GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Medicine and Life Sciences

Article

AI-Driven Clinical Decision Support Optimizes Treatment Accuracy for Mental Illness

Danyating Shen [1,*]

[1] Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
* Correspondence: Danyating Shen, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

Abstract: This paper focuses on exploring the application of AI-based clinical decision support systems in the precise treatment of mental disorders, and analyzes their mechanisms of action, key techniques, and technical solutions. This paper proposes a series of system architecture methods, including unstructured data processing, multi-modal feature fusion, individualized treatment modeling, and interpretable process design, to improve the efficiency and individualization of precise treatment for mental disorders. Meanwhile, an experimental verification system was proposed to comprehensively verify the functional performance and practical value of the system, providing good technical support for the intelligent diagnosis and treatment of clinical mental disorders by this system.

Keywords: artificial intelligence; mental illness; clinical decision support; multimodal fusion; interpretability

1. Introduction

The causes of mental disorders are diverse; the symptoms are complex; and individual differences are significant. Therefore, there are unknown and subjective uncertainties in the process of onset and diagnosis and treatment, making it difficult to improve the precise treatment path. With the application of AI in the medical field, the clinical decision support system (CDSS) based on AI has continuously become a key factor in improving the precise diagnosis of mental disorders. This article mainly systematically summarizes the key issues in the precise treatment of mental disorders by AI, discusses its basic theories, technical frameworks and application effects, and proposes specific solutions for the construction of intelligent mental health management.

Received: 04 June 2025
Revised: 10 June 2025
Accepted: 23 June 2025
Published: 26 June 2025

2. The Theoretical Basis and Key Technologies of AI in the Precise Treatment of Mental Disorders

2.1. Theoretical Roots of the Complexity in the Diagnosis and Treatment of Mental Disorders and the Demand for Precise Intervention

Mental disorders have extremely high heterogeneity. They are caused by multi-source factors including neurobiological, psychosocial, genetic and other factors. The clinical manifestations are diverse, with variable forms and changes, and different patients have inconsistent responses in terms of therapeutic effects. Therefore, the empirical diagnosis and treatment model is difficult to achieve the goal of precise treatment. To solve this problem, precise intervention must be adopted. Besides requiring a higher level of ability to construct specific patient models, it also requires technologies that can provide dynamic and individualized treatment suggestions [1]. Due to its technical advantages such as the ability to process multiple types of information, the ability to establish nonlinear patterns, and the ability to predict, artificial intelligence can become a key means to

overcome the instability and complexity of the diagnosis and treatment of mental disorders, and gradually become an important factor in promoting the development of precise medical treatment for mental disorders.

2.2. Core Principles of Artificial Intelligence Empowering Clinical Decision Support

AI establishes and improves the clinical pathways of mental disorders by mining the hidden connections contained in multimodal clinical big data. The clinical decision support system is essentially to construct a prediction function 1, where 2 represents the multi-dimensional characteristics of the patient (such as symptoms, images, physiological indicators, etc.), and 3 is the diagnostic or therapeutic output. The training objective of the model is:

$$\min_{f \in F} \frac{1}{N} \sum_{i=1}^{N} \vartheta(f(x_i), y_i) + \lambda \Omega(f) \quad (1)$$

Among them, ϑ is the loss function, which measures the prediction error; $\Omega(f)$ is the regularization term, which is used to prevent overfitting. Complex features are constructed through deep learning structures (such as CNN, RNN, Transformer, etc.) to improve the accuracy of diagnosis. The interpretability of this model is also a key factor that can enhance the trust of medical staff in it. Common ones include the application of attention mechanism, SHAP and LIME, which can explain the important factors for the model to make decisions, helping doctors understand and judge the results, and further achieving the intelligent decision-making mode of "human-machine collaboration" [2].

2.3. Key Technical System of AI Decision Support for Mental Illness

The key to the AI clinical decision support system for mental illness lies in the unified modeling and precise reasoning of multi-source heterogeneous data, and its core is to construct a fused, interpretable and individualized prediction model architecture. For different modal data (such as electronic medical records, brain images, biochemical indicators, behavioral trajectories, etc.), the system first extracts feature representations through modal-specific encoders $E_i(x^{(i)})$, and then integrates them into a unified latent representation z through a fusion mechanism:

$$z = F_{fusion}\left(E_1(x^{(1)}), E_2(x^{(2)}), \ldots, E_m(x^{(m)})\right) \quad (2)$$

Among them, F_{fusion} can be fusion strategies such as attention weighting, tensor interaction, or cross-modal Transformer. Finally, diagnostic suggestions, intervention plans or risk assessment results are output through prediction function $f(z)$. This system emphasizes the collaborative expression of cross-modal features, individualized path modeling and the interpretability of prediction results, ensuring that the model has robustness, adaptability and clinical practical value in complex scenarios, and providing intelligent support for the precise diagnosis and treatment of mental disorders [3].

3. Design of Clinical Decision Support System for Mental Disorders and Implementation of Key Technologies

3.1. Heterogeneous Medical Data Collection and Intelligent Preprocessing Mechanism

The data sources of patients with mental disorders are numerous and diverse, including structured electronic medical record information, results of various scales, laboratory test results, as well as unstructured disease course information, medical images, and temporal behavior information obtained from mobile devices and wearable devices [4]. These data have diverse patterns, inconsistent attributes and inconsistent acquisition rates, which bring great problems to the synchronization and accuracy of the input of the decision-making system.

The system has constructed a multi-source and heterogeneous data access module, mapping all heterogeneous data into low-level feature Spaces. Transform structured data into vectors through field matching and normalization [5]. Medical entity recognition and

context-filling technologies are adopted for the text; The image data uses the trained convolutional neural network to capture spatial features; Temporal behavior data generate continuous input sequences through the sliding window model and frequency filling method. By using information gain and minimizing redundancy as the guiding principles for feature selection, the efficiency of the modeling process is improved.

Multimodal data are integrated through the weighted fusion strategy, and the overall representation form is as follows:

$$z_i = \sum_{k=1}^{K} \alpha_k \cdot \psi_k\left(x_i^{(k)}\right) \quad (3)$$

Among them, $x_i^{(k)}$ represents the original feature of the k-th mode, ψ_k is the corresponding feature encoding function, and α_k is the modal weight learned by the model, satisfying $\sum_{k=1}^{K} \alpha_k = 1$. This model can dynamically allocate the importance of corresponding patterns according to the adaptability of the samples, and better solve the problem of the comprehensive representativeness of the entire model. Through the preprocessed consistent feature vectors, they are input into the model as interfaces to ensure the reliability, accuracy and cross-mode adaptability of the modeling.

3.2. Multimodal Feature Modeling and Deep Fusion Strategies

For the diagnosis and treatment suggestions of mental disorders, it is necessary to use information of various modalities to train appropriate predictive models and provide reasonable suggestions. That is, data of different modalities, including demographic parameters, laboratory data, medical records and psychological test results, etc., all need to be incorporated into the system. By adopting a parallel approach to analyze different types of information and setting up corresponding sub-modules, the system can automatically identify the deep implicit meanings of different modal data. Take structured data as an example. Structured data is constructed using MLP. For text, deep networks such as BERT or Clinical BERT are used to learn the medical context of the text. For images, convolutional neural networks such as ResNet and DenseNet are used to capture the spatiotemporal features of the images. Behavioral data uses bidirectional GRU or Transformer to capture time-related semantic background transitions.

After completing the internal modeling of the modalities, the entire system will achieve a unified synthesis method for multiple modalities, mixing the data information of each modality. For example, it can be accomplished through attention-weighted allocation, optimized by a multimodal matching loss function (such as one based on contrastive learning), or by incorporating co-representation space mapping as an intermediate step. In order to increase overall stability, co-attention and self-attention layers are introduced, thereby making the information exchange among various modalities more flexible and retaining the initial characteristic information of each modality. The model then transmits the comprehensive expression to the decision-level network to achieve purposes such as classification, regression and decision suggestions.

3.3. Design of Individualized Intervention Path Planning Model

In this intelligent process model of psychological intervention, the intelligent path selection mode is an intermediate link that connects behavioral and strategic decisions. The model establishes a decision-making path state diagram and generates several possible treatment options (such as adjusting drug dosage, psychological counseling, behavioral therapy, physical therapy, etc.), and uses these options as nodes in the diagram. Each edge represents a possible path from one option to another, and its weight is calculated in real time based on factors such as the possibility of therapeutic effect, side effects, the possibility of compliance, and available medical resources. The patient's state is encoded with multivariate data to form a state vector, which becomes the starting position of graph search.

On this basis, reinforcement learning algorithms or heuristic graph search methods can be adopted to optimize the strategy in order to achieve the path that meets the maximum medical benefits of the goal. During this period, past treatment experiences can be used as environmental feedback, and the reward function can be used to measure and learn the treatment effect to extract useful treatment plans from it. This system has an iterative feedback loop that can receive real-time feedback from patients, thereby updating their state expressions and re-optimizing strategies to enable self-regulation of treatment paths, so as to adaptively adjust to changing disease conditions.

3.4. Decision Support System Architecture and Clinical Interpretability Module

In the AI decision support system for mental illness, it should be suitable for handling extremely complex and multi-source data usage scenarios, capable of efficiently processing data, achieving clear model components, rapid responses, and controllable clinical outputs. This system is based on the idea of hierarchical design and consists of four aspects: the data entry layer, the feature construction layer, the intelligent decision-making layer, and the interpretation output layer. The data entry layer can access data streams from various clinical sources, including structured data (such as diagnostic codes and test results), semi-structured data (such as doctors' written results), unstructured data (such as brain magnetic resonance images), and motion data (such as sleep patterns and wearable device data on movement trajectories). All standardized data are processed by feature extraction modules to achieve consistent representations through embedding, many-to-one mapping, and time series analysis methods.

At this point, multiple modal deep models are fused, such as using the Transformer model to learn time series data, the CNN model to learn medical images, and the BERT-like pre-trained model to learn diagnostic texts, etc., to construct the global model of the patient; Meanwhile, the results of each model are input into the intelligent decision-making link to conduct graph search or reinforcement learning to generate customized intervention paths, or to generate risk scoring indices based on classifiers and regression models; Visualization is performed through causal analysis of interpretable models, including the representation of feature importance and local explanations, such as segmentation with SHAP values or direct textual explanations. Output the diagnostic suggestions, intervention approaches and prediction basis, record the feedback information of doctors in real time for model improvement, and achieve the process of "prediction-explanation-intervention-feedback" in a closed loop, making the system usable and scalable (see Figure 1).

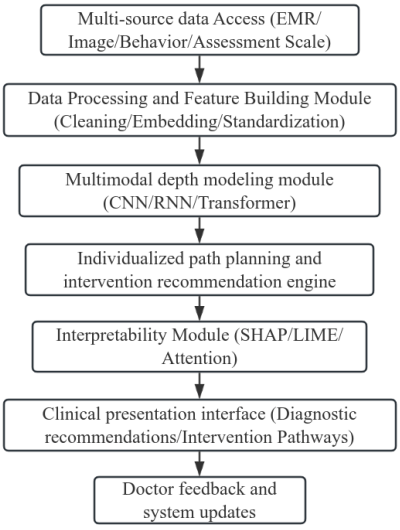

Figure 1. Flowchart of the Overall Architecture of the AI Decision Support System for Mental Illness.

4. Experimental Verification and Performance Evaluation of AI Decision Support System for Mental Illness

4.1. Experimental Data Set and Verification Process Design

In the experimental verification design, the system will construct a universal detection system with data of various types of mental health problems, thereby ensuring richness and verifiability. The experimental data include four types: structured medical reports, brain imaging, event sequences, and natural language. The natural language and event sequence data come from a real psychotherapy conversation database, while the other data are sourced from corresponding clinical repositories. All the data undergo desensitization processing and labeling standardization processes and are input into the data cleaning stage. The interpolation strategy is used to solve the missing problem, and the numerical characteristics are normalized. The word analysis method and the embedding space construction method are adopted to process the text data, thereby ensuring the consistent transfer of different types of data.

At this stage, multimodal deep neural networks are used for joint training, mapping the features of each modal to the shared domain, and the attention mechanism is employed to achieve the mixed trade-offs among different modalities. Five-fold cross-validation was conducted on the training set, and the test set was used for final evaluation. Three outputs were generated: diagnosis category, treatment process sequence, and estimated severity rating. In the evaluation part, the standard classification indicators (precision rate, F1 value, recall rate), path matching accuracy and prediction stability are mainly comprehensively considered. Meanwhile, the interpretability module is provided, which offers SHAP value analysis graphs, attention distributions and text annotations to assist experts in decision-making. Finally, through the doctor's review of the prediction results and analysis contents, the cycle of model accuracy identification and system update is completed (see Figure 2).

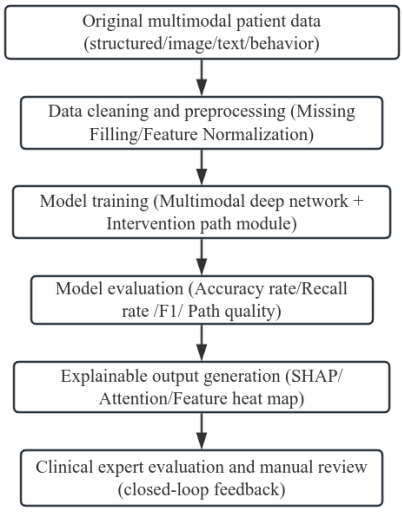

Figure 2. Flowchart of Experimental Verification.

4.2. Comparative Analysis of Multimodal Models and Evaluation of Feature Contributions

Longitudinal studies were conducted on five models, namely the traditional machine learning method, the single-modal CNN, the single-modal Transformer, the multimodal fusion model, and the complete path optimization system. The training and testing of all models adopted the same dataset partitioning scheme to divide the training set and the validation set, and both were used as a measurement scale. The evaluation basis was the accuracy, precision, and the performance improvement brought by feature fusion (see Table 1).

Table 1. Multi-Model Performance Comparison and Analysis Table.

Model type	Accuracy (%)	Precision (%)	ΔF1 (%)
Random Forest	72.3	65.1	-
CNN (Video)	76.8	69.5	+3.1
Transformer	78.4	70.7	+4.5
Multimodal fusion model	84.6	78.2	+9.8
Fusion + path recommendation system	87.2	81.3	+11.6

It can be seen from Table 1 that the feature representation capability of the original random forest model is limited and cannot effectively capture complex pathological associations. The single-modal CNN and Transformer models have improved in the two modalities of image and text, but there are certain limitations due to the lack of cross-modal information interaction. The multimodal fusion model of attention and shared representation has improved its diagnostic accuracy and suggestion precision, increased the F1 value, indicating that the comprehensive multi-information has an optimization effect on the high specificity of mental illness discrimination. Further combined with the route planning of graphics, an overall framework has been formed, achieving the best performance in individualized recommendation tasks. It also indicates that this architecture and the idea of sequential optimization have practicality and technical advantages for individualized treatment.

4.3. Evaluation of Interpretability and Clinical Consistency

The research on interpretability and clinical consistency mainly analyzes whether the logic predicted by the system conforms to medical-related knowledge and whether the model results are reasonably adopted and used in clinical practice, etc. The explanation process integrates both quantitative and qualitative methods. In terms of quantitative aspects, an explanation interface for doctors was generated by using SHAP values, visualization of attention weights, and the order of feature importance, and at the same time, scores were given by psychiatric experts who have held relevant certificates.

An analysis of Table 2 shows that the features calculated by the mixed data model are relatively stable, and the key indicators show consistent distribution. The average feature interpretability score of the model reaches 4.6 points (out of 5), indicating that it can clearly distinguish key influencing factors such as emotional scores and changes in brain region activity.

Table 2. Evaluation Table of Model Interpretability and Clinical Consistency.

Evaluation dimension	Indicator value	Explanation
Average feature interpretability score	4.6/5	The doctor's score for the clarity of feature interpretation
Proportion of diagnostic consistency	85.7%	The sample proportion consistent with the doctor's diagnosis in the model
Intervention pathway adoption rate	79.3%	The proportion of doctors accepting the model-recommended paths
Average clinical confidence score	4.4/5	The doctor's evaluation of the overall trust level of the system

In the consistency verification stage, the manually labeled diagnostic results are compared with the predicted results of the system, and the proportion of cases with consistent judgments is taken as the case-level consistency rate. The consistency between the system's prediction results and the diagnosis results of senior physicians reached 85.7%. Good accuracy was still maintained in severe and difficult cases. Moreover, among the number of treatment plan suggestions given by the system, the number of plans ultimately accepted and adopted by medical staff accounted for 79.3% of the total, indicating that the model has a high degree of reliability and is also practical and credible. In addition,

the overall credibility of the system was evaluated in combination with the feedback forms of medical staff, and the average satisfaction score obtained was 4.4 points.

5. Conclusion

This paper constructs an AI-based clinical decision support system for mental disorders, capable of processing multiple types of data and supporting various modeling approaches. Meanwhile, it can generate personalized intervention paths and provide easily interpretable explanations. The experimental results of this system demonstrate strong diagnostic consistency and intervention guidance, providing practical and operational methods for the precise treatment of mental disorders. It has extremely wide applicability.

References

1. M. Elhaddad and S. Hamam, "AI-driven clinical decision support systems: An ongoing pursuit of potential," *Cureus*, vol. 16, no. 4, 2024, doi: 10.7759/cureus.57728.
2. C. Y. Elgin and C. Elgin, "Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: A qualitative study of healthcare professionals' perspectives," *BMC Med. Ethics*, vol. 25, no. 1, p. 148, 2024, doi: 10.1186/s12910-024-01151-8.
3. A. Pesqueira, M. Sadat, J. Oliveira, M. Ribeiro, R. Teixeira, and C. Costa, et al., "Designing and implementing SMILE: An AI-driven platform for enhancing clinical decision-making in mental health and neurodivergence management," *Comput. Struct. Biotechnol. J.*, vol. 27, pp. 785–803, 2025, doi: 10.1016/j.csbj.2025.02.022.
4. N. S. Mosavi and M. F. Santos, "Enhancing clinical decision support for precision medicine: A data-driven approach," *Informatics*, vol. 11, no. 3, 2024, doi: 10.3390/informatics11030068.
5. S. Jhade, A. Kumar, R. Singh, M. Patel, V. Sharma, and P. Raj, et al., "Smart medicine: Exploring the landscape of AI-enhanced clinical decision support systems," in *MATEC Web Conf.*, vol. 392, 2024, Art. no. 01083, doi: 10.1051/matecconf/202439201083.