I) **AI - Driven Mental Disorders Categorization from Social Media: A Deep Learning Pre-Screening Framework:** [**IEEE link**](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/document/10580665)

* Published in [2024 International Conference on Machine Intelligence and Smart Innovation (ICMISI)](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/xpl/conhome/10579967/proceeding)
* Date of Conference: 12-14 May 2024
* Date Added to IEEE *Xplore*: 11 July 2024
* ISBN Information:
* DOI: [10.1109/ICMISI61517.2024.10580665](https://doi-org.ccny-proxy1.libr.ccny.cuny.edu/10.1109/ICMISI61517.2024.10580665)
* Publisher: IEEE
* Conference Location: Alexandria, Egypt

1. **Early Studies:**

* Logistic regression with n-gram features and feed-forward neural networks has been utilized to identify mental health expressions in Reddit posts. However, these methods struggled with nuanced text analysis and imbalanced datasets.
* CNN-based models have shown promise in mental health content classification but required independent binary classifiers for specific disorders, limiting their scalability.

1. **Transformer-Based Models**:

* Recent advancements in transformer-based models, such as BERT, DistilBERT, RoBERTa, and AlBERT, offer improved contextual understanding of text data.
* For example, the BERT-base model achieved 91% accuracy on the Kim dataset (six disorders) and 87.85% on the Low dataset (eleven disorders), outperforming previous approaches in precision, recall, and F1 score.

1. **Dataset Use**:

* The study employs the Kim and Low datasets, sourced from mental health-related subreddits to train and validate models.
* **Pros**: These datasets include a large volume of user-generated content spanning multiple years, enabling robust model training. They offer a diverse set of mental health-related discussions.
* **Cons/Limitations**: The datasets lack fine-grained annotations for specific mental health conditions, which could limit the granularity of classification. Additionally, the inherent biases in user posts and subreddit-specific language trends may affect generalizability.

1. **Challenges Addressed:**

* Data imbalance has been a persistent limitation in prior research. This study incorporates techniques such as weighted cross-entropy loss and focal loss, which enable models to focus on minority classes and achieve balanced classification performance.

1. **Model Comparison:**

* The research demonstrates that pre-trained BERT models outperform CNNs and other traditional approaches in accuracy, robustness, and scalability. The class-weighted cross-entropy method proves particularly effective for handling imbalanced datasets.

1. **Future Directions**:

* The study suggests expanding the dataset to include more diverse and annotated data to improve classification granularity and diagnostic accuracy.
* **Pros:** Exploring advanced transformer-based models and multi-modal approaches (e.g., combining text with audio or visual data) could enhance the system's ability to diagnose complex mental health disorders.
* **Cons/Limitations**: Reliance on social media data introduces challenges, including language diversity, regional biases, and the risk of misinterpreting slang or colloquial terms. Scaling the system for real-world deployment will require addressing these limitations.

II) **A Hybrid Learning-Architecture for Mental Disorder Detection Using Emotion Recognition:** [**IEEE Link**](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/document/10577973)

* Published in: IEEE Access (Volume: 12)
* Page(s): 91410 - 91425
* Date of Publication: 01 July 2024
* Electronic ISSN: 2169-3536
* PubMed ID: 39054996

1. **Facial Emotion Recognition and Mental Health**:
   * Previous studies have employed datasets such as AffectNet and FER 2013 for emotion recognition, focusing on emotions like anger, sadness, and happiness. These studies utilized AI models, including CNNs and Vision Transformers (ViTs), for detecting facial expressions associated with mental disorders.
2. **Hybrid Learning Architecture**:
   * The current research introduces a hybrid model that combines CNNs, ViTs, and the YOLOv8 object detection algorithm. The ensemble classifier leverages the strengths of these models to improve accuracy and robustness in identifying mental disorders such as anxiety and depression.
3. **Dataset Use**:
   * The study utilizes the AffectNet and FER 2013 datasets to develop and validate its model. These datasets offer a diverse range of facial expressions, ensuring real-world applicability.
   * **Pros**: The datasets cover a variety of emotions, ethnicities, and lighting conditions, enhancing generalizability.
   * **Cons/Limitations**: Despite their size and diversity, these datasets lack annotations specific to mental health conditions. This necessitates generating a new mental disorder dataset from the predictions, which might introduce biases or inconsistencies in labeling.
4. **Explainability and Interpretability**:
   * To ensure healthcare professionals trust the system, techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and saliency maps are incorporated. These tools highlight regions in the input images that significantly influence the model's predictions, providing transparency and aiding informed decisions.
5. **Performance and Limitations**:
   * The ensemble model achieves an overall accuracy of 81%, outperforming traditional methods in terms of precision and recall.
   * **Pros**: The integration of advanced techniques and multiple learning models ensures robust performance and scalability.
   * **Cons/Limitations**: The system's dependency on visual cues may lead to inaccuracies when facial expressions are ambiguous or influenced by external factors (e.g., lighting, occlusion). Additionally, the reliance on static datasets raises concerns about its adaptability to diverse real-world scenarios.
6. **Future Directions**:
   * The research outlines potential improvements, including the use of larger and more comprehensive datasets specifically tailored to mental health conditions.
   * Exploring multi-modal approaches, such as combining facial analysis with speech and text data, could enhance diagnostic accuracy.
   * Advanced transformer models and techniques for reducing dataset biases are suggested to improve generalizability and scalability.

III) **Machine Learning-Based Detection of Post-Traumatic Stress Disorder in Mental Health:** [**IEEE Link**](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/document/10193036)

* Published in: 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)
* Date of Conference: 06-08 July 2023
* Date Added to IEEE Xplore: 01 August 2023
* DOI: 10.1109/ICESC57686.2023.10193036
* Publisher: IEEE
* Conference Location: Coimbatore, India

1. **Model Learning Architecture:**

This study provides a systematic review of various machine learning approaches used to diagnose mental health conditions such as schizophrenia, bipolar disorder, anxiety, depression, and PTSD. It highlights existing obstacles, restrictions, and future opportunities in applying ML to mental health analysis.

1. **Dataset Use:**

* Utilizes the PRISMA methodology to evaluate 30 unique research articles.
* Examines multimodal ML models, including deep learning-based classification techniques.
* Identifies key performance factors and limitations of AI-based mental health screening.

1. **Explainability and Interpretability:**

This study emphasizes the importance of transparency in AI-driven mental health diagnosis by reviewing how different ML models explain their predictions. Techniques such as attention mechanisms and decision visualization tools are suggested to enhance trust among healthcare professionals.

The study also incorporates methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to provide insights into feature importance, ensuring interpretability for practitioners. Furthermore, visualization techniques such as heatmaps and gradient-based saliency maps are recommended to highlight the most influential features in diagnosing PTSD and related disorders.

1. **Performance and Limitations:**

* The reviewed machine learning models have demonstrated varying degrees of effectiveness in diagnosing different mental disorders.
* Some models showed limitations in handling complex emotional states and non-verbal cues from patients.
* The study suggests that a combination of supervised and unsupervised learning approaches may improve overall performance.
* Accuracy values for different models ranged from 78% to 97%, with random forest achieving the highest accuracy in PTSD detection.

1. **Future Directions:**

* The study recommends integrating multi-modal data sources, such as social media text and voice recordings, with facial expression analysis.
* Developing new benchmark datasets specifically tailored for mental health screening to improve AI model reliability.
* Investigating the impact of ethical AI principles and bias mitigation strategies to ensure fair and responsible AI-driven mental health diagnosis.

1. **Alignment with Our Work**:

The emphasis on explainability is crucial for our project, as we aim to build a system that mental health professionals can trust. We plan to incorporate similar techniques (e.g., Grad-CAM and saliency maps) to ensure transparency in our model's predictions.

**IV)** **Multi-task Learning on Mental Disorder Detection, Sentiment Analysis, and Emotion Detection Using Social Media Posts:** [**IEEE Link**](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/document/10845733)

* Published in: 2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)
* Date of Conference: 06-07 December 2024
* Date Added to IEEE Xplore: 22 January 2025
* DOI: 10.1109/ISAS64331.2024.10845733
* Publisher: IEEE
* Conference Location: İstanbul, Turkiye

1. **Model Learning Architecture:**

This study explores the application of multi-task learning (MTL) to mental disorder detection using social media posts. The proposed framework trains on three interconnected tasks: mental disorder detection as the primary task, while sentiment analysis and emotion analysis serve as auxiliary tasks. The authors implement a joint training and semi-shared trunk architecture that allows shared representations across the three tasks to enhance model performance. Compared to single-task learning models, the MTL approach significantly improves classification accuracy by leveraging shared features across related domains.

1. **Dataset Use:**

* The study utilizes the SWMH dataset, which includes posts from five Reddit communities: ‘Anxiety,’ ‘Bipolar,’ ‘Depression,’ ‘Suicide Watch,’ and ‘offmychest.’
* An additional 5,000 instances were manually labeled with sentiment (positive/negative) and emotion labels (anger, fear, hopefulness, hopelessness, joy, sadness, calmness, and disgust).
* The dataset was preprocessed using text normalization, lemmatization, stop-word removal, and tokenization to improve data quality and ensure consistent input representations.

1. **Explainability and Interpretability:**

To enhance model transparency and build trust among mental health professionals, the study integrates attention mechanisms into the MTL architecture. The authors employ SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to provide feature importance insights. Furthermore, heatmaps and gradient-based saliency maps highlight the influential factors contributing to mental disorder classification, making the model’s decisions interpretable to researchers and clinicians.

1. **Performance and Limitations:**

* The multi-task model achieved an F1-score of 62% for mental disorder detection, outperforming its single-task equivalent by 2%.
* Sentiment detection maintained an F1-score of 88% across all models.
* Emotion classification experienced a slight drop in performance, with the MTL model scoring 63%, compared to 66% in single-task models.
* The model faced challenges in detecting complex mental disorders like bipolar disorder due to overlapping symptoms with other conditions.
* Limited training instances for specific emotion labels (e.g., calmness, disgust) reduced classification accuracy in these categories.

1. **Future Directions:**

* Expanding the dataset by labeling more instances for underrepresented classes to improve model performance.
* Integrating additional multi-modal features such as voice and facial expressions to enhance detection accuracy.
* Developing real-world applications such as digital mental health assistants and chatbots that can leverage MTL for automated mental health assessments.
* Addressing ethical concerns regarding data privacy and model bias to ensure fair and responsible AI deployment in mental healthcare.

1. **Alignment with Our Work**:

The use of multi-task learning is relevant to our future plans of integrating multi-modal data (e.g., text and facial expressions). The study's focus on emotion detection aligns with our project's goal of using facial expressions to assess mental health.

V) **Improving Mental Health Assessments: CNN-SVM for Depression Detection:** [**IEEE Link**](https://ieeexplore-ieee-org.ccny-proxy1.libr.ccny.cuny.edu/document/10774790)

* Published in: 2024 3rd International Conference for Advancement in Technology (ICONAT)
* Date of Conference: 06-08 September 2024
* Date Added to IEEE Xplore: 10 December 2024
* DOI: 10.1109/ICONAT61936.2024.10774790
* Publisher: IEEE
* Conference Location: GOA, India

1. **Model Learning Architecture:**

This study proposes a hybrid CNN-SVM approach for diagnosing five mental disorders: **Major Depression, Melancholia, Delusions, Paranoia, and Antenatal Depression**. The model leverages convolutional neural networks (CNNs) for feature extraction and a support vector machine (SVM) for classification. The model aims to improve detection accuracy by learning hierarchical representations from sequential input data.

1. **Dataset Use:**

* The dataset consists of multiple categories of mental disorders, with imbalanced instances across classes.
* Preprocessing steps included normalization, data augmentation, and feature extraction from textual and image-based inputs.
* The model was trained on a combination of structured mental health assessment records and real-world clinical datasets.

1. **Explainability and Interpretability:**

* The model employs confusion matrices and class-wise analysis to validate the differentiation between multiple mental disorder categories.
* A combination of macro-average, weighted-average, and micro-average performance metrics provides a robust evaluation of classification reliability.
* Feature importance analysis identifies key input attributes contributing to disorder classification, ensuring transparency for clinical validation.

1. **Performance and Limitations:**

* The CNN-SVM model achieved a macro-average precision of 91%, recall of 92.67%, and F1-score of 92.27%.
* The weighted-average metrics reported precision of 92.19%, recall of 92%, and F1-score of 92.13%, confirming the model’s robustness across unbalanced class distributions.
* The micro-average across all classes was 92% for precision, recall, and F1-score.
* Individual class performances were notable, with Major Depression achieving 90.31% precision and 97% recall.
* The Antenatal Depression category achieved a precision of 93.48% and recall of 94.80%, demonstrating strong classification accuracy.

1. **Future Directions:**

* Enhancing model generalization by incorporating larger, multi-center datasets for better representation.
* Exploring additional hybrid AI architectures, integrating transformers with CNN-SVM to improve disorder classification.
* Developing clinical decision support systems based on the proposed model to aid mental health professionals in real-time diagnosis.
* Addressing bias and fairness concerns by ensuring equal representation of diverse population groups in training data.

1. **Alignment with Our Work**:

The hybrid approach of combining CNNs with other classifiers (e.g., SVM) is similar to our proposed CNN-LSTM hybrid model for temporal analysis. The study's focus on depression detection is directly relevant to our project.