Westborough High School Mental Health Predictive Analyses

Wesley Lo

Project Advisors

Professor Benjamin C. Nephew[1], PhD

Professor Dmitry Korkin[2,3], PhD

Co-advisor

Professor Angela C Incollingo Rodriguez[1,4], PhD

[1]Department of Neuroscience, Worcester Polytechnic Institute

[2]Department of Mathematical Sciences, Worcester Polytechnic Institute

[3]Department of Bioinformatics and Computational Biology, Worcester Polytechnic Institute

[4]Department of Psychological Science, Worcester Polytechnic Institute

Abstract

With the rise in mental health challenges over the years, especially in student populations, the need to develop targeted diagnostic and treatment tools has increased. Given the substantial research to highlight the tremendous variation in mental health symptomatology, efforts to develop intervention strategies may benefit from insights into the individual. In this study, we investigate the predictive capabilities of random forest classifiers (RFC) in prediction of mental health risk scores in mental health data from the Westborough High School. Our results show variation in mental health outcomes across different student populations and underscore the need for personalized interventions over a one-size-fits-all approach.

Acknowledgements

I would like to express my heartfelt gratitude to Professor Nephew and Professor Korkin for advising me on this project. I am deeply grateful for the weekly meetings and support, especially amidst my own personal challenges. I am also thankful to graduate students Senbao Lu and Justin Polcari for providing valuable guidance throughout the project, and to Professor Rodriguez for her insightful feedback and input. Additionally, I would also like to extend my appreciation to the Westborough High School administration for providing the data used in this study.

Table of contents

Introduction

Background

As mental health challenges become more prevalent among adolescents around the globe, the need for prognostic tools has grown. The *Youth Behavior Risk Survey*, a recent CDC report, highlights these trends over the past decade, from 2010 to 2021. In this report, 6 out of 10 female students reported feeling "persistently sad or hopeless" over the past year. In the general population, a recent meta-analysis discovered a 34% increase in self-reported symptoms of depression (Larry et al, 2020). There were also variations specific to gender and race. The percent of female students with encounters of sexual assault increased for the first time in 10 years, and significantly so by 14%. Compared to male peers, female students twice as often felt persistent feelings of sadness or hopelessness (6 in 10) and suicidality (1 in 10). A recent meta-analysis looked at >350,000 college students across 373 campuses in years 2013-2021 and found students of color to have the lowest rates of mental health service utilization; Asian, Black, and Latinx students had the same or below past-year treatment from White students (Lipson et. al. 2022). Another recent meta-analysis conducted by Eylem et al. (2020) found that mental illness stigma is higher among ethnic minorities compared to majorities.

These trends support the need for diagnosis and intervention of mental health, especially in adolescents. This has been exacerbated in light of major current events such as the COVID-19 global pandemic and its impact on mental health (Hawes et al., 2021; Mariah et al., 2022). The systematic study by Keles et al. (2020) suggests that social media has been a significant contributor, too. To further emphasize the urgency of addressing mental health challenges among students, a local example at Worcester Polytechnic Institute (WPI) is worth noting. By February 2022, 7 students had committed suicide over the preceding 8-month span. This local instance highlights the broader need for mental health interventions in educational institutions (Moody, 2022).

As noted in a recent news article, the 2021 MetroWest Adolescent Health Survey revealed mixed results for Westborough students. Fortunately, drug and alcohol use has decreased since the last survey in 2018, with a substantial decrease in vaping amongst 7-12th graders. The committee attributed this to efforts to spread awareness for the harm of vaping. Conversely, increases were noted in depression, cyberbullying and stress from the 2018 to 2021, particularly higher amongst LGBTQ+, Latina/Latino, and female student groups. The Westborough officials underscored the relationship and support focused nature of their K-12 program, along with analysis of surrounding data (Sullivan, 2022).

This study aims to investigate the relationship between mental health and these factors in the Westborough High School (WHS) freshman population. In particular, data of WHS responses to the GAD-7 and PHQ-9 questionaires were used. These questionnaires have been well received in clinical literature as efficient yet effective diagnostic tools for depression and anxiety respectively (Manea et. al., 2015). (Robert, 2002; Simon & Dean, 2015; Löwe B et al. 2008).

Our primary goal with this study was to predict mental health outcomes using student demographic, academic, and detailed mental health screening data. The data's mental health metrics included PHQ-9 and GAD-7 overall risk scores and question scores. These question responses gave details into the specific mental health symptoms including, anhedonia, hopelessness, and fatigue. We used question data and student attributes to make predictions on total scores using the Random Forest Classifier (RFC), which let us determine which questions were most responsible for predicting mental health outcomes. Our feature set was limited, so our analysis focused on comparing significant predictors within varying gender and race / ethnicity groups. However, this analysis is by no means fully comprehensive, and numerous factors can contribute to symptomatic variations.

Overview of Student and Mental Health Features: GAD-7 and PHQ-9 Scores

The dataset contained information for first year students (class of 2025) at Westborough High School in Massachusetts. The following features were included: academic performance (semester 1 and year end GPA weighted or unweighted), educational statuses (special education (SPED), English-language learner (ELL), and 504 plan), and race / ethnicity and gender. Since the freshman class had not received grades for their core academic subjects for semester 1, analysis was restricted to year end GPA. There were 274 students in the filtered dataset, with breakdowns of student characteristics summarized in Table 1. Additional breakdowns of mental health outcomes in different groups are provided in **Tables 16** and **17** of **Appendix A**.

Table 1 - General Statistics per Subgroup (mean ± standard error)

***Note**: SPED, 504, and ELL are scored 0 or 1. GPA unweighted and weighted are along 4.0 and 5.0 scales respectively*

		SPED status	504 Plan Status	ELL status	GPA Weighted	GPA Unweighted
Race	**A (105)**	0.04 ± 0.02	0.02 ± 0.01	0.03 ± 0.02	3.89 ± 0.04	3.64 ± 0.03
	W (169)	0.18 ± 0.03	0.12 ± 0.03	0.05 ± 0.02	3.31 ± 0.06	3.23 ± 0.05
Gender	**F (145)**	0.08 ± 0.02	0.07 ± 0.02	0.04 ± 0.02	3.63 ± 0.05	3.45 ± 0.04
	M (129)	0.18 ± 0.03	0.10 ± 0.03	0.04 ± 0.02	3.43 ± 0.07	3.31 ± 0.05
Both	**AF(56)**	0.04 ± 0.03	0.02 ± 0.02	0.04 ± 0.03	3.96 ± 0.06	3.68 ± 0.05
	AM (49)	0.04 ± 0.03	0.02 ± 0.02	0.02 ± 0.02	3.81 ± 0.07	3.58 ± 0.05
	WF (89)	0.11 ± 0.03	0.10 ± 0.03	0.04 ± 0.02	3.42 ± 0.07	3.31 ± 0.06
	WM (80)	0.26 ± 0.05	0.15 ± 0.04	0.05 ± 0.02	3.20 ± 0.09	3.14 ± 0.07

Lastly, the dataset had mental health metrics, which were student responses to the GAD-7 and PHQ-9 questionnaires and their summary scores. The 7-question GAD-7 and 9-question PHQ-9 are widely used self-reporting questionnaires for screening GAD and depression, respectively, and are modeled after the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV). The questions listed below in **Table 2**. They are provided with a phrase representing the question's intent, which will be referenced later:

Table 2 - Labeled PHQ-9 and GAD-7 Questions (Spitzer et al. 2006)

<u>**PHQ-9**</u> 1. (**anhedonia**) Little interest or pleasure in doing things 2. (**hopeless**) Feeling down, depressed, or hopeless 3. (**sleep difficulties**) Trouble falling or staying asleep, or sleeping too much 4. (**fatigue**) Feeling tired or having little energy 5. (**appetite**) Poor appetite or overeating 6. (**failure**) Feeling bad about yourself or that you are a failure or have let yourself or your family down 7. (**poor concentration**) Trouble concentrating on things, such as reading the newspaper or watching television 8. **(slow speaking / restless**) Moving or speaking so slowly that other people could have noticed. Or the opposite being so fidgety or restless that you have been moving around a lot more than usual 9. (**suicidal ideation**) Thoughts that you would be better off dead, or of hurting yourself
<u>**GAD-7**</u> 1. (**Nervous/anxiousness**) Feeling nervous, anxious, or on edge 2. (**Worrying**) Not being able to stop or control worrying 3. (**Worrying**) Worrying too much about different things 4. (**Stress**) Trouble relaxing 5. (**Restless**) Being so restless that it is hard to sit still 6. (**Irritability**) Becoming easily annoyed or irritable 7. (**Fearfulness**) Feeling afraid, as if something awful might happen

Methods

Data Availability

The deidentified dataset was provided by the Westborough High School administration and is not publicly accessible. Students, and parents on behalf, were given the option to opt out of being included in the dataset and study, and only 9 out of 329 students did.

Data Analysis

All data analysis and visualizations were provided from Python and its libraries. Pandas and numpy were used to clean the dataset, and also provide correlation matrices. Sklearn was used for binary classification models (random forest, logistic regression). Scipy was used to run anova. Seaborn and Matplotlib were used to make heatmaps and plots. Analysis of variance (ANOVA) was performed using Scipy.

Preprocessing

We filtered the dataset by removing students in groups with fewer than 10 individuals in both race / ethnicity (being Black, Hispanic, and other) and gender (being non-conforming). This left Asian and White for the former, and Male and Female for the latter. We also removed rows with missing values, particularly in GPA. This left 274 students in the year-end dataset.

ANOVA

Two-way ANOVA was performed to discern both main effects and specific interaction effects between Race / Ethnicity and Gender across individual GAD-7 and PHQ-9 questions and full summary scores.

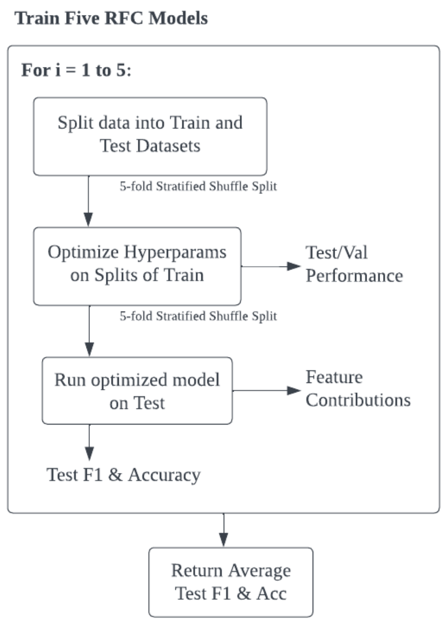
Random Forest Classification & Feature Contributions

For the prediction of mental health status in the WHS mental health dataset, we settled on using the random forest classifier (RFC), an ensemble learning method that leverages decision trees to make binary classifications. One benefit of using RFC over other algorithms is the interpretability through feature contributions. Random forest as a decision-tree-based model can capture causal relationships between features in its prediction process, through feature contributions. These results sum to 1 and represent how well RFC's decision trees divide the data into positive and negative classes via Gini importance. In prediction of each target, feature contributions for each feature were averaged across all five models. The top two features of this aggregate were included in the results section.

Predictive Analysis Pipeline

For predictive analyses, we stratified by gender, race / ethnicity, and both, yielding 8 groups. In each group, 6 total analyses were performed; predictions on the three targets (GAD-7 predictions, PHQ-9 predictions, and endorse Q9 predictions) for question level and overall risk level features.

Figure 1 - Pipeline for training Random Forest Classifiers (RFC) in given group



We sought to avoid two problems in the formation of test and train sets: (1) Keep the proportion of mental health targets the same in both test/train sets (2) re-randomize the set to capture the full population. Problem 1 was addressed using stratification, which encourages the test/train/split shuffling process towards proportions of outcome variables in splits representative of the entire dataset. Problem 2 was addressed by generating five of these models or "splits" then averaging the results, notated by the outer shell in figure 5. Five-fold cross-validation was performed on each split. This process is outlined above in **Figure 1**. This whole procedure was performed once on the whole population, and also once per each of the 8 groups stratified by race / ethnicity, gender, or both.

Scoring PHQ-9 and GAD-7

Responses to each question are scored on a scale of 0 to 3, with higher scores indicating greater severity of symptoms (Spitzer et al. 2006). For PHQ-9, an individual's summary score is obtained by summing all question responses, thus ranging from 0-27. Specific cutoffs indicate increasing severity: minimal (0–4), mild (5–9), moderate (10–14), moderately-severe (15–19), and severe (20–27). The GAD-7 questions are similarly scored 0-3, with the summary score being the sum of all question responses ranging up to 21 with the following cutoffs: minimal (0–4), mild (5–9), moderate (10–14), and severe (15–21). All of these questions are considered “endorsed” for a score above 1, except for PHQ-9 Q9, which is above 0 and used to screen for suicidality. This is referred to as endorse Q9 later in the paper (Kroenke et al., 2001).

Binarization

For classification, we binarized mental health total scores. GAD-7 summary scores of above 10 indicate potential for clinical significance. PHQ-9 has similar recommendations within the range of 5 to 15 as summarized in **Tables 3** and **4**. However, to keep the distributions of PHQ-9 and GAD-7 similar, we binarized according to 10 as well. An approximate 80% cutoff was observed within these two distributions. A cutoff of 1> was used for endorse Q9.

Features Used

The features used in the overall risk level analysis included SPED, ELL, and 504 plan statuses, along with GPA and comorbid overall risks. For instance, GAD-7 overall risk score and endorse Q9 overall risk score were used for predicting binary outcomes for PHQ-9 overall risk. Only the targets were binarized. For instance, in prediction of PHQ-9 risk, in both the overall risk and question level analyses, only PHQ-9 risk would be binarized. The question level analysis included specific responses to both PHQ-9 and GAD-7 in addition to the statuses and GPA. For the plots of results, only the question features are shown.

Table 3 - PHQ-9 Symptoms Scoring (Kroenke et al., 2001)

Score	Symptom Severity	Comments
0-4	Minimal or none	Monitor; treatment may not be needed
5-9	Mild	Consider symptom duration, functional impairment to determine whether treatment is necessary
10-14	Moderate	
15-19	Moderately severe	Active treatment with psychotherapy, medication, or a combination of both is warranted.
20-27	Severe	

Table 4 - GAD-7 Symptoms Scoring (Spitzer et al. 2006)

Score	Symptom Severity	Comments
0-5	None	Monitor; treatment may not be needed
5-9	Minimal	
10-14	Moderate	Possibly clinically significant
>15	Severe	Active treatment is possibly warranted

Figure 2 - Distributions of PHQ-9 and GAD-7 Summary Scores

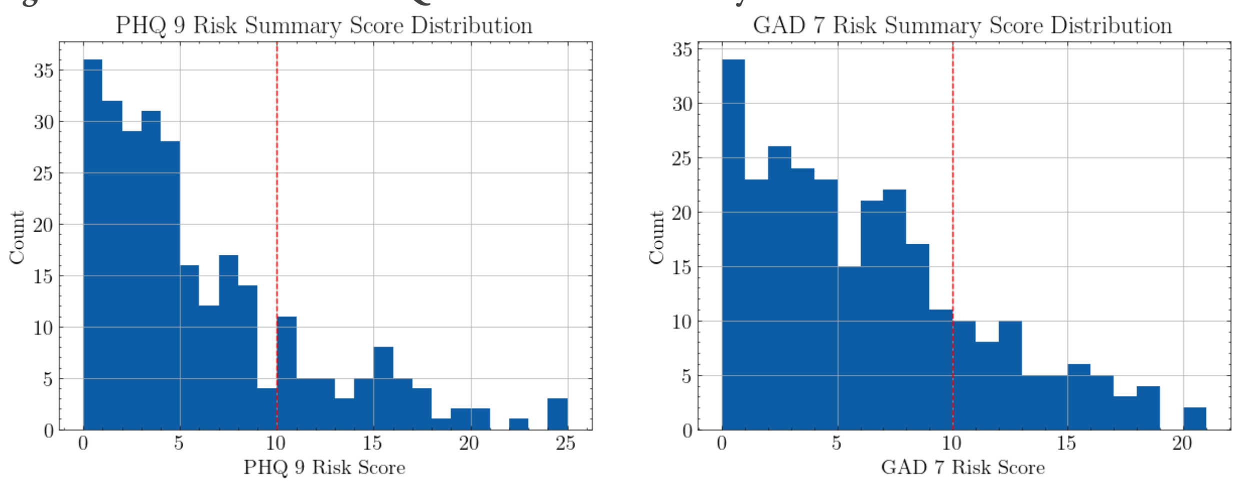


Table 5 - GAD-7 Overall Risk and Question Responses in Different Student Populations (mean ± standard error)

		GAD-7 Overall Risk	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Race	A (105)	5.10 ± 4.09	0.97 ± 0.77	0.74 ± 0.84	0.98 ± 0.89	0.58 ± 0.72	0.41 ± 0.70	0.87 ± 0.91	0.54 ± 0.76
	W (169)	6.34 ± 5.20	1.21 ± 0.95	0.83 ± 0.91	1.17 ± 1.00	0.78 ± 0.90	0.79 ± 0.96	0.95 ± 0.99	0.60 ± 0.91
Gender	F (145)	7.20 ± 5.08	1.39 ± 0.88	1.06 ± 0.93	1.30 ± 0.97	0.88 ± 0.89	0.76 ± 0.94	1.12 ± 1.06	0.69 ± 0.92
	M (129)	4.36 ± 4.06	0.81 ± 0.79	0.51 ± 0.73	0.86 ± 0.89	0.50 ± 0.73	0.52 ± 0.81	0.70 ± 0.78	0.46 ± 0.76
Both	AF (56)	6.30 ± 4.21	1.21 ± 0.76	1.00 ± 0.91	1.21 ± 0.91	0.68 ± 0.74	0.41 ± 0.63	1.14 ± 1.00	0.64 ± 0.77
	AM (49)	3.71 ± 3.51	0.69 ± 0.68	0.45 ± 0.65	0.71 ± 0.79	0.47 ± 0.68	0.41 ± 0.79	0.55 ± 0.68	0.43 ± 0.74
	WF (89)	7.76 ± 5.51	1.51 ± 0.94	1.09 ± 0.95	1.36 ± 1.01	1.01 ± 0.95	0.98 ± 1.03	1.10 ± 1.10	0.72 ± 1.01
	WM (80)	4.76 ± 4.34	0.89 ± 0.84	0.55 ± 0.78	0.95 ± 0.94	0.53 ± 0.76	0.59 ± 0.82	0.79 ± 0.82	0.47 ± 0.78

Table 6 - PHQ-9 Overall Risk and Question Responses in Different Student Populations (mean ± standard error)

		PHQ-9 Overall Risk	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Race	**A (105)**	5.10 ± 4.73	0.48 ± 0.77	0.48 ± 0.72	0.72 ± 0.95	0.98 ± 0.91	0.60 ± 0.88	0.85 ± 0.99	0.53 ± 0.71	0.30 ± 0.61	0.16 ± 0.48
	W (169)	5.76 ± 5.69	0.54 ± 0.82	0.53 ± 0.79	0.82 ± 1.03	1.12 ± 0.99	0.59 ± 0.90	0.59 ± 0.90	0.98 ± 0.94	0.41 ± 0.73	0.17 ± 0.51
Gender	**F (145)**	6.75 ± 5.91	0.62 ± 0.88	0.68 ± 0.85	0.94 ± 1.05	1.31 ± 0.96	0.79 ± 0.99	0.84 ± 1.01	0.95 ± 0.92	0.42 ± 0.72	0.20 ± 0.57
	M (129)	4.11 ± 4.24	0.40 ± 0.69	0.33 ± 0.61	0.60 ± 0.91	0.79 ± 0.88	0.38 ± 0.71	0.52 ± 0.83	0.65 ± 0.81	0.31 ± 0.63	0.12 ± 0.40
Both	**AF (56)**	6.05 ± 5.20	0.54 ± 0.81	0.68 ± 0.86	0.84 ± 0.99	1.16 ± 0.89	0.75 ± 0.96	0.98 ± 1.07	0.64 ± 0.72	0.27 ± 0.49	0.20 ± 0.55
	AM (49)	4.02 ± 3.92	0.41 ± 0.73	0.24 ± 0.43	0.59 ± 0.89	0.78 ± 0.90	0.43 ± 0.76	0.69 ± 0.87	0.41 ± 0.67	0.35 ± 0.72	0.12 ± 0.39
	WF (89)	7.19 ± 6.30	0.67 ± 0.93	0.67 ± 0.85	1.01 ± 1.08	1.40 ± 1.00	0.81 ± 1.02	0.75 ± 0.97	1.15 ± 0.98	0.52 ± 0.83	0.20 ± 0.59
	WM (80)	4.16 ± 4.45	0.40 ± 0.67	0.38 ± 0.70	0.61 ± 0.93	0.80 ± 0.88	0.35 ± 0.68	0.41 ± 0.79	0.80 ± 0.85	0.29 ± 0.58	0.12 ± 0.40

Results

ANOVA Findings

Two-way ANOVA was run between Gender and Race and outcome risk scores, with main (ME) and interaction effects (IE) provided.

Table 7 - Two-way ANOVA performed on Raw GAD-7 Risk Scores against Gender and Race

Source	MS	F	p
Gender (ME)	552.029367	26.028336	***<.001.**
Race / Ethnicity (ME)	103.963042	4.901886	***0.0267**
Gender * Race / Ethnicity (IE)	2.741197	0.129248	0.7

GAD-7 was significantly different across race / ethnicity (Asian and White) and even more so across gender (male and female), as indicated by the low p values and greater F statistic for the former; anxiety was higher in White students than in Asian students, and higher in females than in males. GAD-7 was not significantly different among both variables simultaneously, as noted from the insignificant interaction effect.

Table 8 - Two-way ANOVA performed on Raw PHQ-9 Risk Scores against Gender and Race

Source	MS	F	p
Gender (ME)	478.462696	17.744063	***<.001.**
Race / Ethnicity (ME)	29.104512	1.079357	0.299771
Gender * Race / Ethnicity (IE)	15.978991	0.592590	0.442092

Similarly, PHQ-9 was quite significantly different across gender, as indicated by the higher F statistic and P value; here, depression was more common in females than in males. However, this was not true for race / ethnicity alone or combined with gender.

Table 9 - Two-way ANOVA performed on Raw Endorse Q9 Scores against Gender and Race

Source	MS	F	p
Gender (ME)	0.394251	1.583712	0.209314
Race / Ethnicity (ME)	0.001189	0.004777	0.944947
Gender * Race / Ethnicity (IE)	0.000172	0.000692	0.979036

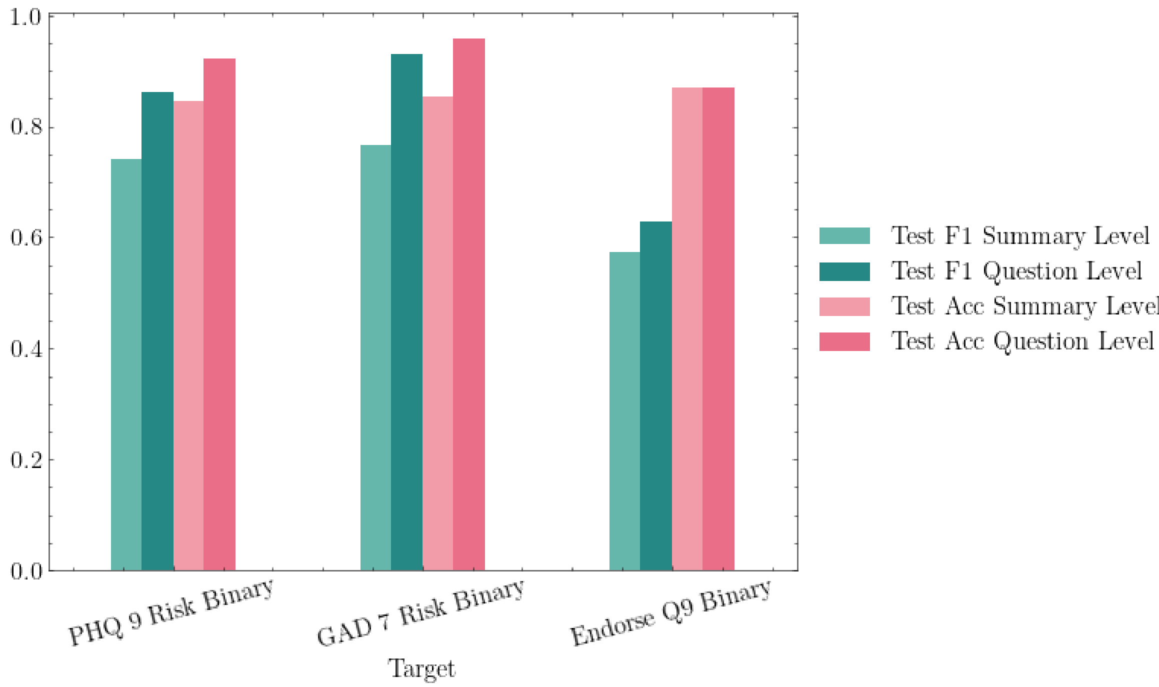
Endorse Q9 was not significantly different across gender, race / ethnicity, or combined.

Overall, the results of **Tables 7-9** show significant differences in GAD-7 scores across gender and race/ethnicity, with gender having a more substantial impact (greater imbalance towards females over males than White over Asian students). However, these differences were not observed when considering both variables simultaneously. PHQ-9 scores were also significantly different across gender but not race/ethnicity or their interaction. Endorse Q9 scores did not show any significant differences across gender, race/ethnicity, or their interaction.

Random Forest: Model performances

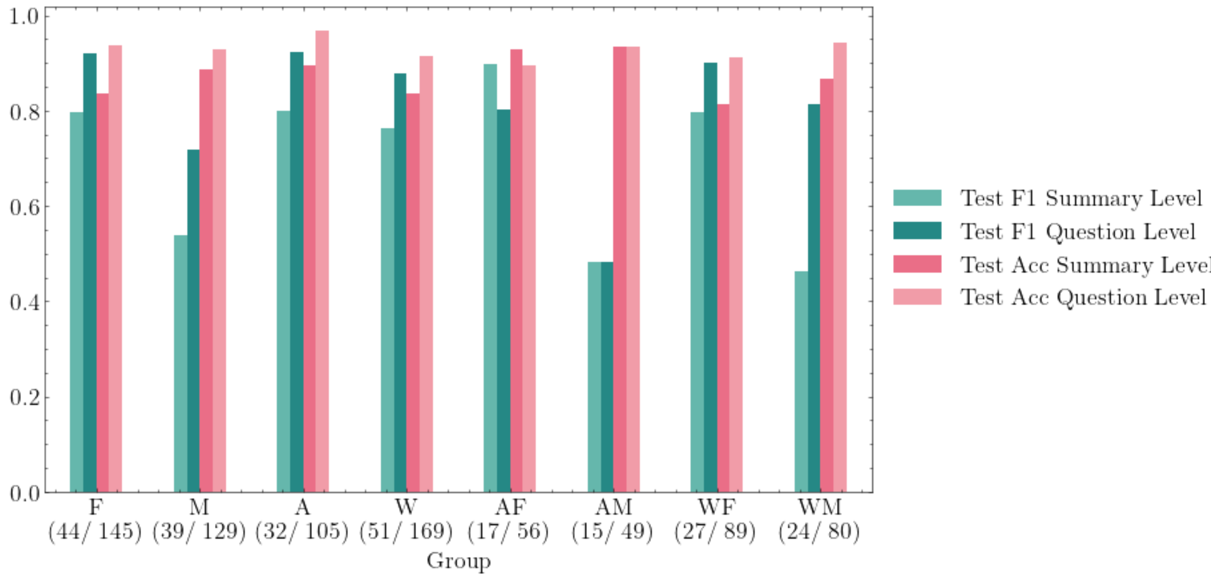
Performances for the random forest classifiers are provided in **Figures 3-6**. A more comprehensive summary of model performance data is provided in **Appendix B**. Overall risk level predictions refer to using comorbid risks to predict outcome variables; for instance, risk scores for PHQ-9 and endorse Q9 were used (along with additional features as listed) to predict a binary outcome for GAD-7, and vice versa.

Figure 3 - Performances of RFC predictions using Summary/Question Level Features Across All Students



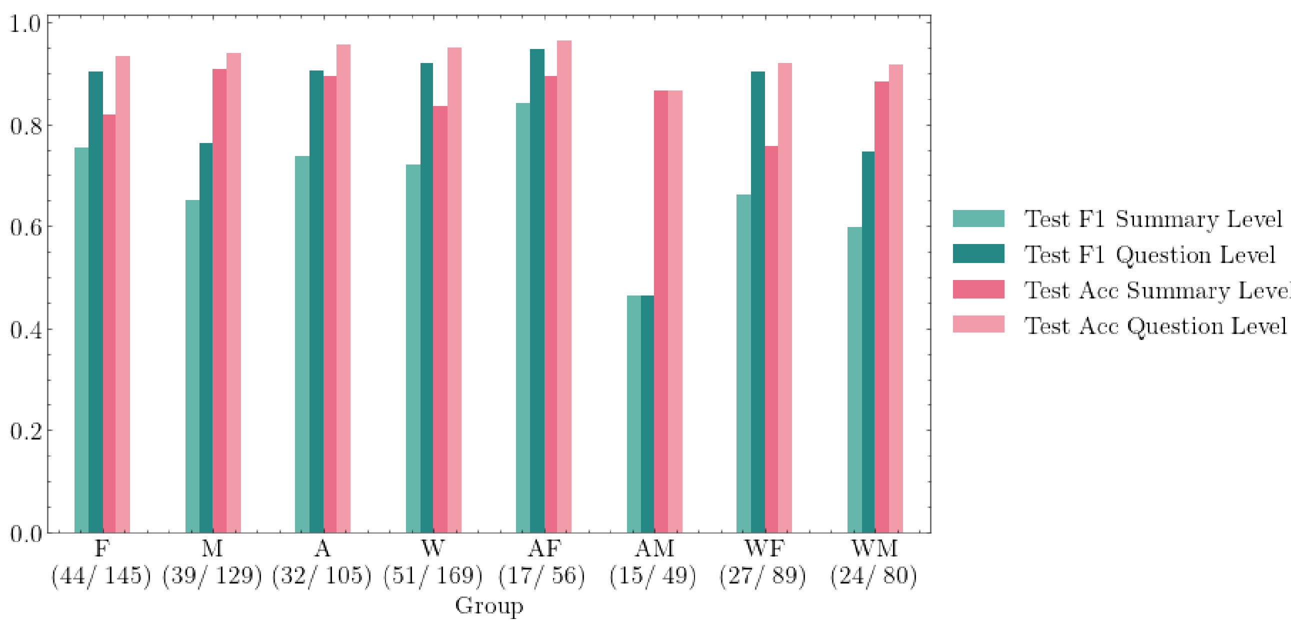
In the whole student population, accuracies for all three predictions (PHQ-9, GAD-7, and endorse Q9) were high in all conditions (>0.8). F1 scores were lower in predicting endorse Q9 (about 0.6). Using overall risk level data, F1 scores for PHQ-9 and endorse Q9 were greater than 0.7, and using question level data, greater than 0.8 for question level data.

Figure 4 - Performance of RFC predictions of GAD-7 in Groups (Test N / Train N)



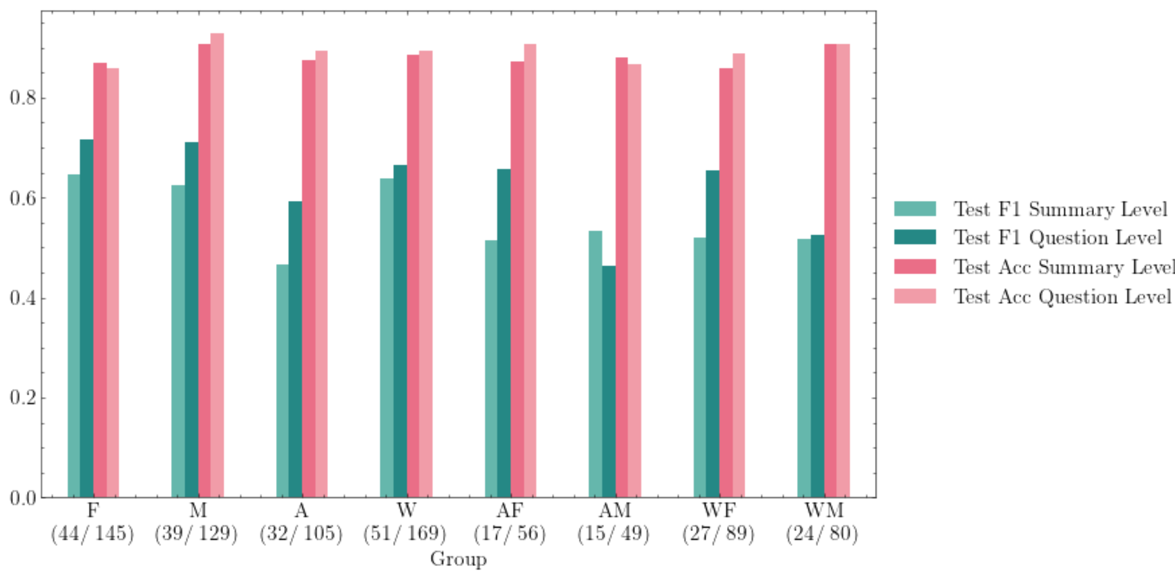
Accuracies in all conditions were high (>0.8). F1 scores for overall risk level and question level data were also in the same range, except for the male group and male subgroups, which had F1 scores of around 0.5.

Figure 5 - Performance of RFC predictions of PHQ-9 in Groups (Test N / Train N)



Accuracies in all conditions were high (>0.8). F1 scores using overall risk level data were in the 0.7 - 0.8 range, except lower for the male group, male subgroups, and female white population. However, F1 scores for using question level data were all above 0.8, except in the male group and male subgroups, which were again in the 0.4-0.6 range.

Figure 6 - Performance of RFC predictions of Endorse Q9 in Groups



Accuracies in all conditions were high (>0.8). However, F1 scores in all conditions were lower (<0.7).

Overall, the models overall performed accurately, between 85% to 95% across the board. F1 scores were high as well between 0.8 to 0.95, except in male and in prediction of endorse Q9. These can be attributed to the low numbers of samples that qualified for endorse Q9 risk, as seen in **Table 18** of **Appendix A**.

RFC: Overall Risk Level Predictions

Predictions were performed using comorbid overall risk scores in addition to student statuses and year end GPA, as in **Figures 7-9**. In **Tables 10-12**, the top two average feature contributions across 5 runs are provided.

Figure 7 - Feature Contributions in GAD-7 Score Prediction with Overall PHQ-9 and GAD-7 Risk Scores and Student Statuses in All Students

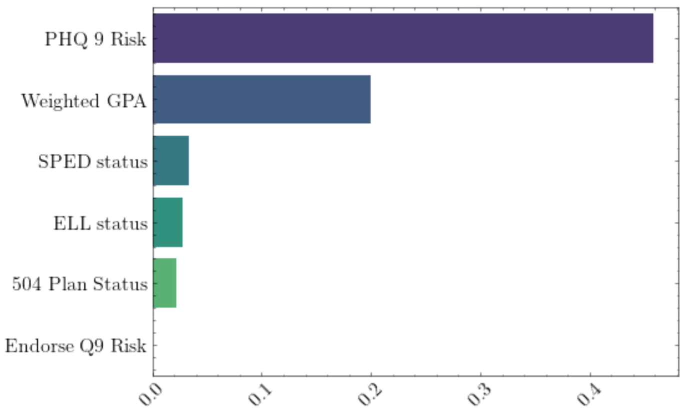


Table 10 - Top 2 Feature Contributions in GAD-7 Prediction with Summary Scores in Subgroups

Top 2 feature contributions in varying stratification groups

		PHQ-9 Risk	Endorse Q9 Risk	Year-end GPA (Weighted)	SPED Status	504 Plan Status	ELL Status
Gender	**F**	1		2			
	M	1		2			
Race	**A**	1		2			
	W	1		2			
Both	**AF**	2		1			
	AM	1		2			
	WF	2		1			
	WM	1		2			

For the full student population, high PHQ-9 risk and low GPA were the top two predictors of GAD-7 risk. In most groups, the top predictor was high PHQ-9, however, low GPA was highest in the female subgroups (Asian and White female).

Figure 8 - Feature Contributions in PHQ-9 Score Prediction with Overall Risk Scores and Student Status in All Students

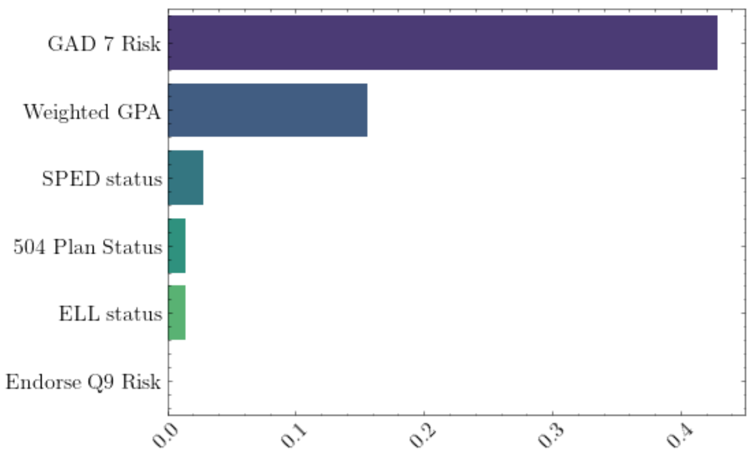


Table 11 - Top 2 Feature Contributions in PHQ-9 Prediction with Summary Scores in Subgroups

Top 2 feature contributions in varying groups stratified by gender and race / ethnicity

		GAD-7 Risk	Endorse Q9 Risk	Year-end GPA (Weighted)	SPED Status	504 Plan Status	ELL Status
Gender	**F**	1		2			
	M	1		2			
Race	**A**	1		2			
	W	1		2			
Both	**AF**	1		2			
	AM	1		2			
	WF			1		2	
	WM	2		1			

For the full student population, high GAD-7 and low GPA were the top two predictors of GAD-7 risk. This was mostly true in all subgroups, except for White females that had GPA and 504 plan status as the top 2 predictors and white males, where low GPA was the top ranked predictor of PHQ-9 risk.

Figure 9 - Feature Contributions in Endorse Q9 Score Prediction with Overall PHQ-9 and GAD-7 Risk Scores and Student Status in All Students

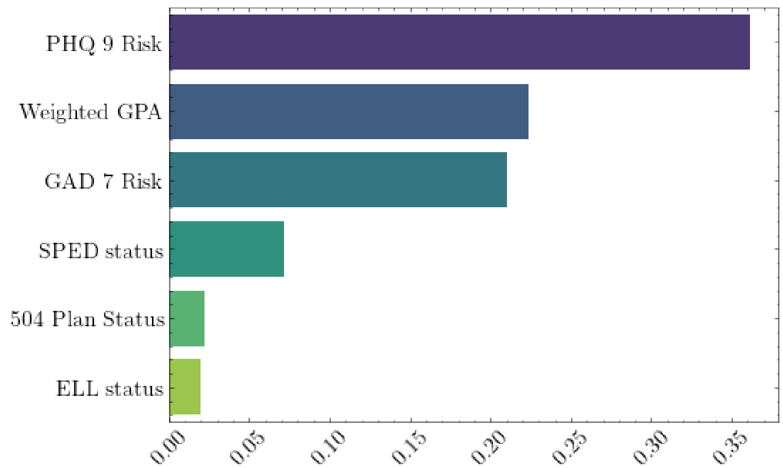


Table 12 - Top 2 Feature Contribution in Endorse Q9 Prediction with Summary Scores in Subgroups

Top 2 feature contributions in varying groups stratified by gender and race / ethnicity

		GAD-7 Risk	PHQ-9 Risk	Year-end GPA (Weighted)	SPED Status	504 Plan Status	ELL Status
Gender	**F**	2		1			
	M		2	1			
Race	**A**		1	2			
	W	1	2				
Both	**AF**		1	2			
	AM		1	2			
	WF		1	2			
	WM		2	1			

For the full student population, high PHQ-9 risk was the greatest predictor for endorse Q9. Top two feature contributions tend to be low GPA and PHQ-9, with both the White and Female groups having high GAD-7 as a stronger predictor. GPA was the top ranked feature, outranking high PHQ-9 risk in the female, male, and White male subgroups.

Overall, for overall risk level predictions in the total student population, one of the comorbid overall risk scores was the highest predictor; GAD-7 for PHQ-9 prediction, PHQ-9 for GAD-7 prediction and PHQ-9 for endorse Q9 prediction. GPA was ranked 2nd in all three cases as well. For stratified predictions, most subgroups followed suit, with a few exceptions; For GAD-7, both Female subgroups had GPAs higher than PHQ-9. For PHQ-9, the White female population had 504 plan status as the 2nd top feature. For endorse Q9, the White population had GAD-7 as a greater predictor than PHQ-9. This suggests anxiety to be a higher predictor for suicidality in the white population over depressive symptoms.

RFC: Question Level Predictions

For the question level analysis, predictions of GAD-7 and PHQ-9 and Endorse Q9 student outcomes were made using student responses to the GAD-7 and PHQ-9 questions as features; both question types were provided, to check for any predictive crossover. The top features are shown for the predictions on the overall population in **Figures 10-12**. Below each figure, **Tables 13-15** respectively show the top two features when predicting on each subpopulation.

Figure 10 - Average RFC Contributions of Question Level Features in GAD-7 Prediction in All Students

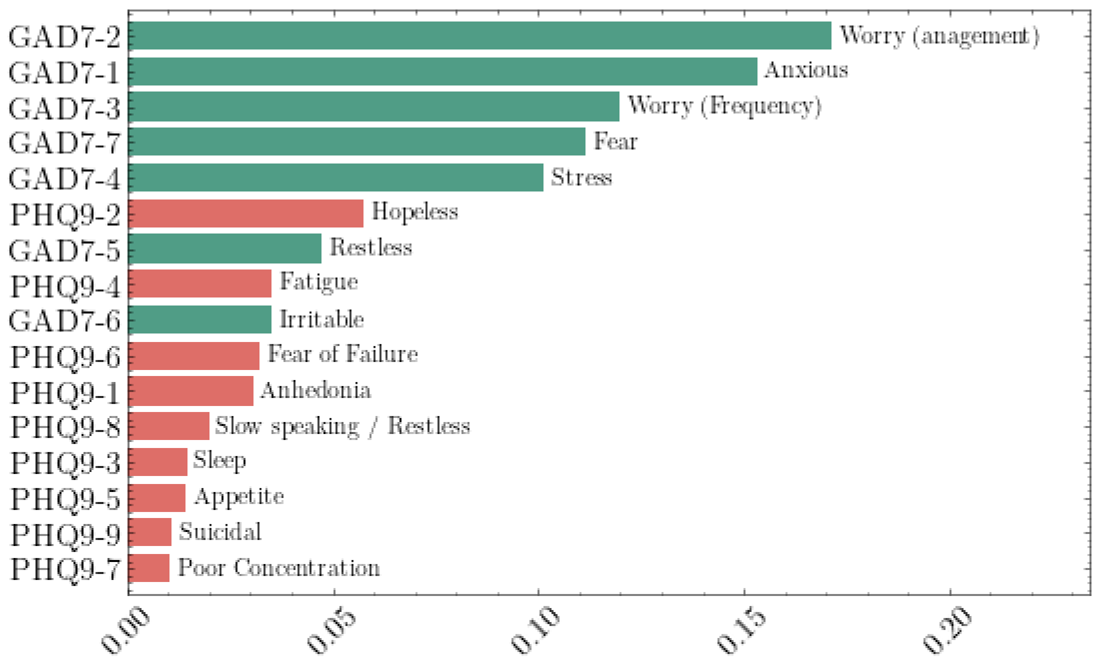


Table 13 - Top 2 RFC Contributions of Question Level Features in GAD-7 Prediction in Groups

		Questionaire Top Features															
		GAD Questions							PHQ Questions								
		1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	9
Gender	F			1	2												
	M				2			1									
Race	A		1		2												
	W	2	1														
Both	AF		2					1									
	AM	1						2									
	WF									2				1			
	WM		1					2									

For each group, the top two feature contributions in GAD-7 prediction are as follows: female (1. worry, 2. stress), male (1. fear, 2. stress), Asian (1. worry, 2. stress), White (1. worry, 2. anxiety), Asian female (1. worry, 2. fear), Asian male (1. anxiety, 2. fear), White female (1. hopeless, 2. failure), and White male (1. worry, 2. fear).

For all students, questions associated with worrying, anxiousness, ranked the highest in prediction of GAD-7. However, across individual student groups, this varied significantly. By gender, this was worrying for female students, and fearfulness in males. Interestingly, for White female students, PHQ-9 questions ranked as the highest predictors of anxiety (hopelessness and failure). Overall, most groups had GAD-7 questions 2, 4 and 7 in their top 2 contributions.

Figure 11 - RFC Contributions of Question Level Features in PHQ-9 Prediction in All Students

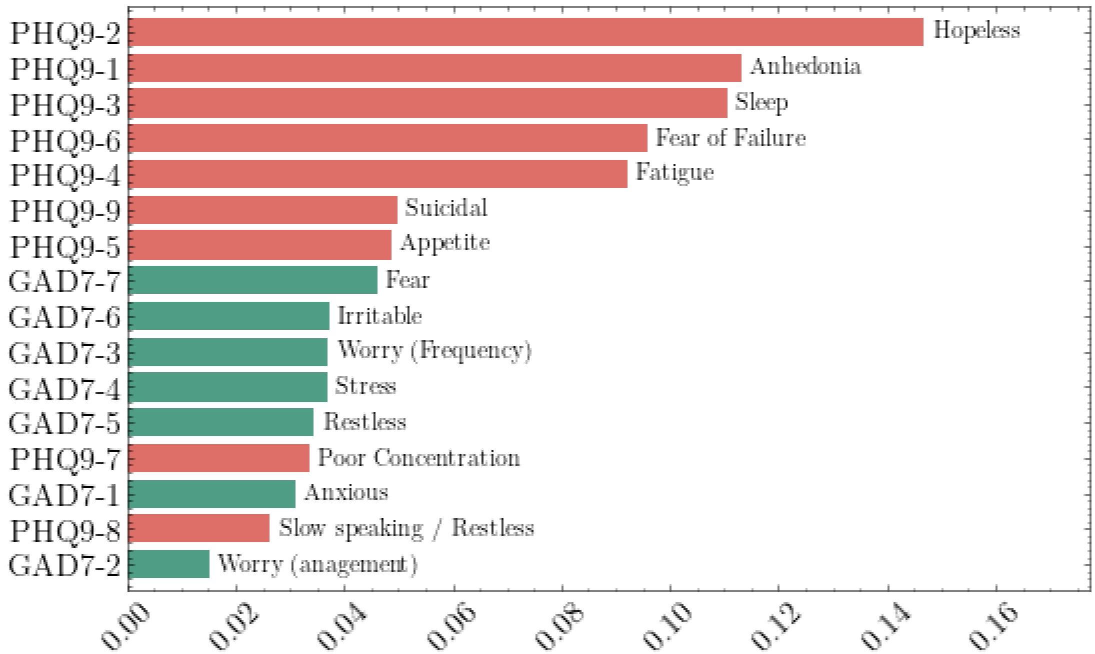


Table 14 - Top 2 RFC Contributions of Question Level Features in PHQ-9 Prediction in Subgroups

		Questionaire Top Features															
		GAD Questions							PHQ Questions								
		1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	9
Gender	**F**								2	1							
	M										2	1					
Race	**A**									1				2			
	W									1	2						
Both	**AF**										1					2	
	AM								2		1						
	WF									2				1			
	WM										2	1					

For each group, the top two feature contributions in PHQ-9 prediction are as follows: female (1. hopeless, 2. anhedonia), male (1. fatigue, 2. sleep), asian (1. hopeless, 2. failure), white (1. hopeless, 2. sleep), Asian female (1. sleep, 2. restless), Asian male (1. sleep, 2. anhedonia), White female (1. failure, 2. hopeless), White male (1. fatigue, 2. sleep).

Overall, PHQ-9 questions were generally linked to PHQ-9 risk in all students. As with before, the top 2 ranked features varied substantially between groups. There was no overlap by gender alone; for females, this was hopelessness & anhedonia, and for males, fatigue and sleep difficulties. This suggests distinct differences in their experiences of depression.

Figure 12 - RFC Contributions of Question Level Features in Endorse Q9 Prediction in All Students

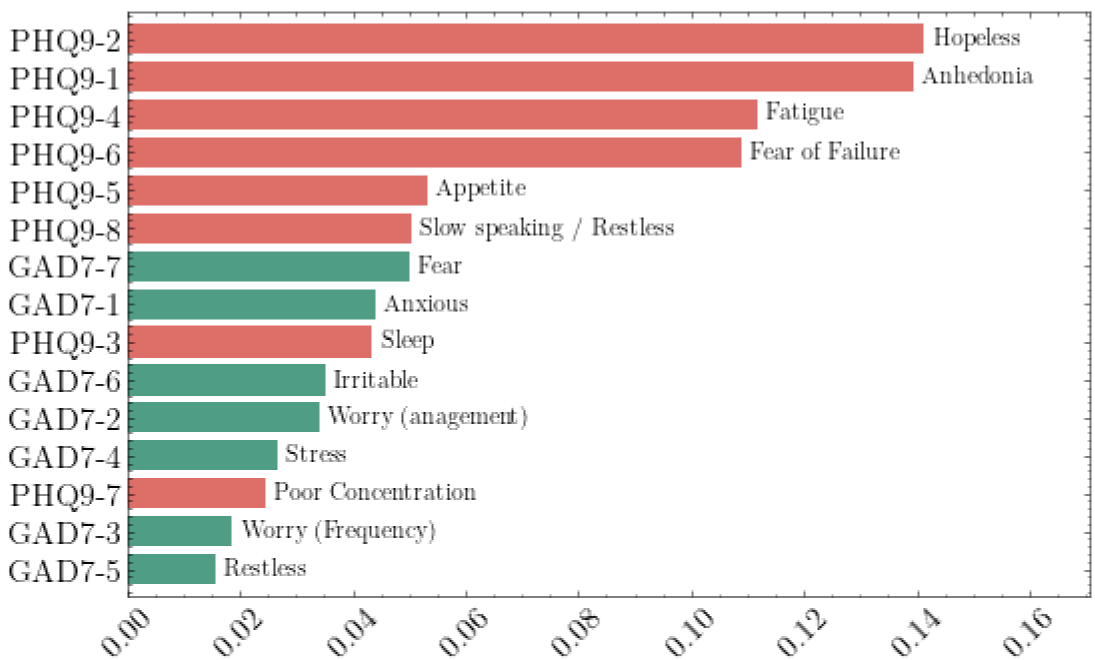


Table 15 - Top 2 RFC Contributions of Question Level Features in Endorse Q9 Prediction in Groups

		Questionaire Top Features														
		GAD Questions							PHQ Questions							
		1	2	3	4	5	6	7	1	2	3	4	5	6	7	8
Gender	F									1				2		
	M								1			2				
Race	A											2		1		
	W								1	2						
Both	AF									1				2		
	AM											1		2		
	WF								2	1						
	WM				2				1							

For each group, the top two feature contributions in endorse Q9 prediction are as follows: female (1. hopeless, 2. failure), male (1. anhedonia, 2. fatigue), Asian (1. failure, 2. fatigue), White (1. anhedonia, 2. hopeless), Asian female (1. hopeless, 2. failure), White female (1. hopeless, 2. anhedonia), Asian male (1. fatigue, 2. failure), White male (1. anhedonia, 2. stress).

For the entire population, Hopelessness and anhedonia ranked the highest, followed by fatigue and failure. These top two features were reflected along several groups, however, variation was still observed. For instance, in females, hopelessness and failure were highest, whereas anhedonia and fatigue were highest in males. In Asian students, failure and fatigue were observed, whereas anhedonia and hopelessness were observed in White students. Within males, Asian males had fatigue and failure, while White males had anhedonia and trouble relaxing (Q4 from GAD-7). Interestingly, the GAD7 question 4 question was in the top 2 features for endorse Q9 prediction. Overall, questions 1, 2, 4, and 6, tended to be the most commonly top 2 features in endorse Q9 prediction.

Discussion

Mental health is known to express itself in a wide array of symptoms, as dependent on a host of factors. The presence of a variety and question types in the PHQ-9 and GAD-7 questionnaires signifies this complexity. As noted earlier, there is a wealth of literature on these differences. Our primary aim with this study was to predict mental health outcomes using student demographic, academic, and detailed mental health screening data. Since our feature set was most effectively divided by gender and race, we geared our analysis towards observing differences in mental health outcomes and predictors between different student subgroups. Through our analysis, we found that mental health symptoms can manifest differently in different student groups, highlighting the importance of considering intersectionality in mental health research and treatment.

Overall Risk Level Findings

Overall, even with a modest dataset, clinical outcomes could still be predicted with high accuracy throughout the analysis (**Figures 3 - 6**). Although a general metric, differences were observed in the overall risk score analysis. The high performances of the models in predicting GAD-7 with PHQ-9 and vise versa, supports the comorbidity of these disorders in addressing the student population. In thinking about suicidality, our results of prediction endorse Q9 have shown GAD-7 to rank higher than PHQ-9 for White females, suggesting the role of considering anxiety in the development of targeted interventions for suicidal ideation, particularly for this population (**Tables 12**). This aligns with previous findings that have found a greater risk for women to develop anxiety disorders (Pigott 1999; Källström et al., 2022; McLean et al., 2011). Accordingly, the work of Aronen et al. (2005) has suggested the need to holistically consider anxious and depressive traits as they relate to cognition and academic performance. Likewise, a study by Owens et al. (2012) found lower academic performance associated with depression and anxiety levels in adolescents.

Our analysis revealed GPA as a strong predictor of anxiety in female students (**Tables 10**). GPA may have a particularly negative impact on female students and their anxiety levels, and was also strongly linked to depression in White males and females. The link between anxiety and academic performance has been explored throughout the literature (Aronen et al., 2005; Christensen, 1979; Keogh et al. 2004). A paper from Maloney et al. (2014) proposes that both trait-anxiety and state-anxiety both lead to academic performance deficits, mediated through working memory (WM) inhibition. They found that high math anxious subjects were 20% more likely to make mistakes in conditions with high WM requirements, over low math anxious subjects.

While resources are generally segmented along mental health and academics, a more integrated approach in the form of enhanced coordination and communication could provide benefit to both dimensions. For instance, interventions on the adverse effects of anxiety and GPA may have mutual benefits for both academics and mental health. The efforts of WHS to have counseling accessible towards students, support these goals. Academic performance is known to be a significant source of stress in highschool students (Pascoe et al., 2019; Prabu, 2015; Barbayannis et al. 2022). This impact has also been established for state-anxiety as well, suggesting state-anxiety Fernández-Castillo et al. (2015) looked at university students and found selective attention and concentration to be high with the absence of academic stressors. To address this, promoting awareness for these challenges can be a source of destigmatizing these issues and improve anxiety and depression symptomatology.

Question level findings

In comparison, the question level data provided a more nuanced view of the variety of anxiety and depression manifestation. For instance, analysis on all students (**Figure 10**) didn't predict fearfulness as a relevant predictor of anxiety in males (**Table 13**). This issue with missing key predictors in the male population is systematic due to higher PHQ and GAD scores in the female population, as observed in **Tables 16 and 17** in **Appendix A**. We also observed a masking effect; in PHQ prediction, females matched the overall rankings (**Figure 11**) of the student body in hopelessness and anhedonia (**Table 14**), while for males, fatigue and sleep were most significant. Furthermore, even within the female population, Asian females had sleep difficulties and slow speech/restlessness as a greater predictor of PHQ-9 scores (**Table 13**). Thus, recommendations made on the whole population may align with certain subgroups (such as females), while *masking* the needs/concerns of other subgroups (such as males and Asian females). This phenomenon raises concerns, given the differences in suicidality trends in males and females (Rick et al., 1998; CDC, 2023).

Variety in mental health symptoms is also highlighted by depression questions being unexpectedly found as strong predictors of anxiety in the White female population (**Table 13**). This further implies that depression-specific interventions could be an essential consideration for this group, and that comprehensively addressing the anxiety/depression comorbidity is crucial. These findings indicate the importance of evaluating different metrics when assessing mental health among diverse student populations, allowing for more personalized and effective support and interventions.

In the case of males, who may be more reluctant to disclose mental health challenges (McKenzie et al., 2022; Chatmon, 2020; Clark et al. 2020), our findings indicate that fatigue and sleep difficulties are crucial factors to consider when assessing their mental well-being. Research from Matos et al. (2015), demonstrates that factors such as skipping class, school achievement, disliking school, pressure with school work, contribute to these fatigue. Furthermore, these factors may be causal and relatively more actionable within the home setting, such as implementing earlier bedtimes, reducing screen time, and removing screens from bedrooms in the evening. An increased understanding of adolescents' sleep needs has already prompted many school districts across the US to adopt later start times (Kelman, 1999; Malone, 2011). Ming et. al (2011) found sleeping less than 7 hours on weekdays or weekends was associated with poorer performance in school, and several other studies have found significant detrimental effects of poor sleep quality in decreased academic performance (Shin et al., 2003; Kang et al. 2012; Ahberg et al., 2012).

In females, the consistently robust role of hopelessness and anhedonia might be associated with social media consumption. The social media habits between genders have been explored in the literature. As cited earlier in the paper, Mazman & Usluel et al., (2011) discovered differences in social media behavior on the basis of gender. Social media was also found to be higher in perceived usefulness and perceived risk in females than in males (Idemudia et al., 2017). In Haferkamp et al. (2012), Females were found more likely to to compare themselves with other females, while males used social media to connect. A large study of 221,096 adolescents, age 13 to 18-year-old in the U.S. and UK, found adolescent girls to spend more time on social media, texting, and smartphones (Twenge & Martin, 2020). They also found increased technology use to be twice as likely to be associated with reduced well-being, increased mental health issues, and suicidal risk. These concerns bring to mind the crucial consideration of gender differences in social media use and its potential impact on mental health, particularly among adolescent girls. Strategies to promote healthy and balanced use of technology should be considered. While anhedonia is a general and nonspecific symptom that may not be directly influenced by the inherent rewarding aspects of a student's activities, it is possible that reevaluating and modifying the selection, scheduling, and management of academic and non-academic pursuits could lead to a reduction in anhedonia and overall depression levels. By addressing the potential impact of social media and other lifestyle factors on these symptoms, interventions can be tailored to support the mental well-being of female students.

Though these concerns are applicable to all groups, it may be more effective to identify intervention strategies that specifically address the most pressing issues faced by different student populations in order to optimize treatment outcomes. Additionally, evaluation of student mental health globally may overshadow critical group specific needs, decreasing efficiency and effectiveness in identifying and treating these at-risk groups. For students with unique symptoms that are not being addressed, this could create an adverse attitude/response towards efforts to support mental health. A study by Steketee et al. (2005) found effects of perceived criticism during behavioral treatment of anxiety to exacerbate depressive/anxious symptoms. Such differences support the need for personalized recommendations, data collection, and care for the individual student. Enhanced communication and coordination among members of specific social spheres, through academics and counseling. Beyond the scope of this project, non-academic factors such as disciplinary actions, social interaction (withdrawal, aggression) and impulsivity should be considered.

It is worth noting that there was no specific objective to identify race and gender-specific patterns in the dataset. Despite the limited descriptive statistics on the students, our analyses provided several unbiased patterns in gender and race. Analogous to the benefits of teachers getting to know students in an academic setting, detailed data collection can improve mental health programs and interventions. As physical examinations are required for scholastic sports participation and standardized exams for graduation, standardized mental health data collection or formal mental health screening is warranted, given the concerning increase in mental illness prevalence among children and adolescents.

Conclusion

In conclusion, our study has emphasized the importance of considering intersectionality in mental health research and treatment among student populations. By utilizing machine learning tools, we were able to identify unique patterns in mental health symptoms and predictors across different student groups. We found that anxiety and depression can manifest differently in different student groups. This underscores the need for personalized recommendations, data collection, and care for individual students. Enhanced communication and coordination among members of specific social spheres, through academics and counseling, can help address these unique needs.

Limitations and Next Steps

Despite these findings, there were several limitations in our study. Firstly, the amount of individuals who qualified for some features was small, which restricted the depth of our analysis and models. In many groups, SPED status, ELL status, and 504 plan status had fewer than 10 samples, which likely constrained the predictive impact of these variables. This was similarly true for the number of students who qualified for high GAD-7 and PHQ-9 in the student subgroups. Additionally, other student populations had less than 10 individuals along gender (gender non-conforming) and race (Black, Latinx, and Hispanic individuals), so these populations were left out. As cited in the introduction, these populations are minority groups, and thus it is important for future studies to consider more diverse samples to better understand the experiences of these underrepresented populations. It should be noted that the overall F1 scores in both the summary level and question level predictions, stratified and unstratified, were relatively lower, despite the high accuracies. These low sample sizes could explain these differences. As shown in the **Appendix A**, some subgroups did not have high counts of students with significant risk, which was apparent in the low F1 score (0.5) in the male Asian group. Despite this, F1 and accuracy scores were high across the board, suggesting the robustness of RFC and their capabilities to predict mental health outcomes in student groups in a targeted way.

The feature set provided was restricted to a small set of variables that likely did not capture the comprehensive complexity of mental health experiences. Additionally, mental health measures were not taken continuously, so we were unable to track how mental health changed over time. Due to the way the data was conducted, we were making predictions on past mental health based on future GPA data, which may not reflect desired predictions. As such, it should be noted that mental health is complex and multifaceted and additional features such as social interactions, extracurricular activities, or family dynamics may also be considered for more nuanced predictions in future research. Likewise, a targeted survey of underrepresented individuals or school districts with greater representation of such students. Furthermore, exploring these topics on the basis of the individual student, could provide personalized support and interventions to address a student's individual needs and challenges.

References

Ahrberg, K., Dresler, M., Niedermaier, S., Steiger, A., & Genzel, L. (2012). The interaction between sleep quality and academic performance. Journal of psychiatric research, 46(12), 1618-1622. doi: 10.1016/j.jpsychires.2012.09.008. Epub 2012 Oct 3. PMID: 23040161.

Alfonso, S. V., & Lonigan, C. J. (2021). Trait anxiety and adolescent's academic achievement: The role of executive function. Florida State University, Department of Psychology, 1107 W. Call Street, Tallahassee, FL 32306-4301, USA. https://doi.org/10.1016/j.lindif.2020.101941

Aronen, E. T., Vuontela, V., Steenari, M. -R., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. Neurobiology of Learning and Memory, 83(1), 33–42. https://doi.org/10.1016/j.nlm.2004.06.010

Barbayannis G, Bandari M, Zheng X, Baquerizo H, Pecor KW, Ming X. (2022) Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19. Front Psychol. May 23;13:886344. doi: 10.3389/fpsyg.2022.886344. PMID: 35677139; PMCID: PMC9169886.

Centers for Disease Control and Prevention. (2023). Suicide data & statistics. Retrieved April 26, 2023, from https://www.cdc.gov/suicide/suicide-data-statistics.html

Chatmon, B. N. (2020). Males and mental health stigma. Psychology of Men & Masculinities, 21(4), 619-620. https://doi.org/10.1177/1557988320949322

Christensen, H. (1979). Test anxiety and academic achievement in high school students. Perceptual and Motor Skills, 49(2), 648. https://doi.org/10.2466/pms.1979.49.2.648

Clark, L.H., Hudson, J.L. & Haider, T. Anxiety Specific Mental Health Stigma and Help-Seeking in Adolescent Males. J Child Fam Stud 29, 1970–1981 (2020). https://doi.org/10.1007/s10826-019-01686-0

DeFreitas, S. C., Crone, T., DeLeon, M., & Ajayi, A. (2018). Perceived and personal mental health stigma in Latino and African American college students. Frontiers in public health, 6, 49. https://doi.org/10.3389/fpubh.2018.00049

Fernández-Castillo, A., & Caurcel, M. J. (2015). State test-anxiety, selective attention and concentration in university students. International Journal of Psychology, 50(4), 265-271. https://doi.org/10.1002/ijop.12092

Haferkamp, N., Eimler, S.C., Papadakis, A.-M., & Kruck, J.V. (2012). Men Are from Mars, Women Are
from Ve-nus? Examining Gender Differences in Self-Presentation on Social Networking Sites. Cyberpsychology, Behavior, and Social Networking, 15(2), 91-98.

Hawes, M., Szenczy, A., Klein, D., Hajcak, G., & Nelson, B. (2022). Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychological Medicine, 52(14), 3222-3230. doi:10.1017/S0033291720005358

Idemudia, E. C., Raisinghani, M. S., Adeola, O., & Achebo, N. (2017). The effects of gender on the adoption of social media: An empirical investigation. Proceedings of the 23rd Americas Conference on Information Systems (AMCIS 2017), Boston, MA, USA. Retrieved from https://aisel.aisnet.org/amcis2017/SocialComputing/Presentations/11/

Källström, M., Nousiainen, N., Jern, P., Nickull, S., & Gunst, A. (2022). Mental health among sexual and gender minorities: A Finnish population-based study of anxiety and depression discrepancies between individuals of diverse sexual orientations and gender minorities and the majority population. PloS one, 17(11), e0276550. doi: 10.1371/journal.pone.0276550. PMID: 36331970; PMCID: PMC9635740.

Kang, V., Shao, J., ..., & Wagner, G. C. (2012). Sleep deficiency and sleep health problems in Chinese adolescents. Clinical Medicine Insights: Pediatrics, 6, 11-19. https://doi.org/10.4137/CMPed.S8407

Keles, B. Y., McCrae, N., & Grealish, A. (2020). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth, 25(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

Kelman BB. (1999) The sleep needs of adolescents. J Sch Nurs. Aug;15(3):14-9. PMID: 10745797.

Keogh, E., Bond, F. W., French, C. C., Richards, A., & Davis, R. E. (2004). Test anxiety, susceptibility to distraction and examination performance. Anxiety, Stress & Coping: An International Journal, 17(3), 241–252. https://doi.org/10.1080/10615300410001703472

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of general internal medicine, 16(9), 606-613. doi: 10.1046/j.1525-1497.2001.016009606.x. PMID: 11556941; PMCID: PMC1495268.

Lipson, S. K., Zhou, S., Abelson, S., Heinze, J. E., Jirsa, M., Morigney, J., Patterson, A., Singh, M., & Eisenberg, D. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. Journal of Affective Disorders, 306, 138–147. https://doi.org/10.1016/j.jad.2022.03.038

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Medical care, 46(3), 266-274. doi: 10.1097/MLR.0b013e318160d093. PMID: 18388841.

Malone SK. (2011) Early to bed, early to rise?: an exploration of adolescent sleep hygiene practices. J Sch Nurs. 27(5):348-54. doi: 10.1177/1059840511410434.

Maloney, E. A., Sattizahn, J. R., & Beilock, S. L. (2014). Anxiety and cognition. Wiley Interdisciplinary Reviews: Cognitive Science, 5(4), 403-411. https://doi.org/10.1002/wcs.1299

Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen Hosp Psychiatry. 2015 Jan-Feb;37(1):67-75. doi: 10.1016/j.genhosppsych.2014.09.009. Epub 2014 Sep 23. PMID: 25439733.

Matos, M. G., Gaspar, T., Tomé, G., & Paiva, T. (2015). Sleep variability and fatigue in adolescents: Associations with school-related features. International Journal of Psychology, 50(5), 350-356. https://doi.org/10.1002/ijop.12167

Mazman, G., & Usluel, Y. K. (2011). Gender differences in using social networks. ResearchGate. https://www.researchgate.net/publication/279622067_Gender_differences_in_using_social_networks

McKenzie, S. K., Oliffe, J. L., ..., & Collings, S. (2022). Men's experiences of mental illness stigma across the lifespan: A scoping review. Psychology of Men & Masculinities. Advance online publication. https://doi.org/10.1177/15579883221074789

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. Journal of Psychiatric Research, 45(8), 1027-1035. https://doi.org/10.1016/j.jpsychires.2011.03.006

Ming, X., Koransky, R., Kang, V., Buchman, S., Sarris, C. E., & Wagner, G. C. (2011). Sleep insufficiency, sleep health problems and performance in high school students. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine, 5. https://doi.org/10.4137/CCRPM.S7955

Moody J., Inside Higher Ed. (2022, February 3). Multiple suicides leave WPI reeling. https://www.insidehighered.com/news/2022/02/03/multiple-suicides-leave-wpi-reeling

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. Journal of Educational Psychology, 33(4), 1-12. https://doi.org/10.1177/0143034311427433

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2019). The impact of stress on students in secondary school and higher education. Educational Psychology in Practice, 35(2), 104-112. https://doi.org/10.1080/02673843.2019.1596823

Pigott, T. A. (1999). Gender differences in the epidemiology and treatment of anxiety disorders. The Journal of clinical psychiatry, 60(Suppl 18), 4-15. PMID: 10487250.

Prabu, P. S. (2015). A study on academic stress among higher secondary students. International Journal of Humanities and Social Science Invention, 4(10), 63-68. ISSN (Online): 2319-7722, ISSN (Print): 2319-7714.

Rich, C., Ricketts, J. E., Fowler, R. C., & Young, D. R. (1988). Some differences between men and women who commit suicide. American Journal of Psychiatry, 145(6), 718–722. https://doi.org/10.1176/ajp.145.6.718

Shin, C., Kim, J., Lee, S., Ahn, Y., & Joo, S. (2003). Sleep habits, excessive daytime sleepiness and school performance in high school students. Psychiatry and Clinical Neurosciences, 57(5), 451-453. https://doi.org/10.1046/j.1440-1819.2003.01146.x

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine, 166(10), 1092-1097. doi: 10.1001/archinte.166.10.1092. PMID: 16717171..

Steketee, G., Lam, J. N., Chambless, D. L., Rodebaugh, T. L., & McCullouch, C. E. (2007). Effects of perceived criticism on anxiety and depression during behavioral treatment of anxiety disorders. *Behaviour Research and Therapy, 45*(1), 11–19. https://doi.org/10.1016/j.brat.2006.01.018

Sullivan, M. (2022, November 21). Symptoms of depression on the rise among Westborough High School students. Community Advocate. https://www.communityadvocate.com/2022/11/20/symptoms-of-depression-on-the-rise-among-westborough-high-school-students/

Twenge, J. M., & Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. Journal of Adolescence, 79, 91-102. https://doi.org/10.1016/j.adolescence.2019.12.018

Vytal, K., Cornwell, B., Arkin, N., & Grillon, C. (2012). Describing the interplay between anxiety and cognition: From impaired performance under low cognitive load to reduced anxiety under high load. Psychophysiology, 49(6), 843-862. https://doi.org/10.1111/j.1469-8986.2012.01358.x

Appendix A - Descriptive Statistics

Descriptive Stats

Table 16 - GAD-7 Risk Distribution

Gender	**F**		**M**		**All**
Race / Ethnicity	**A**	**W**	**A**	**W**	
Minimal/None (0-4)	8.03% (22)	9.85% (27)	4.74% (13)	8.76% (24)	31.39% (86)
Mild (5-9)	8.03% (22)	10.95% (30)	11.68% (32)	16.79% (46)	47.45% (130)
Moderate (10-14)	3.28% (9)	7.30% (20)	1.09% (3)	2.19% (6)	13.87% (38)
Severe (>15)	1.09% (3)	4.38% (12)	0.36% (1)	1.46% (4)	7.30% (20)
All	20.44% (56)	32.48% (89)	17.88% (49)	29.20% (80)	100.00% (274)

Table 17 - PHQ-9 Risk Distribution

Gender	**F**		**M**		**All**
Race / Ethnicity	**A**	**W**	**A**	**W**	
Minimal/None (0-4)	16.06% (44)	21.90% (60)	16.06% (44)	25.91% (71)	79.93% (219)
Mild (5-9)	1.82% (5)	5.47% (15)	1.09% (3)	2.19% (6)	10.58% (29)
Moderate (10-14)	2.55% (7)	3.28% (9)	0.73% (2)	0.73% (2)	7.30% (20)
Moderately Severe & Severe (>15)	0.00% (0)	1.82% (5)	0.00% (0)	0.36% (1)	2.19% (6)
All	20.44% (56)	32.48% (89)	17.88% (49)	29.20% (80)	100.00% (274)

Table 18 - Endorse Q9 Risk Distribution

Gender	F		M		All
Race / Ethnicity	A	W	A	W	
None	17.52% (48)	28.10% (77)	16.06% (44)	26.28% (72)	87.96% (241)
Endorsed	2.92% (8)	4.38% (12)	1.82% (5)	2.92% (8)	12.04% (33)
All	20.44% (56)	32.48% (89)	17.88% (49)	29.20% (80)	100.00% (274)

Appendix B - Model Performances

Question Level RFC Performances (No Stratification)

Table 19 - Performances of RFC predictions using Overall Risk Level Features in All Students
(83 test / 191 train)

Target	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
PHQ-9 Risk	0.742	0.846	0.770	0.839	0.884	0.923
GAD-7 Risk	0.766	0.853	0.801	0.861	0.850	0.897
Endorse Q9	0.573	0.870	0.552	0.841	0.734	0.903

Table 20 - Performances of RFC predictions using Question Level Features in All Students
(83 test / 191 train)

Target	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
PHQ-9 Risk	0.863	0.923	0.910	0.936	0.998	0.999
GAD-7 Risk	0.932	0.957	0.932	0.954	0.985	0.990
Endorse Q9	0.628	0.870	0.706	0.877	0.916	0.962

Overall Risk Level RFC Performances (with Stratification)

Table 21 - Performance of RFC predictions of GAD-7 using Overall Risk Level Features in Subgroups

	Group	Train n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.538	0.887	0.506	0.889	0.760	0.938
	F	101	44	0.797	0.836	0.823	0.850	0.867	0.882
Race / Ethnicity	A	73	32	0.799	0.894	0.737	0.893	0.926	0.969
	W	118	51	0.763	0.835	0.807	0.853	0.847	0.888
Both	MA	34	15	0.483	0.933	0.811	0.967	0.796	0.983
	MW	56	24	0.464	0.867	0.473	0.864	0.551	0.871
	FW	62	27	0.797	0.815	0.753	0.766	0.815	0.833
	FA	39	17	0.898	0.929	0.858	0.907	0.916	0.942

Table 22 - Performance of RFC predictions of GAD-7 using Overall Risk Level Features in Subgroups

	Group	Train n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.650	0.908	0.530	0.886	0.795	0.938
	F	101	44	0.754	0.818	0.796	0.827	0.856	0.879
Race / Ethnicity	A	73	32	0.737	0.894	0.814	0.909	0.922	0.950
	W	118	51	0.721	0.835	0.718	0.824	0.851	0.908
Both	MA	34	15	0.464	0.867	0.568	0.884	0.507	0.889
	MW	56	24	0.598	0.883	0.553	0.887	0.749	0.948
	FW	62	27	0.662	0.756	0.673	0.733	0.807	0.839
	FA	39	17	0.840	0.894	0.844	0.910	0.923	0.948

Table 23 - Performance of RFC predictions of Endorse Q9 using Overall Risk Level Features

	Group	Train n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.624	0.908	0.501	0.892	0.740	0.938
	F	101	44	0.647	0.868	0.631	0.805	0.896	0.942
Race / Ethnicity	A	73	32	0.467	0.875	0.484	0.847	0.584	0.879
	W	118	51	0.638	0.886	0.549	0.874	0.708	0.901
Both	MA	34	15	0.534	0.880	0.483	0.865	0.584	0.911
	MW	56	24	0.516	0.908	0.555	0.908	0.589	0.906
	FW	62	27	0.519	0.859	0.527	0.827	0.701	0.901
	FA	39	17	0.515	0.871	0.607	0.810	0.798	0.908

Question Level RFC Performances (without Stratification)

Table 24 - Performance of RFC predictions of GAD-7 using Question Level Features in Subgroups

	Group	Trai n n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.719	0.928	0.902	0.967	0.979	0.992
	F	101	44	0.921	0.936	0.907	0.921	0.999	0.999
Race / Ethnicity	A	73	32	0.923	0.969	0.923	0.964	0.999	0.999
	W	118	51	0.879	0.914	0.906	0.928	0.996	0.997
Both	MA	34	15	0.483	0.933	0.853	0.975	0.959	0.997
	MW	56	24	0.814	0.942	0.736	0.925	0.984	0.993
	FW	62	27	0.900	0.911	0.847	0.855	0.998	0.998
	FA	39	17	0.803	0.894	0.847	0.907	0.981	0.988

Table 25 - Performance of RFC predictions of PHQ-9 using Question Level Features in Subgroups

	Group	Train n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.763	0.938	0.749	0.930	0.998	0.999
	F	101	44	0.902	0.932	0.920	0.932	0.975	0.979
Race / Ethnicity	A	73	32	0.904	0.956	0.903	0.951	1.000	1.000
	W	118	51	0.919	0.949	0.915	0.947	0.992	0.995
Both	MA	34	15	0.464	0.867	0.589	0.887	0.841	0.953
	MW	56	24	0.747	0.917	0.779	0.944	0.961	0.989
	FW	62	27	0.903	0.919	0.895	0.907	0.990	0.991
	FA	39	17	0.946	0.965	0.957	0.973	0.987	0.991

Table 26 - Performance of RFC predictions of Endorse Q9 using Question Level Features in Subgroups

Strat	Group	Train n	Test n	Test F1	Test Acc	Val F1	Val Acc	Train F1	Train Acc
Gender	M	90	39	0.711	0.928	0.638	0.914	0.920	0.983
	F	101	44	0.715	0.859	0.761	0.867	0.940	0.965
Race / Ethnicity	A	73	32	0.591	0.894	0.599	0.871	0.805	0.925
	W	118	51	0.665	0.894	0.641	0.891	0.864	0.952
Both	MA	34	15	0.464	0.867	0.483	0.865	0.682	0.925
	MW	56	24	0.526	0.908	0.575	0.906	0.726	0.929
	FW	62	27	0.655	0.889	0.582	0.827	0.825	0.940
	FA	39	17	0.658	0.906	0.769	0.877	0.901	0.954