Statistical Analysis of Mental Health: U.S v/s India

SPOORTI ANIL BANDIKATTE[1], RAO DEEKSHA UMESH[2], TANUSH N GUNDAWAR[3], SHRADHA ANAND MULIMANI[4]

[1, 2, 3, 4]*Ajeet Kumar Srivastava, R V College of Engineering, Bangalore*

Abstract— Mental health has gained significant attention in recent decades, with an increasing number of reported cases worldwide. This study conducts a comparative analysis of mental health in India and the United States using advanced statistical models. Employing these models, we investigate relationships between various mental health variables in both countries. Additionally, sampling models are utilized to obtain representative data from diverse population ensuring a comprehensive understanding of mental health across Bangalore (India). By combining these methodologies, our research aims to provide valuable insights into the similarities, differences, and unique challenges in the mental health landscapes of India and the United States, contributing to more effective strategies for mental health care and policy development.

Index Terms- Statistical analysis, bar plots, mental health, Bangalore (India), US

I. INTRODUCTION

Mental health refers to a state of emotional, psychological, and social well-being that affects how individuals think, feel, and act. It plays a crucial role in determining how we handle stress, relate to others, and make choices throughout our lives. Good mental health is essential at every life stage, from childhood through adulthood, and it impacts various aspects of life, including personal relationships, work, and educational performance [1] [2].

Mental health disorders can have a substantial impact on a person's day-to-day functioning by compromising their capacity to manage stress and engage in social interactions. People may exhibit a range of symptoms, including anxiety, mood swings, and behavioral changes, which can make it challenging to maintain relationships in both personal and professional contexts. Long-term mental health problems can affect one's physical health as well, raising the risk of diseases like diabetes and heart disease [1][2][3].

Mental health conditions can be generally classified into the following categories due to several contributing factors: Biological Factors: These comprise brain chemistry abnormalities, genetic predispositions, and physical health conditions that may have an impact on mental health.

Psychological Factors: Personal views, coping strategies, and traumatic experiences in the past can all have a big impact on mental health. Social factors: A person's mental health is greatly influenced by their relationships, family dynamics, cultural influences, socioeconomic status, and surroundings, including violence and poverty [4].

The decision to concentrate on mental health stems from the topic's important influence on individuals, families, and communities and the growing awareness of mental health issues in modern society. Addressing mental health issues urgently emphasizes the need for intervention strategies, education, and awareness, particularly in the context of public health [5].

While not as common as it is now, mental health problems were more common in the past [6]. Although diagnostic standards and practices for mental health disorders were still developing, awareness of these conditions increased in the early 2000s. Updates to the Diagnostic and Statistical Manual of Mental Disorders (DSM) and increased research into mental health conditions led to a better understanding of disorders like major depression and generalized anxiety disorder [7].

By employing correlation models, we investigate relationships between various mental health variables in both countries. Additionally, sampling models are utilized to obtain representative data from diverse demographic groups, ensuring a comprehensive understanding of mental health across different populations. Higher SES is frequently correlated with better mental health outcomes, according to correlation studies [8].

Research conducted in the United States has revealed that adolescents from affluent families have easier access to mental health services, which has a big influence on their mental health [9]. By combining these methodologies, our research aims to provide valuable insights into the similarities, and differences, and increase the likelihood of developing successful intervention techniques that cater to the particular requirements of various customer groups, and unique challenges in mental health landscapes of India and the United States, contributing to more effective strategies for mental health care and policy development.

II. METHODOLOGY

A. Data Collection and Preprocessing:

The data was acquired from the survey and CDC (Center for Disease Control and Prevention) a popular platform for the datasets [10]. Two specific data were targeted. One of the data was collected from the survey in Bangalore. Two google forms one for the teens and other for the working classes were distributed throughout the city through mails and other social platforms. Each of the form consisted of around 10 to 15 questions related to the mental health among teens and working classes. And the other data was collected from a popular platform know as CDC from where the US mental health data was collected. Ones the CDC website is opened, select the location, topic and other essential parameters.

This selection was crucial as it provides the foundation for a comparative study or classification task in medical image analysis. When choosing these datasets, factors such as the credibility of the contributors, dataset ratings, comments, and update frequency were considered to ensure data quality and relevance.

Ones the data was collected, they were converted to csv files and survey results were collected in the graphical form. After successful download, the integrity of the files was verified by checking their sizes and ensuring completeness. The downloaded files were then moved to a dedicated project directory on the local machine, preparing them for extraction.

Once extraction is complete, a thorough check of the extracted files is necessary. This involves examining the structure, counting the files and columns in the excel sheet to ensure all expected columns are present, from each dataset to verify they open correctly and are not corrupted.

B. Model Designing:

With the datasets now extracted and verified on the local machine, the focus shifts to setting up the working environment in Google Colab, a cloud-based Jupyter notebook environment. The process begins by opening Google Colab in a web browser and creating a new notebook or opening an existing one dedicated to this project. An essential step here is to integrate Google Drive with Colab, as this will serve as the storage and access point for the datasets. This integration is achieved through Python code that authenticates and mounts the Google Drive within the Colab environment.

Once Google Drive is mounted, an organized folder structure was created within Drive to house the datasets. This involves creating a main project folder, with separate subfolders for healthy cell images and malignant cell images. The extracted CSV files from the local machine were then uploaded to these respective folders in Google Drive. This upload process can be time-consuming depending on the size of the datasets and the internet connection speed.

Ones the upload is done, now begins the crucial step. With all the data now available in the cloud environment, including the training datasets and the test image, the Colab notebook can be prepared for data processing and analysis. This preparation includes installing necessary libraries or dependencies, and importing required Python libraries such as TensorFlow, OpenCV, or numpy. A code was then written and executed to load the images from Google Drive into the Colab environment, including both the training datasets and the test image. This often includes implementing data preprocessing steps like resizing images to a uniform dimension or normalizing pixel values.

C. Model training:

Finally, the main analysis is run on the Google Collab, by giving the pathway of the folder uploaded in the drive. The first pathway uploaded was of the Indian teen csv file. This provided the results with various graphical forms like heat maps, distribution plots, bar graphs and many more. Similarly, the paths of Indian working csv file, and U.S mental data path both teens and adults were uploaded in the code which provided the results in the form of various graphical forms. The results obtained inferred that the model was trained properly with no errors.

With the obtained graphs from both Bangalore and U.S, a comparative study was done between the results obtained from Indian teen data and U.S teen data, Bangalore working data and U.S adults from the various graphs obtained. This gave us an idea on the mental health condition in teens and working classes in both Bangalore and U.S. Various inference were obtained from the results, which provided various factors causing mental health issues like, family pressure, academic pressure, working hours and so on.

From these results and inference, various mental health issues can be solved. This can help in maintaining a good mental health status among teens and the working classes which directly assist and help out in the development of the country's status and improves the productivity.

III. RESULTS AND DISCUSSION

Using the data obtained through surveys and official website of U.S Centers for Disease control and Prevention, mental health trends between the U.S. and India, focusing on stress levels, prevalence of mental health conditions, and the impact of demographic and socioeconomic factors, were compared.

Taking India's Mental Health data into consideration, few statistical analyses has been done. Firstly, the Indian mental health data of teenagers was analyzed. A Horizontal bar chart (Fig.1) has been plotted to understand the different challenges faced by individuals of different genders and to explore the reasons behind these trends.

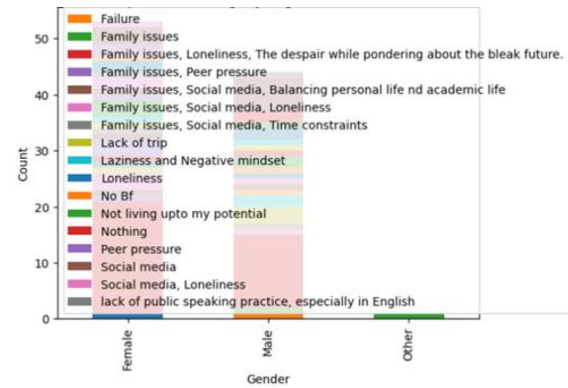

Fig.1

Here, the x-axis represents different gender categories—female, male, and other and the y-axis indicates the number of individuals reporting each issue.

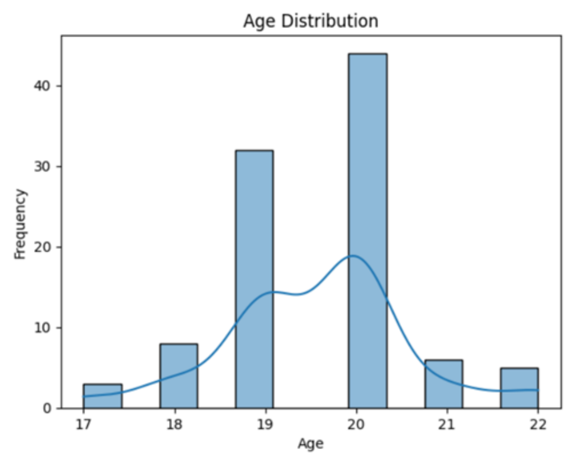

Fig.2

Fig.2 is a histogram with a kernel density estimate (KDE) overlay, representing the age distribution of a certain group of individuals. Here, the x-axis represents the ages of individuals, ranging from 17 to 22 years and y-axis **i**ndicates the number of individuals (frequency) at each age. This type of visualization is useful for understanding the distribution of a single variable (in this case, age) and can help identify trends or patterns within the data.

This graph (Fig.4) is titled as "Distribution of Mental Health Rating". This is a bar chart with a superimposed line graph, showing the distribution of mental health ratings among a surveyed population. The x-axis represents different mental health rating categories, ordered as Good, Fair, very good, Poor, and Excellent. Note that this ordering is not in a typical ascending or descending order of quality and the y-axis shows the frequency or count of individuals for each mental health rating category.

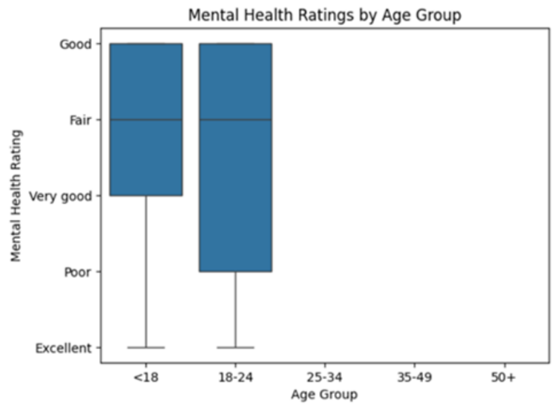

Fig.5

This graph (Fig.5) is titled "Mental Health Ratings by Age Group". This is a box plot (also known as a box and whisker plot). This graph aims to compare the distribution of mental health ratings across different age groups, but the data is limited to the younger age categories. The x-axis represents different age groups: <18, 18-24, 25-34, 35-49, and 50+ and the y-axis shows the mental health ratings, ranging from Excellent at the bottom to good at the top, with poor, very good, and fair in between.

After teenagers, the mental health data of adults was analyzed. Firstly, a correlation heatmap (Fig.6) was plotted. Correlation heatmap is a type of graph used to visualize the correlations between different variables in a dataset. The x-axis and y-axis both contain the same variables: Age, Gender_Encoded, Occupation_Encoded and MHR_Encoded. These variables are likely features or attributes in a dataset, with the "_Encoded" suffix suggesting that categorical variables (Gender, Occupation, and MHR) have been numerically encoded for analysis.

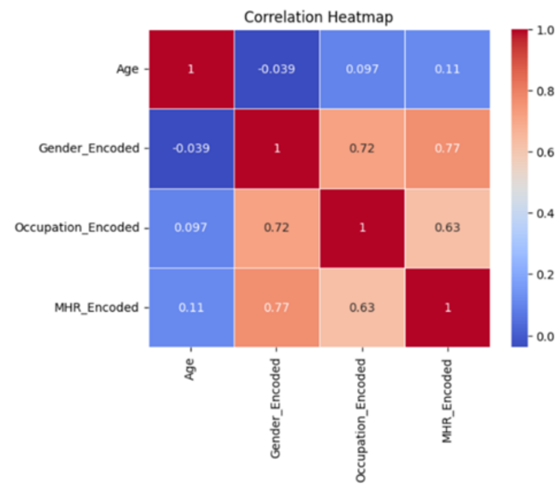

Fig 6

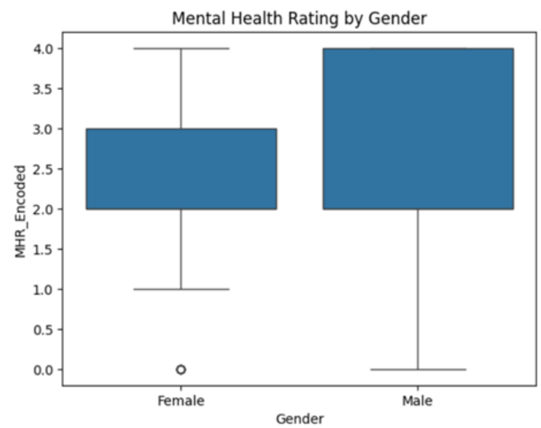

Fig.7

This graph (Fig.7) is a box plot, also known as a box-and-whisker plot. It compares the distribution of Mental Health Ratings (MHR) between genders. The x-axis represents Gender, with two categories - female and male, and the y-axis represents MHR_Encoded (Mental Health Rating Encoded), ranging from 0 to 4.0. The boxes represent the interquartile range (IQR), which contains the middle 50% of the data. The horizontal line within each box is the median. The whiskers extend to show the rest of the distribution, excluding outliers.

Individual points beyond the whiskers represent outliers. From the graph, we can observe that males appear to have a higher median MHR than females. The male distribution has a larger box, suggesting more variability in the middle 50% of their MHR scores. The female distribution shows more spread overall, with longer whiskers and a visible outlier at the bottom. Both distributions are skewed, as the medians are not centered in the boxes.

There's an outlier in the female distribution, represented by a circle at the bottom of the plot. This type of graph is useful for comparing distributions between groups, showing central tendencies, spread, and potential outliers in the data.

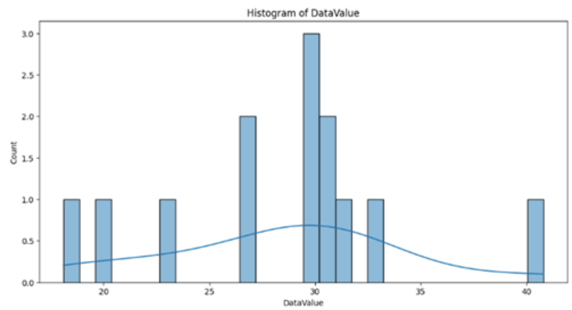

Fig.11

The histogram and KDE curve effectively illustrate the distribution of DataValues from the dataset, which reflects poor mental health statistics among U.S. high school students across various demographic groups. The x-axis ranges from 18 to 42 corresponding to percentages reported in the dataset, with notable values like 18.1% for males and 40.8% for females. The bell-shaped curve, peaking around 30, aligns with common DataValues such as 30.3% for Grade 11, while smaller peaks near 20 and 25 represent groups like "Asian, non-Hispanic" (22.8%). The right-skewed distribution highlights that lower DataValues are more frequent, indicating disparities across demographics. The KDE's peak around 30 suggests a concentration of similar prevalence rates, especially among older students. Overall, this analysis underscores significant trends in mental health prevalence, with distinct variations across sex, race, and grade levels.

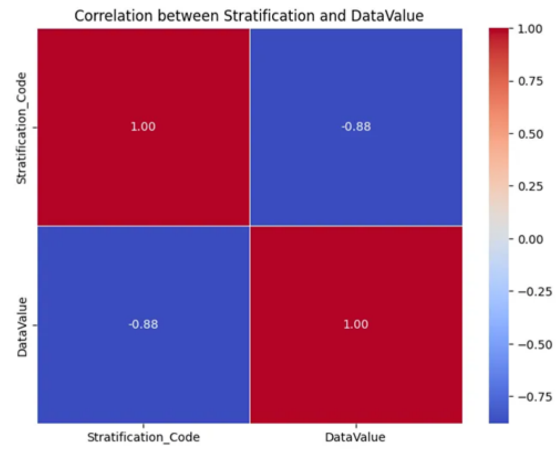

Fig.14

The correlation matrix (Fig.14) reveals a significant inverse relationship between Stratification_Code,(age groups) and DataValue (prevalence of mental distress) in the dataset, reflecting mental health trends among adults across various U.S. states. A strong negative correlation of -0.88 indicates that as the age group increases, the prevalence of frequent mental distress tends to decrease. For instance, younger adults (ages 18-44) report higher levels of mental distress, with percentages like 26.7% in Arkansas, compared to older adults (ages 65+), who show much lower rates, such as 9.4%. The matrix's symmetry and the perfect positive correlation of 1.00 along the diagonal further validate these findings. This analysis suggests that younger adults may face more significant mental health challenges, possibly due to different stressors or life circumstances, emphasizing the need for targeted mental health interventions for this demographic.

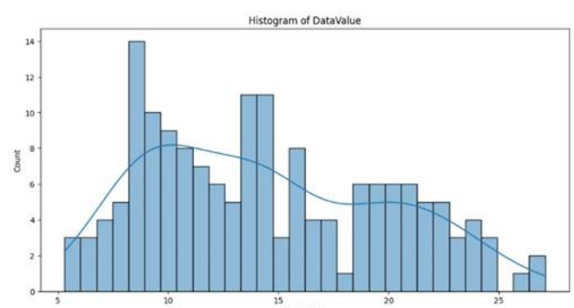

Fig.16

The histogram (Fig.16) provides a clear view of the distribution of DataValue, representing frequent mental distress among adults across various states and age groups. The values span from approximately 5 to

27, with a right-skewed distribution. The most common prevalence is around 8-10, indicated by a prominent peak, while a secondary peak around 14-15 suggests a potential bimodal distribution. The highest frequency bar, with 14 counts, occurs in the 8-10 range, showing that many states report distress prevalence within this range. The right skew indicates that while most individuals report lower levels of distress, there are notable cases of higher prevalence, consistent with younger adults' trends seen in previous analyses.

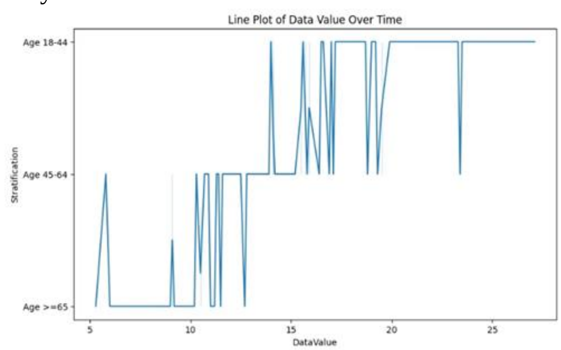

Fig.17

The line plot (Fig.17) vividly demonstrates the relationship between the prevalence of frequent mental distress (DataValue) and different age groups over time. It shows that younger adults (ages 18-44) consistently experience higher levels of mental distress, with DataValues ranging from 18 to 27, compared to older adults (ages 65 and above), who report much lower values, typically below 12. The middle-aged group (ages 45-64) falls in between, with moderate distress levels. This pattern, supported byIV. data from the attached CSV file, highlights distinct age-related differences in mental health, with a clear trend of increasing distress among younger individuals. The plot also reveals significant variability within each age group and a sharp transition between them, emphasizing the critical need for age-specific mental health interventions to address these disparities effectively.

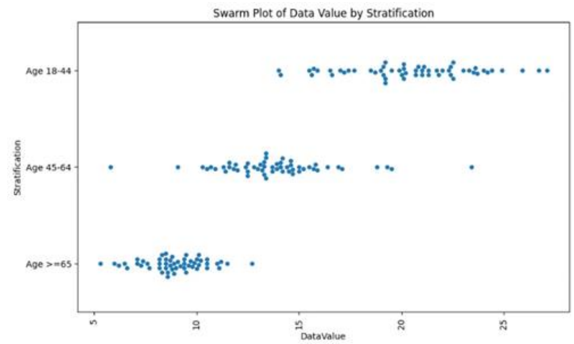

Fig.18

The swarm plot (Fig.18) effectively visualizes the distribution of mental distress prevalence (DataValue) across three age groups, reinforcing the trend that younger adults (ages 18-44) report higher levels of distress, while older adults (ages 65 and above) report lower levels. The youngest group shows the widest spread, with DataValues ranging from 17 to 27, reflecting significant variability in distress levels. In contrast, the oldest group exhibits more compact values between 7 and 12, indicating less variability. The middle age group (ages 45-64) presents moderate distress levels with some overlap with the oldest group, but little overlap with the youngest group. The plot underscores distinct differences across age demographics, with the youngest consistently showing higher distress. This visualization complements previous analyses and highlights the importance of targeted mental health interventions, particularly for younger adults who face greater challenges with mental distress.

V. COMPARISON

The statistical analysis of mental health data from the US and Bangalore reveals distinct trends and disparities between the two regions. In the US, mental health issues among teens and the working class are prominently characterized by a high prevalence of anxiety, depression, and stress-related conditions, especially in urban areas with access to mental health services. Conversely, the data from Bangalore shows a more varied distribution of mental health conditions, with significant differences observed between age groups and socio-economic classes.

One key finding is the higher reported levels of mental health issues among the working class in Bangalore compared to their counterparts in the US, suggesting that workplace stress and socio-economic factors may play a larger role in Bangalore. However, among teenagers, the mental health conditions were relatively similar between the two regions, though cultural factors and differing levels of mental health awareness might have influenced the results.

Discrepancies in the data highlight the subjective nature of mental health and how it is perceived differently by individuals. The varying definitions of mental health and its impact on personal situations underscore the importance of contextual factors such as the facilities available at schools, universities, and workplaces. Additionally, the seriousness with which mental health is regarded in different localities significantly influences the reported conditions.

These findings suggest that while there are commonalities in mental health challenges across regions, tailored interventions are necessary to address the unique needs of each population. Future research should focus on exploring these contextual differences further to develop more effective mental health policies and programs.

VI. FUTURE SCOPE

1. Cross-Cultural Psychometrics
Future research may leverage advanced psychometric techniques to assess mental health constructs across diverse populations. This includes the development of culturally validated scales and measures that allow for standardized assessments of mental health disorders, facilitating cross-national comparisons [11].
2. Biostatistical Modelling
Employing sophisticated biostatistical models, such as multilevel modelling and structural equation modeling, can help elucidate the relationships between social determinants of health and mental health outcomes. These models can account for hierarchical data structures, allowing for a nuanced understanding of how individual, community and societal factors interact [12].
3. Longitudinal Studies
Implementing longitudinal designs can provide insights into the temporal dynamics of mental health issues, identifying causal pathways and long-term effects of interventions. This approach is vital for understanding mental health disorders' progression and preventive measures' efficacy [11].
4. Machine Learning and Predictive Analytics
Integrating machine learning algorithms can enhance predictive analytics in mental health research[11]. By analyzing large datasets, researchers can identify risk factors and predict mental health outcomes, enabling targeted interventions for at-risk populations
5. Health Economics
Future studies may incorporate health economic evaluations to assess the cost-effectiveness of mental health interventions. This includes analyzing the economic burden of mental health disorders and the potential savings from early intervention and prevention strategies [8].

VII. APPLICATIONS

1. Clinical Epidemiology
Research findings can inform clinical epidemiology by identifying prevalence rates and incidence trends of mental health disorders across different demographics. This data is crucial for resource allocation and the development of targeted public health initiatives [13].
2. Public Health Policy
Insights from comparative studies can guide the formulation of evidence-based public health policies. By understanding the mental health landscape in different cultural contexts, policymakers can design interventions that are culturally appropriate and effective [14].
3. Community-Based Participatory Research (CBPR)
Engaging communities in the research process through CBPR can enhance the relevance and impact of mental health interventions. This approach ensures that the voices of diverse populations are included in the development of mental health programs [15].
4. Intervention Trials
Future research may focus on randomized controlled trials (RCTs) to evaluate the effectiveness of specific mental health interventions [14]. These trials can provide robust evidence regarding the efficacy of treatments and inform clinical practice guidelines
5. Translational Research
Bridging the gap between research and practice through translational research can facilitate the implementation of evidence-based mental health interventions in community settings. This includes adapting interventions to fit the cultural and contextual

needs of different populations [12].

CONCLUSION

The research utilized survey data from Bangalore and information from the CDC for the US, employing correlation models and sampling techniques for analysis. The findings revealed various factors influencing mental health, including family pressure, academic stress, and working hours. By identifying similarities, differences, and unique challenges in the mental health landscapes of both countries, the study provides valuable insights that can potentially inform the development of targeted interventions. These interventions could help improve mental health status among teens and working adults, ultimately contributing to increased productivity and national development.

The research underscores the crucial role of mental health in overall well-being and its impact on various aspects of life, including relationships, work, and education.

VIII. ACKNOWLEDGEMENT

The authors express their appreciation to the RSST Trust Bangalore for their continuous support and encouragement.

REFERENCES

[1] World Health Organization, "Mental health," Jun. 17, 2022. [Online]. Available:https://www.who.int/news-room/fact-sheets/detail/mental-health

[2] Mental health. (2024). who. int. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

[3] Adam Felman, "Mental health: Definition, common disorders, early signs, and more," *Medical News Today*, 2024. [Online]. Available: https://www.medicalnewstoday.com/articles/154543.

[4] Deschler and M. Lübbert, "Acute myeloid leukemia: Epidemiology and etiology," *Cancer*, vol. 107, no. 9, pp. 2099–2107, Oct. 2006, doi: 10.1002/cncr.22233.

[5] A.Lu, "Understanding Mental Health as a Public Health Issue," School of Public Health, Tulane University, 2024. [Online]. Available: https://publichealth.tulane.edu/blog/mental-health-public-health/.

[6] American Psychological Association, "Mental health issues increased significantly in young adults over last decade," 2019. [Online]. Available: https://www.apa.org/news/press/releases/2019/03/mental-health-adults.

[7] R. Mojtabai, "National Trends in Mental Health Disability, 1997–2009," *American Journal of Public Health*, vol. 101, no. 1, pp. 191-197, Jan. 2011, doi: 10.2105/ajph.2011.300258.

[8] R. Meghrajani, M. Marathe, R. Sharma, A. Potdukhe, M. B. Wanjari, and A. B. Taksande, "A Comprehensive Analysis of Mental Health Problems in India and the Role of Mental Asylums," *Cureus*, vol. 15, no. 8, p. e42559, 2023. [Online]. Available: https://doi.org/10.7759/cureus.42559

[9] Rajkumar, J. G. Julia, S. Lakshmi K., N. V. Ranjana, P. K. Manjima, R. R. Devi, D. Rukmini, G. Christina, J. Romate, J. G. Allen, J. Abraham, and A. M. Jacob, "Prevalence of mental health problems among rural adolescents in India: A systematic review and meta-analysis," *Scientific Reports*, vol. 12, no. 1, pp. 1-15, 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-19731-2

[10] Patel, A. J. Flisher, S. Hetrick, and P. McGorry, "Mental health of young people: a global public-health challenge," *The Lancet*, vol. 369, no. 9569, pp. 1302-1313, Apr. 2007.

[11] R. S. Murthy, "National Mental Health Survey of India 2015-2016," *Indian Journal of Psychiatry*, vol. 59, no. 1, pp. 21-26, Jan. 2017.

[12] M. L. Hatzenbuehler, J. C. Phelan, and B. G. Link, "Stigma as a fundamental cause of population health inequalities," *Am. J. Public Health*, vol. 103, no. 5, pp. 813-821, May 2013.

[13] R. K. Chadda and K. S. Deb, "Indian family systems, collectivistic society and psychotherapy," *Indian Journal of Psychiatry*, vol. 55, no. Suppl 2, pp. S299–S309, 2013.

[14] R. Kohn, S. Saxena, I. Levav, and B. Saraceno, "The treatment gap in mental health care," *Bull. World Health Organ.*, vol. 82, no. 11, pp. 858-866, Nov. 2004.