**Title:**

**Depression Classification and Prediction System from User Tweets Uaing ML Techniques**

***2ND******RESEARCH PAPER ANAYSIS REPORT BY:***

**VIVEK PANDEY**

***AKTU***

**COMPUTER SCIENCE AND ENGINEERING**

**April 4th 2025**

**BASE PAPER :**

Received October 23, 2019, accepted November 8, 2019, date of publication November 14, 2019, date of current version November 26, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2953087*

**A Novel Co-Training-Based Approach for**

**the Classification of Mental Illnesses**

**Using Social Media Posts**

SUBHAN TARIQ 1, NADEEM AKHTAR 2, HUMAIRA AFZAL 3, SHAHZAD KHALID 4, MUHAMMAD RAFIQ MUFTI 5, SHAHID HUSSAIN 6, ASAD HABIB 6, AND GHUFRAN AHMAD 1

1Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan

2Department of Computer Science and IT, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

3Department of Computer Science, Bahauddin Zakariya University, Multan 60000, Pakistan

4Department of Computer Engineering, Bahria University, Islamabad 44000, Pakistan

5Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan

6Department of Computer Science, Kohat University of Science and Technology, Kohat 26000, Pakistan

Corresponding author: Shahid Hussain ([shussain@comsats.edu.pk](mailto:shussain@comsats.edu.pk))

**Research Overview :**

This research has demonstrated the social media data and big data analytics to enhance the identification and prediction of mental health conditions, such as depression, anxiety, bipolar disorder, and ADHD.

Traditional methods require extensive historical data or constant patient monitoring, which can be impractical.

To address this limitation, **the study proposes a novel semi-supervised approach** **using Co-training, which combines the discriminative power of machine learning classifiers—Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes (NB)** to analyze **data extracted from Reddit posts** and their associated **comments**.

Experimental results show **that this method outperforms existing state-of-the-art classifiers by an average margin of 3%, highlighting its efficacy in mental health prediction**.

**DATASET:**

* **REDDIT :**

Reddit is a website where people share posts, links, and pictures. Other users can vote on these posts **VOTE UP if they like them or VOTE DOWN if they don't**. The most popular posts with more upvotes move to the top of their category.

The site is divided into different communities called "subreddits," each focused on a specific topic like news, science, movies, games, health, or food. Users can join these subreddits to discuss their interests. If a post gets enough upvotes, it can even reach Reddit’s front page, where more people see it

In this research, They focus on mental health subreddits (like those for Depression, Anxiety, Bipolar Disorder, and ADHD) to analyze users' posts for signs of mental illness. By training a machine learning model on this data.

**OBJECTIVE** **is to automatically detect mental health concerns in future posts.** The main objective of proposed methodology is to classify the mental illnesses, based on the training data from the posts on clinical sub-reddits.

**METHODOLOGY:**

This study involves training three machine learning models (SVM, Naive Bayes, and Random Forest) to classify posts from four mental health-related subreddits: ADHD, Depression, Bipolar, and Anxiety.

1. **SVM (Suport vector machine ) for text classification :**

**Text Preprocessing**

Before training an SVM model, the text data needs to be preprocessed:

1. Tokenization: Break down the text into individual words or tokens.

2. Stopword removal: Remove common words like "the", "and", etc. that don't add much value to the text.

3. Stemming or Lemmatization: Reduce words to their base form (e.g., "running" becomes "run").

4. Vectorization: Convert the text data into numerical vectors that can be fed into the SVM model.

**Vectorization Techniques**

There are several vectorization techniques that can be used:

1. Bag-of-Words (BoW): Represent each document as a bag, or a set, of its word frequencies.

2. Term Frequency-Inverse Document Frequency (TF-IDF): Weight word frequencies by their importance in the entire corpus.

3. Word Embeddings (e.g., Word2Vec, GloVe): Represent words as dense vectors in a high-dimensional space, capturing semantic relationships.

**SVM Model**

Once the text data is preprocessed and vectorized, it can be fed into an SVM model:

1. Linear Kernel: Use a linear kernel to classify the text data. This is suitable for high-dimensional data.

2. Non-Linear Kernel: Use a non-linear kernel (e.g., polynomial, radial basis function (RBF)) to classify the text data. This is suitable for non-linearly separable data.

3. Soft Margin: Use a soft margin to allow for some misclassifications. This is suitable for noisy data.

1. **Naive Bayes for Reddit Dataset**

Naive Bayes is a simple but powerful machine learning algorithm that is great for text classification like sorting Reddit posts into categories (e.g., ADHD, Anxiety, Depression, Bipolar).

NB PROCCES FLOW:

1. **Counts Words**:

It scans all the posts and notes how often words (e.g., "stress," "therapy," "sleep") appear in each subreddit. Example: The word "mania" might appear a lot in **Bipolar** posts but rarely in **Depression** posts.

1. **Makes Predictions Using Probability**:

When a **new post** comes in, Naive Bayes checks its words and calculates:

*"How likely is this post ADHD vs. Anxiety vs. Depression based on the words it uses?"*

It picks the category with the **highest probability**.

1. **"Naive" Assumption**:

It treats every word as **independent** (even though words in real life depend on each other). Example: It assumes "panic attack" has the same meaning as "panic" + "attack" separately. This simplifies calculations but isn’t perfectly accurate.

1. **RF (Random Forest):**

Random Forest is like a team of decision trees working together to classify Reddit posts (e.g., into ADHD, Anxiety, Depression, or Bipolar categories). Instead of relying on one "expert," it combines the opinions of many trees to make a smarter, more accurate decision.

1. **Builds Multiple Decision Trees**:

Each tree is trained on a **random subset** of the data (like giving each tree a different sample of Reddit posts). Each tree also uses a **random subset of words** (features) to make decisions.

1. **Voting System**:

When a new post arrives, **every tree votes** on which subreddit it belongs to.

The final prediction is the **most popular vote** (e.g., 30 trees say "ADHD," 10 say "Anxiety" → Result = "ADHD").

1. Randomness ensures trees don’t all make the same mistakes. Some trees might focus on words like "focus" (ADHD), while others check "sadness" (Depression).

**EXECUTION APPROACH:**

* + 1. **DATA COLLECTION :** the first step in a research methodology focused on gathering mental health-related data from Reddit.

1. Tool Used: Python with PRAW (Python Reddit API Wrapper)

PRAW is a popular Python library that allows easy access to Reddit's API for data collection

2. Subreddits Targeted:

* r/Depression
* r/Anxiety
* r/ADHD
* r/Bipolar
* These were chosen as they represent major mental health conditions

3. Post Collection:

* Collected the "top" 1000 posts from each subreddit (sorted by upvotes/hotness)
* Expected total: 4000 posts (1000 × 4 subreddits)
* Actual total: 3922 posts due to some being deleted by users or moderators

4. Comment Collection:

* For each post, gathered the top 5 comments (likely most upvoted or most relevant)
* Comments were stored separately from the main posts
  + 1. **DATA PROCESSING:**
* Eliminate empty rows or missing values from the dataset.
* Standardize all text to lowercase to ensure consistent word treatment (e.g., "Dog" → "dog").
* Split sentences into individual words (tokens).
* Filter out common, low-meaning words (e.g., "the", "is", "and").
* Reduce words to their root forms (e.g., "running" → "run") to group similar words.
* Eliminate numbers, symbols, and punctuation, keeping only letters.
  + 1. **FEATURE EXTRACTION :**

1. **Feature Extraction (TF-IDF)**

* **Term Frequency (TF)**: Measures how often a word appears in a document.
* **Inverse Document Frequency (IDF)**: Reduces the weight of overly common words (e.g., "the," "is") to improve accuracy.
* **Output**: Word-weight vectors representing document features.

1. **Feature Selection**

**Chi-Squared (χ²) Method**:

* Measures the association between features (words) and target labels (e.g., mental health conditions).
* Selects the most relevant features based on highest χ² scores (strongest statistical relationships).
* Formula:

χ2 = ∑ (Observed−Expected)2

Expected

1. **Train-Test Split (80-20)**

* **80%** of data → **Training set** (model learning).
* **20%** of data → **Testing set** (evaluation).
  + 1. **CLASSIFICATION :**

The fourth phase involved was classification using machine learning techniques, comparing traditional models with co-training approaches.

Three base models (SVM, RF, NB) were first implemented with parameter tuning. Then, co-training versions were developed using a combined dataset of labeled posts and unlabeled comments.

The co-training process split features into two views, training weak classifiers on each view using only labeled post data. These classifiers iteratively predicted labels for the most confident unlabeled comments, expanding the training set.

The process repeated through multiple iterations, with model predictions being compared at each stage to progressively improve classification performance by both labeled and unlabeled data.

**RESULT:**

The evaluation compared traditional machine learning models (NB, SVM, RF) with their co-training versions using cross-validation. Among base models, NB showed the weakest performance (68% test accuracy), while RF and SVM achieved around 70% accuracy, though RF exhibited overfitting. The co-training approach significantly improved results, particularly for anxiety classification (F-measure: 0.84 for SVM co-training), demonstrating consistent gains across all mental health categories. While SVM co-training showed the best overall performance, NB and RF co-training versions also outperformed their base counterparts, except in cases of potential feature-label noise affecting ADHD, depression, and bipolar classifications. The study confirmed that semi-supervised co-training effectively enhances model performance when working with limited labeled data.

**CONCLUSTION OF RESEARCH:**

The experimental results indicate that SVM, NB, and RF outperformed with Co-training technique as compared to their individual use in terms of Precision.

**RESEARCH GAP :**

This study's Reddit-based approach faces key limitations compared to Twitter-based depression detection. While Reddit offers structured, long-form posts, Twitter provides noisier but more linguistically diverse real-time data with self-reported labels, reducing ambiguity. The reliance on TF-IDF lacks depth compared to advanced embeddings (BERT, word2vec) used in Twitter studies, and the lower F1-score for depression (0.67 vs. anxiety’s 0.84) highlights class imbalance issues.  The system worked worse for depression than anxiety, and didn't use important clues like how often people post or what emojis they use. Unlike Twitter research, this work didn't look at how often people post and it also didn't address potential biases. Future work should integrate contextual embeddings, multimodal data.