

Article

Advancing Mental Health Care: Intelligent Assessments and Automated Generation of Personalized Advice via M.I.N.I and RoBERTa

Yuezhong Wu ¹, Huan Xie ¹, Lin Gu ², Rongrong Chen ³, Shanshan Chen ⁴, Fanglan Wang ⁴, Yiwen Liu ⁵, Lingjiao Chen ¹ and Jinsong Tang ^{4,*}

- ¹ School of Rail Transit, Hunan University of Technology, Zhuzhou 412007, China; wuyuezhong@hut.edu.cn (Y.W.); m22085400045@stu.hut.edu.cn (H.X.); chenlingjiao@hut.edu.cn (L.C.)
- ² RIKEN AIP (RIKEN Center for Advanced Intelligence Project (AIP)), Rigaku Kenkyujo Kakushin Chino Togo Kenkyu Senta, Tokyo 103-0027, Japan; lin.gu@riken.jp
- ³ School of Business, Hunan University of Technology, Zhuzhou 412007, China; chenrongrong@hut.edu.cn
- ⁴ School of Medicine, Zhejiang University, Hangzhou 310020, China; sschen33@zju.edu.cn (S.C.); 12118265@zju.edu.cn (F.W.)
- ⁵ School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China; m23085400026@stu.hut.edu.cn
- * Correspondence: tangjinsong@zju.edu.cn

Abstract: As mental health issues become increasingly prominent, we are now facing challenges such as the severe unequal distribution of medical resources and low diagnostic efficiency. This paper integrates finite state machines, retrieval algorithms, semantic-matching models, and medical-knowledge graphs to design an innovative intelligent auxiliary evaluation tool and a personalized medical-advice generation application, aiming to improve the efficiency of mental health assessments and the provision of personalized medical advice. The main contributions include the following: (1) Developing an auxiliary diagnostic tool that combines the Mini-International Neuropsychiatric Interview (M.I.N.I.) with finite state machines to systematically collect patient information for preliminary assessments; (2) Enhancing data processing by optimizing retrieval algorithms for efficient filtering and employing a fine-tuned RoBERTa model for deep semantic matching and analysis, ensuring accurate and personalized medical-advice generation; (3) Generating intelligent suggestions using NLP techniques; when semantic matching falls below a specific threshold, integrating medical-knowledge graphs to produce general medical advice. Experimental results show that this application achieves a semantic-matching degree of 0.9 and an accuracy of 0.87, significantly improving assessment accuracy and the ability to generate personalized medical advice. This optimizes the allocation of medical resources, enhances diagnostic efficiency, and provides a reference for advancing mental health care through artificial-intelligence technology.

Keywords: mental health; artificial intelligence; natural language processing; medical-knowledge graph; automatic generation



Citation: Wu, Y.; Xie, H.; Gu, L.; Chen, R.; Chen, S.; Wang, F.; Liu, Y.; Chen, L.; Tang, J. Advancing Mental Health Care: Intelligent Assessments and Automated Generation of Personalized Advice via M.I.N.I and RoBERTa. *Appl. Sci.* **2024**, *14*, 9447. <https://doi.org/10.3390/app14209447>

Academic Editors: Giovanni Costantini and Daniele Giansanti

Received: 21 September 2024

Revised: 12 October 2024

Accepted: 15 October 2024

Published: 16 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With modern life accelerating and societal competition increasing, mental and psychological health issues have become significant global public health challenges. A World Health Organization report indicates that over one billion people globally suffer from various mental health issues [1]. However, many individuals still lack access to proper detection and treatment. These issues not only profoundly affect individuals but also place a substantial burden on families and societies [2]. Globally, there is a severe imbalance in the distribution of mental and psychological healthcare resources across various countries and regions [3]. This imbalance is especially pronounced in developing countries and

low-income areas, marked by a lack of professional personnel and medical facilities, complicating access to high-quality medical services. In developed countries, the distribution of mental health services frequently fails to meet population needs, leading to many patients struggling to receive timely and effective treatment.

Current medical practice faces multiple challenges in assessing, diagnosing, and treating mental and psychological disorders [4]. Firstly, the uncertainty of symptoms and subjectivity of patient perception make accurate disease state assessment difficult for doctors in clinical practice. Secondly, the lack of objective indicators means these subjective perceptions often compromise the accuracy of assessment, diagnosis, and treatment. Despite abundant online professional medical knowledge, existing knowledge retrieval systems often provide low-quality information and lack precision. Some specialized platforms provide reliable online consultation services with professional psychiatrists, but these services are typically time-consuming and labor-intensive.

To tackle these challenges, this paper leverages the latest computer technology and artificial-intelligence algorithms. Specifically, we integrate the RoBERTa model, based on BERT [5] (Bidirectional Encoder Representations from Transformers), for effective processing and analysis of medical literature and patient records. Additionally, we use the M.I.N.I tool [6], a validated, simple, and effective diagnostic tool for psychiatric disorders. The combination of these methods aims to improve the accuracy of assessments and medical-advice generation for mental and psychological disorders.

Existing research mainly relies on doctors' clinical experience and traditional diagnostic tools, which are often lacking in flexibility and personalization [7]. The limited understanding of complex and variable mental disorder characteristics constrains the effectiveness and adaptability of medical advice. Therefore, this paper aims to develop a novel auxiliary assessment tool and an application for generating personalized medical advice. The goal is to enhance the accuracy of mental disorder assessments and treatment efficiency through technological innovation. Our approach aims to comprehensively and objectively assess patients' mental states and provide tailored medical advice, improving medical resource allocation and patient quality of life. We believe this innovation will benefit patients directly and offer new perspectives and tools for medical practice.

2. Literature Review

The assessment and diagnosis of mental and psychological disorders are complex and multidimensional processes. Traditionally, this field has relied on physicians' clinical experience and standardized assessment tools, such as DSM-5 [8] and ICD-10 [9], which provide guidelines for diagnosing mental disorders. Chakraborty, N., Ali, A., et al. [10] highlighted the World Health Organization's mhGAP Intervention Guide's effectiveness and practicality in assessing and providing mental health services. However, their practical application is limited by the need for professional knowledge and flexibility in the assessment and diagnostic process. The M.I.N.I questionnaire, popular for its brevity and structured format, allows for quicker assessment of major mental disorders. Arrow, K., Resnik, P., Michel, H., et al. [11] demonstrated these tools' feasibility and effectiveness in assessing mental and psychological symptoms.

In medicine, text-generation technology [12–14] is an emerging research area. Researchers like Hasani, A.M. and Singh, S. [15] have successfully used advanced pre-trained deep-learning models like GPT-4 for generating and standardizing medical reports. However, this method, while flexible, has limitations in content control, and its reliability and safety in medicine are not fully established. A more prudent approach is using information-retrieval technology to find relevant information and extract keywords and key information for generating medical advice.

Most text-retrieval technologies [16,17] rely on full-sentence matching, which is fast but lacks specificity in key information extraction. To address this issue, some researchers have proposed semantic retrieval-based models [18–20]. While these improve knowledge-acquisition precision, their effects vary, and they are computationally intensive. They also

do not fully consider the patient’s unique situation for personalized medical advice. We propose a keyword-weight [21,22] matching algorithm for preliminary information filtering. In addition, Liu, Y., Ott, M., Goyal, N., et al. [23] confirmed that RoBERTa, a variant of BERT, performed exceptionally well in semantic sentence-matching tasks. Therefore, we adopted a RoBERTa model based on the Cross-Encoder [24,25] network structure to further enhance the accuracy of information retrieval.

This combination lays the foundation for generating personalized medical advice. The introduction of medical-knowledge graphs [26–29] offers a structured way to organize and query medical information, showing great potential in medical-advice generation.

Building on existing literature and technologies, this paper aims to combine finite state machines, the RoBERTa model, and medical-knowledge graphs to improve mental disorder assessment accuracy and the personalization and universality of medical-advice generation. Our method focuses on assessment accuracy and emphasizes medical-advice personalization and specificity, addressing issues like uneven resource distribution and variable service quality. The following sections will elaborate on our research methods, experimental design, and results and discuss how these innovations can enhance the quality and efficiency of mental and psychological health services.

3. Materials and Methods

3.1. Materials

The dataset, named “Huatuo-26M” [30] is one of the largest Chinese medical Q&A datasets currently available, compiled from multiple reliable sources, as detailed in Table 1. It contains over 26 million high-quality medical Q&A pairs, primarily sourced from public medical Q&A forums on the internet. The forums feature responses from certified medical professionals, with verifiable personal and employment details, ensuring data reliability and credibility.

Table 1. Huatuo-26M dataset information.

Source	Data Fields
Online Medical Encyclopedia	Huatuo_encyclopedia_qa
Medical-Knowledge Graph	Huatuo_knowledge_graph_qa
Public Online Medical Q&A Forums	Huatuo_consultation_qa

The dataset includes diverse Q&A pairs, with internet forum data being particularly valuable due to their detailed patient-related information, aligning well with this study’s experimental environment. For this paper, data from these forums were meticulously gathered and cleaned to build a preliminary dataset. By applying a set of keywords specific to the mental and psychological domain, high-quality Q&A data were filtered and saved as pairs, with questions containing detailed information such as names, symptoms, medications, and patients’ personal information. The answer section consists of expert treatment methods, guidance, and suggestions, totaling approximately 1 million entries.

To date, several studies have utilized the “Huatuo-26M” dataset to construct and evaluate medical question-and-answer systems [31–33], demonstrating its advantages in improving the accuracy and practicality of medical Q&A. The “Huatuo-26M” dataset encompasses a wide range of medical fields, including internal medicine, surgery, pediatrics, and obstetrics and gynecology, among others. The content of the Q&A covers the diagnosis, treatment, and prevention of common diseases, primarily focusing on general medical knowledge. Additionally, it includes a substantial number of Q&A pairs in both the mental and psychological domains, thereby forming a comprehensive medical question-and-answer knowledge base.

Data for constructing the domain-knowledge graph in mental and psychological health were sourced from online medical encyclopedias, medical websites, books, academic papers, and other resources.

The data used to construct the knowledge graph in the field of mental and psychological health were collected from various sources, including online medical encyclopedias, medical information websites, medical books, and academic papers. The entire graph includes 5 types of entities, with a total of 2430 instances; 8 types of attributes; and 9 types of entity relationships, with 5575 instances, as shown in Table 2 below.

Table 2. Knowledge graph data.

Data Category	Category Count	Total Count
Entity	5	2430
Attribute	8	N/A
Public Online	9	5140

3.2. Methods

The methodology proposed in this article consists of three main components: the development and implementation of an auxiliary assessment tool for mental and psychological health, the generation of personalized medical advice, and the application of a medical-knowledge graph. Firstly, patient information is gathered using the auxiliary assessment tool and is then evaluated. This information is then used to generate personalized medical advice. In instances where the data quality is poor or the model matching accuracy falls below a predetermined threshold, making it insufficient for generating personalized advice, the system employs the knowledge graph to produce universally applicable medical advice. The overall architecture is depicted in Figure 1.

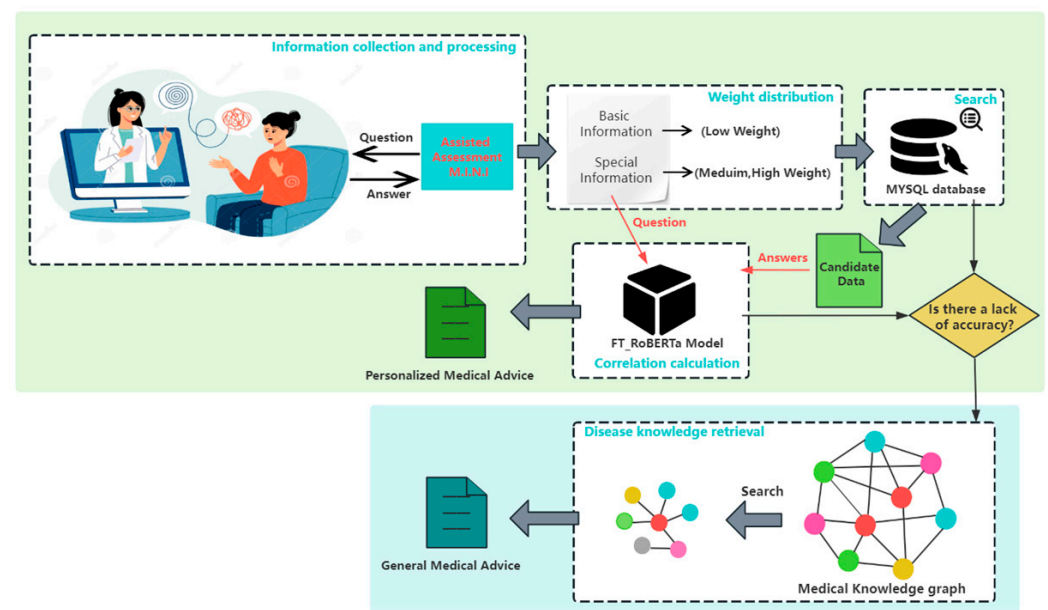


Figure 1. Overall architecture.

3.2.1. Construction and Implementation of the M.I.N.I-Based Diagnostic Tool

With advancements in medical technology and the growth of digitalization, an increasing number of diagnostic tools are utilizing computer technology to boost their efficiency and accuracy. In this study, M.I.N.I version 6.0.0 serves as the basis for the auxiliary assessment of mental and psychological disorders. It is supported by technologies like finite state machines and web interaction for automated and intelligent information support.

1. Overview of M.I.N.I

M.I.N.I is an extensive diagnostic assessment tool, comprising 16 independent modules from A to P. Each module covers a broad range of content and employs intricate logical

calculations and judgment mechanisms. Considering the substantial differences in the assessment methods of diagnostic results among these modules, simplifying M.I.N.I into a standard questionnaire is not recommended.

To realize the objectives of automation and intelligence, this paper adopts a specialized modeling approach for each module. By applying the finite-state-machine model, precise logic calculations and logical jumps based on user responses are possible. This approach not only facilitates the flexible collection of multiple indicator information but also enables comprehensive result assessment based on this data. Once information collection is completed in all modules, the system can generate an integrated preliminary assessment report. The use of this method is expected to significantly enhance the precision and efficiency of diagnoses. The implementation effect is illustrated in Figure 2. The picture presents three parts from left to right: an overview of the overall auxiliary diagnosis module, the information collection process, and the evaluation results.

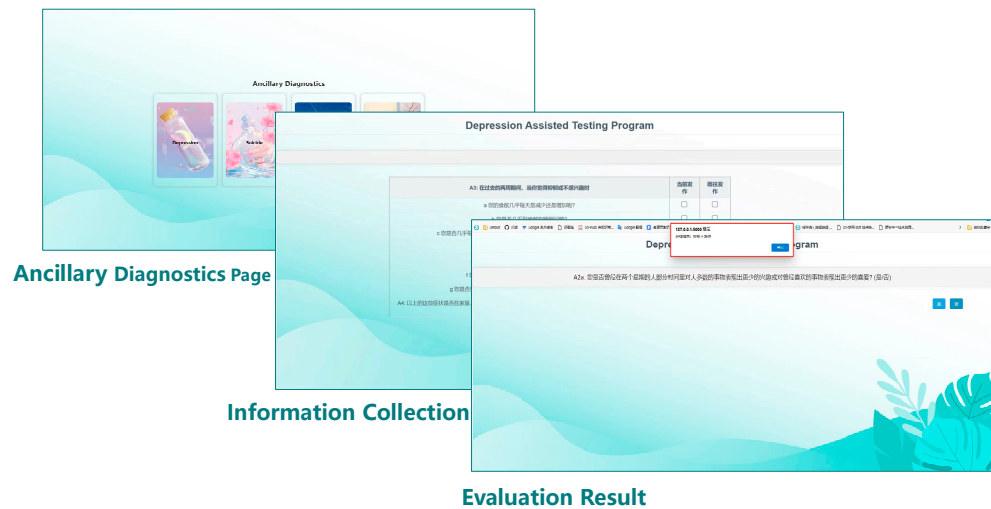


Figure 2. Auxiliary assessment effect.

2. Finite state machine and its application in this study

State machines [34,35] are computational models that describe the behavior or state transitions of objects or systems in response to external events or conditions. In event-driven systems, their behavior or state changes in response to external events or conditions. Table 3 shows the basic components of a state machine. A simple diagram illustrating this concept is shown in Figure 3.

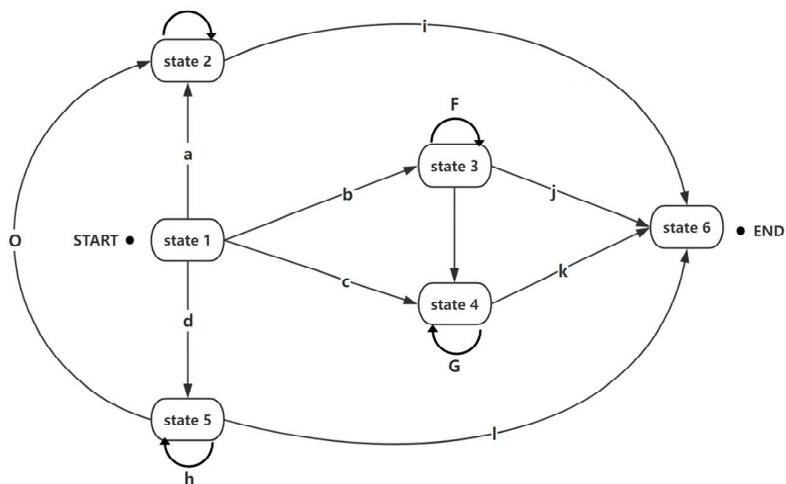
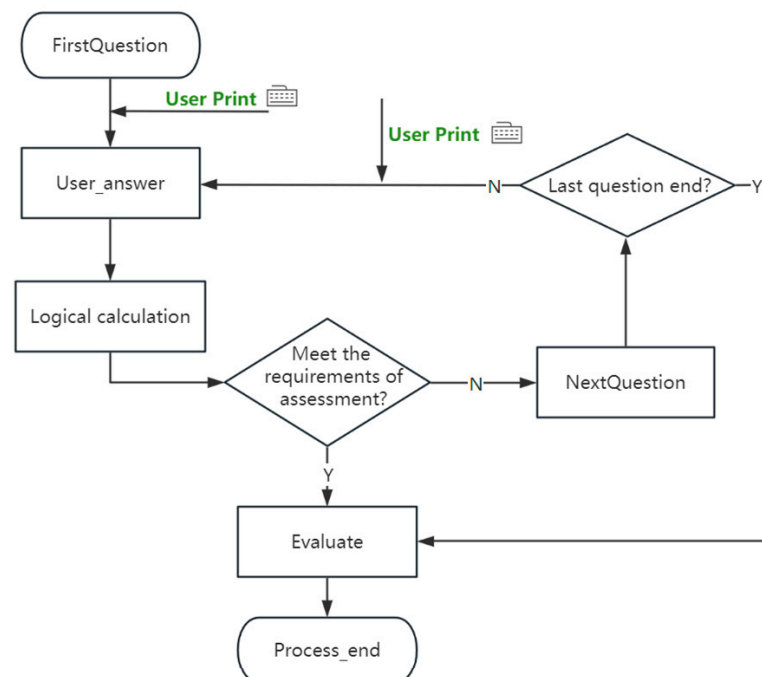


Figure 3. Working principle diagram of finite state machine.

Table 3. State-machine component table.

Component	Category Count
States	Different operational phases of the system, representing specific behaviors or attributes.
Events	External factors that trigger state changes, such as user input or sensor feedback.
Transitions	Rules for transitioning between states, associated with specific events and conditions.
Actions	Specific operations performed during state transitions.

In this study, we utilize the principles of the state-machine model to structure each module of M.I.N.I 6.0.0. Each user's response to a question triggers an event, determining the next question to be asked. This process continues until either the assessment criteria are met or all questions in the module are answered, followed by result evaluation. The flowchart for this process is depicted in Figure 4.

**Figure 4.** Program flowchart.

3.2.2. Generation of Personalized Medical Advice

In the age of big data, the increasing demand for personalized medical care matches well with the growing volume of medical data. The internet abounds with high-quality medical advice from medical experts, tailored to patients' conditions. By leveraging advanced computer technology, we can effectively gather and share this knowledge, providing robust support for generating personalized medical advice.

1. Information collection

The generation of personalized medical advice primarily depends on two types of information:

Basic Information: Includes gender, age, etc., collected before using the tool.

Special Information: Includes diagnostic results, personal experiences, drug allergy history, concurrent symptoms, etc., gathered during the use of the diagnostic tool.

2. Retrieval matching strategy

The evolution of text-retrieval technology has progressed from the initial Boolean retrieval models [36] to modern deep-learning-based models, such as the Transformer architecture [37]. These technologies have strong information-retrieval capabilities. However, in specific scenarios, particularly with limited data volumes, they may not fully meet precise retrieval requirements.

This study adopts a keyword-weight-matching method. Initially, patient-related information is collected via the M.I.N.I tool, and this is then cleaned and filtered to extract keywords and assign weights. Using these keywords, a preliminary match is performed on a pre-stored high-quality mental and psychological Q&A dataset in MySQL 5.7, forming subsets of question–answer pairs. These subsets are then ranked according to their weight scores, and the top N-ranked data are selected as the candidate set. The advantage of this method is its focus on key information, significantly enhancing the quality of the final data, while also substantially reducing the computational load of subsequent semantic matching. The process schematic is illustrated in Figure 5.

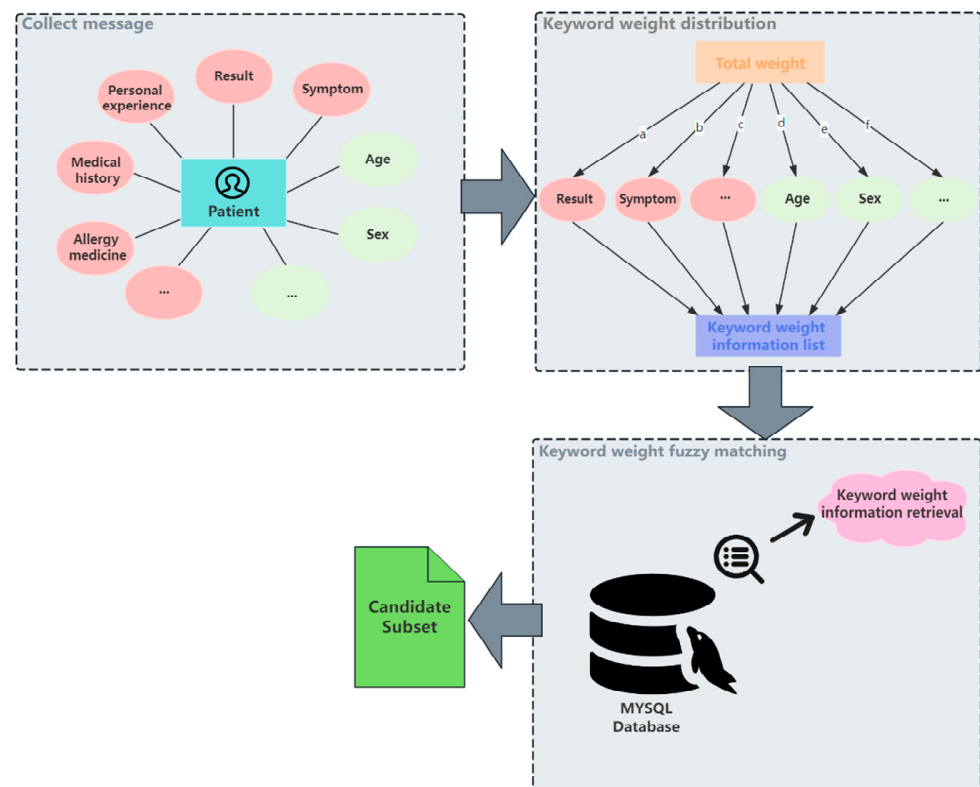


Figure 5. Keyword-weight matching. Letters a–f represent different weight values.

3. Semantic matching model

Semantic matching, the process of measuring textual similarity, has evolved from early neural network-based methods, like Word2Vec [38] and GloVe [39], to Long Short-Term Memory networks (LSTM/BiLSTM) [40]. In 2018, Devlin et al. [5] introduced the BERT model, based on the Transformer architecture, which excels in learning deep language representations from unlabeled text and exhibits remarkable multi-tasking capabilities. Its Cross-Encoder structure, depicted in Figure 6, assesses the relationship between two texts by providing a direct relational score, considering both inputs in their entirety. A variant of BERT, RoBERTa, shows superior performance in tasks like semantic sentence matching.

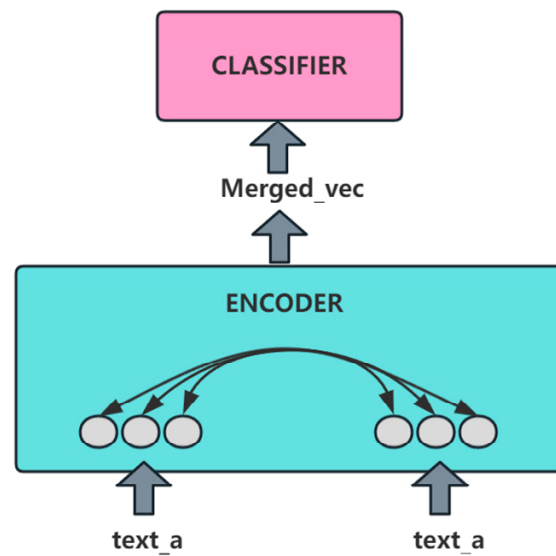


Figure 6. Cross-Encoder structure.

This study employs the RoBERTa model and a Cross-Encoder structure to develop a semantic-matching model, fine-tuned with knowledge in mental and psychological health. It evaluates the relevance between Text_Question (question) and Text_Answer (answer), as shown in Figure 7. Equations (1) and (2) illustrate the ReLU and Sigmoid activation functions, respectively, commonly used in neural networks and deep learning.

$$ReLU(x) = \max(0, x) \tag{1}$$

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

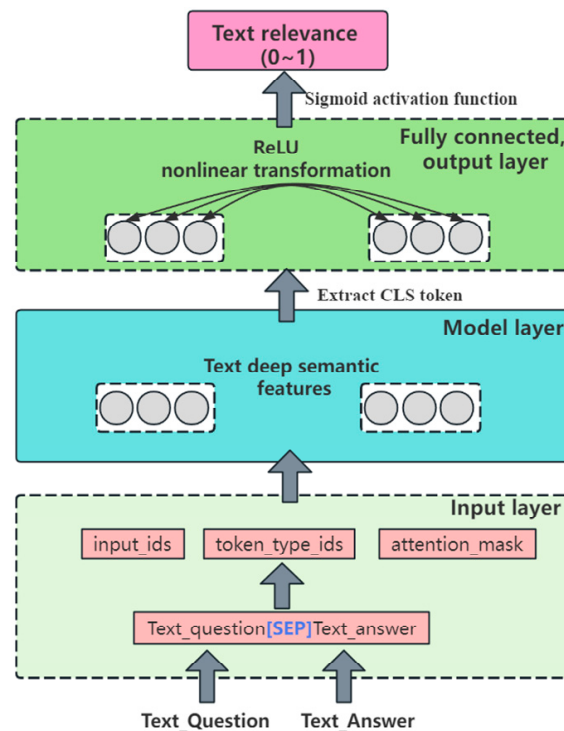


Figure 7. RoBERTa model based on Cross-Encoder.

The model architecture includes the following four main layers, as detailed in Table 4.

Table 4. Model structure.

Layer	Function	Description
Input	Text Preprocessing	Merges text and uses a tokenizer to convert it into the three standard inputs for RoBERTa.
Model	Semantic Extraction	Uses the pre-trained RoBERTa model to extract semantic information from text, outputting token embedding vectors.
Fully Connected	Feature Transformation	Extracts the CLS token vector from RoBERTa's output and performs a nonlinear transformation using Equation (1).
Output	Relevance Scoring	Uses a fully connected layer with the activation function in Equation (2) to map to a relevance score between 0 and 1.

After training and fine-tuning with mental and psychological disorder-related data, the model outputs a relevance score ranging from 0 (unrelated) to 1 (closely related), effectively measuring the connection between Text_Question and Text_Answer. This architecture is designed to deeply understand text content and capture complex relationships.

4. Generate medical advice

We propose an innovative framework for generating personalized medical advice, integrating natural language processing technology with expert medical knowledge. Initially, patient information collected via the M.I.N.I tool is used to form a preliminary subset of candidate advice using a keyword-weight-matching algorithm. This subset is then processed through the semantic-matching model, which scores the advice based on relevance, selecting the top M pieces. The refinement process includes data cleaning and extraction, filtering out various treatment methods, medications, and suggestions, and information filtering and analysis, identifying commonly and uniquely occurring treatment plans, removing duplicates, and retaining unique, relevant advice.

Finally, new templates are designed based on the filtered information to generate personalized medical advice containing core treatment plans. The detailed process of generating medical advice is illustrated in Figure 8.

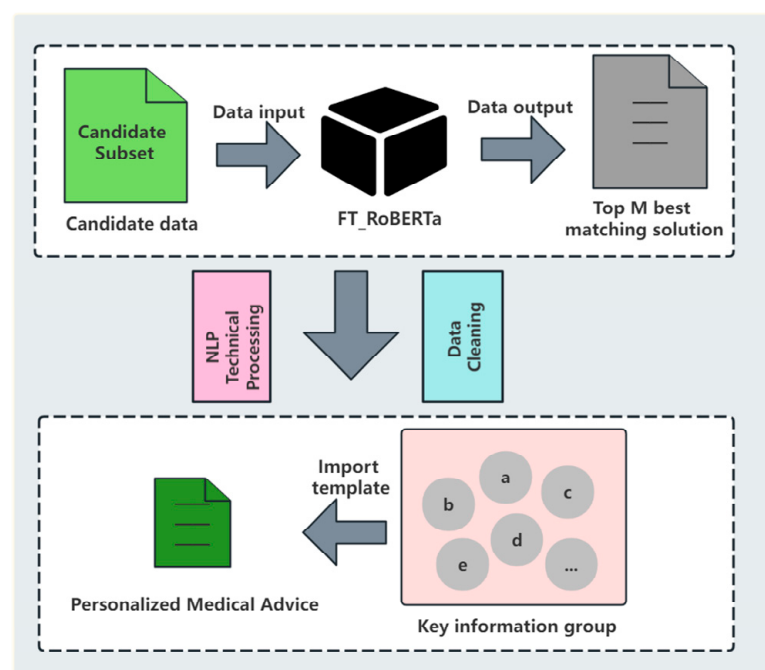


Figure 8. Medical advice generation process flowchart. Letters a–e represent different keywords.

Compared to advice generated solely by text generation technology, this method ensures data reliability and effectiveness, as all medical advice originates from human medical experts and undergoes professional review. It also guarantees the quality of the final medical advice by extracting key information from multiple sources, providing patients with more trustworthy and personalized advice.

3.2.3. Application of Medical-Knowledge Graph

A knowledge graph is a structured management tool for knowledge, with medical-knowledge graphs extensively used to integrate and organize a vast array of medical and health-related information. Leveraging the foundations of artificial intelligence and data science, these graphs aim to provide comprehensive medical knowledge and insights and have broad applications.

1. Knowledge generation

When encountering complex cases with scarce relevant Q&A examples, difficulties arise in matching appropriate medical advice due to a lack of precision. In such instances, the characteristics of medical-knowledge graphs can be utilized to provide universal and structured medical information as a supplement, achieving the goal of generating universally applicable medical advice. The process of knowledge generation is shown in Figure 9.

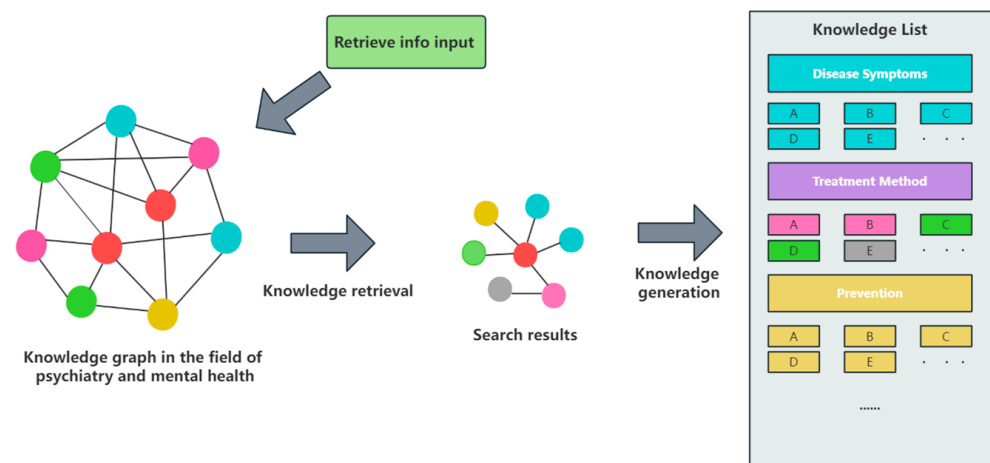


Figure 9. Knowledge generation process in the mental and psychological health domain-knowledge graph. Different colors represent entity categories, and different letters represent subcategories of the same entity.

Initially, the patient's disease assessment results obtained from the M.I.N.I auxiliary assessment tool are used as input. Then, multiple combinational searches are conducted in the mental and psychological health domain-knowledge graph. The search results include information on disease symptoms, treatment methods (including treatment modalities, medications, recommended diets), prevention measures, etc. Finally, after the classification and organization of knowledge, the information is compiled into a knowledge list.

4. Experiment Details

4.1. RoBERTa Model Fine-Tuning

To train an efficiently fine-tuned FT_RoBERTa model for accurately determining the relevance of question–answer pairs, the following steps were employed:

4.1.1. Data Preparation

- **Positive Examples:** We used original question–answer pairs as highly relevant instances, where each question was paired with its correct answer. These pairs were labeled as positive (e.g., label 1), representing high relevance.

- **Negative Examples:** To generate low-relevance or irrelevant instances, we deliberately disrupted the original question–answer pairings. This was done by randomly selecting mismatched answers for the questions, creating instances that were unrelated to the specific questions. These mismatched instances were assigned negative labels (e.g., label 0), representing low relevance or irrelevance.

4.1.2. Model Training

The model was fine-tuned using the labeled positive and negative examples, with the optimization guided by the cross-entropy loss function, as shown below:

$$L = -[y \cdot \log(p) + (1 - y) \cdot \log(1 - p)] \quad (3)$$

where:

- y represents the true label (0 or 1),
- p is the predicted probability from the model.

To prevent overfitting, we employed an early stopping strategy during training. Additionally, learning rates and batch sizes were adjusted to achieve optimal performance.

4.2. Data Processing in the System

Before introducing the FT_RoBERTa model for relevance calculation, the keyword information collected by the M.I.M.I. assistant evaluation tool was flexibly assigned different weights based on importance. The total weight was normalized to 1. Then, fuzzy keyword matching was applied to the question portion of the dataset stored in the MySQL database, scoring the matches based on the assigned weights. To ensure data reliability, only the top-ranked data with a score greater than 0.7 were selected as the candidate dataset.

After filling the keyword information into a question template to form a complete question, the FT_RoBERTa model was used to compute the relevance between this question and the candidate dataset. Instances with a relevance score greater than 0.85 were selected for the final dataset.

The final dataset underwent data cleaning using NLP techniques. Common and unique treatment methods, medications, and recommendations were extracted from the medical advice within the dataset. The filtered information was then applied to a pre-designed medical recommendation template to generate personalized medical suggestions.

5. Results and Discussion

5.1. Keyword-Weight-Matching Algorithm

By assigning different weights to keywords related to diagnostic results, symptoms, and patient basic information, we can effectively enhance the focus and accuracy of matching. For example, the keyword information group ['depression': 0.4, 'insomnia': 0.2, 'low spirits': 0.1, 'appetite': 0.1, 'motor inhibition': 0.1, 'female': 0.05, '24': 0.05] is weighted as shown in Figure 10, with an average precision score of 7.5.

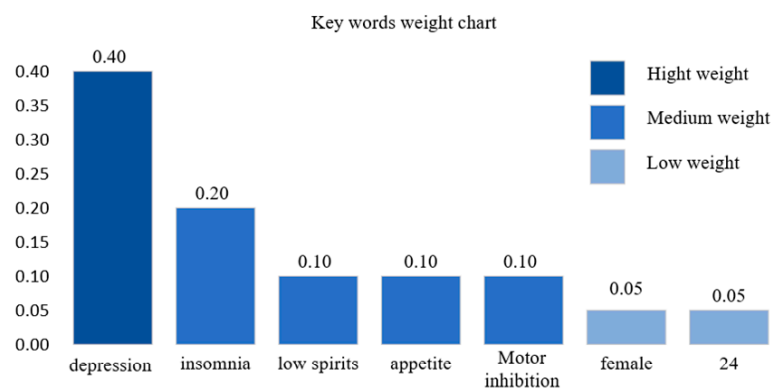


Figure 10. Weight distribution.

Taking 30 different keyword information groups to the MySQL database for fuzzy matching and scoring by weight, the top 15 Q&A pairs averaged 0.7 (out of 1), as shown in Figure 11, indicating rich content and high quality.

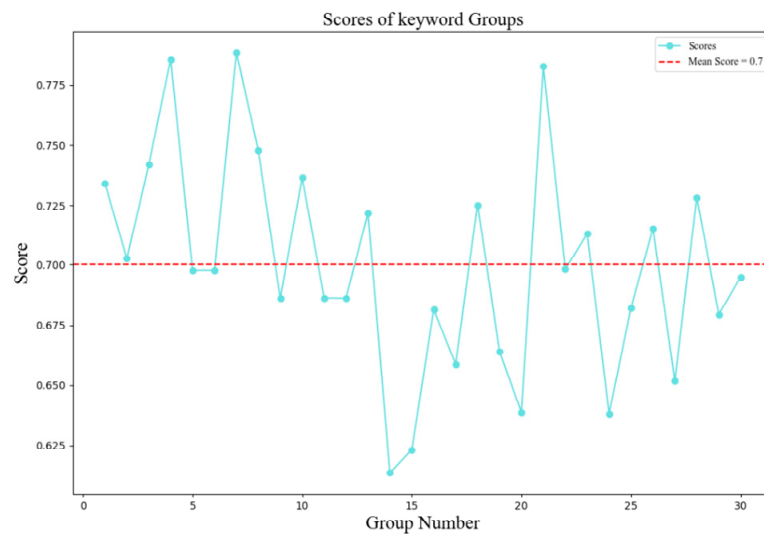


Figure 11. Weight-matching score.

5.2. Semantic-Matching Models

The assessment metrics for this experiment were relevance and precision. We used a pre-trained RoBERTa-based model to generate embeddings for template questions and for each answer. For the template question (Text1) and answers (Text2. . .n) in the dataset, corresponding vector representations v_1 and $v_1. . .n$ were generated. The relevance between these vectors was measured by calculating the Manhattan Distance, Euclidean Distance, and Cosine Similarity between them. The formulas for these calculations are as follows: (4)–(6), where P and Q represent two points and p_i and q_i are their coordinates on the i th dimension. Manhattan Distance sums the absolute differences in each dimension, reflecting the cumulative difference between two points; the larger the value, the greater the difference. Euclidean Distance calculates the “straight-line” distance between two points, representing the spatial distance; the larger the value, the further apart the points. Cosine Similarity calculates the angle between vectors, with a larger value indicating greater similarity.

$$D_{\text{manhattan}}(P, Q) = \sum_{i=1}^n |p_i - q_i| \quad (4)$$

$$D_{\text{euclidean}}(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (5)$$

$$\text{Cosine}(\vec{v}_1, \vec{v}_2) = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \times \|\vec{v}_2\|} \quad (6)$$

Precision is the ratio of correctly retrieved relevant Q&A pairs ($U1$) to all retrieved Q&A pairs ($U2$). A Q&A pair is defined as correctly retrieved if its cosine similarity with the template question is greater than 0.85, calculated using Formula (7):

$$\text{Accuracy}(U1, U2) = \frac{U1}{U2} \quad (7)$$

The candidate data obtained from the keyword-weight-matching step are further processed using the semantic-matching model, significantly enhancing the precision and quality of answers. For instance, with the template question “[Gender: female, age: 24], suffering from [depression] with symptoms of [insomnia, low spirits, loss of appetite, motor inhibition], what treatment should be taken? What medication? Any suggestions?”,

we conducted comparative experiments on several advanced models with 15 different template questions and their corresponding candidate answer datasets. The average results are presented in Table 5.

Table 5. Comparative experimental results of different methods.

Model	Manhattan Distance	Euclidean Distance	Cosine Value
BERT	158	7.32	0.84
RoBERTa	52	4.26	0.88
ALBERT [41]	278	6.67	0.73
XLNet [42]	767	69.72	0.87
FT_RoBERTa	50	4.46	0.90

Figures 12–14 illustrate the line graphs showing the scores of various models across 15 sets of data for different metrics. From Figure 14, it is clear that RoBERTa’s cosine similarity score (0.88) is the closest to XLNet’s (0.87), and both significantly outperform BERT (0.84) and ALBERT (0.73). Figures 12 and 13 show that RoBERTa maintains lower scores for Manhattan distance (52) and Euclidean distance (4.62), while XLNet’s scores for Manhattan distance (767) and Euclidean distance (69.72) remain extremely high. Based on these analyses, we conclude that RoBERTa outperforms the other models in terms of both vector similarity and spatial distance when processing our dataset, demonstrating its superior performance. This further indicates that RoBERTa exhibits higher precision and robustness in capturing semantic similarity.

Specifically, RoBERTa’s deeper pre-training and more optimized training strategies enable it to better understand and represent complex medical question–answer pairs, leading to higher relevance and accuracy in semantic-matching tasks. Furthermore, the method employed in this study, which incorporates a keyword-weight-matching algorithm, not only significantly reduces computational costs compared to previous semantic retrieval models but also enhances the focus and accuracy of the matching process. This strategy prioritizes highly relevant question–answer pairs during the initial candidate data screening, providing more precise input for subsequent semantic matching. This is one of the key reasons behind the superior overall performance of our approach.

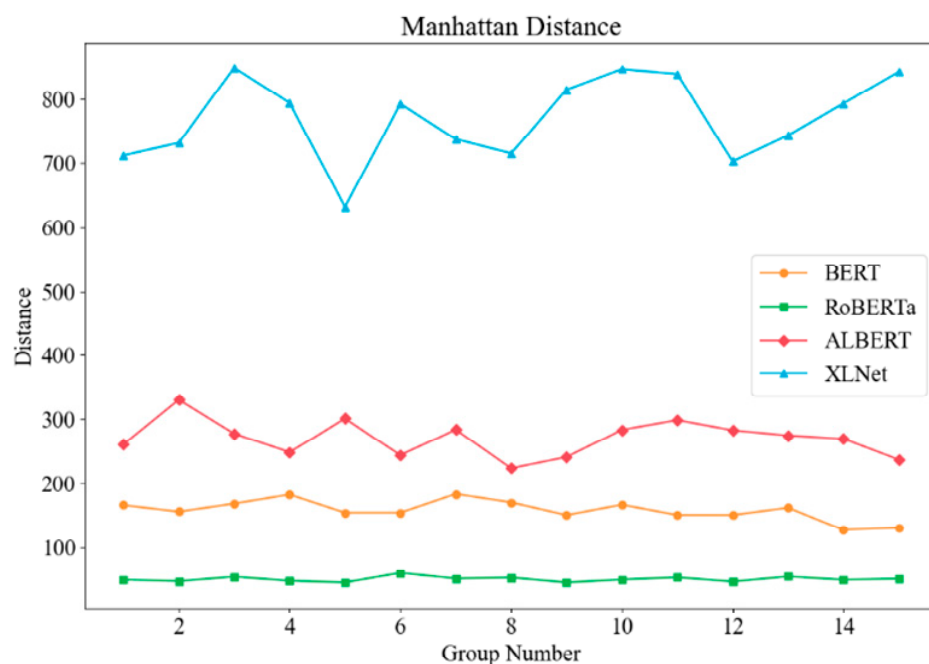


Figure 12. Manhattan distance score.

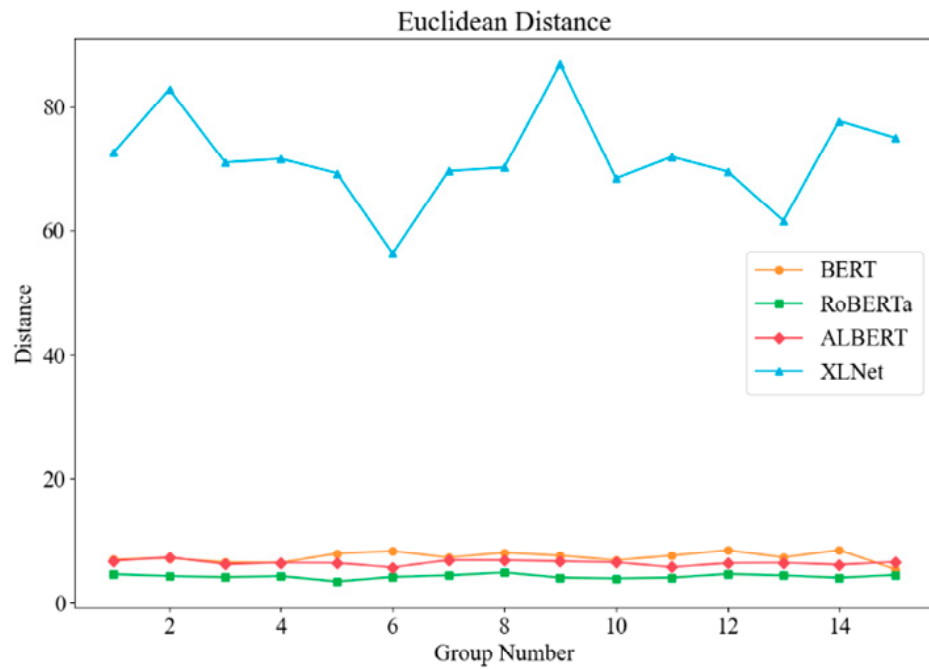


Figure 13. Euclidean distance score.

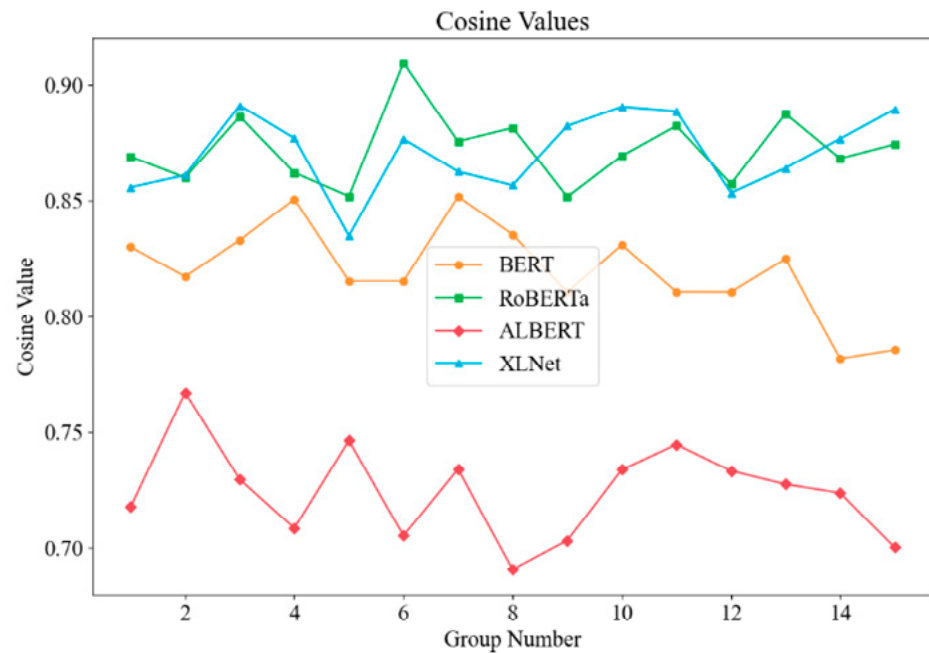


Figure 14. Cosine value score.

Using domain-specific data from the mental health field, we fine-tuned the RoBERTa model to create a new FT_RoBERTa model, which was then tested on the same 15 datasets. As shown in the scatter plot in Figure 15, the fine-tuned FT_RoBERTa model achieved an average cosine similarity score of 0.9 for semantic relevance, an improvement of 0.02 compared to the original RoBERTa model (0.88). According to the calculation in Equation (7), both models achieved a precision score of 0.87, indicating that the fine-tuned FT_RoBERTa model performed well and showed a slight improvement.

Several groups of data selected through the semantic-matching model are combined with advanced natural language processing (NLP) technology to generate personalized medical advice. This advice includes treatment suggestions, recommended medications, lifestyle advice, and mental health guidance, all highly accurate, complete, and consistent,

matching the patient’s personal circumstances and derived from solutions provided by experts in the relevant medical field.

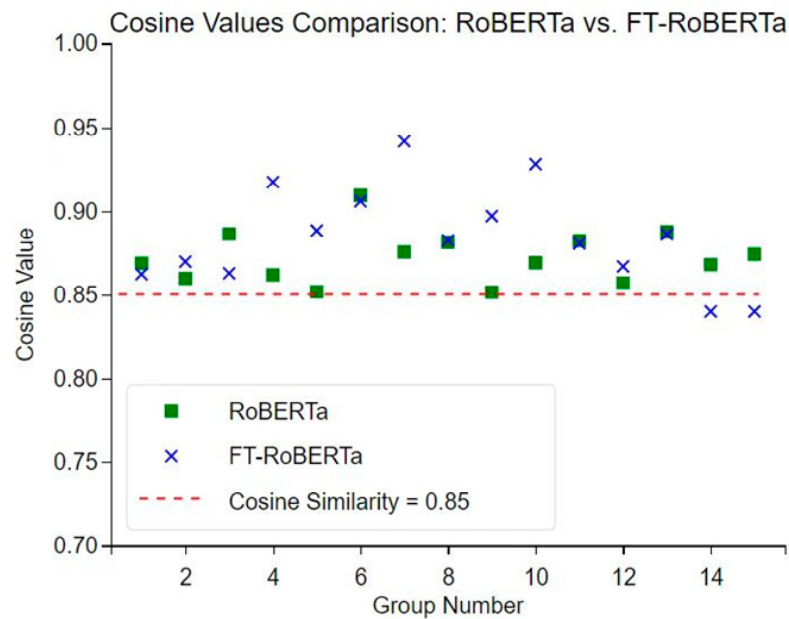


Figure 15. RoBERTa vs. FT-RoBERTa.

5.3. Construction of Medical-Knowledge Graph

To support knowledge retrieval and the generation of universally applicable medical advice in the field of mental and psychological health, we specifically constructed a domain-knowledge graph. The mental and psychological health domain medical-knowledge graph developed in this experiment, as illustrated in Figures 16–19, displays information related to entity categories, entity attributes, and entity relationship types. It includes five entity types with 2430 entities, eight attribute types, and nine relationship types with 5140 relationships, showing a well-developed initial scale.

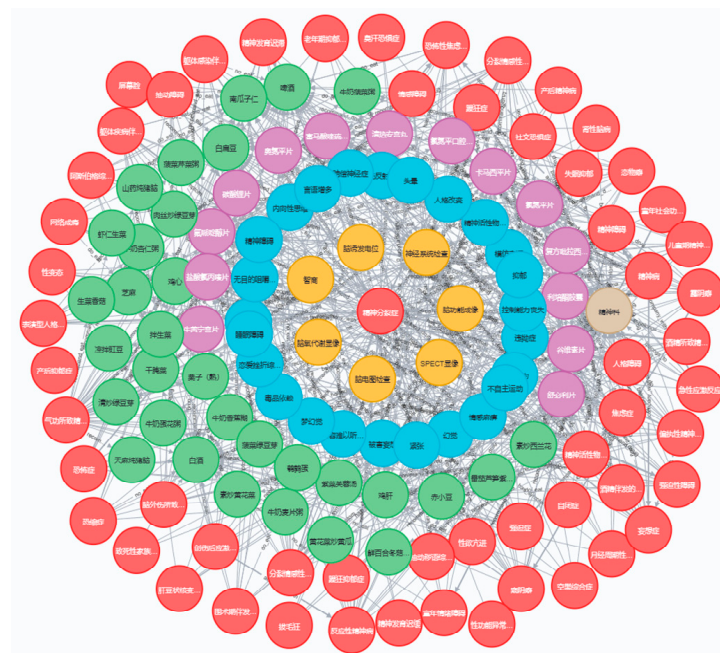


Figure 16. Mental and psychological health domain medical-knowledge graph.

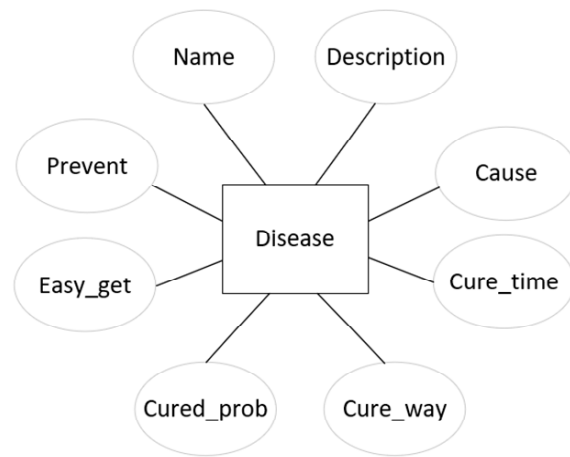


Figure 17. Entity-related info.

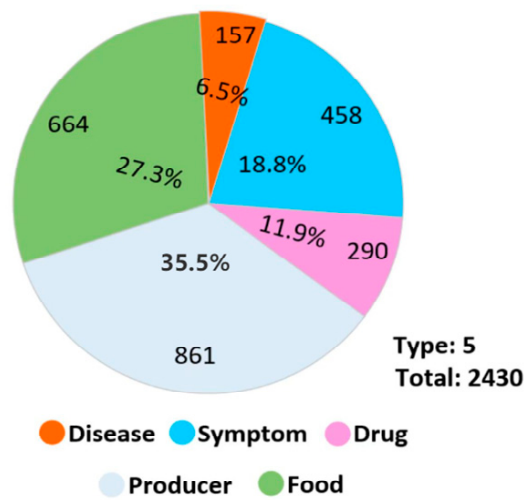


Figure 18. Attribute-related info.

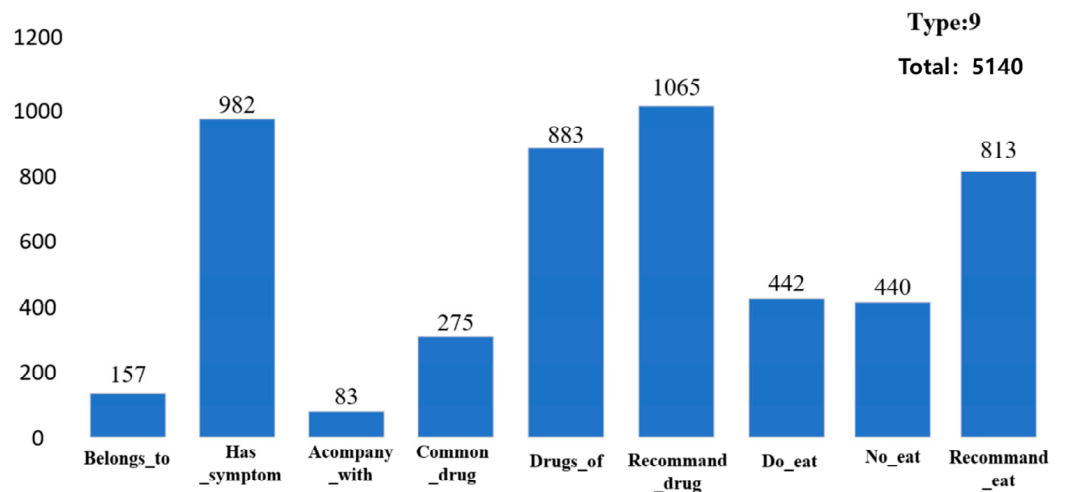


Figure 19. Entity relationship-related Info.

This knowledge graph enables the retrieval of extensive knowledge related to mental and psychological illnesses through specific query templates. For diseases such as depression and mania, the knowledge graph provides detailed information, including symptoms, comorbidities, susceptible populations, treatment duration, cure probabilities, treatment methods, medication use, recommended diets, foods to avoid, and preventive measures.

Figures 20 and 21 show the effectiveness of using query statements to retrieve data from the knowledge graph and filling the results into prepared templates, effectively generating universally applicable medical advice.



Figure 20. Depression knowledge generation.



Figure 21. Mania knowledge generation.

5.4. Limitations and Future Work

While our methods have shown promising results on a subset of the Huatuo-26M dataset, several limitations exist. Firstly, although models like BERT and RoBERTa capture deep semantic meanings, they are not specifically tailored for the mental and psychological health domain, potentially limiting precision. Future work includes fine-tuning these models with larger domain-specific datasets to enhance accuracy. Secondly, the use of the Cross-Encoder structure with RoBERTa incurs high computational costs and poses challenges in parallelization. Exploring more efficient network architectures is planned to reduce computational overheads. Thirdly, the dataset and the mental health-knowledge graph used are not comprehensive. Expanding these resources and testing with more diverse datasets are future objectives.

Importantly, our methods have not yet been validated in real clinical settings, which may affect clinician adoption due to concerns about clinical validity. To address this, future work will focus on the following:

- **Clinical Validation:** Collaborating with mental health professionals to deploy our tools in clinical environments and compare their performance against standard clinical practices.
- **Feedback Collection:** Gathering insights from clinicians and patients on usability and practical utility.
- **Evaluating Key Metrics:** Assessing improvements in assessment time, diagnostic accuracy, and patient satisfaction.

By undertaking these steps, we aim to demonstrate the practical applicability of our work, addressing concerns about clinical validity and encouraging adoption in real-world settings.

6. Conclusions

This paper developed a strategy for auxiliary assessment and personalized medical-advice generation in the mental and psychological health domain. Initially, an effective auxiliary assessment tool using M.I.N.I was constructed, enabling the automatic and intelligent collection of patient data and providing preliminary diagnostic results. Subsequently, a keyword-weight-matching retrieval algorithm and an improved semantic-matching model based on RoBERTa were proposed. This approach enhances data quality by focusing on key information through keyword weighting, and the semantic model filters the most relevant data. A medical-advice generation template was then designed, with NLP technology extracting key information from various pieces of advice. After cleaning, calculation, and filtering, this resulted in personalized medical advice. Additionally, a method using a knowledge graph was introduced to generate universally applicable advice in cases of data insufficiency. The proposed algorithms and models were validated on a public benchmark dataset and demonstrated good performance and reliability in generating medical advice. This work aims to alleviate resource distribution issues in mental and psychological health care and enhance diagnostic and treatment efficiency, ultimately improving mental health services' quality and efficiency. Future efforts will focus on optimizing algorithm performance and expanding the knowledge graphs for broader clinical application.

Author Contributions: Conceptualization, Y.W. and J.T.; Methodology, Y.W. and H.X.; Software, H.X.; Validation, H.X.; Formal analysis, R.C.; Investigation, L.G.; Data curation, S.C.; Writing—original draft, Y.W. and H.X.; Writing—review & editing, L.G. and J.T.; Visualization, F.W.; Supervision, Y.L.; Project administration, L.C.; Funding acquisition, J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Key RD Program of China under Grant no.2022YFE010300; in part by the Natural Science Foundation of Hunan Province under Grant no.2021JJ50050; in part by the Scientific Research Fund of Hunan Provincial Education Department under Grant no.22A0422; in part by the University IUR Innovation Foundation of China under Grant no.2022IT052; National Natural Science Foundation of China (62106074, Research on influence mechanism of research collaboration networks from the network evolution), Ling-Jiao Chen.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/FreedomIntelligence/Huatuo-26M>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. *World Mental Health Report: Transforming Mental Health for All*; World Health Organization: Geneva, Switzerland, 2022. Available online: <https://www.who.int/publications/i/item/9789240049338> (accessed on 6 February 2023).
2. Vigo, D.; Thornicroft, G.; Atun, R. Estimating the true global burden of mental illness. *Lancet Psychiatry* **2016**, *3*, 171–178. [[CrossRef](#)] [[PubMed](#)]
3. World Health Organization. *WHO Report Highlights Global Shortfall in Investment in Mental Health*; World Health Organization: Geneva, Switzerland, 2023. Available online: <https://www.who.int/news/item/08-10-2021-who-report-highlights-global-shortfall-in-investment-in-mental-health> (accessed on 6 December 2023).
4. Patel, V.; Saxena, S.; Lund, C.; Thornicroft, G.; Baingana, F.; Bolton, P.; Chisholm, D.; Collins, P.Y.; Cooper, J.L.; Eaton, J.; et al. The Lancet Commission on global mental health and sustainable development. *Lancet* **2018**, *392*, 1553–1598. [[CrossRef](#)] [[PubMed](#)]
5. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.

6. Heehan, D.V.; Lecrubier, Y.; Sheehan, K.H.; Amorim, P.; Janavs, J.; Weiller, E.; Hergueta, T.; Baker, R.; Dunbar, G.C. 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. S20), 22–33; quiz 34–57. [[PubMed](#)]
7. Jones, S.R.; Fernyhough, C. A new look at the neural diathesis–stress model of schizophrenia: The primacy of social-evaluative and uncontrollable situations. *Schizophr. Bull.* **2007**, *33*, 1171–1177. [[CrossRef](#)]
8. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Publishing, Inc.: Washington, DC, USA, 2013.
9. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders*; World Health Organization: Geneva, Switzerland, 1993.
10. Chakraborty, N.; Ali, A.; Alakpa, C. Diagnostic categories of mental illness in a rural African setting—the mhGAP experience in Edawu (Nigeria). *Int. J. Ment. Health* **2021**, *50*, 91–95. [[CrossRef](#)]
11. Arrow, K.; Resnik, P.; Michel, H.; Kitchen, C.; Mo, C.; Chen, S.; Espy-Wilson, C.; Coppersmith, G.A.; Frazier, C.; Kelly, D. Evaluating the Use of Online Self-Report Questionnaires as Clinically Valid Mental Health Monitoring Tools in the Clinical Whitespace. *Psychiatr. Q.* **2023**, *94*, 221–231. [[CrossRef](#)]
12. Pan, Y.; Chen, Q.; Peng, W.; Wang, X.; Hu, B.; Liu, X.; Chen, J.; Zhou, W. MedWriter: Knowledge-Aware Medical Text Generation. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; International Committee on Computational Linguistics: New York, NY, USA, 2020; pp. 2363–2368.
13. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.-Y. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409. [[CrossRef](#)]
14. Kumichev, G.; Blinov, P.; Kuzkina, Y.; Goncharov, V.; Zubkova, G.; Zenovkin, N.; Goncharov, A.; Savchenko, A. MedSyn: LLM-Based Synthetic Medical Text Generation Framework. In *Machine Learning and Knowledge Discovery in Databases*; Applied Data Science Track; ECML PKDD 2024; Lecture Notes in Computer Science; Bifet, A., Krilavičius, T., Miliou, I., Nowaczyk, S., Eds.; Springer: Cham, Switzerland, 2024; Volume 14950. [[CrossRef](#)]
15. Hasani, A.M.; Singh, S.; Zahergivar, A.; Ryan, B.; Nethala, D.; Bravomontenegro, G.; Mendhiratta, N.; Ball, M.; Farhadi, F.; Malayeri, A. Evaluating the performance of Generative Pre-trained Transformer-4 (GPT-4) in standardizing radiology reports. *Eur. Radiol.* **2023**, *34*, 3566–3574. [[CrossRef](#)]
16. Thakur, N.; Reimers, N.; Rucklé, A.; Srivastava, A.; Gurevych, I. BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models. *arXiv* **2021**, arXiv:2104.08663.
17. Zhu, Y.; Yuan, H.; Wang, S.; Liu, J.; Liu, W.; Deng, C.; Dou, Z.; Wen, J. Large Language Models for Information Retrieval: A Survey. *arXiv* **2023**, arXiv:2308.07107.
18. Tang, X.; Luo, Y.; Xiong, D.; Yang, J.; Li, R.; Peng, D. Short text matching model with multiway semantic interaction based on multi-granularity semantic embedding. *Appl. Intell.* **2022**, *52*, 15632–15642. [[CrossRef](#)]
19. Cai, Y.; Fan, Y.; Guo, J.; Sun, F.; Zhang, R.; Cheng, X. Semantic Models for the First-Stage Retrieval: A Comprehensive Review. *ACM Trans. Inf. Syst.* **2021**, *54*, 66.
20. Nigam, S.; Goel, N. Nigam@COLIEE-22: Legal Case Retrieval and Entailment using Cascading of Lexical and Semantic-based models. *arXiv* **2022**, arXiv:2204.07853.
21. Zou, Y.; Liu, H.; Gui, T.; Wang, J.; Zhang, Q.; Tang, M.; Li, H.; Wang, D. Divide and Conquer: Text Semantic Matching with Disentangled Keywords and Intents. *arXiv* **2022**, arXiv:2203.02898.
22. Gupta, S.; Mishra, A. Publisher Side Profit Optimization Using Adaptive Keyword Weighted Sponsored Search Technique. *J. Web Eng.* **2022**, *21*, 1449–1469. [[CrossRef](#)]
23. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **1907**, arXiv:1907.11692.
24. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.1008.
25. Li, Z.; Yang, N.; Wang, L.; Wei, F. Learning Diverse Document Representations with Deep Query Interactions for Dense Retrieval. *arXiv* **2022**, arXiv:2208.04232.
26. Zhu, Y.; Li, Y.; Cui, Y.; Zhang, T.; Wang, D.; Zhang, Y.; Feng, S. A Knowledge-Enhanced Hierarchical Reinforcement Learning-Based Dialogue System for Automatic Disease Diagnosis. *Electronics* **2023**, *12*, 4896. [[CrossRef](#)]
27. Smith, J.; Johnson, A.B. Medical knowledge graphs: An in-depth review. *Big Data Min. Anal.* **2023**, *6*, 201–217. [[CrossRef](#)]
28. Wu, X.; Duan, J.; Pan, Y.; Li, M. Medical Knowledge Graph: Data Sources, Construction, Reasoning, and Applications. *Big Data Min. Anal.* **2023**, *6*, 113–128. [[CrossRef](#)]
29. Chandak, P.; Huang, K.; Zitnik, M. Building a knowledge graph to enable precision medicine. *Sci. Data* **2023**, *10*, 67. [[CrossRef](#)] [[PubMed](#)]
30. Li, J.; Wang, X.; Wu, X.; Zhang, Z.; Xu, X.; Fu, J.; Tiwari, P.; Wan, X.; Wang, B. Huatuo-26m, a large-scale chinese medical qa dataset. *arXiv* **2023**, arXiv:2305.01526, 2023.
31. Li, W.; Yu, L.; Wu, M.; Liu, J.; Hao, M.; Li, Y. DoctorGPT: A Large Language Model with Chinese Medical Question-Answering Capabilities. In Proceedings of the 2023 International Conference on High Performance Big Data and Intelligent Systems (HDIS), Macau, China, 6–8 December 2023; pp. 186–193. [[CrossRef](#)]
32. Ye, Q.; Liu, J.; Chong, D.; Zhou, P.; Hua, Y.; Liu, A. Qilin-med: Multi-stage knowledge injection advanced medical large language model. *arXiv* **2023**, arXiv:2310.09089.

33. Yang, S.; Zhao, H.; Zhu, S.; Zhou, G.; Xu, H.; Jia, Y.; Zan, H. Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 19368–19376.
34. Jung, S.; Han, S.-Y.; Kang, B. Design of a Variable-Mode Sync Generator for Implementing Digital Filters in Image Processing. *J. Inst. Korean Electr. Electron. Eng.* **2023**, *27*, 273–279.
35. Islam, K.Z.; Murray, D.; Diepeveen, D.; Jones, M.G.K.; Sohel, F. LoRa-based outdoor localization and tracking using unsupervised symbolization. *Internet Things* **2023**, *25*, 101016. [[CrossRef](#)]
36. Steven, W. Boolean operations. In *Information Retrieval Data Structures & Algorithms*; Prentice-Hall, Inc.: Saddle River, NJ, USA, 1992; ISBN 0-13-463837-9.
37. Aswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
38. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
39. Pennington, J.; Socher, R.; Manning, C.D. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014.
40. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
41. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *arXiv* **2019**, arXiv:1909.11942.
42. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems* (Vol. 32). *arXiv* **2019**, arXiv:1906.08237.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.