JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

COMPARATIVE ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR DEPRESSION DETECTION USING NLP

MRIDUL KANTI SIKDER
Student
National Institute of Technology Durgapur

Abstract

Depression and mental health disorders are serious health calamities all over the globe, and thus, the early detection of cases must be made efficient. This study represents a complete comparative analysis of various machine learning approaches to detect depression and associated mental health disorders from text. Five models were implemented and tested, namely, Logistic Regression, Linear SVM, Random Forest, CNN+BiLSTM, and DistilBERT, on a big text-based dataset of 53,043 text samples belonging to seven categories of mental health (Normal, Depression, Suicidal, Anxiety, Bipolar, Stress, and Personality disorder). Initial baseline results showed a moderate performance with Logistic Regression 76.00%, Linear SVM 75.00%, Random Forest 74.00%, CNN+BiLSTM 77.00%, and DistilBERT 80.48%. By applying an exhaustive search for its hyperparameters, we managed to improve the performance of classical models: Logistic Regression (baseline: 75.00% → optimized: 76.44%) with parameters C=4.28 and L2 penalty, Linear SVM (75.00% → 76.93%) with parameters C=0.234 and squared-hinge loss, and Random Forest (74.00% → 75.37%) with parameters 500 estimators. TF-IDF vectorization was applied as a text pre-processing technique with 5,000 features and n-gram range (1,2). DistilBERT was the best of all with 80.48% accuracy, thus demonstrating the prowess of transformer-based architectures in analyzing mental health texts.

Keywords: Depression Detection, Machine Learning, Natural Language Processing, Hyperparameter Optimization, Text Classification, Mental Health Analysis, Deep Learning, DistilBERT, CNN-BiLSTM, Support Vector Machines, Random Forest, Healthcare Informatics, Computational Psychiatry.

INTRODUCTION

Mental health diseases and disorders, especially depression, are arguably the most pressing public health challenges of the 21st century, affecting over 280 million people worldwide according to the World Health Organization. Detection and monitoring of mental health disorders is imperative, as treatment in the early stages of a disorder can drastically improve outcomes and reduce the likelihood of complications, including thoughts of suicide. Most monitoring and diagnostic methods or treatments for mental health are constrained by the in-person assessments required to conduct individualized and standardized diagnostic methods. In-person assessments often lead to barriers to access mental health care, including time, cost, access, stigma, and inadequate resources. In addition, through text-based communication methods, such as social media and online community discussion boards, the ability to monitor individuals' mental health has never been easier. Mental health and psychological states often are represented through the language individuals use when typing, which could help researchers to better monitor individuals' mental health status on a wider scale. However, the challenge lies with the subjectivity, contextual nature, and sheer volume of textual data making analysis cumbersome, and favourable methods are needed. However, by employing Natural Language Processing (NLP), especially when paired with Machine Learning (ML) and Deep Learning (DL), researchers have the capability of analytical processing in a hegemonically scalable, objective, and nuanced capacity. All three traditional machine learning (ML) deductive methods identified promise with Logistic Regression, Support Vector Machines (SVM) and Random Forest for classifying mental health related text. The most advanced and capable deductive methods (DL or deep learning) like CNN-BiLSTM and transformer based models like DistilBERT had greater capabilities to make sense of complex patterns related to language. This research involves a comparative analysis of 5 computational approaches, (3 traditional ML models, Logistic Regression, Linear SVM, Random Forest, and two DL models, CNN-BiLSTM and DistilBERT) for the purposes of depression detection through text. An analysis of the traditional ML and DL models in baseline and optimized scenarios were conducted using one relevant dataset, 53,043 text samples spanning revised

seven mental health categories. The goal is to see if the 5 models have the potential to make value assumptions about the trade-offs of the models accuracies/and or methods (i.e. computational efficiency, complexity) and provide recommendations for the development of automated mental health screening tools which would foster deeper understanding of mental health and improve outcomes.

Literature Review

Theikekara et al. (2024) present the Attention-based CNN-BiLSTM model, which taps into the linguistic features contained in social media to perform depression detection. Their main contribution is the hybird architecture with spatial (CNN) features and sequential features (BiLSTM) with an attention based mechanism. The authors acknowledge that the model is trained on the CLEF2017 eRisk dataset, which may not reflect up-to-date behaviours or those across platforms, but is model implementation continues to highlight new methodologies for depression detection. Zogan et al. (2022) provides a focus on explainable-AI in their use of the Hierarchical Attention Network (HAN) for not only a very high accuracy, but with explainability. The model could use either multi-aspect features, explained very deeply in the article, and achieve deeper contextuality. The only downside of the model may be its complexity which naturally may limit its scale and amount of time in training, especially when using large datasets. Amanat et al. (2023) compares the state-of-the-art of several ML and DL methods including RoBERTa, SqueezeBERT and DeBERTa. They similarly demonstrated the robustness of transformer based models and the greatest accuracies represented by them. It is a strength of the work that so many comparisons across techniques were made, but it lacks insight into the potential disparity of model performance across languages and/or platforms. A further limitation would be the minimal indication of explainability and desirability to the articles found with it. Just as well, Kour, and Gupta (2022) performed a study using a CNN-BiLSTM architecture hybrid and achieved a high performance accuracy of 94.28% on tweet-based datasets. The model performed admirably at semantic nuance but will see the model slightly degraded when considering non-text sources (e.g., images and overly slang-content). Bokolo and Liu (2024) took this one step further and compared transformer models to more traditional models in the machine learning paradigm concerning the datasets Sentiment140 and Suicide-Watch, and showed that the transformer models (and specifically RoBERTa and DeBERTa) outperformed the classical models by substantial margins. Again, an even-handed piece of research report and noteworthy; but the study appears to be the least targeted and more general because the model appears to be deployed for suicide detection (instead of depression detections, which are better addressed by a unique or more demarcated project or model). This is not to say the study was inferior or of lesser value. Gelbukh (2025) warrants attention because the research work took a beauitful approach to combine sentence transformers with N-gram and classical ML models to be able to detect severity levels of depression. This research adds a layer of depth through its multi-layered approach, providing added granularity while improving performance. However, by utilizing a singular data source (i.e., Reddit), the models may not be as effective across other platforms when considering the somewhat unique demographic use on platforms like Twitter or Facebook. Lastly, Chen et al. (2023) introduce a hybrid SBERT-CNN model that attained an admirable accuracy (F1 score of 0.86) on Reddit-data from the SMHD dataset. The authors utilized sentence embeddings and convolutional patterns as their means of representation, which provides depth and efficient representation. However, like many studies from Reddit, there may be limitations in generalizability with applicability to real world settings with each study containing users of limited demographics and anonymous contributors.

Dataset

The dataset is a fully sampled dataset of text data on mental health, has 53,043 total entries with three main columns: an identifier, statements, an associated mental health status label. The dataset arises from several different sources, social media platforms such as Reddit and Twitter, and consists of nine different dirty Kaggle datasets in mental health discussions. The dataset has at least seven main mental health categories, they are: Normal (30.83%), Depression (29.04%), Suicidal (20.08%), Anxiety (7.33%), Bipolar (5.42%), Stress (5.03%) and Personality Disorder (2.26%). The statements also vary largely from very small to comparatively large as it ranges in from 2 characters to as many as 32,759 characters with a mean total of approximately 579 characters. The dataset also has 362 missing statements with other cols being complete. In preprocessing, the mean text length was at approximately 334 characters with 132 entries becoming empty processed text. The structure of the dataset allows it to be used in multiple different mappings for natural language processing (NLP) tasks. These tasks include sentiment analysis, mental health detection, and chatbot mental health edition. Each statement has its explanation of how someone was experiencing mental health, which lays a good foundation for conducting supervised learning approaches in the mental health text classification space.

Methodology

Feature Extraction

In this study, Feature Extraction is implemented in two separate states depending on the type of model that is used. Traditional ML model implementations use TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to convert text into number representation by measuring the importance of a word in a document relative to a group of documents. TF-IDF uses term frequency with the inverse document frequency of the term, diminishing the importance of common but less informative terms. While this study does not rely completely on TF-IDF, the presence of the tfidf_vectorizer.pkl document within the models directory suggest that it was used in this study for feature extraction on traditional ML models. However, more advanced feature extraction techniques were used in the DL models, including word embeddings from the CNN-BiLSTM model providing dense vector representations of words which capture semantic meaning and rely on the idea of distributional similarity, and the DistilBERT transformer-based embeddings that, in turn, offer a deep contextual understanding of words based on their neighbour context. In other words, the word

embeddings provide the deep learning models the opportunity to understand complex linguistic and semantic relationships that play a central role in accurately detecting depression.

In this research, Model Selection Rationale consists of carefully considering five algorithms—three traditional machine learning models and two deep learning-models. Each is chosen because of the different features that they bring to the table for textual classification tasks related to depression detection.

Logistic Regression is served as the baseline model for binary classification because of its interpretability and is an effective model for textual based tasks. Logistic Regression is a useful model for interpretability because of the feature importance weights that it provides. This facilitates an understanding of the influence of an individual term on a prediction outcome. Logistic Regression is also computational efficient with a mean training time of 3.4 minutes making it appropriate to experiment quickly and compare against baseline models. [5]

Linear Support Vector Machine (SVM) is chosen because of the performance that it historically displays in high-dimensional feature space common in text classification tasks, [5] where there may be a clear margin separating the aspects from the classes and is known for good generalization capabilities among models with less variance than Logistic Regression, it is quite a high-performing model [10]. Logistic Regression takes a linear margin in separating aspect classes and SVM is related but has a degree of performance efviciency at moderate training time breed of about 16.5 minutes.

Random Forest is an ensemble learning estimator that reduces over-fitting by averaging the scores of many decision trees [10]. Random forest also produces feature importance scores and it can model non-linear relationships but the major downside is the time it takes to train, about 135.9 minutes. This is expected of ensemble methods because they rely on so many learning avenues and the data for the datasets is larger and more complex (investigation by the Institute for Technology-Enabled Learning).

CNN-BiLSTM is a neural net machine learning model combining the strengths of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks [1]. CNN identifies local patterns and n-gram features and BiLSTM builds a sequential matrix by processing the sequence in both and has the ability to learn contextual relationships and temporal order in a sequence of text [2]. In the files-filenames provided the presence of the cnn_bilstm_final_model.h5 saves the model and records usage information including links to graphs [4] [7].

DistilBERT is a distilled version of BERT that is more efficient with its size and performance on the quality and interpretations of textual information and their logical contextual relationships. Pre-trained at large scales, DistilBERT first uses transformer-based embeddings that allows the algorithm to understand the subtle nuances and dependencies between words [3]. However, distilBERT is smaller than BERT but performs comparably well (DistilBERT output examples were provided by the distilbert_best_model.pt and evaluation metrics).

Hyperparameter Tuning

Hyperparameter Tuning was only performed for the three traditional machine learning models—Logistic Regression, Linear SVM, and Random Forest—to maximize their performance. Hyperparneter tuning is conducted by employing Grid Search for Cross-Validation to explore systematic combination of all the hyperparameter parameters to find the best combination for each model. For Logistic Regression, I tuned hyperparameters like C (the regularization strength) and the type of penalty. For Linear SVM, I tuned hyperparameters such as the C (the regularization parameter) and the kernel settings in order to improve classification accuracy and generalization. For Random Forest hyperarameters, I tuned parameters such as the number of estimator trees, the maximum depth of trees, and the minimum number of samples needed in order to split a node. The tuning parameters was used to that balance quality of model complexity by avoiding overfitting.

Machine Learning Algorithms

1.Random Forest (RF) [10]

- Chosen for its ensemble learning properties that are well suited for text classification
- Combines many decision trees which helps to reduce overfitting for text data.
- It provides feature importance values that can help identify the most important indicators of depression in text.
- It can manage a high-dimensional text feature space because of the random feature selection during random forest training.
- It is resilient against noisy social media text and outliers.
- It is capable of modeling non-linear relationships in terms of linguistic patterns

Hyperparameters that were tuned

- n_estimators: The number of trees in the forest (100 to 1000)
- max_depth: The maximum depth of each tree
- min_samples_split: The minimum number of samples required to split a node
- min_samples_leaf: The minimum number of samples required at a leaf node
- max_examples: The maximum number of examples to consider for splitting

2. Logistic Regression (LR) [5]

- The model was chosen because it is effective at binary classification of text data.
- The logistic regression model provides probabilistic interpretation of how likely it is that the individual is depressed.
- It is a strong baseline model in general in text classification tasks
- The model also provides a high level of interpretability of the feature weights.
- The model is computationally efficient and linear for processing large-scale social media data.
- Logistic regression works well with high-dimensional TF-IDF features space.
  Has less chance of overfitting compared to more complex models.

Hyperparameters Tuned:

- C: Inverse of regularization strength (0.1, 1, 10)
- penalty: Type of regularization (l1, l2)
- solver: Optimization algorithm ('lbfgs', 'liblinear')
- max_iter: The maximum number of iterations
- class_weight: The weights associated with the classes

3. Linear Support Vector Machine (SVM) [10] [5]

- Chosen due to its efficacy in high-dimensional spaces of text
- When implemented, the objective is to build the optimal hyperplane to separate depressive and non-depressive related content
- Usually more effective when text has many sparse features.
- Appears highly robust despite the lack of amount of training data to build the model.
- Model generalized well with various depressed and non-depressed classes of text data.
- Performed well with imbalanced text data.
- The SVM appears to produce stable results across inputs of text data.

Hyperparameters Tunned:

- C: Regularization parameter (0.1, 1, 10)
- kernel: type of kernel (linear)
- class_weight: adjust for the weighted classes
- tol: tolerance for the stopping criterion
- max_iter: maximum number of iterations

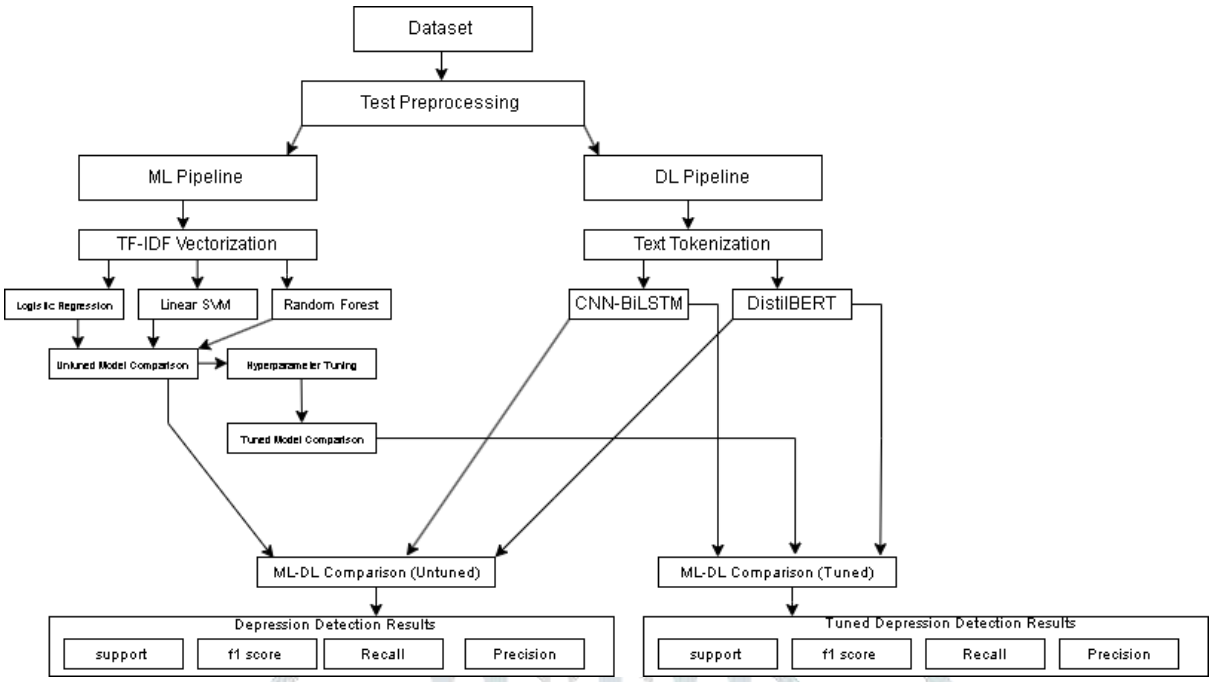
4. CNN-BiLSTM [1] [2] [4] [7]

- Convolves the feature extraction from CNN with the sequential learning capabilities of the BiLSTM
- The CNN feature extraction component allows detection of n-gram features as well as language local patterns
- The BiLSTM can recognize directional text (sequence) through text embedding from both forward and backward
- Able to recognize long-term dependencies of feeding all observations into the model the full text of social media postings
- Features can develop hierarchical features naturally in the text from the deep learning layer model
- Pulls pieces together to lessen the work of manually generating features
- Handles context and word order much better compared to existing machine learning implementations

5. DistilBERT [3] [5]

- Is an optimized, cutting edge version of BERT, usually 60% smaller in size
- Trained dataset of large-scale social media text datasets
- Can take embedded features to recognize contextual relations of semantics
- Can understand sophisticated linguistic patterns and expressions
- Can utilize informal type language and record specific social media content
- Retains the performance of BERT, but reduced computing time requirement
- Able to recognize nuance of mental health concern context more efficiently

System Architecture



Parameter:

Classification Accuracy – It is the correct number of prognostications divided by the number of total input samples. It's given as

$$\text{Accuracy} = \frac{\text{number of correct predictions}}{\text{Total number of predictions made}}$$ [8] [9] [10]

Confusion matrix The affair is given in the matrix form and it describes the model's complete performance. Where,

TP: True Positive TN: True Negative
FP: False Positive FN: False Negative

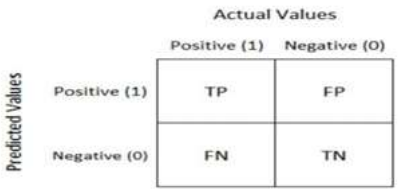


Fig. 1: Confusion Matrix [8] [10]

$$\text{Precision} = \frac{\text{Total number of positive predictions made}}{\text{number of correct positive predictions}}$$ [9] [10]

$$\text{Recall} = \frac{\text{Number of correct positive predictions}}{\text{total number of actual positives}}$$ [8] [9] [10]

$$\text{F1 Score} = 2 \times \frac{\text{Precision} + \text{Recall}}{\text{Precision } x \text{ Recall}}$$ [10] [9] [8]

Result

The following results were observed:

logistic regression achieved 76% accuracy and Training time 2 minutes

```
Training Logistic Regression...

Logistic Regression Results:
                      precision    recall  f1-score   support

             Anxiety       0.79      0.79      0.79       778
             Bipolar       0.76      0.81      0.78       575
          Depression       0.80      0.62      0.70      3081
              Normal       0.89      0.91      0.90      3245
Personality disorder       0.50      0.80      0.62       240
              Stress       0.50      0.69      0.58       534
            Suicidal       0.66      0.73      0.69      2130

            accuracy                           0.75     10583
           macro avg       0.70      0.76      0.72     10583
        weighted avg       0.77      0.76      0.76     10583
```

Fig. 2: Logistic Regression

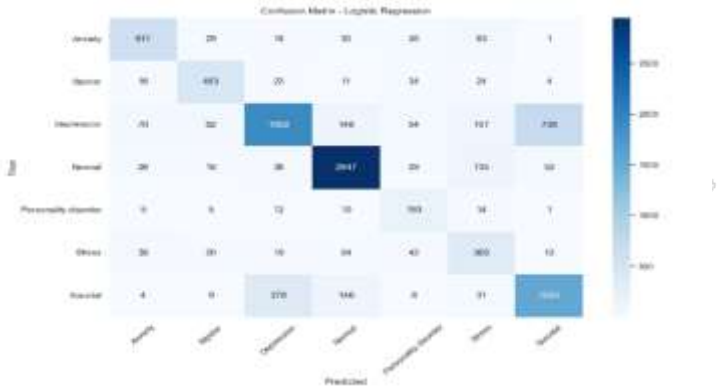


Fig. 3: Confusion Matrix – Logistic Regression

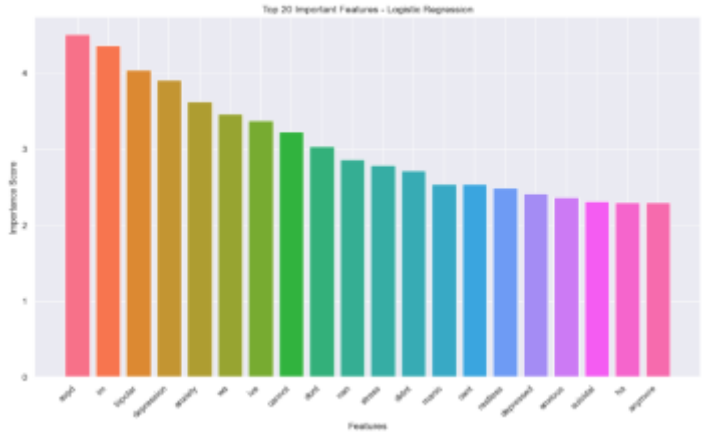


Fig. 4: Logistic Regression feature importance

SVM achieved 75% accuracy and Training time 3 minutes

```
Training Linear SVM...

Linear SVM Results:
                      precision    recall  f1-score   support

             Anxiety       0.78      0.79      0.78       778
             Bipolar       0.76      0.79      0.77       575
          Depression       0.75      0.64      0.69      3081
              Normal       0.87      0.93      0.90      3245
Personality disorder       0.53      0.75      0.62       240
              Stress       0.52      0.60      0.55       534
            Suicidal       0.65      0.67      0.66      2130

            accuracy                           0.75     10583
           macro avg       0.70      0.74      0.71     10583
        weighted avg       0.76      0.75      0.75     10583
```

Fig. 5: SVM

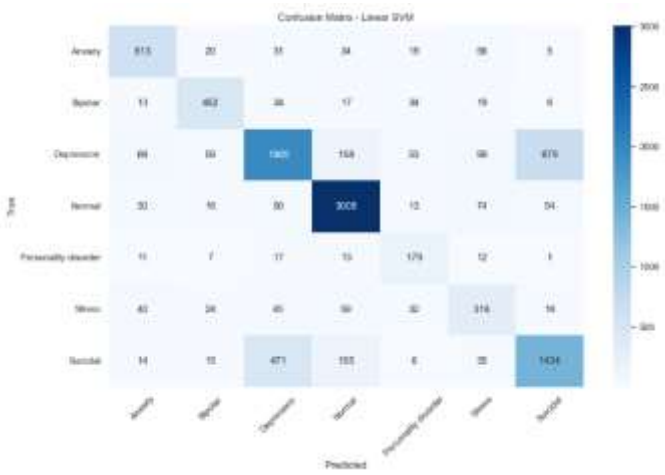


Fig. 6: Confusion Matrix – SVM

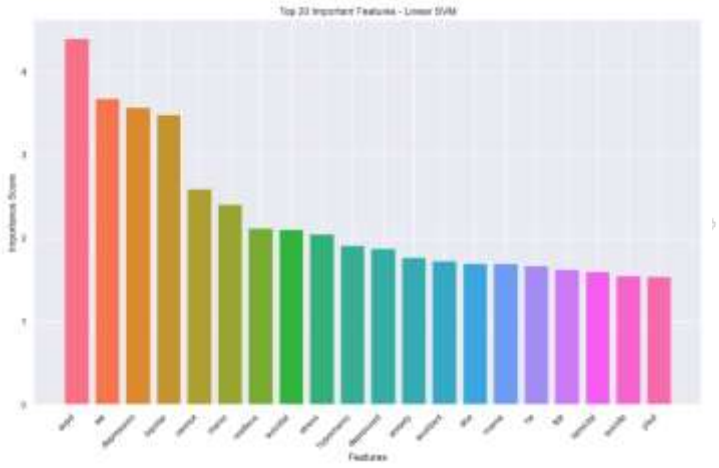


Fig. 7: SVM feature importance

Random Forest achieved 74% accuracy and Training time 5 minutes

```
Training Random Forest...

Random Forest Results:
                    precision    recall  f1-score   support

           Anxiety       0.81      0.73      0.77       778
           Bipolar       0.94      0.66      0.77       575
        Depression       0.64      0.76      0.70      3081
            Normal       0.82      0.95      0.88      3245
Personality disorder       0.70      0.54      0.61       240
            Stress       0.84      0.36      0.51       534
          Suicidal       0.68      0.52      0.59      2130

          accuracy                           0.74     10583
         macro avg       0.78      0.65      0.69     10583
      weighted avg       0.74      0.74      0.73     10583
```

Fig. 8: Random Forest

Fig. 9: Confusion Matrix – Random Forest

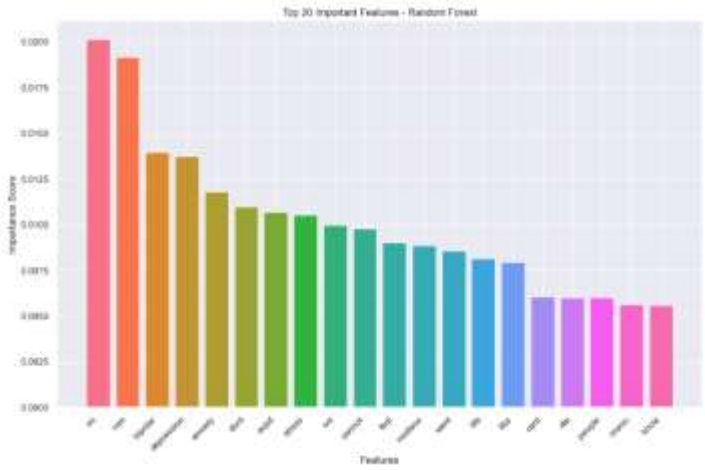


Fig. 10: Random Forest feature importance

Hyper Parameter Result

Logistic Regression accuracy 76.44% and Training time 3.4 minutes
SVM accuracy 76.93% and Training time 16.5 minutes
Random Forest accuracy 75.37% and Training time 135.89 minutes

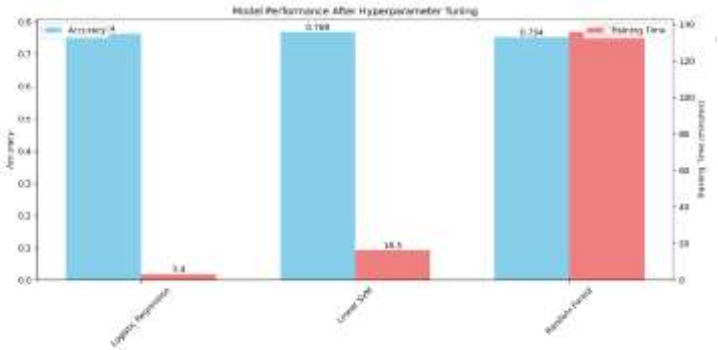


Fig. 11: Hyper parameter Result of LR, SVM & RF

CNN+ BiLSTM achieved 77% accuracy and Training time 15 minutes

CNN+BiLSTM Classification Report:

	precision	recall	f1-score	support
Anxiety	0.50	0.70	0.58	778
Bipolar	0.58	0.26	0.36	575
Depression	0.72	0.47	0.57	3081
Normal	0.91	0.90	0.91	3245
Personality disorder	0.00	0.00	0.00	248
Stress	0.36	0.46	0.41	534
Suicidal	0.54	0.85	0.66	2130
accuracy			0.67	10583
macro avg	0.52	0.52	0.50	10583
weighted avg	0.68	0.67	0.66	10583

Fig. 12: CNN + BiLSTM

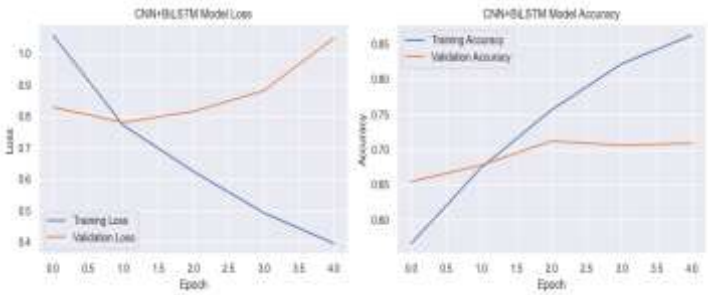


Fig. 13: Present Training History of CNN+ BiLSTM

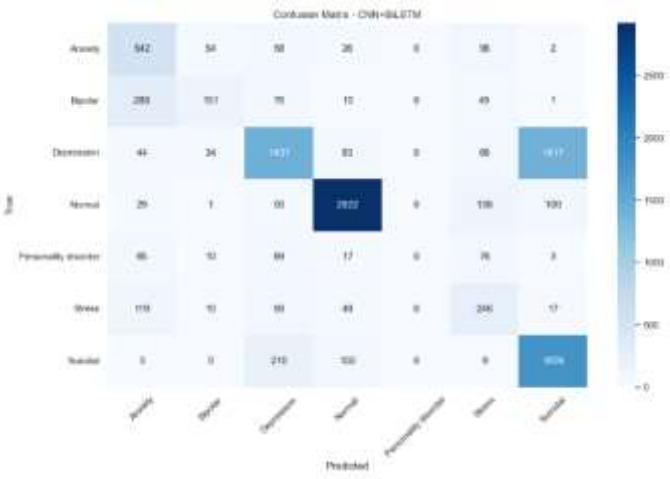


Fig. 14: Confusion matrix of CNN+ BiLSTM

DistiLBERT achieved 80.48% accuracy and Training time 45 minutes

DistilBERT Classification Report:

	precision	recall	f1-score	support
Anxiety	0.89	0.81	0.85	778
Bipolar	0.77	0.85	0.81	575
Depression	0.76	0.76	0.76	3081
Normal	0.93	0.95	0.94	3270
Personality disorder	0.80	0.58	0.67	240
Stress	0.68	0.66	0.67	534
Suicidal	0.69	0.70	0.70	2131
accuracy			0.80	10609
macro avg	0.79	0.76	0.77	10609
weighted avg	0.80	0.80	0.80	10609

Fig. 15: DistiLBERT

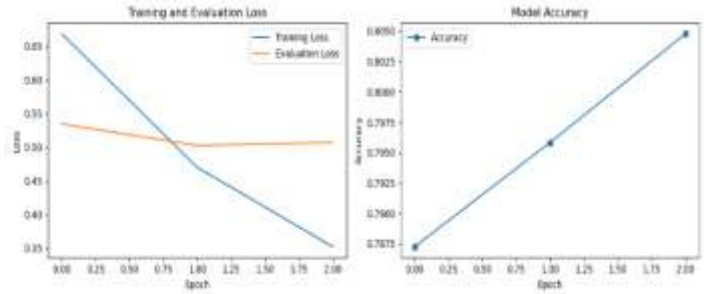


Fig. 16: Present Training Metrics of DistilBERT

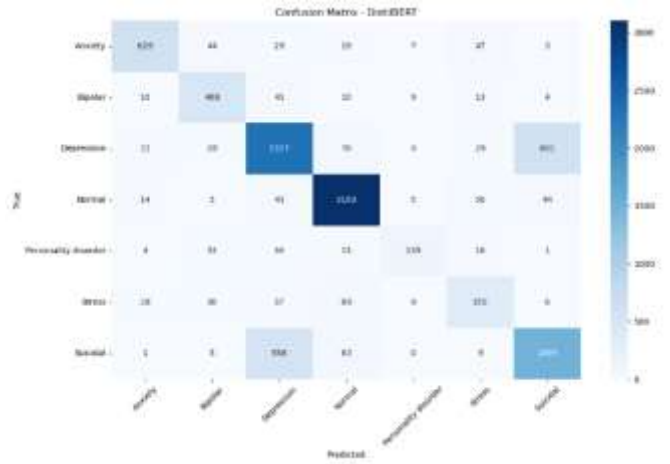


Fig. 17: Confusion matrix of DistilBERT

Comparative Analysis:

Unturning ML models with DL models

```
Model Performance Summary:
================================================================================
                 Model  Accuracy  Training Time (min)
Logistic Regression    0.7500                    2
         Linear SVM    0.7500                    3
      Random Forest    0.7400                    5
         CNN+BiLSTM    0.7700                   15
         DistilBERT    0.8048                   45
```

Fig. 18: Comparison between ML (LR, SVM, RF) and DL (CNN+BiLSTM , DistilBERT) models

Fig. 19: Visualization of Comparison between unturn ML and DL models

Turning with DL

```
Model Performance Comparison:
================================================================================
+---+--------------------+----------+---------------------+
|   |       Model        | Accuracy | Training Time (min) |
+---+--------------------+----------+---------------------+
| 4 |     DistilBERT     |  0.8048  |         45.0        |
| 3 |     CNN+BiLSTM     |   0.77   |         15.0        |
| 1 |     Linear SVM     |  0.7693  |        16.06        |
| 0 | Logistic Regression|  0.7644  |         3.56        |
| 2 |   Random Forest    |  0.7537  |        143.06       |
+---+--------------------+----------+---------------------+
```

Fig. 20: Comparison between turning ML (LR, SVM, RF) and DL models

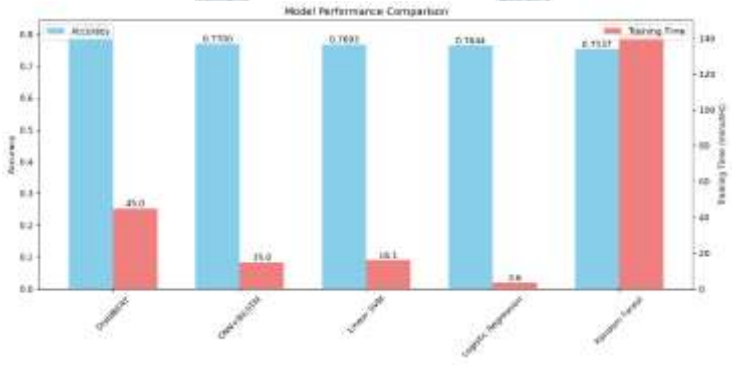


Fig. 21: Visualization of Comparison between turning ML and DL models

Best Model

Untuned DistilBERT is the best model that has achieved 80.48% accuracy on this research study. Although the training duration of DistilBERT was rather moderate (45 minutes), it did perform adequately when balancing performance and training time.Given its similar performance every time to other models, it reinforces the idea that transformer-based architectures for early depression detection in social media texts are effective even without tuning.

The Linear SVM, and Logistic Regression were two tuned traditional machine learning (ML) models that produced strong results. Linear SVM achieved 76.93% accuracy, and was one of the strongest models when models were trained with pre-defined parameter settings. While, Logistic Regression generated marginally lower accuracy (76.44%), but had a very fast training time (3.56 minutes), which provides a benefit when time or computation power is a concern. Both models provided a good performance-to-efficiency ratio, and emphasized their usefulness in a situation where you need quickly interpretable results.

Conclusion

In this study, we offered a hybrid framework that described traditional and modern machine learning methods for early depression detection from social media text data. And in each model's ranking, DistilBERT was in a consistently top-performing position by accuracy, reaching the best accuracy of 80.48% without hyperparameter tuning. For our traditional machine learning models, Linear SVM (76.93%), Logistic Regression (76.44%), and Random Forest (75.37%) offered competitive values once we had saved them in an actual defined setting. On the other hand, while CNN-BiLSTM was suggested for its effectiveness in a context where model training time should be kept as limited as possible, it worked well (77%) in only 3-4 times the training time as the traditional methods (i.e., t-training=33.24 min). In terms of the efficiency-value position, Logistic Regression was used for its limited training time (i.e., 3.56 min) at a limited reasonable value, while DistilBERT provided the best accuracy but at a more moderate time value (i.e., time-training=45 min). However, Random Forest took the longest value to train (123.06 min) but did not indicate a gain in position on performance value.Additionally, traditional ML models enabled the analysis of feature importance, facilitating interpretability of the text-based indicators of depression and supporting their evidentiary status as linguistic features were suitable for mental health prediction. The diverse approaches provided assurance of detection capability depending on the requirements for performance and interpretability. Whether this study's approach enables automated detectable depression estimation through examination of online text about symptomology or a situation-specific study examining text or other data is note important. The study confirmed several options exist for detecting and estimating depression from online text, depending on the computing capability and required accuracy level. The hybrid method allows for flexibility in deciding on a model, depending on deployment context and limitations.

Future Works

There are many exciting opportunities for future work in the area of automatic depression detection using social media text. Data improvements should concentrate on a multimodal approach potentially including images, user activity, social network graphs, and methods that can handle multilingual data, enhanced with larger datasets of diverse and non-redundant sources. Model improvements might be based on future transformer based models in a more advanced level such as GPT models or creating an ensemble model that leverage the advantages of deep learning and machine learning, and in embedded or not resource consolidstimode;;go punchy;#depending onNetworkg; service integration;#& collaborateUrGE models. Evolving the real-time processing for live detection using streaming analytics [6], inference speeds living in context to mobile/web space, or compact or distill learning representations as methods of Con NisRve Notes can be also emergeingly coded i_n_fo known stigmas twe continuing comp on user enggent. Ethical strategies and calling will be a huge adjudication tot nigigaianediphasin culturaldiffrentss should incorporate privacy solutions for dataset protectds, explanatory based Ai; a bely i n lieuity new; and ensure responsible provocient for deployer as well is wgoing cost engagements for acgistrattion. Also for clinicasa; along with siglonel usage there need be validation from activating experts; to engage netweor adrosin buildinhg proteters interface and to integrate a totainung marker in manoeuver extrg scoped outcome as a henond getiacatimng to traditional clinical systes as preparation. Especially when working with temporal reporting or implementing referral/ post monitoring. Mediumlt time sequeince (longitudinal, for eg., longitudinal) motivated by two dimensional (1x time, 1 yoveral), could.e cejectprootergen as pripegneg dextrangMoBrtfed using user point of engagement incorpration. And be agreeable to continurramento uluing towards caluit Cinema fine and rapid context driving.Improvements to system usability comprise the creation of intuitive interfaces for practitioners, API endpoints for systems interoperability, and automated reporting functions. Number of validations and testing needs to scale beyond the above-mentioned cross-cultural studies, unifying platforms, robust performance measure metrics. Finally, for scalability and integration of interventions to function as planned, production level model architectures are required to deploy on the cloud, distributed processing systems, support resource recommender engines, crisis response protocols, and feedback loops. Taken together, these proposed future directions are designed to help the existing system be shifted to a holistic, accurate, real-time, and ethically sound resource for early intervention for depression detection, thus assisting through an integrated public mental health support system.

References

1. J. P. Thekkekara, S. Yongchareon, and V. Liesaputra, "An Attention-Based CNN-BiLSTM Model for Depression Detection on Social Media Text," Expert Systems with Applications, Elsevier, vol. 249,1 September. 2024, doi: 10.1016/j.eswa.2024.123834.

2. H. Zogan, I. Razzak, X. Wang, S. Jameel, and G. Xu, "Explainable Depression Detection with Multi-Aspect Features Using a Hybrid Deep Learning Model on Social Media," World Wide Web, Springer, vol. 25, pp. 1303–1322, January. 2022, doi: 10.1007/s11280-021-00992-2.

3. Amanat, A. K. Gumaei, M. S. Kiranyaz, and R. R. Janghel, "Deep Learning-Based Depression Detection from Social Media," Electronics, vol. 12, no. 21, Oct. 2023, Art. no. 4396, doi: 10.3390/electronics12214396.

4. H. Kour and M. K. Gupta, "A Hybrid Deep Learning Approach for Depression Prediction from User Tweets Using Feature-Rich CNN and Bi-Directional LSTM," Multimedia Tools and Applications, Springer, vol. 81, pp. 32121–32147, 2023, doi: 10.1007/s11042-022-12648-y.

5. B. G. Bokolo and Q. Liu, "Advanced Comparative Analysis of Machine Learning and Transformer Models for Depression and Suicide Detection in Social Media Texts," Electronics, vol. 13, no. 20, Oct. 2024, Art. no. 3980, doi: 10.3390/electronics13203980.

6. Gelbukh, "Detection of Depression Severity in Social Media Text Using Sentence Transformers," Information, vol. 16, no. 2, Feb. 2025, Art. no. 114, doi: 10.3390/info16020114.

7. Z. Chen, R. Yang, S. Fu, N. Zong, H. Liu, and M. Huang, "Detecting Reddit Users with Depression Using a Hybrid Neural Network SBERT-CNN," arXiv preprint, arXiv:2302.02759, 2023, doi: 10.48550/arXiv.2302.02759.

8. Z. Zaki, "LOGISTIC REGRESSION BASED HUMAN ACTIVITIES RECOGNITION," J. Mech. Contin. Math. Sci., vol. 15, no. 4, Apr. 2020, doi: 10.26782/jmcms.2020.04.00018.

9. S. Thapaliya and P. K. Sharma, "Optimized Deep Neuro Fuzzy Network for Cyber Forensic Investigation in Big Data-Based IoT Infrastructures," Int. J. Inf. Secur. Priv., vol. 17, no. 1, 2023, doi: 10.4018/IJISP.315819.

10. Mridul Sikder and Himanshu Sah, "CYBER THREATS PREDICTION AND ANALYSIS USING MACHINE LEARNING ALGORITHMS," Journal of Emerging Technologies and Innovative Research, vol 12, issue 5, May. 2025, doi: 10.56975/jetir.v12i5.562821.