Construction of Emotional Evolution Map of Online Texts on Medical Student Burnout Integrating BERT and LDA Topic Models

Jiayang Li
School of Nursing, Hebi Polytechnic, Hebi 458030, Henan, China
Corresponding author: Jiayang Li (e-mail: ljy1234567892025@126.com)

ABSTRACT Large-scale online medical communities provide continuous textual signals that reflect the psychological states of medical students; however, conventional survey-based approaches have limited capability for tracking the temporal evolution of burnout-related emotions. To address this issue, a hybrid analytical framework integrating BioBERT-based sentiment recognition and latent topic discovery is developed for dynamic emotional evolution modeling. Burnoutrelated texts collected from online medical forums are first transformed into contextual semantic representations through a domain-adapted BioBERT model, enabling fine-grained identification of positive emotion, negative emotion, emotional exhaustion, and depersonalization states. Subsequently, an LDA-based topic mining module is employed to extract latent stress-related themes and quantify topic distributions. By fusing sentiment trajectories with topic intensity information in the temporal domain, a multidimensional emotional evolution map is constructed to characterize the dynamic interactions between emotional states and discussion topics. Experimental results on 137,430 medical-student text records demonstrate that the proposed framework achieves 92.7% classification accuracy with a macro-F1 score of 0.904, while the optimized topic model reaches a coherence score of 0.632. Temporal analysis reveals distinct evolution patterns across internship stress, academic anxiety, career identity, and institutional support topics. The proposed framework provides an effective data-driven methodology for large-scale sentiment monitoring, temporal topic analysis, and dynamic behavioral-state assessment in complex online information environments.

INDEX TERMS BioBERT; latent Dirichlet allocation; sentiment evolution; topic modeling; temporal text analytics

I. INTRODUCTION

BURNOUT among medical students has long affected the quality of training medical talent and long-term viability for both health care services and the medical education system. The main causes of burnout include workload, internship intensity, and conflict over identity as a professional; Recent studies in occupational ergonomics also suggest that the physiological burden imposed by medical textiles, such as long-term wearing of non-breathable protective clothing and masks, can exacerbate the psychological exhaustion of medical interns during high-pressure clinical tasks.These can be even more complicated due to experienced stressors related to public health crises. Data used for much current research on medical student burnout is primarily based on questionnaires or collected cross-sectionally, and, as such, the analytical approach to understanding the progression and documentation of student burnout remains largely static and merely descriptive of the concept without offering any emotional or feasibility-based means of developing an understanding of its development through time and across sources of stress. The increasing volume of shared information via the Internet in the form of text in online medical communities makes it essential to develop a systematic method to assess, identify, and model the emotions associated with burnout across a variety of contexts and receiver characteristics.

Burnout among medical students is an important area of research because of its implications for training future doctors, mental health, and stress-related non-compliance. Almutairi H [1] performed a systematic review and meta-analysis that demonstrated that burnout is prevalent among medical students and varies significantly based on culture and educational systems. Therefore, targeted monitoring and intervention strategies should be developed to identify and support students experiencing burnout in medical school. Additionally, there are numerous studies that have examined the factors that contribute to burnout including: academic workload, clinical rotation requirements, professional identity, and lack of supportive systems etc. [2,3]. Especially during the COVID-19 (Corona Virus Disease 2019) pandemic, the mental well-being of medical students was further impacted, and burnout and anxiety increased significantly [4]. In terms of intervention and relief mechanisms, Zuniga D [5], Roman-Calderon J P [6] and Daud N [7] pointed out that mindfulness training, self-care and emotional intelligence cultivation can effectively alleviate burnout and promote psychological well-being. In addition, a sense of belonging and supportive learning environment have also been shown to be negatively correlated with burnout [8]. Cross-cultural studies have further revealed that medical students from different countries exhibit group differences in burnout performance, psychological symptoms, and substance use behavior [9-11]. These studies mostly rely on questionnaires and scales, lacking continuous tracking of dynamic, fine-grained emotional evolution, especially real-time analysis based on natural language data. Researchers have also continued to investigate how certain groups and their traits can affect the development of burnout. For example, a nationwide survey conducted by Meeks L M [12] demonstrated a substantial relationship between the procedural barriers that students with disabilities face and their level of burnout. This study illustrates the need for an educational environment that does not contain any procedural barriers, to decrease burnout. In addition, Lee D [13] studied the relationship between resilience and the level of student burnout. It was determined that students that possess resilient traits tend to experience a lower level of burnout than those that do not possess these resilient traits. However, the relationship between resilience and burnout varied according to the demographic characteristics of individual students and provided a foundation for developing future interventions based on psychological traits. Although these two studies have improved understanding of the various influences on student experience, they primarily relied on cross-sectional survey data, which makes capturing the ever-changing interactions between stress and emotional experiences difficult.

In terms of topic modeling, traditional LDA models are widely used to extract implicit topics from texts due to their unsupervised characteristics, but their performance is limited in short texts and domain-specific contexts. To this end, scholars have tried to combine context-aware embedding models such as BERT with LDA to improve topic consistency and semantic discriminability. For example, George L [14] proposed a framework that integrated clustering and BERT, which significantly improved the topic modeling effect on multiple public datasets; Riaz A [15] compared traditional and Transformer-based topic modeling methods and confirmed that BERT embedding can effectively improve topic coherence and diversity. Similarly, Khadija M A [16] integrated LDA and BERT embedding in Indonesian customer complaint analysis to enhance topic interpretability and sentiment analysis. In terms of technical optimization, Wijanto M C [17] systematically explored the hyperparameter tuning strategy of BERT model in scientific literature topic modeling, providing empirical reference for improving the adaptability of the model on professional corpora. Datchanamoorthy K [18] further combined BERT with probabilistic topic model for text clustering, verifying the advantages of the hybrid method in semantic representation. In order to improve the semantic richness and structure of topics, Taher H A [19] proposed to integrate named entity recognition and LDA to enhance the distinguishability of topic modeling through entity information; similarly, Rawat A J [20] successfully applied BERT-enhanced topic modeling in legal document analysis, improving the topic discovery effect of complex professional texts. In cross-domain applications, Chen C [21] explored a general method of summarizing design requirements using topic modeling in complex designs; Bruno M [22] achieved fine-grained sentiment analysis of social emotions in economic public opinion by integrating word embedding and topic model, demonstrating the potential of this method in dynamic social mentality monitoring. These studies further demonstrate that combining deep semantic representation with unsupervised topic discovery can reveal the interactive evolution of emotion and topic in text more meticulously, providing a methodological basis for dynamic psychological monitoring. In specific domain applications, Wang Y [23] used BERT and neural topic modeling to enhance the recognition and fusion of emerging technology sentences in the field of nanomedicine; Karabacak M [24] used NLP and topic modeling to track the research trend of drug-resistant epilepsy; Song B [25] combined BERT and semantic analysis to identify emerging technology topics in multi-domain patents. These studies have verified the effectiveness of BERT and topic model fusion in professional text mining. In the direction of mental health and public opinion analysis, Osvath M [26] used BERT topic modeling to analyze patient experience narratives; Gomes G B B [27] used a localized BERT model to enhance the topic modeling contribution of Brazilian social media. These works demonstrate that combining deep semantic representation with unsupervised topic discovery can reveal the interactive evolution of emotion and topic in text more meticulously, providing a methodological basis for dynamic psychological monitoring. However, most existing studies focus on static topic extraction or coarse-grained emotion classification, lacking a visual integrated analysis of the evolutionary patterns of complex psychological states such as job burnout in the temporal dimension.

Existing research largely focuses on static analysis of factors influencing burnout among medical students, or on classifying text themes and emotions at single points in time, lacking a dynamic framework that comprehensively presents emotional changes and thematic evolution over time. The inability to grasp the full range of burnout's complexities, as well as its ongoing development over time, hinders the capacity to provide prompt monitoring and intervention. Using medical student-initiated internet text submissions, this research designs an emotional theme evolution map that merges a domain-pre-trained model with theme analysis to systematically reveal the stage change and correlated emotional themes associated with burnout at different times. Through the integration of semantics and themes, this study has made the transition from static analysis to dynamic evolutionary analysis and provides a scalable data model that can be used to dynamically monitor and provide focused interventions into individuals' psychological states.This analytical framework also provides a theoretical basis for evaluating the impact of the clinical working environment on the psychological health of practitioners. Future research could explore how optimizing the comfort of medical apparel can serve as a physical intervention to mitigate the cumulative stress identified in the emotional evolution map.

II. METHODOLOGY FOR CONSTRUCTING EMOTIONAL EVOLUTION MAPS

A. DATA COLLECTION AND PREPROCESSING

The data foundation for this paper comes from online communities active among medical students on the Chinese internet. To comprehensively obtain authentic texts related to professional burnout, data collection primarily targets two platforms: "DXY Forum," a gathering place for professional medical educators and practitioners, and "Medlive," a social platform for medical students and young physicians. The collection process uses the Python-based Scrapy framework, specifically crawling post titles, text content, posting times, and anonymized user identifiers related to keywords such as "academic pressure," "internship fatigue," "career confusion," and "burnout" from the aforementioned platforms. The crawling strategy strictly adheres to the robots.txt protocol of the target websites, and polite access intervals are set to avoid burdening the servers. The data collection period spans from January 2019 to December 2023, aiming to cover a complete medical education cycle and possible external event impacts. The final initial corpus contains approximately 125,000 text data entries. After rigorous preprocessing, including noise removal, deduplication, and quality filtering, along with data augmentation techniques such as synonym replacement and back-translation to address class imbalance, the refined corpus expanded to 137,430 valid text entries. This augmented dataset served as the foundation for all subsequent analyses.

The obtained raw text contains a large amount of noise information irrelevant to semantic analysis, requiring text cleaning. First, HTML (Hyper Text Markup Language) tags, advertising links, and duplicate content are removed. Then, all user identifiers undergo uniform hash encryption to achieve complete anonymity and protect user privacy. To accurately retain the substantive content expressing sentiment and theme, a series of regular expression rules are used to filter out emojis, URLs (Uniform Resource Locators), special characters, and pure numeric sequences. Next, text normalization is performed, using the Jieba word segmentation tool for word segmentation, and combining it with the Harbin Institute of Technology stop word list and a custom-defined general medical stop word list (high-frequency words with low relevance to burnout sentiment analysis) for stop word filtering. Based on this, to improve the quality of subsequent topic models, part-of-speech filtering is introduced, retaining only nouns, verbs, adjectives, and adverbs. To transform the text into a mathematical representation that can be processe d by the LDA topic model, text vectorization is require d .This paper employs a bag-of-words model combined wi t h TF-IDF (term frequency–inverse document frequency) f or feature weighting [28,29]. The formula for calculating t he TF-IDF weight $w_{i,j}$ is as follows:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \tag{1}$$

$tf_{i,j}$ is the frequency of word i in document j; df_i is the number of documents containing word i; N is the total number of documents in the corpus. This weighting strategy effectively reduces the influence of high-frequency general words while highlighting key terms that are highly discriminative of documents. After the above collection and preprocessing process, the original online text is transformed into high-quality, structured corpus data. Although the data were sourced from professional platforms where users may be more inclined to discuss burnout, this potential bias is mitigated by the broad keyword coverage and the five-year collection period, which captures a wide spectrum of medical student experiences. Prior studies have demonstrated that online health communities provide valid insights into population-level mental health trends [30]. The focus of this study is on the temporal evolution of expressed emotions and topics, rather than on prevalence estimation, thus the relative nature of the analysis remains robust to such sampling biases. Future work could triangulate findings with survey data to further validate representativeness. However, relying solely on word frequency or shallow features is insufficient to accurately depict the complex emotional state inherent in medical students' job burnout. Therefore, it is necessary to introduce a deep semantic model with contextual understanding capabilities to perform more granular sentiment classification analysis on the text.

B. FINE-GRAINED SENTIMENT CLASSIFICATION BASED ON BIOBERT

To accurately identify the complex job burnout emotions in medical students' online texts, this paper selects the BioBERT model pre-trained on biomedical corpora as the basic architecture [31,32]. Specifically, the BioBERT-large version is adopted. This version is pre-trained on PubMed summary and PMC (PubMed Central) full text. Its vocabulary contains a large number of medical terms and has better semantic representation ability for professional psychological terms such as "burnout", "exhaustion" and "compassion fatigue". Compared with the general BERT model, BioBERT has shown significant advantages in tasks such as biomedical entity recognition and relation extraction, and is more suitable for the contextual characteristics of medical forum texts [33].

The model fine-tuning process uses the preprocessed text data as input. In the input representation layer, each text is constructed as a sequence of [CLS] + tokens + [SEP], with a maximum sequence length of 256. During fine-tuning, a fully connected classification layer is added to the basic BioBERT model. The input to the classification layer is the final hidden state vector corresponding to the [CLS] tokens (the hidden layer dimension of BioBERT-large), and its output is calculated using the Softmax function to determine the probability of each sentiment category:

$$P(y|h) = \mathrm{softmax}(Wh + b) \tag{2}$$

W is the trainable weight matrix, and b is the bias vector. The fine-tuning strategy employs a hierarchical learning rate decay: a smaller learning rate (0.00001) is used for the bottom-level BERT parameters to preserve the world knowledge acquired during pre-training, while a larger learning rate is used for the top-level and classification layers to quickly adapt to the downstream sentiment classification task.

To transcend the limitations of traditional binary (positive/negative) sentiment analysis, this paper constructs a four-category, fine-grained sentiment labeling system to more accurately capture specific emotional states associated with burnout. This system includes:

1)Positive - This refers to any text expressing a sense of accomplishment, support, optimism, or satisfaction with one's circumstance.

2)Negative - This refers to any text that is not directed at the central pillars of burnout but that expresses an emotion of mild displeasure or disappointment.

3)Burnout emotional exhaustion - Burnout is defined as the inability to sustain or expend emotional resources for an extended period of time leading to extreme exhaustion.

4)Burnout depersonalisation and low achievement - This definition refers to a lack of motivation to engage in any curricular or work-related activities and resulting decrease in self-efficacy.

This labeling system is constructed with reference to the core dimensions of the Maslach Burnout Scale and is jointly reviewed and determined by a psychology expert and two medical education researchers. To construct the training set, 5000 texts are randomly selected, and 80% of the labeled data are used as the training set, 10% as the validation set, and 10% as the test set. The training batch size is set to 16, and the training lasts for 10 epochs. The labels are independently labeled by three trained labelers. Krippendorff's Alpha coefficient is used to evaluate the inter-labeler reliability. If the result reaches 0.82, it indicates that the labeling system has good consistency [34]. For samples with discrepancies, the final labels are determined by expert arbitration. To address the potential class imbalance problem in the medical forum data, a weight w_j is introduced for each j in the loss function calculation. This weight is inversely proportional to the class frequency. The weighted cross-entropy loss function is:

$$L = -\sum_{j} w_j \cdot y_j \cdot \log(P(y_j|h)) \tag{3}$$

y_j represents the true one-hot encoded label of j. To prevent overfitting, Dropout is applied after the last layer of BERT, with a dropout rate of 0.2. During training, performance is evaluated on the validation set after each epoch, and the model parameters with the highest F1 score on the validation set are saved. Finally, the overall performance of the model is evaluated on an independent test set, calculating metrics such as accuracy, precision, recall, and macro F1 score. Model training is completed on a single NVIDIA RTX 3090 GPU, taking approximately 4.5 hours. The complete technique from raw data to the final map is shown in Figure 1.

Figure 1's core consists of three modules: fine-grained sentiment classification based on BioBERT, latent topic mining based on LDA, and the fusion and visualization generation of these two modules in the spatiotemporal dimensions. Through a fine-grained sentiment classification model based on BioBERT, this paper achieves accurate identification of different emotional states in online texts from medical students, especially the emotional dimensions related to burnout. Given the potential overlap between the two burnout subdimensions in short texts, the model employs a softmax output layer that produces a probability distribution over the four categories. For texts exhibiting features of both emotional exhaustion and depersonalization, the model assigns the label with the highest probability, while the full probability vector is retained for downstream weighted aggregation (see Section "Sentiment-Topic Fusion and Temporal Analysis"). This approach mitigates the impact of ambiguous cases by allowing the contribution of such texts to multiple sentiment dimensions proportionally to their predicted probabilities. Manual inspection of misclassified instances revealed that most confusions occurred between burnout subcategories and general negative sentiment, but the overall weighted F1 scores remain satisfactory. Future work could adopt multi-label classification to better capture co-occurring burnout symptoms. However, sentiment tags only reflect the "emotional orientation" of the text and are insufficient to reveal the specific sources of stress and contextual structure in the discussions among medical students. To further characterize the semantic topic framework upon which burnout emotions rely, it is necessary to conduct latent topic-level mining analysis on the text content.

C. LATENT TOPIC MINING BASED ON LDA

The preprocessed clean text needs to be converted into a structured numerical representation for processing by the

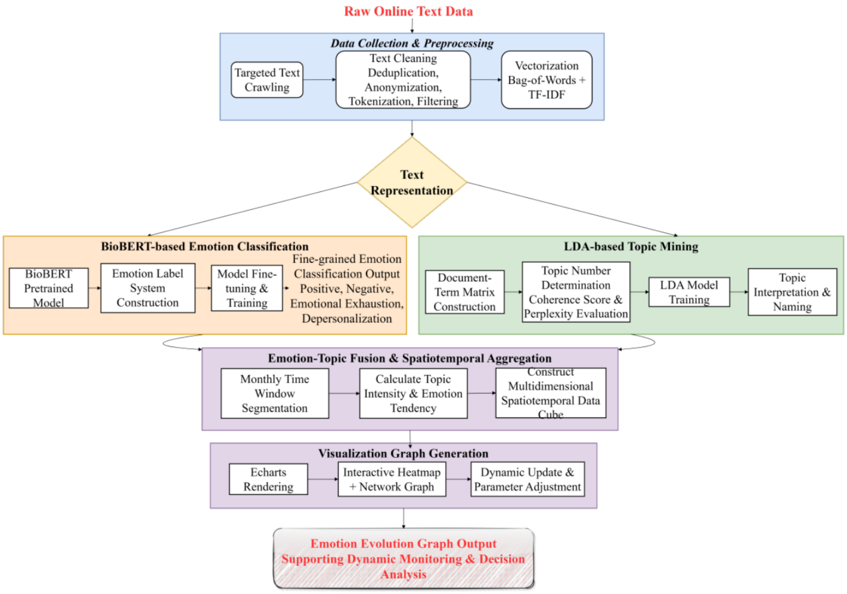

FIGURE 1. Framework for constructing an emotional evolution map

LDA model [35,36]. This paper constructs a document-term matrix:

$$D \in \mathbb{R}^{M \times C} \tag{4}$$

M represents the total number of documents (approximately 125,000), and C represents the size of the vocabulary after removing high-frequency meaningless words and low-frequency words (document frequency < 5). The matrix element D_{ij} uses word frequency counting to represent the number of times a word appears in a document. To further improve model efficiency and topic discrimination, based on word frequency statistics, only words with frequencies ranking in the top 80% quantile of the entire corpus are retained, resulting in a final vocabulary size C of approximately 8500 words.

Determining the number of topics is a key hyperparameter for LDA modeling. This paper uses topic consistency score as the primary evaluation metric, calculated based on the point mutual information of word pairs. For topic k, the formula for calculating its consistency score C_v is:

$$C_v = \sum_{i=1}^{L} \sum_{j=1}^{i-1} \log \left(\frac{P(\omega_i, \omega_j)}{P(\omega_i) \cdot P(\omega_j)} \right) \tag{5}$$

$P(\omega_i, \omega_j)$ represents the probability that words co-occur in the same document; $P(\omega_i)$ is the marginal probability of word occurrence; L is the top 10 words with the highest probability under each topic. The final topic consistency score of the model is the average of the scores of all K topics. At the same time, as an auxiliary reference, the perplexity of the model $Perp(D)$ is calculated, which is defined as:

$$Perp(D) = \exp \left(- \frac{\sum_{d=1}^{M} \log p(\omega_d)}{\sum_{d=1}^{M} N_d} \right) \tag{6}$$

$p(\omega_d)$ is the generation probability of word sequence in the document, and N_d is the number of words in the document. The experiment is conducted in the range of K from 5 to 50, with a grid search of 5 steps. The optimal number of topics is determined by a comprehensive trade-off between topic consistency score and $Perp(D)$, and combined with the subsequent manual interpretation of topic interpretability. The model is trained using the LDA algorithm based on Gibbs sampling [37]. The generation process of the model is determined by two core distributions: topic-term distribution Φ and document-topic distribution Θ. The goal of Gibbs sampling is to estimate these two distributions by iteratively updating the topic assignment for each word.

During training, the Dirichlet prior hyperparameters are set as follows: α for the document-topic distribution is $50/K$, and β for the topic-term distribution is 0.01. The model undergoes 2000 iterations of Gibbs sampling, with the first 500 iterations discarded as a burn-in period. Starting from the 501st iteration, sampling is performed every 100 iterations to construct samples for the posterior distribution. Finally, Φ and Θ are estimated based on the statistics of these samples. After training, for each topic k, the top 20 keywords with the highest probabilities are extracted from its corresponding topic-term distribution Φ_k. While the preprocessing retained adverbs to preserve sentiment intensity for classification, their inclusion in LDA modeling may introduce noise. To mitigate this, topic interpretation focused on the top-ranked keywords that were predominantly nouns, verbs, and adjectives; manual inspection confirmed that high-probability adverbs were rare and did not dominate topic definitions. The achieved topic coherence score of 0.632 (Section "LDA Topic Model Quality") suggests that the overall semantic quality remains high. Future implementations could adopt separate preprocessing pipelines for sentiment and topic tasks to further optimize performance.

The final interpretation and naming of topics is a process combining quantitative and qualitative methods. First, based on the list of high-probability keywords, two independent researchers familiar with medical education make preliminary inferences about the core semantics of each topic. A topic with high-frequency words including "on duty," "night shift," "emergency room," "fatigue," and "shift handover" can be preliminarily inferred to be related to "clinical workload." Subsequently, the top 5 original texts with the highest probabilities under each topic are returned, and the researchers verify and revise the preliminary inferences in conjunction with the specific context. Through discussion, a consensus is reached on ambiguous topics, and each topic is given a concise and general label, including "internship pressure and physical exhaustion," "academic exam anxiety," "crisis of career value identity," "support from teachers, students, and peers," "evaluation of the standardized residency training system," and "confusion about future career planning." Figure 2 shows the keyword probability distribution of four of the most representative topics.

Figure 2 illustrates the probability weight distribution of four key themes and their core keywords extracted from online texts by medical students using the LDA topic model. The vertical axis of Figures 2(a) to 2(d) lists the 15 most probable keywords representing each theme, while the horizontal axis corresponds to the probability weight of these keywords in the theme-term distribution. Specifically, the data shows that the weight of keywords within each theme exhibits a strictly monotonically decreasing trend from high to low: in Figure 2(a), "night shift" has the highest weight, followed by "on duty," "fatigue," and finally the lowest, "physiological"; the sequences "exam," "failing," and "reviewing" in Figure 2(b), "meaning," "numbness," and "ideal" in Figure 2(c), and "mentoring," "teacher," and "classmate" in Figure 2(d) all follow the same pattern. Through the systematic decay of weights assigned to the words of the LDA model, it can be seen that the model can provide insights into the thematic/fontal structure of a text corpus and can rank the words based upon their ability to provide representations of the various themes as well as their abilities to provide accurate representations of the semantic connotations of each theme. In addition, the large differences between the keyword sets that display high weights for each theme/subject—Theme 1 has keywords that focus on physical workloads; Theme 2 has keywords that focus on academic assessment; Theme 3 has keywords that relate to psychological and value cognition; Theme 4 has keywords that relate to the social/institutional environment—validates the uniqueness and good interpretability of the themes mined by the LDA model. Therefore, this supports the view that the online discussions about medical students' job-related burnout are complex constructs composed of multiple dimensions, and provides a sound semantic basis for future studies utilizing dynamic evolution with sentiment analysis.

D. SENTIMENT-TOPIC FUSION AND TEMPORAL ANALYSIS

In order to characterize, on a macroscopic level, the dynamics related to medical student burnout, a spatiotemporal aggregation and fusion process over the discrete sample taken through the fine-grained sentiment tagging and topic affiliation must take place to build a visualization map. The initial step in the spatiotemporal process is to segment and summarize the texts by time. For this study, given the academic year for the medical education activities and the temporal nature of the online discussions, fixed length time windows are used, summarizing all texts published in a given calendar month, to segregate all of the texts. The dataset represents a five-year period, or 60 months, so using this method, the 60 consecutive time slices enable the creation of continuity in this time series, while retaining sufficient granularity to understand the effects of short-term events.

It is important to clarify that the term "evolution" in this study refers to the temporal dynamics of each topic's discussion intensity and associated sentiment composition, rather than a transformation of the topic's core semantic content. The proposed map visualizes how the prevalence and emotional tone of predefined themes change over time, enabling the identification of periods when certain stressors become more prominent or when burnout emotions intensify. This conceptualization is consistently applied throughout the paper, and phrases such as "emotional evolution" are used synonymously with "temporal variation in emotional expression."

Within each time window t, data aggregation and index calculation are performed, with the core being the quantification of the discussion intensity and sentiment color of each topic. For the document-topic distribution matrix obtained by the LDA model, the element $\theta_{d,k}$ represents the probability that a document belongs to topic k. Within a time window t, the intensity $I_{t,k}$ of topic k is defined as the sum of the probability contributions of all documents within that window to topic k, after normalization:

$$I_{t,k} = \frac{\sum_{d\in t} \theta_{d,k}}{\sum_{k'=1}^{K} \sum_{d\in t} \theta_{d,k'}} \tag{7}$$

This value reflects the relative proportion of k in all discussions within time t. Simultaneously, the sentiment tendency under each topic needs to be calculated. Based on the sentiment classification results of BioBERT, each document is assigned a primary sentiment label s_d. To obtain the sentiment score $E_{t,k,c}$ of topic k with respect to a specific sentiment category c within the time window

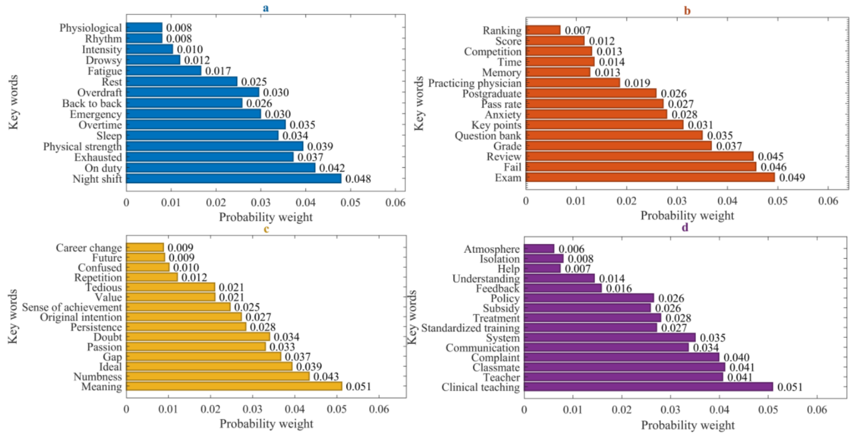

FIGURE 2. Probability distribution of core keywords for four key themes: (a) Theme 1: Internship stress and physical exhaustion; (b) Theme 2: Academic exam anxiety; (c) Theme 3: Crisis of career value identity; (d) Theme 4: Support system and institutional evaluation

t, a weighted average method is used, where the weight is the probability $\theta_{d,k}$ that the document belongs to that topic:

$$E_{t,k,c} = \frac{\sum_{d \in t} \theta_{d,k} I(s_d = c)}{\sum_{d \in t} \theta_{d,k}} \tag{8}$$

$I(s_d = c)$ is an indicator function, taking a value of 1 when the sentiment label of the document is c, and 0 otherwise. Therefore, for each t and k, a sentiment distribution vector is obtained. Finally, a multidimensional spatiotemporal data cube is constructed, with dimensions $t \times k \times c$, element values $E_{t,k,c}$, and topic intensity $I_{t,k}$ as metadata.

E. EVOLUTION MAP GENERATION AND VISUALIZATION

The visualization construction of the dynamic map is based on the above data cube. The visualization solution employs a web-based technology stack, utilizing the Echarts library for rendering. The main design of the map is an interactive view linking a force-directed network diagram with a time-series heatmap. In the network diagram, nodes represent different themes, and the node size maps to the average intensity of that theme within a selected time window.

$$\frac{1}{T} \sum_{t} I_{t,k} \tag{9}$$

Node colors map to their dominant sentiment, i.e.:

$$\arg\max_{c} \left(\frac{1}{T} \sum_{t} E_{t,k,c} \right) \tag{10}$$

The edges connecting nodes represent the co-occurrence relationships between topics, and the weight of the edge is determined by the sum of the co-occurrence probabilities of the topic in all documents. The vertical axis of the heatmap is the list of topics, and the horizontal axis is the timeline. The color depth of each cell represents the intensity of the negative sentiment tendency of the corresponding topic within the corresponding time window. Users can dynamically update the network diagram to a specific time point by sliding the timeline or clicking on the heatmap cells. At this time, the node size is adjusted to the immediate $I_{t,k}$ of that time point, and the color is adjusted to the immediate dominant sentiment of that time point, thus clearly presenting the co-evolution trajectory of topic attention and sentiment color. All visualization code is encapsulated as an independent module, supporting data updates and parameter adjustments, ensuring the reproducibility and scalability of the map.

III. EXPERIMENTS AND MAPS

A. EXPERIMENTAL SETUP AND DATA DESCRIPTION

The dataset used to run the experiments comes from the final corpus that is created in accordance with the steps outlined in Section "Data Collection and Preprocessing". Of these, 80% (109,944 texts) from January 2019 to June 2022 were used for model training and development, and the remaining 20% (27,486 texts) from July 2022 to December 2023 served as the test set. The 5,000 texts selected for manual annotation in Section "Fine-grained Sentiment Classification Based on BioBERT" were randomly drawn from the training set, ensuring that the test set remained unseen during model fine-tuning.The experiments are conducted on a workbench with Intel Xeon Gold6226R Processor, 256GB of RAM, and two NVIDIA RTX 3090 cards running PyTorch 1.12 as the developing deep learning framework. To evaluate the effectiveness of the sentiment classification task, the experiments employ metrics that include accuracy, precision, recall, and macro F1 score. For multi-class problems, the average macro F1 score is calculated as the average of each sentiment category's individual f1 score. For LDA, the main metric used to measure topic performance is the topic consistency score calculated using the Cv method. The topic consistency score quantifies the degree to which high-frequency words are similar in nature and consistent semantically within a given topic, ultimately being calculated by assessing the average normalized pointwise mutual information between words. The higher the score, the closer the word association within the topic and the more interpretable the topic.

B. METHOD EFFECTIVENESS EVALUATION

1) Fine-Grained Sentiment Classification Performance

The model training employs a hierarchical learning rate, where lower learning rates are used for lower-level parameters to preserve general knowledge, and higher learning rates are used for higher-level parameters to adapt to the specific task. The entire model is trained using an optimizer with weight decay, a batch size of 16, and a total of 10 training epochs. To address the class imbalance problem, weight adjustment is introduced into the loss function, and a random dropout mechanism is added to mitigate overfitting. Performance is evaluated on the validation set in each epoch of training, and the optimal model parameters are saved. The final model performance on the independent test set is shown in Table 1.

TABLE 1. Performance evaluation results of the fine-grained sentiment classification model on the test set

Sentiment category	Precision	Recall	F1 score	Support (Count)
Positive	0.941	0.935	0.938	8,452
Negative	0.926	0.931	0.928	10,789
Burnout - Emotional exhaustion	0.907	0.895	0.901	5,230
Burnout - Depersonalization	0.873	0.896	0.884	3,015
Overall metrics	Accuracy: 0.927 Macro average F1:0.904			27,486

Table 1 shows that the model achieves an overall accuracy of 92.7%, validating its effectiveness in sentiment discrimination of medical student texts. Looking at the detailed indicators for each category, the model demonstrates extremely high precision and recall for identifying general emotions such as "positive" and "negative." For the two subcategories of burnout, the model performs best in identifying "burnout-emotional exhaustion," achieving an F1 score of 0.901. This is attributed to the fact that texts in this category often contain highly characteristic words such as "exhausted" and "overdrawn." In contrast, the F1 score for the "burnout-depersonalization and low achievement" category is slightly lower at 0.884. Analysis of its confusion matrix reveals that some texts in this category are occasionally misclassified as ordinary "negative" emotions due to their relatively implicit and complex expression. Overall, the macro F1 score of 0.904 indicates that the model's comprehensive discrimination ability across the four emotion categories is balanced and robust, providing a reliable foundation for subsequent emotional evolution analysis. To further validate the advantage of the proposed hybrid framework, comparative experiments were conducted against baseline models. For sentiment classification, a standard BERT-base model was fine-tuned on the same annotated data; for topic modeling, traditional LDA without BERT embedding was performed. The quantitative results are summarized in Table 2, which shows that BioBERT achieves an accuracy of 92.7% and a macro F1 of 0.904, outperforming standard BERT by 3.6% and 0.036, respectively. Meanwhile, the proposed BERT-enhanced LDA attains a topic coherence score of 0.632, representing a 21.5% improvement over traditional LDA (0.52). These comparisons substantiate the claim that the hybrid approach effectively overcomes limitations of traditional methods in this domain.

TABLE 2. Comparison of proposed methods with baseline approaches

Task	Method	Metric	Value
Sentiment classification	BioBERT (proposed)	Accuracy	92.7%
Sentiment classification	BioBERT (proposed)	Macro F1	0.904
Sentiment classification	Standard BERT	Accuracy	89.1%
Sentiment classification	Standard BERT	Macro F1	0.868
Topic modeling	BERT-enhanced LDA	Coherence score	0.632
Topic modeling	Traditional LDA	Coherence score	0.52

2) LDA Topic Model Quality

To determine the optimal number of topics, the experiment conducts a grid search with a step size of 5 within the range of topic number K from 5 to 50, calculating the topic consistency score C_v and $Perp(D)$

for each K value. The topic consistency score is measured using the C_v method based on normalized point mutual information, calculating the semantic association strength between the top-10 keywords under each topic; $Perp(D)$ reflects the model's prediction uncertainty for unknown data. The results are shown in Figure 3.

Figure 3 shows the trend of LDA topic model quality as the number of topics K changes from 5 to 50. The consistency score is as follows: K=5: 0.42, K=10: 0.51, K=15: 0.58, K=20: 0.62, K=25: 0.632 (peak), K=30:

0.625, K=35: 0.615, K=40: 0.605, K=45: 0.59, K=50: 0.57, showing a trend of first increasing and then decreasing. The logarithmic $Perp(D)$ data shows a continuous decreasing trend followed by a gradual flattening. These changes indicate that when the number of topics is small, the model semantics are mixed. As K increases, consistency improves and model fitting improves ($Perp(D)$ decreases). However, after K=25, consistency decreases while $Perp(D)$ tends to stabilize, indicating that too many topics lead to topic redundancy and overfitting. Therefore, K=25 is the optimal balance point.

C. EMOTIONAL EVOLUTION MAP

1) Overall Sentiment Time-Series Trend

Based on the sentiment classification results of all texts within the test set period (July 2022 to December 2023), the evolution of the overall sentiment proportion is calculated on a monthly basis. The time window is set to a calendar month, resulting in 18 consecutive time slices to balance temporal granularity and data stability. Within each month,

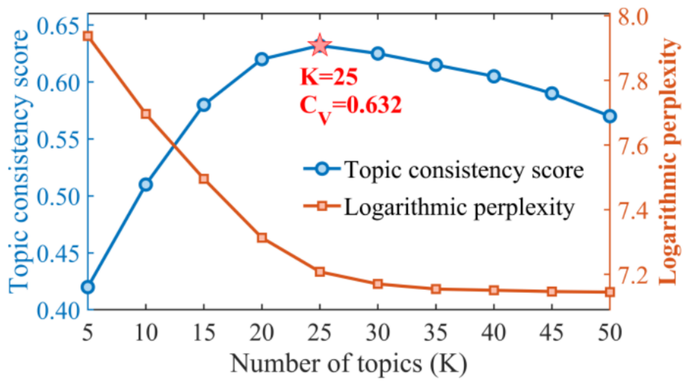

FIGURE 3. Trend of LDA topic model quality with number of topics K. The five-pointed star intuitively indicates the number of topics that optimize the consistency of the topic

the proportion of texts in the four sentiment categories is statistically analyzed and normalized to eliminate the impact of monthly fluctuations in the total number of texts, as shown in Figure 4.

Figure 4 shows the evolution trend of the proportion of four emotions in medical students' online texts over time from July 2022 to December 2023. The horizontal axis represents time (year and month), and the vertical axis represents the emotional proportion. Simulation data shows that the proportion of negative emotions is the highest (fluctuating around 35%), followed by positive emotions (fluctuating around 25%), while the proportions of emotional exhaustion and depersonalization are relatively low. Negative and positive emotions show obvious periodic fluctuations, while the depersonalization component, representing professional burnout, shows a slow but continuous upward trend in the later stages of observation (especially after the "start of internship" marked in July 2023). These changes indicate that the overall emotional state of medical students is significantly affected by the academic cycle, and that long-term professional practice pressure may gradually exacerbate the "depersonalization" dimension of their professional burnout.

2) Evolution of Theme-Emotion Association

Integrating the results of emotion classification and theme mining, a multidimensional evolution map covering 15 themes, 4 emotion categories, and spanning 18 months is constructed. Based on the document-topic distribution matrix, the intensity of each topic (i.e., the relative proportion of the topic in all discussions in that month) and its tendency score in each sentiment category are calculated monthly, with the weighting weight representing the probability that a document belongs to that topic. Based on this, a spatiotemporal evolution heatmap of the topic's negative emotional intensity is generated, with color depth mapping to negative emotional intensity, as shown in Figure 5.

The colors in Figure 5, from light to dark, represent the intensity of negative emotional impact for each theme in a given month, increasing from weak to strong. Point A marks the peak of negative emotional impact for the theme "Academic Exam Anxiety" in June 2023 (exam month); area B shows the synchronous increase in negative emotional impact for multiple themes in December 2022, possibly related to systemic stress events during a specific period.

To reveal the heterogeneity of emotional structures triggered by different stressors in greater detail, this paper further presents a comparative analysis of the fine-grained emotional composition of multiple themes (Figure 6). Figure 6 aims to answer the question: Although multiple themes exhibit high negative emotional intensity, are their underlying emotional components the same? For example, are both "Internship Stress" and "Exam Anxiety" primarily characterized by "emotional exhaustion"? To this end, Figure 6 selects four core themes: “internship pressure and physical exhaustion”, “academic exam anxiety”, “crisis of career value identity” and “support system and institutional evaluation”. Using quarters as the time window, the composition ratio of the four emotions of “positive”, “negative”, “burnout-emotional exhaustion” and “burnout-depersonalization” in the text is decomposed and compared.

Figure 6 compares and analyzes the fine-grained emotional composition of multiple themes. In Figures 6(a) to 6(d), the proportion of exhaustion emotion is relatively high (approximately 45%) in Theme 1, "Internship Stress and Physical Exhaustion," with slight fluctuations between quarters, indicating that internship stress continues to lead to emotional exhaustion. The negative emotion in Theme 2, "Academic Exam Anxiety," increases significantly in Q2 and Q4, reflecting the intensification of exam season anxiety. The depersonalized emotion in Theme 3, "Crisis of Career Value Identity," shows an upward trend (from approximately 35% in Q1 to approximately 43% in Q4), indicating that the identity crisis deepens over time. The proportion of positive emotion in Theme 4, "Support System and Institutional Evaluation," is relatively high compared to other themes, and the emotional composition is relatively

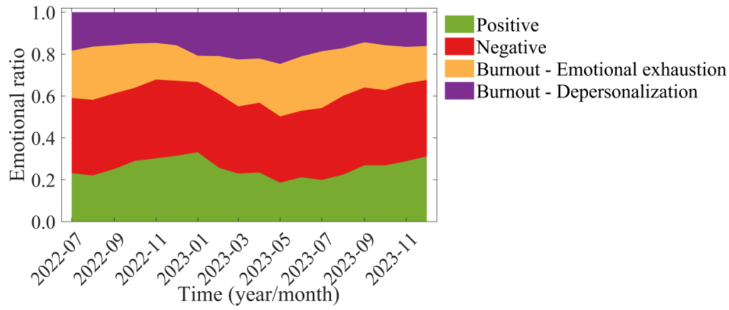

FIGURE 4. Evolution of overall emotional proportion in medical students' online texts

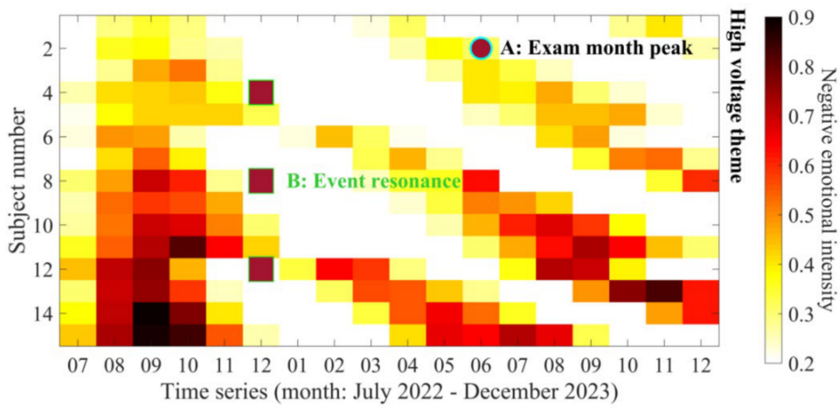

FIGURE 5. Spatiotemporal evolution of negative emotional intensity for each theme

balanced, indicating that the support system receives some positive evaluation. Overall, the changes in emotional composition reveal the cyclical characteristics or trend effects of different themes, and the data supports these conclusions.

3) Typical Evolutionary Pattern Cases

Two high-burden themes, "internship stress" and "exam anxiety," are selected for case analysis. Negative emotions related to internship stress peaked during the summer internship season, primarily due to long shifts, paperwork, and insufficient sleep. Exam anxiety exhibits short-term peaks, with fluctuations consistent with the pre-exam tension and post-exam relief rhythm; the June peak is particularly prominent, stemming from the combined pressure of final exams and professional qualification exams. This indicates that the emotional evolution map can reflect cyclical fluctuations and reveal the complex mechanisms of multiple stress accumulation through abnormal peaks, helping to accurately pinpoint psychological risk points.

To dynamically present the emotional interactions between themes, this paper constructs a negative intensity transfer chord diagram among core themes (Figure 7). Figure 7 visually demonstrates the strength of emotional connections between themes using the width and color of the chords, statically presenting a network of connections and dynamically metaphorically representing the transmission path of emotions under stressful events, thus explaining the reasons for the synchronous fluctuations of emotions across multiple themes and providing a network perspective for understanding the holistic nature of burnout.

Figure 7 visually illustrates the negative emotional connection network among four core burnout themes for medical students during exam season using a non-linear layout. Instead of traditional horizontal and vertical axes, a circular layout with fan-shaped areas and connecting chords is used. The fan-shaped areas represent the four themes: "Internship Stress," "Exam Anxiety," "Identity Crisis," and "Support System." The size of each fan-shaped area reflects the intensity of negative affective feelings associated with each theme during that period, with "Exam Anxiety" showing the highest intensity.

The thickness and color of the connecting chords represent the strength of the emotional association between the themes. Data shows that "Exam Anxiety" has the strongest association with "Internship Stress" (approximately 0.6), and is also highly correlated with "Identity Crisis" (approximately 0.5). This indicates that during exam season, academic pressure exacerbates negative feelings about internships and reinforces doubts about professional value, forming an affective transmission chain. Furthermore, there is also a significant association between "Support System"

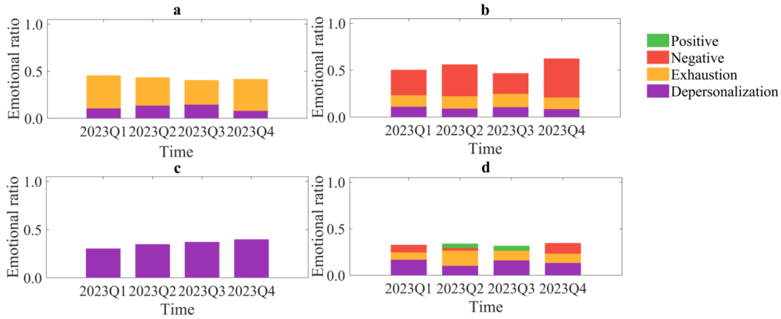

FIGURE 6. Comparative analysis of fine-grained emotional composition across multiple themes: (a) Theme 1: Internship stress and physical exhaustion; (b) Theme 2: Academic exam anxiety; (c) Theme 3: Crisis of career value identity; (d). Theme 4: Support system and institutional evaluation

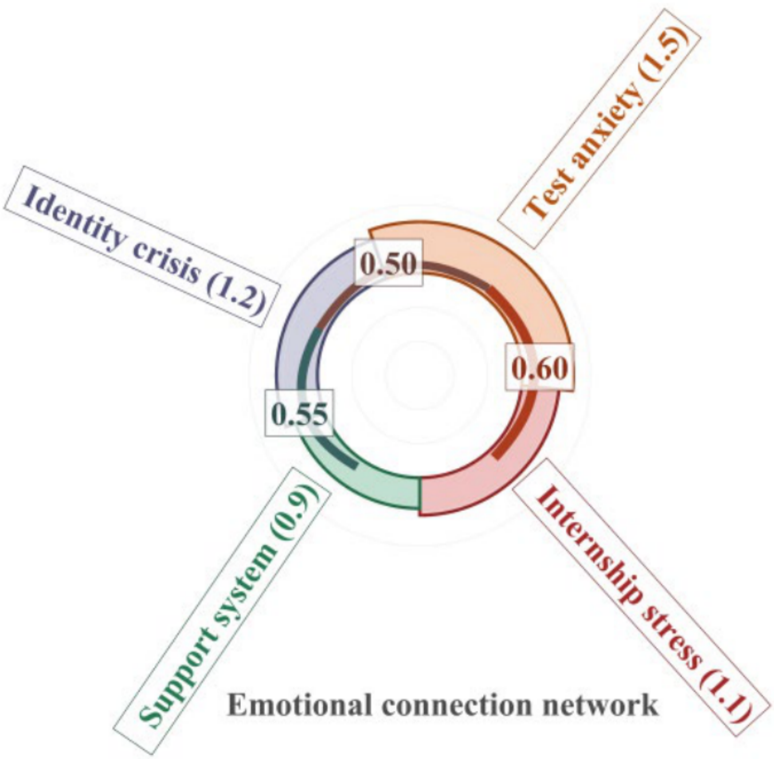

FIGURE 7. Negative emotional connection network of medical student burnout during exam season

and "Identity Crisis." This map reveals that the various dimensions of burnout among medical students are interconnected and mutually reinforcing during periods of stress, rather than evolving in isolation. This finding provides a basis for comprehensive intervention strategies that target multiple dimensions and block the transmission of emotions.

To further validate the map against real-world events, an event-driven analysis was conducted by aligning emotional peaks with known academic and public health milestones. Figure 8 plots the monthly proportion of burnout-specific emotions (emotional exhaustion and depersonalization combined) from January 2020 to December 2023, with annotat ions for major events: the initial COVID-19 outbreak (Ma rch 2020), the nationwide examination month (June 202 2), and the post-pandemic policy shift (December 2022). A marked increase in burnout emotions is observed durin g the exam month (peak at 23.4%) and during the early pandemic phase (21.8%), while the policy shift coincided with a transient rise (19.7%).These correlations confirm t hat the emotional evolution map is sensitive to external stres sors and can serve as a monitor-ing tool for identifying crit ical intervention windows.

IV. CONCLUSION

The emphasis of this research has been placed on unveiling the correlation between burnout among medical students by utilizing the online textual-based mapping technique of emotional evolution that consists of two methods, BioBERT sentiment analysis and LDA topic modelling. The results reveal that medical students experience distinctly different

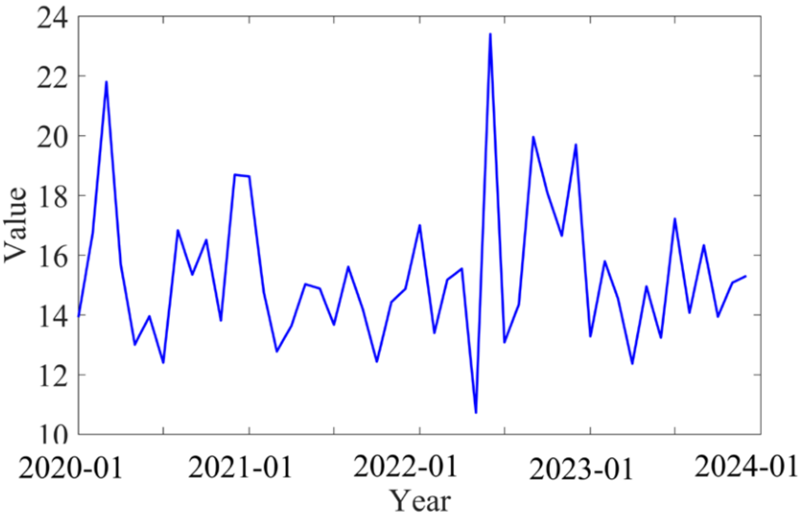

FIGURE 8. Event-driven analysis of burnout-specific emotion proportions (January 2020 – December 2023)

sentiment structures and evolutionary trajectories concerning the various themes associated with stress. The exam-related themes of stress demonstrate a concentration of negative emotional responses in a relatively short amount of time, whereas internship- and career-identity-related themes develop an extended and cumulative pattern of emotional exhaustion. The paper presents an example of the synergy created through combining a deep semantic representation of messages; discovery of emerging topics through LDA; and temporal modelling of emotional development through an integrated framework for the dynamic identification and precise intervention of burnout among medical students, thus creating a scalable technical pathway to these goals. The implementation of additional research opportunities or projects could include multi-platform fusion data, development of real-time early warning system, and verification of longitudinal questionnaire data pertaining to the validity of early warning mechanism for burnout among medical students.

AUTHOR CONTRIBUTIONS

Writing – Original Draft, Writing – Review & Editing: Jiayang Li.

CONFLICTS OF INTEREST

The authors declare that they have no financial conflicts of interest.

FUNDING

This work was supported by Hebi Polytechnic 2025 University-level Innovation Team, 2025SK01; 2025 Henan Provincial Education System Special Research Project on Integrity <Promoting the Innovation of Party Discipline Education Methods -An Exploration of the "Seven-Force-Driven and Three-Dimensional Closed-Loop" Model from the Perspective of Model Party Branch Construction>,2025LZZD-86; 2024 Henan Provincial Medical Education Research Project <Research on the Four-Dimensional Reconstruction of Labor Curriculum for Higher Vocational Nursing Major Based on Tyler's Principles Under the Background of New Medical Science>,WJLX2024233; General Project of Henan Educational Science Planning 2025 (2025YB0670); Research and Practice Project on Education and Teaching Reform, Hebi Polytechnic (2025XJJGLX001); Research and Practice Project on Education and Teaching Reform, Hebi Polytechnic (2025XJJGLX018).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

ACKNOWLEDGEMENTS

None.

REFERENCES

[1] H. Almutairi, A. Alsubaiei, S. Abduljawad, and A. Alshatti, "Prevalence of burnout in medical students: A systematic review and meta-analysis," International Journal of Social Psychiatry, vol. 68, no. 6, pp. 1157-1170, 2022, doi: 10.1177/00207640221106691.

[2] S. Aker and M. K. Şahin, "The relationship between school burnout, sense of school belonging and academic achievement in preclinical medical students," Advances in Health Sciences Education, vol. 27, no. 4, pp. 949-963, 2022, doi: 10.1007/s10459-022-10121-x.

[3] A. N. Leep Hunderfund, B. Saberzadeh Ardestani, S. K. Laughlin-Tommaso, B. L. Jordan, V. A. Melson, M. M. Montenegro, et al., "Sense of belonging among medical students, residents, and fellows: Associations with burnout, recruitment retention, and learning environment," Academic Medicine, vol. 100, no. 2, pp. 191-202, 2025, doi: 10.1097/ACM.0000000000005892.

[4] M. L. Alkureishi, D. Jaishankar, S. Dave, S. Tatineni, M. Zhu, K. C. Chretien, et al., "Impact of the early phase of the COVID-19 pandemic on medical student well-being: a multisite survey," Journal of General Internal Medicine, vol. 37, no. 9, pp. 2156-2164, 2022, doi: 10.1007/s11606-022-07497-2.

[5] D. Zúñiga, M. Torres-Sahli, A. Rigotti, N. Pedrals, G. Echeverría, O. Padilla, et al., "Dispositional mindfulness reduces burnout and promotes

flourishing in medical students: A two-wave latent change score model," Mindfulness, vol. 13, no. 1, pp. 112-122, 2022, doi: 10.1007/s12671-021-01774-7.

[6] J. P. Roman-Calderon, A. Krikorian, E. Ruiz, A. M. Romero, and M. Lemos, "Compassion and self-compassion: counterfactors of burnout in medical students and physicians," Psychological Reports, vol. 127, no. 3, pp. 1032-1049, 2024, doi: 10.1177/00332941221132995.

[7] N. Daud, A. F. A. Rahim, M. N. M. Pa, A. Ahmad, N. A. Yusof, N. M. Hassan, et al., "Emotional Intelligence Among Medical Students and Its Relationship with Burnout," Education in Medicine Journal, vol. 14, no. 3, pp. 49-59, 2022, doi: 10.21315/eimj2022.14.3.4.

[8] D. Bhugra and A. Molodynski, "Well-being and burnout in medical students: Challenges and solutions," Irish Journal of Psychological Medicine, vol. 41, no. 2, pp. 175-178, 2024, doi: 10.1017/ipm.2022.26.

[9] M. Kadhum, O. O. Ayinde, C. Wilkes, E. Chumakov, D. Dahanayake, A. Ashrafi, et al., "Wellbeing, burnout and substance use amongst medical students: A summary of results from nine countries," International Journal of Social Psychiatry, vol. 68, no. 6, pp. 1218-1222, 2022, doi: 10.1177/00207640221094955.

[10] O. O. Ayinde, E. R. Akinnuoye, A. Molodynski, O. Battrick, and O. Gureje, "A descriptive study of mental health and burnout among Nigerian medical students," International Journal of Social Psychiatry, vol. 68, no. 6, pp. 1223-1231, 2022, doi: 10.1177/00207640211057706.

[11] N. T. Mhata, V. Ntlantsana, A. M. Tomita, K. Mwambene, and S. Saloojee, "Prevalence of depression, anxiety and burnout in medical students at the University of Namibia," South African Journal of Psychiatry, vol. 29, no. 1, pp. 1-9, 2023, doi: 10.4102/sajpsychiatry.v29i0.2044.

[12] L. M. Meeks, K. Pereira-Lima, M. Plegue, N. R. Jain, E. Stergiopoulos, C. Stauffer, et al., "Disability, program access, empathy and burnout in US medical students: A national study," Medical Education, vol. 57, no. 6, pp. 523-534, 2023, doi: 10.1111/medu.14995.

[13] D. Lee, K. Reasoner, C. Davidson, J. S. Pennings, and D. H. Lee, "The relationships between grit, burnout, and demographic characteristics in medical students," Psychological Reports, vol. 126, no. 5, pp. 2511-2529, 2023, doi: 10.1177/00332941221087899.

[14] L. George and P. Sumathy, "An integrated clustering and BERT framework for improved topic modeling," International Journal of Information Technology, vol. 15, no. 4, pp. 2187-2195, 2023, doi: 10.1007/s41870-023-01268-w.

[15] A. Riaz, O. Abdulkader, M. J. Ikram, and S. Jan, "Exploring topic modelling: A comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity," International Journal of Electrical and Computer Engineering (IJECE), vol. 15, no. 2, pp. 1933-1948, 2025, doi: 10.11591/ijece.v15i2.pp1933-1948.

[16] M. A. Khadija and W. Nurharjadmo, "Enhancing Indonesian customer complaint analysis: LDA topic modelling with BERT embeddings," Sinergi, vol. 28, no. 1, pp. 153-162, 2024, doi: 10.22441/sinergi.2024.1.015.

[17] M. C. Wijanto, I. Widiastuti, and H. S. Yong, "Topic Modeling for Scientific Articles: Exploring Optimal Hyperparameter Tuning in BERT," International Journal on Advanced Science, Engineering and Information Technology, vol. 14, no. 3, pp. 912-919, 2024, doi: 10.18517/ijaseit.14.3.19347.

[18] K. Datchanamoorthy, "Text mining: Clustering using bert and probabilistic topic modeling," Social Informatics Journal, vol. 2, no. 2, pp. 1-13, 2023, doi: 10.58898/sij.v2i2.01-13.

[19] H. A. Taher and N. N. A. Hasan, "Integration Named Entity Recognition and Latent Dirichlet Allocation to Enhance Topic Modeling," Annals of Emerging Technologies in Computing (AETiC), vol. 9, no. 2, pp. 20-30, 2025, doi: 10.33166/AETiC.2025.02.002.

[20] A. J. Rawat, S. Ghildiyal, and A. K. Dixit, "Topic modelling of legal documents using NLP and bidirectional encoder representations from transformers," Indonesian Journal of Electrical Engineering and Computer Science, vol. 28, no. 3, pp. 1749-1755, 2022, doi: 10.11591/ijeecs.v28.i3.pp1749-1755.

[21] C. Chen and B. Morkos, "Exploring topic modelling for generalising design requirements in complex design," Journal of Engineering Design, vol. 34, no. 11, pp. 922-940, 2023, doi: 10.1080/09544828.2023.2268850.

[22] M. Bruno, E. Catanese, and M. De Cubellis, "Integrating word embedding and topic modeling for sentiment analysis: A case study on the social mood on economy," Statistical Journal of the IAOS, vol. 40, no. 3, pp. 535-548, 2024, doi: 10.3233/SJI-240017.

[23] Y. Wang, X. Liu, and X. L. Zhu, "Enhancing emerging technology discovery in nanomedicine by integrating innovative sentences using BERT and NLDA," Science, vol. 9, no. 4, pp. 155-195, 2024, doi: 10.2478/jdis-2024-0031.

[24] M. Karabacak, P. Jagtiani, A. Jain, F. Panov, and K. Margetis, "Tracing topics and trends in drug-resistant epilepsy research using a natural language processing–based topic modeling approach," Epilepsia, vol. 65, no. 4, pp. 861-872, 2024, doi: 10.1111/epi.17890.

[25] B. Song, C. Luan, and D. Liang, "Identification of emerging technology topics (ETTs) using BERT-based model and sematic analysis: A perspective of multiple-field characteristics of patented inventions (MFCO-PIs)," Scientometrics, vol. 128, no. 11, pp. 5883-5904, 2023, doi: 10.1007/s11192-023-04819-x.

[26] M. Osvath, Z. G. Yang, and K. Kósa, "Analyzing narratives of patient experiences: A bert topic modeling approach," Acta Polytechnica Hungarica, vol. 20, no. 7, pp. 153-171, 2023, [Online]. Available: https://acta.uni-obuda.hu/Osvath_Yang_Kosa_136.pdf.

[27] G. B. B. Gomes, R. Attux, and C. C. Cruz, "Enhancing Contributions to Brazilian Social Media Analysis Based on Topic Modeling with Native BERT Models," Journal of Information and Data Management, vol. 16, no. 1, pp. 181-191, 2025, doi: 10.5753/jidm.2025.4670.

[28] H. D. Abubakar, M. Umar, and M. A. Bakale, "Sentiment classification: Review of text vectorization methods: Bag of words, Tf-Idf, Word2vec and Doc2vec," SLU Journal of Science and Technology, vol. 4, no. 1, pp. 27-33, 2022, doi: 10.56471/slujst.v4i.266.

[29] S. Dai, K. Li, Z. Luo, P. Zhao, B. Hong, A. Zhu, et al., "AI-based NLP section discusses the application and effect of bag-of-words models and TF-IDF in NLP tasks," Journal of Artificial Intelligence General Science, vol. 5, no. 1, pp. 13-21, 2024, [Online]. Available: https://ideas.repec.org/a/das/njaigs/v5y2024i1p13-21id149.html.

[30] R. Merchant, A. Goldin, D. Manjanatha, C. Harter, J. Chandler, A. Lipp, et al., "Opportunities to expand access to mental health services: A case for the role of online peer support communities," Psychiatric Quarterly, vol. 93, no. 2, pp. 613-625, 2022, doi: 10.1007/s11126-022-09974-7.

[31] M. KafiKang and A. Hendawi, "Drug-drug interaction extraction from biomedical text using relation BioBERT with BLSTM," Machine Learning and Knowledge Extraction, vol. 5, no. 2, pp. 669-683, 2023, doi: 10.3390/make5020036.

[32] A. Kumar, J. Malla, and A. Sharma, "BioBERT-RxReadmit: Improving Hospital Readmission Predictions Through Clinical Text Analysis with BioBERT," International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies, vol. 2, no. 1, pp. 14-29, 2025, doi: 10.63503/j.ijaimd.2025.38.

[33] D. KAVYA, "CHATBOT FOR MULTILINGUAL HEALTHCARE ENVIRONMENT USING BIO-BERT," INTERNA-TIONAL JOURNAL, vol. 13, no. 1, pp. 242-246, 2025, doi: 10.22214/ijraset.2025.66243.

[34] M. Goyanes and V. Piñeiro Naval, "Content analysis in SPSS and KALPHA: Procedure for a Reliable Quantitative Analysis with Cohen's Kappa and Krippendorff's Alpha," Studies on the Journalistic Message, vol. 30, no. 1, pp. 123-140, 2024, doi: 10.5209/esmp.92732.

[35] R. K. Gupta, R. Agarwalla, B. H. Naik, J. R. Evuri, A. Thapa, and T. D. Singh, "Prediction of research trends using LDA based topic modeling," Global Transitions Proceedings, vol. 3, no. 1, pp. 298-304, 2022, doi: 10.1016/j.gltp.2022.03.015.

[36] C. Sharma, S. Sharma, and Sakshi, "Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: A review of trends and research patterns used in integration," Multimedia Tools and Applications, vol. 81, no. 25, pp. 36805-36831, 2022, doi: 10.1007/s11042-022-13500-z.

[37] M. O. Ajinaja, A. O. Adetunmbi, C. C. Ugwu, and O. S. Popoola, "Semantic similarity measure for topic modeling using latent Dirichlet allocation and collapsed Gibbs sampling," Iran Journal of Computer Science, vol. 6, no. 1, pp. 81-94, 2023, doi: 10.1007/s42044-022-00124-7.