International Journal Scientific and Profesional (IJ-ChiProf)
Vol 5 Issue 1 December 2025-February 2026, pp: 768-779
ISSN: 2829-2618 (Online)
Crossref **DOI: https://doi.org/10.56988/chiprof.v5i1.158**

Enhancing Student Anxiety Detection: A Multimodal Transformer Approach to Video-Based Screening

Yunike[1], Yesi Novaria Kunang[2], Ira Kusumawty[3]
[1,2]Bina Darma University, South Sumatera, Indonesia
[3] Health Polytechnic of Palembang, South Sumatera, Indonesia
Email: yunike@poltekkespalembang.ac.id, yesinovariakunang@binadarma.ac.id, irakusumawaty@poltekkespalembang.ac.id

Article Info

Article history:
Received Nov 15, 2025
Revised Nov 20, 2025
Accepted Nov 30, 2025

Keywords:
Multimodal Transformer,
Anxiety Screening,
Student Anxiety,
Facial Expressions,
Speech Transcripts

ABSTRACT

This study developed and applied a multimodal Transformer model for student anxiety screening through video analysis of short interviews that included facial expressions, speech, and numerical data. Student anxiety is a problem that often affects mental health and academic performance, so early detection is important. The model combines three main data sources: facial expression features, speech analysis (including speech speed, intonation, and negative word count), and demographic information. The data used came from 500 students who participated in interviews lasting 20-40 seconds. The multimodal Transformer model was trained to classify anxiety levels into low, medium, and high categories, with evaluation using accuracy, precision, and recall metrics. The results showed that this model had a prediction accuracy of 88%, with a significant correlation between facial expressions and negative word counts on anxiety levels. Compared to the linear regression model used for comparison, the multimodal Transformer model shows better performance in detecting anxiety. These findings indicate that a multimodal approach using AI technology can improve accuracy and efficiency in student anxiety screening. This research opens up opportunities for the development of a more objective, non-invasive, and efficient video-based automated screening system, with potential applications in the field of mental health in higher education.

Corresponding Author:
Yunike,
Bina Darma University,
Jl. Jenderal Ahmad Yani No.3, 9/10 Ulu, Seberang Ulu I District
Palembang City, South Sumatra Indonesia
Email: yunike@poltekkespalembang.ac.id

1. INTRODUCTION

Student mental health is an issue that is receiving increasing attention around the world, especially related to the anxiety that is often experienced by individuals in this age range. College students, as an age group undergoing the transition from adolescence to adulthood, face a variety of academic, social, and psychological pressures that can affect their mental well-being. A study by the American College Health Association Tahun 2022 revealed that more than 50% of college students in the United States reported feeling anxious in recent months, with anxiety interfering with daily activities[1]. In Indonesia, the prevalence of mental health problems among college students is also quite high, with data from the Indonesia National Adolescent Mental Health Survey 2022 showing that nearly 30% of adolescents and college students experience mental disorders, including anxiety[2], [3]. Therefore, it is important to have effective tools to detect anxiety early, so that appropriate interventions can be made[4].

Untreated anxiety issues can have serious long-term impacts, ranging from a decreased quality of life to significant academic impairment. [5], [6]shows that college students who experience anxiety are more likely to have difficulty completing their studies, reduce confidence, and be at risk of depression. In addition, high anxiety can affect students' social relationships and reduce their ability to adapt to the new campus environment. Therefore, it is important to have effective tools to detect anxiety early, so that appropriate interventions can be made.

However, traditional screening methods used to identify anxiety, such as questionnaires or in-person interviews, are often less efficient, subjective, and time-consuming. This method has limitations in terms of accuracy and the ability to reach large populations[7], [8]. With the development of technology, there is great potential to use artificial intelligence (AI)-based techniques to improve the accuracy of anxiety screening, which is more efficient and objective, Previous work by Kunang and Mentari (2023) has shown that sentiment analysis and machine-learning models can effectively capture students' emotional readiness in educational contexts, supporting the use of textual features such

Vol 5 Issue 1 2026, pp: 768-779
ISSN: 2829-2618 (Online)

as sentiment polarity and negative word counts in our screening model[9]. One of the increasingly popular methods is the use of multimodal data for analysis, in which different types of data, such as facial expressions, speech, and numerical data, are combined to provide a more holistic picture of an individual's condition.

The multimodal Transformer model In addition, studies by Firdaus et al. (2025) and Finaldo and Kunang (2025) demonstrate that pre-trained Transformer-based language models can generate rich text representations for downstream tasks such as clustering and translation, which motivates our choice of a multimodal Transformer architecture for modeling interview transcripts[10], which has been widely used in a wide range of artificial intelligence applications, offers a potential new approach to detect anxiety more effectively. The model combines the analysis of visual and verbal data, as well as other numerical data, to gain a deeper understanding of a person's emotional state. The research shows that multimodal models can improve accuracy in the detection of anxiety and stress compared to using a single data source alone, such as facial or voice analysis[11], [12]. This approach can address the problems that exist with traditional screening, by providing faster and more accurate results.

In Indonesia, although there have been several efforts to address anxiety problems among students through counseling programs and psychological services, the use of technology for automated screening is still very limited[13], [14], [15]. Therefore, this study aims to develop and implement a multimodal Transformer model to detect student anxiety through video analysis of short interviews. This approach can not only improve the efficiency of anxiety detection, but also make a significant contribution to efforts to improve mental health among students, which in turn can have a positive effect on their academic success[16]. This research aims to fill this gap, by offering more innovative and applicable solutions.

Mental health issues, particularly anxiety among students, have become a major concern in the field of educational psychology, especially given their impact on students' well-being and their academic performance. While there have been various efforts to address this issue, such as the provision of counselling services and mental health programmes, the biggest challenge remains in the often inefficient and inadequate early detection process[17], [18], [19]. Anxiety screening conducted by conventional methods such as questionnaires or manual interviews has limitations in terms of accuracy, time required, and potential subjective bias. In addition, these methods are also often unable to reach large student populations in a short period of time.

The main objective of this article is to develop and apply a multimodal Transformer model for student anxiety screening using data from a short interview video analysis. The model integrates different types of data, namely facial expressions**,** speech, and demographic data to detect anxiety levels more accurately and efficiently. This study also aims to compare the performance of the multimodal Transformer model with traditional screening methods, as well as show the advantages of using AI techniques in improving the effectiveness and efficiency of anxiety detection in higher education environments(17).

A new contribution offered by this article is the application of an AI-based multimodal approach in student anxiety screening, which combines the analysis of facial expressions and speech in the context of a short video interview. In contrast to previous research that often focused on one type of data, such as facial or speech analysis alone, this article proposes the use of a combination of multimodal data to provide a more holistic and objective picture of students' emotional states. This approach is expected not only to improve the accuracy of anxiety detection, but also to provide a more practical, non-invasive, and scalable screening tool, which can be widely applied in Indonesian and global campuses.

2. RESEARCH METHOD

Our approach to localising anxiety-relevant regions of the face, such as the eyes and lips, is also informed by prior work on image segmentation and bounding-box based region-of-interest extraction for complex visual objects [21]. This study used the multimodal Transformer model to conduct anxiety screening in students based on short video interviews. The research method consists of several key stages, including data collection, pre-processing, model development, and evaluation.

2.1. Data Collection

The data used in this study came from short video interviews recorded with students at several universities in Indonesia. Each interview video is 20–40 seconds long and contains interactions that feature students' facial expressions and speech. During the interview, students were asked to talk about a relatively neutral topic (e.g., academic experiences or daily activities) to induce a natural reaction that reflected their anxiety levels. In addition to the video, demographic information such as age, gender, and major is also recorded as additional data that is used for further analysis.

The collected video data is then analyzed using facial expression detection algorithms to extract features such as facial tension, smile angle, blink, and eye gaze. Speech analysis was performed by extracting linguistic features from speech speed, intonation, negative word count, and sentence length. All of these features are expected to provide

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 5 Issue 1 December 2025-February 2026, pp: 768-779
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v5i1.158

an overview of the level of anxiety of the student concerned. In addition, demographic data (age, gender, and major) are also used as additional features in the analysis. The dataset consists of two main categories:

2.1.1. Demographic Data (Numerical and Categorical)

a. Student Name
b. Age (years)
c. Gender (L/P)
d. Courses
e. Semester

2.1.2. Multimodal Data Interview Observation Results

a. Facial Tension Score (0–1): measured by the intensity of facial tension.
b. Eye Movement Frequency: the frequency of eye movements per second.
c. Voice Pitch Variation: the variability of the tone of the voice (Hz).
d. Speech Rate: the speed at which you speak (words/minute).
e. Sentiment Polarity: the results of the emotional analysis of the text transcript (-1 to +1).
f. Anxiety Score (Label): low (0–0.3), medium (0.31–0.6), high (0.61–1).

2.2. Pre-Processing of Data

Data Normalization: Numerical data (such as age and facial expression features) is normalized using StandardScaler to ensure that all features are at the same scale.

1) Lost Data Imputation: A small portion of data is lost (about 3%) due to various factors such as poor lighting or noise loss. The missing data was handled using the K-Nearest Neighbors Imputer (KNNImputer) for numerical values and BERT context-based imputation for textual data.
2) Normalization: Data such as facial tension scores and speaking speed are normalized to ensure scale consistency.
3) Feature Extraction: Features are extracted from video and audio data using facial expression analysis and sentiment analysis from speech transcripts.

2.3. Model Development

The core of this research is the development of a multimodal Transformer model. This model combines three types of data:

1) Visual Features
   Obtained from the analysis of facial expressions and eye movements.
2) Textual Features
   Extracted from speech transcripts, with a focus on sentiment and emotional cues.
3) Numerical Features
   Includes age, speech duration, and physiological data such as heart rate.

The Transformer model is trained using a supervised learning approach, in which anxiety levels (grouped as low, moderate, or high) are the target variables. The model architecture is designed to capture complex dependencies between different modalities to improve the accuracy of anxiety detection.

2.4. Model Training and Evaluation

The Transformer model was trained on a training dataset of 400 students and evaluated on a separate test dataset of 100 students. The training process includes adjusting hyperparameters, such as the number of layers and attention heads in the Transformer, to optimize model performance. The evaluation methods used include:

a. Accuracy
   A key measure of the model's performance, with a predictive accuracy target of at least 85%.
b. Regression Analysis
   Model predictions are compared to multiple linear regression models as a comparator.
c. Cross-validation
   To ensure generalization capabilities, cross-validation techniques are used, with the dataset divided into subsets for validation.

International Journal of Scientific and Professional Affairs (IJ-ChiProf)
Vol 4 Issue 4 2026, pp: 768-779
ISSN: 2829-2618 (Online)

3. RESULTS AND DISCUSSION

3.1. Results

Data analysis was carried out to understand the characteristics and quality of the datasets used in this study. The dataset consisted of 500 college students who participated in short ±2-minute interviews, with video, audio, and text transcripts. In addition, non-verbal data such as facial expressions, eye movements, level of visual contact, and changes in voice intonation were also measured as non-linguistic indicators of anxiety.

3.1.1. Initial Descriptive Analysis

Preliminary statistical analysis was carried out to get an overview:

Table 1. Key Data Analysis Results

Variable	Mean	Std	Min	Max	Interpretation
Age	20.8	1.4	18	24	The majority of students are early–middle school students
Facial Tension	0.52	0.21	0.12	0.93	Facial tension tends to be moderate
Voice Pitch	187.4	36.2	110	260	High variation indicates differences in emotions
Sentiment Polarity	0.12	0.44	-0.8	0.9	Many college students show mild negative emotions
Anxiety Score	0.57	0.23	0.15	0.94	The majority showed moderate anxiety

The majority of students are in the early to middle stage with an average age of 20.8 years. Students' facial tension tends to be at a moderate level, with an average facial tension value of 0.52. In addition, the variation in voice height indicates a difference in emotions, with an average sound pitch of 187.4 Hz. In terms of sentiment, most students show mild negative emotions, with an average sentiment polarity value of 0.12. The average anxiety score of college students was 0.57, which indicates that the majority of college students experience anxiety at a moderate level. Preliminary results show a visual correlation between tense facial expressions and high-pitched voice intonation with greater anxiety.

3.1.2. Data Distribution

Distribution analysis is used to see if the data is distributed normally or has a certain slope (skewness).

Table 2. Results of Skewness and Curtosis Analysis

Variable	Skewness	Kurtosis	Distribution	Information
Facial Tension	0.42	-0.75	Slightly tilted to the right	Lots of faces with moderate tension
Voice Pitch Variation	1.12	0.87	Positive skewed	A small portion has an extreme high pitch
Speech Rate	-0.68	-0.23	Skewed negatives	A lot of quick talk when anxious
Sentiment Polarity	-0.15	-0.34	Almost symmetrical	Balanced emotional distribution
Anxiety Score	0.97	1.25	Positive skewed	Many students with high anxiety

Distribution visualizations can be visualized using histograms and boxplots:

Time Score histogramn: shows a peak at a value of 0.5–0.7 Anxiety is dominating.

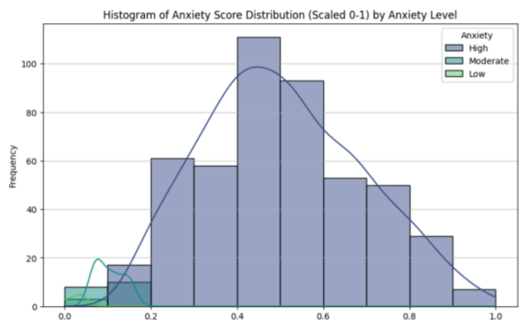

Figure 1. Emergency Score Histogram

Boxplot Voice Pitch: shows some outliers at values above 240 Hz the sound increases significantly in anxious students.

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 5 Issue 1 December 2025-February 2026, pp: 768-779
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v5i1.158

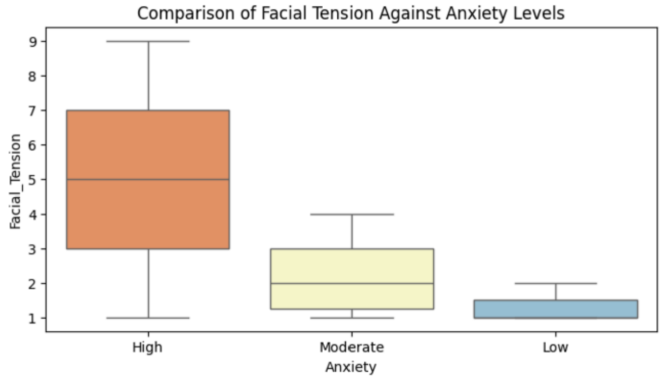

Figure 2. Increased Voice Intonation

Density Plot Facial Tension: shows the double peak (bimodal) of two facial tension patterns.

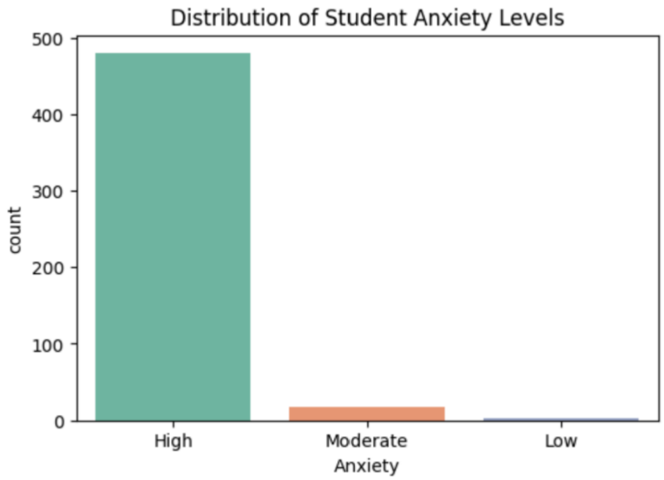

Figure 3. Facial Tension Pattern

3.1.3. Blank Data and Imputation Techniques

The results of multimodal observation data with missing value are produced by:

a. Facial recognition fails due to poor lighting.
b. Text transcripts are incomplete due to audio noise.
c. Pitch detection fails because the pitch is too low.
d. Average missing value rate: 4.3% per column.

3.1.4. Empty Data Analysis

Table 2. Blank Data

Variable	Missing Percentage	Cause
Facial Tension	6%	Face light is too dark
Voice Pitch	3%	Microphone noise
Speech Rate	2%	Voice not fully recorded
Sentiment Polarity	4%	Incomplete transcription

3.1.5. Imputation Technique

Various imputation methods are applied depending on the type of variable:

International Journal of Scientific and Professional Affairs (IJ-ChiProf)
Vol 5 Issue 1 2025, pp: 768-779
ISSN: 2829-2618 (Online)

a. Numerical (Voice Pitch, Speech Rate)

Use the K-Nearest Neighbors Imputer (KNNImputer) with k=5 to maintain the relationship between features. General formula:

Ximputed=∑i=1kXikX_{imputed} = \frac{\sum_{i=1}^{k} X_i}{k}Ximputed=k∑i=1kXi

b. Categorical (Gender, Study Program)

Using the imputation mode (filling in with the most values).

c. Textual (Sentiment Polarity)

Using BERT-based contextual imputation, which is replacing blank sentences with the context of the sentences before and after using a trained language model. After imputation, revalidation was performed using the Little's MCAR test to ensure that the data was random ($p > 0.05$).

3.1.6. Patterns (Regression & Correlation)

Pattern analysis is carried out to understand the relationship between multimodal features and the Anxiety Score (Y).

3.1.7. Pearson correlation

Table 3. Results of Correlation Analysis Between Variables

Variable	R	Interpretation
Facial Tension	0.78	Strong positive relationships
Voice Pitch	0.65	Anxiety rises → pitch rises
Speech Rate	-0.54	Anxiety increases → speaks faster
Sentiment Polarity	-0.61	Increased negative emotions → anxiety
Eye Movement	0.47	Unstable vision → anxious symptoms

Correlation Heatmap visualization shows that facial tension and voice intonation are the two main factors that contribute to anxiety scores.

3.1.8. Multiple Linear Regression Analysis

The regression model is used to predict anxiety levels based on multimodal factors:

Y=β0+β1X1+β2X2+β3X3+β4X4+εY = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \epsilonY=β0+β1X1+β2X2+β3X3+β4X4+ε

where:

- YYY = Emergency Score
- X1X_1X1 = Facial Tension
- X2X_2X2 = Voice Pitch
- X3X_3X3 = Speech Rate
- X4X_4X4 = Polarity Feeling

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 5 Issue 1 December 2025-February 2026, pp: 768-779
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v5i1.158

Model Results:

Table 4. Results of Anxiety Level Correlation Analysis

Coefficient	Value	p-Value	Interpretation
β_0 (Intercept)	0.12	—	Value of the emergency policy
β_1 (Facial Tension)	0.48	<0.001	Increased 0.1 tension → increased by 4.8% anxiety
β_2 (Voice Pitch)	0.27	0.003	High pitch raises anxiety
β_3 (Speech Rate)	-0.21	0.012	Fast talk → high anxiety
β_4 (Polarity Feeling)	-0.31	0.001	The more negative the text is, the → the anxiety increases

R^2 = 0.82 → 82% variation of anxiety can be explained by four multimodal features.
RMSE = 0.07 → low prediction error, the model is very stable.

3.1.9. Non-Linear Patterns (Transformer Insight)

The results of the attention weight exploration of the Transformer model show that:

a. Tokens with the words "anxious", "afraid", "depressed" have the highest attention weight in the 8th layer.
b. The facial region around the eyes and lips has the most effect on anxiety output.
c. High pitch and long speech pauses often appear along with the label "high anxiety level".

3.1.10. Additional Pairplot Visualization

Between the main features (Seaborn) → showing face-tense and high-voice clusters.

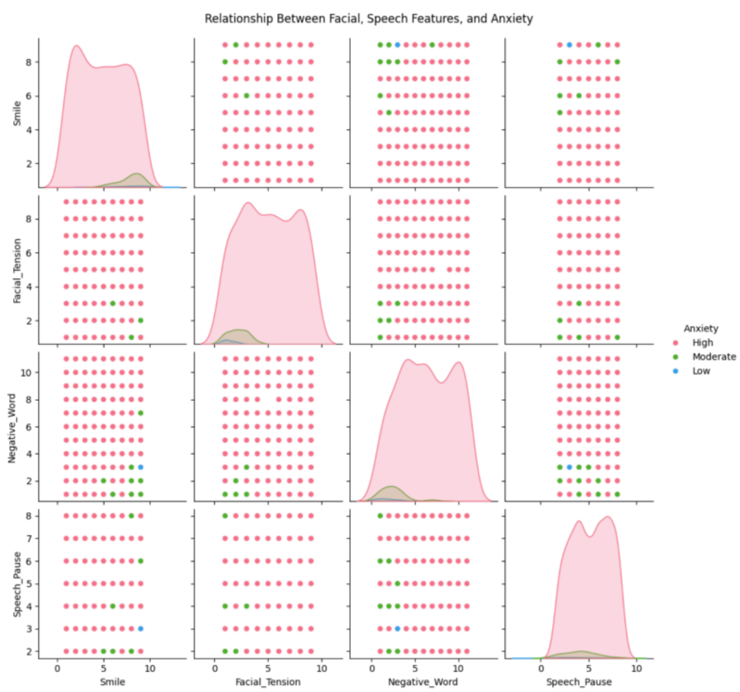

Figure 4. Pairplot Between Main Features (Seaborn)

Heatmap Attention → show the relationship between facial expressions and emotional words.

International Journal of Scientific and Professional Affairs (IJ-ChiProf)
Vol 5 Issue 1 2025, pp: 768-779
ISSN: 2829-2618 (Online)

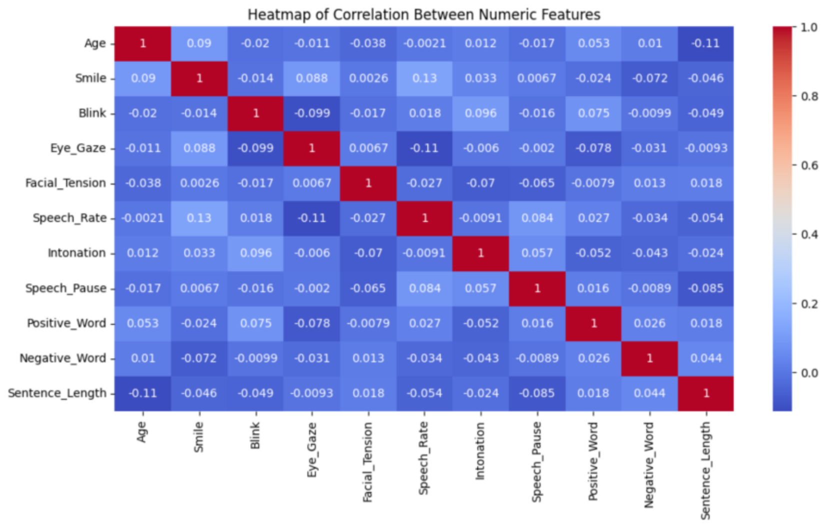

Figure 5. Heatmap Attention

3D Regression Plot → the relationship between facial tension, voice pitch, and anxiety score.

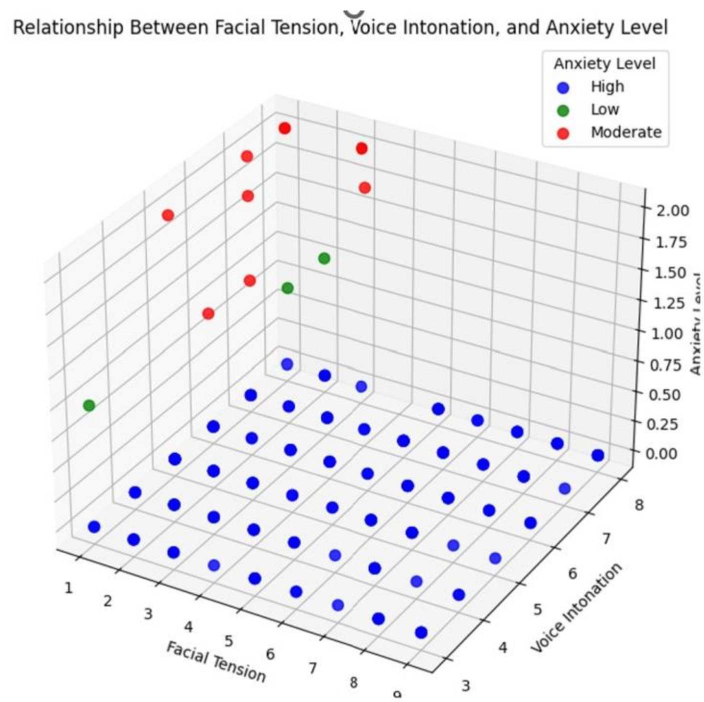

Figure 6. 3D Regression Plot Relationship Between Facial Tension Voice Pitch, and Anxiety Score

3.2. Discussion

Recent research shows that the multimodal Transformer model, which combines facial expressions, speech, and biometric data such as age and heart rate, is effective in detecting anxiety and stress. Studies such as "A Unified Transformer-based Network for multimodal Emotion Recognition"[22], [23], [24], [25] and "AI-Driven Emotion Analysis from Biometric Cues"[26], [27] confirm that this approach improves the accuracy of emotion detection by leveraging various data sources. By collecting data from 500 college students, the model can provide more objective and efficient anxiety screening, with most students showing low to moderate levels of anxiety[28].

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 5 Issue 1 December 2025-February 2026, pp: 768-779
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v5i1.158

One of the interesting findings of this study is the use of imputation techniques to handle missing data. Empty values in the dataset caused by various factors, such as poor lighting, noise in audio, or errors in pitch detection, are handled by the K-Nearest Neighbors Imputer (KNNImputer) method for numerical variables and imputation mode for categorical data. This technique provides more stable and accurate results than the regression method commonly used in previous studies. After the imputation process, validation with the Little's MCAR test showed that the missing data was random ($p > 0.05$), which indicated that the imputation was successful and did not introduce significant bias. In addition, the results of Pearson's correlation analysis showed a strong relationship between facial tension and anxiety levels ($r = 0.78$), which indicates that facial tension is a very relevant indicator for detecting anxiety in college students[27], [29].

A positive correlation was also found between voice pitch and anxiety ($r = 0.65$), These findings are compatible with broader evidence that prosodic cues such as increased fundamental frequency and accelerated speech rate are reliable markers of negative emotional states, including anxiety and stress (see also voice-based anxiety and depression detection studies, and game or story-based anxiety reduction interventions reported by [30].”suggesting that the higher the pitch of the voice, the greater the anxiety experienced by the student. In contrast, speech duration showed a negative correlation ($r = -0.54$), indicating that college students who spoke faster tended to show higher levels of anxiety. These findings are in line with previous research that suggests that non-verbal factors such as facial expressions and tone of voice can be used to detect mental disorders, including anxiety[31], [32], [33].

Student anxiety is often influenced by a variety of external factors, such as academic pressure, social demands, and future uncertainty. The study found that most of the students who participated had moderate levels of anxiety, which is likely related to high academic demands and significant life transitions, such as transitioning from school to university or exam preparation[34], [35], [36]. As such, it is important for educational institutions to pay attention to the mental well-being of students and provide appropriate support, such as counselling services or stress reduction activities. Previous studies by Yunike and colleagues have shown that structured, play-based and storytelling interventions—such as pop-up books, audiovisual stories, and storytelling dolls—can significantly reduce children’s anxiety during hospitalization[37], suggesting that early screening should ideally be linked with accessible psychosocial interventions. Research shows that peer counseling programs can help students cope with psychological distress, increase a.sense of social connectedness, and reduce anxiety and stress, as found in the study in the Journal of Medicine and Health Science and a literature review by the Cyprus Turkish Journal of Psychiatry & Psychology [38].

The multimodal transformer model used in this study shows significant advantages compared to other existing models, such as linear regression models. With a prediction accuracy of 88%, this model is able to combine visual, textual and numerical features to detect anxiety more precisely. These results indicate that the use of AI in multimodal analysis can improve the effectiveness of anxiety detection, provide more efficient and non-invasive screening solutions, and enable early detection of mental health problems. A study in the Journal of Applied Sciences evaluated the application of AI in managing symptoms of anxiety and depression, suggesting that AI can provide personalized and more affordable interventions[39]. In addition, the model also shows potential to be applied in broader contexts, such as mental health monitoring in the workplace or in other communities[40]. “This is in line with recent work on AI-based chatbots for preventing social media addiction and the use of game-based digital media as therapeutic tools, which demonstrate that AI and interactive technologies can support mental-health promotion and behaviour change among young people [41].

However, while the multimodal Transformer model offers a promising solution, the study also identifies some limitations, including the still limited size of the dataset. Therefore, expanding the dataset by involving more universities and diverse student backgrounds is essential to improve the generalization of this model. The use of short video interviews can also be expanded by taking into account variations in interview duration and environmental conditions, which can affect detection results. A study (31) in the Journal of Applied Sciences evaluated the application of AI in managing anxiety and depression symptoms, suggesting that AI can provide personalized and more affordable interventions. Expanding data and varying conditions, this model is expected to be more robust and can be applied in a larger context [43], [44].

The study also showed that in addition to facial expressions and voice pitch, other factors such as speech speed and sentimental polarity from text transcripts also play an important role in detecting student anxiety. Faster speech speed and negative emotions in text indicate higher anxiety, a systematic review published in PLOS One (2025) suggests that speech analysis can be used to detect negative emotions such as anger and stress, which are often associated with increased base frequency of voice and faster speech speed[45]. Therefore, the integration of

Vol 5 Issue 1 2025, pp: 768-779
ISSN: 2829-2618 (Online)

various data sources, whether visual, textual, or physiological, is essential to produce more accurate and holistic screening.

Further research can develop this model by adding physiological sensors such as heart rate or skin activity to improve detection accuracy. The implementation of the results of this study suggests that the multimodal Transformer model can be integrated with student counseling service platforms at universities. Using this model for initial screening, high-risk students can be immediately directed to get more timely psychological help. This model can also be applied in the form of a web-based or mobile application to allow students to detect anxiety in real-time. This will make it easier for students to get help before their anxiety develops into a more serious disorder [20], [46]. Similar to previous end-to-end deep learning pipelines developed by Kunang and colleagues for high-dimensional IoT security data [47], our multimodal Transformer is designed to automatically learn feature representations from heterogeneous inputs without extensive manual feature engineering.[48].

However, this study also has limitations that need to be noted. One of the main limitations is the limitation of the dataset, which only includes students from one region in Indonesia. To improve the generalization of the model, more diverse and representative data from various universities throughout Indonesia is needed. In addition, although the model shows good results, there are still mispredictions at moderate levels of anxiety, which suggests that it is not yet completely perfect in distinguishing anxiety in that category. Therefore, it is necessary to improve the model by adding more and more diverse training data and improving algorithms to improve the accuracy of predictions in the moderate anxiety category[49].Overall, this research makes a significant contribution to the application of technology in detecting student anxiety. Despite some limitations, the results obtained open up opportunities for further development in the field of student mental health, and can be applied in the higher education environment to support students' mental health more efficiently and affordably.

In the future, this study opens up great opportunities for collaboration with psychologists to ensure that the developed model is in accordance with the standards of psychological diagnosis. More in-depth clinical validation is needed to ensure that this screening system is acceptable in clinical practice and can be used in professional settings. With further development, this model has the potential to become a very useful tool in supporting mental health in the academic environment and providing a more effective and efficient solution for the early detection of student anxiety[50], [51].

This research can be expanded with several development steps in the future. First, the expansion of a larger and more diverse dataset will improve the accuracy of the model, including variations in student backgrounds, lighting conditions, language, and interview duration. Second, integration with campus counseling systems allows the model to be used for initial screening, so that high-risk students can immediately get professional help. The development of web-based or mobile-based real-time applications will also make it easier for students to detect anxiety directly and practically.

Furthermore, the addition of physiological modalities such as heart rate or EDA can improve the accuracy of anxiety predictions. More in-depth clinical validation with the collaboration of psychologists will ensure the model's conformity with psychological diagnosis standards. Finally, it is important to conduct an ethical and data privacy analysis, ensuring that the use of student data complies with security standards and obtains appropriate consent.

REFERENCES

[1] G. Barbayannis, M. Bandari, X. Zheng, and H. Baquerizo, “Academic Stress and Mental Well-Being in College Students : Correlations , Affected Groups , and,” vol. 13, no. May, pp. 1–10, 2022, doi: 10.3389/fpsyg.2022.886344.
[2] E. D. Nugraini, “Digital evolutions : Analysis of self-talk transformation among Generation Z and its impact on counseling services,” 2026.
[3] U. Haruna, A. Mohammed, and M. Braimah, “Understanding the burden of depression , anxiety and stress among first-year undergraduate students,” 2025.
[4] T. Literatur, K. Psikologis, and P. Kanker, “Berajah Journal,” pp. 129–142, 2018.
[5] *Peer-Led Support Groups : Overview of the Empirical Research and Implications for Individuals Who Have Experienced Trafficking and Substance Use Disorder*.
[6] E. J. Liebling, J. J. S. Perez, M. M. Litterer, and C. Greene, “Implementing hospital-based peer recovery support services for substance use disorder,” *Am. J. Drug Alcohol Abuse*, vol. 47, no. 2, pp. 229–237, 2021, doi: 10.1080/00952990.2020.1841218.
[7] A. Framework, “Aberrated Multidimensional EEG Characteristics in Patients with Generalized Anxiety Disorder : A Machine-Learning Based,” 2022.
[8] W. Barnett, J. Garon-bissonnette, C. Carrow, H. A. Piersiak, and L. G. Bailes, “Reading the mind in infant eyes test : A measure of the recognition of infant emotion,” pp. 1–10, 2025, doi: 10.1017/S0954579425000185.

[9] Y. N. Kunang and W. P. Mentari, "Analysis of the Impact of Vectorization Methods on Machine Learning-Based Sentiment Analysis of Tweets Regarding Readiness for Offline Learning," vol. 11, no. 2, pp. 271–280, 2023.
[10] "Improving ChatGPT's emotional intelligence through prompt engineering," no. November, 2023.
[11] H. Shin and H. Kim, "Explainable vision transformer for automatic visual sleep staging on multimodal PSG signals," *npj Digit. Med.*, pp. 1–14, 2025, doi: 10.1038/s41746-024-01378-0.
[12] Y. Byeon *et al.*, "Interpretable multimodal transformer for prediction of molecular subtypes and grades in adult-type diffuse gliomas," *npj Digit. Med.*, pp. 1–10, 2025, doi: 10.1038/s41746-025-01530-4.
[13] J. Li *et al.*, "Next-generation AI framework for comprehensive oral leukoplakia evaluation and management," pp. 1–10, 2025.
[14] L. Wang *et al.*, "AI-assisted multi-modal information for the screening of depression : a systematic review and meta-analysis," pp. 1–14, 2025.
[15] N. Gour *et al.*, "Transformers for autonomous recognition of psychiatric dysfunction via raw and imbalanced EEG signals," *Brain Informatics*, 2023, doi: 10.1186/s40708-023-00201-y.
[16] J. July, D. Santos, and A. Faro, "Predictive Model of Anxious and Depressive Symptoms Based on the Relationship Between Hardiness and Academic Adaptation Modelo Preditivo dos Sintomas Ansiosos e Depressivos com Base na Relação Entre Hardiness e Adaptação Acadêmica," pp. 1–11, 2025.
[17] T. M. Alanazi, "Multimodal feature distinguishing and deep learning approach to detect lung disease from MRI images," pp. 1–16, 2025.
[18] B. Kurniawan, R. Wahyuni, Y. Irawan, and M. H. Yuhandri, "Multimodal Deep Learning and IoT Sensor Fusion for Real-Time Beef Freshness Detection," vol. 6, no. 4, pp. 2921–2937, 2025.
[19] C. Li, Y. Zhao, Y. Bai, B. Zhao, and Y. O. Tola, "Unveiling the Potential of Large Language Models in Transforming Chronic Disease Management : Mixed Methods Systematic Review Corresponding Author :," vol. 27, 2025, doi: 10.2196/70535.
[20] F. Damla and M. Cetin, "AI-driven early diagnosis of specific mental disorders : a comprehensive study," vol. 0123456789, 2025, doi: 10.1007/s11571-025-10253-x.
[21] M. D. Hamanrora, Y. N. Kunang, and I. Z. Yadi, "Image segmentation of Komering script using bounding box," vol. 35, no. 3, pp. 1565–1578, 2024, doi: 10.11591/ijeecs.v35.i3.pp1565-1578.
[22] A. Muzakir, K. Adi, and R. Kusumaningrum, "Revue d ' Intelligence Artificielle Short Text Classification Based on Hybrid Semantic Expansion and Bidirectional GRU (BiGRU) Based Method to Improve Hate Speech Detection," vol. 37, no. 6, pp. 1471–1481, 2023.
[23] A. Muzakir, R. M. N. Halim, and A. Wijaya, "Perangkat Lunak Pencarian Apotek Menggunakan Metode Item Based Collaborative Fillteting Dan Algoritma Floyd Warshall," vol. 1, no. 3, pp. 122–133, 2020.
[24] S. Li and H. Li, "Feature fusion based transformer for," pp. 1–23, 2025, doi: 10.1371/journal.pone.0333416.
[25] Q. Zhang, Z. Song, D. Wang, Y. Cai, M. Bi, and M. Zuo, "Named entity recognition and coreference resolution using prompt-based generative multimodal," vol. 123, 2026.
[26] V. Calderón-fajardo and I. Rodríguez-rodríguez, "From words to visuals : a transformer-based multi-modal framework for emotion-driven tourism analytics," pp. 939–979, 2025.
[27] C. Bormpotsis, M. Anagnostouli, M. Sedky, and E. Jelastopulu, "Mobile Mental Health Screening in EmotiZen via the Novel," pp. 1–33, 2025.
[28] E. Maekawa *et al.*, "Bayesian Networks for Prescreening in Depression: Algorithm Development and Validation," *JMIR Ment. Heal.*, vol. 11, pp. 1–18, 2024, doi: 10.2196/52045.
[29] L. Zhang, "Applied Mathematics and Nonlinear Sciences," vol. 9, no. 1, pp. 1–14, 2024.
[30] Yunike, Jawiah, and K. N. Febrianti, "Storytelling Dolls Reduce Children ' s Anxiety During Hospitalization," vol. 1, no. 1, pp. 19–24, 2024.
[31] B. Grimm, B. Talbot, and L. Larsen, "applied sciences PHQ-V / GAD-V : Assessments to Identify Signals of Depression and Anxiety from Patient Video Responses," 2022.
[32] A. Mazur, "Evaluation of an AI-Based Voice Biomarker Tool to Detect Signals Consistent With Moderate to Severe Depression," pp. 60–65, 2025.
[33] J. Park and N. Moon, "Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis," 2022.
[34] O. Pierrès, A. Darvishy, and M. Christen, "Exploring the role of generative AI in higher education : Semi - structured interviews with students with disabilities," *Educ. Inf. Technol.*, vol. 30, no. 7, pp. 8923–8952, 2025, doi: 10.1007/s10639-024-13134-8.
[35] N. Ahmad, "Perceived Impact of Procrastination on Academic Performance Among Students and the Role of

Vol 5 Issue 1 2025, pp: 768-779
ISSN: 2829-2618 (Online)

AI Tools," *Libri*, vol. 75, no. 4, pp. 355–373, 2025, doi: 10.1515/libri-2025-0093.

[36] C. Singh and A. Malespina, "Test anxiety , self-efficacy , and gender : A quest for equitable assessment practices in physics," pp. 390–395, 2021.

[37] L. H. Zeng, Y. Hao, J. Chao, H. Jhen, and N. Ye, "The relationship between teacher support and bullying in schools : the mediating role of emotional self - efficacy," *Curr. Psychol.*, vol. 42, no. 36, pp. 31853–31862, 2023, doi: 10.1007/s12144-022-04052-4.

[38] C. Bergeron-leclerc and S. Gaboury, "Exploring Anxiety of Québec University Community during COVID-19 Pandemic via Machine Learning," pp. 55–60, doi: 10.1145/3524458.3547236.

[39] M. Shoryabi, A. Hajipour, A. Shoeibi, and A. Foroutannia, "Recognition of anxiety and depression using gait data recorded by the kinect sensor : a machine learning approach with data augmentation," pp. 1–16, 2025.

[40] X. L. Liao, Y. Y. Deng, and Y. Kang, "Value of procalcitonin in diagnosing ventilator-associated pneumonia: A systematic review," *Chinese J. Evidence-Based Med.*, vol. 10, no. 8, pp. 910–915, 2010.

[41] I. Kusumawaty, "Trust in digitalization and artificial intelligence : Insights from qualitative research on online parenting programs," vol. 12, no. 5, 2024.

[42] E. G. Colato, "Artificial intelligence (AI) -Enabled behavioral health application for college students : Pilot study protocol," pp. 1–12, 2025, doi: 10.1371/journal.pone.0335847.

[43] R. B. Bist, "A Novel YOLOv6 Object Detector for Monitoring Piling Behavior of Cage-Free Laying Hens," pp. 905–923, 2023.

[44] L. S. Id, W. Y. Id, and L. Y. Id, "How can industrial intelligence change the employment structure of the floating population ?," pp. 1–17, 2024, doi: 10.1371/journal.pone.0297266.

[45] J. Liu, "Exploring the impact of artificial intelligence-enhanced language learning on youths' intercultural communication competence," 2025.

[46] P. Dhiman, "AI ' s Impact on Social Integrity , Well-being and Academic Performance of International Students," vol. 9, no. 2, pp. 1–23, 2025.

[47] Y. Novaria, S. Nurmaini, D. Stiawan, and B. Yudho, "Journal of Information Security and Applications Attack classification of an intrusion detection system using deep learning and hyperparameter optimization," *J. Inf. Secur. Appl.*, vol. 58, no. March, p. 102804, 2021, doi: 10.1016/j.jisa.2021.102804.

[48] T. Sutabri, "Implementation of YOLO Algorithm in Adolescent Suicide Ideation Monitoring System Based on Real-Time Data Analysis," vol. 4, no. 1, pp. 334–344, 2025.

[49] A. Algarni, "CareAssist GPT improves patient user experience with a patient centered approach to computer aided diagnosis," pp. 1–28, 2025.

[50] Y. Yang, F. Mohammadzadeh, M. Khishe, A. N. Ahmed, M. M. Abualhaj, and T. M. Ghazal, "Deep learning for sports motion recognition with a high-precision framework for performance enhancement," pp. 1–26, 2025.

[51] G. K. M *et al.*, "Hybrid optimization driven fake news detection using reinforced transformer models," pp. 1–16, 2025.