Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 4s (2025)

Implementing Sentiment Analysis with Applied Nonlinear Analysis: Predicting Mental Disorders from Diverse Online Social Network Datasets

Mr Sushant.A.Patinge[1] and Dr. Vijaya.K.Shandilya[2]
[1,2] Sipna College of Engineering & Tech. Amravati, Maharashtra, India
[1]spatinge@gmail.com

Article History:

Received: *15-10-2024*

Revised: *24-11-2024*

Accepted: *03-12-2024*

Abstract:

Introduction: WHO projected that the depression disorder will increase over the next two decades; therefore, it should be diagnosed early and the sooner it can be detected and prevented. Several machine learning algorithms that may predict depression will be compared in this paper: CNN, LSTM, Bidirectional LSTM, Naive Bayes, Logistic Regression, Random Forest, AdaBoost, and Support Vector Machine. The results show that the models of deep learning, mainly variants of CNN and LSTM, exhibit high accuracy, and traditional classifiers perform exceptionally well. These results highlight how machine learning can contribute towards developing useful tools for the prevention of mental health disorders early in life. In modern life, online social platforms have very subtly integrated into the daily routine of a significant proportion of global citizens.

Keywords: Depression, Machine Learning, CNN, LSTM, Early Detection.

1. Introduction

Growth of social media has dramatically transformed the look and pattern of modern communication and connection around the world, with its users estimated to have shot from 2.46 billion in 2017 to 3.02 billion in 2021. Such growth underlines how deep such places are for people in life, with places exchanging things ranging from personal updates and experiences up to contacts and conversation over issues involving the world. Since it increased connectivity and sharing of information, the more complex issues, particularly psychological one, was established. One such is the so-called "Social Network Mental Disturbance" that consolidates all those bad effects in psychology that people face for excessive exposure to social networking sites. These include higher degrees of anxiety and depression along with other mental problems that are frequently and incessantly caused by online relations.

The rising use of social media has, therefore led to the realization of the urgent need to confront the mental implications of internet interaction process. With the inclusion of digital systems in modern life, their associated mental risks have to be looked at closely and attended to. The necessary kind of interdisciplinary research that takes into consideration the pervasive influence of social media on emotional well-being needs to be carried out to understand such challenges. This further opened avenues for researchers and practitioners of mental health to target early indicators for psychological problems and ways through which risks may be actively averted. The current research will try filling this gap by putting forward a framework for mental health disorders detection through content of social media, thus giving room for prevention of mental health disorders.

This research attempts to find methodologies for using advanced data analytics to identify potential symptoms of mental disorders through online social interactions. This will use state-of-the-art machine learning and deep learning techniques to detect subtle behavioral and linguistic patterns that indicate psychological vulnerabilities. This would leave the final aim of coming up with risk-predicting models that can predict the risks of the individuals, hence giving the opportunities for timely intervention sets. This is well aligned with the broader objectives of increasing mental health awareness and utilizing technology to address one of the most significant challenges facing the digital ages.

Proposed Work

The proposed work is about developing a predictive framework that integrates advanced machine learning and deep learning methods to identify patterns indicating mental health concerns in social media users. The methodology begins with systematic data collection from various platforms such as Twitter and Reddit, encompassing textual posts, comments, user interactions (e.g., likes, shares, retweets), and network connections. At this stage, anonymity of the user and strict observance of all laws related to privacy are ensured to adhere to legal and ethical standards. Cleaning the raw input includes removal of special characters, URLs, and stopwords as noises. Other procedures also include tokenization and lemmatization processes, and handling missing data to structure data in a way that analysis will be easier. In this step, features include natural language process, network analysis, as well as behavioral modeling applied to data extracts so as to set the extractions of features. Moving by this, NLP-based methods, such as sentiment analysis, word embeddings that contain models like Word2Vec, GloVe, and BERT can be capable enough to create emotional tone and semantic linkage representing the text information about the early warns relating to mental illness. Techniques of network analysis simultaneously build social graphs, which capture user interactions; metrics like centrality and clustering coefficients give insights into group dynamics and user influence. This framework based on behavioral modeling will be taken into consideration to explain the temporal analysis and interaction patterns like posting frequency, activity timing, and response behaviors that might indicate anomalies showing signs of psychological distress.

Hybrid Predictive Model Architecture

Hybrid Architecture for Predictive Model - Deep Learning + Traditional ML techniques, Using RNNs, LSTMs to comprehend the complicated nature of time and sequences of mental health; it detects the pattern in the complexity of words and phrases which CNNs are supposed to do, but it is very hard to concentrate attention on these using critical input features. Hybridization of traditional models like Random Forests and SVMs gives this system an additional benefit in the form of accuracy and interpretability in classification. Thus, the framework could provide an overall understanding of the behavior and linguistic patterns of the users, thus constituting a strong tool for early mental health detection. Cross-validation strategies include k-fold cross-validation. Validating the efficiency of the proposed model includes all measures like accuracy, precision, recall, F1 score, and AUC-ROC.

2. Methodology

This paper introduces a systematic approach to predicting mental disorders using advanced deep learning and machine learning techniques applied to various data sets that have been collected from online social networks. The ultimate goal is for an inclusive system that might measure, gauge, and assess a person's mental health conditions based solely on online behavior. It entails a methodology that integrates natural language processing, network analysis, and behavioral modeling in giving a wide-angled view of the complexities of the psyche in the digital ecosystem.
The methodology begins with collecting user-generated multilingual content from all kinds of social media such as Twitter, Reddit, and more. This is the most important step because the model needs to generalize to all kinds of languages and cultures. Collected data in the format of a CSV file undergoes EDA, which identifies any kind of patterns, distribution, or possible anomalies that could exist. EDA also detects missing data, and techniques for handling are applied to that. Text data is preprocessed with techniques like tokenization, removal of stopwords, stemming, and lemmatization to transform the raw input into a structured format. Then, features are extracted using methods like TF-IDF and word embeddings to represent textual data in a numeric format for easy analysis. All of these steps enable building of robust foundations that develop prediction models.

Preprocessed data are further divided into two subsets that represent the training and testing. This usually happens in a ratio of 80 to 20 while judging how much generalization actually occurs over the unseen data samples. Applications done are in the form of Machine learning as well as Deep learning algorithms applied to these subsets: Logistic regression, SVM, Random Forest, Artificial neural networks, and LSTMs. These models are selected based on their complementary strengths. The classical ML models like SVM and Random Forest provide a good baseline performance, besides being very interpretable in their decision-making process. However, DL models such as LSTMs can capture extremely complex temporal patterns. All of the models are trained and optimized over the evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The best performing model is serialized and deployed as a web service with Flask and similar frameworks, which allows real-time prediction for practical applications.

This research also encompasses several algorithms that, in itself, deals with different aspects of the problem at hand. CNNs have been used for feature data extraction especially when there is text evidence with spatial or composition-based patterns of writing related to psychological illnesses. It has more advanced variants, which is Bidirectional LSTMs and LSTM complexes. That will cover time dependencies, which involve the capturing of both forward and backward contexts in data samples. To ensure its robustness and increase its classification accuracy, it involves probabilistic models, including Naive Bayes and ensemble techniques, such as Random Forests and AdaBoost. SVM is used on the basis of its aptness in binary classification, particularly in a high-dimensional feature space for the process.

This research aims at attaining some of the main outcomes and achieves some of its key objectives by providing a multilingual data classification system. It builds up towards a greater accuracy rate for the detection of mental disorders on online platforms. This research is rich in fetching multimedia as well as textual content by means of numerous NLP techniques, hence realising an integrated user's mental state. Finally, it is including state-of-the-art deep learning-based classification models to predict mental illnesses using datasets that are diverse from each other. Finally, the analysis conducted in this work is in comparison with models proposed to present their relative effectiveness and to identify areas requiring improvement based on the analysis. This study contributes significantly to proactive mental health intervention in the digital age by taking the concepts of machine learning from ivory towers out into the world.

3. Result

The two tables in this section demonstrate the performance metrics of different algorithms used within the study to predict mental disorders based on online social network data samples. These metrics include accuracy, error rate, precision, recall, and F1 Score, all of which are quite important to assess the performance of each model. The proposed algorithms have been categorized into deep learning models and traditional machine learning models. A detailed explanation and comparison of the algorithms mentioned against these metrics are given below

Table 5.1: Performance Metrics of Deep Learning Algorithms

Sr No	Algorithm	Accuracy	Error Rate	Precision	Recall	F1 Score
1	CNN Complex	83.55	16.44	83.70	83.55	83.62
2	LSTM	82.87	17.12	83.03	82.87	82.88
3	LSTM Complex	81.75	18.24	81.83	81.75	81.75
4	Bidirectional LSTM	82.82	17.17	82.95	82.82	82.82
5	CNN	82.61	17.39	82.76	82.61	82.61

Table 5.1 Summary Table for various performance metrics of deep learning algorithms, in relation to applying the above on mental disorder predictions. Inclusion of CNN Complex, LSTM, LSTM Complex, Bidirectional LSTM, CNN is present with metrics being, Accuracy, Error Rate, Precision, Recall, and F1 Score. Among the models, it was the CNN Complex that came out to be the best performer, realizing the highest accuracy of 83.55%, with precision and recall at 83.70% and 83.55%, respectively. Also, its F1 score of 83.62 indicated an optimal balance between both precision and recall, ensuring it to be highly capable of identifying mental health conditions effectively. The LSTM model closely followed this one with 82.88%, supported by good precision, recall, but more importantly F1 Score at 82.88%, which is in fact what has captured richly temporal patterns for sequential data. Similarly, the Bidirectional LSTM scored 82.83% for accuracy by using forward and backward sequence processing to attain precision and a recall score of 82.95% and 82.83%, respectively, in building a strong F1 score of 82.83%. The LST Complex was not that accurate as the simpler version but attained 81.76% accuracy with precision and recall around 81.8%, and an F1 Score of 81.75% indicating room for further improvement. Finally, the simple CNN model performed with reliable results: it had an accuracy of 82.61% with precision, recall, and F1 Score all at 82.61% making it an effective yet lesser complex model as compared to CNN Complex. Such results unveil that deep learning algorithms are promising predictors for mental disorders, where CNN Complex exhibits the best performance in terms of overall results because of the possibility of extracting complex hierarchical features, while LSTM-based models show robust handling of dependencies.

Table 5.2: Performance Metrics of Traditional Machine Learning Algorithms

Sr No	Algorithm	Accuracy	Error Rate	Precision	Recall	F1 Score
1	Naive Bayes_cv	66.94	33.05	66.76	66.94	65.71
2	Logistic Regression_cv	77.72	22.27	77.46	77.72	77.58
3	Random Forest Classifier_cv	62.52	37.47	65.31	62.52	58.29
4	AdaBoost Classifier_cv	73.66	26.33	74.20	73.66	72.57
5	Support Vector Machine_cv	78.16	21.83	77.66	78.16	77.59

Table 5.2 Comparison of traditional machine learning algorithms using metrics such as Accuracy, Precision, Recall, and F1 Score for Naive Bayes, Logistic Regression, Random Forest, AdaBoost, and Support Vector Machine SVM Performance: The best performance in terms of accuracy was from SVM with an accuracy of 78.16% and closely aligned precision and recall of 77.66% and 78.16% respectively, giving a robust F1 Score of 77.60%. This performance highlights SVM's robustness in handling high-dimensional feature spaces. Logistic Regression follows closely with an accuracy of 77.73% and balanced precision and recall around 77.5%, yielding an F1 Score of 77.58%, making it a dependable choice for datasets with linear separations. AdaBoost performs with moderate success with an accuracy of 73.66%, precision of 74.21%, and F1 Score of 72.58%. It therefore tends to amplify weak classifiers, but a bit less than generalizing SVM and Logistic Regression. Whereas Naive Bayes and Random Forest underachieve against other models. The accuracy achieved by Naive Bayes is the lowest at 66.95% with precision and recall both approximately around 66.77% and an F1 Score of 65.72% that indicates the weakness due to the conditional independence assumption. The results of Random Forest were 62.53% for accuracy, precision is at 65.31%, and an F1 Score of 58.30%, which indicates that overfitting is a problem and the data has very high dimensions in this process. Thus, it can be said that traditional algorithms like SVM and logistic regression work well but deep learning has shown better performance in such complex datasets where the variables' relationship is nonlinear as well, and it is a crosswise interaction process.

1. Comparative Analysis

Deep learning models tend to perform better than traditional machine learning models on all the metrics. CNN Complex outperforms others in terms of accuracy and F1 Score. This clearly shows that deep learning models perform better when the data patterns are complex, especially when dealing with high-dimensional and sequential data.

Accuracy: CNN Complex has the highest accuracy at 83.55%, whereas SVM is leading in the traditional models at 78.16%. Deep learning models are generally more accurate than traditional models. It may be because they can spot more complex patterns in larger datasets that may not be possible with traditional models.

Error Rate: Traditional machine learning models tend to have higher error rates compared to deep learning models, which points out the potential limitations of traditional models in dealing with the complexity and variability of social media data. Error rates inversely mirror accuracy scores, and as might be expected, CNN Complex has the lowest error rate at 16.45%. Precision and Recall: On average, the deep learning models generally do better in terms of precision and recall when compared to the traditional models. However, CNN Complex outperforms again marginally. This indicates that there is the potential for these models to classify positive cases even more effectively while reducing false positives. For the traditional model, the SVM is able to provide the best trade-off between precision and recall.

F1 Score: The F1 score, like recall and precision, provides an averaged measure of the strength of a model. CNN Complex is the best of these four, with an F1 score of 83.62%. Among the baseline models, SVM was a clear winner at 77.60%, followed by logistic regression (77.58%).

In all metrics comparison between deep learning and traditional models for mental disorder prediction, the former outperformed the latter. Among all traditional models trained, CNN Complex proved to outperform the others in each metric: the highest in accuracy at 83.55%, the lowest in error rate at 16.45%, and best F1 Score at 83.62%. Other traditional models are SVM, leading among all traditional models at an accuracy of 78.16% with an F1 Score of 77.60%, thus still behind the performance of models using deep learning.

2. Conclusion and Future Scope:

This promises huge potential for the further progress of research in mental health with this application of techniques in deep learning and machine learning for the prediction of mental disorders from online social networks and also underlines the prospect of being able to use the great amount of user-generated rich content of social media for supporting and intervening early for a population at risk. These models are shown to have excellent predictability but require interdisciplinary cooperation and careful attention to ethical issues, such as data privacy and informed consent. Interpreting the models is still very challenging, as well as the dynamic nature of social networks; however, this approach will be a crucial step toward bettering global mental health and responsible and ethical use of the field as it evolves. The future scope for deep learning and machine learning in the prediction of mental disorders by the use of diverse datasets of online social networks is massive, with the potential scope of advancements in algorithm development, multimodal data integration, real-time monitoring, personalized interventions, and addressing bias and fairness.

Such innovations will make integrating various forms of data such as text, images, and audio boost the predictive precision and better understand mental health. Further, the generative AI can be used in synthesizing training data and modeling complicated interactions and producing unique content in interventions, therefore making capabilities higher for predictive models. Such approaches demand rigorous validation, ethical considerations, and collaboration with mental health professionals so that they may have relevance for clinical practice and for users once they are incorporated into the healthcare system. This is a future vision for potential transformation in mental health care, but an ethical process should be followed to make sure it is responsibly implemented for the process.

References

[1] Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68.

[2] Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 25(1), 79-93.

[3] O'Reilly, M., Dogra, N., Whiteman, N., Hughes, J., Eruyar, S., & Reilly, P. (2018). Is social media bad for mental health and well-being? Exploring the perspectives of adolescents. *Clinical Child Psychology and Psychiatry*, 23(4), 601-613.

[4] Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. Y., Rosen, D., Colditz, J. B., Radovic, A., & Miller, E. (2017). Social media use and perceived social isolation among young adults in the U.S. *American Journal of Preventive Medicine*, 53(1), 1-8.

[5] Statista. (2018). Number of social media users worldwide from 2010 to 2021. *Statista*. Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

[6] Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society*, 4(1), 2056305118763366.

[7] Zhang, J., & Wallace, B. C. (2017). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*.

[8] Medhat, A., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.

[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186).

[10] Newman, M. E. (2018). Networks. *Oxford University Press*.

[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

[12] Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

[13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

[14] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

[15] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (Vol. 14, No. 2, pp. 1137-1145).

[16] Choudhury, M. D., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media* (pp. 128-137).

[17] Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing*. Pearson.

[18] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

[19] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[20] Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media, Inc.

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

[22] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.

[23] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

[24] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

[25] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6), 602-610.

[26] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[27] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681.

[28] McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. *AAAI-98 workshop on learning for text categorization*, 752(1), 41-48.

[29] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

[30] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.

[31] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.