2025
05

ORIGINAL ARTICLE

Clustering anxiety and depression among international student athletes during study abroad using transformer-based embeddings

Shuoyu Jing[ABCDE], Mohd Mahzan Awang[ABCD], Wan Ahmad Munsif Wan Pa[ABCD]
Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia

Authors' Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

Abstract

Background and Study Aim Student-athletes studying abroad experience increased risks of anxiety and depression as they balance academic and athletic responsibilities in cross-cultural environments. These psychological challenges may vary depending on individual adaptation, motivation, and environmental stressors. Although previous studies have applied network analysis to explore the structure of anxiety and depression symptoms, the relative effectiveness of advanced semantic approaches in identifying subgroup heterogeneity remains of practical interest. As a follow-up study, this research aimed to apply a transformer-based semantic embedding approach to cluster the mental health profiles of international student-athletes and to compare the model's performance with traditional clustering methods.

Material and Methods Data were collected from 219 Chinese international student-athletes who completed the Generalized Anxiety Disorder 7-item (GAD-7) and the Patient Health Questionnaire-9 (PHQ-9). Three models were compared: (1) K-means clustering on raw item scores, (2) K-means clustering after dimensionality reduction using Principal Component Analysis (PCA), and (3) K-means clustering on pseudo-text representations embedded via a transformer model, followed by PCA and K-means. Internal validity was assessed with silhouette scores. Between-cluster differences were analyzed using t-tests with Holm correction, effect sizes (Cohen's d), and cluster profiles.

Results Model 3 (Transformer embeddings + PCA + K-means) outperformed Models 1 and 2, achieving the highest silhouette score (0.391). Visualization in 2D and 3D projections confirmed clearer separation. Three clusters were identified: Cluster 2 (high symptoms), Cluster 0 (intermediate), and Cluster 1 (low symptoms). Pairwise comparisons revealed significant differences across nearly all items.

Conclusions Transformer-based semantic embeddings provide an effective approach to clustering psychological symptoms, outperforming traditional numerical methods. The results indicate the heterogeneity of anxiety and depression subgroups among student-athletes during study abroad, offering valuable insights for targeted screening, early identification, and long-term monitoring.

Keywords: anxiety, depression, machine learning clustering, transformer embeddings, student-athlete.

Introduction

International student-athletes represent a unique population facing complex psychological and social challenges during their studies abroad. The combination of academic responsibilities, competitive athletic demands, and adaptation to new cultural environments creates conditions that may heighten emotional vulnerability and affect overall mental well-being. Managing this dual role requires sustained cognitive, physical, and emotional resources, often leading to increased stress and potential development of anxiety or depressive symptoms. Taken together, these interrelated factors determine the mental state and academic performance of student-athletes in international educational settings.

In this context, previous research on anxiety and depression among college students or athletes has been well established [1, 2, 3], but studies on the mental health of student-athletes during their study abroad period remain scarce [4]. This unique group, with dual identities as both college students and athletes, must manage the combined pressures of academic progress and athletic demands [5]. Additionally, they encounter the extra challenges of adapting to overseas academic and living environments [6, 7]. Within the context of studying abroad, multiple interacting factors such as academic pressure, training demands, cultural differences, language barriers, and insufficient support make student-athletes more susceptible to anxiety, depression, and emotional fluctuations [4, 7, 8]. Previous research has shown that issues such as cross-cultural adaptation, time differences, and limited social networks may increase mental health risks and emotional vulnerability during study abroad [6, 8, 9, 10].

Previous studies on anxiety and depression among college student-athletes have largely

doi:10.15561/20755279.2025.0506

relied on traditional statistical methods such as factor analysis, regression analysis, and structural equation modeling [2, 3]. In recent years, emerging technologies like machine learning have increasingly been applied to mental health research, expanding the exploration of complex symptom networks and underlying mechanisms [11, 12]. However, studies on subpopulation heterogeneity remain limited, and most still rely on clustering analysis based on numerical matrices, which fails to fully capture the semantic differences behind symptom expressions. At the same time, the rapid development of artificial intelligence, particularly large language models (LLMs) based on deep learning, has created new methodological opportunities for mental health and sports research [12, 13, 14]. These models can capture contextual information within language through semantic embeddings, transforming symptom differences from simple numerical data into meaningful health characteristics with richer semantic dimensions [15]. This growing research direction opens new possibilities for identifying symptom patterns, detecting potential subgroups, and improving the interpretability of mental health assessment clusters.

The ongoing development in mental health emphasizes dimension-based approaches to psychology, moving beyond traditional diagnostic categories to capture the heterogeneity of symptoms across different populations [16, 17]. Concurrently, the stepped care model has emerged as a research framework in psychology, education, and sports medicine, advocating that individuals receive interventions commensurate with their symptom severity [18, 19, 20]. This perspective aligns with the present study's use of machine learning-based clustering to analyze anxiety and depression symptoms among student-athletes. Based on these theoretical models, a data-driven machine learning paradigm grounded in symptom dimensions and stepped care principles was applied. By utilizing transformer-based semantic embeddings with widely used screening tools [Generalized Anxiety Disorder 7-item scale (GAD-7), Patient Health Questionnaire-9 (PHQ-9)], the analysis sought to identify natural groupings among student-athletes across a spectrum of mental health risks. This approach extends the new model to educational and sports contexts and demonstrates how machine learning can be integrated into actionable analytical frameworks through symptom-based perspectives [17, 21]. It also provides opportunities for early monitoring and tiered intervention of mental health issues among student-athletes.

Recent advances in mental health and sports research increasingly integrate traditional statistical approaches with deep learning-based analytical frameworks [12, 13, 14, 22, 23]. Conventional clustering techniques, including K-means and dimensionality reduction methods such as Principal Component Analysis, remain effective for exploring structural relationships within psychological data [22, 23]. However, emerging transformer architectures based on self-attention mechanisms provide a means to capture subtle semantic variations underlying psychological responses, offering a more context-sensitive reflection of health status [15, 24, 25]. Through semantic embeddings, these models represent scale items as multidimensional linguistic constructs rather than isolated numerical scores, enriching the interpretability of psychological symptom patterns [25]. The feasibility and reliability of such semantic approaches have been validated across studies in mental health, sports science, and diverse populations [12, 14, 26]. As large language models continue to evolve, they open possibilities for analyzing extensive datasets, identifying latent clusters, and supporting continuous monitoring in sports and mental health domains [12, 13, 27]. This conceptual shift toward language-based representations of psychological data provides a foundation for deeper exploration of individual variability in mental health profiles among student-athletes.

Analysis of research findings has shown that the mental health of student-athletes studying abroad is influenced by a complex interplay of emotional, academic, and cultural factors. Researchers emphasize that while traditional approaches have revealed the structural characteristics of anxiety and depression symptoms, their ability to capture subgroup variability remains limited. Deep learning and semantic embedding methods offer promising directions for advancing the precision of psychological analysis and for identifying latent patterns within symptom dimensions. Therefore, the present study aims to apply transformer-based semantic embedding methods to cluster the mental health profiles of international student-athletes and to compare their performance with traditional clustering approaches.

Materials and Methods

Participants

This study is a follow-up to a previous investigation that examined anxiety and depression among 219 Chinese student-athletes during their study abroad period [4]. Anxiety and depression were assessed using the GAD-7 and PHQ-9 scales [28, 29]. Participants were Chinese college students aged 18 years and above who were enrolled in undergraduate or postgraduate programs in Malaysia and held athlete qualifications recognized by the General Administration of Sport of China [30]. Recruitment of student-athletes was conducted voluntarily through the distribution of recruitment information via university channels and overseas student communities. The study was approved by the Faculty of Education at the National University of Malaysia. All participants provided informed consent online prior to completing the questionnaire. Detailed sociodemographic characteristics are presented in Table 1.

Table 1. Demographic information of participants

Variable	Detailed characteristics	N (%)
Sports program	Ball	77 (35.1)
	Track and field	102 (46.5)
	Water sports	8 (3.7)
	Combat sports	3 (1.4)
	Dance sports	10 (4.6)
	Gymnastics	7 (3.1)
	Strength sports	12 (5.5)
Educational level	Degree	120 (54.7)
	Master	42 (19.2)
	Doctoral	57 (26.1)
Gender	Male	131 (59.8)
	Female	88 (40.2)
Athlete level	Elite athlete	3 (1.4)
	Division I	25 (11.4)
	Division II	161 (73.5)
	Division III	30 (13.7)
	Total	N = 219

This research adopts a novel analytical perspective by utilizing machine learning clustering methods to explore potential symptom subgroups. Three distinct models were compared:

1. K-means clustering applied directly to raw item datasets;
2. K-means clustering performed after dimensionality reduction using Principal Component Analysis (PCA);
3. K-means clustering conducted after converting item scores into pseudo-text input for a large language model (BERT), extracting semantic embeddings, and subsequently applying PCA and K-means clustering.

Specifically, traditional numeric item scores were transformed into semantically meaningful text representations to enable language model embeddings. For example, for the GAD1 item, a participant's response was reformulated as "This student-athlete felt nervous, anxious, or on edge nearly every day during the last two weeks." Due to space constraints in this article, the complete embedded text based on the GAD-7 and PHQ-9 scales is provided as code available in the GitHub link below. This approach moves beyond purely numeric inputs by converting responses into semantic text that can be modeled and vectorized through the transformer architecture [31]. Model 3 represents a Transformer-based semantic embedding model (sentence-transformers/all-roberta-large-v1-5.0.0) used to encode the data. The present research focuses on comparing different methodological paradigms rather than developing new NLP models; therefore, open-source, reproducible, general-purpose models were utilized. The use of mature models ensures transparency and reproducibility of analyses, and such applications are common in structured text response research. This model is a variant of RoBERTa pre-trained on a large corpus [32]. It employs the Transformer (4.53.1) architecture and a multi-head self-attention mechanism to capture contextual semantics and linguistic nuances [33]. Compared to raw numerical matrices, Transformer embeddings represent the semantic meaning of responses (e.g., "not at all" vs. "almost every day") in a continuous vector space with interpretable distances [33], facilitating clearer interpretation of the meaning conveyed by pseudo-text statements.

In Models 2 and 3, principal component analysis (PCA) was applied prior to clustering to reduce the dimensionality of the original feature space. PCA identifies latent components that explain the maximum variance in the data, improving computational efficiency while reducing noise [14]. In addition to dimensionality reduction, PCA allows intuitive two- and three-dimensional projections of high-dimensional embeddings, facilitating visual inspection of clustering structures and inter-group separation [34]. Furthermore, interpretability was enhanced by evaluating the clustering effectiveness of each model through internal validity measures and pairwise t-tests, as well as by generating cluster portraits to illustrate anxiety and depression patterns.

Statistical Analysis

All statistical analyses and model construction were conducted in Python (3.12.2) using the packages *scipy* (1.14.0), *statsmodels* (0.14.5), *scikit-learn* (1.7.0), and other supporting libraries. Independent-samples t-tests were used to compare clusters, with multiple comparisons corrected using the Holm method. Cohen's d was calculated as the standardized effect size.

Internal validity was evaluated using the silhouette score, which quantifies cohesion within clusters and separation between clusters. Silhouette values range from −1 to 1, with higher values indicating more compact and well-separated clusters [35]. Silhouette scores were compared across the three analytic models to assess relative clustering quality. For each item within clusters, descriptive statistics were calculated, including means, standard deviations (SD), and 95% confidence intervals (95% CI). These results provided an overview of symptom levels and variability across clusters. Confidence intervals were computed using t-distribution–based estimates [36].

Pairwise comparisons between clusters were performed using independent-samples *t*-tests, with Holm correction applied to adjust for multiple comparisons [37]. To complement significance testing, pairwise effect sizes were reported as Cohen's *d* [38], with 95% bootstrap confidence intervals based on 1,000 resamples, quantifying the magnitude and precision of between-cluster differences. Results were summarized in tables, and forest plots were generated to visually represent effect sizes across items [39].

Results

Model Validity Comparison

As shown in Table 2, Model 1 (Raw + K-means) and Model 2 (Raw + PCA + K-means) achieved their highest silhouette scores when the number of clusters was set to two, with values of 0.208 and 0.356, respectively. In contrast, Model 3 (Raw + Transformer-based embeddings + PCA + K-means) demonstrated its best performance with three clusters, yielding the highest silhouette score of 0.391. These results indicate that the semantic embedding approach produced more coherent and better-separated clusters when allowing for a slightly higher cluster solution.

Table 2. Silhouette coefficient for the optimal number of clusters for each model

	Model	Cluster	Silhouette Score
1	Raw + K-means	2	0.208
2	Raw + PCA + K-means	2	0.356
3	BERT (TRAN) + PCA + K-means	3	0.391

To further illustrate the clustering structure of Model 3, the three-cluster solution was projected into a reduced feature space using PCA. As shown in Figure 1 (2D projection) and Figure 2 (3D projection), the three clusters formed relatively distinct groups with minimal overlap, suggesting improved separation compared with the raw data–based models. Both the quantitative validity indices and the visual inspection of the embedding space consistently demonstrated that Model 3 provided the most effective clustering solution.

Cluster Characterization and Differences

Descriptive statistics and between-cluster comparisons are summarized in Table 3. Clear differentiation was observed between clusters across all GAD and PHQ items. Cluster 2 consistently demonstrated the highest mean scores, while Cluster 1 showed the lowest scores across most items. Cluster 0 generally fell between these two groups. For example, for GAD1, the mean scores were 2.17 (SD = 0.68, 95% CI [1.98–2.36]) for Cluster 2, 1.19 (SD = 0.44, 95% CI [1.10–1.28]) for Cluster 0, and 0.72 (SD = 0.52, 95% CI [0.60–0.85]) for Cluster 1.

Pairwise comparisons confirmed statistically significant differences across clusters, with large effect sizes in several items. Cluster 2 scored significantly higher than Cluster 1 on nearly all items, with Cohen's *d* ranging from 1.30 to 2.45 (all $p < .001$). Cluster 2 also scored higher than Cluster 0 on multiple items, with effect sizes ranging from moderate to large ($d \approx 0.45–1.85$). In addition, Cluster 0 scored significantly higher than Cluster 1 on most items ($d \approx 0.42–1.99$).

Figures 3 and 4 summarize the pairwise effect sizes across all GAD-7 and PHQ-9 items. The forest plots provide an intuitive visualization of the magnitude and direction of between-cluster differences, with horizontal bars indicating the 95% confidence intervals. Values to the right of the zero line indicate higher scores in the first-named cluster, whereas values to the left indicate higher

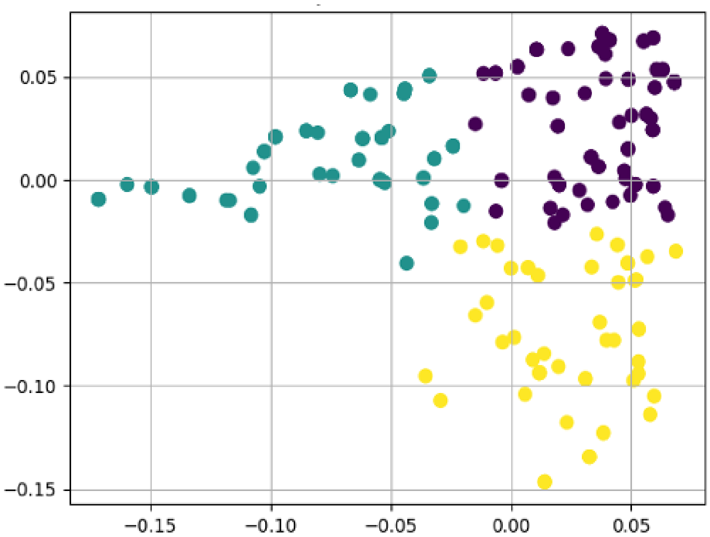

Figure 1. Cluster visualization of Model 3 (2D PAC projection)

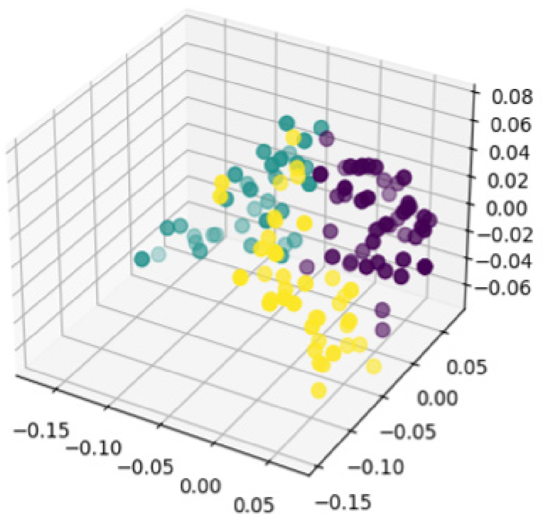

Figure 2. Cluster visualization of Model 3 (3D PAC projection)

Table 3. Cluster Characterization and Between-Cluster Differences.

Item	Cluster 0 (Mean ± SD, 95% CI)	Cluster 1 (Mean ± SD, 95% CI)	Cluster 2 (Mean ± SD, 95% CI)	Group differences (pairwise d)
GAD1	1.19±0.44 [1.10–1.28]	0.72±0.52 [0.60–0.85]	2.17±0.68 [1.98–2.36]	2>1 (d=2.44***); 2>0 (d=1.85***); 0>1 (d=0.99***)
GAD2	1.03±0.30 [0.97–1.09]	0.40±0.49 [0.28–0.52]	1.75±0.95 [1.49–2.01]	2>1 (d=1.85***); 0>1 (d=1.63***); 2>0 (d=1.20***)
GAD3	1.22±0.50 [1.12–1.32]	0.78±0.48 [0.66–0.90]	2.19±0.77 [1.98–2.41]	2>1 (d=2.25***); 2>0 (d=1.61***); 0>1 (d=0.88***)
GAD4	1.10±0.56 [0.99–1.21]	0.71±0.79 [0.51–0.90]	1.98±0.64 [1.80–2.16]	2>1 (d=1.76***); 2>0 (d=1.50***); 0>1 (d=0.60***)
GAD5	1.07±0.45 [0.98–1.16]	0.62±0.70 [0.44–0.79]	1.44±0.54 [1.29–1.59]	2>1 (d=1.30***); 0>1 (d=0.81***); 2>0 (d=0.77***)
GAD6	0.97±0.30 [0.91–1.03]	0.35±0.54 [0.22–0.49]	1.38±0.84 [1.15–1.62]	0>1 (d=1.50***); 2>1 (d=1.49***); 2>0 (d=0.76**)
GAD7	0.96±0.40 [0.88–1.04]	0.52±0.64 [0.36–0.68]	1.69±0.64 [1.51–1.87]	2>1 (d=1.82***); 2>0 (d=1.48***); 0>1 (d=0.86***)
PHQ1	1.03±0.48 [0.94–1.12]	0.58±0.83 [0.38–0.79]	1.92±0.68 [1.73–2.11]	2>1 (d=1.75***); 2>0 (d=1.61***); 0>1 (d=0.70***)
PHQ2	0.86±0.40 [0.78–0.94]	0.12±0.33 [0.04–0.21]	1.40±0.69 [1.21–1.60]	2>1 (d=2.45***); 0>1 (d=1.97***); 2>0 (d=1.05***)
PHQ3	1.22±0.58 [1.10–1.33]	0.69±0.71 [0.52–0.87]	2.00±0.71 [1.80–2.20]	2>1 (d=1.84***); 2>0 (d=1.25***); 0>1 (d=0.83***)
PHQ4	1.20±0.47 [1.11–1.29]	0.68±0.64 [0.52–0.84]	2.02±0.67 [1.83–2.21]	2>1 (d=2.05***); 2>0 (d=1.50***); 0>1 (d=0.96***)
PHQ5	0.85±0.62 [0.73–0.97]	0.35±0.57 [0.21–0.50]	1.19±0.97 [0.92–1.46]	2>1 (d=1.08***); 0>1 (d=0.83***); 2>0 (d=0.45*)
PHQ6	0.79±0.41 [0.71–0.87]	0.22±0.41 [0.11–0.32]	1.48±0.83 [1.25–1.71]	2>1 (d=2.00***); 0>1 (d=1.41***); 2>0 (d=1.18***)
PHQ7	1.02±0.28 [0.96–1.08]	0.72±0.67 [0.56–0.89]	1.62±0.63 [1.44–1.79]	2>0 (d=1.38***); 2>1 (d=1.36***); 0>1 (d=0.62**)
PHQ8	0.93±0.41 [0.85–1.01]	0.15±0.36 [0.06–0.24]	1.27±0.77 [1.05–1.48]	0>1 (d=1.99***); 2>1 (d=1.92***); 2>0 (d=0.61**)
PHQ9	0.10±0.30 [0.04–0.16]	0.00±0.00 [0.00–0.00]	0.35±0.68 [0.16–0.54]	2>1 (d=0.76**); 2>0 (d=0.53*); 0>1 (d=0.42**)
	N=101	N=66	N=52	

Note. *** p < 0.001, ** p < 0.01, * p < 0.05; d = Cohen's d for pairwise group comparisons

scores in the second cluster. Larger absolute effect sizes reflect more pronounced differences between clusters. For instance, in the GAD1 plot shown in Figure 3, Cluster 2 scored substantially higher than both Cluster 0 and Cluster 1, with large effect sizes ($d > 1.8$), while Cluster 0 also exceeded Cluster 1 with a moderate effect size ($d \approx 1.0$). These plots offer a visual summary of the robustness and consistency of between-cluster differences across the full item set.

Discussion

The purpose of this study was to apply transformer-based semantic embedding methods to cluster the mental health profiles of international student-athletes and to compare their performance with traditional clustering approaches. The analysis revealed that the transformer-based model, combined with PCA and K-means clustering, achieved the highest internal validity, as indicated by the silhouette score. This model produced more coherent and better-separated clusters compared with those based on raw or numerically reduced data. Three distinct subgroups of student-athletes were identified, reflecting high, intermediate, and low levels of anxiety and depression symptoms. These findings demonstrate the potential of semantic embedding techniques to enhance the interpretation of psychological symptom structures in the context of mental health assessment among student-athletes studying abroad.

This study presents a clustering analysis

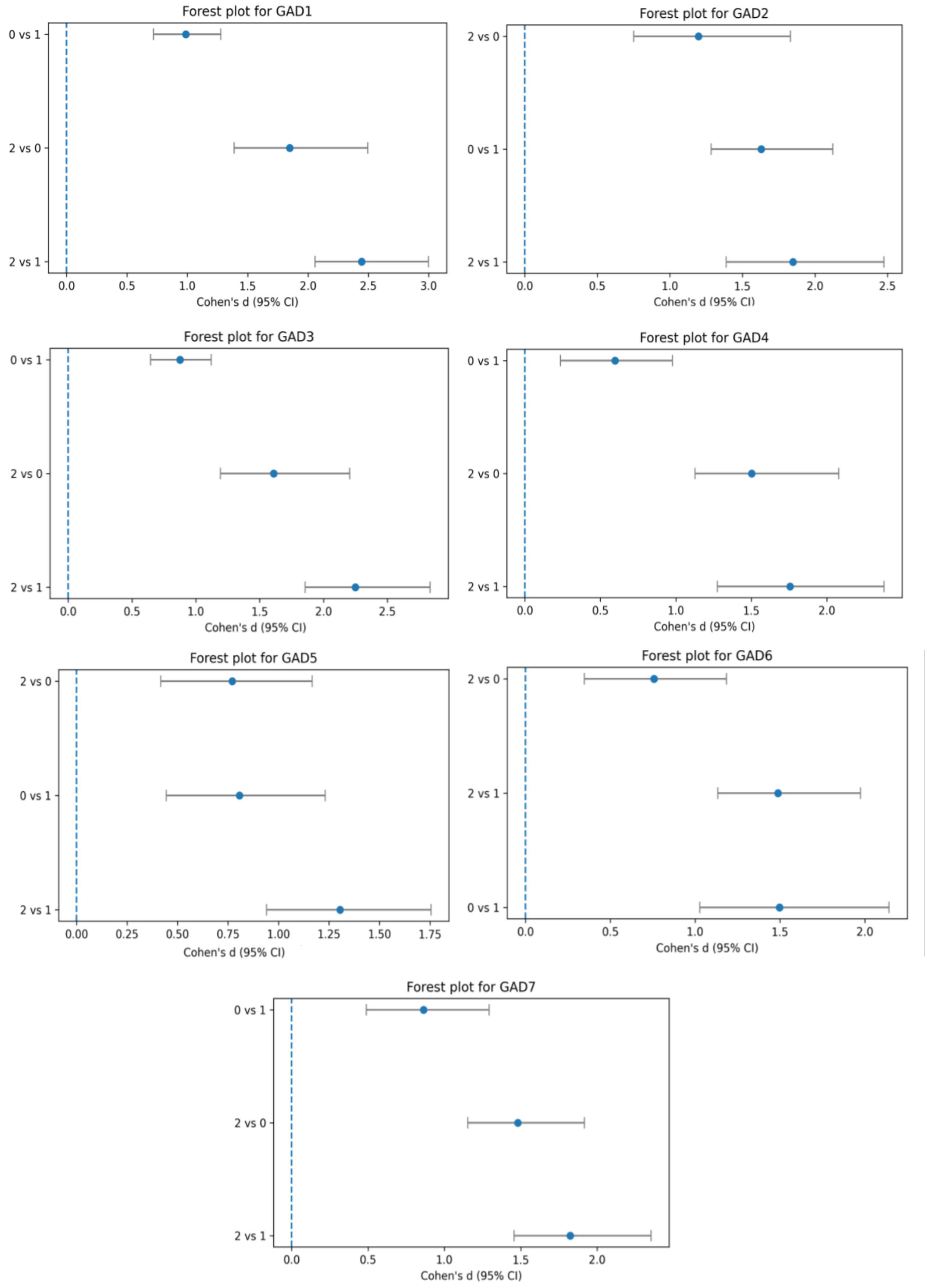

Figure 3. GAD7 Pairwise Effect Size Forest Plot

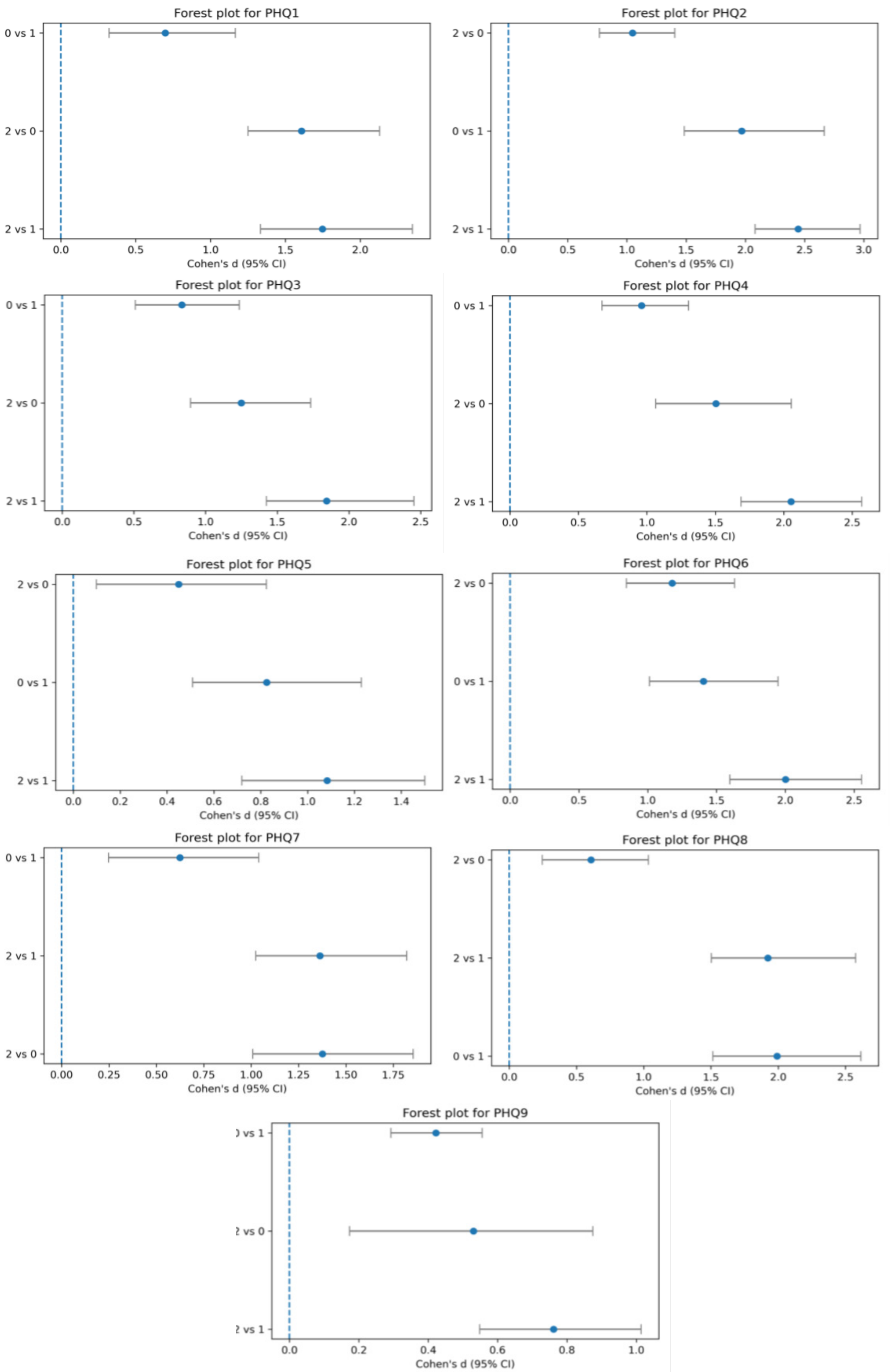

Figure 4. PHQ9 Pairwise Effect Size Forest

approach in the field of student-athlete mental health, applying a machine learning method based on Transformer embeddings (Model 3) to group anxiety and depression symptoms during study abroad. This approach addresses the limitations of traditional statistical and network analysis models in capturing the semantic content of symptoms. The results showed that the Transformer (BERT)-embedding-based model achieved higher internal validity and clearer visual separation compared with raw numerical clustering methods. Model 3 identified three distinct clusters of psychological symptom distribution: Cluster 2 exhibited elevated anxiety and depression levels, representing a potential high-risk group; Cluster 1 showed the lowest symptom levels, indicating a relatively low-risk group; and Cluster 0 represented an intermediate level, possibly corresponding to subclinical or transitional states. These findings reflect the heterogeneity of psychological symptoms and provide typological evidence relevant to understanding the mental health characteristics of student-athletes at different performance levels. Furthermore, the clustering results suggest that the use of large language models may enhance the precision of screening for anxiety and depression risks among student-athletes in educational contexts. The subdivision of clusters also points to subtle differences between subgroups, implying that machine learning–based monitoring approaches can complement, rather than replace, traditional clustering methods in this area.

Previous research revealed the network structure of anxiety and depression symptoms among student-athletes during their study abroad period [4]. This follow-up study incorporates a machine learning–based clustering approach, addressing the limitation of network analysis being restricted to symptom correlations [40]. Unlike traditional clustering methods, the current analysis applies Transformer-based semantic embeddings to convert item responses into vectorized semantic representations [14, 32]. This approach integrates structured data with a linguistic-semantic dimension, allowing for more precise differentiation of symptom patterns and identification of detailed mental health profile clusters [41]. Compared with modeling that relies solely on PCA or raw matrices, Transformer embeddings capture semantic gradient differences across scales [42]. For example, responses such as "not at all" and "almost every day" are positioned at natural and interpretable distances within the semantic space, reflecting meaningful variation in symptom intensity. This advantage is reflected not only in quantitative indices (e.g., higher silhouette scores) but also in the visual clarity of boundaries within 2D and 3D projections, illustrating the distinct separation between clusters [34]. Overall, this study highlights the applicability of natural language processing techniques in sports and mental health research and contributes methodological tools for future exploration of psychological characteristics among student-athletes.

The three-cluster structure identified in this study has both clinical and practical relevance. The high-symptom cluster (Cluster 2) represents student-athletes with markedly elevated levels of psychological distress. These individuals may be at increased risk of developing anxiety or depressive disorders and require timely attention and support from educational institutions and professionals [43]. The low-symptom cluster (Cluster 1) demonstrates relatively mild symptom levels. When considered alongside the GAD-7 and PHQ-9 scale thresholds, this cluster can serve as a reference group for mental health benchmarking and stratification. The intermediate cluster (Cluster 0) reflects a subclinical level, suggesting that this group may be in a transitional stage of symptom progression or recovery. This provides an important observation window for continuous monitoring and early intervention during study abroad [44]. From a practical standpoint, this analytical paradigm may be applied in mental health screening and monitoring systems within higher education and sports environments, similar in concept to electronic log monitoring used in clinical settings [15].

Early identification of student-athletes experiencing abnormal psychological states can safeguard both their academic progress and athletic performance. It also serves as a preventive measure against potential psychological fluctuations that may arise in complex overseas learning and living environments [45, 46]. Existing research indicates that international students are more vulnerable to mental health risks due to factors such as cultural adaptation, language barriers, and limited social support [8, 47, 48]. Student-athletes, who face dual pressures from academic and athletic commitments, experience even greater risks during study abroad [4, 7, 49]. Therefore, applying deep learning–based mental health profiling and clustering in daily monitoring can facilitate early intervention and precise resource allocation [15]. Moreover, this approach may help prevent the negative effects of anxiety and depression on both academic and athletic performance.

The findings of this study indicate that a machine learning–based, step-up care-oriented research paradigm can be applied to managing the mental health of student-athletes in overseas educational contexts. From a theoretical perspective, the clustering results provide empirical support for the psychopathology and psychological dimension model, showing that mental health symptoms among student-athletes form clustering patterns beyond a simple dichotomy. When integrated with the stepped care framework, this approach enables systematic alignment of support levels with the corresponding clustering tiers [17]. For example, preventive education may be offered to low-symptom groups, continuous monitoring and counseling to transitional groups, and specialized treatment and care to high-risk groups [17, 21]. Placing this machine learning approach within the framework of stepped care emphasizes not only the methodological contribution of semantic embeddings but also their conceptual integration into established theoretical models applied to new populations and domains [17, 21, 50]. In educational and sports environments, timely allocation of psychological support resources and athlete mental health monitoring systems remains essential [51, 52], with this paradigm serving as a bridge between monitoring precision and traditional analytical approaches.

This research highlights three main aspects of contribution. First, at the methodological level, Transformer-based semantic embeddings were introduced into the study of student-athletes' mental health during overseas education, showing the potential of combining natural language processing with traditional psychometric scales. Second, regarding research participants, the focus on the distinctive group of international student-athletes provides additional evidence for understanding the heterogeneity of their mental health during study abroad. Third, at the empirical level, the comparative analysis of three modeling approaches demonstrated that Transformer-based clustering achieved higher validity and discriminative capacity, offering practical implications. Compared with previous network analysis emphasizing the structural interdependence of symptoms [4], the present clustering method shifts the focus to the hierarchical grouping of anxiety and depression at the individual level. This complementary approach links emerging subgroups to potential intervention tiers, increasing the applicability of the findings in educational settings. Furthermore, it extends the scope of previous research from a traditional mental health framework to practical applications within educational and athletic environments, such as stratified monitoring and targeted intervention.

Overall, the findings of this study contribute to the growing body of evidence supporting the integration of machine learning techniques into mental health research among student-athletes. The application of transformer-based semantic embeddings offers a data-driven framework for analyzing symptom heterogeneity while maintaining consistency with established theoretical models such as stepped care and dimensional psychopathology. These methodological advancements illustrate how modern computational tools can complement traditional assessment approaches, deepening the understanding of mental health variability within sports and educational contexts.

Limitations and Future Directions

However, this study also has several limitations. First, the sample size was relatively small and limited to student-athletes from a specific background, which may restrict the external applicability of the findings. Second, as a cross-sectional study, it cannot establish causality or capture dynamic changes in symptoms over time. Third, the Transformer model "sentence-transformers/all-roberta-large-v1" was trained on general-purpose corpora, requiring further validation in alternative contexts. Additionally, text transformation and embedding using different large language models or psychological scales may produce varying clustering outcomes, with their generalizability remaining uncertain. Future research could fine-tune large language models on specialized sports psychology corpora to improve their performance across different athlete populations.

Conclusions

This study identified heterogeneous patterns of anxiety and depression symptoms among student-athletes during their study abroad period using Transformer-based semantic embedding clustering analysis. The results showed that the Transformer embedding model performed better than traditional methods in terms of internal validity and visual separability, forming three distinct groups: high-symptom, intermediate-level, and low-symptom. These findings provide additional evidence for understanding and stratifying the mental health management of student-athletes. The study also illustrates the applicability of natural language processing techniques in psychological and sports research, expanding methodological approaches for mental health monitoring and intervention. Despite limitations related to sample size, context, and the need for further validation across different language models and scale conversions, this research offers useful directions for future investigation. Further studies should validate these findings in larger and more diverse samples, employ longitudinal designs to observe symptom changes over time, and test model robustness when trained on domain-specific corpora. Through these efforts, deep learning–based methods may contribute to improving the efficiency of mental health screening and intervention in sports and educational environments.

Acknowledgments

The authors express their sincere gratitude to all participants and experts who contributed to this research. The core code and dataset supporting this study are available from the corresponding author upon reasonable request.

Conflict of Interest

The author declares no conflict of interest.

Code Availability

The core code for this study, along with the complete pseudo-text embeddings and construction details, is available at: https://github.com/Ech0Jing/Athlete-cluster.

References

4. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders*, 2015;173: 90–96. https://doi.org/10.1016/j.jad.2014.10.054
5. Beisecker L, Harrison P, Josephson M, DeFreese JD. Depression, anxiety and stress among female student-athletes: a systematic review and meta-analysis. *British Journal of Sports Medicine*, 2024;58(5): 278–285. https://doi.org/10.1136/bjsports-2023-107328
6. Armstrong SN, Burcin MM, Bjerke W, Early J. Depression in student athletes: A particularly at-risk group? A systematic review of the literature. *Athletic Insight*. 2015;7(2):177-193.
7. Jing S, Wan Pa WAM, Awang MM. Anxiety and depression among Chinese international student-athletes during study abroad: a psychological network approach. *Physical Education of Students*, 2025;29(1): 27–38. https://doi.org/10.15561/20755279.2025.0103
8. Lopes Dos Santos M, Uftring M, Stahl CA, Lockie RG, Alvar B, Mann JB, et al. Stress in Academic and Athletic Performance in Collegiate Athletes: A Narrative Review of Sources and Monitoring Strategies. *Frontiers in Sports and Active Living*, 2020;2: 42. https://doi.org/10.3389/fspor.2020.00042
9. Sawir E, Marginson S, Deumert A, Nyland C, Ramia G. Loneliness and International Students: An Australian Study. *Journal of Studies in International Education*, 2008;12(2): 148–180. https://doi.org/10.1177/1028315307299699
10. Newell EM. International Student–Athlete Adjustment Issues: Advising Recommendations for Effective Transitions. *NACADA Journal*, 2015;35(2): 36–47. https://doi.org/10.12930/NACADA-14-015
11. Misirlis N, Zwaan MH, Weber D. *International students' loneliness, depression and stress levels in COVID-19 crisis. The role of social media and the host university*. 2020. https://doi.org/10.48550/ARXIV.2005.12806
12. Brown L. Language and Anxiety: An Ethnographic Study of International Postgraduate Students. *Evaluation & Research in Education*, 2008;21(2): 75–95. https://doi.org/10.1080/09500790802152167
13. Tan Y, Wu Z, Qu X, Liu Y, Peng L, Ge Y, et al. Influencing Factors of International Students' Anxiety Under Online Learning During the COVID-19 Pandemic: A Cross-Sectional Study of 1,090 Chinese International Students. *Frontiers in Psychology*, 2022;13: 860289. https://doi.org/10.3389/fpsyg.2022.860289
14. Rossi A, Pappalardo L, Cintia P. A Narrative Review for a Machine Learning Application in Sports: An Example Based on Injury Forecasting in Soccer. *Sports*, 2021;10(1): 5. https://doi.org/10.3390/sports10010005
15. Lu L. Big data analysis of mental health intervention effects in student-athletes: based on data mining techniques and affective computing algorithms. *Multimedia Tools and Applications*, 2024;84(17): 18547–18565. https://doi.org/10.1007/s11042-024-19786-5
16. Zhao Z, Wang J. Exploring the Potential of Large Language Model in Predictive Mental Health Diagnosis of Athletes. *Advances in Education, Humanities and Social Science Research*, 2024;12(1): 342. https://doi.org/10.56028/aehssr.12.1.342.2024
17. Yun HJ, Jang N, Jeon M. Deep learning-based tennis match type clustering. *BMC Sports Science, Medicine and Rehabilitation*, 2025;17(1): 104. https://doi.org/10.1186/s13102-025-01147-w
18. Niu H, Omitaomu OA, Langston MA, Olama M, Ozmen O, Klasky HB, et al. EHR-BERT: A BERT-based model for effective anomaly detection in electronic health records. *Journal of Biomedical Informatics*, 2024;150: 104605. https://doi.org/10.1016/j.jbi.2024.104605
19. De Boeck P, Wilson M, Acton GS. A Conceptual and Psychometric Framework for Distinguishing Categories and Dimensions. *Psychological Review*, 2005;112(1): 129–158. https://doi.org/10.1037/0033-295X.112.1.129
20. Delgadillo J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified Care vs Stepped Care for Depression: A Cluster Randomized Clinical Trial. *JAMA Psychiatry*, 2022;79(2): 101. https://doi.org/10.1001/jamapsychiatry.2021.3539
21. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency: Narrative literature review. *British Journal of Psychiatry*, 2005;186(1): 11–17. https://doi.org/10.1192/bjp.186.1.11
22. Van Straten A, Hill J, Richards DA, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. *Psychological Medicine*, 2015;45(2): 231–246. https://doi.org/10.1017/S0033291714000701
23. Kaiser J, Walter N, Oppitz L, Braun B, Schmitz J. 2Steps4Health project: Promotion of mental health in junior professional sports: Study protocol. *Sports Psychiatry*, 2024;3(1): 23–30. https://doi.org/10.1024/2674-0052/a000066
24. Scodari BT, Chacko S, Matsumura R, Jacobson NC. Using machine learning to forecast symptom changes among subclinical depression patients receiving stepped care or usual care. *Journal of Affective Disorders*, 2023;340: 213–220.

https://doi.org/10.1016/j.jad.2023.08.004
25. Kodinariya TM, Makwana PR. Review on Determining Number of Cluster in K-Means Clustering. *Int J Adv Res Comput Sci Manag Stud.* 2013;1(6):90-95.
26. Ding C, He X. *K* -means clustering via principal component analysis. In: *Twenty-first international conference on Machine learning - ICML '04*, Banff, Alberta, Canada: ACM Press; 2004. P. 29. https://doi.org/10.1145/1015330.1015408
27. Wu W, Wang W, Jia X, Feng X. Transformer Autoencoder for K-means Efficient clustering. *Engineering Applications of Artificial Intelligence*, 2024;133: 108612. https://doi.org/10.1016/j.engappai.2024.108612
28. Holmes B, Raymer M, Banerjee T. Extraction of patients subpopulations with psychiatric symptoms using a transformer architecture. In: *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA: IEEE; 2024. P. 1–4. https://doi.org/10.1109/EMBC53108.2024.10781648
29. Torres-Luque G, Ramirez A, Cabello-Manrique D, Nikolaidis TP, Alvero-Cruz JR. Match analysis of elite players during paddle tennis competition. *International Journal of Performance Analysis in Sport*, 2015;15(3): 1135–1144. https://doi.org/10.1080/24748668.2015.11868857
30. Gotthardt MU. *Clustering Large-Scale 3D Football Player Skeleton Data: Investigating Differences in Player Pose Distributions and their Correspondence to Tracking Performance Levels* [Master's thesis]. Stockholm (Sweden): KTH, School of Electrical Engineering and Computer Science; 2024.
31. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 2001;16(9): 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 2006;166(10): 1092. https://doi.org/10.1001/archinte.166.10.1092
33. Song B, Martínez-Aranda LM, Leiva-Arcas A, Sánchez-Pato A. The evolution of Chinese high-performance student-athletes' admission, cultivation and management policies. *International Journal of Sport Policy and Politics*, 2024;16(1): 151–175. https://doi.org/10.1080/19406940.2023.2273350
34. Bhatia S. Exploring variability in risk taking with large language models. *Journal of Experimental Psychology: General*, 2024;153(7): 1838–1860. https://doi.org/10.1037/xge0001607
35. Li B, Zhou H, He J, Wang M, Yang Y, Li L. *On the Sentence Embeddings from Pre-trained Language Models*. 2020. https://doi.org/10.48550/ARXIV.2011.05864
36. Khosa S, Mehmood A, Rizwan M. Unifying Sentence Transformer Embedding and Softmax Voting Ensemble for Accurate News Category Prediction. *Computers*, 2023;12(7): 137. https://doi.org/10.3390/computers12070137
37. Salloum S, Alhumaid K, Salloum A, Shaalan K. K-means Clustering of Tweet Emotions: A 2D PCA Visualization Approach. *Procedia Computer Science*, 2024;244: 30–36. https://doi.org/10.1016/j.procs.2024.10.175
38. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 1987;20: 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
39. Cumming G, Finch S. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist*, 2005;60(2): 170–180. https://doi.org/10.1037/0003-066X.60.2.170
40. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979;6(2):65-70.
41. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.*. 0 edn Routledge; 2013. https://doi.org/10.4324/9780203771587
42. Viechtbauer W. Conducting Meta-Analyses in *R* with the metafor Package. *Journal of Statistical Software*, 2010;36(3). https://doi.org/10.18637/jss.v036.i03
43. McNally RJ. Can network analysis transform psychopathology? *Behaviour Research and Therapy*, 2016;86: 95–104. https://doi.org/10.1016/j.brat.2016.06.006
44. Tareaf RB, AbuJarour M, Engelman T, Liermann P, Klotz J. Accelerating Contextualization in AI Large Language Models Using Vector Databases. In: *2024 International Conference on Information Networking (ICOIN)*, Ho Chi Minh City, Vietnam: IEEE; 2024. P. 316–321.https://doi.org/10.1109/ICOIN59985.2024.
45. Miranda O, Kiehl SM, Qi X, Brannock MD, Kosten T, Ryan ND, et al. Enhancing post-traumatic stress disorder patient assessment: leveraging natural language processing for research of domain criteria identification using electronic medical records. *BMC Medical Informatics and Decision Making*, 2024;24(1): 154. https://doi.org/10.1186/s12911-024-02554-8
46. Geiger S, Jahre LM, Aufderlandwehr J, Krakowczyk JB, Esser AJ, Mühlbauer T, et al. Mental health symptoms in German elite athletes: a network analysis. *Frontiers in Psychology*, 2023;14: 1243804. https://doi.org/10.3389/fpsyg.2023.1243804
47. Roitblat Y, Cleminson R, Kavin A, Schonberger E, Shterenshis M. Assessment of anxiety in adolescents involved in a study abroad program: a prospective study. *International Journal of Adolescent Medicine and Health*, 2020;32(2): 20170101. https://doi.org/10.1515/ijamh-2017-0101
48. Minutillo S, Cleary M, P. Hills A, Visentin D. Mental Health Considerations for International Students. *Issues in Mental Health Nursing*, 2020;41(6): 494–499. https://doi.org/10.1080/01612840.2020.1716123
49. Ansari Lari S, Zumot MS, Fredericks S. Navigating mental health challenges in international university students: adapting to life transitions. *Frontiers in Psychiatry*, 2025;16: 1574953. https://doi.org/10.3389/fpsyt.2025.1574953
50. Liw L, Ciftci A, Kim T. Cultural values, shame and guilt, and expressive suppression as predictors of depression. *International Journal*

of Intercultural Relations, 2022;89: 90–99. https://doi.org/10.1016/j.ijintrel.2022.05.005
51. Akiba D, Perrone M, Almendral C. Study Abroad Angst: A Literature Review on the Mental Health of International Students During COVID-19. *International Journal of Environmental Research and Public Health*, 2024;21(12): 1562. https://doi.org/10.3390/ijerph21121562
52. Wolanin A, Gross M, Hong E. Depression in Athletes: Prevalence and Risk Factors. *Current Sports Medicine Reports*, 2015;14(1): 56–60. https://doi.org/10.1249/JSR.0000000000000123
53. Wilhelm C, Steckelberg A, Rebitschek FG. Benefits and harms associated with the use of AI-related algorithmic decision-making systems by healthcare professionals: a systematic review. *The Lancet Regional Health - Europe*, 2025;48: 101145. https://doi.org/10.1016/j.lanepe.2024.101145
54. Balcombe L, De Leo D. Psychological Screening and Tracking of Athletes and Digital Mental Health Solutions in a Hybrid Model of Care: Mini Review. *JMIR Formative Research*, 2020;4(12): e22755. https://doi.org/10.2196/22755
55. Duffy A. University Student Mental Health: An Important Window of Opportunity for Prevention and Early Intervention. *The Canadian Journal of Psychiatry*, 2023;68(7): 495–498. https://doi.org/10.1177/07067437231183747

Information about the authors:

Shuoyu Jing; (Corresponding Author); https://orcid.org/0009-0001-1155-7504; p119216@siswa.ukm.edu.my; Faculty of Education, Universiti Kebangsaan Malaysia; Bangi, Malaysia.

Mohd Mahzan Awang; https://orcid.org/0000-0003-1791-7707; mahzan@ukm.edu.my; Faculty of Education, Universiti Kebangsaan Malaysia; Bangi, Malaysia.

Wan Ahmad Munsif Wan Pa; https://orcid.org/0000-0001-6218-4376; munsif@ukm.edu.my; Faculty of Education, Universiti Kebangsaan Malaysia; Bangi, Malaysia.

Cite this article as:
Jing S, Awang MM, Wan Pa WAM. Clustering anxiety and depression among international student athletes during study abroad using transformer-based embeddings. *Physical Education of Students,* 2025;29(5):383–394. https://doi.org/10.15561/20755279.2025.0506

Received: 17.09.2025

Accepted: 18.10.2025; Published: 30.10.2025