



Age-Stratified Mental Health Prediction Using SHAP: An Explainable Artificial Intelligence Framework

Tsholofelo Mokheleli

Department of Applied Information Systems, University of Johannesburg, Johannesburg, South Africa

✉ tsholofelom@uj.ac.za

KEYWORDS

age stratification;
explainable artificial
intelligence; machine
learning; mental
health; shapley
additive explanations

ABSTRACT

Mental health disorders present a growing global concern, yet predictive models often overlook age-specific variations in risk factors. This study introduces an explainable artificial intelligence (XAI) framework for age-stratified mental health risk prediction using Shapley Additive Explanations (SHAP). Using the Open Sourcing Mental Illness (OSMI) dataset, the study stratifies individuals into five different age groups (18-55+ years) and applies machine learning models - Random Forest, extreme gradient boosting, and support vector machine - enhanced with SHAP to identify key predictors across life stages. The findings reveal significant variations in risk factors: younger adults (18-24) are influenced by social support, while familial history and past mental health disorders gain prominence in middle-aged groups (25-54). For older adults (55+), social networks and environmental stressors become critical. Unlike traditional «black box» AI models, SHAP provides interpretable insights, ensuring transparency in predictive decision-making. This study contributes to the literature by demonstrating that mental health risks are not static but evolve with age, necessitating tailored interventions. The framework advances age-specific predictive modelling and offers actionable insights for policymakers and clinicians, particularly in resource-constrained settings. By addressing the limitations of conventional AI approaches, this research establishes a foundation for personalised, explainable, and effective mental health risk assessment across diverse populations. The integration of XAI with age stratification sets a new benchmark for mental health research, highlighting the transformative potential of AI-driven, context-sensitive solutions in addressing the global burden of mental health disorders.

Tsholofelo Mokheleli

Age-Stratified Mental Health Prediction Using SHAP:
An Explainable Artificial Intelligence Framework



1. Introduction

Mental health is a global health concern affecting people of all ages and demographics, with far-reaching social and economic implications. According to the World Health Organisation (WHO), mental health disorders account for 15 % of the global burden of disease, particularly among people aged 10 to 19, with depression and anxiety among the leading causes of disability globally (World Health Organisation, 2024). The situation is particularly severe in low- and middle-income countries (LMICs) due to a lack of resources, stigma surrounding mental health disorders, and limited access to mental health services. Approximately 80 % of people with mental health disorders in LMICs do not receive proper treatment, which leads to worse outcomes and continuing cycles of poverty and inequality (Javed et al., 2021; Nadkarni et al., 2024; Naslund & Deng, 2021). Suicide rates are also worryingly high, with LMICs accounting for nearly 77 % of global suicides annually, often due to untreated mental health disorders. The challenge lies in providing care and identifying at-risk individuals early to prevent the progression of these disorders (Shiraly et al., 2024).

Artificial intelligence (AI) has emerged as a transformative tool for addressing these challenges, mainly through machine learning (ML) and deep learning (DL) approaches. These technologies allow for the analysis of complex datasets to identify patterns and predict mental health risks with unprecedented accuracy (Mohamed et al., 2023). AI models have successfully predicted depression, anxiety, and other mental health disorders using diverse data sources such as clinical records, social media, and self-reported surveys (Byeon, 2023; Wang et al., 2021). However, many of these models are «black boxes» which limits their practical application in sensitive domains such as mental health (Pinto et al., 2023). Explainable AI (XAI) addresses this limitation by making AI models interpretable, allowing users to understand how predictions are made, and which features impact outcomes the most (Dwivedi et al., 2023). Techniques such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are particularly valuable as they allow clinicians and stakeholders to trust and validate AI predictions while also ensuring their alignment with ethical and practical considerations (Ikram et al., 2022; Islam et al., 2024).

Despite these advances, there is still a critical gap in understanding how mental health factors differ across different age groups. Age is often overlooked but is a key determinant of mental health, influencing both risk factors and resilience. For example, young adults may be more vulnerable to mental health challenges caused by academic pressure or social media use. In contrast, older adults may struggle with chronic illness or social isolation. Age-stratified analysis, which divides populations into different age groups for targeted analysis, provides a way to uncover these differences (Kusuma et al., 2024). However, most AI models treat age as a homogeneous factor, failing to account for the subtle differences between age groups. This oversight can lead to generalised interventions that may be less effective for certain groups, particularly in diverse settings such as LMICs, where age-related social and cultural variables play a significant role.

This study addresses these gaps by introducing a framework for age-stratified mental health prediction using ML and SHAP for interpretable AI models. By stratifying data into different age groups and applying SHAP to identify feature importance within each group, the study provides a transparent and interpretable analysis of how risk factors differ across age groups. For example, predictors such as peer communication or internet addiction may have a more significant impact on younger individuals. At the same time, physical health or financial stability might weigh more heavily on older adults. This approach improves the interpretability of AI models and aligns predictive insights with real-world applications, enabling targeted and effective interventions.

This study makes several key contributions to the literature on AI-driven mental health prediction:

1. It introduces a novel framework that integrates explainable AI (XAI) techniques, specifically SHAP, with age-stratified analysis to uncover age-specific predictors of mental health risks. This addresses a critical gap in the literature, as most existing models treat age as a homogeneous factor, leading to generalised and less effective interventions.
2. The study demonstrates that mental health predictors are dynamic and evolve across different life stages, providing a different understanding of how risk and protective factors change with age.
3. Using XAI ensures the model's predictions are interpretable and actionable, bridging the gap between data science and practical mental health care.

These contributions have significant implications for research and practice, offering a roadmap for developing targeted, age-appropriate mental health interventions, particularly in resource-constrained settings.

The rest of the paper is structured as follows: Section 2 discusses relevant work, including the existing literature on AI in mental health, the function of XAI, and age stratification. Section 3 outlines the dataset, stratification process, and analytical framework; Section 4 - results and discussion presents the findings, highlighting SHAP-based interpretations for each age group; Section 5 presents the implication of the study, which explores the practical applications of these insights in improving mental health outcomes; Section 6 - limitations and future work address areas for refinement and further exploration; and Section 7, the conclusion.

2. Literature Review

AI advancements have revolutionised mental health prediction, allowing novel techniques to identify risk factors and improve intervention strategies. However, the complex interplay between age and mental health outcomes is still underexplored, with few studies addressing age-specific variations in predictive modelling. XAI techniques, such as SHAP, have emerged as effective tools for increasing transparency and understanding in ML models, particularly in high-stakes domains such as mental health. While research has shown the effectiveness of AI and XAI in identifying critical predictors across diverse contexts, the absence of age-stratified analysis has limited the granularity of insights, limiting the development of tailored mental health solutions. This section explores existing literature on AI in mental health, the role of XAI in improving interpretability, and the value of age-stratified risk assessment, highlighting gaps that this study aims to address.

2.1. AI in Mental Health

AI has increasingly been applied to mental health prediction, using ML and DL techniques. Ding et al. (2025) explored ML models such as Backpropagation Neural Networks and adaptive boosting (AdaBoost) to predict mental health risks among youths aged 14-25 in China. They used SHAP to guide feature selection and found that emotional intelligence, coping strategies, and learning pressure were key predictors with strong predictive performance (C-index = 0.9120). However, the study did not stratify outcomes within the age range, leaving gaps in understanding age-specific variations. Islam et al. (2024) applied voting-based classifiers and XAI techniques (LIME and SHAP) to model anxiety

among Bangladeshi university students, achieving 96.5 % accuracy. Although age was categorised into broad groups, the analysis lacked detailed stratification, leaving variations in anxiety predictors unexamined.

Kerz et al. (2023) used interpretable bidirectional long short-term memory (BiLSTM) models and a pre-trained model named MentalRoBERTa transformers to detect mental health conditions through language behaviour. While MentalRoBERTa outperformed other models, XAI techniques such as LIME and Attention Gradient Analysis (AGRAD) were critical in highlighting predictors. The study emphasised the utility of XAI in mental health applications. Byeon (2023) reviewed ML applications in depression prediction, emphasising the importance of SHAP in increasing transparency. Despite research into predictors such as social isolation and financial distress, the review highlighted the absence of studies using age-stratified analysis for mental health prediction.

2.2. Explainable AI

Explainable AI (XAI) has transformed predictive modelling by clarifying the importance of features and providing actionable insights. SHAP has proven to be useful in healthcare applications. Wang et al. (2021) used SHAP to analyse environmental stressors contributing to mental disorders, demonstrating that nitrogen dioxide (NO₂) and sulphur dioxide (SO₂) have significant, complex interactions. Ikram et al. (2022) applied extreme gradient boosting (XGBoost) and SHAP to predict COVID-19-related mental health risks, identifying predictors such as sleep disturbances and income levels. Although age was a significant factor, the study did not stratify results to explore variations in predictive patterns.

Khan et al. (2024) used SHAP to analyse life satisfaction determinants with ensemble ML models and found that health and financial stability were key predictors across all age groups. While the study stratified participants into age categories, it focused on global feature importance rather than detailed age-specific analyses. Kerz et al. (2023) showed how SHAP improved interpretability in language-based mental health prediction, emphasising stylistic and lexical features as critical for conditions such as Attention-Deficit/Hyperactivity Disorder (ADHD) and depression. These studies highlight the adaptability of the XAI technique SHAP while emphasising the underuse in age-stratified mental health analysis.

2.3. Age-Stratified Risk Assessment

Age-stratified analysis has been used in various domains to improve prediction accuracy and reveal different trends among subgroups. Domínguez-Olmedo et al. (2022) developed XGBoost models for age-stratified COVID-19 mortality prediction, revealing different biomarkers for different age groups, such as D-Dimer for younger patients and C-reactive protein (CRP) for older patients. González-Nóvoa et al. (2021) tailored intensive care unit (ICU) alarm systems using SHAP, identifying age-specific thresholds for critical variables such as Glasgow Coma Scale scores and respiratory rate. These studies demonstrated the benefits of stratified modelling in healthcare, but their limitations regarded physical health outcomes. In agriculture, Kabir et al. (2024) introduced an age-stratified tea leaf dataset to demonstrate how stratification improves model accuracy for quality prediction.

Similarly, Lee et al. (2019) conducted an age-stratified analysis of internalised homophobia and mental health outcomes among Korean lesbian, gay, and bisexual (LGB) adults, identifying different patterns of depressive symptoms and suicidal ideation across age groups. While impactful, the study used traditional statistical methods, limiting predictive capacity. Zhang et al. (2024) investigated

loneliness among adolescents (11-16 years old), identifying predictors such as peer communication and internet addiction without detailed age subgroup analyses. These studies underline the potential of age stratification but highlight its lack of application in mental health AI research.

Despite advances in AI and XAI for mental health, the integration of age-stratified analysis is still underexplored. To our knowledge, no study has comprehensively used SHAP to reveal age-specific patterns in mental health prediction. This study addresses this gap by demonstrating that predictors of mental health risks differ across age groups and uses SHAP to provide interpretable insights. By integrating age stratification, this work advances methodology and practical applications, providing a novel contribution to understanding the interplay between age and mental health outcomes.

3. Methodology

This section describes the dataset used, data preprocessing methods, the age groups for stratification, model selection and hyperparameter tuning, evaluation metrics, and SHAP analysis.

3.1. Dataset

This study uses the Open Sourcing Mental Illness (OSMI) dataset, which combines data collected between 2016 and 2023 (OSMI, 2023). Due to its diversity and depth, the OSMI dataset is an important resource in mental health research. It includes a wide range of features such as demographic information, workplace-related factors, perceptions of mental health, productivity impacts, and experiences of stigma. Although initially designed as a technology sector survey, its findings and insights are applicable across various environments, enabling generalisability in mental health studies.

The consolidated dataset contains 3,439 instances and 47 consistent features before data preprocessing. These selected features capture critical dimensions such as mental health benefits provided by employers, perceptions of anonymity and support in the workplace, individual mental health history, and the effect of mental health on productivity. Notably, while the dataset includes a variety of workplace-related factors, its insights extend to broader mental health applications. The dataset was downloaded from the official OSMI repository, and appropriate preprocessing steps were applied to clean and format the data for analysis. This involved handling missing values and encoding categorical variables as required for ML models.

3.2. Data Preprocessing

The data preprocessing stage involved combining all datasets, as described in the dataset subsection, followed by thorough cleaning and preparation for analysis. This process addressed inconsistencies, missing values, and unrealistic entries to ensure the integrity and suitability of the dataset for the study.

Gender presented a unique challenge, as the survey allowed participants to provide free-text responses, resulting in a wide range of entries. To standardise this feature, gender was simplified into three categories: «Male», «Female» and «Other». The age variable also required significant cleaning due to the presence of outliers, including an unrealistic value of 223. The age range was restricted to 15-70 years. The minimum age of 15 was chosen as the youngest plausible respondent likely to have taken the survey, even though 18 is a more common minimum age for survey participation. This

broader range was used to ensure a more comprehensive representation of age groups in the age-stratified analysis. The maximum age of 70 was selected to exclude outliers and ensure the relevance of the data to the general population.

Categorical variables with fewer than 10 unique options were manually label-encoded to ensure consistency and usability in ML models. For example, features such as mental health benefits and perceptions were converted into numeric codes. Country-related features, such as the respondent's country of residence and work, were encoded using Label Encoder, converting categorical data into numeric data and allowing for efficient handling in the modelling process (Mbuya et al., 2023). Missing data was addressed using a Simple Imputer with the mode (most frequent value). This method ensured that the dataset remained consistent while retaining the interpretability necessary for the study (Mokheleli et al., 2024). Once the data was cleaned and processed, it was ready for subsequent stages of analysis, supporting a robust age-stratified mental health risk assessment.

Handling imbalanced data was a critical step in the preprocessing stage to ensure strong and unbiased model performance. The study employed a hybrid approach combining the Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links, implemented as SMOTETomek. Individually, SMOTE addresses class imbalance by synthetically generating new samples for the minority class through interpolation between existing samples, thereby increasing its representation (Balogun et al., 2020; Mokheleli et al., 2023). Conversely, Tomek Links identifies and removes ambiguous samples nearest neighbours of differing classes and lies close to the decision boundary, effectively cleaning the dataset by reducing class overlap (Leng et al., 2024). SMOTETomek combines these techniques, balancing the class distribution and improving dataset quality by reducing noise and overlap (Ratantja Kusumajati et al., 2024). This combined technique was selected for its ability to improve the discriminative power of ML models, particularly in datasets where class imbalances can obscure meaningful patterns (Talukder et al., 2024). The balanced dataset ensured fairness in predictions and supported the generation of reliable insights across all age-stratified groups.

3.3. Age Stratification

Age stratification is important in this study because it allows for the analysis of mental health trends across different life stages. Stratifying the data by age enables the identification of specific mental health challenges and risk factors unique to each demographic group, thereby providing different insights that can inform targeted interventions. This approach is particularly important for understanding how mental health disorders manifest and are perceived within different age groups, ensuring that the findings remain relevant to a diverse population.

During the exploratory data analysis, an initial age stratification was employed to group participants into the following categories: '0-17', '18-24', '25-34', '35-44', '45-54', '55-64', and '65+'. These groups were designed to reflect key developmental and life stages, including adolescence, early adulthood, midlife, and older adulthood. However, the '0-17' was excluded from subsequent analyses due to their low representation in the dataset, which limited the statistical power and generalisability of findings for the group. The age group 55-64 was combined with age 64+ to have more representation. Therefore, the final age stratification for the analysis focused on five groups: '18-24', '25-34', '35-44', '45-54', and '55+'.

By employing this age stratification, the study ensures a detailed examination of mental health trends across different phases of adulthood. This approach improves the interpretability of the findings and facilitates the development of age-specific strategies to address mental health concerns effectively.

3.4. Model Selection and Hyperparameter Tuning

This study employed three ML classifiers: random forest (RF), extreme gradient boosting (XGBoost), and support vector machine (SVM). These classifiers were selected due to their sophisticated algorithms, which allow them to model complex patterns in the data but at the expense of direct interpretability (Sagi & Rokach, 2021; Sun et al., 2024). RF is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions for classification tasks. It is robust enough to overfit and effectively handle high-dimensional datasets (Sun et al., 2024). XGBoost, a scalable and efficient implementation of gradient-boosted decision trees, is known for optimising performance through regularisation techniques, making it highly suitable for datasets with intricate relationships between features (Sagi & Rokach, 2021). SVM, a powerful supervised learning algorithm, constructs a hyperplane in a high-dimensional space to separate classes with maximum margin. It is particularly effective in handling non-linear classification problems using kernel functions (Montesinos López et al., 2022).

Hyperparameter tuning was conducted using Grid search cross-validation (GridSearchCV), a systematic approach to identify each classifier's optimal combination of hyperparameters (Pirjatullah et al., 2021). Given the computational intensity of GridSearchCV, it was executed once to determine the best-performing hyperparameters, which were then consistently applied in subsequent model evaluations. This approach ensured computational efficiency while maintaining the reliability of the results. The integration of these advanced classifiers and systematic hyperparameter tuning provided a robust foundation for the predictive modelling in this study.

3.5. Evaluation Metrics

The evaluation of the predictive models in this study relied on a range of metrics to ensure a comprehensive assessment of performance across different aspects of classification. These metrics included Accuracy, Precision, Recall, F1 Score, and Receiver Operating Characteristic Area Under the Curve (ROC AUC) (see Table 1).

Table 1. Evaluation metrics

Metric	Formula
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$
Recall	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
F1 Score	$2 \times \frac{Precision * Recall}{Precision + Recall}$
ROC AUC	$Integral \int_0^1 ROC$

Accuracy measures the proportion of correctly predicted instances out of the total, offering a general overview of model performance (Fergus & Chalmers, 2022). Precision quantifies the ratio of true

positive predictions to all positive predictions, indicating the model's reliability in identifying positive cases. At the same time, recall, also known as sensitivity, calculates the proportion of correctly identified true positives out of all actual positives, reflecting the model's ability to detect relevant cases (Fergus & Chalmers, 2022). The F1 Score, the harmonic mean of Precision and Recall, balances these two metrics, making it particularly useful when dealing with imbalanced datasets (Fergus & Chalmers, 2022). ROC AUC evaluates the ability of the model to distinguish between classes, providing an aggregate measure of performance across all classification thresholds (Bowers & Zhou, 2019).

3.6. SHAP Analysis

To improve the interpretability of the predictive models employed in this study, SHapley Additive exPlanations (SHAP) analysis was used. SHAP is a state-of-the-art explainability technique that quantifies the contribution of each feature to a model's predictions (Letoffe et al., 2024). Rooted in cooperative game theory, SHAP assigns a Shapley value to each feature, representing its marginal contribution to the prediction when combined with other features (Li et al., 2024). This ensures a consistent and theoretically grounded approach to understanding how individual features influence model outcomes, addressing the ML algorithms' complex «black box» nature.

The SHAP visualisation was used in this study, specifically, the summary plot, which is instrumental in identifying key features influencing mental health predictions. The summary plot provides a high-level overview of feature importance across the dataset by combining feature contribution magnitudes and directions (positive or negative impacts on predictions). Each point on the plot represents a Shapley value for an individual instance, enabling the identification of features that consistently drive predictions across the population. This visualisation is crucial for understanding the global behaviour of the model and determining which features most significantly influenced mental health risk assessments across different age groups. This granular insight allows to interpret model predictions case-by-case, facilitating a deeper understanding of how the model processes unique feature combinations.

In the context of this study, SHAP analysis was important in identifying key features that influence mental health outcomes, enabling the interpretation of their significance across age-stratified groups. By providing actionable insights into the relationships between these features and mental health outcomes, SHAP analysis supports the development of targeted interventions. It improves the transparency and trustworthiness of the predictive models.

4. Results and Discussion

4.1. Exploratory Data Analysis

Exploratory data analysis (EDA) is a critical step in any data-driven study, uncovering patterns, identifying anomalies and providing a comprehensive understanding of the dataset before applying advanced modelling techniques (Páez & Boisjoly, 2022). Using visualisations and descriptive statistics, EDA helps contextualise the data, guiding the selection of appropriate analysis methods and ensuring the robustness of subsequent results. In this study, EDA was conducted to explore the distribution of key demographic and mental health-related variables across the age-stratified groups, providing foundational insights into the dataset's characteristics and preparing it for predictive modelling.

The EDA focuses on three key aspects relevant to this study: the distribution of diagnosed mental health conditions across the defined age groups, the gender distribution within each age group, and the impact of addressing class imbalance using resampling techniques.

Figure 1 illustrates the distribution of mental health diagnoses across age groups. The Yes category represents individuals diagnosed with a mental health condition, while the No category reflects those without a diagnosis. The highest prevalence of diagnosed mental health conditions is observed in the 25-34 and 35-44 age groups, with the 25-34 group showing the most significant disparity between diagnosed and non-diagnosed individuals. Conversely, the 18-24 and 55+ age groups have significantly lower counts, indicating reduced representation or fewer diagnoses in these groups. This distribution shows the importance of age-stratified analysis in identifying key demographic trends.

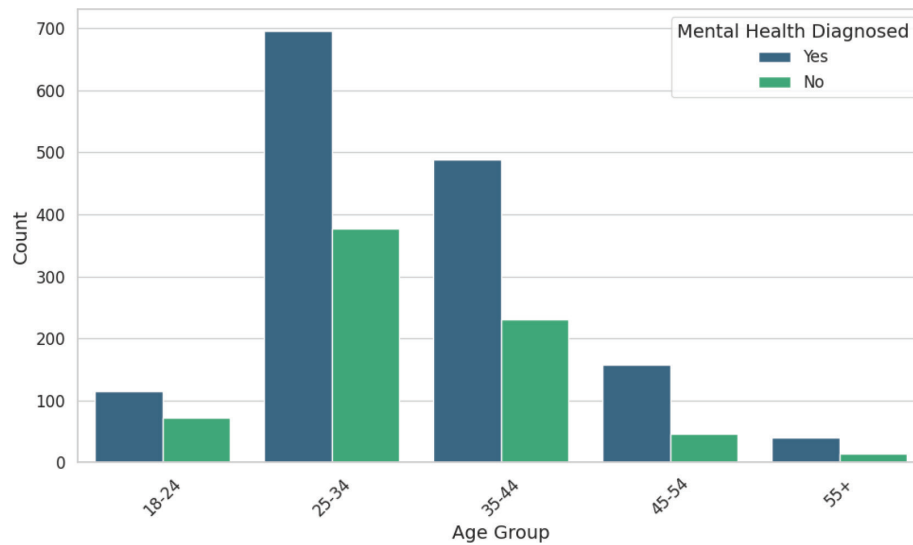


Figure 1. Distribution of diagnosed mental health conditions by age group

Figure 2 illustrates the class distribution before and after applying the SMOTETomek technique, which addressed the significant imbalance in the dataset. Before resampling, the dataset was heavily skewed, with 1,498 instances in the «Yes» class and only 740 instances in the «No» class. This imbalance posed a serious challenge to the predictive models, as they would have been biased toward the majority class («Yes»), leading to misleadingly high accuracy scores while failing to detect the minority class («No») adequately. Such an outcome would have diminished the model's ability to generalise and its reliability in real-world applications, where correctly identifying minority outcomes is critical.

SMOTETomek, a hybrid resampling technique, was applied to address this issue. This method combines SMOTE, which creates synthetic instances of the minority class, with Tomek Links, which removes ambiguous data points close to the class decision boundary. As shown in Figure 2, this approach successfully balanced the dataset, resulting in an equal distribution of 1,186 instances for both classes. By addressing class imbalance, the resampling process improved the model's ability to

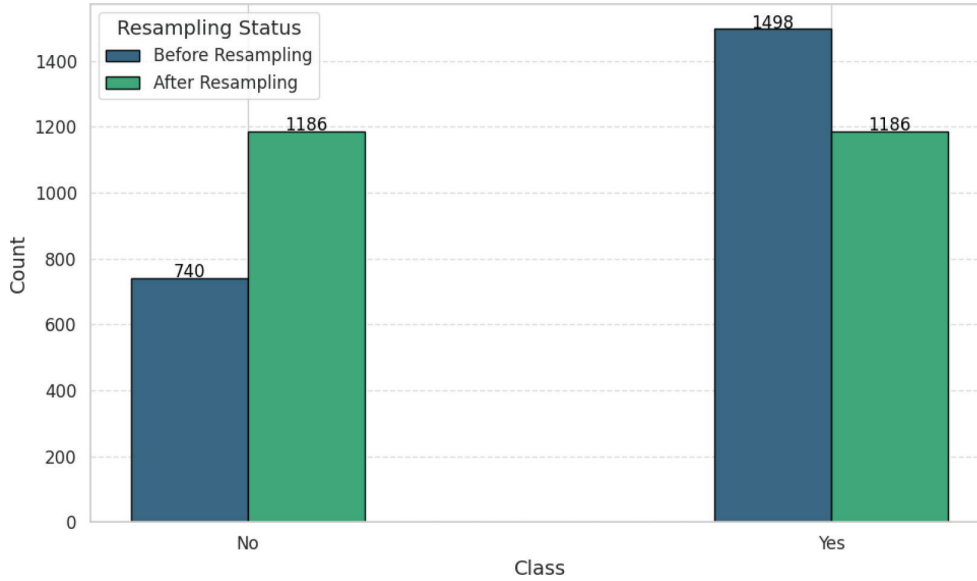


Figure 2. Class distribution before and after resampling using SMOTETomek

detect patterns in both classes, ensuring fairness, reducing bias, and enhancing the overall predictive performance. This step was critical for developing reliable insights, particularly in age-stratified mental health analysis.

4.2. Age-Stratified Results

The age-stratified analysis evaluated the predictive performance of models across different age groups, offering insights into demographic-specific mental health outcomes (see Table 2).

Table 2. Results of all age groups

Age Group 18 - 24					
Classifier	Accuracy	Precision	Recall	F1 Score	ROC AUC
RF	0.92	1.0	0.88	0.93	0.94
XGBoost	0.84	1.0	0.76	0.86	0.88
SVM	0.84	0.93	0.82	0.87	0.84
Age Group 25 - 34					
RF	0.91	0.91	0.95	0.93	0.90
XGBoost	0.93	0.93	0.96	0.94	0.92
SVM	0.88	0.87	0.96	0.91	0.86

Age Group 35 - 44					
RF	0.91	0.93	0.93	0.93	0.89
XGBoost	0.91	0.93	0.93	0.93	0.89
SVM	0.90	0.93	0.92	0.92	0.89
Age Group 45 - 54					
RF	0.95	0.94	1.0	0.97	0.90
XGBoost	0.97	0.97	1.0	0.98	0.95
SVM	0.91	0.94	0.94	0.94	0.88
Age Group 55+					
RF	0.71	0.75	0.90	0.82	0.58
XGBoost	0.71	0.75	0.90	0.95	0.58
SVM	0.85	0.83	1.00	0.90	0.75

For the *18-24 age group*, RF outperformed other models with an accuracy of 92 %, precision of 1.0, and F1 score of 93 %. It exhibited no false positives and only two false negatives, indicating high reliability in detecting positive cases. XGBoost and SVM achieved lower accuracies (84 %), with XGBoost showing a recall of 76 % (four false negatives). RF's robustness is particularly significant for this group, as young adults are at a critical stage for early mental health intervention. In the *25-34 age group*, XGBoost emerged as the top performer with an accuracy of 93 %, precision of 93 %, and F1 score of 94 %. RF followed closely with an accuracy of 91 %, while SVM lagged slightly (88 % accuracy) due to higher false positives (18). This group, well-represented in the dataset, is critical for mental health predictions due to its impact on productivity and societal outcomes.

For the *35-44 age group*, RF and XGBoost delivered identical performance, with accuracies of 91 %, F1 scores of 93 %, and balanced accuracy of 90 %. Both models showed seven false positives and seven false negatives, demonstrating consistent predictive ability. SVM also performed well, with an accuracy of 90 %. Often balancing professional and familial responsibilities, this group benefits from accurate mental health predictions for targeted interventions. The *45-54 age group* saw exceptional performance, with XGBoost achieving the highest accuracy (97 %), precision (97 %), and F1 score (98 %), with only one false positive and no false negatives. RF also performed strongly (95 % accuracy), while SVM trailed slightly (91 % accuracy, two false negatives). This group's openness to discussing mental health enhances prediction reliability.

In the *55+ age group*, smaller sample sizes introduced variability. RF and XGBoost achieved accuracies of 71 %, with RF showing a recall of 90 % but a low ROC AUC (0.58), indicating limited discriminatory power. SVM performed better with an accuracy of 85 % and a recall of 100 %, though its ROC AUC (0.75) remained moderate. These results highlight the challenges of small sample sizes in generalising predictions for this group.

Overall, RF and XGBoost consistently demonstrated strong performance across age groups, while SVM showed competitive but slightly lower results. The analysis underscores the importance of age stratification in identifying demographic-specific trends and tailoring mental health interventions. The subsections that follow explore the use of SHAP analysis for further model prediction interpretation and the identification of key predictive features.

4.3. SHapley Additive exPlanations Results

The SHAP analysis provides insights into features' importance and contributions to model predictions. The summary plot highlights global feature impacts, while the force plot explains individual predictions. These visualisations improve model transparency, revealing key drivers of mental health outcomes across age groups.

Age Group 18-24

The SHAP analysis identified key factors influencing mental health risk predictions for the 18-24 age group, a demographic often facing transitional challenges such as academic pressures, early career demands, and increased independence. This analysis aims to uncover actionable insights for targeted mental health interventions.

The SHAP summary plot revealed that «*Interference with daily activities due to lack of treatment*», «*Interference with daily activities despite treatment*», and «*Seeking treatment*» were the most influential features (see Figure 3). High levels of interference were strongly associated with elevated mental health risks, showing the burden of untreated or inadequately managed conditions. These findings suggest that disruptions in daily routines are critical indicators of mental health deterioration in this group. In addition, «*Current mental health disorder*» emerged as a consistently impactful predictor, reflecting the high prevalence of ongoing mental health challenges among young adults, likely exacerbated by social pressures and identity development.

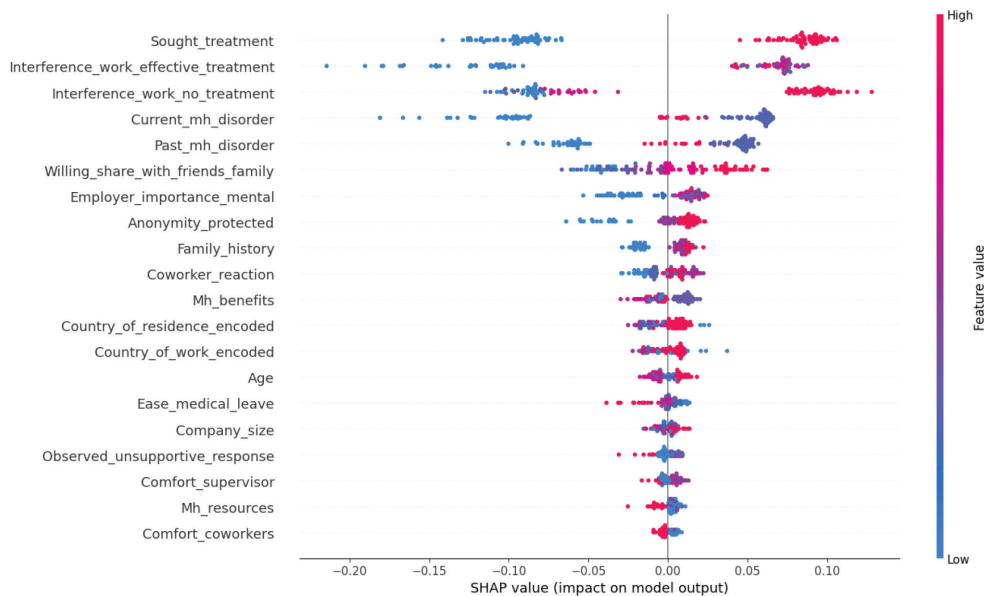


Figure 3. SHAP summary plot for age group 18-24

These results highlight the importance of early intervention and accessible support systems for this demographic. Reducing stigma around mental health and promoting open communication is essential

for encouraging help-seeking behaviour. The insights also provide a foundation for designing tailored interventions that address the unique needs of young adults, enhancing the effectiveness of mental health strategies. This analysis shows the value of XAI in advancing age-specific mental health research and interventions.

Age Group 25-34

The SHAP analysis for the 25-34 age group identified key factors influencing mental health risk predictions, revealing continuity and divergence from the 18-24 age group. The top contributors included «*Seeking treatment*», «*Interference with daily activities due to lack of treatment*», and «*Interference despite treatment*» (see Figure 4), highlighting the persistent burden of untreated or poorly managed mental health conditions and their disruptive impact on daily functioning. «*Current mental health disorder*» remained a significant predictor, reflecting the ongoing prevalence of mental health challenges.

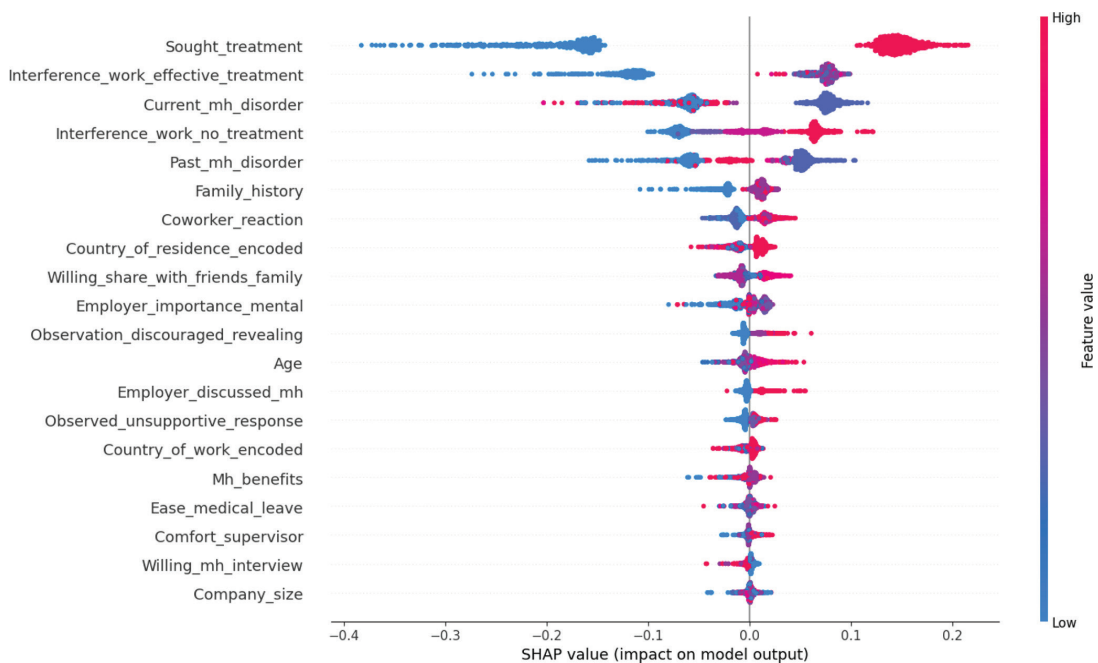


Figure 4. SHAP summary plot for age group 25-34

A notable divergence for this age group was the heightened importance of «*Family history of mental health disorders*», suggesting that individuals in their late twenties and early thirties may become more aware of or impacted by familial predispositions as they navigate increased responsibilities and life transitions. In addition, «*Willingness to share mental health concerns with friends or family*» persisted as a protective factor, though its relative importance diminished compared to the 18-24 group. This shift may reflect a growing reliance on professional support or self-directed coping mechanisms as individuals balance career development, personal relationships, and financial independence.

These findings show the evolving dynamics of mental health determinants as individuals transition into a phase marked by greater self-reliance and accountability. While treatment-related interference remains a consistent burden, familial and genetic factors gain prominence, highlighting the dual pressures faced by emerging professionals. By linking these insights to the younger age group, the analysis provides a more precise trajectory of how mental health risks evolve across life stages, reinforcing the need for nuanced, age-specific interventions. These results contribute to a deeper understanding of mental health determinants and offer a foundation for targeted prevention strategies tailored to the unique challenges of this demographic.

Age Group 35-44

The SHAP analysis for the 35-44 age group revealed a different evolution in mental health risk predictors, reflecting the unique pressures and responsibilities of this life stage. While some predictors remained consistent with younger age groups, notable differences emerged in the relative importance of certain features.

The key contributors included «*Seeking treatment*», «*Interference with daily activities due to lack of treatment*», and «*Current mental health disorder*» (see Figure 5), reaffirming the persistent burden of untreated or poorly managed mental health conditions. However, «*Past mental health disorder*» gained greater significance in this group compared to younger groups, suggesting that prior mental health experiences play an increasingly pivotal role as individuals age. «*Family history of mental health disorders*» also remained a significant predictor, reflecting its enduring influence on mental health outcomes.

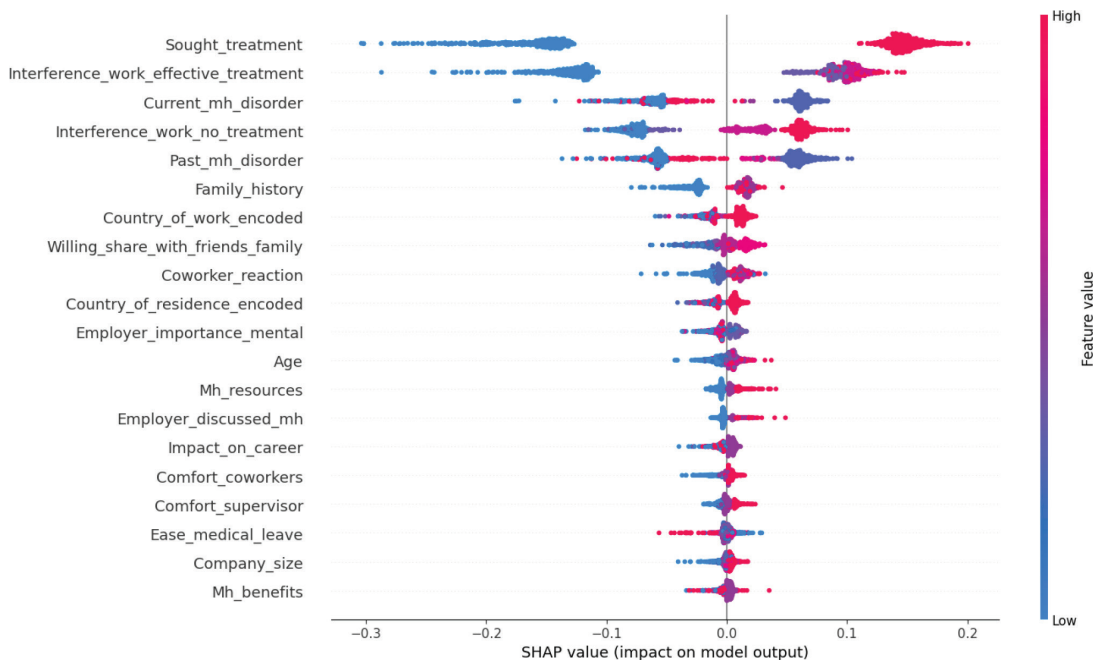


Figure 5. SHAP summary plot for age group 35-44

A notable divergence was the diminished role of «*Willingness to share concerns with friends or family*» as a protective factor, which was more prominent in younger age groups. This shift may indicate a transition toward self-reliance or greater reliance on professional resources, likely driven by heightened professional demands and family caregiving responsibilities that limit opportunities for informal support systems.

These findings highlight the increasing emphasis on historical and genetic factors in this age group, coupled with a relative decline in the influence of social support. While treatment-related disruptions remain a consistent challenge, the results underscore the importance of interventions that address both current and long-term mental health risks tailored to the complex realities of this demographic. By identifying these evolving predictors, the analysis reinforces the need for age-specific mental health strategies, offering actionable insights for targeted interventions that align with the unique challenges faced by individuals aged 35-44.

Age Group 45-54

The SHAP analysis for the 45-54 age group revealed evolving patterns in mental health risk predictors, reflecting the unique challenges of midlife. Consistent with younger groups, features such as «*Seeking treatment*», «*Interference with daily activities due to lack of treatment*», and «*Current mental health disorder*» remained among the top contributors (see Figure 6), showing the persistent impact of untreated or poorly managed mental health conditions across age groups.

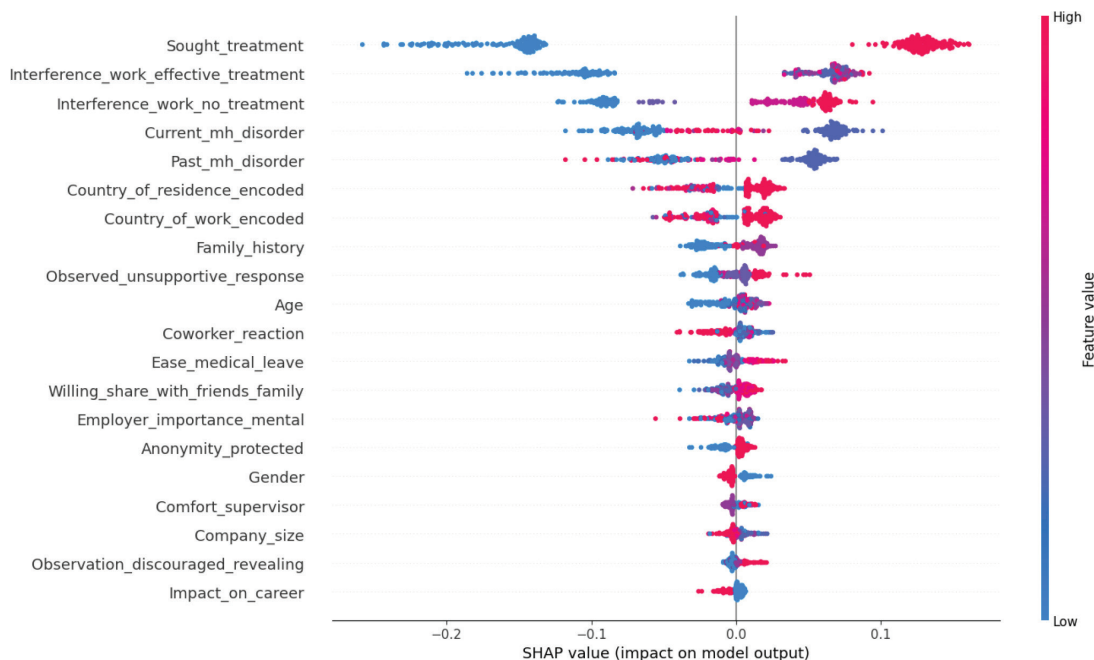


Figure 6. SHAP summary plot for age group 45-54

However, notable differences emerged in the relative importance of specific predictors. While «*Past mental health disorder*» and «*Family history of mental health disorders*» continued to play critical roles, their influence slightly diminished compared to the 35-44 group. This may reflect a reduced direct impact of familial and historical factors as individuals approach later life stages and shift focus on personal well-being. Conversely, «*Country of residence*» and «*Country of work*» gained prominence, suggesting that environmental and contextual factors, such as access to mental health resources and societal attitudes, play an increasingly significant role in this demographic.

These findings highlight a shift in the balance of mental health risk factors for the 45-54 age group. While treatment-related disruptions remain a dominant challenge, the growing influence of environmental and contextual variables shows the importance of addressing external factors, such as societal support systems and resource availability. This shift may be tied to midlife transitions, including caregiving responsibilities, health management, and preparation for later life.

The results emphasise the need for tailored interventions that address both shared and evolving risk factors, providing context-sensitive support for individuals navigating the complexities of midlife mental health challenges. These insights offer a foundation for designing targeted strategies that align with the unique needs of this demographic.

Age Group 55+

The SHAP analysis for the 55+ age group revealed distinct shifts in mental health risk predictors, reflecting the unique challenges of later adulthood. Consistent with earlier age groups, «*Seeking treatment*», «*Interference with daily activities due to lack of treatment*», and «*Current mental health disorder*» remained critical predictors, underscoring the ongoing impact of untreated or persistent mental health conditions.

However, notable differences emerged in feature importance. While «*Past mental health disorder*» maintained a strong presence, its influence slightly diminished compared to the 45-54 group, suggesting a shift toward more immediate factors. Notably, «*Willingness to share mental health concerns with friends or family*» regained prominence as a protective factor, reversing the downward trend observed in younger groups. This may reflect increased reliance on personal support networks as individuals retire or reduce work responsibilities. In addition, «*Observation of discouraging behaviour from others*» emerged as a significant predictor, highlighting this group's sensitivity to social cues and interpersonal dynamics. Contextual variables, such as «*Country of residence*» and «*Country of work*», remained influential, reflecting the role of societal attitudes and resource availability in shaping mental health outcomes (see Figure 7).

These findings indicate a shift in mental health predictors for the 55+ group, with greater emphasis on personal and social support systems alongside enduring challenges related to treatment and daily functioning. The results show the need for interventions that address both immediate mental health needs and the evolving significance of social connections. By capturing these dynamics, the analysis enhances understanding of age-specific mental health challenges and informs targeted strategies to support this population.

4.4. Uncovering Lifespan-Specific Drivers Using Explainable AI

This study's age-stratified SHAP analysis provided a comprehensive understanding of the predictors influencing mental health risks across different life stages. By stratifying by age, the study recognises that age-specific experiences, priorities, and stressors shape mental health. Using XAI techniques

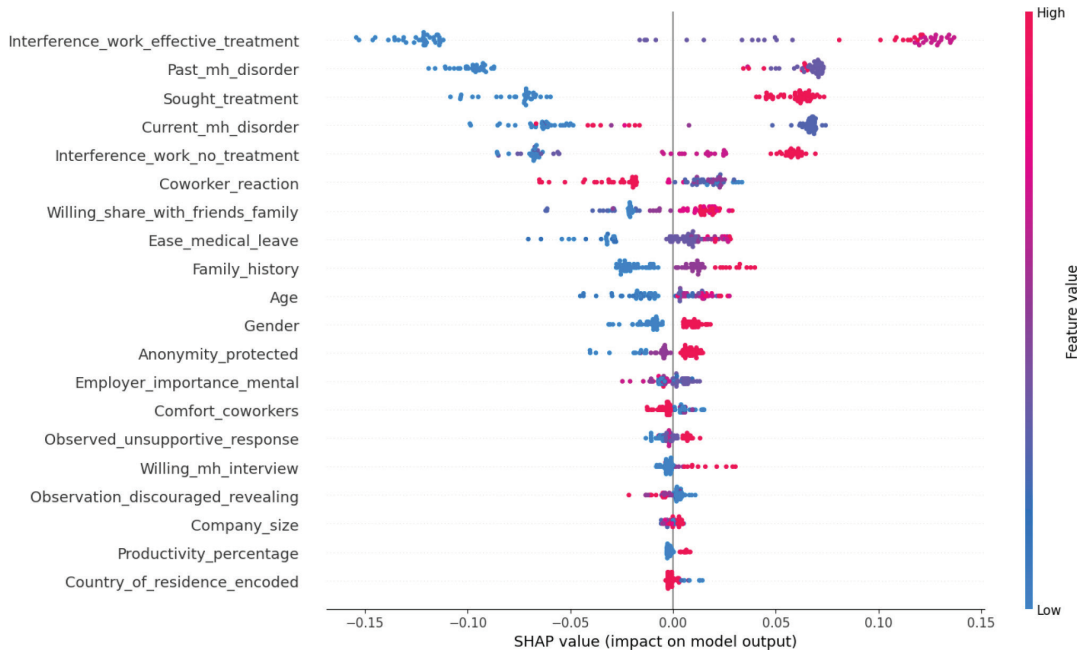


Figure 7. SHAP summary plot for age group 55+

enhances the interpretability of mental health prediction models and uncovers evolving risk patterns and protective factors unique to each age group.

The results highlight consistent predictors, such as «*Interference with daily activities due to lack of treatment*», «*Current mental health disorder*», and «*Seeking treatment*», showing the persistent burden of untreated or poorly managed mental health conditions across all age groups. However, significant differences emerged: social support was pivotal for younger adults (18-24), familial and historical factors gained prominence in middle-aged groups (25-54), and reliance on social support resurged in later adulthood (55+), reflecting shifting coping mechanisms as individuals transition into retirement or reduced work responsibilities. In addition, contextual factors such as «*Country of residence*» and «*Country of work*» became increasingly influential in older groups, emphasising the role of societal and environmental conditions in shaping mental health outcomes.

These findings have implications, demonstrating that mental health risks are dynamic and evolve with age. This necessitates targeted interventions that address both shared and age-specific challenges. The study establishes age-stratified analysis as a critical methodological approach by revealing these different dynamics. Traditional, one-size-fits-all models often fail to capture the distinct needs of different age groups, limiting their effectiveness. In contrast, this study highlights how age stratification enhances the precision of mental health predictions and informs tailored prevention and intervention strategies.

The novelty of this work lies in its integration of XAI with age-specific mental health research, offering both methodological and practical contributions. Methodologically, it demonstrates the utility of SHAP in uncovering age-specific patterns and providing interpretable insights into complex ML models. Practically, the findings offer a roadmap for designing interventions that align with individuals' evolving mental health needs across their lifespan.

Using XAI to move beyond generic risk models, this study enables age-sensitive approaches that account for the unique drivers of mental health risks at different life stages. Combining robust ML techniques with age stratification, this study advances the field of mental health prediction, offering actionable insights for researchers, policymakers, and practitioners. This contribution fills a critical gap in the literature, setting a benchmark for future research at the intersection of AI, mental health, and demographic specificity and paving the way for more effective, globally relevant mental health strategies.

5. Implications of the Study

5.1. Theoretical Implications

The study makes a significant theoretical contribution by demonstrating the effectiveness of XAI techniques, specifically SHAP, in identifying age-stratified mental health predictors. Moving beyond traditional, age-agnostic mental health models, this study introduces a framework that acknowledges the dynamic interplay between age and mental health risk factors. It demonstrates that mental health predictors are not static but differ significantly across life stages, necessitating tailored approaches to prediction and intervention. This advancement emphasises the importance of age as a critical variable in mental health research, bridging the gap in the existing literature where age-specific differences have been underexplored. Furthermore, the integration of age stratification with XAI improves the interpretability of ML models, paving the way for more inclusive and actionable theoretical models in AI-driven mental health research.

5.2. Practical Implications

The study has far-reaching practical implications, providing actionable insights for mental health practitioners, policymakers, and researchers. The age-stratified findings provide a roadmap for developing targeted interventions that align with the unique challenges faced at different life stages, from early adulthood to pre-retirement. For example, the increased role of social support in younger adults emphasises the need for community-based initiatives. At the same time, the prominence of familial and historical factors in middle-aged groups suggests the importance of long-term mental health monitoring. In later adulthood, the resurgence of reliance on personal networks points to the value of promoting social connectivity.

In addition, the study shows how XAI can bridge the gap between predictive insights and real-world applications by increasing transparency and trust. This is especially critical for deploying AI tools in sensitive contexts such as mental health care, where stakeholders demand interpretable and ethical solutions. Policymakers can use these findings to allocate resources more effectively, addressing specific age-related needs and reducing the global burden of untreated mental health conditions. Overall, this work advances the theoretical understanding of age-specific mental health risks and provides a practical foundation for developing more effective, context-sensitive interventions.

6. Limitations and Future Work

While the study provides significant insights into age-stratified mental health risk prediction, it has limitations. The study used the OSMI dataset, which, although diverse and comprehensive, was

initially designed for the technology sector, potentially limiting the generalisability of the findings to other industries or broader societal contexts. Furthermore, the sample sizes for specific age groups, particularly the 55+ category, were relatively small, introducing prediction variability and potentially reducing the strength of insights for these groups. While the SHAP methodology provides powerful interpretability, the analysis may be influenced by the inherent biases present in the self-reported nature of the dataset, including underreporting or misrepresentation of mental health conditions.

Future work could build on this study by incorporating datasets from diverse sectors and geographic regions to improve generalisability and provide a more comprehensive understanding of age-stratified mental health risks. Addressing sample size limitations, particularly for underrepresented age groups, would also strengthen the reliability of findings. In addition, integrating other forms of data, such as clinical records or social media activity, could complement the self-reported survey data, offering a richer and more nuanced perspective on mental health predictors. Finally, developing dynamic models that adapt to evolving mental health trends across different demographics could further refine the applicability of XAI in mental health interventions.

7. Conclusion

The study represents a pioneering effort to integrate XAI techniques with age-stratified analysis for mental health prediction, addressing a critical gap at the intersection of AI and mental health research. By using SHapley Additive Explanations (SHAP), the study identifies age-specific predictors of mental health risks, offering a novel framework that enhances both the interpretability and applicability of ML models in this domain.

The findings reveal that mental health predictors are dynamic and vary significantly across different life stages. Social support networks are pivotal for younger adults (18-24), while familial and historical mental health factors gain prominence in middle-aged groups (25-54). For older adults (55+), personal and social connections reemerge as critical protective factors. These different insights show the importance of tailoring mental health interventions to the unique needs of each demographic group, moving beyond the one-size-fits-all approaches often employed in AI models.

Beyond predictive insights, this study highlights the value of XAI in sensitive mental health domains. By translating complex model outputs into interpretable insights, SHAP bridges the gap between data science and practical application, fostering trust and accountability in AI-driven interventions. Policymakers and practitioners can leverage these findings to design targeted, age-appropriate strategies that address the global burden of mental health disorders, particularly in resource-constrained settings such as LMICs.

While the study has limitations, including using a technology-sector-focused dataset and sample size disparities, it lays a strong foundation for future research. Expanding to more diverse datasets, integrating longitudinal data, and refining dynamic AI models will enhance the generalisability and impact of age-stratified mental health predictions.

In conclusion, this study advances the theoretical understanding of how mental health predictors evolve with age and provides a practical roadmap for using XAI to improve mental health outcomes. Demonstrating the power of age stratification in predictive modelling sets a new benchmark for AI's responsible and effective application in mental health. This study inspires future research to explore the transformative potential of AI-driven, context-sensitive solutions in this critical domain, contributing to more personalised and effective mental health care globally.

References

- Balogun, A. O., Lafenwa-Balogun, F. B., Mojeed, H. A., Adeyemo, V. E., Akande, O. N., Akintola, A. G., Bajeh, A. O., & Usman-Hamza, F. E. (2020). *SMOTE-Based Homogeneous Ensemble Methods for Software Defect Prediction* (pp. 615-631). https://doi.org/10.1007/978-3-030-58817-5_45
- Bowers, A. J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, 24(1), 20-46. <https://doi.org/10.1080/10824669.2018.1523734>
- Byeon, H. (2023). Advances in Machine Learning and Explainable Artificial Intelligence for Depression Prediction. *International Journal of Advanced Computer Science and Applications*, 14(6). <https://doi.org/10.14569/IJACSA.2023.0140656>
- Ding, H., Li, N., Li, L., Xu, Z., & Xia, W. (2025). Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study. *Journal of Affective Disorders*, 368, 537-546. <https://doi.org/10.1016/j.jad.2024.09.111>
- Domínguez-Olmedo, J. L., Gragera-Martínez, Á., Mata, J., & Pachón, V. (2022). Age-Stratified Analysis of COVID-19 Outcome Using Machine Learning Predictive Models. *Healthcare*, 10(10), 2027. <https://doi.org/10.3390/healthcare10102027>
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*, 55(9), 1-33. <https://doi.org/10.1145/3561048>
- Fergus, P., & Chalmers, C. (2022). *Performance Evaluation Metrics* (pp. 115-138). https://doi.org/10.1007/978-3-031-04420-5_5
- González-Nóvoa, J. A., Busto, L., Rodríguez-Andina, J. J., Fariña, J., Segura, M., Gómez, V., Vila, D., & Veiga, C. (2021). Using Explainable Machine Learning to Improve Intensive Care Unit Alarm Systems. *Sensors*, 21(21), 7125. <https://doi.org/10.3390/s21217125>
- Ikram, M., Shaikh, N. F., Vishwanatha, J. K., & Sambamoorthi, U. (2022). Leading Predictors of COVID-19-Related Poor Mental Health in Adult Asian Indians: An Application of Extreme Gradient Boosting and Shapley Additive Explanations. *International Journal of Environmental Research and Public Health*, 20(1). <https://doi.org/10.3390/ijerph20010775>
- Islam, M. T., Ashraf, K., Hosen, Md. H., Nawar, S., & Asgar, S. (2024). Predictive Modeling of Anxiety Levels in Bangladeshi University Students: A Voting-Based Approach with LIME and SHAP Explanations. *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (ICACCESS)*, 01-06. <https://doi.org/10.1109/iCACCESS61735.2024.10499576>
- Javed, A., Lee, C., Zakaria, H., Buenaventura, R. D., Cetkovich-Bakmas, M., Duailibi, K., Ng, B., Ramy, H., Saha, G., Arifeen, S., Elorza, P. M., Ratnasingham, P., & Azeem, M. W. (2021). Reducing the stigma of mental health disorders with a focus on low -and middle- income countries. *Asian Journal of Psychiatry*, 58, 102601. <https://doi.org/10.1016/j.ajp.2021.102601>
- Kabir, M. M., Hafiz, M. S., Bandyopadhyay, S., Jim, J. R., & Mridha, M. F. (2024). Tea leaf age quality: Age-stratified tea leaf quality classification dataset. *Data in Brief*, 54, 110462. <https://doi.org/10.1016/j.dib.2024.110462>

- Kerz, E., Zanwar, S., Qiao, Y., & Wiechmann, D. (2023). Toward explainable AI (XAI) for mental health detection based on language behavior. *Frontiers in Psychiatry, 14*. <https://doi.org/10.3389/fpsyt.2023.1219479>
- Khan, A. E., Hasan, M. J., Anjum, H., Mohammed, N., & Momen, S. (2024). Predicting life satisfaction using machine learning and explainable AI. *Heliyon, 10*(10), e31158. <https://doi.org/10.1016/j.heliyon.2024.e31158>
- Kusuma, K., Larsen, M., Quiroz, J. C., & Torok, M. (2024). Age-stratified predictions of suicide attempts using machine learning in middle and late adolescence. *Journal of Affective Disorders, 365*, 126-133. <https://doi.org/10.1016/j.jad.2024.08.043>
- Lee, H., Operario, D., Yi, H., Choo, S., & Kim, S.-S. (2019). Internalized Homophobia, Depressive Symptoms, and Suicidal Ideation Among Lesbian, Gay, and Bisexual Adults in South Korea: An Age-Stratified Analysis. *LGBT Health, 6*(8), 393-399. <https://doi.org/10.1089/lgbt.2019.0108>
- Leng, Q., Guo, J., Tao, J., Meng, X., & Wang, C. (2024). OBMI: oversampling borderline minority instances by a two-stage Tomek link-finding procedure for class imbalance problem. *Complex & Intelligent Systems, 10*(4), 4775-4792. <https://doi.org/10.1007/s40747-024-01399-y>
- Letoffe, O., Huang, X., Asher, N., & Marques-Silva, J. (2024). *From SHAP Scores to Feature Importance Scores*. <http://arxiv.org/abs/2405.11766>
- Li, M., Sun, H., Huang, Y., & Chen, H. (2024). Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems, 4*(1), 2. <https://doi.org/10.1007/s43684-023-00060-8>
- Mbuya, E., Mokheleli, T., Bokaba, T., & Ndayizigamiye, P. (2023). *A Multiclass Approach to Predicting Diabetes Using Machine Learning*. <https://aisel.aisnet.org/acis2023/140>
- Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V., & Hussain, A. (2023). A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. *Healthcare Analytics, 3*, 100185. <https://doi.org/10.1016/j.health.2023.100185>
- Mokheleli, T., Bokaba, T., & Museba, T. (2023). *An In-Depth Comparative Analysis of Machine Learning Techniques for Addressing Class Imbalance in Mental Health Prediction*. <https://aisel.aisnet.org/acis2023/15>
- Mokheleli, T., Bokaba, T., Museba, T., & Ntshingila, N. (2024). *A Machine Learning Approach to Mental Disorder Prediction: Handling the Missing Data Challenge* (pp. 93-106). https://doi.org/10.1007/978-3-031-63999-9_6
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Support Vector Machines and Support Vector Regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 337-378). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_9
- Nadkarni, A., Hanlon, C., & Patel, V. (2024). Mental Health Care Models in Low- and Middle-Income Countries. In *Tasman's Psychiatry* (pp. 3347-3393). Springer International Publishing. https://doi.org/10.1007/978-3-030-51366-5_156
- Naslund, J. A., & Deng, D. (2021). Addressing mental health stigma in low-income and middle-income countries: A new frontier for digital mental health. *Ethics, Medicine and Public Health, 19*, 100719. <https://doi.org/10.1016/j.jemep.2021.100719>
- OSMI. (2023). *About OSMI*. Open Sourcing Mental Health. <https://osmihelp.org/about/about-osmi.html>

- Páez, A., & Boisjoly, G. (2022). *Exploratory Data Analysis* (pp. 25-64). https://doi.org/10.1007/978-3-031-20719-8_2
- Pinto, M. F., Batista, J., Leal, A., Lopes, F., Oliveira, A., Dourado, A., Abuhaiba, S. I., Sales, F., Martins, P., & Teixeira, C. A. (2023). The goal of explaining black boxes in EEG seizure prediction is not to explain models' decisions. *Epilepsia Open*, 8(2), 285-297. <https://doi.org/10.1002/epi4.12748>
- Pirjatullah, Kartini, D., Nugraha, D. T., Muliadi, & Farmadi, A. (2021). Hyperparameter Tuning using GridsearchCV on The Comparison of The Activation Function of The ELM Method to The Classification of Pneumonia in Toddlers. *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 390-395. <https://doi.org/10.1109/IC2IE53219.2021.9649207>
- Ratantja Kusumajati, F., Rahmat, B., & Junaedi, A. (2024). IMPLEMENTATION OF BALANCING DATA METHOD USING SMOTETOMEK IN DIABETES CLASSIFICATION USING XGBOOST. *Jurnal Ilmiah Kursor*, 12(4), 201-212. <https://doi.org/10.21107/kursor.v12i4.410>
- Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, 572, 522-542. <https://doi.org/10.1016/j.ins.2021.05.055>
- Shiraly, R., Jazayeri, S. A., Seifaei, A., Jeihooni, A. K., & Griffiths, M. D. (2024). Suicidal thoughts and behaviors among untreated illicit substance users: a population-based study. *Harm Reduction Journal*, 21(1), 96. <https://doi.org/10.1186/s12954-024-01015-9>
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549. <https://doi.org/10.1016/j.eswa.2023.121549>
- Talukder, Md. A., Sharmin, S., Uddin, M. A., Islam, M. M., & Aryal, S. (2024). MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs. *International Journal of Information Security*, 23(3), 2139-2158. <https://doi.org/10.1007/s10207-024-00833-z>
- Wang, C., Feng, L., & Qi, Y. (2021). Explainable deep learning predictions for illness risk of mental disorders in Nanjing, China. *Environmental Research*, 202, 111740. <https://doi.org/10.1016/j.envres.2021.111740>
- World Health Organisation. (2024, October 10). *Mental health of adolescents*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>
- Zhang, J., Feng, X., Wang, W., Liu, S., Zhang, Q., Wu, D., & Liu, Q. (2024). Predicting the Risk of Loneliness in Children and Adolescents: A Machine Learning Study. *Behavioral Sciences*, 14(10), 947. <https://doi.org/10.3390/bs14100947>