# Title: Deep Learning Approach for Depression Classification and Prediction from User Tweets

* **JUSTIFICATION:**
* Social media platforms, like Twitter, provide real-time and extensive data for analyzing depression through user expressions​.
* Small transformer-based models ensure high accuracy in depression intensity classification while being computationally efficient for low-resource systems.
* Multi-class classification (mild, moderate, severe) offers a refinement understanding of depression, surpassing binary classifications​.
* **What the Topic Wants to Say**:

**Importance of Depression Detection**:

* Depression is a global mental health crisis, often leading to severe consequences like suicide if not addressed early.
* Social media has become a critical platform for individuals to express emotions, including signs of depression.

**Leveraging Social Media Data**:

* Platforms like Twitter are a rich source of data for identifying depression through user posts.
* Users often share sentiments and mental states that can be analyzed to detect depression intensity.

**Small Deep Transfer Learning Models**:

* The research emphasizes the use of lightweight transformer-based models for depression classification.
* These models are efficient in terms of computational resources, making them accessible for low-powered devices.

**Multi-Class Classification**:

* Moving beyond binary classification (depressed vs. non-depressed), the study categorizes depression into three intensities: mild, moderate, and severe.

**Practical Applications**:

* The study highlights the effectiveness of models like ESG (Electra Small Generator), which achieve high accuracy with minimal computational cost.
* Such advancements can aid in building accessible tools for early detection of depression.

REQUIRED WRITE UP:

The research highlights the potential of social media as a resource for identifying depression through advanced NLP techniques. It explores small deep transfer learning models that classify tweets into varying intensities of depression (mild, moderate, severe). These lightweight models balance accuracy and computational efficiency, making them suitable for real-world applications, including deployment on low-resource devices. This work demonstrates how integrating cutting-edge transformer models with social media analysis can contribute to public health by enabling early detection and intervention in mental health crises.

# DATA SET USED IN THE TOPIC

The dataset used in the research paper titled **"Depression Classification From Tweets Using Small Deep Transfer Learning Language Models"** comprises Twitter data.

* **Source**: Depression-related tweets were collected using Twitter's public APIs.
* **Collection Method**: Tweets were gathered by applying hashtags related to depression as seed words. Tools like **VADER** (Valence Aware Dictionary and sEntiment Reasoner) and **TextBlob** were used for sentiment analysis to calculate polarity and subjectivity scores.
* **Labeling**: Tweets were categorized into three depression intensity classes—'mild,' 'moderate,' and 'severe'—based on the **ICD-10 depression diagnostic criteria**.
* **Dataset Size**: A total of **73,355 labeled tweets** were included, with class-wise tweet counts detailed in the paper.
* **What is VADER and TEXT BLOB:**
* **VADER (Valence Aware Dictionary and Sentiment Reasoner):**

VADER is a lexicon-based sentiment analysis tool designed to recognize sentiment in text, particularly from social media. It assigns polarity scores to text—how positive, negative, or neutral the sentiment is. Key features include:

1. **Pre-trained Lexicon:** VADER includes a dictionary of words rated for their emotional intensity.
2. **Scoring System:**
   * Positive, Negative, Neutral, and Compound Scores are provided.
   * The compound score is a normalized, aggregated sentiment score ranging from -1 (most negative) to +1 (most positive).
3. **Context Sensitivity:**
   * It considers the intensity of punctuation (e.g., exclamation marks) and capitalization.
   * Modifiers like "very" or "not" influence the sentiment score.
4. **Use Cases:** Popular for analyzing tweets, product reviews, and informal texts where standard sentiment tools might struggle.

* **TextBlob**

TextBlob is a Python library for **natural language processing (NLP)** tasks. It's simple and intuitive, offering functionalities for a range of tasks:

1. **Sentiment Analysis**:
   * It uses rule-based methods to classify text as **positive**, **negative**, or **neutral** based on pre-trained models.
   * Provides two metrics: **Polarity** (ranges from -1 to 1) and **Subjectivity** (ranges from 0 to 1).
   * **Polarity**: Indicates the sentiment's direction (negative or positive).
   * **Subjectivity**: Measures the degree to which a statement is opinion-based versus fact-based.
2. **Text Processing**:
   * Tokenization
   * Part-of-speech tagging
   * Language translation and detection
3. **Ease of Use**: It is beginner-friendly, making it a common choice for quick sentiment analysis and prototyping.

* **A list of relevant datasets that can be used for the task:**
* **eRisk 2017 Dataset:**
* **Reason**: Specifically designed for early detection of depression and related mental health conditions. It includes text-based data from Reddit with annotations for depression severity and suicidal ideation.
* **Use Case**: Enhances models' ability to detect early signs of depression with nuanced classification.
* **DAIC-WOZ Dataset (Distress Analysis Interview Corpus):**
* **Reason**: A high-quality clinical dataset that includes transcriptions and audio data for depression detection. It enables multi-modal approaches, which improve model accuracy by combining text and audio cues.
* **Use Case**: Refines models for detecting depression severity, especially in conversational contexts.
* **Sentiment140:**
* **Reason**: Large-scale Twitter dataset labeled for sentiment polarity. It helps improve the foundational sentiment analysis capability of the model, especially for informal text like tweets.
* **Use Case**: Strengthens baseline sentiment understanding for short, noisy, and informal texts similar to tweets.

***WHAT IS Small Deep Transfer Learning Language Models ?***

**Definition**: Small deep transfer learning language models are **pre-trained transformer-based models** with fewer parameters and computational requirements compared to large-scale models (like BERT or GPT-3). These models are designed to balance computational efficiency with high performance for various Natural Language Processing (NLP) tasks, including text classification, sentiment analysis, and depression detection.

**Key Concepts**

**1. Deep Learning Language Models**

* Deep learning language models are neural networks designed to understand and generate human language by learning patterns in text data.
* These models use a **transformer architecture**, which leverages self-attention mechanisms to process entire input sequences simultaneously.
* Examples: BERT, GPT, Electra, RoBERTa.

**2. Transfer Learning**

* **Transfer learning** refers to taking a model trained on a large, generic dataset (e.g., Wikipedia or Common Crawl) and fine-tuning it on a smaller, task-specific dataset.
* The pre-trained model already understands the general structure and meaning of the language, so it requires fewer resources and less data to adapt to a specific task.
* Example: A pre-trained BERT model fine-tuned on depression-related tweets for classification.

**3. "Small" Language Models**

* These are smaller versions of deep transfer learning models with fewer parameters, making them lightweight and faster without significantly sacrificing performance.
* They are particularly useful in scenarios with limited computational resources (e.g., mobile devices, IoT systems) or when quick training is necessary.

**Advantages of Small Deep Transfer Learning Models**

1. **Reduced Computational Complexity**:
   * Fewer parameters mean faster training and inference times.
   * Suitable for devices with limited memory or processing power.
2. **Efficiency**:
   * Lower training costs without a significant drop in performance.
3. **Fine-Tuning Flexibility**:

* Can be effectively fine-tuned for specific tasks (e.g., depression classification from tweets).

1. **Energy Efficient**:
   * Requires less energy for training and deployment, making it environmentally friendlier.

* **ABOUT DEPRESSION**
* **What is Depression?**
* Depression, also known as major depressive disorder (MDD), is a mental health condition characterized by persistent feelings of sadness, hopelessness, and a lack of interest or pleasure in activities once enjoyed.
* It affects both mental and physical well-being, influencing how a person thinks, feels, and behaves.

**Causes of Depression**

Depression is a complex condition with multiple contributing factors. These can be broadly categorized into:

**1. Biological Factors**

* Genetics
* Neurotransmitter Imbalance
* Hormonal Changes
* Brain Structure

**2. Psychological Factors**

* Trauma
* Negative Thought Patterns
* Stress

**3. Environmental and Social Factors**

* Life Events
* Social Isolation
* Cultural and Societal Pressures

**4. Lifestyle and Behavioral Factors**

* Substance Abuse
* Poor Diet
* Lack of Sleep and Exercise

**5. Medical Conditions**

* Chronic Illnesses
* Side Effects of Medications

### **RESEARCH PAPER (BASE PAPER) ANALYSIS**

The Detection of Depression Using Multimodal Models Based on Text and Voice Quality Features

Hanadi Solieman

*Saint Petersburg Electrotechnical University "LETI"*

St. Petersburg, Russia [khsoliman@stud.etu.ru](mailto:khsoliman@stud.etu.ru)

**Writer objective:**

The objective of the research is to develop an intelligent, automated system for depression detection using a multimodal approach that combines text and voice quality features. By leveraging deep learning techniques, the study aims to analyze linguistic and acoustic markers of depression extracted from the DAIC-WOZ database. The research focuses on building separate models for text and voice data, enabling a comprehensive evaluation of both verbal and non-verbal indicators of depression. Ultimately, the goal is to create a system that facilitates early, efficient, and accessible mental health diagnostics without the need for clinical visits, addressing the limitations of traditional diagnosis methods such as subjectivity, inconsistency, and cost.

**Technologies Proposed By the Authors:**

* **Multimodal Models:**
* Utilized a combination of text and voice features for depression detection.
* Developed separate unimodal models for text analysis and voice quality analysis.
* **Deep Learning Techniques:**
* Employed deep neural networks for feature extraction and classification.
* Implemented sequential models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) layers.

** Natural Language Processing (NLP):**

* Applied word-level text analysis for semantic content extraction using word embeddings.
* Focused on sentiment analysis to identify depressive indicators in textual transcripts.
* **Voice Quality Analysis:**
* Used acoustic features such as Normalized Amplitude Quotient (NAQ), Quasi-Open Quotient (QOQ), H1-H2, Maxima DispersionQuotient (MDQ), and Peak Slope.
* Analyzed voice qualities like tenseness and breathiness to detect signs of depression.
* **Transfer Learning:**
* Leveraged pre-trained word embeddings for the text analysis model to enhance accuracy and reduce training requirements.
* **Data Preprocessing Techniques:**
* Employed dropout layers to prevent overfitting.
* Used focal loss functions to handle class imbalance issues in the DAIC-WOZ dataset.
* **Speaker-Independent Modeling:**
* Designed models that are independent of individual voice characteristics, ensuring generalizability across different speakers.

**The technologies proposed by the authors, including deep learning, NLP, voice quality analysis, focal loss, and transfer learning, work together to build robust models for depression detection**.

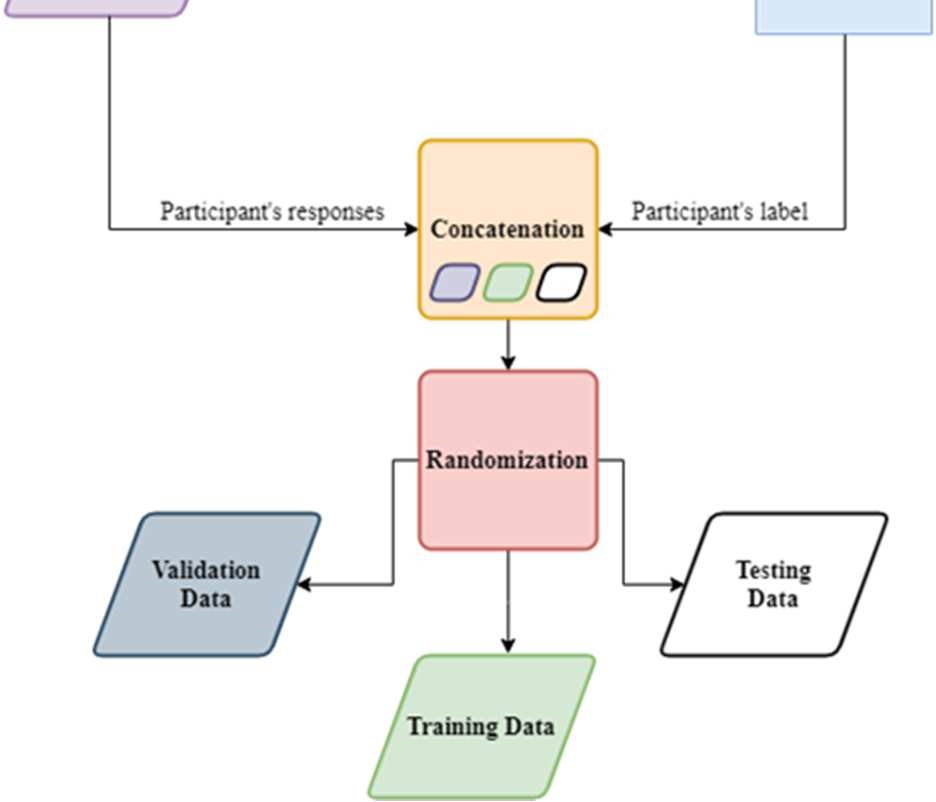
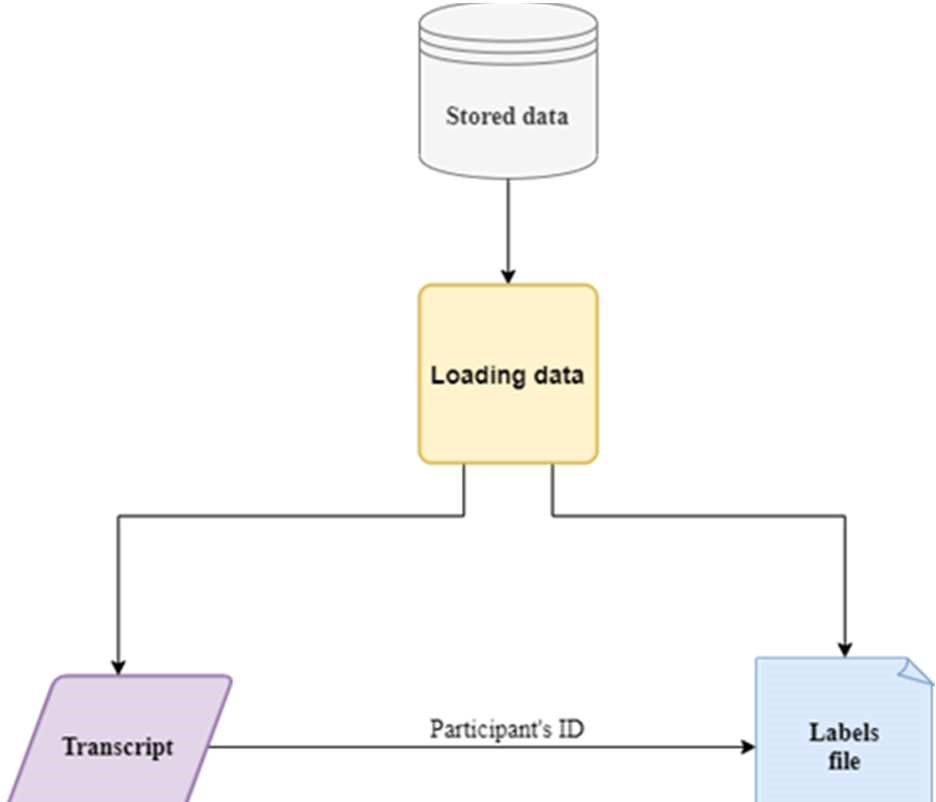
**Problem statement:**

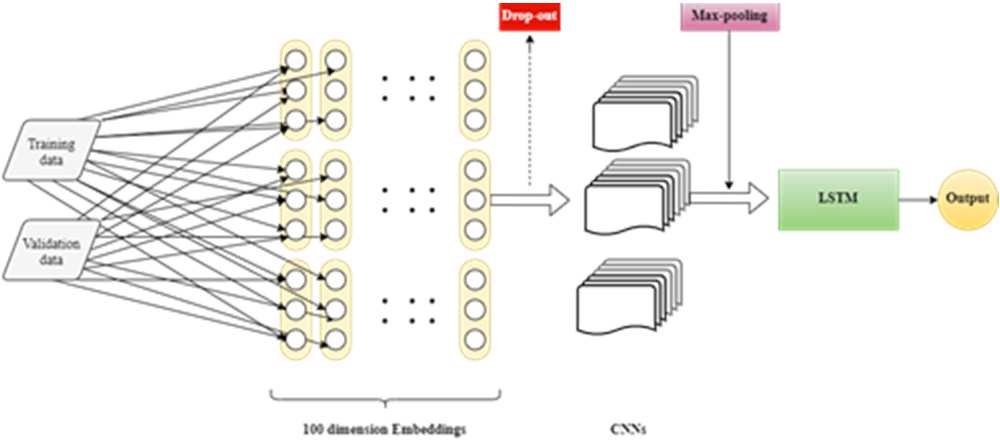
This research addresses these issues by proposing a multimodal system that integrates textual and voice quality features to detect depression using deep learning models. The system aims to provide an efficient, reliable, and accessible alternative to traditional diagnostic methods, facilitating early detection and improved mental health outcomes.

**PROPOSED MODEL:**

The proposed model in *"The Detection of Depression Using Multimodal Models Based on Text and Voice Quality Features"* introduces a sequential multimodal approach to detecting depression by analyzing text and voice data. The system integrates deep learning techniques to process and classify data efficiently.

* 1. **Multimodal Approach:**
* The model incorporates two distinct unimodal sub-models:
  + **Text Analysis Model**: Processes textual data from interview transcripts.
  + **Voice Quality Model**: Analyzes vocal features extracted from audio recordings.
* Both models operate independently but provide complementary insights, enabling a comprehensive analysis of verbal and non-verbal depression markers.
  + 1. **Text Analysis Model**
* Utilizes **Natural Language Processing (NLP)** techniques to analyze word-level semantic content.
* Employs pre-trained word embeddings to create vector representations of words.
* Architecture:
  + An **embedding layer** with pre-trained static weights for semantic representation.
  + A **CNN layer** extracts local patterns in word sequences.
  + An **LSTM layer** captures temporal dependencies in text.
  + A **dense layer** outputs the classification result (depressed or non-depressed).
* Focuses on sentiment analysis, ignoring syntactic and lexical features.
  + 1. **Voice Quality Model**
* Processes audio features like normalized amplitude quotient (NAQ), quasi-open quotient (QOQ), H1-H2, MDQ, and peakSlope, extracted using the **COVAREP toolbox**.





Remarkably, it performed better on the five glottal flow features (NAQ, QOQ, H1-H2, MDQ, and peakSlope) in comparison with the whole set of COVAREP. This contradicts the previous results of multiple studies. Still, our optimization was on just the 5 features, and that makes the results applicable just for it than any other set.

The best results of both models on the testing data are described in the table below.

TABLE. THE RSULTS ON TEST DATA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Focal Loss** | **Accuracy** | **Class** | **Precision** | **Recall** | **F1score** |
| Text analysis | 0.35 | 0.70 | Nondepres  sed | 0.74 | 0.88 | 0.81 |
|  |  |  | Depressed | 0.50 | 0.29 | 0.36 |
| Voice quality | 0.67 | 0.66 | Nondepres  sed | 0.77 | 0.73 | 0.75 |
| Depressed | 0.44 | 0.50 | 0.47 |
| Concatena  tion | 0.07 | 0.68 | Nondepres  sed | 0.74 | 0.85 | 0.79 |
| Depressed | 0.44 | 0.29 | 0.35 |

The used programming language to code both of those models is Python, implemented on Google Colab. Both models used focal loss function in a try to overcome the imbalance problem in the DAIC-WOZ database. The text analysis model detected depression with an F1-score equal to 0,8 on the non-depressed group of the test set. While the voice quality model made it with 0,75.

**OBJECTIVE TO INCREASE THE ACCURACY AND PERFORMANCE:**

* **. Vision Transformers (ViT) for Image Analysis:**
  + Vision Transformers have shown state-of-the-art performance in image classification tasks. They handle large image datasets well and can capture detailed visual cues that might indicate mood or emotional state.
  + Apply a ViT model to analyse visual content associated with tweets, which can be helpful for capturing depressive themes in images shared by users.
* **. Multimodal Attention Mechanisms:**
* Attention mechanisms highlight important parts of each input (text, image, video) and can help focus on depression-relevant aspects in complex, multimodal content.
* Implement attention layers that weigh key features across modalities, helping the model prioritize depressive signals in text, imagery, and video content for improved prediction.

**RESEARCH GAP:**

The research gaps in the paper include the use of a small and imbalanced dataset (DAIC-WOZ), limiting generalizability and performance. The models lack advanced contextual understanding for text and broader integration of multimodal features. Real-world scalability and robustness against variability in voices and noise are not addressed. Additionally, the focus on binary classification misses opportunities to detect depression severity or temporal progression.

**Improvement:**

Incorporating advanced models like transformers for contextual understanding and deeper multimodal fusion techniques can enhance feature integration. Testing on real-world noisy datasets will improve scalability and robustness. Expanding to multi-class classification or regression tasks can capture depression severity and progression.

**Parameter:**

The model uses 224×224×3 input images, five convolutional layers with 3×3 kernels, ReLU activation, and 2×2 max pooling. It includes two fully connected layers, a 0.5 dropout rate, and a softmax output. Training is done using the Adam optimizer (learning rate 0.0001), batch size 32, for 50 epochs, with categorical cross-entropy loss.