Embedding-Based Classification of Athlete-Focused Mental Health Responses: A Preliminary Study

Adhuna Devineni[1]

Received April 11, 2026
Accepted August 13, 2026
Electronic access September 15, 2026

Prior mental health NLP work used social-media data, sport language models, sentence embeddings, and tweet-sized transformer output for the purposes of classification and representation. This initial comparative experiment aims to determine if various models may be able to cluster sport focused athlete mental health responses into four specific research labels: anxiety, ADHD, depression and PTSD. This experiment was inspired by a context where athlete response is influenced by sport-related stressors such as the pressure to perform and the sport stigma that athlete mental health and athlete help seeking studies reflect as impacting athletes' well-being. A new dataset was created which consisted of 33 pairs of prompts and answers, drawn from various resources related to athletes and sports psychology. Five classification approaches were compared: majority-class baseline, a candidate voting ensemble approach, a regulated LoRA fine-tuning run, sentence embeddings using all-MiniLM-L6-v2 sentence embeddings with a LinearSVC classifier, and MiniLM-L6-v2 with LinearSVC. The findings do not filter out any stop words and have not applied paraphrasing augmentation techniques. An accuracy of 87.9% and 95% CI between seventy-three percent and ninety-five percent with a macro-F1 of 0.87 and weighted-F1 of 0.88 were achieved in the MiniLM/LSVC classification run; 29 of 33 held-out examples were labeled correctly here. The regulated LoRA model achieved a 75% accuracy score, while the candidate voting approach achieved 73%, and the majority-class baseline achieved 39% accuracy. In general, this paper contributes a reproducibility focused comparison of several possible models at organizing sport-focused athlete mental health responses from a small corpus and supports the usage of the MiniLM/LSVC as a viable tool among the models investigated.

Keywords: athlete mental health, sentence embeddings, text classification, support vector classifier, sports psychology, NLP

Introduction

Athlete's mental health & help-seeking has been investigated into the following aspects: psychological balance, stigma, social support, training disruption, & aspects of the coach athlete relationship[1–5]. Within this study these context aspects in the same domain are utilized to analyze athlete-focused mental health responses using pressures, fatigue, focus issues, criticism, avoidance of emotions, stigma, coping, and distress.

The research investigates if athlete focused mental health responses can be classified into four research labels based on a small original data set with 33 prompt-response pairs. The creation of this data set is explained by the fact that no available public data set with the specified athlete-focused educational prompt-response format could be found. Transparency, cross-validation, complete performance metrics, and comparison of several models under consideration, such as majority class baseline, candidate voting ensemble, controlled LoRA, and MiniLM embedding/classifier approach[6,7] are emphasized in this study. The underlying motivation behind this research is self-awareness and self-help education. Better organization of responses will allow researchers and readers to recognize language patterns regarding the pressure, fatigue, inability to focus, stigma, coping, and distress.

This study is guided by 3 research questions. First, how well can the athlete-focused mental health responses be categorized into the anxiety, ADHD, depression, and PTSD labels based on the small, labeled data set? Secondly, how does each measure perform under same cross-validation conditions? Third, do the labels have any overlap and which labels are the hardest to distinguish?

Literature Review

Prior NLP work on mental health tends to involve social media corpora, benchmark data sets, or mental-health pretrained models. This work is different from the current work in the type of data, definition of the labels, and the evaluation scale used. Work that has been done on Twitter and Reddit involves classifying users or posts based on social media language indicative of depression, PTSD, stress, or self-diagnosis of men-

[1] *Cambridge Centre for International Research (CCIR)*

© The National High School Journal of Science 2026

tal problems[8–10]. The model MentalBERT measures the effect of mental health pretraining on mental health detection benchmarks, which require much more text compared to the current data set[11]. Sentence-BERT and MiniLM provide effective sentence representations, providing reasons for incorporating a fixed embedding classifier into this study, along with LoRA and other baselines[12,13].

Analysis of language in social media suggests that there is a correlation between the linguistic characteristics and mental health classifications in large-scale data that does not involve athletes[14]. Experimental studies that involve testing conversational agents to promote mental health provide results that are based on human assessments; hence, the current study does not make any claims related to response creation without performing similar analysis[15,16]. Further, research on athletes identifies linguistic characteristics that are linked to psychological balance and sport-specific mental health profiles of elite athletes. The study of help seeking behaviors includes identification of obstacles, stigmas, attitudes towards counseling, and help seeking experience of athletes[4,17–19]. The research on athletes also involves such areas as teammate interaction, identity, social support, interruption of training, and the relationship between coach and athlete[2,3,5,20]. Altogether, these studies justify the need for caution and motivate the study, with the current dataset being used to test the classification system.

The main contribution of this paper is the development of an athlete focused prompt-response benchmark based on educational categories.

Methods

Dataset

Dataset Provenance: The data set was developed through YouTube videos that are accessible on the Internet from various channels on topics such as athletes' mental health, sport psychology, pressure of sports, stigma, coping strategies, attention difficulties, symptoms of depression, anxiety, and trauma related stress. To qualify for inclusion in the data set, a video should have contained either sports or athlete-specific information and language referring to anxiety, ADHD, depression, or PTSD. This experiment used prompt-answer combinations to be manually collected as educational examples based on the subjects of the source examples. Each data set record includes a prompt id, a prompt, an answer, an adjudicated label, the source video URL, and, if applicable, source channel metadata. The URLs to the sources and associated source channel metadata are used for the purposes of the dataset metadata and not as peer-reviewed scientific evidence. These types of classifications relate to research areas and not the diagnoses made from clinical analysis.

The examples analyzed from this dataset include the issues associated with anxiety, depression, emotion regulation, social media pressure, problems with attentional regulation, burnout, trauma-induced stress, stigmas, social support, and coping in the context of sports.

Final labeling includes such categories of anxiety, ADHD, depression, and PTSD. These labels are used for the classification of texts and are based on the prompt and answers text. The data set has the following class distribution: 7 for anxiety, 7 for ADHD, 13 for depression, and 6 for PTSD.

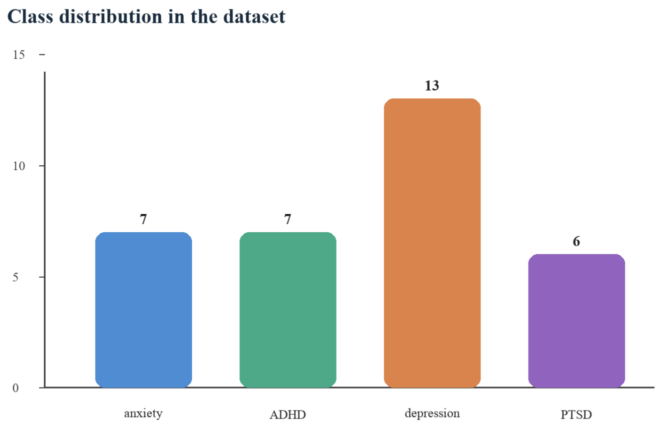

Figure. 1 Distribution of the dataset used in the cross-validation study.

The model inputs included the response texts of the question-response pairs. The labels were joined to the prompt_id by the adjudicated labels file. Each row had been assigned to a certain held-out fold according to the predefined cross-validation file. The fold seeds were set to 3407.

Annotation Procedure

The guidelines for annotation defined the presence of anxiety in terms of worry, fear, panic, nervousness, racing thoughts, or physiological symptoms of anxiety; ADHD as attentional difficulties such as distraction, disorderliness, impulsiveness, shifting, and burden on the executive functions; depression as ongoing depressed state, numbness, emptiness, demotivation, lack of hope, fatigue, seclusion, or exhaustion, in the case where depressed state was the most important; and PTSD as traumatic experience, intrusive memories, avoidance, hyper vigilance, traumatic stress, trigger, or reaction from trauma experience.

Each case was annotated by two annotators separately based on the annotation guidelines provided. Prior to adjudication, each case was labeled by both annotators with one major category. Agreement between the two before adjudication was 29 of 33 instances (87.9%), where Cohen's kappa was 0.828. Four disagreements were reconsidered according to the label

© The National High School Journal of Science 2026

Table 1 A comparison of Literature used to position the current study

Related Work Area	Dataset or Source	Model Family	Limitation for this Study
Twitter Depression/PTSD Benchmark	Twitter users marked as either suffering from depression or PTSD;	lexical and supervised classifiers	Present benchmark combines already existing NLP datasets on mental health topics defined in various fields and formats.
SMHD	large collection of Reddit texts with self-reported mental health diagnostics and controls;	Language features and supervised classification	Significantly larger dataset of social media data at user level with self-reported diagnosis labels; it should be noted that the current analysis makes use of education athlete examples
Dreaddit	Reddit texts about stress in various subreddits;	Traditional and neural supervised models	useful for analyzing stress language and uses a different labeling scheme and data domain.
MentalBERT	Mental health corpora for pre-training domain-specific language models	BERT/RoBERTa-style pretrained models	Shows advantages of domain-specific pretraining but it should be noted that this work is based on significantly larger corpora than the current 33 examples dataset.
Sentence-BERT and MiniLM	Benchmarks for general sentence representation.	fixed sentence embeddings with lightweight downstream models	While the strategy of efficient embeddings is well supported in this current work, validation of the athlete mental health labels needs to be done separately.
Mental Health Research among Elite Athletes	Surveys, qualitative research, and cohort studies of athletes;	non-NLP sports psychology literature.	This forms the motivation for the domain but does not address the NLP dataset development and classification performance for future work.

definition provided and the disagreements were adjudicated to provide the label file to be used in the model evaluation. This is an agreement between two researchers; The clinical reliability of these findings requires further expert review.

The map contained in Table 3 gives the annotated cues used to develop the research labels for the study. The map describes how some lines are repeated in different categories.

Preprocessing Audit

In the reported model comparison, the original answer text was retained without alterations in terms of negation, punctuation, intensifiers, word order, etc. This preprocessing step was required as some mental health language depends on the above aspects. Previously, stop word removal was considered as part of preprocessing experiments; however, words such as no, again, too was excluded from the evaluated runs since its impact would change the meaning in the mental-health related answers. The use of paraphrasing augmentation as a preprocessing step was also excluded from evaluated runs due to lack of matched leakage-safe ablation evidence.

Strongest Evaluated Model

In the strongest evaluated model, all-MiniLM-L6-v2 sentence-transformers library was used to encode the answer in a sentence embedding. Within each fold, the linear SVM classifier implemented in scikit-learn with the parameter of C set to 0.3 was trained and evaluated on holdout examples of the corresponding fold. Family of support-vector classifier algorithms is introduced by Cortes and Vapnik, with the concrete implementation utilized in the current analysis described by scikit-learn[21,22].

Use of this model is dictated by the size of the dataset. MiniLM is essentially a small transformer-based sentence encoder, which takes an input sentence and transforms it to a numeric vector representing the sentence semantics. On the top of that, LinearSVC is a simple classifier that uses those vectors as an input to learn the way examples are grouped into anxiety, ADHD, depression, and PTSD. Thus, there is a separation between language representation performed by MiniLM and classification performed by the SVM.

For the small dataset such an approach is computationally efficient as only the lightweight classifier must be trained using pretrained sentence encoder to generate fixed representations. Previous Sentence-BERT paper provides evidence supporting the use of sentence embeddings as a fixed efficient text representation method[12]. MiniLM was designed as a compact transformer library with lower computation cost[13]. Small datasets are susceptible to instability and overfitting risks thus the efficiency claim is only valid for this particular case.

LoRA, or Low-Rank Adaptation, is a parameter-efficient fine-tuning algorithm that adapts a pretrained language model in terms of small low-rank matrices[23]. It was chosen for this work since the size of the dataset is rather small and the full transformer fine-tuning would cost additional computational resources and could be prone to overfitting. LoRA is just one representative of a wide range of parameter-efficient adaptation techniques, including BitFit and other approaches that intend to adapt pretrained transformers with a small fraction of updated parameters[24]. Controlled LoRA comparison was performed on the same 33 examples, adjudicated labels, predefined cross-validation folds, fold seed 3407, and held-out prediction denominator as in the embedding/classifier run. In the result, it got 25 correct predictions out of 33, or 75.8% accuracy. The predictions on the row level, metrics, and confusion

Table 2 Examples that show the connection between label and appropriate text

ID	Label	Textual Cues	Annotation Reason
P002	Anxiety	tightening chest, restricted or shallow breathing, suffocation	key factors include physical arousal and breathing difficulty.
P004	ADHD	flooded thoughts, distractions, hard to focus, sidetracked, staying organized	attention problems and executive function are among the issues that are important to consider.
P008	Depression	weighs you down, drains your energy, emptiness, numbness, disconnected from joy	low mood, numbness, fatigue, and loss of joy dominate the answer
P015	PTSD	suppressed feelings, pushed through injury, emotional distress, avoidance of needs	This label is dependent upon the experience of trauma or stress and represents one of its limitations because of the few instances of trauma cues.

Table 3 Annotations feature map for the four research labels

Label	Label-based language cues	Athlete Context Cues	Common Overlap
Anxiety	worry, fear, panic, racing thoughts, tight chest, shallow breathing	pressure to perform, criticism by others, pre-game pressure, uncertainty.	can overlap with PTSD if language of fear or stress is used without trauma cues
ADHD	difficulty focusing, distractibility, forgetfulness, disorganized, impulsiveness, switching tasks	missed alarms, balance in class, flooded thoughts, time management	can overlap with depression if fatigue, low motivation, or concentration difficulties are present
Depression	sadness, hopelessness, loss of interest, fatigue, isolation, lack of self-worth, numbness	burnout, heaviness of emotions, withdrawal from sport, lack of motivation	can overlap with ADHD if there are problems with concentration and focus.
PTSD	trauma, triggers, flashbacks, nightmares, hyper vigilant, avoidant behaviors, intrusive memories.	Culture of injury, emotional suppression, toughing it out, trauma related stress	can overlap with anxiety or depression if trauma cues are not direct.

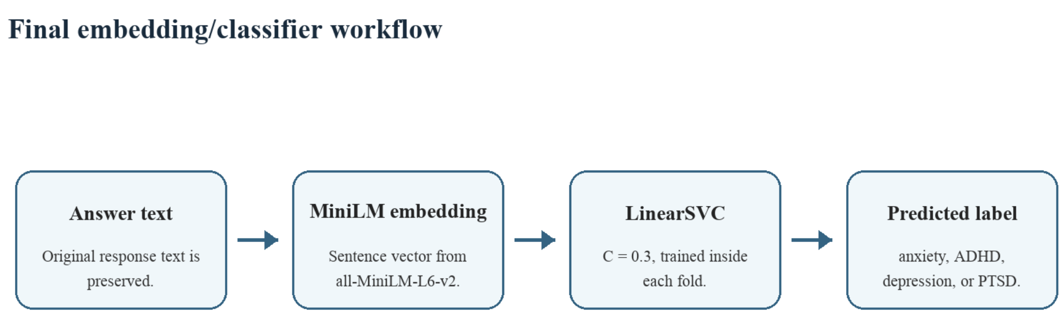

Figure. 2 Embedding/classifier workflow used for the 87.9% model run.

matrix can be found in the supplementary materials.

LoRA result: 25/33 correct predictions.

Results

With a total of 33 test predictions split into cross-validation folds, the model that used the embeddings generated by MiniLM alongside the LinearSVC classifier predicted accurately 29 predictions, which gives 87.9% accuracy. Considering that the dataset is quite small, this estimate comes with some uncertainty: the accuracy 95% Wilson confidence interval is equal to 72.7%-95.2%. Also, Macro-F1 is 0.869, and weighted-F1 is 0.876, which means that the performance was reasonably strong across all labels.

Table 4 Per-class metrics for the embedding/classifier run

Label	Precision	Recall	F1	Support
Anxiety	0.857	0.857	0.857	7
ADHD	1.000	0.857	0.923	7
Depression	0.812	1.000	0.897	13
PTSD	1.000	0.667	0.800	6
Accuracy			0.879	33
Macro Avg	0.917	0.845	0.869	33
Weighted Avg	0.896	0.879	0.876	33

There are different trends observed among the four labels in

© The National High School Journal of Science 2026

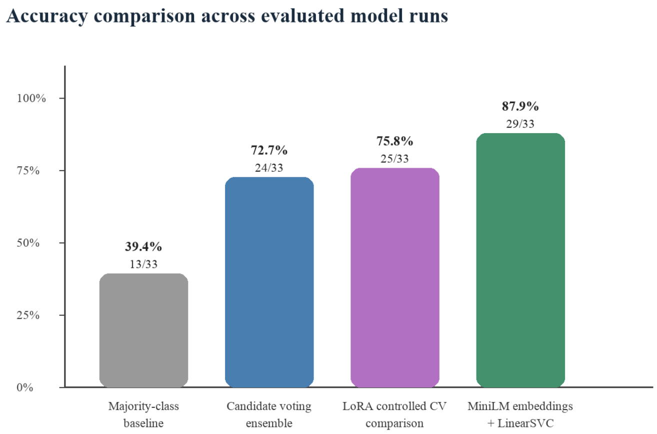

Figure. 3 Precision, recall, and F1 by class for the embedding/classifier run.

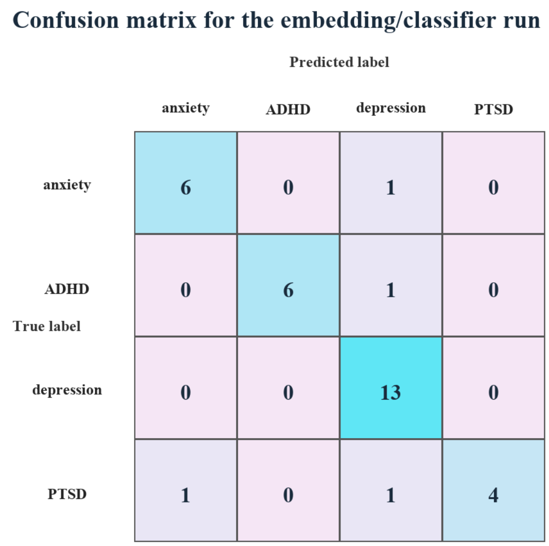

Figure. 4 Confusion matrix for the embedding/classifier run.

the class level measures. Recall was the highest for the case of depression, since all the thirteen depression samples were captured correctly by the system. On the other hand, its precision was not as high as its recall due to the inclusion of some responses belonging to other labels in the depression cluster. There is good classification for both anxiety and ADHD categories, where one example was missed in each of the two categories.

Table 5 Confusion matrix for the embedding/classifier run

True / Predicted	Anxiety	ADHD	Depression	PTSD
Anxiety	6	0	1	0
ADHD	0	6	1	0
Depression	0	0	13	0
PTSD	1	0	1	4

While the four misclassifications give more insightful information compared to accuracy, these mistakes are found among the instances where response towards athletes uses generic terms referring to problems with tiredness, inability to concentrate, stress control, inability to suppress emotions, worrying, and being in distress. For instance, the example labeled with ADHD has been misclassified as depression because it refers to being tired, heavy, and unable to concentrate. Another two examples labeled with PTSD have been wrongly classified as anxiety or depression, as they refer to the stress management process and the emotional control in the response without any trauma indication. Another instance labeled with anxiety has been classified as depression because it refers to worries about a child's health and depression-related feelings.

This suggests a lack of uniform boundary between the four labels. Depression is the easiest condition to identify because of the examples labeled with depression containing specific terms indicating feeling low, helpless, exhausted, isolation, and decreased activity. It can be suggested that PTSD is the most difficult disorder to distinguish because there are several examples of responses describing distress in indirect terms while not containing specific cues of trauma like triggers, flashbacks, avoidance of the place related to the traumatizing experience, insomnia, nightmares, or hypervigilance.

Comparing these results allows providing additional information about the primary finding. Majority class baseline reached 13 out of 33, as it classified all the instances with depression label. The voting ensemble classified 24 out of 33. Controlled LoRA comparison reached 25 out of 33 (75.8%) with the same set of examples, folds, seed, and evaluation denominator. The best approach tested in this study is MiniLM embedding/classifier reaching 29 out of 33.

The embedded scatter plot explains the error distribution graphically. From the scatter plot with true labels, several samples with different labels exist near one another in the vector space of MiniLM. According to the correctness overlay plot, all the four samples that were incorrect appear in areas with overlap between symptom languages of labels.

The audit of ablation limit clearly defines the interpretation of the comparison. The LoRA training used the same data, folds, seed, and denominator for evaluation. Methodological decisions such as the use of structured prompting, Alpaca format, stop-word filtering, paraphrasing, and changes to hyperparameters during development need to be analyzed using a fixed-split ablation analysis. Therefore, the results justify comparing different models, but the contribution of each development decision needs further study.

Overlap was highest between the categories of fatigue, concentration problems, stress management, emotional suppres-

Table 6 Analysis of errors in four wrong held-out predictions

ID	True -> Predicted	Reason for Overlap	Boundary Description
P003	ADHD -> Depression	Fatigue, low energy, feeling heavy, and poor cognitive functioning relate to the language of depression; while poor sustained attention relates to the language of ADHD.	Symptoms related to fatigue and attention can make distinction between depression and ADHD obscure.
P013	PTSD -> Anxiety	The suggested approach focuses on relaxation, stress management, controlled breathing, and preparations before performing an activity without mentioning any traumatic events.	It shows how language used for talking about stress management can cover the problem of PTSD by translating it into anxiety.
P015	PTSD -> Depression	Depression is linked to emotional suppression, subjective suffering, coping with difficulties, and poor well-being without any mention of trauma specific symptoms.	It emphasizes the importance of being more accurate when describing trauma experiences, triggers, avoidance, and hypervigilance.
P025	Anxiety -> Depression	Worries about children's psychological state and the emotional load associated with it may be similar to the signs of depression.	It demonstrates how worries and thoughts about depression can be combined in the context of the suggested solution.

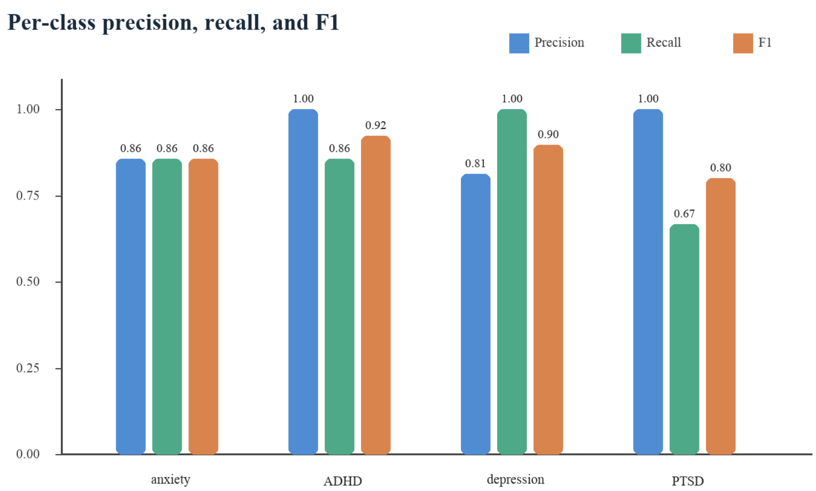

Figure. 5 Accuracy comparison between evaluated model runs

sion, worry, and distress in general. The most difficult category was PTSD, where many cases that were marked as PTSD did not have any trauma indicators such as a trigger, intrusive thoughts, avoidance or hypervigilance. The simplest category was depression; all 13 of the cases were categorized correctly.

Discussion

The main conclusion here is that a fixed-embedding solution fits better in this case in comparison with heavy adaptation. MiniLM generated generalizable sentence embeddings, while the classifier had to learn a comparatively small classification boundary based on the set of research labels. It is a local result limited to our dataset and explains why the embedding/classifier model can be regarded as the reasonable choice for our research.

Sports-related language is used quite often by athletes to describe distressing symptoms using terms like pressure, toughness, coping, fatigue, and emotional regulation. Moreover, these descriptions may cover several symptom categories, especially if there are no clear signals of trauma in the case of PTSD. That is why the development of the database should focus on those cases where the difference between anxiety, depression, ADHD, and PTSD becomes more obvious. Self-awareness motivation is still exploratory. Our research shows that the outputs of classifiers can be used to structure responses to analyze them but does not examine if self-awareness is improved by them.

Limitations

There are limitations associated with this study. The dataset consists of 33 examples and therefore limits generalization

 © The National High School Journal of Science 2026

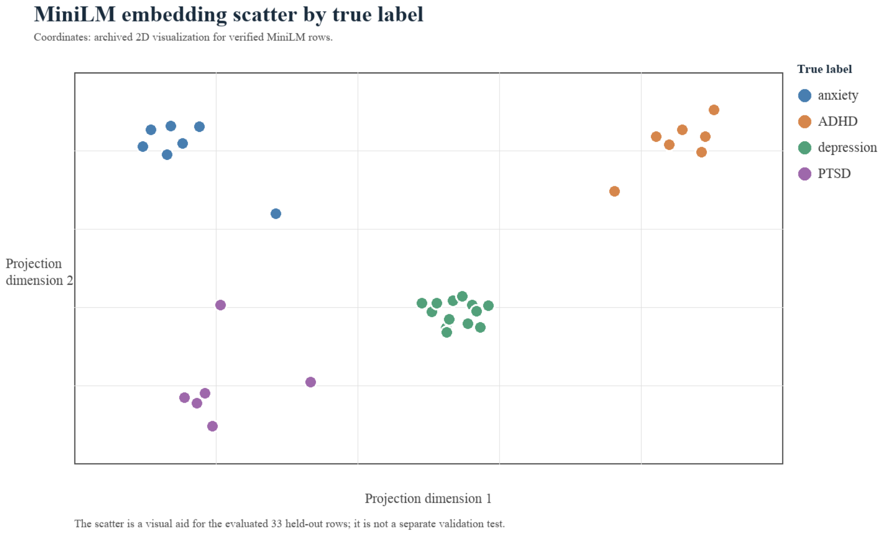

Figure. 6 MiniLM embedding-space scatter visualization colored by true label.

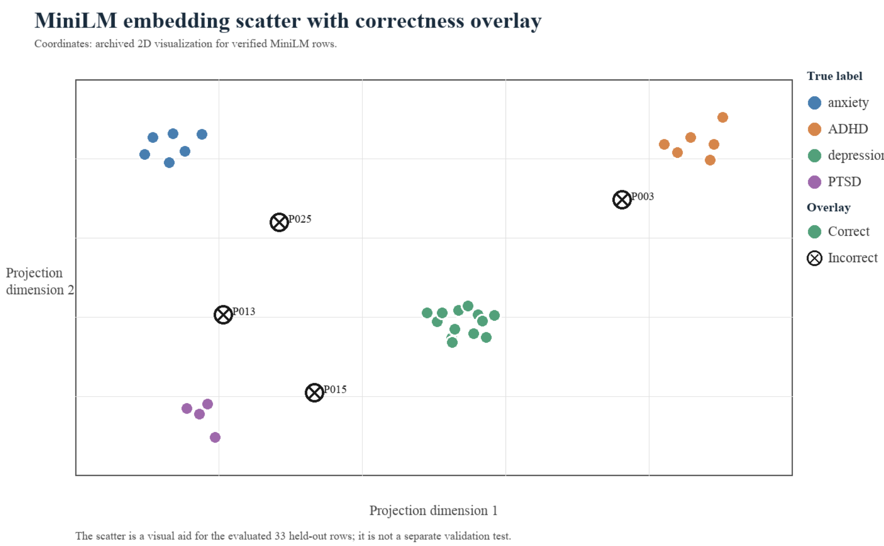

Figure. 7 correct and incorrect predictions overlaid on the MiniLM embedding-space scatter visualization.

and makes it too early to make strong conclusions about fine-tuning transformers. The dataset was created because there was no appropriate publicly available dataset for this task. Therefore, it should be viewed only as preliminary baseline for model building. Class distribution is imbalanced with 13 examples of depression and 6 examples of PTSD. Examples have been selected from very limited set of YouTube related to athletes and their mental health. Some of the provenance information is limited, access to the raw transcript is required.

Athletes, clinicians, mental health professionals and other independent experts have not been involved in the model output evaluation. This paper shows how the classifier works with respect to the curated dataset under the specified cross-validation setting. Real-life validation with athletes, clinicians, and teams is still needed. Therefore, the classifier should be viewed as a research model used for this specific task in particular setting. In the future, it is possible to build a larger dataset preserving provenance of transcripts and sources; use more independent annotators or domain experts; use more stable statistical methods of measuring agreement; use classical baselines (e.g., TF-IDF+Logistic Regression); and test the approach on independent athlete mental health re-

Table 7 Ablation-limit audit for factors considered during model development

Item	Status	Interpretation
Multifactor exploratory run	Context of development	Preprocessing, formatting, LoRA tuning, epochs, weight decay, and paraphrase tuning were performed simultaneously.
Structured prompting or Alpaca formatting	Needs ablation study	An ablation study using a fixed data split method is needed to determine the effect of formatting on other factors.
LoRA tuning	Same fold control test	25/33 correct predictions using the same data set, folds, seed, and evaluation denominator used for the embedding/classifier tests.
TF-IDF and logistic regression	Future baseline	Suggested that developing a baseline using the same fold method could be the future task.
Stopword removal	Excluded	There is no information that suggests that similar ablations or negation elimination can help interpret the results of mental health classification.
Paraphrastic augmentation	Excluded	The folds used in cross-validation consist of the original 33 examples.
Hyperparameter changes	Model-level result	The results are shown for each model; the evaluation of individual contributions requires fixed-split ablation studies.

sponses.

Ethical Considerations

As for ethical concerns, the paper describes the proposed model as an algorithm classifier for categories within organized research. The detection of illness, its treatment, replacement of professionals, and prioritizing intervention are beyond the scope of this research. Within this context, the text classification may be helpful to the researcher in organizing the response pattern in the controlled dataset. The demographic distribution feature is left out since the videos do not have any information on the athlete's age, ethnic group, gender, type of sport, geographic area, previous illnesses, and competition level. The labeling bias might be accounted for through the fact that labeling is done with the help of educational research categories based on the text. The domain bias might be introduced through the narrow context of examples (YouTube).

Conclusion

In general, this study serves as a reproducible beginning to the classification of responses related to mental health in athletes. In addition, further research may be conducted on how symptoms are linguistically associated in athlete contexts.

References

1 K. Schaal, M. Tafflet, H. Nassif, V. Thibault, C. Pichard, M. Alcotte, M. Guillet, E. El Helou, S. Berthelot, G. Simon, and J. F. Toussaint. Psychological balance in high level athletes: Gender-based differences and sport-specific patterns. PLOS ONE. Vol. 6, e19007, 2011, `https://doi.org/10.1371/journal.pone.0019007`.

2 G. Hagiwara, T. Tsunokawa, T. Iwatsuki, H. Shimozono, and T. Kawazura. Relationships among student-athletes' identity, mental health, and social support in Japanese student-athletes during the COVID-19 pandemic. International Journal of Environmental Research and Public Health. Vol. 18, article 7032, 2021, `https://doi.org/10.3390/ijerph18137032`.

3 Y. Karrer, S. Fröhlich, S. Iff, J. Spörri, J. Scherr, E. Seifritz, B. B. Quednow, and M. C. Claussen. Training load, sports performance, physical and mental health during the COVID-19 pandemic: A prospective cohort of Swiss elite athletes. PLOS ONE. Vol. 17, e0278203, 2022, `https://doi.org/10.1371/journal.pone.0278203`.

4 R. S. Wahto, J. K. Swift, and J. L. Whipple. The role of stigma and referral source in predicting college student-athletes' attitudes toward psychological help-seeking. Journal of Clinical Sport Psychology. Vol. 10, pg. 85-98, 2016, `https://doi.org/10.1123/jcsp.2015-0025`.

5 M. Powers, J. Fogaca, R. A. R. Gurung, and C. M. Jackman. Predicting student-athlete mental health: Coach-athlete relationship. Psi Chi Journal of Psychological Research. Vol. 25, pg. 172-180, 2020, `https://doi.org/10.24839/2325-7342.jn25.2.172`.

6 A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson. Machine learning algorithm validation with a limited sample size. PLOS ONE. Vol. 14, e0224365, 2019, `https://doi.org/10.1371/journal.pone.0224365`.

7 G. Varoquaux. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage. Vol. 180, pg. 68-77, 2018, `https://doi.org/10.1016/j.neuroimage.2017.06.061`.

8 G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology. Pg. 31-39, 2015, `https://doi.org/10.3115/v1/W15-1204`.

9 A. Cohan, B. Desmet, A. Yates, L. Soldaini, S. MacAvaney, and N. Goharian. SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. Proceedings of COLING. Pg. 1485-1497, 2018, `https://aclanthology.org/C18-1126/`.

10 E. Turcan and K. McKeown. Dreaddit: A Reddit dataset for stress analysis in social media. Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis. Pg. 97-107, 2019, `https://doi.org/10.18653/v1/D19-6213`.

11 S. Ji, T. Zhang, L. Ansari, J. Fu, P. Tiwari, and E. Cambria. MentalBERT: Publicly available pretrained language models for mental healthcare. Proceedings of the Thirteenth Language Resources and Evaluation Conference. Pg. 7184-7190, 2022, `https://aclanthology.org/2022.lrec-1.778/`.

12 N. Reimers and I. Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Pg. 3982-3992,

 © The National High School Journal of Science 2026

2019, https://doi.org/10.18653/v1/D19-1410.
13 W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou. MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Advances in Neural Information Processing Systems. Vol. 33, pg. 5776-5788, 2020, https://proceedings.neurips.cc/paper/2020/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.
14 J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preotiuc-Pietro, D. A. Asch, and H. A. Schwartz. Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences. Vol. 115, pg. 11203-11208, 2018, https://doi.org/10.1073/pnas.1802331115.
15 K. K. Fitzpatrick, A. Darcy, and M. Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent: A randomized controlled trial. JMIR Mental Health. Vol. 4, e19, 2017, https://doi.org/10.2196/mental.7785.
16 R. Fulmer, A. Joerin, B. Gentile, L. Lakerink, and M. Rauws. Using psychological artificial intelligence to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Mental Health. Vol. 5, e64, 2018, https://doi.org/10.2196/mental.9782.
17 A. Gulliver, K. M. Griffiths, and H. Christensen. Barriers and facilitators to mental health help-seeking for young elite athletes: A qualitative study. BMC Psychiatry. Vol. 12, article 157, 2012, https://doi.org/10.1186/1471-244X-12-157.
18 M. D. Bird, G. M. Chow, and B. T. Cooper. Student-athletes' mental health help-seeking experiences: A mixed methodological approach. Journal of College Student Psychotherapy. Vol. 34, pg. 59-77, 2020, https://doi.org/10.1080/87568225.2018.1523699.
19 R. C. Hilliard, L. A. Redmond, and J. C. Watson. The relationships among self-compassion, stigma, and attitudes toward counseling in student-athletes. Journal of Clinical Sport Psychology. Vol. 13, pg. 374-389, 2019, https://doi.org/10.1123/jcsp.2018-0027.
20 S. Graupensperger, A. J. Benson, J. R. Kilmer, and M. B. Evans. Social (un)distancing: Teammate interactions, athletic identity, and mental health of student-athletes during the COVID-19 pandemic. Journal of Adolescent Health. Vol. 67, pg. 662-670, 2020, https://doi.org/10.1016/j.jadohealth.2020.08.001.
21 C. Cortes and V. Vapnik. Support-vector networks. Machine Learning. Vol. 20, pg. 273-297, 1995, https://doi.org/10.1007/BF00994018.
22 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. Vol. 12, pg. 2825-2830, 2011, https://jmlr.org/papers/v12/pedregosa11a.html.
23 E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. LoRA: Low-rank adaptation of large language models. International Conference on Learning Representations, 2022, https://doi.org/10.48550/arXiv.2106.09685, https://openreview.net/forum?id=nZeVKeeFYf9.
24 E. Ben Zaken, Y. Goldberg, and S. Ravfogel. BitFit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Pg. 1-9, 2022, https://doi.org/10.18653/v1/2022.acl-short.1.

© The National High School Journal of Science 2026