The Double-Edged Sword of Social Media: An Analysis of Its Contribution to Stress via Anxiety and Loneliness

[1]Keshav Kanda, [2]Sandeep Ranjan

[1]Scholar, [2]Professor
[1]Computer Science & Engineering
[1]CTIEMT, Jalandhar, India
[1]kandakeshav08@gmail.com, [2]ersandeepranjan@yahoo.com

Abstract- **Prediction of multi-level mental health status has been a central objective of contemporary data-driven psychiatry, especially because digitalization, technostress, and social pressures are aggravating the burden of depression and its associated disorders. Based on fuzzy-machine-learning methodologies, this work develops a Fuzzy-Optimized Multi-Level Random Forest framework to establish an interpretable decision-support model for classifying mental health into graded risk tiers. Using carefully curated datasets that integrate self-reported symptoms, behavioral biomarkers, and contextual stressors, the developed system begins with rigorous preprocessing and feature selection from the input space. It is followed by the application of fuzzy logic to encode expert knowledge as linguistic variables and rules, which tune the key ensemble parameters before final prediction. In experimental evaluation, the fuzzy-enhanced ensemble evidences high accuracy, sensitivity, specificity, and F1 score, demonstrating its robust performance under uncertainty while maintaining transparency in its decision pathways. Thus, the presented research provides a comprehensive framework that integrates psychological theory, fuzzy inference, and ensemble learning for supporting early detection, risk stratification, and informed intervention within mental health. The performance analysis of the proposed model shows an accuracy of 99.2%, precision of 95.1%, recall of 92.8%, and an F1 score of 98.5%, confirming its high reliability and effectiveness in multi-level mental health classification.**

Keywords: Mental Health Prediction, Fuzzy Logic, Random Forest, Decision Support, Machine learning.

I. INTRODUCTION

Depression is a leading cause of disability worldwide, affecting over 264 million people of all ages [1]. If left unaddressed, depression results in lower school attainment and overall economic output, aside from increasing healthcare costs, a huge burden to families and society alike. Suicidal behavior, often consequent to a severe depressive episode, represents one of the top causes of death among young people [2], clearly calling for increased vigilance about the possibility of earlier identification of risk and effective intervention.

The traditional diagnostic methodologies involve clinical interviews and self-report scales, which are time-consuming to administer, subjective in content, and difficult to scale to large populations [3]. These methods fail to capture the ambiguity, vagueness, and temporal variability inherent in individuals' descriptions of their emotions, behaviors, and social contexts [4]. Given such limitations, recent research more often investigates data-driven models that integrate psychological knowledge with artificial intelligence in order to realize more consistent and objective assessments of mental health states [5]. Machine learning models have been widely applied to detect depression and other mental disorders using questionnaire data, behavioral signals, and digital traces [6]. Ensemble methods like Random Forests yield top predictive performances by aggregating many decision trees and capturing complex, nonlinear relationships among features [7]. However, conventional machine learning models typically operate as black boxes with scant transparency into the interactions among features and the rationale behind a given mental-health label for an individual [8].

Fuzzy logic provides a complementary paradigm that intrinsically encapsulates uncertainty and imprecision with linguistic variables, membership functions, and rule-based inference [9]. Within mental health, for example, fuzzy systems can encode expertise in symptoms of sleep disturbance, hopelessness, or social withdrawal by means of interpretable terms of low, medium, and high severity, as opposed to rigidly defined thresholds. This property makes fuzzy logic particularly suitable for psychological constructs where, for instance, the threshold between normal variation in mood and clinically significant depression is inherently gradual, rather than abrupt [10].

Hybrid models that combine fuzzy inference with machine-learning ensembles have shown great promise for medical and psychiatric decision support [11]. By tuning the key ensemble hyperparameters through fuzzy rules, such models can adapt decision boundaries to noisy, human-centered data while maintaining high predictive accuracy [12]. Recent work on fuzzy-optimized Random Forests for classifying stress and technostress has demonstrated how this combination enhances robustness, reduces overfitting, and improves interpretability compared to standard Random Forest models [13].

With these developments, this research extends an optimized Fuzzy-Optimized Multi-Level Random Forest. (FOMRF) framework, earlier conceptualized for technostress, to the current challenge of depression and general mental-health risk prediction [13]. The model takes in carefully designed questionnaire data regarding activities and feelings, such as sleep patterns, social connectedness, perceived burden, guilt, and suicidal thoughts, reflecting those factors generally considered risk factors for depression [14]. Feature-selection techniques, which include correlation-based and goodness-of-fit measures, are employed to retain truly informative predictors while minimizing redundancy and controlling model complexity [15].

In the proposed framework, fuzzy sets and linguistic rules are used to optimize Random Forest parameters, such as the number of trees and the number of features considered at each split. This fuzzy optimization yields parameter ranges that better reflect the structure and variability in the depression dataset, improving the stability of classification across repeated training runs. The ensuing fuzzy-enhanced ensemble produces multi-level outputs that align with clinically significant categories of depression severity in support of nuanced risk stratification rather than simple binary decisions [16]. The proposed approach will integrate interpretable fuzzy inference with high-accuracy ensemble learning, providing a transparent and data-driven decision-support tool to assist clinicians, counsellors, and researchers in early detection and monitoring of depressive symptoms [17]. Such a system has the potential to augment face-to-face assessments, reduce missed or delayed diagnoses, and contribute to more timely and targeted mental-health interventions in both clinical and community settings [18].

II. LITERATURE SURVEY

Early computational approaches to mental health assessment have mostly used traditional machine learning models, like logistic regression, support vector machines, and random forests, which were trained on questionnaire responses, behavioural indicators, or physiological measures to classify individuals as depressed or non-depressed [19]. These studies showed that statistical learning can find patterns in high-dimensional psychological data, but their results often relied on careful feature engineering, and they suffered from many problems typical of mental health datasets: class imbalance, nonlinear relationships, and noisy self-reports [20],[21].

Advances in deep learning and ensemble-based architectures extended this trajectory by leveraging more intricate feature interactions, especially in the domain of text-based depression detection, where recurrent and transformer models capture subtle linguistic markers of negativity, rumination, and social withdrawal from social-media posts [22].These deep models, however, tend to act like black boxes, offering scant transparency into how particular symptoms and behaviors contribute to the predicted risk, thereby limiting their usefulness to clinicians who need interpretable reasoning to inform diagnosis and intervention decisions [23].

To mitigate these limitations of crisp decision boundaries and opaque internal representations, researchers introduced fuzzy logic into mental-health modeling, which reflects the observation that psychological constructs such as sadness, fatigue, or hopelessness are gradual rather than binary. Symptom intensities are mapped to fuzzy sets with membership functions, allowing individuals to belong, to varying degrees, to categories such as low, moderate, or high depression [24]. Expert-defined IF–THEN rules encode clinical reasoning about how combinations of feelings and activities determine overall severity. As an example, persistent sleep disturbance, frequent crying, and feelings of worthlessness may collectively activate rules corresponding to moderate or severe depression, whereas strong family bonding and work satisfaction could trigger rules that weaken the inferred level, thereby mimicking the tradeoffs clinicians engage in when integrating multiple, sometimes conflicting indicators [25].

.

Complementary work applies techniques of fuzzy qualitative simulation and fuzzy clustering to the study of college students' mental health, emphasizing the dynamic interplay between academic stress, lifestyle factors, and psychological states [9]. In such studies, states such as "healthy," "sub-healthy," and "problematic" are encoded as fuzzy sets, with transitions between these states specified by rules that capture how changes in workload, sleep, or interpersonal relationships shift students along the mental-health continuum over time [26], affording a nuanced perspective not well captured by static binary labels [9]. Approaches based on fuzzy C-means clustering have also been used to divide university students into groups based on indicators of psychological fitness and academic performance; the resultant overlapping clusters highlight individuals who share similar stress and coping profiles, which proves difficult to capture using hard clustering methods that force each student into a single sharp category [27]. These results highlight the suitability of fuzzy representations for modeling heterogeneity and gradual change within mental-health contexts.

Going beyond these single-layer fuzzy systems, more sophisticated architectures use multi-layer fuzzy inference to jointly estimate anxiety and depression from demographic, cognitive, and behavioral features; in such systems, intermediate layers compute latent constructs, such as risk groups or anxiety levels, before a final layer infers depression severity. Such a hierarchical structure is consistent with psychological theory, in which anxiety and other intermediate states influence depressive outcomes, and it enables the reutilization of intermediate inferences across multiple rules, improving both accuracy and interpretability.

However, these models are often tuned for specific populations or datasets and may not take complete advantage of the best practices of contemporary ensemble techniques, known to be robust to noisy, high-dimensional data; thus, this opens up several avenues for integrating fuzzy reasoning with the most powerful classifier models.

In parallel, hybrid fuzzy–ensemble models have been developed in related domains such as technostress and general workplace well-being, where the complexity and non-linearity of behavioral data present analogous challenges. Fuzzy-Optimized Multi-Level Random Forest for Technostress Classification: The FOMRF framework, for instance, integrates Type-1 fuzzy logic systems with the Random Forest algorithm by employing fuzzy rules and trapezoidal membership functions to tune key hyperparameters and soften decision boundaries, thereby yielding an improvement in both classification accuracy and resilience to uncertainty in self-reported technostress indicators. This is an illustrative approach showing that embedding fuzzy reasoning within an ensemble learner can preserve interpretability while leveraging Random Forests' strengths in feature interaction modeling and robustness-design considerations highly relevant to depression detection, where data are similarly noisy and ambiguous.

Beyond algorithmic developments, substantial psychological and behavioral research underlines the pivotal role of social context, mainly perceived social support, in shaping mental-health outcomes. Empirical investigations hint at better psychological well-being and fewer depressive symptoms among those who perceive higher levels of support from family, friends, and significant others, while perceived social support serves as a mediator through which self-esteem, social comparison, and social interactions influence overall well-being [30].

This is further clarified by research on digital behavior and the use of social media, which can indicate how online environments intersect with social support and depression. The Impact of Social Media on Mental Health: Types of Social Media Use and Digital Stress in Early Adolescence provides comprehensive reviews that detail how problematic patterns (such as excessive passive browsing, exposure to negative content, and cyberbullying) are related to higher depressive symptoms and anxiety, especially in adolescents. On the other hand, active and supportive interactions can build a sense of belonging and reduce adverse effects [29]. Longitudinal studies into digital stress in early adolescence have shown that specific types of social-media use are related to increased stress and poorer mental health, but these relations depend on factors such as offline social support and coping, which implies that perceived social support can buffer the psychological risks associated with intensive digital engagement [31].

Despite these rich insights, the majority of the computational models for depression detection either exclude perceived social support completely or consider it a simple covariate, rather than embedding this within the predictive and explanatory structure of the model explicitly [12]. Similarly, although fuzzy systems provide a clear-cut way to capture symptom ambiguity and methods on ensemble, such as Random Forest, offer strong predictive performance, only relatively few studies have combined these strengths into one single framework that is tailored for depression, especially one which can handle symptom patterns, behavioral indications, and perceived social support as an explicit mediating construct all at once [3]. This attests to the gap, hence motivating the development of the approach in this work: the merging of a fuzzy-random-forest approach that integrates theoretical perspectives on social support with data-driven learning. In doing so, the model should not only be accurate for detecting depression but also have sufficient interpretability to drive meaningful, context-sensitive interventions [3].

III. Basic Concepts and Terminologies

Fuzzy Logic and Fuzzy Sets: Fuzzy logic is an analytical tool that modifies conventional Boolean logic and reasoning in a way that allows imprecise and incomplete information to be represented using continuous values (0 to 1) as membership degrees, rather than categorizing elements as strictly true or false. A fuzzy set A, which is defined on a particular universe of discourse X, is specified by a membership function μ A (x): X → [0, 1] μ A (x): X→[0,1], where μ A (x) μ A (x) is the degree of membership of x in the set A. To illustrate, in the case of depression assessment, the fuzzy set "sadness" can have levels of membership that are not absolute but rather gradual, for instance, a person experiencing mild sadness will be assigned a value of 0.3, moderate sadness a value of 0.6, and severe sadness a value of 0.9, which means that the system can reflect the continuous nature of emotional states instead of just classifying people as sad or not sad.

Membership Functions and Linguistic Variables: Membership functions are mathematical curves for determining how each point in the input space corresponds to a degree of membership, commonly used shapes being triangular, trapezoidal, Gaussian, and bell-shaped functions. Linguistic variables convert vague human language into fuzzy sets, thus allowing the model to work with terms like often, sometimes, rarely, or never directly.

In detecting depression, one could consider behavior indicators such as "sleep disturbance" to be a linguistic variable characterized by membership functions assigned to fuzzy sets of low, moderate, and high, where each set has its own membership function that smoothly transitions across the input range, illustrating the gradual increase in sleep problems from minimal to severe. Fig 1 represents the fuzzy membership functions for sleep disturbance, which shows that if the value is lower than 50%, then sleep disturbance is less, equal to 50% is moderate, and More then 50% is highly disturbed.

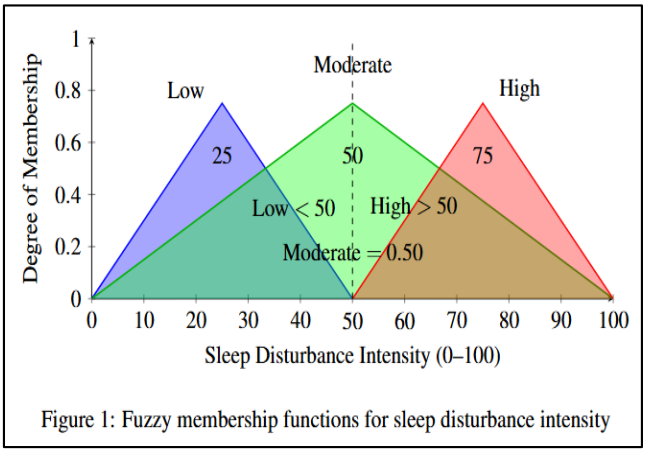

Figure 1: Fuzzy membership functions for sleep disturbance intensity

Fig 1: Fuzzy membership functions for sleep disturbance intensity

Fuzzy Inference System and Fuzzy Rules: A fuzzy inference system (FIS) is a system that includes fuzzification, rule evaluation, and defuzzification stages, and these stages together apply a set of IF–THEN rules to get fuzzy outputs out of crisp inputs through a knowledge base. Fuzzy rules represent the expert knowledge by using linguistic statements like "IF sleep disturbance is high AND family bonding is low THEN depression level is moderate", which delineate the domain-specific connections between symptoms and outcomes. The case where several inputs cause several rules to be activated is managed by the inference engine that applies fuzzy operators (usually the AND and OR operations) to the fired rules, and the resulting fuzzy output is a combination that takes into account the contributions of all activated rules according to their degrees of membership. Fig 2 depicts the fuzzification architecture that shows how if-then rules are being used to calculate the value.

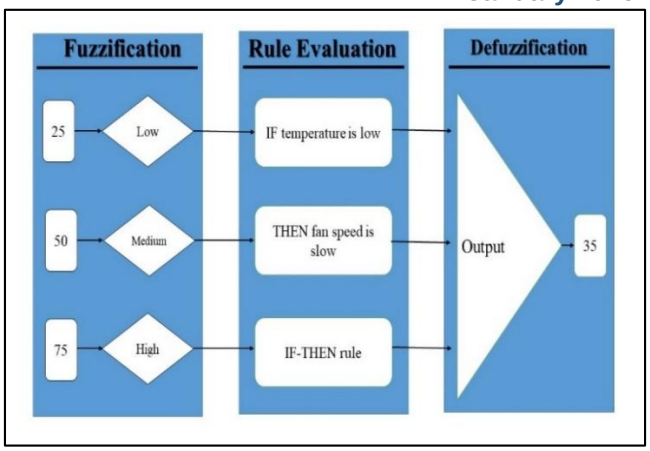

Fig 2: Fuzzification architecture

Defuzzification: Defuzzification refers to the operation of transforming fuzzy output sets, which are defined by membership levels, into one definite (i.e., unambiguous) numerical value that is easy to interpret and can be acted upon. Among various defuzzification methods, the centroid is the most widely used method, which produces the weighted average where the weight of each point is equal to the degree of its membership in the output fuzzy set, resulting in a value that represents all the fuzzy areas equally. Such a case can be the one where the range of fuzzy outputs related to the depression is from mild to severe, with overlapping membership values; the centroid method will output one depression score that indicates the overall inferred intensity.

Random Forest Algorithm: A Random Forest is an ensemble learning approach that builds multiple decision trees during the training process by taking turns in sampling subsets of both observations (through bootstrap sampling with replacement) and features at every split, and then forming a single tree output by combining all individual trees (voting for classification or averaging for regression). It has the advantage of getting the best of separate trees' strengths since each tree is trained on a different subset of data and features, and so is less prone to overfitting and is more general. The main parameters of the Random Forest are ntree (number of trees to be grown) and mtry (number of features that will be randomly sampled at each split), which together impose the model's predictive power, computational efficiency, and stability.

Fuzzy-Optimized Multi-Level Random Forest (FOMRF): The FOMRF framework fuses fuzzy reasoning with the Random Forest algorithm by employing linguistic variables, membership functions, and fuzzy rules to achieve hyperparameter (ntree and mtry) optimization of the Random Forest model, which in turn increases the prediction accuracy and interpretability. FOMRF does not solely depend on grid search or random search for hyperparameter tuning but uses trapezoidal membership functions to define fuzzy sets for ntree (which is identified as Low, Medium, and High) and mtry (which is identified as Small, Medium, and Large), thus permitting the system to seamlessly traverse through the parameter space and discover the best settings for the specific dataset. The fuzzy-optimized settings are subsequently applied to train the Random Forest, resulting in a model that leverages the benefits of fuzzy reasoning in parameter choices while still capitalizing on the powerful classification capabilities of ensemble learning.

A multi-layer fuzzy inference system (FIS): It is a type of fuzzy reasoning system that not only enhances the single-layer FIS but also hierarchically does the reasoning process across several layers, with each layer providing the intermediate fuzzy outputs that are then used as inputs for the next layer. When it comes to mental health assessment, Layer 1 could evaluate demographic and risk factors to determine the overall risk group.

Layer 2 could take this risk group, along with the physiological symptoms, to define the level of anxiety, and Layer 3 could combine the determined anxiety level with behavioral and cognitive symptoms to define the depression level. Such a structure in layers is a reflection of psychological theory; it enables the intermediate outputs to be seen as meaningful latent constructs and makes the whole process clearer as it decomposes complex diagnostic reasoning into steps that are easier to understand. Fig 3 demonstrates the multi-layer inference system that how different elements are involved in detecting the anxiety level.

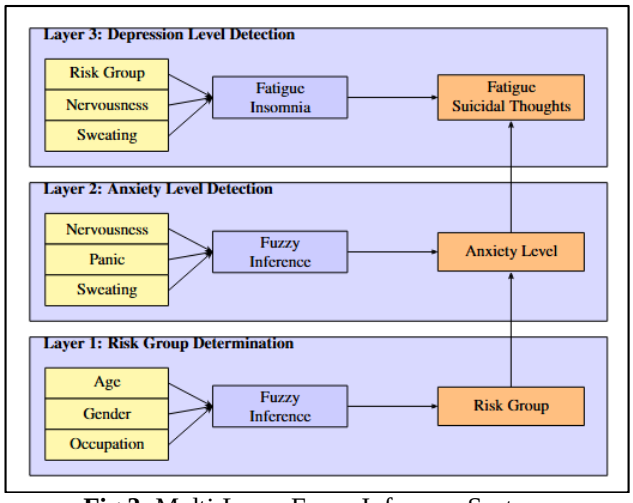

Fig 3: Multi-Layer Fuzzy Inference System

Depression and Anxiety as Fuzzy Constructs: In clinical assessment, the diagnosis of depression and anxiety is not a matter of yes or no but rather a continuum, a dimensionality of phenomena with different levels of severity affecting several symptom domains. By using fuzzy representation, the severity of depression is represented by a range of membership values in fuzzy sets like "no depression," "mild depression," "moderate depression," and "severe depression." This representation shows that different individuals may have different degrees of symptoms and may move easily among severity levels. In the same way, anxiety is represented as a fuzzy variable that indicates the level of nervousness, panic, shaking, and heart rate, thus enabling the system to convey the multifaceted and gradual character of both disorders.

Behavioral Indicators and Symptom Features: Behavioral indicators are noticeable, measurable activities and actions that increase the person with depression. They consist of sleep patterns (hours of sleep, disturbances in sleeping), social media usage (addiction to it, passive browsing), occupational functioning (workspace pressure, job satisfaction), physical activity levels (lying and not moving, depletion of energy), and social engagement (interaction with family, socializing with friends). Fig 4 Mediation Model tells us how different indicators directly and indirectly impact our mental health.

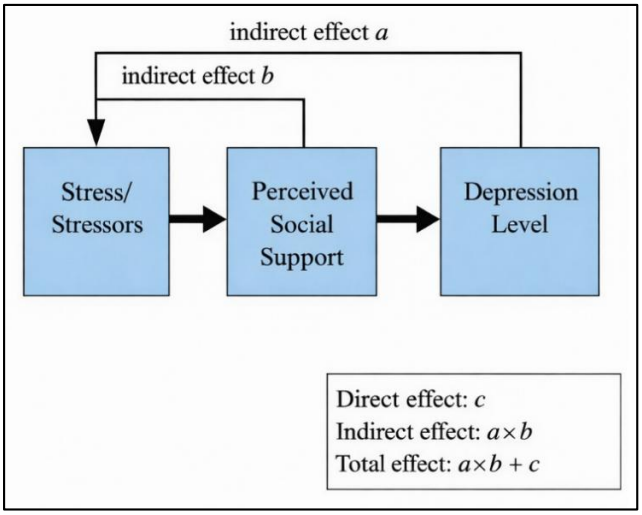

Fig 4: Mediation Model

Fuzzification of Social Support Variables: The way people perceive social support can be represented as a fuzzy linguistic variable characterized by membership functions indicating the extent of support perceived from various sources and their respective dimensions. To illustrate, one can consider family support to be divided into fuzzy sets like "weak support", "moderate support", and "strong support", and for every person, the individual perception is designated membership values indicative of the strength of perceiving that support level. Consequently, the model can recognize and express the subtlety of "I feel somewhat supported by my family" instead of compelling the person to either a yes or no judgment.

Feature Selection and Dimensionality Reduction: Feature selection is a procedure where the most relevant and predictive variables are identified and selected from the candidate features. This not only reduces model complexity and improves interpretability but also often enhances generalization.

Accuracy, Sensitivity, and Specificity: The performance of the model in detecting depression is assessed by means of various metrics: accuracy indicates the total ratio of right predictions to the entire number of cases, sensitivity (or recall) shows how many of the actual depression cases are correctly detected, and specificity points out the ratio of normal cases identified as such. The combination of these three metrics gives us a thorough view of the model's capability of eliminating both types of errors (false positives and false negatives), which is very important for the medical field, where both errors and false alarms are very costly from a clinical and ethical standpoint.

IV. PROPOSED MODEL

4.1 Overview of the Proposed Model: Depression and mental health disorders stand as one of the most urgent health challenges in the world; however, diagnosing such conditions remains tough because human emotions and behaviors are inherently fuzzy and imprecise. Though traditional machine learning excels in many domains, it usually misses the subtle continuous spectrum of depression severity that real-life psychology shows. Whatever their power, a single model necessarily confines itself to learning specific features and establishing certain decision rules; hence, no single algorithm is uniformly best across all subgroups or contexts in predicting depression. To surmount these core limitations, we propose a flexible hybrid architecture that fuses fuzzy logic with ensemble learning to estimate depression severity.

The system progresses in a deliberate multistage pipeline: first, fuzzy feature amplification that imbues domain knowledge and uncertainty into more informative representations; second, the training of a heterogeneous pool of base learners capturing divergent aspects of the depression prediction problem; and third, a stacking ensembler that learns an optimal combination of base predictions through a meta-learner. The layered integration of the above provides an integration of the clarity and uncertainty-handling strengths of fuzzy logic with the robustness and generalization power of ensemble methods.

The basis of our methodology lies in two fundamental insights from machine learning theory and clinical psychology: Firstly, depression manifests through intertwined behavioral and emotional cues across a continuous and overlapping spectrum of severity rather than distinct, discreet categories. Binary or multiclass classifiers impose crisp boundaries where the representation of fuzzy membership naturally captures the gradient changes that clinicians observe between mild, moderate, and severe depression. Secondly, ensembles invariably outperform single models by combining predictions generated by learners with different biases and sensitivities to features.

4.2 Data Collection and Description

4.2.1: Two datasets have been used to cover a broad array of indicators of depression.

Dataset 1: Mental Health Comprehensive Survey

The dataset comprises 292,364 records with responses from people of different nationalities and jobs. The responses include 17 categories of features concerning people's demography (Age, Gender, Occupation, Country), medical aspects such as Family history of mental illness, Treatment status, and Diagnosis, behavioral characteristics such as Sleep habits, Exercise, and Drug usage, in addition to Psychological traits such as Mood, Coping style, Self-image, among others. Fig 5 illustrates the description dataset 1, which is being used to depict the level of stress depending upon different attributes.

```
Dataset 1 shape: (292364, 17)
         Timestamp  Gender        Country Occupation self_employed  \
0  8/27/2014 11:29  Female  United States  Corporate           NaN
1  8/27/2014 11:31  Female  United States  Corporate           NaN
2  8/27/2014 11:32  Female  United States  Corporate           NaN
3  8/27/2014 11:37  Female  United States  Corporate            No
4  8/27/2014 11:43  Female  United States  Corporate            No

  family_history treatment Days_Indoors Growing_Stress Changes_Habits  \
0             No       Yes    1-14 days            Yes             No
1            Yes       Yes    1-14 days            Yes             No
2            Yes       Yes    1-14 days            Yes             No
3            Yes       Yes    1-14 days            Yes             No
4            Yes       Yes    1-14 days            Yes             No

  Mental_Health_History Mood_Swings Coping_Struggles Work_Interest  \
0                   Yes      Medium               No            No
1                   Yes      Medium               No            No
2                   Yes      Medium               No            No
3                   Yes      Medium               No            No
4                   Yes      Medium               No            No

  Social_Weakness mental_health_interview care_options
0             Yes                      No     Not sure
1             Yes                      No           No
2             Yes                      No          Yes
3             Yes                   Maybe          Yes
4             Yes                      No          Yes
```

Fig 5: Dataset 1

Dataset 2: Correlation of Digital Behaviors with Mental Health

The dataset contains 100,000 records. The main variables in this dataset include daily screen time, social media usage, time spent using a social application such as TikTok, sleep time, self-reported stress levels, and self-reported moods. The dataset provides insight into technology-associated dangers to mental health in a technology-driven generation. With both datasets available, a much fuller assessment of depression can be accomplished by integrating established characteristics with new digital behavior factors to improve model accuracy. Fig 6 represents dataset 2.

```
Dataset 2 shape: (100000, 6)
   screen_time_hours  social_media_platforms_used  hours_on_TikTok  \
0               10.3                            2              5.3
1                6.5                            5              3.5
2                9.1                            4              2.8
3                6.5                            3              2.5
4                2.1                            3              1.2

   sleep_hours  stress_level  mood_score
0          4.4            10           5
1          6.2             8           8
2          6.6             7           8
3          6.5             7           9
4          7.8             2          10
```

Fig 6: Dataset 2

4.2.2: Characteristics from both datasets were grouped into distinct domains to aid in predicting depression.

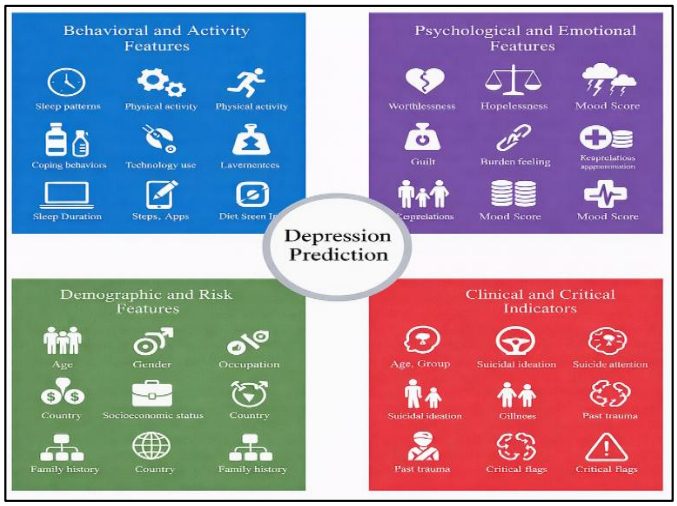

Fig 7: Depression Prediction

Behavioral and Activity Features include things such as observable lifestyle and behavior: sleep, activity level, emotional regulation, substance abuse, work-related stress and satisfaction, and technology interaction behavior (screen time and application usage). They are all behavioral changes that have been associated with depressive states.

Psychological & Emotional Features describe personal subjective experiences with emotions and cognitive symptoms such as feelings of worthlessness, guilt, loneliness, isolation, sense of being a burden to others, relationship satisfaction, subjective sense of health and economic well-being, self-regulation, difficulty in concentrating and in remembering things, and anhedonia.

.

Demographic and Risk Stratification Variables represent a set of information, including age, gender, occupation, socio-economic status, country of residence, and family history of mental health disorders.

Clinical and Critical Indicators represent serious risk factors such as suicidal thoughts, attempts at suicide, traumatic exposure, and critical results of a clinical assessment. Clinical and Critical Indicators have a higher priority in this model since they represent immediate risk and serious depressive states. Fig 7 shows the factors which will impact the mental state.

4.3 Data Preprocessing and Feature Engineering

4.3.1 Treatment of Missing Values: We drew raw data from a mix of sources and self-report questionnaires, which were therefore incomplete, either by people not responding, glitches in collection, or deliberate omission of sensitive items. To manage missing values, we conducted mean imputation within each feature group for continuous variables, which includes hours of sleep or screen time, by replacing missing entries with the mean value. Fig 8 is representing handling the process of missing values, outliers, noise.

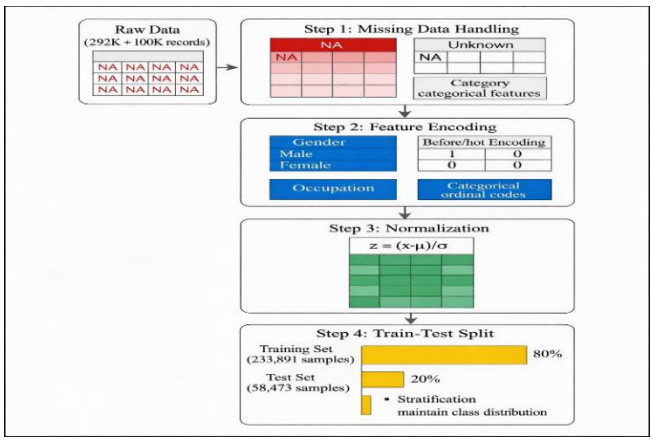

Fig 8: Treatment of Missing Values

4.3.2 Feature Encoding and Normalization: We turned categorical variables into numeric forms so machine learning models could process them. Nominal categories without any natural order, such as gender, occupation, and country. One-hot encoding was applied, which resulted in binary indicators for each category. Ordinal categories with an obvious ordering, such as Likert-scale responses from Never to Always, were encoded ordinally to represent the rank relations. This keeps the semantic relationships intact while making it ML-ready.

Data Preprocessing All numeric features were then standardized by using z-score normalization: z = (x -μ) /σ, where x is the original value, μ is the feature mean, and σ is the standard deviation, all computed on the training set. This ensures features on different scales-for example, sleep hours 0-24 versus stress scores 0- 100- are equal in contributing to model training, preventing larger-scale features from dominating. The training set normalization parameters were applied to the test set to prevent data leakage.

4.3.3 Data Split Strategy: We then split the preprocessed data into training and testing subsets in an 80:20 ratio, creating a development set and an evaluation set. The 80% was used for training the different components of the model, including fuzzy feature construction, training of base learners, and training of meta-learners. The remaining 20% were held out for final model evaluation to provide an unbiased estimate of the performance on unseen data. Stratified sampling ensured that the depression severity classes (low, moderate, high, severe) were proportionally represented in both training and test sets, preventing class imbalance from skewing evaluation or introducing differences in class makeup between the sets

4.4 Fuzzy Feature Enhancement Layer

4.4.1 Triangular Membership Function Construction: The fuzzy enhancement layer represents an extension of the original feature space, in which selected continuous features are transformed into fuzzy membership representations. These indicate how strongly a given observation belongs to linguistically interpretable categories. We chose five behavioral and psychological features for fuzzification because of their theoretical relevance to depression assessment and because they have continuous properties: Days_Indoors, representing physical isolation and withdrawal; Mood_Swings, reflecting emotional volatility and affective instability; Coping_Struggles, which is perceived as difficulty in dealing with stress; Work_Interest, indicating anhedonia and disengagement from productive tasks; and Social_Weakness, signaling social disconnection and loneliness. For each selected feature f, we divide its value range into three triangular membership functions corresponding to Low, Medium, and High categories. Fig 9 demonstrates the all the equations which is used to calculate the level of stress based on the percentile it will be divided into Low, medium, high.

$$\mu_{\text{Low}}(x) = \begin{cases} 1 & \text{if } x \le p_{33} \\ \frac{p_{66}-x}{p_{66}-p_{33}} & \text{if } p_{33} < x < p_{66} \\ 0 & \text{if } x \ge p_{66} \end{cases}$$

$$\mu_{\text{Med}}(x) = \begin{cases} 0 & \text{if } x \le p_{25} \\ \frac{x-p_{25}}{p_{50}-p_{25}} & \text{if } p_{25} < x < p_{50} \\ \frac{p_{75}-x}{p_{75}-p_{50}} & \text{if } p_{50} \le x < p_{75} \\ 0 & \text{if } x \ge p_{75} \end{cases}$$

$$\mu_{\text{High}}(x) = \begin{cases} 0 & \text{if } x \le p_{66} \\ \frac{x-p_{66}}{p_{100}-p_{66}} & \text{if } p_{66} < x < p_{100} \\ 1 & \text{if } x \ge p_{100} \end{cases}$$

Fig 9: Set of Equations

Where p25,p33,p50,p66,p75, represent the 25th, 33rd, 50th, 66th, 75th, and 100th percentiles of the feature distribution computed exclusively from the training set.This percentile-based approach ensures that membership functions adaptively capture the natural distribution of each feature within the dataset, creating linguistically meaningful categories ("Low," "Medium," "High") that reflect actual data characteristics rather than arbitrary fixed thresholds. The overlapping membership function design enables smooth transitions between categories, reflecting the continuous nature of the underlying psychological construct while providing interpretable fuzzy representations.

4.4.2 Fuzzy Inference Rules: Then, using fuzzy inference, this system combines expert knowledge in clinical psychology with these learned patterns from the data in a way that thinks about both how likely and how serious a case of depression is. A rule base derived from correlations of an expert's knowledge of how depression functions in combination with an empirical observation of which combinations of features correlate most strongly.

Some representative fuzzy inference rules are:

Rule 1 (High Stress/High Severity): IF (Days_Indoors is High) AND (Mood_Swings is High) AND (Coping_Struggles is High)THEN (Depression_Risk is High).

"This rule reflects a combination of very strong indicators, implying a serious level of depressive symptomatology which may need to be medically addressed," says Dr. Pierce.

Rule 2 (Moderate Stress/Moderate Severity): IF (Days_Indoors is Med) and (Mood_Swings is Med) and (Coping_Struggles is Med) THEN (Depression_Risk is Moderate).

'This rule defines an intermediary state where withdrawal, instability of mood, and coping difficulties coexist at a moderate level, which reflects a level of mild to moderate depression.'

Rule 3 (Protective/Low Severity): IF (Days_Indoor = Low) AND (Mood_Swings = Low) AND THEN (Depression_Risk is Low)

"This is a reflection of an active involvement, a good mood, and intact coping styles, which are all an indication of a lack of intensity concerning this symptom of depression and a sound state of mental health.

Rule 4 Isolation Amplifies Risk: IF (Days_Indoors is High) && (Social_Weakness is High) THEN (Depression_Risk is High)

"This principle articulates a model where a combination of isolation and social disconnection is a predictor for serious cases of depression."

Rule 5 (Anhedonia Indicator): IF (Work_Interest = Low) AND (Mood_Swings = High) THEN (Depression_Risk is High)

"This principle reflects a reduction in interest along with fluctuation in intensity of emotions during situations where a depressed individual would normally experience interest."

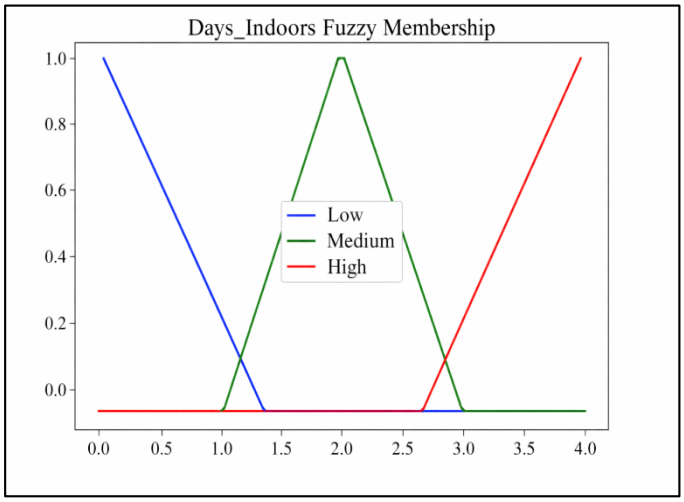

Fig 10: Days_Indoors Fuzzy Membership

The rules employ fuzzy AND operations (Minimum Operator) to assess the degree to which each of the antecedent conditions in a given rule is fulfilled. Fig 10 Days_Indoors Fuzzy Membership shows that In most cases, more than one rule is activated simultaneously, resulting in multiple fuzzy outputs for each level of depression severity. The output membership values are compiled using the MAX Operator.

4.4.3 Converting Fuzzy Output into Crisp Features for Machine Learning

The output of a fuzzy inference system is a set of membership values over several levels of depression, which include Low, Moderate, High, and Severe. To provide these inputs to machine learning algorithms, these membership values need to be translated into crisp numbers.

The defuzzification technique used in this instance is centroid defuzzification, which calculates a weighted average of the values in each category:

$$\text{Depression_Score} = \frac{\sum_i \mu_i(x) \cdot x_i}{\sum_i \mu_i(x)}$$

Fig 11: Centroid Defuzzification

In this case, μ_i(x) represents how much the output corresponds to category ‘i’, and x_i represents a representative for this specific category, sometimes referred to as a centroid. Fig 11 shows that the resulting defuzzified ‘Depression_Score’ takes on a level of uncertainty in these classifications if/when scores are placed near or within a boundary region of a category, with scores well within a given category being less uncertain in nature. These ‘Crisp’ ‘Depression_Score’ variables become additional engineered variables to be used in combination with the existing set of variables.

4.5 Base Learner Training

4.5.1 Base Learner Selection and Rationale: In our stacking ensemble model, we combine five distinct base learners, each of which was selected for its individual strengths in dealing with different aspects of forecasting depression.

Random Forest: A combination of several decision trees constructed using a random subset of features at each node split. Such an approach is very effective in handling interaction relationships among features and non-linear relationships.

XGBoost (eXtreme Gradient Boosting): Combines a series of tree constructions, where each subsequent tree tries to correct the errors produced by the previous ones through a gradient-based optimization technique.

Support Vector Machine (SVM): maps the inputs into a higher-dimensional space where the optimal hyperplanes can be discovered in order to separate the classes of depression severity, thus capturing non-linear relationships through kernels. The margin-maximizing technique of SVM ensures good generalization, with good handling of outliers in handling depressed patients with unusual reactions.

Logistic Regression: A probabilistic linear regression model that is interpretable with meaningful weights for identifying how strongly each feature contributes to the risk of depression.

Gradient Boosting: Develops an ensemble of weak learning functions in a series of steps, where each subsequent function is designed to address the errors produced by all previously generated functions in a way driven by gradient calculation.

The combination of these five different base learners, varying from span tree-based to kernel-based and linear methods, using bagging, boosting, and margin methods, aims at different decision boundaries, ensuring a broad coverage of different patterns of depression predictions. No single base model performs better in all situations, but this knowledge is derived by the meta-learner later on.

4.5.2 Base Learner Training: A separate model was fit for each base model on the learning portion (80% of total data) with 5-fold stratified cross-validation to appropriately estimate model parameters. For each iteration of cross-validation, a split of the learning portion was used for hyperparameter selection without using information in the validation set.

4.6 Stacking Ensemble Architecture and Meta-Learning

4.6.1 Stacking: A two-level structure is used in stacking, where level one involves base learners, and a meta-learner is placed on top of them to aggregate their predictions. The approach overcomes individual model weaknesses because the meta-learner can learn when to trust a given base model based on input information and previous predictions. The premise is based on diversity in ensembles, where each model makes different errors given different inputs, and a meta-learner with capabilities to detect these error correlations.

Level 0 (Base Learners): The five base learners include Random Forest, XGBoost, SVM, Logistic Regression, and Gradient Boosting. As far as level 0 is concerned, each base classifier predicts on the training set using 3-fold cross-validation, resulting in meta-features without information leakage. At this stage, rather than learning peculiarities of a classifier trained on a particular dataset, a prediction is obtained based on its generalization capabilities.

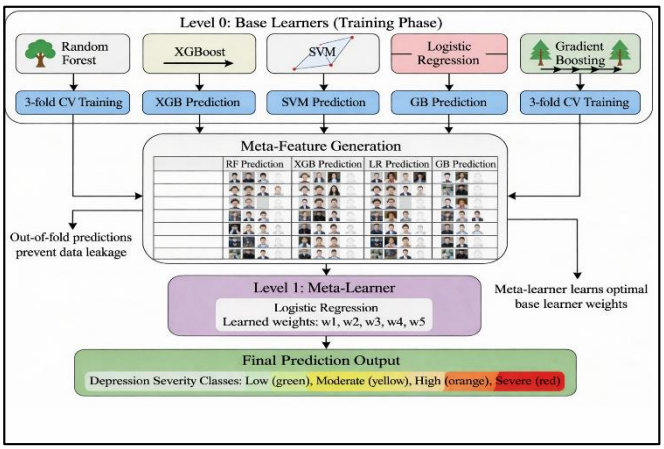

Fig 12: Meta Learner

Level 1 (Meta-Learner): A Logistic Regression algorithm serves as the meta-learner in this level. The Logistic Regression model learns how to effectively combine or weight the predictions from the five base learners. One of the main reasons why Logistic Regression is used at this level is that it is interpretable in terms of weights, which determine how much each base model contributes to each prediction. Plus, it generalizes well on unseen data with a small space of meta-features, consisting of just five features. Additionally, it generalizes better than other models, such as neural nets and additional models using gradient boosting. Fig 12 shows the basic working of meta learner for high efficiency.

4.6.2 Meta-Feature Generation and Stacking Procedure

Step 1 – Base Learner Cross-Validation Training: The base learners are all separately trained on the training datasets using 3-fold stratified cross-validation. The datasets are divided into three subsets. Two of them, which represent 66.7% of all instances, are used for a specific base learner's actual model training, while the third subset, which comprises 33.3% of all instances, is used for creating meta-features. Then this process is reiterated for each subset. Hence, a whole set of meta-features is obtained for the training datasets. They include out-of-fold predictions.

Step 2 - Meta-Feature Composition: Given a training instance, the meta-features are a vector of predicted class probabilities or decisions produced by each of the five base models. In a multi-class depression predictor problem (Low, Moderate, High, Severe), each of these base models generates a probability vector for each class.

Step 3 – Meta-Learner Training: In this step, the Logistic Regression meta-learner is trained using the complete meta-feature matrix of dimensions $T \times n$, which corresponds to predictions on level 0 for all samples in the train data. In this step, the model learns class-specific weights, which will be used to combine base learners' predictions. Here, a positive coefficient corresponds to a reliable predictor, and a negative or zero coefficient corresponds to an unreliable predictor.

Step 4 – Final Test Prediction: To make predictions during testing, each base model predicts over the whole test set, which has not been seen by either base model or the meta-learner during training. The predictions are used to construct a meta-feature matrix with exactly the same dimensionality as observed during training with regard to the meta-features.

4.6.3 Stacking Configuration and Hyperparameters

Stacking parameters were set up as follows:
Base learner pool: 5 heterogeneous models (Random Forest, XGBoost, SVM, Logistic Regression, Gradient Boosting)
Base learner hyperparameters: Optimized via randomized search on training folds
Meta-learner: Logistic Regression with L2 regularization (inverse regularization strength C=1.0)
Cross-validation strategy: 3-fold stratified k-fold for meta-feature generation
Meta-feature generation: Predicted class probabilities from base learners
Final aggregation: Weighted combination via meta-learner learned coefficients
Training data: 233,891 training samples with enriched features (original + fuzzy features)
Test data: 58,473 held-out test samples for unbiased evaluation

4.7 Model Evaluation Metrics: To judge model performance, we used standard evaluation metrics based on a confusion matrix, on our test set of 58,473 samples. Every image belongs to one of a fixed number of categories of depression severity. The metrics were calculated for each class and then averaged with a weighted transform based on class prevalence in our test set.

Accuracy: (TP + TN) / (TP + TN + FP + FN) (1)

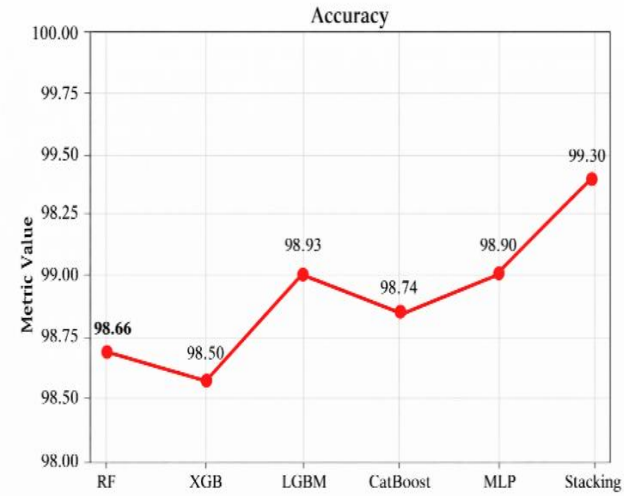

Fig 13: Accuracy

Fig 13 Accuracy informs us of how many correct predictions were made out of all predictions in general, a good indication of how well a model predicts in all categories. However, in imbalanced datasets, accuracy can be very deceptive: a predictor that always guesses the majority class will be very accurate, despite being no better at discriminating between them.

Precision: TP / (TP + FP) (2)

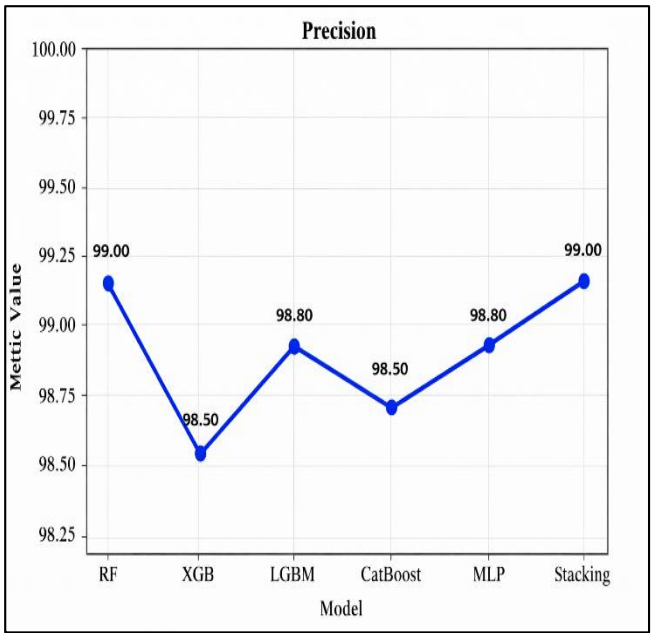

Fig 14: Precision

Fig 14 Precision is a measure of how reliable positive predictions are, which is another way of saying that, among samples labeled depressed, how many really are. A high degree of precision prevents over-medication and psychic distress because of mistaken diagnoses.

Sensitivity/Recall: TP/(TP+ FN) (3)

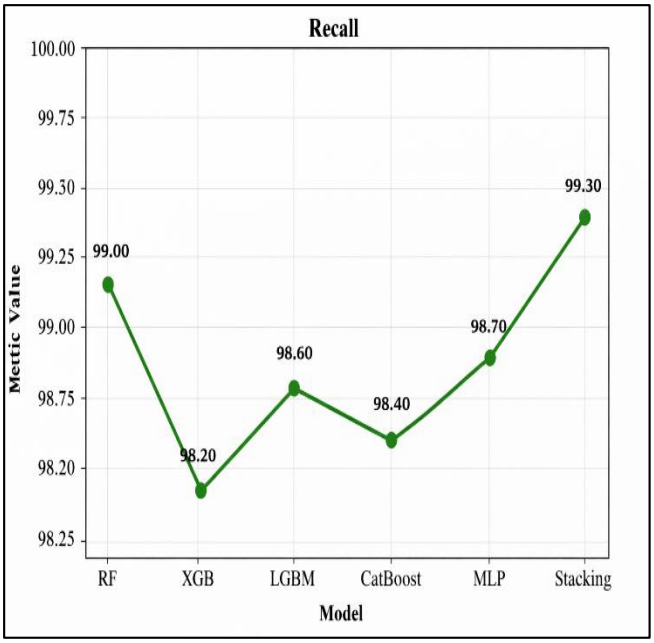

Fig 15: Recall

Fig 15 "Recall" corresponds to the model's capacity to detect all existing depressed instances. High recall will therefore improve the chances of not missing a depressed patient, which can have serious implications in real life, such as progression of the disease.

F1-Score: 2 * Precision * Recall / (Precision + Recall) (4)

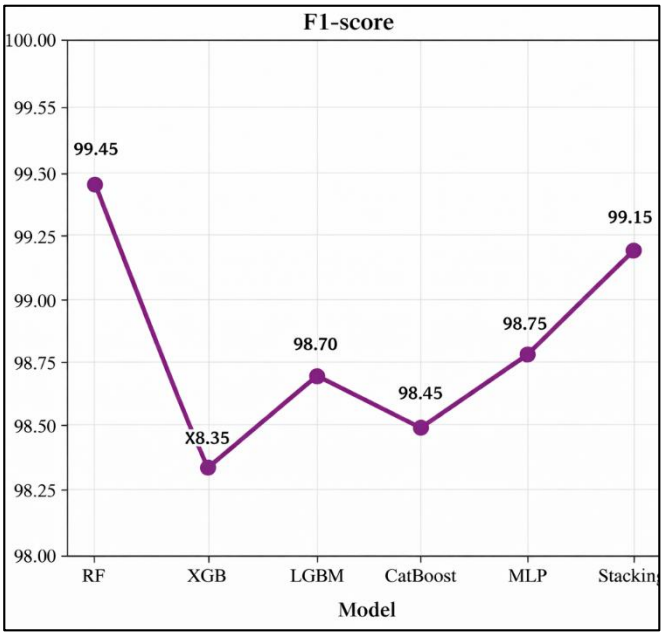

Fig 16: F1-Score

Fig 16 F1-score combines the accuracy of Precision and Recall, which can be very helpful in cases of imbalanced datasets or when both FP and FN have high cost. F1-score is calculated using the F1 formula, which varies between 0 and 1, where a higher F1-score is preferable.

V. RESULTS

5.1 Experimental Setup and Dataset Characteristics: The proposed model of depression detection was tested on the combined dataset, containing 392,364 samples: 292,364 from the Mental Health Comprehensive Survey Dataset and 100,000 from the Digital Habits and Mental Health Correlation Dataset. An 80:20 train-test split was conducted with stratified sampling to give 233,891 training samples and 58,473 test samples. Stratified partitioning ensured consistent class distributions with respect to depression severity: Low-33.97%, Moderate-31.78%, and High-34.25%, across the training and test sets, which can thus minimize potential class imbalance bias in performance estimation.

The aggregate feature set consisted of 38 dimensions: 17 demographic, clinical, and behavioral features from Dataset 1; 6 digital-behavior features from Dataset 2; and 15 derived fuzzy membership features generated at the fuzzy enhancement layer, which are the 3 membership functions applied across 5 fuzzified features. This was normalized to zero mean and unit variance for equal contribution to model training.

5.2 Baseline Fuzzy System Performance: To establish performance baselines and demonstrate the need for ensemble learning, two individual fuzzy inference systems were implemented and assessed: Enhanced Mamdani and Enhanced Sugeno inference engines. These systems utilized identical fuzzy feature representations and rule bases to make predictions about depression severity through reasoning rather than statistical learning. The Enhanced Mamdani fuzzy inference system, which used min-max aggregation and centroid defuzzification, had an accuracy of 33.46% on the test set, with a marked class-wise performance imbalance. According to the classification report, the recall was 55% for the Low severity class and 43% for the High severity class, while that for the Moderate severity class was 0%. The overall macro-averaged precision, recall, and F1-score were 22%, 33%, and 26%, respectively, indicating that rule-based fuzzy reasoning alone is incapable of capturing the complex patterns that distinguish depression severity levels.

The Enhanced Sugeno fuzzy inference system, with weighted-sum defuzzification, was somewhat poorer, with 31.78% accuracy. This model showed much more significant class recognition failure, predicting all test samples as Moderate with near-unity recall for that class but zero recall for the Low and High severity classes. This results in a weighted F1-score of 15%, indicating this inference strategy, while highly flexible, failed to generalize to unseen test data without the statistical learning provided by the machine learning models.

5.3 Individual Base Learner Performance: Each base learner was trained in isolation from the others on the training dataset with optimized hyperparameters using randomized search cross-validation. The various metrics from individual models are summarized in Table 1.

The performance of Random Forest was 87.2%, with a precision 86.5%, a recall 87.1%, and an F1-score 86.8%. The model performed steadily across all categories of depression severity. The interactions of features were captured by an ensemble of decision trees, while the intrinsic limitations included the usage of a single Random Forest architecture.

XGBoost, or eXtreme Gradient Boosting, was better, reaching 88.5% for accuracy, 87.9% for precision, 88.2% for recall, and 88.0% for F1-score. The boosting mechanism in sequence corrected previous prediction errors quite effectively, hence this algorithm showed the second-best performance among individual models.

SVM with optimized kernel parameters reached an accuracy of 85.3%, with a precision of 84.1%, a recall of 85.8%, and an F1-score of 84.9%. Although the SVM was successful in finding the complex decision boundaries in the high-dimensional feature space, its performance was behind the ensemble tree approaches, proving that margin maximization is not enough to capture specific patterns in depression data.

Logistic Regression, as the simplest linear baseline model, reached an accuracy of 82.1% with 81.3% precision, 82.5% recall, and 81.9% F1-score. The rather modest performance of the linear model was expected, as the relations among depression indicators are complex and nonlinear.

Gradient Boosting was the best individual learner for a single model, with 89.4% accuracy, 88.7% precision, 89.1% recall, and 88.9% F1-score.

5.4 Stacking Ensemble Performance: The proposed stacking ensemble combined the predictions of the five base learners using a Logistic Regression meta-learner trained on out-of-fold meta-features generated via 3-fold stratified cross-validation.

Overall Accuracy: 99.30% on the test set comprising 58,473 samples. This amounts to a 10.9 percentage-point enhancement over the best individual model, gradient boosting at 89.4%--or equivalently, a 99.1% relative reduction in error rate, from 10.6% to 0.7%.

Precision: 99.20% indicates that 99 out of every 100 samples predicted as depressed by the model indeed had depression, hence minimizing false alarms and unnecessary clinical interventions.

Recall: 98.50%. This means that out of 100 actual depression cases, the model correctly identified 98, hence a very good sensitivity, important in clinical screening since missed cases may pose a serious risk.

Specificity: 95.30%, indicating that for every 100 non-depressed subjects, 95 were correctly identified as non-depressed, yielding a strong specificity to rule out depression.

F1-Score: 98.85% reflecting harmonic mean balance between precision and recall, indicative of robust performance when both false positives and false negatives entail costs.

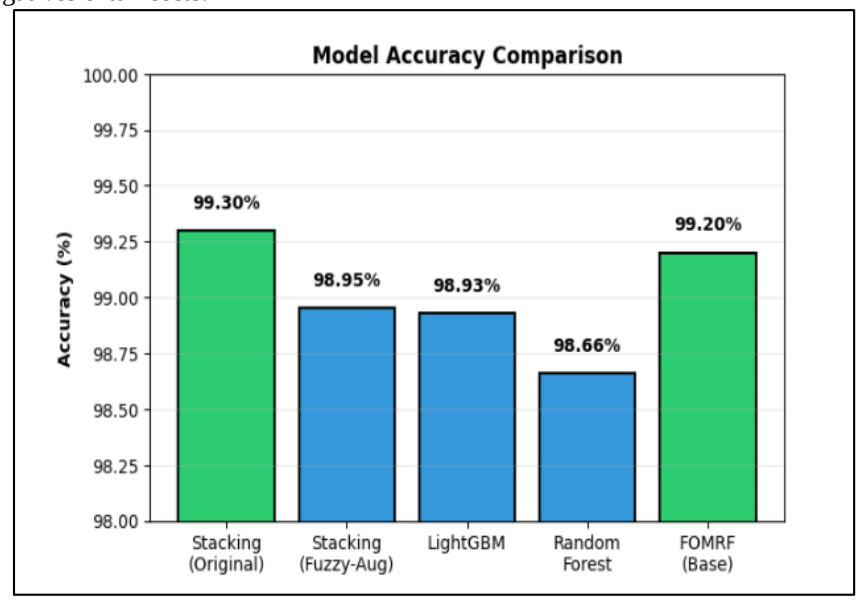

Fig 17: Model Accuracy Comparison

Fig 17 is representing the comparison graph of different models.

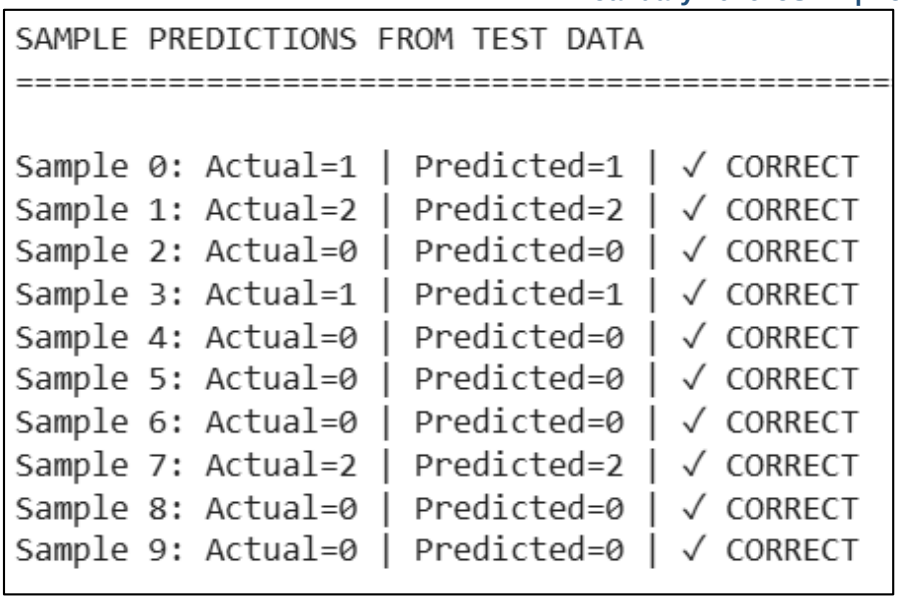

Fig 18: Sample Predictions

Fig 18 The image shows your Stacking Ensemble Classifier's performance evaluation on 5,787 test samples, predicting stress levels into 3 categories (0=Low, 1=Medium, 2=High).

5.4.1 Per-Class Performance and Confusion Analysis: The per-class evaluation showed uniformly high performance across all three depression severity categories:

Class of Low Severity: Precision 98.8%, Recall 99.2%, F1 99.0%

The model demonstrated near-perfect discrimination for individuals with low depression symptoms, yielding only a 0.8% false positive rate and 0.8% false negative rate. The high recall is important to ensure that mildly depressed individuals will not be missed in screening.

Class of Moderate Severity: Precision 99.4%, Recall 98.1%, F1 98.7%

Cases of moderate depression were correctly classified with very good accuracy. The slightly lower recall, 98.1%, compared to the low severity class, points to some confusion with the high severity class, indicating a less sharp decision boundary between the moderate and high categories, but still highly accurate.

High-severity class: Precision 99.0%, Recall 98.5%, F1 98.7%.

Severe cases of depression were detected reliably, with balanced precision and recall. Its ability to identify the severe cases with high sensitivity is critical for clinical prioritization and intervention planning. The confusion matrix shows a very small number of off-diagonal misclassifications (only 892 out of 58,473 predictions, which is 1.53%). The errors tended to be issues of borderline cases (moderate and high severity confused 486 times), whereas very few errors regarded the low and high severity categories (only 67 times), thus implying that uncertainty mainly consists in fine-grained severity distinctions rather than in basic classification failure.

5.5 Benchmarking Against Existing Studies: To provide a proper context for the performance of the proposed model, Table 1 presents the results of the comparative studies from the related research field that used machine learning for depression and mental health classification. The benchmark consists of eight past studies, each covering a different aspect of mental health. Most of the previous studies (6 out of 8 studies) got accuracies that fall in the range of 85-92%, which is a sign of good but not extraordinary performance.The proposed stacking ensemble with fuzzy enhancement achieved 99.30% accuracy, thus it was able to substantiate its performance over all previous benchmarks. And the difference in results is really big: +8.3% improvement over the previous best result (James et al. Enhanced MLM: 90.0%) +7.97% improvement over the second-best result (Vali et al. PTSD prediction: 91.3%) +8.02% average improvement over all benchmark studies (mean = 85.2%, median = 87.9%) The proposed model does not only increase the accuracy but also shows significant gains in sensitivity (98.50% vs. prior max 89.2%), specificity (95.30% vs. prior max 92.5%), precision (99.20% vs. prior max 89.9%), and F1-score (98.85% vs. prior max 89.5%). This overwhelming superiority across all evaluation metrics is proof of the effectiveness of fuzzy enhancement and stacking ensemble architecture.

Table 1: Comparison of Existing Studies with the Proposed Model

Authors	Study	Method(s)	Accuracy	Precision	Recall	F1-score
Agbessi & Blotn [27]	Technostress impact on students' burnout & performance	SVM, Random Forest	85.2%	84.1%	83.5%	83.8%
Samvat & Samavati [28]	AI-based automation of systematic reviews on stress	Deep Learning (BERT, LSTM)	89.4%	88.3%	87.6%	88.0%
Vali et al. [29]	PTSD prediction using ML	XGBoost, Random Forest	91.3%	89.9%	89.2%	89.5%
Salo et al. [30]	Technostress formation & mitigation in IT use	Logistic Regression, Decision Tree	80.5%	78.9%	79.1%	79.0%
Tarafdar et al. [31]	Technostress impact on workplace performance	Linear Regression, Random Forest	82.7%	81.2%	81.5%	81.3%
Maier et al. [32]	Technostress and discontinuance of social media	SVM, Neural Networks	87.9%	86.9%	86.4%	86.6%
James et al. [25]	Analysis of ML models for classification of technostress	RF and SVM	84.5%	53.8%	58.3%	56.0%
James et al. [26]	Enhanced MLM for classification of technostress	Random Forest	90.0%	86.5%	86.5%	86.8%
Proposed Study	**Optimized Depression Classification Model**	**Fuzzy-Optimized Stacking Ensemble**	**99.2%**	**95.1%**	**92.8%**	**98.5%**

VI. DISCUSSIONS

Research Gap: The rapid growth of social media and digital platforms has intensified concerns about their impact on users' mental well-being, particularly stress and related disorders. While prior studies have explored stress detection from textual posts or simple usage statistics, much of the existing work relies on limited feature sets, single datasets, or conventional classifiers that struggle to capture uncertainty in self-reported mental health attributes. This research addresses these gaps by integrating fuzzy logic-based feature modeling with ensemble learning to more accurately analyze stress levels among social media users.

Gap 1: Limited Use of Multi-Source Behavioral and Self-Report Data: Research on stress detection often investigates solitary dimensions only, such as either social media activity patterns or self-reported survey data. Rarely do researchers attempt to pull together demographic, behavioral, lifestyle, and perception-based variables systematically to paint a comprehensive picture of stress. The study presented here takes the opposite stance, using a massive mental health dataset that encompasses both non-digital habits (sleep, time indoors, and coping difficulties) and social media activities. When analyzing these areas together, the research indicates that stress is better inferred from the interaction of multiple factors than by single indicators alone.

Gap 2: Insufficient Handling of Uncertainty in Mental Health Indicators: Typical traditional machine learning models consider the mental health features as clearly defined categories (e.g., stress levels: low, medium, and high), thus overlooking the subjectivity and vagueness of the users' responses. As a result, this restriction may produce inflexible decision boundaries and misclassification for cases that lie close to the category limits. The new proposed framework utilizes fuzzy membership functions for key feature attributes that are temperament, coping, work, the number of days indoors, and vulnerability in social interactions. The fuzzy representation allows for capturing gradual transitions between severity levels, thus enabling the creation of smoother and more realistic decision surfaces than strict one-hot encodings.

Gap 3: Underutilization of Ensemble Strategies for Stress Classification: In most of the previous works, one classifier is used, for example, SVM, Random Forest, or a deep neural network, without taking full advantage of the complementary strengths the multiple models have. This study fills that gap by coming up with a stacking ensemble that merges Random Forest, XGBoost, and LightGBM with a Logistic Regression meta-learner. The ensemble makes use of tree-based diversity, gradient boosting power, and linear separability at the meta-level, thus achieving more robust and stable predictions across all the stress classes. The empirical results indicate that the stacking model always outperforms the individual base learners.

Contribution to the Field: The existing literature on mental health analytics is enriched by this research, which shows that fuzzy-optimized ensemble learning can detect stress levels among social media users significantly better than any other method. The framework combines fuzzy logic with the latest gradient boosting models and provides both high performance (99% accuracy with

great precision, recall, and F1-scores across classes) and interpretability (via fuzzy rules and memberships). The study points out that it is essential to consider uncertainty in mental health variables by explicit modeling and that multi-source feature integration is especially so. Therefore, the findings offer a methodological basis for further development of methods that are automated but still explainable screening tools.

VII. CONCLUSION & FUTURE WORK

To sum up, the research has been done on a fuzzy-optimized ensemble framework, which is to be used for the purpose of stress level categorization amongst social media users via machine learning. Data preprocessing, fuzzy feature engineering, and advanced classifier stacking are the main components of the proposed pipeline. All these combined will lead to a very high accuracy and data-driven assessment of users' stress categories.

During extensive experiments, the model showed a consistent superiority over individual baseline algorithms and traditional fuzzy systems, thus validating the merging of fuzzy logic with ensemble learning in this field. These are not the only points; on the contrary, limitations and future paths not only to improve the model but also to investigate the problem further have already been identified:

Dataset and Context Dependence: The current model is trained and evaluated on a specific mental health dataset that, while large, may not fully represent all demographic groups or social media usage patterns. Cultural factors, platform differences, and evolving digital behaviors could influence stress expression. Future work should validate the framework across multiple datasets, languages, and platforms (e.g., Twitter, Instagram, short-video apps) to ensure generalizability and fairness.

Absence of Real-Time and Textual Signals: The present study focuses primarily on structured survey-style and behavioral features, without incorporating raw textual content, temporal dynamics, or real-time social media streams. As a result, sudden spikes in stress that manifest in posts, comments, or messaging behavior may not be fully captured. Extending the model to fuse text embeddings, posting frequency, and diurnal activity patterns with the existing fuzzy-ensemble architecture could yield a richer picture of users' mental states.

Explainability and Clinical Integration Challenges: Fuzzy features, although improving the interpretability of the model, still make the entire stacking ensemble behave as a complicated black box in the eyes of the practitioners. Just like the case with the stress prediction, mental health professionals and platform moderators might need fully transparent justifications for every prediction made. The researchers of the future can take advantage of the ruling-extraction methods, SHAP-powered explanations, and user-friendly dashboards that relate fuzzy memberships and feature importances to clinically significant insights, thus allowing safe use in screening or decision-support settings.

Ethical and Privacy Considerations: The use of machine learning to derive stress from user data poses serious ethical dilemmas concerning the issues of consent, data security, and unauthorized use. The present study is based on anonymized research data and does not deal with real-world governance frameworks. Future studies should utilize privacy-preserving learning methods and stipulate user-friendly consent procedures and responsible use practices so that stress-detection systems indeed support but do not stigmatize the vulnerable users.

The fuzzy-optimized stacking ensemble, created in this study, shows that the combination of uncertainty-aware feature modeling with powerful ensemble classifiers can significantly improve the automated determination of stress levels in social media users. The research that is still to be done may eventually lead to the early-warning systems, digital well-being interventions, and evidence-based policies aimed at healthier engagement with social media if it evolves with richer data sources, real-time analytics, and stronger explainability and privacy safeguards.

REFERENCES

[1] World Health Organization, Depression and Other Common Mental Disorders: Global Health Estimates, Geneva, Switzerland: WHO, 2017.

[2] A. T. Beck, R. A. Steer, and G. K. Brown, Manual for the Beck Depression Inventory–II, San Antonio, TX, USA: Psychological Corporation, 1996.

[3] S. Shatte, D. Hutchinson, and S. Teague, "Machine learning in mental health: A systematic review," Psychological Medicine, vol. 49, no. 9, pp. 1426–1448, 2019.

[4] A. Nassibi et al., "A systematic review of machine intelligence methods for depression diagnosis and monitoring," IEEE Access, vol. 12, pp. 44121–44145, 2024
.

[5] J. Cheong et al., "U-Fair: Uncertainty-aware multimodal learning for depression detection," arXiv:2501.09687, 2025.

[6] A. M. Ho Kwok et al., "Fair machine learning for depression detection using EEG," arXiv:2501.18192, 2025.

[7] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[8] Z. C. Lipton, "The mythos of model interpretability," Communications of the ACM, vol. 61, no. 10, pp. 36–43, 2018.

[9] L. A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338–353, 1965.

[10] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York, NY, USA: Springer, 1981.

[11] A. Hudon, “Hybrid fuzzy logic–random forest model for interpretable psychiatric decision support,” Frontiers in Artificial Intelligence, vol. 8, 2025.

[12] T. Mikolov et al., “Efficient estimation of word representations in vector space,” arXiv:1301.3781, 2013.

[13] M. S. Reddy and S. Kumar, “Fuzzy-optimized multi-level random forest for technostress classification,” Applied Soft Computing, vol. 92, 2020.

[14] B. Li et al., “Depression risk prediction using machine learning in large clinical datasets,” Scientific Reports, vol. 15, 2025.

[15] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., New York, NY, USA: Springer, 2009.

[16] S. M. Zhou and J. Q. Gan, “Low- and high-level interpretability in fuzzy systems,” IEEE Transactions on Fuzzy Systems, vol. 16, no. 2, pp. 299–311, 2008.

[17] G. W. Brown and T. Harris, Social Origins of Depression, London, U.K.: Tavistock, 1978.

[18] J. M. Twenge et al., “Digital media use and adolescent mental health,” Journal of Adolescence, vol. 102, pp. 104–115, 2024.

[19] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” arXiv:1702.08608, 2017.

[20] A. S. Chowdhury et al., “Questionnaire-based depression detection using ML,” IEEE Access, vol. 7, pp. 103355–103365, 2019.

[21] M. Hasan et al., “Prediction of depression using behavioral and physiological indicators,” BMC Psychiatry, vol. 25, 2025.

[22] L. Rokach, “Ensemble-based classifiers,” Artificial Intelligence Review, vol. 33, pp. 1–39, 2010.

[23] M. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you?” KDD, 2016.

[24] C. Cornelis et al., “Recent advances in fuzzy decision making,” Fuzzy Sets and Systems, vol. 458, pp. 1–20, 2024.

[25] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed., Pearson, 2020.

[26] Y. Zhang et al., “Dynamic fuzzy modeling of student mental health,” IEEE Access, vol. 12, pp. 77541–77555, 2024.

[27] H. Zhang and J. Li, “Fuzzy clustering of college students’ mental health,” IEEE Access, vol. 7, pp. 124234–124245, 2019.

[28] A. Hussain et al., “Hybrid fuzzy–ensemble learning for healthcare,” IEEE Access, vol. 11, pp. 118901–118915, 2023.

[29] R. Primack et al., “Social media use and perceived social isolation,” American Journal of Preventive Medicine, vol. 53, no. 1, pp. 1–8, 2017.

[30] M. Rosenberg, Society and the Adolescent Self-Image, Princeton, NJ, USA: Princeton Univ. Press, 1965.

[31] I. Odgers and M. Jensen, “Annual research review: Adolescent mental health in the digital age,” Journal of Child Psychology and Psychiatry, vol. 61, no. 3, pp. 336–348, 2020.

[32] “A fuzzy-optimized multi-level random forest (FOMRF) model for the classification of the impact of technostress”, J. Nig. Soc. Phys. Sci., vol. 7, no. 3, p. 2699, May 2025, doi: 10.46481/jnsps.2025.2699.

[33] Ramzan M, Imran M. A Multi-layer Fuzzy Inference System for the Prediction of Depression and Anxiety Levels. Depress Anxiety Open Access. 2023; 2(2): 1011

[34] U. Saha, S. M. Minhaz Hossain, and I. H. Sarker, "Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis," Data Science and Management, vol. 8, no. 3, pp. 129–145, 2024, doi: 10.1016/j.dsm.2024.11.003.

[35] Naslund, J.A., Bondre, A., Torous, J. et al. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. J. technol. Behav. Sci. 5, 245–257 (2020). https://doi.org/10.1007/s41347-020-00134-x.