Proc. of International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA 2025)
7-9 August 2025, Antalya-Türkiye

Machine Learning Meets the Subconscious: Can AI Decode Dreams?

Erika Dickson
Department of Computer Science
California State University, Fullerton
Fullerton, CA
erikald@csu.fullerton.edu

Dulce Funez Chinchilla
Department of Computer Science
California State University, Fullerton
Fullerton, CA
dfunez1@csu.fullerton.edu

Drashti Mehta
Department of Computer Science
California State University, Fullerton
Fullerton, CA
dumehta@csu.fullerton.edu

Kanika Sood
Department of Computer Science
California State University, Fullerton
Fullerton, CA
kasood@fullerton.edu

***Abstract*—This study examines the relationship between the structured features of dream narratives and the presence of negative emotional content. We extract and engineer features that include demographic indicators (e.g., gender), character composition (e.g., proportions of family, friends, animals, dead or imaginary figures) and emotional behavior indices (Aggression/Friendliness). To model the relationship between dream characteristics and negative emotions, we implement a classification algorithm, k-nearest neighbors (KNN) and multiple regression algorithms: linear regression, k-nearest neighbors (KNN), support vector machine (SVM), random forest, and extreme gradient boost (XGBoost), and evaluate performance using multiple evaluation metrics such as Mean Squared Error (MSE). XGBoost achieved the lowest MSE (0.2065), outperforming other models in capturing non-linear patterns. Among the features, Aggression/Friendliness demonstrate the strongest associations with negative emotions, while Friendliness and sexuality show moderate correlations. Gender has minimal impact, suggesting that dream content may be largely independent of the dreamer's gender. Although the overall predictive performance of the models is modest, the findings provide insight into the dream elements that can influence the emotional tone. This work highlights both the promise and limitations of using machine learning for effective dream analysis and suggests future directions that integrate psychological or contextual variables for improved emotion prediction.**

***Index Terms*—Dreams, negative emotions, nightmares, dream characteristics, machine learning, regression**

I. Introduction

Dreaming is a nearly universal human experience, yet despite its prevalence in our lives, so little is known about dreams and what aspects of our waking lives influence the course of them [1]. This research aims to change this; by using artificial intelligence, we strive to determine if there is any relation between the independent factors observed in our dataset and the dependent feature, the dreamer's overall perception of the dream. The significance of this determination lies in its ability to determine the extent to which popular themes present within a dream positively or negatively influence the course of our dreams. In discovering this relationship, we seek to understand the underlying social and societal perspective of the features of particular interest observed by the dream reports, as revealed by the reporters' subconscious contrivances and their described emotions while undergoing their dreams.

Several machine learning techniques from Python's Scikit-Learn machine learning library are utilized to gauge and process the data, from predicting the percentage of negative emotions present in a dream based on the values of the explanatory variables examined to discovering any patterns of association between the features. Regression algorithms are used primarily due to the nature of our research topic and the data set, which consists largely of continuous numerical values that detail the percentage of the observed themes and emotions present within the dream. Our study aims to understand the relationship between the factors observed by the observers within their dreams and the target variable, a continuous numerical value. Upon determining a relationship between the attributes and target feature, we endeavor to train a model to consistently and accurately predict the degree of negative emotions present within a dream based on the independent variables. Thus, regression machine learning models are principally utilized in this work.

The remainder of the paper is structured as follows. Section II details the background of our research to provide context to the issue at hand; Section III provides an overview of the dataset and data preprocessing performed. Section IV outlines the innovation of this work, followed by the methodology utilized to process the data in the next section. Section VI concludes our research and presents the chronicles of potential future work.

979-8-3315-3562-9/25/$31.00 © 2025 IEEE

II. Background

Dreams are a universal phenomenon that most individuals have experienced. They are complex and confusing, but occasionally we can notice a relationship between our inner thoughts and dreams. Even if dreams are commonly experienced, there are still many aspects about them that remain a mystery. There is so much work that can be done to better understand their meaning. Researchers in the field might pose some questions, such as the following: What can we learn from our dreams? How are our dreams related to events that occur in our daily lives? Will certain factors determine the emotions of dreams?

Extensive research has been done on predicting the type of dream someone might have based on different factors. In "The Typical Dreams of Canadian University Students", Nielsen et al. discuss the differences in the dreams of Canadian students based on their gender and the region they live in [2]. In "Incorporation of Pain in Dreams of Hospitalized Burn Victims", Raymond et al. discuss the effect of past memories of pain on the dreams of burn victims [3]. Different researchers have presented numerous theories and research about dreams, but there is still much to be understood, such as the emotions in dreams. In dreams, people can experience a range of positive and negative emotions. Positive emotions can include happiness, joy, and love, while negative emotions can include fear, sadness, and distress. Negative emotions can cause a dream to become what we refer to as a nightmare. Other dreams may elude any strict categorization, neither positive nor negative, but neutral. Neutral dreams are dreams that are not marked or distinguished by strong emotions that cause any kind of visceral or deep feelings. Some researchers believe that dream emotions come from unexpressed emotions of the day. Sigmund Freud discusses this topic in his book *The Interpretation of Dreams* (1900) [4]. Other researchers feel that dreams just combine random thoughts. Hobson and McCarley hypothesize in "The brain as a dream state generator: an activation-synthesis hypothesis of the dream process", that the dream is the brain making sense of random neural signals while we sleep [5].

The most recent work on [6] showcases the potential of dreams to identify psychological connections. [7] translates dreams into a visual story using brain signals. While these research findings demonstrate potential of the use of AI in decoding dreams, they rely on the brain waves data which can be a challenge for majority of the potential users of the system since majority of the individuals may not have the hardware to gather that form of data.

This work utilizes predictive models to determine if a dream is positive, neutral, or a negative dream/nightmare, based on its percentage. Some of the models used in the research include linear regression, K-nearest neighbor, support vector machine, random forest regressor, and extreme gradient boosting (XGB). We consider the most suitable features to utilize with these predictive models and evaluate their performance. We then check the performance metrics to determine which models work best to predict the negative emotion feature. Different factors of the dream, such as family characters, the presence of animal characters, dead characters, seem useful to predict what type of emotion an individual might have in their dream. Much research has been done to better understand dreams based on words from the dream description. We aim to better identify how some factors in dreams may impact the emotions presented in them.

III. Dataset & Data Preprocessing

The dataset [8] examined within this research is obtained from DreamBank's collection of more than 20,000 dream reports, compiled over the course of several years through a variety of sources and research studies featuring a diverse array of dream reporters of varying ages and backgrounds [9]. Subjects responsible for populating the dataset consist of approximately 70 dreamers who were repeat reporters that provided abundant recounts of their dream descriptions, from which the Hall and Van de Castle (HVdC) coding system is applied to standardize a system of emotional interpretation [9], [10]. This dataset is selected as (1) it is one of the largest datasets currently available on dream studies, both in terms of observations and features and (2) based on Dryad's policies, datasets involving human subjects are required to adhere to strict ethical guidelines and anonymization protocols. The initial dataset is a .tsv file that consists of more than 21,000 data points and 21 features, which is largely altered during the preprocessing stage [8]. Before any preprocessing can be done, the original dataset must first be converted to a .csv file for ease of importing and use. After finishing the file conversion to a .csv file and importing it into a Jupyter Notebook, the data preprocessing can begin. The data preprocessing phase includes inputting and encoding a 'gender' variable as a binary numerical value to indicate the sex of the dreamer, partitioning a train and test set with an 80:20 ratio, removing the majority of features for reasons of irrelevance, inconsistency, and/or incompatibility, visualizing the data with a range of graphs and matrices, as well as normalizing the data to be within a range of 0-1, and supplying derived features to the dataset to aid in extrapolating potentially significant correlations between variables.

The resulting processed dataset is composed of ten input features (Gender, Male, Animal, Friends, Family, Dead&Imaginary, Aggression/Friendliness, A/C Index, F/C Index, S/C Index) and the initial singular target variable, Negative Emotions. The intent behind retaining these features is to determine if there is a relationship between the dreamer's characteristics (gender), the presence of males, animals, friends, family, dead and imaginary characters, as well as the level of aggression per number of characters (A/C Index), friendliness per number of characters (F/C Index), and sadness per number of characters (S/C Index), and the propensity for these traits to contribute to the amount of negative emotions reported by the dreamer [11].

IV. Innovation

Analysis and research have been conducted on dreams for centuries, from Grecian philosophers contemplating the responsible factors of dream induction [12], to modern technology being utilized to conduct neuroimaging on individuals with nightmare disorders [13]. Despite a long history of study, dreams remain a confounding event in which no all-encompassing, conclusive answer to their occurrences have been found. However, many attempts by earlier researchers have sought to explore the meaning behind them, the roles they play within our mental and psychological states, what dreams reveal about how brains operate, and the correlation between situational burdens and personality traits to the tendency toward nightmares, thus advancing our understanding of the presence of dreams in our lives and what they could be indicative of [1], [14], [15]. Our research diverges from traditional dream studies in that we aim to reverse the course of traditional dream analysis. We choose to examine how different factors within our dreams can offer perspectives, thoughts, and feelings of our dreams, as opposed to examining how a person's conscious life can influence the course of their dreams and making inferences from that respect. In doing so, we aim to better understand the emotions and psychological states our dreams are conveying that may otherwise remain suppressed or obscured, even from the dreamers' own knowledge.

V. Approach and Findings

This section outlines our methodology and discusses the dataset we utilize, the approach to handling the data, the feature space we explore, and the regressors we select for this analysis.

A. Features

The feature selection process focuses on dimensionality reduction and data quality. Several metadata columns, such as dream_id, dream_language, and text descriptions, are removed as they are not directly relevant to the analysis. Additionally, coded columns with missing values (characters_code, emotions_code, etc.) are dropped to ensure data completeness. The remaining numerical indices (A/C Index, F/C Index, and S/C Index) are normalized to a 0-1 range using MinMaxScaler, standardizing their impact on the analysis. The final processed dataset retains only the most relevant features for analyzing dream patterns, which are as follows: [11]:

- Gender: Indicates the binary classification of the dreamer's gender, where Female is represented as 1 and Male as 0.
- Male: Represents the proportion of male characters in the dream, calculated as the number of males divided by the sum of males and females.
- Animal: Reflects the proportion of animal characters in the dream, determined by dividing the number of animals by the number of characters.
- Friends: Measures the proportion of friends among human characters in the dream, calculated as the number of friends divided by all humans.
- Family: Shows the proportion of family members and relatives among human characters, derived by dividing the sum of family and relatives by all humans.
- Dead & Imaginary: Represents the proportion of dead or imaginary characters in the dream, calculated as the sum of dead and imaginary characters divided by all characters.
- Aggression/Friendliness: Quantifies the proportion of dreamer-involved aggression compared to the total of dreamer-involved aggression and friendliness.
- A/C Index: Indicates the proportion of all aggression instances relative to the total number of characters in the dream.
- F/C Index: Represents the proportion of all friendliness instances relative to the total number of characters in the dream.
- S/C Index: Measures the proportion of all sexual interactions relative to the total number of characters in the dream.
- NegativeEmotions (target variable): Captures the proportion of negative emotions relative to the total number of emotions in the dream.

As a matter of scientific exploration, we briefly attempt to utilize a classifier in order to determine if classification techniques are more responsive to our attributes and target feature. The target feature is converted from a continuous numerical variable to a binary numeric classifier of zero or one by establishing a threshold, in which values that fall within a specific range are assigned a specific class. In this manner, we utilize a boundary for the values of Negative Emotions, considering values between 0.0 - 0.49 to be a dream that is "not a nightmare" and classifying values between 0.5 - 1.0 as a "nightmare." We utilize a KNN classifier model to evaluate our transformed target feature and discover that with a k-value of ten, our performance reaches an accuracy score of approximately 0.60.

Regression is chosen to explore the impact of features on negative emotions because regression models are well-suited for analyzing continuous correlations between input variables and a desired outcome. Using regression, we can assess how each factor contributes to forecasting changes in unpleasant feelings. This method enables us to measure the effect of various aspects, such as character interactions, emotions, and other pertinent features, on the overall levels of negative emotions in the data. Regression allows us to grasp the underlying patterns better and forecast emotional outcomes. These features indicate the presence of specific entities in dreams. These features are analyzed in relation to the target variable, NegativeEmotions, which represents negative sentiments within dreams. Visualizations such as scatter plots and histograms are implemented to examine trends, distributions, and potential relationships between the features. For instance, the presence of male characters (Male) and family members (Family) is analyzed to assess their influence on negative emotions. To enhance feature selection, we use Weka to correlate features with the target feature and for better visualization. The top three correlated attributes identified are:

- Aggression/Friendliness - 0.18
- A/C Index - 0.16
- F/C Index - 0.08

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (numeric):
        Correlation Ranking Filter
Ranked attributes:
 0.18198    8 Aggression/Friendliness
 0.16435    9 A/CIndex
 0.07681   10 F/CIndex
 0.06377    2 Gender
 0.04157   11 S/CIndex
 0.02571    6 Family
 0.01731    3 Male
 0.00738    5 Friends
-0.00469    4 Animal
-0.00933    7 Dead&Imaginary
-0.10903    1 0

Selected attributes: 8,9,10,2,11,6,3,5,4,7,1 : 11
```

Fig. 1. Correlation of features gathered through Weka.

These findings highlight the role of aggression and friendliness factors and composite indices in shaping the emotional tone of dreams. In addition to tuning our features, we generate a correlation matrix to identify relationships among features and guide the selection process. This correlation matrix reveals several interesting findings in the relationships between different features. It identifies various patterns and correlations in the dataset, emphasizing the independence of numerous dream elements.

Finding 1: Strongest Correlations Among Behavioral Variables.

The analysis indicates moderately positive correlations between the behavioral markers. The most substantial connection is 0.42 between A/C Index (Aggression/Characters Index) and Aggression/Friendliness, and about 0.45 between S/C Index (Sexuality/Characters Index) and F/C Index (Friendliness/Characters Index). These findings imply that certain behavioral features in dreams, such as aggression, friendliness, and sexuality, frequently co-occur and may reflect underlying tendencies in dream narratives.

Finding 2: Gender has Minimal Impact on Most Features.

Gender exhibits relatively weak associations with most features, indicating that it has little influence on the content of dreams, as reflected in the dataset. There is a minor negative

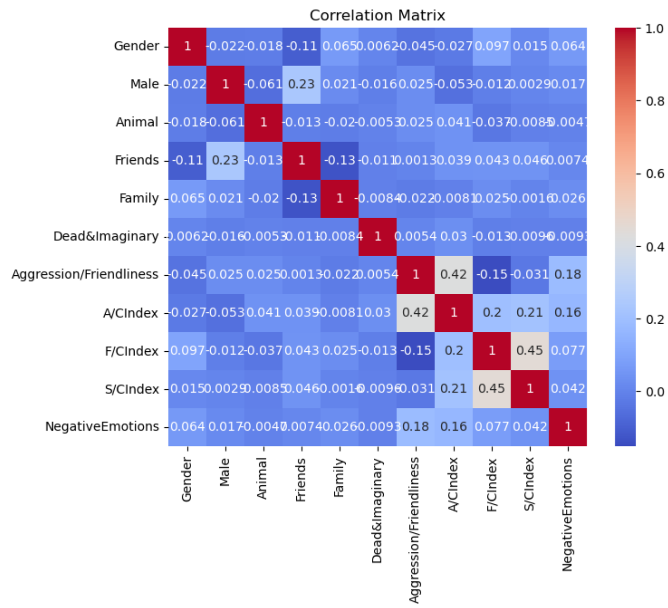

Fig. 2. Correlation matrix.

connection (r = -0.11) between gender and the occurrence of friends in dreams, indicating a subtle gender difference in this area.

Finding 3: Aggression Features are Linked to Negative Emotions.

NegativeEmotions has the strongest associations with Aggression/Friendliness (r $\approx$ 0.18), and A/C Index (r $\approx$ 0.16). While these relationships are small, they indicate that aggressive dreams are more likely to have unpleasant emotional content.

Finding 4: Dream Features Operate Independently.

Most features have weak relationships with one another, showing that many dream aspects function relatively independently. For example, the appearance of animals, family members, or dead/imaginary characters has little link with other dream aspects, implying that the features present in dreams are independent of their context. With the dataset and features defined, next, we implement the machine learning algorithms in our workflow.

B. Regressors

We utilize the following models: linear regression, KNN regressor, support vector machine regressor, extreme gradient boosting regressor, and random forest regressor. Next, we discuss the performance of each of these regressors. Linear regression captures linear correlations between characteristics and the target variable, revealing simple dependencies. The KNN regressor predicts outcomes based on their similarity to the k-nearest data points; its performance is affected by the value of k and the necessity for feature scaling. The optimal value of k, which in our case is 10, is chosen using the elbow plot technique. The SVM regressor effectively sorts data points into groups using decision boundaries, which proves helpful in finding complex patterns. The XGB regressor is an ensemble method that constructs and improves several decision trees consecutively; it excels at dealing with complex patterns and decreases mistakes using gradient-based techniques. Finally, the Random Forest regressor merges many decision trees to create strong predictions, demonstrating its capacity to handle complex patterns while minimizing the risk of overfitting. These models present a variety of techniques for data analysis and interpretation.

To assess model performance and improve prediction accuracy, we evaluate findings using the Mean Squared Error (MSE), a metric that quantifies how close forecasts are to actual data, with lower values indicating greater performance. Among the models, the XGB regressor had the lowest MSE of 0.2065, indicating higher prediction accuracy. Linear regression has an MSE of 0.2172, demonstrating its efficacy in capturing linear correlations. The Random Forest and KNN regressors got MSE values of 0.2267 and 0.2282, indicating they can model more complex patterns. The SVM regressor has the highest MSE of 0.2492, suggesting a relative underperformance in this scenario. These findings help us determine that the XGB regressor is the most accurate model for our dataset.

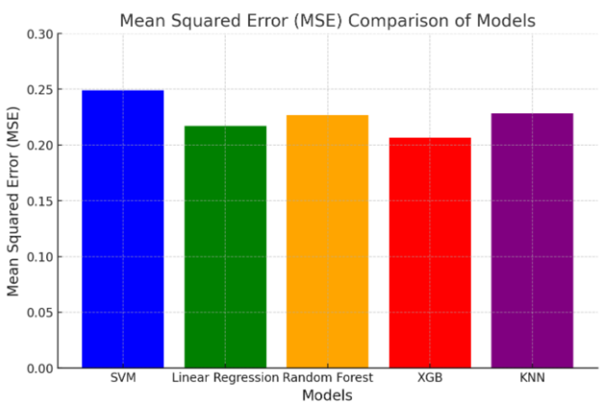

Fig. 3. The bar graph illustrates the Mean Squared Error (MSE) values for five machine learning models: SVM regressor, Linear Regression, Random Forest regressor, XGB regressor, and KNN regressor.

These results suggest several important insights about predicting negative emotions in dreams. First, the low or negative performance across different modeling approaches indicates that the relationship between dream features and negative emotions is highly complex and may not be effectively captured by traditional regression techniques. Second, the slightly better performance of the simpler linear regression model over the other models could indicate either that the selected features do not contain sufficient information to predict negative emotions accurately or that the relationship between dream characteristics and emotional content is too nuanced or subjective to be modeled effectively with the current approach.

Several tests are conducted on the dataset by utilizing supervised learning algorithms. Algorithms we are exploring include: Linear regression, K-Nearest Neighbors regression, Random Forest regression, Huber regression, Lasso regression, Ridge regression, Elastic Net regression, and SVM regression models for their ability to process continuous data for prediction purposes. However, due to the poor performance of certain models, such as the Huber regressor, Lasso regressor, Ridge regressor, and Elastic Net regressor, we forego further analysis on these models. In addition, ensemble learning techniques are utilized to further develop the performance of models. These boosting techniques include an Extreme Gradient Boosting (XGBoost) algorithm and Scikit-Learn's histogram-based boosting regressor. We deploy multiple evaluation techniques to test the performance of each machine learning model, with our best-performing model determined to be the XGB gradient-boosted decision tree.

One of the key issues is the inherent subjectivity of dream data. Emotions and dream features are subjective, which might contribute to unpredictability and noise in the dataset. Furthermore, the weak correlations between numerous attributes and unpleasant emotions indicate that the dataset may be missing some informative variables that can provide more insights that the connections with the machine learning models are too complicated to capture correctly.

VI. Conclusion and Future Work

We conclude that the principal determinant of performance would be the MSE metric, as our research is a regression problem, and the degree to which the models accurately predict the target feature is most significant. From this, we determine that the XGB regressor has the best performance because it has the comparatively lowest MSE value at approximately 0.2065. Of the ones we examine, the worst performer of the algorithms tested is the SVR model with an MSE value of 0.2492.

Our findings provide new insight into the association between dream aspects and unpleasant feelings. Features such as Aggression/Friendliness and A/C Index have the most significant connections with negative emotions, implying that aggression-related aspects in dreams may significantly influence emotional tone. Behavioral variables like friendliness and sexuality also have moderate positive associations, indicating that certain dream traits tend to co-occur. Gender has little effect on most aspects, implying that dream content may be entirely independent of the dreamer's gender. Our machine learning models perform at varied levels, with the XGB regressor achieving the highest accuracy, as evidenced by the lowest Mean Squared Error (MSE) of 0.2065, demonstrating its ability to represent complicated relationships in data successfully. In the future, we would include psychological or contextual variables to analyze their impact on dreams.

This research emphasizes the importance of regression-based approaches in analyzing dream patterns and their emotional consequences. Future studies could build on this work by including larger datasets, investigating upscale feature selection approaches, and using neural networks for more intricate modeling. Furthermore, incorporating deep learning techniques like Recurrent Neural Networks (RNNs) could help capture temporal and sequential patterns in dream descriptions or emotional data. This technique can potentially find more complicated links while also improving predictability. Our findings add to the growing field of machine learning dream analysis, paving the way for additional investigation into the complex relationships between emotions and dream content.

References

[1] S. Givrad. "Dream Theory and Science: A Review." Psychoanalytic Inquiry, vol. 36, no. 3, Apr. 2016, pp. 199–213. EBSCOhost, https://doi-org.lib-proxy.fullerton.edu/10.1080/07351690.2016.1145967.

[2] T. Nielsen, et al. "The Typical Dreams of Canadian University Students." Dreaming, vol. 13, no. 4, 2003, pp. 211–235, https://doi.org/10.1023/b:drem.0000003144.40929.0b.

[3] I. Raymond, et al. "Incorporation of Pain in Dreams of Hospitalized Burn Victims." Sleep, vol. 25, no. 7, Oct. 2002, pp. 765–770, https://doi.org/10.1093/sleep/25.7.765.

[4] S. Freud. *The Interpretation of Dreams*. New York, Modern Library, Cop, 1900.

[5] J. Hobson, and R. McCarley. "The Brain as a Dream State Generator: An Activation-Synthesis Hypothesis of the Dream Process." The American Journal of Psychiatry, vol. 134, no. 12, 1977, pp. 1335–48, www.ncbi.nlm.nih.gov/pubmed/21570, https://doi.org/10.1176/ajp.134.12.1335.

[6] F. Rasheed, M. Waldman. and P.S. Onkar, 2024, November. Dreamcatcher a Real-Time Generative Art Studying Subconscious Imagery. In Indian Conference on Human-Computer Interaction Design and Research (pp. 363-369). Cham: Springer Nature Switzerland.

[7] Y. Fu, J. Gao, B. Yang and J. Feng, 2025. Making Your Dreams A Reality: Decoding the Dreams into a Coherent Video Story from fMRI Signals. arXiv preprint arXiv:2501.09350.

[8] L. Aiello, D. Quercia, & A. Fogli (n.d.). Our dreams, our selves: Automatic interpretation of Dream Reports. Dryad. https://datadryad.org/stash/dataset/doi:10.5061/dryad.qbzkh18fr

[9] Our dreams, our selves: Automatic analysis of dream reports — Royal Society Open Science. (n.d.). https://royalsocietypublishing.org/doi/10.1098/rsos.192080

[10] "Coding Rules for the Hall/Van de Castle System of Quantitative Dream Content Analysis." Dreamresearch.Net, dreams.ucsc.edu/Coding/.

[11] "DreamResearch.net: Measurement and Statistical Issues." Ucsc.edu, 2018, dreams.ucsc.edu/Library/methodological_appendix.html.

[12] H. McCurdy. "The History of Dream Theory." Psychological Review, vol. 53, no. 4, July 1946, pp. 225–33. EBSCOhost, https://doi.org/10.1037/h0062107.

[13] C. Shen, et al. "Waking-hour Cerebral Activations in Nightmare Disorder: A Resting-state Functional Magnetic Resonance Imaging Study." Psychiatry and Clinical Neurosciences, vol. 70, no. 12, Dec. 2016, pp. 573–81. EBSCOhost, https://doi.org/10.1111/pcn.12455

[14] M. Blechner. "The Grammar of Irrationality: What Psychoanalytic Dream Study Can Tell Us About the Brain." Contemporary Psychoanalysis, vol. 41, no. 2, Apr. 2005, pp. 203–21. EBSCOhost, https://doi-org.lib-proxy.fullerton.edu/10.1080/00107530.2005.10745859.

[15] J. Gessert, and M. Schredl. "The Association between Personality Traits and Stress with Nightmare Frequency: An Online Study." Dreaming, vol. 33, no. 3, Sept. 2023, pp. 244–51. EBSCOhost, https://doi-org.lib-proxy.fullerton.edu/10.1037/drm0000241.