14, 15]. However, despite repeated reports of HRV reduction in depressed patients, the causal role of pharmacological antidepressant treatment in autonomic dysfunction remains uncertain, highlighting the complexity of the relationship between psychological symptoms and physiological states.

Recent studies have also suggested that restlessness is a key factor contributing to lower HRV, emphasizing that stationarity is a prerequisite for short-term HRV (ST-HRV) assessments [16-18]. As the HRV signal has been used as an indicative measure to estimate the level of stress, anxiety and depression, the advantages of artificial intelligence (AI)-based learning systems in data processing and logical reasoning capabilities are widely recognized. In terms of predictive capabilities, assessment of factors with impact on the accuracy must be taken into consideration. Thus, the evaluation of effective techniques, the choice of validation method in particular, should be the top priority in enhancing performance.

To address these challenges, this study investigates the associations between symptoms of depression and HRV in clinically supervised outpatients as well as multiple comparisons with normal controls. To enhance data reliability, we applied a validity labeling method based on the standard deviation of stress induced vascular response index (SDsVRI), a photoplethysmography (PPG) derived parameter reflecting beat-to-beat vascular responsive fluctuations, for observation of stationarity throughout HRV testing. This method ensures that the high-quality HRV recordings could be identified from all data obtained and labelled reliable for further analysis.

In order to further evaluate the performance of validation method, we employ in-context learning (ICL) with a large language model (LLM) [19] for depression prediction based on HRV metrics. It is assumed that the data used to train the model and the testing data to make predictions are free of errors. But rarely a data set is absolutely clean, thus a threshold of non-stationarity has been used to define the data of poor reliability in the process of HRV data extraction. By comparing model performance before and after labeling, we assess the effectiveness of the PPG-derived validity labels both in statistical analysis and AI-driven classification. Notably, this study represents the first attempt to integrate LLMs based on the electrocardiogram (ECG) derived HRV metrics incorporated with the validation method using a PPG based label. As a candidate method for hemodynamic sensing, PPG has been used in clinical settings for quantitative observation and analysis of changes in peripheral arterial function [20-26]. The experimental results provided evidence for evaluation of the labelling performance and applicability through comparison of observational accuracy in predictive models.

2. Subjects and Methods

2.1. Study Subjects

The current study was performed at the psychological rehabilitation department of Dongguan Rehabilitation Hospital, Guangdong, China. A total of 139 participants were recruited. Among them, the depression subjects were recruited from the outpatient attendees (n=40) who had been diagnosed with

depression in their first and second interview, while the healthy control subjects ($n=99$) were recruited from the students of Guangdong Vocational and Technical College. The diagnosis should be confirmed using DSM-5 [2]. The study was approved by the Institutional Review Board of Dongguan Rehabilitation Hospital (Approval No: EC-20201211-1003), and all subjects had signed their written informed consent.

The inclusion criteria were: i) youth age group between the ages of 13 and 21 years; ii) Body mass index (BMI) is between 14.5 to 20.7; iii) no pharmacological antidepressant treatment on cardiac autonomic dysregulation. iv) All subjects refrained from caffeine and drugs that could alter the cardiac autonomic function on the day before the tests.

The exclusion criteria were: i) Current or prior history of known heart problems; ii) the use of regular medication; and iii) sports-based intervention or physical training program.

According to the strict inclusion criteria, only 28 of the 40 patients were eventually enrolled and 34 of 99 healthy controls were enrolled from the students in Guangdong Vocational and Technical College.

2.2. Measurement of ECG and PPG

In this study, both ECG and PPG signals were measured using JM2020, a HRV analytical system incorporating both ECG and PPG sensors (JintzMed, Shenzhen, China) at a sampling rate of 500 Hz. The HRV metrics were obtained from ECG electrode and a PPG derived index, stress induced vascular response index (sVRI) [27, 28], was obtained from finger clip sensor. The value of sVRI throughout testing represents the varying beat to beat stressful condition and its trend map exhibits quantitatively a dynamic process in vascular response [29].

ECG measurements were performed with electrodes applied in a three-lead (chest-mounted configuration with one electrode under each clavicle and the third on the lower left rib cage) to extract HRV metrics of all subjects during resting state at a sitting position. In parallel, PPG signals were recorded from the right index finger with the right hand being held at heart level. After the subjects rested for 15 minutes to ensure cardiac autonomic stability, ECG and PPG recordings were conducted simultaneously at a duration of 300 seconds, during which all the subjects were asked to calm down and breathe normally with an averaging period covering at least ≥ 300 RR intervals. Although measurements were performed when all participants were sitting still, validation method embedded in JM2020 as a strategy to identify non-stationary recordings is systematically applied.

2.3. Data Validity Labeling

Statistical analyses were performed using SPSS software (version 22.0; IBM Corp.). Results are presented as the mean \pm SD. Comparisons between groups were performed using t-test. The default threshold of non-stationarity in JM2020 is as the followings:

$SDsVRI \geq 0.026$.

$SDsVRI CV \geq 14$.

Being used in the error estimation procedure, all HRV recordings with both $SDsVRI$ and $SDsVRI CV$ values exceeding the threshold range were labeled as poor reliability. In our paper, we call this procedure as “validity labeling”. Thus, by activating validity labeling, the data is added a label that indicates whether the data is reliable or not.

This study will provide the evidence of the non-stationarity (i.e., poor reliability) impact on the statistical analysis of HRV metrics, as well as on the AI-based prediction performance. We hypothesize that if the validity labeling is disabled, the data will compromise the ability of AI prediction model.

2.4. AI-Driven Prediction Model

Traditional machine learning models (e.g., SVMs, random forests) have been used for depression prediction, which often rely on structured clinical questionnaires, and may lead to potential bias. Deep learning models, which were applied to text, audio, and video data, have shown promise in depression prediction task, while ECG derived HRV incorporated with PPG derived Validity Label remains underexplored. Unlike conventional AI methods, ICL enables LLMs to generalize from limited examples without retraining, making it particularly useful for small clinical datasets.

Here, we introduced an LLM-based prediction model to classify the control and depression groups. Given that the logic inference ability of LLM heavily affects the overall performance, we adopt a typical very large LLM, i.e., the GPT-4-0125 version that has a strong logic inference ability. In this model, we provide a simplified prompt instructing the LLM to estimate depression likelihood based on HRV features, and a fine-tuned prompt incorporating preprocessing steps, feature selection, and stepwise reasoning (Chain-of-Thought approach) in the feature selection module, we applied principal component analysis (PCA) to reduce dimensionality while retaining key physiological variations.

3. Results

3.1. HRV Statistical Analysis Results

Table 1. Descriptive characteristics of two groups.

Characteristic	Included participants (n=62)	
	Control (n=34)	Depression (n=28)
Gender (F/M)	16/18	19/9
BMI	18.86 \pm 2.10	18.87 \pm 2.55
Height	163.88 \pm 5.46	162.29 \pm 4.40
Weight	50.65 \pm 6.16	49.57 \pm 5.95

Table 1 summarizes the descriptive characteristics of the study population.

In Table 1, both depression subjects (16.00 ± 2.45) and control subjects (17.67 ± 3.16) were from the same age group

of 13-21 years. No significant differences were observed in BMI, height, or weight ($p > 0.05$). A chi-square test showed no significant difference in gender distribution between groups ($p > 0.05$), suggesting a similar gender composition.

Table 2. Comparison of HRV metrics (computed over the data in the first phase of the study) between control and depression groups before and after applying the validation labeling.

Characteristic	Validity labeling DISABLED Included participants (n=28)				Validity labeling ACTIVATED Included participants (n=22) Exclusion of both SDsVRI ≥ 0.026 and SDsVRI CV ≥ 14			
	Control (n=12)	Depression (n=16)	t-value	p-value	Control (n=11)	Depression (n=11)	t-value	p-value
SDNN (ms)	72.27 \pm 17.51	83.80 \pm 58.88	-0.741	0.468	71.04 \pm 17.81	50.26 \pm 23.47	2.34	0.03
Heart Rate (bpm)	81.33 \pm 9.62	91.69 \pm 14.49	-2.141	0.042	81.81 \pm 9.94	90.45 \pm 17.24	-1.44	0.17
pNN50 (%)	23.90 \pm 16.98	28.31 \pm 31.98	-0.472	0.642	23.32 \pm 16.68	14.41 \pm 23.28	1.01	0.32
RMSSD (ms)	58.36 \pm 29.55	94.52 \pm 91.44	-1.482	0.155	58.67 \pm 30.79	42.96 \pm 37.68	1.07	0.30
VLF Power (ms ²)	799.71 \pm 552.23	679.94 \pm 627.28	0.526	0.604	729.67 \pm 520.27	397.04 \pm 207.26	1.97	0.07
LF Power (ms ²)	957.51 \pm 556.75	1341.89 \pm 1838.14	-0.698	0.491	920.92 \pm 568.60	631.34 \pm 452.05	1.32	0.20
HF Power (ms ²)	1470.52 \pm 870.40	3586.26 \pm 5319.38	-1.563	0.137	1473.39 \pm 912.83	1084.58 \pm 2218.75	0.54	0.60
Total Power (ms ²)	3227.74 \pm 1693.03	5608.08 \pm 7616.87	-1.211	0.243	3123.96 \pm 1735.18	2112.96 \pm 2554.39	1.09	0.29
LF/HF Ratio	0.91 \pm 0.75	1.21 \pm 1.11	-0.782	0.441	0.91 \pm 0.79	1.61 \pm 1.13	-1.69	0.11
LF norm	2.48 \pm 0.94	2.83 \pm 2.27	-0.501	0.621	2.40 \pm 0.95	1.87 \pm 0.73	1.45	0.16
HF norm	8.08 \pm 4.99	19.99 \pm 29.32	-1.594	0.13	8.17 \pm 5.23	6.21 \pm 12.45	0.48	0.64
LF/HF norm	0.51 \pm 0.45	0.79 \pm 0.84	-1.056	0.3	0.51 \pm 0.47	1.10 \pm 0.86	-1.99	0.06
SD1	41.26 \pm 20.89	66.83 \pm 64.66	-1.482	0.155	41.49 \pm 21.90	30.37 \pm 26.64	1.07	0.30
SD2	92.45 \pm 19.76	95.14 \pm 57.55	-0.174	0.864	90.41 \pm 19.35	62.95 \pm 23.95	2.96	0.01
SD1/SD2	0.43 \pm 0.17	0.58 \pm 0.29	-1.707	0.1	0.44 \pm 0.19	0.45 \pm 0.21	-0.11	0.91
SampEn	1.35 \pm 0.36	1.18 \pm 0.42	1.091	0.285	1.34 \pm 0.38	1.21 \pm 0.36	0.79	0.44

Table 3. Comparison of HRV metrics (computed over the total data of the study) between depression and control groups before and after applying the validation labeling.

Characteristic	Validity labeling DISABLED Included participants (n=62)				Validity labeling ACTIVATED Included participants (n=46) Exclusion of both SDsVRI ≥ 0.026 and SDsVRI CV ≥ 14			
	Control (n=34)	Depression (n=28)	t-value	p-value	Control (n=26)	Depression (n=20)	t-value	p-value
SDNN (ms)	61.35 \pm 18.53	72.21 \pm 48.78	-1.11	0.27	63.18 \pm 19.08	56.26 \pm 24.25	1.08	0.29
Heart Rate (bpm)	79.44 \pm 9.36	91.91 \pm 12.99	-4.28	0.01	80.34 \pm 9.25	90.95 \pm 14.97	-2.78	0.01
pNN50 (%)	19.91 \pm 18.10	22.95 \pm 27.32	-0.53	0.60	19.81 \pm 17.00	16.16 \pm 21.44	0.64	0.52
RMSSD (ms)	47.16 \pm 25.26	76.71 \pm 75	-1.99	0.06	48.88 \pm 25.24	51.21 \pm 37.69	-0.25	0.80
VLF Power (ms ²)	780.51 \pm 517.59	576.72 \pm 512.86	1.55	0.13	773.70 \pm 543.23	439.68 \pm 243.93	2.79	0.01

Characteristic	Validity labeling DISABLED Included participants (n=62)				Validity labeling ACTIVATED Included participants (n=46) Exclusion of both SDsVRI ≥ 0.026 and SDsVRI CV ≥ 14			
	Control (n=34)	Depression (n=28)	t-value	p-value	Control (n=26)	Depression (n=20)	t-value	p-value
LF Power (ms ²)	827.57±521.07	1086.51±1508.47	-0.87	0.39	883.99±553.08	744.11±720.61	0.75	0.46
HF Power (ms ²)	1026.82±875.75	2361.41±4241.88	-1.64	0.11	1090.33±911.96	959.53±1681.84	0.34	0.74
Total Power (ms ²)	2634.90±1633.65	4024.63±6059.07	-1.18	0.25	2748.01±1731.55	2143.32±2146.40	1.06	0.30
LF/HF Ratio	1.50±1.27	1.33±0.99	0.59	0.56	1.41±1.17	1.53±0.96	-0.35	0.73
LF norm	2.28±0.88	2.50±1.97	-0.56	0.58	2.35±0.91	2.06±1.16	0.97	0.34
HF norm	5.63±4.80	13.22±23.41	-1.69	0.10	5.97±5.01	5.51±9.45	0.21	0.83
LF/HF norm	0.92±0.97	0.91±0.82	0.02	0.99	0.81±0.77	1.01±0.74	-0.85	0.40
SD1	33.35±17.86	54.24±53.03	-1.99	0.06	34.56±17.85	36.21±26.65	-0.25	0.80
SD2	79.25±22.51	84.24±48.48	-0.50	0.62	81.71±22.89	69.39±26.10	1.70	0.10
SD1/SD2	0.41±0.17	0.55±0.27	-2.35	0.02	0.41±0.15	0.49±0.23	-1.43	0.16
SampEn	1.44±0.28	1.19±0.43	2.68	0.01	1.41±0.31	1.16±0.41	2.40	0.02

Next, we show the impact of data reliability in term of validity labeling, where the detailed results are presented in Table 2 and Table 3. In left side of Table 2, we present the statistical results based on the data collected in the early phase of the study (28 participants, left half in Table 2). The right side presents the results after the validity labeling and throwing away the data with poor reliability label. As study keeps going, more cases were added to our data set. Finally, 62 participants were accounted and 46 are kept after the labeling and exclusion, as shown in Table 3. Such expanded dataset was used to assess whether the observed effects remain consistent with increased sample size.

From Table 2 and Table 3, we may observe that when validity labeling was disabled (including the data with all types of quality), HRV metrics in the depression group appeared higher level than those in the healthy control group. When the validity labeling was activated and the low reliability data was excluded, the statistical results of HRV metrics in the depression group become lower than in the control group, which was consistent in trend with most of the previous reports.

This shift suggests that HRV data of poor reliability may have contributed to error estimation in measurement, particularly in small sample size. After such validation method being used, the results aligned more closely with expectations, highlighting the importance of validation method in improving the reliability of HRV-based depression analysis.

3.2. AI-based Prediction Results

To evaluate the impact of validity labeling in AI-based depression prediction, we also trained traditional machine learning models as the benchmarks, including linear regression (LR)

model, a neural network model of multi-layer perception (MLP), random forest (RF), and gradient boosting (GB).

The training/evaluation dataset was categorized into four subsets:

Set 1: The data set with validity labeling DISABLED (n=28), corresponding to left side of Table 2.

Set 2: The data set with validity labeling DISABLED (n=62), corresponding to left side of Table 3.

Set 3: The data set with validity labeling ACTIVATED over Set 1.

Set 4: The data set with validity labeling ACTIVATED over Set 2.

Different with statistical analysis method, where the data with poor reliability label is excluded, the AI-based prediction methods can utilize the data of all kinds. To further verify such idea, we evaluated the prediction accuracy for the AI-based methods, which are trained to predict whether the given sample belongs to control group or depression group.

As shown in Table 4, the designed LLM with ICL method outperformed all benchmarks across different data subsets. However, the performance of the methods varied between un-labeled data sets (Set 1, Set 2) and validity-labeled data sets (Set 3, Set 4).

Table 4. Overall prediction performance over different data sets.

Method	Set 1	Set 2	Set 3	Set 4
LR	1	0.6	0.7	0.6

Method	Set 1	Set 2	Set 3	Set 4
MLP	0.6	0.7	0.6	0.6
GB	0.6	0.7	0.5	0.7
RF	0.9	0.8	0.8	0.7
LLM with ICL (Proposed)	1	0.8	0.7	0.8

Specifically, comparing the results over Set 1 and Set 3, the LLM's accuracy dropped from 100% to 70% after introducing validity labeling. This reduction likely reflects that lacking the label of noisy or non-stationary data may easily cause overfitting. In contrast, the comparison of Set 2 vs. Set 4 shows that the stable accuracy is approximately around 0.8, indicating improved robustness under broader sample conditions. Furthermore, Set 3 vs. Set 4 suggests that validity labeling provides consistent performance even as the dataset expands. These comparisons demonstrate that adding the sVRI-based reliability label as a feature enhances the prediction model's stability, interpretability, and alignment with physiological expectations, especially in AI-driven depression prediction tasks with small sample size.

Table 5. Ablation study.

Method	Set 1	Set 2	Set 3	Set 4
Fine-tuned prompt + GPT-4	1	0.6	0.7	0.6
Simplified prompt + PCA + GPT-4	0.6	0.7	0.6	0.6
Fine-tuned prompt + PCA + GPT-4	1	0.8	0.7	0.8

Further ablation studies (Table 5) show that feature selection module (i.e., PCA) and fine-tuned prompts significantly contribute the LLM's accuracy. Applying PCA increased accuracy from 0.6 to 0.8 in Set 2 and Set 4, highlighting the importance of removing redundant features.

Additionally, Table 6 demonstrates the effect of sample size and feature selection size: Unlike traditional models, increasing the number of samples did not linearly improve LLM's accuracy due to token length limitations. The optimal number of features was around 5, balancing informativeness and overfitting risks.

Table 6. Impact of sample and feature size.

Number of Samples	10	20	30
LLM with ICL (Proposed)	0.4	0.7	0.6
Number of Features	3	5	10
LLM with ICL (Proposed)	0.7	0.8	0.5

The above findings suggest that validity labeling improves AI-driven HRV-based depression prediction. While sample diversity enhances model generalization, ensuring measurement reliability is crucial for improving prediction accuracy and aligning AI analysis with the expectations on physiological and psychological effects.

4. Discussion

HRV is determined by the periods of time between successive heart beats, known as RR intervals (named for the heartbeat's R-phase) [30]. Although HRV manifests as a function of heart rate, it actually originates from the nervous system [31]. The autonomic nervous system is a component of the peripheral nervous system that regulates the involuntary aspects of physiology and regulates bodily functions [32]. The involuntary physiologic processes including heart rate, blood pressure, respiration, digestion, and sexual arousal. It contains three anatomically distinct divisions: sympathetic, parasympathetic and enteric [33].

While the Sympathetic and Parasympathetic responses are one of the stronger influences on HRV, they might not be the only things that can affect it. There are many factors, when under mental load states such as stress and fear, relaxation and cognitive load, tension or exciting events, could cause HRV value to change in unexpected ways [34]. If it happens when someone is having a 5 min ST-HRV testing, the effects of momentary states even transient emotional changes are quite possible to drown out any effect the Sympathetic and Parasympathetic nervous system has. Thus, false signals are possibly given off when all subjects remain steady and still throughout testing. As no method being used to quantify how strong their impacts are, HRV measurement had sometime been called a tricky metric. In fact, the imperceptibility of non-stationarity and unpredictability in noticing a factor for its HRV outcome highlight the challenges in modelling of predictive AI.

In the context of using HRV analysis for diagnosing and assessing depression, the resting state throughout testing is of paramount importance as it more accurately mirrors stress levels. Therefore, validation strategy is necessarily needed as a labeling method embedded in HRV analytical system capable of identifying and rejecting non-stationary recordings in particular.

The sVRI, a PPG derived parameter representing beat-to-beat stress-induced vascular response, was employed in this study to improve model training, particularly at the very beginning of experiment with small sample size. In this study, a standardized validation method was proposed, which had been incorporated within JM2020, the threshold of standard deviation of sVRI (SDsVRI) and the coefficient of variation of SDsVRI (SDsVRI CV) had been embedded as default. By monitoring the standard deviation and its variations in segments within the 5 min ST-HRV testing process, the sVRI data demonstrated that quantifying the resting state

and assessing the results in term of validation label significantly improved the modelling accuracy.

5. Conclusions

This study demonstrates that depression is associated with reduced HRV although small sample size when reliable data is used, which aligns with findings reported in the literature. The validity method we introduced, using the beat-to-beat sVRI derived from PPG signals, played a key role in improving the HRV-based prediction performance. By integrating this validation strategy, we identified and mitigated the impact of poor quality data which contributed to measurement instability and potential error estimation of HRV in model training.

Moreover, we investigated the performance of a large language model (LLM) in depression prediction under conditions of limited sample size. As an effort to rigorously evaluate the model's effectiveness, we applied both the validation strategy as bias control measures and professional psychiatric supervision to ensure the experiment implemented in accordance with Clinical Practice Guidelines. We proposed an in-context learning (ICL) framework that enables the LLM to process both linguistic and numerical data for depression classification. The experimental results demonstrate that the validation method enhances prediction performance, particularly in conditions of small sample size, by leveraging validity labeling to refine model learning. In addition, the improvements in the data analyzing highlight the potential of sVRI as a robust validity metric for HRV-based mental health assessments.

Future research should focus on expanding the data sets with a larger and more diverse sample population, further validating the role of labeling in both statistical analysis and predictive AI. Additionally, optimizing feature selection and refining structured prompts for LLM-based predictions could enhance model generalization and clinical applicability. As more validated HRV recordings and psychiatric assessments accumulate, our proposed framework will continue to evolve, providing a scalable and effective approach for AI-assisted depression diagnosis in medical big data applications.

Abbreviations

HRV	Heart Rate Variability
PPG	Photoplethysmography
ICL	In-Context Learning
LLM	Large Language Model
sVRI	Stress-induced Vascular Response Index
SD	Standard Deviation
CV	Coefficient of Variation
ST	Short-Term
PHQ	Patient Health Questionnaire
BMI	Body Mass Index

SDNN	Standard Deviation of NN Intervals
RMSSD	Root Mean Square of Successive RR Interval Differences
pNN50	Percentage of Successive NN Intervals That Differ by More Than 50 ms
VLF	Very-Low-Frequency
LF	Low-Frequency
HF	High-Frequency
SampEn	Sample Entropy
PCA	Principal Component Analysis
LR	Linear Regression
MLP	Multi-layer Perception
RF	Random Forest
GB	Gradient Boosting

Acknowledgments

The authors would like to thank Guangdong Vocational and Technical College for the recruitment of research participants and the data collection.

Funding

This work was supported in part by a research grant from Guangdong Health Information Net Association with Project No. ZX-202408-0003.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Cassano, P., Fava, M. Depression and public health: an overview, *Journal of psychosomatic research*. 2002, 53(4): 849-857. [https://doi.org/10.1016/S0022-3999\(02\)00304-5](https://doi.org/10.1016/S0022-3999(02)00304-5)
- [2] Diagnostic and statistical manual of mental disorders, fifth Edition, Washington, DC: American psychiatric association; 1980, pp. 205-224.
- [3] Jesulola, E., Micalos, P., Baguley, I. J. Understanding the pathophysiology of depression: From monoamines to the neurogenesis hypothesis model-are we there yet? *Behavioural brain research*. 2018, 341: 79-90. <https://doi.org/10.1016/j.bbr.2017.12.025>
- [4] Anisman, H., Zacharko, R. M. Depression: The predisposing influence of stress, *Behavioral and brain sciences*. 1982, 5(1): 89-137. <https://doi.org/10.1017/S0140525X00010633>
- [5] Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., Beekman, A. T. F. Prevalence and predictors of recurrence of major depressive disorder in the adult population, *Acta Psychiatrica Scandinavica*. 2010, 122(3): 184-191. <https://doi.org/10.1111/j.1600-0447.2009.01519.x>

- [6] Wells, T. T., Beevers, C. G. Biased attention and dysphoria: Manipulating selective attention reduces subsequent depressive symptoms, *Cognition & Emotion*. 2010, 24(4): 719-728. <https://doi.org/10.1080/02699930802652388>
- [7] Maybery, D, Goodyear, M., Reupert, A. The family-focused mental health practice questionnaire, *Archives of Psychiatric Nursing*. 2012, 26(2): 135-144. <https://doi.org/10.1016/j.apnu.2011.09.001>
- [8] Kemp, A. H., Quintana, D. S., Gray, M. A., Felmingham, K. L., Brown, K., & Gatt, J. M. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis, *Biological psychiatry*. 2010, 67(11), 1067-1074. <https://doi.org/10.1016/j.biopsych.2009.12.012>
- [9] Jangpangi, D., Mondal, S., Bandhu, R., Kataria, D., & Gandhi, A. Alteration of heart rate variability in patients of depression, *Journal of clinical and diagnostic research: JCDR*. 2016, 10(12), CM04. <https://doi.org/10.7860/JCDR/2016/22882.9063>
- [10] Schumann, A., Andrack, C., & Baer, K. J. Differences of sympathetic and parasympathetic modulation in major depression, *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2017, 79, 324-331. <https://doi.org/10.1016/j.pnpbp.2017.07.009>
- [11] Hartmann, R., Schmidt, F. M., Sander, C., & Hegerl, U. Heart rate variability as indicator of clinical state in depression, *Frontiers in psychiatry*. 2019, 9, 735. <https://doi.org/10.3389/fpsy.2018.00735>
- [12] Hopwood, M. J., Malhi, G. To screen for depression or not?, *The Medical Journal of Australia*. 2016, 204(9): 329. <https://doi.org/10.5694/mja16.00217>
- [13] Beauchaine, T. P., Thayer J. F. Heart rate variability as a transdiagnostic biomarker of psychopathology, *International Journal of Psychophysiology*. 2015, 98(2): 338-350. <https://doi.org/10.1016/j.ijpsycho.2015.08.004>
- [14] Pham, T., Lau, Z. J., Chen, S. H. A. Heart rate variability in psychology: A review of HRV indices and an analysis tutorial, *Sensors*. 2021, 21(12): 3998. <https://doi.org/10.3390/s21123998>
- [15] Koch, C., Wilhelm, M., Salzmann, S., Rief, W., Euteneuer, F. A meta-analysis of heart rate variability in major depression, *Psychological Medicine*. 2019, 49(12): 1948-1957. <https://doi.org/10.1017/s0033291719001351>
- [16] Pinna, G. D., Maestri, R., Torunski, A., et al. Heart rate variability measures: a fresh look at reliability, *Clinical Science*. 2007, 113(3): 131-140. <https://doi.org/10.1042/CS20070055>
- [17] Chang, C., Metzger, C. D., Glover, G. H., Duyn, J. H., Heinze, H.-J., Walter, M. Association between heart rate variability and fluctuations in resting-state functional connectivity, *NeuroImage*. 2013, 68: 93-104. <https://doi.org/10.1016/j.neuroimage.2012.11.038>
- [18] Campbell, J., Ehler, U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*. 2012, 37(8): 1111-1134. <https://doi.org/10.1016/j.psyneuen.2011.12.010>
- [19] Vacareanu, R., Negru, V. A., Suci, V., Surdeanu, M. From words to numbers: Your large language model is secretly a capable regressor when given in-context examples, *First Conference on Language Modeling*, Philadelphia, USA, 2024.
- [20] Sun, X., Su, F., Chen, X., Peng, Q., Luo, X., & Hao, X. Doppler ultrasound and photoplethysmographic assessment for identifying pregnancy-induced hypertension. *Experimental and Therapeutic Medicine*. 2020, 19(3), 1955-1960. <https://doi.org/10.3892/etm.2019.8405>
- [21] Su, F, Li Z, Sun, X, et al. The pulse wave analysis of normal pregnancy: investigating the gestational effects on photoplethysmographic signals, *Bio-med Mater Eng*. 2014; 24: 209-19. <https://doi.org/10.3233/bme-130801>
- [22] Tang, S., Luo, X., Meng, S., Wang, Z., & Zhao, A. The impact of maternal vasodilatation as pregnancy progress on peripheral arterial tonometry in assessment of endothelial function. *IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, USA. 2020, pp. 241-246. <https://doi.org/10.1109/BIBE50027.2020.00047>
- [23] Zhao, A., Chong, Y., Zhong, H., Ma, J., Zhang, H., Luo, Z., & Luo, X. Clinical Assessment of Brachial-Ankle Pulse Wave Velocity and Stiffness Index: Hypertriglyceridemia Effects on Arterial Stiffness. *IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, Washington, DC, USA, 2017, pp. 537-541. <https://doi.org/10.1109/BIBE.2017.00008>
- [24] Xiao-Min, L., Lei, W., & Lei, Y. Identification of susceptibility to acute mountain sickness by detecting vascular tone using a photoplethysmographic sensor. *Chinese Journal of Applied Physiology*. 2016, 32(6), 494-498. <https://doi.org/10.13459/j.cnki.cjap.2016.06.003>
- [25] Luo, X., Wang, L., & Yang, L. Influence of induced altitude acclimatization on development of acute mountain sickness associated with a subsequent rapid ascent to high altitude. *IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, Taichung, Taiwan, China. 2016, pp. 289-292. <https://doi.org/10.1109/BIBE.2016.41>
- [26] Liu, Z., Zhou, Y., Yi, R., He, J., Yang, Y., Luo, L., & Luo, X. Quantitative research into the deconditioning of hemodynamic to disorder of consciousness carried out using transcranial Doppler ultrasonography and photoplethysmography obtained via finger-transmissive absorption. *Neurological Sciences*. 2016, 37, 547-555.
- [27] Luo, L., Xiao, L., Miao, D., Luo, X. The Relationship between Mental Stress Induced Changes in Cortisol Levels and Vascular Responses Quantified by Waveform Analysis: Investigating Stress Dependent Indices of Vascular Changes, *International Conference on Biomedical Engineering and Biotechnology*, Macao, China, 2012. <https://doi.org/10.1109/iCBEB.2012.437>
- [28] Lyu, Y., Luo, X., Zhou, J., et al. Measuring Photoplethysmogram-Based Stress-Induced Vascular Response Index to Assess Cognitive Load and Stress, *The 33rd annual ACM conference on human factors in computing systems*, Seoul, Korea, 2015, pp. 857-866. <https://doi.org/10.1145/2702123.2702399>

- [29] Zhang, X., Lyu, Y., Qu, T., Qiu, P., Luo, X., Zhang, J., Fan, S., Shi, Y. Photoplethysmogram-based cognitive load assessment using multi-feature fusion model. *ACM Transactions on Applied Perception (TAP)*, 2019, 16(4), 1-17. <https://doi.org/10.1145/3340962>
- [30] Draghici, A. E., Taylor J. A. The physiological basis and measurement of heart rate variability in humans. *Journal of physiological anthropology*. 2016, 35: 22. <https://doi.org/10.1186/s40101-016-0113-7>
- [31] Grégoire, J. M., Gilon, C., Carlier, S., Bersini, H. Autonomic nervous system assessment using heart rate variability, *Acta cardiologica*. 2023, 78(6): 648-662. <https://doi.org/10.1080/00015385.2023.2177371>
- [32] Gaskell, W. H. The involuntary nervous system. Longmans, Green, 1920.
- [33] Wehrwein, E. A., Orer, H. S., Barman, S. M. Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system, *Comprehensive Physiology*. 2016, 6(3): 1239-78. <https://doi.org/10.1002/cphy.c150037>
- [34] Sammito, S., Böckelmann, I. Factors influencing heart rate variability, *International Cardiovascular Forum Journal*. Vol. 6. 2016.