

# An End-to-End Machine Learning Solution for Anxiety and Depressive Disorder Symptom Occurrence During COVID-19: A New York Case Study

Nikhil Yadav  
St. John's University  
[yadavn@stjohns.edu](mailto:yadavn@stjohns.edu)

Christopher Singh  
St. John's University  
[christophersingh86@outlook.com](mailto:christophersingh86@outlook.com)

Kajal Srivastava  
Jaipuria Institute of Management  
[kajal.srivastava@jaipuria.ac.in](mailto:kajal.srivastava@jaipuria.ac.in)

## Abstract

*Anxiety and depression during the COVID-19 pandemic have heightened as evidenced by the rapidly growing corpus of research articles suggesting a link between the pandemic and mental health. This paper proposes a unique end-to-end user-centric machine learning (ML) architecture, capable of assessing the quality of ML predictions about the occurrence of anxiety and/or depression symptoms. A case study is presented using official New York State COVID-19 data, highlighting the plug-and-play capabilities of this architecture for both external features, and newer ML models. This is demonstrated through the formal design of a custom weighted clustering algorithm which outperforms conventional unsupervised techniques in grouping symptomatic cases. The ability to augment external sentiment data mined from social media platforms like Twitter, increases the predictive power of this architecture. This work serves as a blueprint to build a practical ML solution to better gauge the effect of future pandemic waves on mental health.*

## 1. Introduction

Anxiety and depressive disorders are very prevalent in the US and affect close to 29.5% of the US population [1, 2]. In 2017, the Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration reported that roughly 20% of the US adult population had been affected by a mental illness [3]. Anxiety symptoms involve sustained worry on a variety of one's life aspects; depression symptoms involve disturbance of motivation, appetite, sleep, persistent low mood, and suicidal tendencies [4]. Both MDD (Major depressive disorder) and GAD (Generalized anxiety disorder) are psychiatric disorders resulting from persistent symptoms, and are the leading causes of disabilities in the US [5, 6]. Either or both of these symptoms can affect the well-being of an individual leading to devastating psychological effects

on themselves, their relatives, and communities [7].

The ongoing COVID-19 pandemic has amplified the adverse effects of these symptoms on individuals as evidenced in recent literature [8]. Data collected by the CDC (Centers for Disease Control) in the US, shows that the prevalence of anxiety disorder symptoms for the second quarter of 2020, was approximately three times those reported in the same period in 2019 (25.8% versus 8.1%), while the prevalence of depressive disorder was approximately four times more (24.3% versus 6.5%) [9, 10]. The increase in hospitalization and infection rates, deaths, and case counts during the pandemic have aggravated the reported symptoms. Patients hospitalized during the peak of COVID-19 in 2020, reported increased anxiety due to identified factors such as uncertainty in disease progression, strict restriction of visitation from family members, and the inability to develop an open connection with health care workers due to the use of personal protective equipment [11]. The clinical effect of the pandemic on mental health will need to be closely scrutinized over time. Only recently have initial mental health issues associated to the pandemic been identified. These include stress, experiencing episodes of anxiety and depression, alcohol consumption, eating disorders, hunger, and uncertainty about the future, among others. [12]. In addition, COVID-19 related social media usage in times of disasters exacerbate physiological outcomes (i.e. negative effect, depression and anxiety) [13]. Coronavirus rumors and content on social media can amplify negative emotions such as panic, fear, and distress, which may in turn induce anxiety symptoms [14]. In [15], the authors indicate how increased social media use can contribute to undesirable life experiences such as reduced creativity and withdrawal from reality.

Recently, there have been multiple studies highlighting research results of the pandemic on mental health. In particular, the use of machine learning (ML) and artificial intelligence (AI) techniques have been explored and reviewed [16]. The majority of these studies focus on: diagnosis and prognosis based on

clinical data, data analysis for disease detection, using social media data to estimate risk of mental illness, and genomics data analysis for understanding mental health illness [17]. There exists a need for a mechanism to model the symptom occurrence of anxiety and/or depression based on the external environment of a user constituted by: hospitalizations, infection, death rates, and sentiment in social media content.

The purpose of this work is to provide a conceptual and practical implementation of an end-to-end machine learning architecture that aids in answering the underlying research question - *“Can the occurrence of anxiety and/or depressive disorder symptoms be accurately modeled from official state hospital data and social media sentiment using machine learning?”*

Utilizing official New York State, CDC, and Twitter data during the pandemic, the implemented system demonstrates effectiveness in the comparative studies of both supervised learning classification tasks when labels for these symptoms exist, together with unsupervised clustering techniques when such labels are unavailable.

The remainder of this paper is organized as follows; Section 2 gives a comprehensive background on prior work performed in the application of machine learning to mental health data. Section 3 proposes a research architecture implementation and methodology. Section 4 discusses the machine learning approaches implemented throughout this framework. Section 5 includes results from the application of ML models using the defined architecture. To conclude, Section 6 envisages a future research direction for the developed machine learning architecture.

## 2. Related Work

The increasing availability of data pertaining to mental health collected through self-reporting questionnaires has presented several opportunities. By answering questions related to feelings or social interactions, it is now possible to apply AI and ML techniques to gain knowledge on mental health conditions and disorders. This can assist medical practitioners and clinical decision makers [18, 19].

There have been several previous studies related to the application of machine learning to predict depression and anxiety. Shatte et al. conducted a literature review to categorize ML techniques in mental health according to: diagnosis, prognosis, public health, treatment, and research and clinical administration [20]. In Wang et al. [21], the authors accurately predict postpartum depression (PPD) using extracted features from electronic health records (EHR) such as the use of antidepressant drugs and pain diagnosis. Several

ML research efforts have focused on the diagnostic prediction of anxiety and depression from clinical data.

Solutions using human voice have been realized to develop automatic mental health monitors assisting in early diagnosis and longitudinal monitoring of anxiety and depression symptoms in everyday speech conversation [22, 23, 24]. Accelerometers have also been used to detect the presence and level of depression from motor activity recordings [25]. Nguyen et al. apply text mining within online social communities to better understand linguistic-related topics in the context of mental health [26]. Fatima et al. use ML to identify depression and its severity from social media textual posts [27]. Tweets are used by [28] to mine user generated content longitudinally, while developing a statistical model for unobtrusive monitoring of clinical depressive symptoms in social media. The use of structured data in the form of questionnaires and health records are explored in several studies [29, 30, 31, 32]. These studies develop well-known models like Bayesian networks to better understand the severity of illness, and identify individuals with anxiety symptoms like stress for appropriate clinical intervention.

There are several papers that use multimodal data to study signs of depression e.g., through multimodal communication with a social robot [33]. In [34], a survey of ML in mental health reveals that out of 54 research papers, few describe the conduct of empirical studies of an end-to-end ML system [35, 36] or assess the quality of ML predictions. Newer deep learning (DL) technology approaches aim to address this by developing an end-to-end mechanism mapping raw input features directly into outputs through multi-layered neuron network structures. These networks can capture the hidden patterns within the data [37]. In recent years, DL applications have been used widely in healthcare settings due to their superior performance in terms of predicting outcomes [38, 39, 40]. However, there is less work on the design implications for user-centric, deployable ML systems [41]. This research effort is focused on filling this gap. The aim through this work is not to detect specific symptoms for diagnosis, or improve treatment; rather, it is to realize a practical end-to-end user-centric ML architecture, capable of studying the efficacy of plug-and-play data and ML models to predict mental health symptom occurrence from official state hospital statistics, records, and social media sentiment.

## 3. Methodology

To be able to compare the quality of ML models in the domain of mental health, a number

of stages need to be realized in an end-to-end fashion. These range from: data gathering and consolidation, to feature engineering for input selection, to model implementation and evaluation through predictor accuracies. The methodological sections that follow, document the proposed end-to-end ML architecture pipeline highlighting these stages.

### 3.1. Proposed ML Architecture

Figure 1 shows the proposed ML Architecture which was designed by undertaking an experimental approach. Firstly, systematic reconfiguration of input data fed into the predictive models was conducted. In the initial attempts, only mental health and COVID-19 data was explored. Secondly, the social media aspect of mental health was explored since the link between the consumption of social media platforms and mental illnesses has been well documented. To this end, Twitter sentiment counts were fed into the model along with CDC and New York COVID-19 data. More specifically, the architecture was used to examine the affect of pandemic-related trigger words linked to positive and negative social media content, on observable mental health symptoms for anxiety and depression.

The results presented in Section 5, show how the use of external plug-and-play Twitter data was used to improvise a specific ML model which outperformed others in the unsupervised learning gambit. A subsequent attempt prioritized the use of only statistically significant features in order to prove that the findings of the previous iterations are meaningful and not merely by chance. The stages forming the pipeline are highlighted in the subsections that follow.

### 3.2. Input Datasets

Involves selecting appropriate datasets, which in this case study, were procured from three distinct sources, namely: CDC mental health data [42], New York City COVID-19 data [43], and Twitter count data [44]. The mental health dataset was formed by the CDC in conjunction with the National Center for Health Statistics (NCHS) and produced the Household Pulse Survey. This documents the social, economical and psychological effect that the COVID-19 pandemic has had on individuals. More specifically, the survey reports the affects of the pandemic on various lifestyle changes such as employment status, educational disruption, food insecurities, spending habits and finally both psychical and mental well-being [42]. The variance of the survey data is seen in patient demographics such as age, gender, race, educational level and disability status. The dispersal of each subject's background ensured an

unbiased mental health dataset [45]. The resulting tabulated data is reflective of a 95% confidence interval within the surveyed population.

The New York COVID-19 time series dataset was made open source by John Hopkins University and reports the daily infection rates dating back to the beginning of the pandemic [43]. The accumulation of these daily indicators are documented by the state's various hospitals, testing sites and private medical centers. This data provides information on the positive COVID cases, hospitalization and death rates.

Social media data was obtained via a secondary data source, and captures the most frequent coronavirus-related tweets and hashtags. This data was collected from a daily stream of tweets and was filtered out by language and location to include data for only English tweets from New York. At the height of the pandemic in April 2020, nearly 4.4 million COVID-19 related tweets originated on a daily basis before ultimately subsiding to approximately 2.6 million tweets per day [44]. For natural language processing (NLP) purposes, standard data cleansing was performed and sentiment analysis techniques were conducted using the NLTK package [46]. The polarity values of each tweet were computed in which the probability of assigned classes were determined based on the class with the higher probability. This revealed the dataset had approximately nine times more negative content compared to positive content during this period. This algorithmic process is outlined in Algorithm 1.

```

Input : clean Twitter data ( $T_d$ )
Output : positive or negative sentiment label
1 Start NLTK process
2   polarity = sentimentAnalyzer.polarity( $T_d$ )
   // return key-value pair
3
4   positive = polarity["positive"]
5   negative = polarity["negative"]
   // determine sentiment based
   on polarity scores
6
7   if negative > positive then
8     | sentiment = negative connotation
9     | negative_sentiment_count+ = 1
10  else
11    | sentiment = positive connotation
12    | positive_sentiment_count+ = 1
13 End

```

**Algorithm 1. Twitter Labeling Sentiment Analysis Based On The Polarity Score**

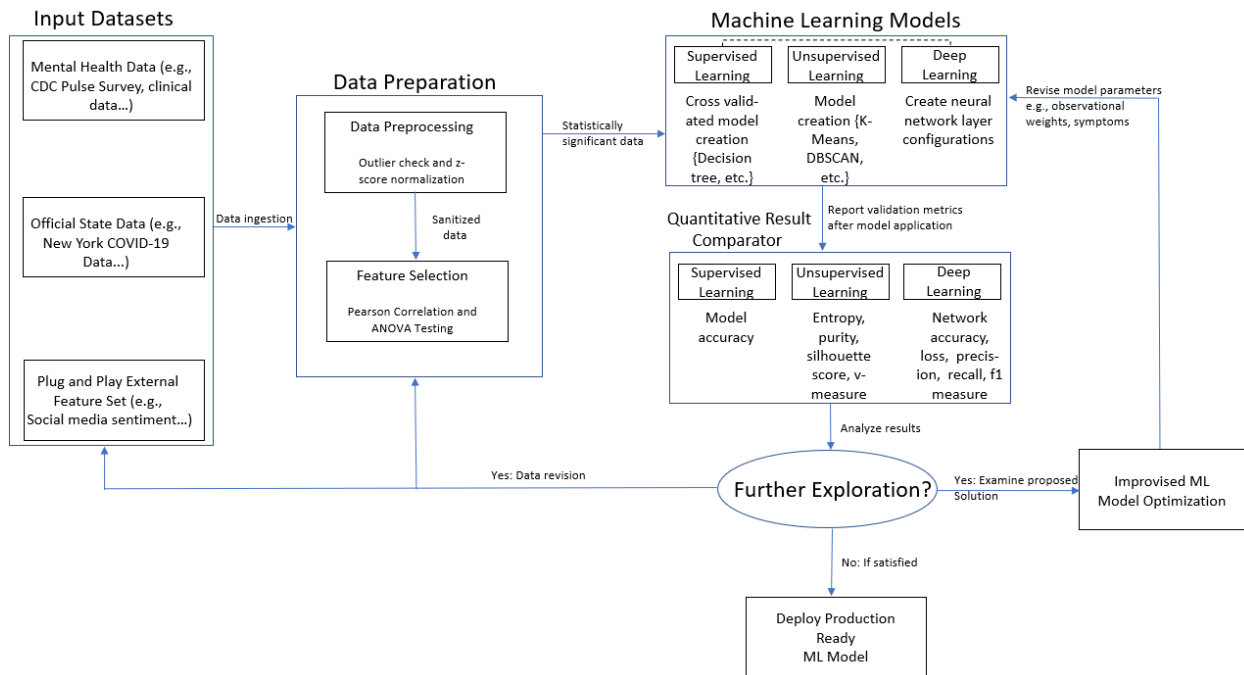


Figure 1. Proposed End-to-End ML Mental Health Architecture For Predicting Anxiety and/or Depression

The final dataset was created by simply merging all three data sources together by date. This case study had nearly 36,000 unique individual observations. Table 1 lists the features that were used.

### 3.3. Data Preparation

The final merged dataset is ingested into this phase for sanitization. The undertaking of building a robust mental health ML model focuses on the use of the most important parameters in model construction. As

such, the need to reduce dimensionality was crucial so that predictive power could be improved. Therefore, various tests for statistical significance were carried out to identify which features had the greatest influence. First, the Pearson Correlation was used to determine the linear relationship between the variables in the dataset. The correlation threshold that was used in this experiment to filter out features was 0.9. The situation of multicollinearity was handled by removing one feature from the pair that had the greater correlation average with respect to other features. The issue of

Variable	Definition
phase	The state of the pandemic lockdown and reopening.
time period	The number of days since the start of the study (April 23, 2020 - February 1, 2021).
mental health value	A computed field from the Household Pulse Survey results conducted by the CDC.
low case influence	The negative influence of social determinants caused by the pandemic.
high case influence	The extremely rare positive influence of social determinants on everyday life.
age	The age measured in years. This study considers all adults over the age of 18.
case rate	The amount of daily positive COVID-19 cases.
hospitalization rate	The amount of daily hospitalizations and overcrowding caused by COVID-19.
death rate	The amount of daily COVID-19 related deaths.
positive sentiment	The amount of positive coronavirus-related tweets and hashtags aggregated daily.
negative sentiment	The amount of negative coronavirus-related tweets and hashtags aggregated daily.
mental health indicator	The target variable of either: anxiety, depression, or the worst-case-scenario of both.

Table 1. Dataset Input Features (Source: CDC, New York COVID-19, Twitter Sentiment Counts) and Target Variable (Indicator of Anxiety and/or Depression)

non-linear data relationships not being captured by the Pearson Correlation was handled by transforming specific variables to their linear form.

Feature selection was performed by first filtering out features based on the correlation factor. To be more selective, p-values were obtained by fitting the resulting filtered dataset onto the Ordinary Least Squares (OLS) regressor. The threshold used to determine final feature selection was  $\alpha = 0.05$ , in which any feature that had a smaller p-value was considered statistically significant. The statistically significant features were: the time period, mental health value, case and hospitalization rate, positive and negative sentiment counts.

These results were verified by the ANOVA f-test feature selection method. The f-test found the same set of features as the previous attempt. Table 2 reports the p-values and f-statistics for significant data only:

	<i>p-value</i>	<i>f-statistic</i>
Time period	0.0318	62.337
Mental health value	0.0147	187.12
Case rate	0.0225	103.54
Hospitalization rate	0.0472	25.846
Positive sentiment	0.0386	37.010
Negative sentiment	0.0283	88.743

**Table 2. P-values and F-statistic Measures For Statistically Significant Features**

## 4. Machine Learning Models

A wide variety of supervised, unsupervised, and deep machine learning models can be plugged and played to the sanitized data. For this case study, all models were implemented using best practices.

### 4.1. Conventional Supervised Learning

The baseline models that were implemented include: Logistical Regression, Support Vector Machines, Decision Trees, K-Nearest Neighbors, Linear Discriminant Analysis, Naive Bayes Classifier and Random Forests. The first iteration of this implementation involved using only CDC mental health and COVID-19 data. The second iteration introduces the Twitter data component and illustrates the social media influence on mental health. The main metric that was used to measure the performance of these conventional machine learning models was accuracy in terms of predictive power. This process was systematically performed by computing the average accuracy of each fold from the cross validated output.

### 4.2. Unsupervised Learning

K-Means, Hierarchical, DBSCAN, and Gaussian Mixture Model (GMM) were applied and studied. The main reasoning behind the exploration of various clustering algorithms was to examine the grouping of records from the following point of views: Euclidean centroid distance partition, hierarchy distance matrix, density regional space, and a finite amount of Gaussian distributions. The parameters used in each clustering algorithm were extracted using benchmark standards. For all of the clustering algorithms, the target variable was removed. Similar to the supervised learning implementation, this was repeated both with and without the social media (external feature) component.

### 4.3. Deep Supervised Learning

Relies on implementing and exploring neural network based models using similar implementation iterations as in supervised learning. For this case study, three different neural network implementations which contained a varying number of hidden layers were tested. The first model only contained an input layer and an output layer, while the second model had a similar architecture with the lone addition of a hidden layer. The third model had an extra hidden layer. Although model 1 might not hold the properties of a true “*deep learning*” model, it helps to understand the effectiveness of a more layered deep approach as seen in models 2 and 3. All the hidden layers had 11 neurons, with varying input layer neurons based on whether Twitter sentiment was used (11 neurons) or not (9 neurons). In all models, dense layers were used along with a Relu activation function except for the output layer which had a Sigmoid activation function. A total of 68 epochs were used with a batch size of 256. The neural networks were analyzed using a categorical cross-entropy loss function in which model accuracy and loss were reported.

### 4.4. Quantitative Result Comparator

The quantitative results from all these models are transferred to this module in the architecture. This allows for visualization and comparison across the different models. The outcome of this process can help a user of this architecture ask the underlying research question: “*Is further exploration desired?*”. If the user responds with a NO, they can use their findings to deploy the appropriate ML model. If the answer to this question is YES, the user can either 1) choose to go back to explore the input datasets to edit existing data or add other external data sources, or 2) fine tune the data preparation phase by implementing (new) algorithmic

enhancements, or 3) use the knowledge gained from the results to improvise their own ML models.

#### 4.5. Improvising a ML Solution

For this case study, the architecture was used to decide on formalizing a new clustering solution implementation to effectively group subjects with anxiety and/or depression symptoms. The unique clustering algorithm was built from the list of statistically significant features (excluding the target variable), allowing for unseen raw data to be input to the model and determine a label of the symptom occurrence. This utilizes a weighted K-Means partition-based technique which outputs a continuous value that is then used in an observational weight formula. The main reason for the use of a weighted approach is that the original K-Means clustering algorithm assumes each feature is equally important when determining class labels. The weights were determined iteratively by executing the formula illustrated in Equation 1 below for each feature:

$$w_i = \sigma \cdot \left[ \log_2 \left( \sum_{f_i}^{s_i} \frac{(f_i - \mu)}{\sigma} \right) \right]$$

**Equation 1. Formula to Obtain Weights ( $s_i$  is the total number of significant features and  $f_i \in s_i$ )**

This formula was used to assign the feature weight by first finding the mean and the standard deviation of the respective column. Once those values were obtained, the logarithmic sum of each observation was computed. The final step was to take the standard deviation of the resulting sum. The decision to use  $\log_2$  is to support two-way branching. Furthermore, it served as a unique inverse exponentiation function in which it decreased the output values over the set of positive integers instead of increasing their magnitude. The following table documents the assigned weight of each feature.

Feature	Weight
Time period	0.5083
Mental health value	0.9746
Case rate	0.7132
Hospitalization rate	0.3529
Positive sentiment	0.4926
Negative sentiment	0.5317

**Table 3. Output Of Calculated Weights For Statistically Significant Features**

The feature weight has a range from 0 (lowest influence) to 1 (highest influence). Once the weight

assignments were completed, the following clustering algorithm was used to determine the observational weight of the instance in question.

$$O_w = \frac{x_i - \mu}{\sigma} \cdot \sqrt{\left( \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right) \cdot \left( \phi \sum_{x_i \in \mathbb{R}}^k w_i(x_i) \right)}$$

**Equation 2. Formula to Obtain Weight of Instance**

Table 4 explains the mathematical symbols of the observational weighted equation in detail:

Symbol	Description
$x_i$	input feature value
$\mu$	arithmetic mean
$\sigma$	population standard deviation
$N$	total number of observations
$\sum$	summation operator
$w_i$	feature weight
$\in$	existence of element in a vector
$\mathbb{R}$	set of rational and irrational numbers
$k$	unique count of target class members
$\phi$	golden ratio scalar
$O_w$	resulting observation weight of instance

**Table 4. Definition Of Mathematical Symbols**

The expression outside the square root is simply a z-score normalization. The first term inside the square root is the standard deviation which is used to indicate the dispersion of the new value in relation to previous observations. The second term in the square root is the partitioning element. In this portion, a summation is applied to the result of the new value multiplied by both the weight and the golden ratio. The golden ratio is used as a continuous proportional fractal in which the limit of the consecutive values is roughly 1.618 as the series approaches  $\infty$ . This summation is repeated k-times generically, to consider the total possible mental health diseases. In this experiment, k was set to 3 since there are 3 class variables; however, this can be scaled up or down at the clinician's discretion dependent on the various types of mental health disorders in their respective practice. The end result of this algorithm is a continuous value that is then used for clustering.

After the observational weight was determined, the final phase of this approach was to assign a classification label to the data. The thresholds for determining class membership were found by computing the mean of the mental health value column, and then adding and subtracting the standard deviation. This additional step is carried out because the dataset takes on the

properties of a normal distribution. Algorithm 2 shows the partitioning step to determine symptom occurrence.

```

Input :  $\mu$  and  $\sigma$  of mental health value,
          observational weight ( $O_w$ )
Output : classification of symptom occurrence
1 Start Clustering Assignment Threshold Layout
2    $high = \mu + \sigma$ 
3    $low = \mu - \sigma$ 
   // use threshold to partition
4
5   if  $0 \leq O_w \leq low$  then
6     class = Depression
7   else if  $low < O_w \leq high$  then
8     class = Anxiety
9   else
10    class = Anxiety and Depression
11 End

```

**Algorithm 2. Systematic Labeling Of Mental Health Symptoms Based On Observational Weight**

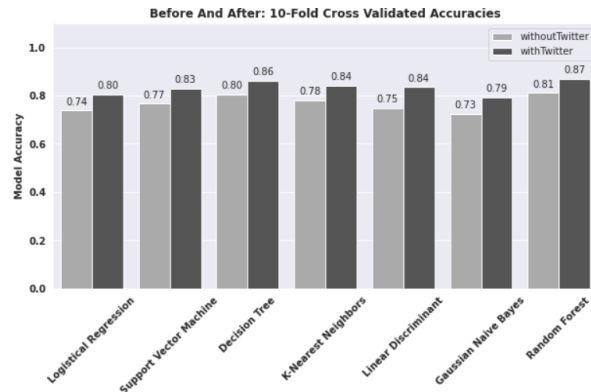
## 5. Results

The performance metrics of all ML models were analyzed using best practices and the superior models are highlighted in each table. This was done in terms of their respective category and the impact of adding Twitter data. For the supervised learning models, the accuracy of the 10-fold cross validation output is reported. In the unsupervised models, cluster entropy, purity, silhouette score and v-measure validation metrics are reported. Finally, in the deep learning approach, the model accuracy and categorical cross-entropy loss is documented. In certain instances, a classification report was generated which provided insight on key metrics including precision, recall and f1-score.

### 5.1. Supervised Learning Validation Metrics

In this experiment, cross-validated conventional supervised machine learning models were explored. Figure 2 illustrates the importance of Twitter sentiment data because in all cases, model accuracy improved. Tree-based algorithms in general outperformed the other classifiers with both random forest and decision tree outputting the highest accuracies at 87.02% and 86.03% respectively. These results suggest that tree-based models are strong candidates for deployment using the proposed architecture. Their high accuracy can be attributed to determining the appropriate splits based on information gain. Aside from tree ML models, logistical regression, k-nearest neighbors, support vector

machine and linear discriminant, nearly all had identical accuracies (about 82%). The Gaussian Naive Bayes probabilistic classifier performed the worst.



**Figure 2. Supervised Learning Classifier Accuracies**

### 5.2. Cluster Validation Metrics

Similar to the supervised approach, all clustering algorithms improved with Twitter data. Table 5 shows the benefits of adding Twitter data. Both DBSCAN and K-Means produced decent clustering solutions when compared to Agglomerative and GMM. This, however, was overshadowed by the improvised ML solution which surpassed all algorithms in every performance metric. This advantage is reflective in the entropy and purity measures (0.2595 and 0.8614 respectively). According to the proposed architecture, the plug-and-play improvised ML model should be the preferred deployable solution in this specific domain.

	Entropy	Purity	Silhouette Score	V-Measure	With Twitter
K-Means	0.4526	0.5301	0.3982	0.3382	No
	0.3511	0.6256	0.6318	0.4927	Yes
Agglomerat.	0.4637	0.5162	0.3591	0.3194	No
	0.3782	0.5970	0.4817	0.3827	Yes
DBSCAN	0.3123	0.6912	0.6395	0.4385	No
	0.3527	0.7283	0.6947	0.5924	Yes
Gauss. MM	0.4943	0.5163	0.4430	0.4782	No
	0.4102	0.5408	0.4917	0.4831	Yes
<b>Impr-ovised</b>	<b>0.2595</b>	<b>0.8614</b>	<b>0.7390</b>	<b>0.6709</b>	<b>Yes</b>

**Table 5. Cluster Validation Metrics**



### 5.3. Deep Learning Results

The evaluation metrics used to determine the performance of the neural networks were mainly model accuracy and loss. For specific models which had a superior performance, further classification metrics such as precision, recall and f-1 measures are reported. Table 6 documents the performance measures for Models 1, 2, and 3, in which accuracy increased while loss decreased with respect to the presence of Twitter data. Figure 3 compares the accuracy and loss of model 3 with its 10-fold cross validated version. Both versions excelled when Twitter data was used. Twitter-based models continue to increase in accuracy and decrease in loss, while the metrics of the non-Twitter models tend to stabilize as the learning process continues over each epoch. Model 3 with Twitter had the best accuracy (95%) and loss (0.07) in all experiments. This ultimately

	<i>Accuracy Without Twitter</i>	<i>Loss Without Twitter</i>	<i>Accuracy With Twitter</i>	<i>Loss With Twitter</i>
Model_1	0.7982	0.3696	0.8837	0.2506
Model_2	0.8351	0.3203	0.9232	0.1552
<b>Model_3</b>	<b>0.8620</b>	<b>0.2875</b>	<b>0.9527</b>	<b>0.0760</b>

Table 6. Accuracies and Loss of Artificial Neural Networks Before and After Adding Twitter Data

highlights the effectiveness of a more deeply layered network as shown in the classification report in table 7.

<i>Model_3 Label</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Measure</i>	<i>With Twitter</i>
Anxiety	0.94	0.78	0.85	No
Depress.	0.95	0.92	0.84	No
Both	0.85	0.77	0.91	No
<b>Anxiety</b>	<b>0.96</b>	<b>0.96</b>	<b>0.97</b>	<b>Yes</b>
<b>Depress.</b>	<b>0.99</b>	<b>0.98</b>	<b>0.99</b>	<b>Yes</b>
<b>Both</b>	<b>0.97</b>	<b>0.95</b>	<b>0.96</b>	<b>Yes</b>

Table 7. Classification With/Without Twitter

As seen, the model 3 version with Twitter surpassed the models without it across all metrics. The non-Twitter model faltered when classifying anxiety and depression. This is depicted in the precision and recall, resulting in a lower harmonic mean f1-score. For the 10-fold cross-validated model, adding Twitter data improved predictive power by over 10% while reducing loss as shown in Table 8.

	<i>Accuracy</i>	<i>Loss</i>
Without twitter	0.8409	0.3294
<b>With twitter</b>	<b>0.9435</b>	<b>0.1168</b>

Table 8. Performance of Cross Validated Model

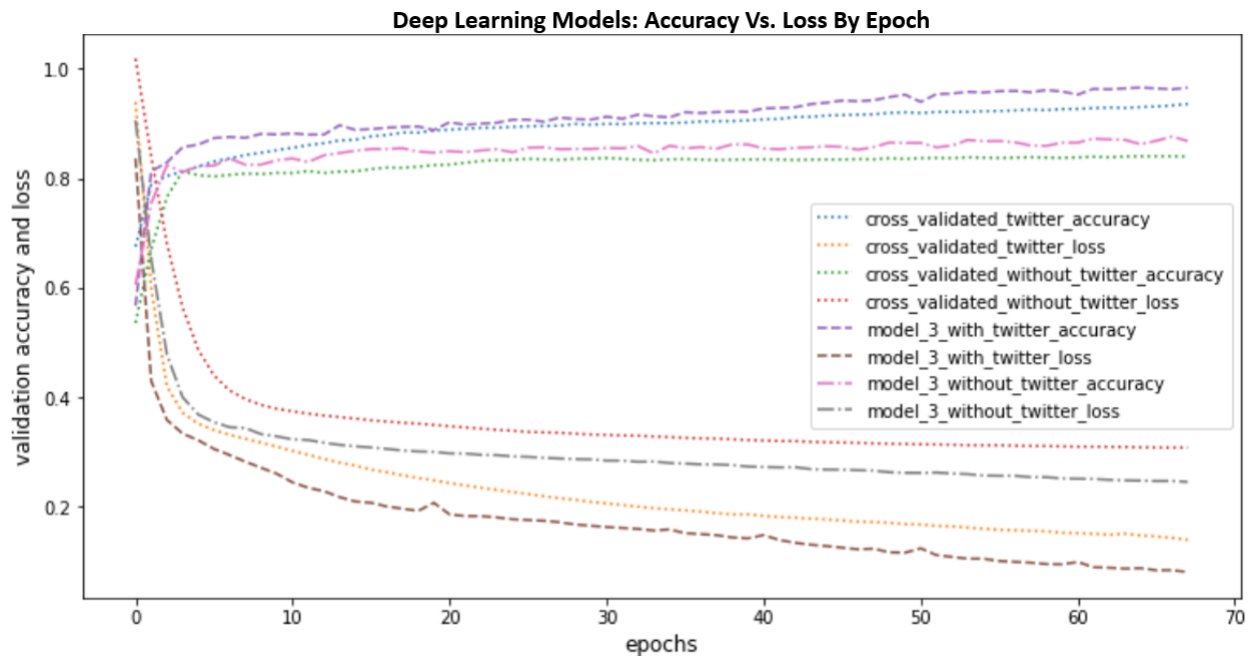


Figure 3. Accuracy vs Loss of Model 3 and its 10-Fold Cross Validated Version (accuracy gradually increases after 20 epochs while loss continues to decrease)



## 6. Discussion and Future Work

The ML architecture proposed in this research serves to provide valuable insight on mental health as evidenced through the New York case study. A limitation of this work is the non-availability of data pertaining to subjects without mental health symptoms in the pulse survey explored; however, our solution is not limited due to this. It can easily be leveraged to accommodate newer subject data or any future mental health data sets as they become more readily accessible. The use of social media sentiment data greatly improved the accuracy of the ML models in predicting mental health symptomatic occurrence. The effective use of this methodology indicates that there is an associative link between psychological state and social media usage.

For future work, investigations using more specific pulse-like surveys asking subjects how and when they were more active on social media platforms, what specific platform they used, and how their mood was affected longitudinally based on time, can provide more conclusive evidence between social media use and anxiety and depressive symptoms. In this research, neural networks outperformed conventional supervised learning algorithms, albeit this is expected for the smaller data size. Development of such tools using the proposed approach, is beneficial for clinicians and decision makers alike to successfully identify issues of concern in mental health, especially in disaster scenarios such as the COVID-19 pandemic. In conclusion, this architecture is practical and capable of scaling to much larger studies with richer data sets.

## References

- [1] A. Bystritsky, S. S. Khalsa, M. E. Cameron, and J. Schiffman, "Current diagnosis and treatment of anxiety disorders," *P & T : a peer-reviewed journal for formulary management*, 38(1), p. 30–57, 2013.
- [2] R. C. Kessler and et al., "The epidemiology of major depressive disorder: results from the national comorbidity survey replication (ncs-r)," *JAMA* 289(23), p. 3095–3105, 2003.
- [3] National Survey On Drug Use And Health, "Samhs (substance abuse and mental health services administration). key substance use and mental health indicators in the united states," *HHS Publication No. SMA 18-5068, NSUDH Series H-53*, 2017.
- [4] M. Lader, "Generalized anxiety disorder," *Stolerman IP, Price LH, eds. Encyclopedia of Psychopharmacology*, p. 699–702, 2015.
- [5] C. Mathers and et al., "The global burden of disease: 2004 update," *World Health Organization*, 2008.
- [6] L. Ghio and et al., "Duration of untreated depression influences clinical outcomes and disability," *J Affect Disord*, no. 175, pp. 224–228, 2015.
- [7] M. S. Reddy, "Depression: the disorder and the burden," *Indian J Psychol Med.* 32(1), p. 1–2, 2010.
- [8] B. Javed, A. Sarwer, E. B. Soto, and Z. U. Mashwani, "The coronavirus (covid-19) pandemic's impact on mental health," *The International journal of health planning and management*, no. 35(5), p. 993–996, 2020.
- [9] M. Czeisler, R. Lane, E. Petrosky, and et al., "Mental health, substance use, and suicidal ideation during the covid-19 pandemic — united states," *MMWR Morb Mortal Wkly Rep*, no. 69, p. 1049–1057, 2020.
- [10] CDC, "National center for health statistics. early release of selected mental health estimates based on data from the january–june 2019 national health interview survey," *National Center for Health Statistics*, 2020.
- [11] K. Roth, R. Upadhyay, and V. Paul, "The effects of covid-19 on mental health during inpatient hospitalization," *Chest*, no. 158(4), A341, 2020.
- [12] E. Altena, C. Baglioni, C. A. Espie, J. Ellis, D. Gavriloff, B. Holzinger, A. Schlarb, L. Frase, S. Jernelöv, and D. Riemann, "Dealing with sleep problems during home confinement due to the covid-19 outbreak: Practical recommendations from a task force of the european cbt-i academy," *Journal of Sleep Research*, vol. 29, no. 4, p. e13052, 2020.
- [13] N. Zhao and G. Zhou, "Social media use and mental health during the covid-19 pandemic: Moderator role of disaster stressor and mediator role of negative affect," *Applied psychology. Health and well-being*, no. 12(4), p. 1019–1038, 2020.
- [14] R. Goodwin, Y. Palgi, Y. Hamama-Raz, and M. Ben-Ezra, "In the eye of the storm or the bullseye of the media: Social media use during hurricane sandy as a predictor of post-traumatic stress," *Journal of Psychiatric Research*, no. 47(8), p. 1099–1100, 2013.
- [15] M. Altuwairiqi, E. Arden-Close, N. Jiang, G. Powell, and R. Ali, "Problematic attachment to social media: the psychological states vs usage styles," in *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, pp. 1–6, 2019.
- [16] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychol. Med.*, no. 49, p. 1426–1448, 2019.
- [17] M. Nemesure, M. Heinz, and et al., "Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence," *Sci Rep* 11, 1980, 2021.
- [18] D. B. Dwyer, P. Falkai, and N. Koutsouleris, "Machine learning approaches for clinical psychology and psychiatry," *Annu. Rev. Clin. Psychol.*, no. 14, p. 91–118, 2018.
- [19] C. A. Lovejoy, V. Buch, and M. Maruthappu, "Technology and mental health: the role of artificial intelligence," *Eur. Psychiatry*, no. 55, pp. 1–3, 2019.
- [20] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychol. Med.*, no. 49, p. 1426–1448, 2019.
- [21] S. Wang, J. Pathak, and Y. Zhang, "Using electronic health records and machine learning to predict postpartum depression," *Studies in Health Technology and Informatics, MEDINFO 2019: Health and Wellbeing e-Networks for All*, pp. 888–892, 2019.
- [22] K.-h. Chang, M. K. Chan, and J. Canny, "Analyze this: Unobtrusive mental health monitoring by voice," in *CHI*

- '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, p. 1951–1956, Association for Computing Machinery, 2011.
- [23] A. Salekin, J. W. Eberle, J. J. Glenn, B. A. Teachman, and J. A. Stankovic, “A weakly supervised learning framework for detecting social anxiety and depression,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, July 2018.
- [24] V. Mitra, E. Shriberg, M. McLaren, A. Kathol, C. Richey, D. Vergyri, and M. Graciarena, “The sri avec-2014 evaluation system,” in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, AVEC '14*, (New York, NY, USA), p. 93–101, Association for Computing Machinery, 2014.
- [25] J. I. Frogner, F. M. Noori, P. Halvorsen, S. A. Hicks, E. Garcia-Ceja, J. Torresen, and M. A. Riegler, “One-dimensional convolutional neural networks on motor activity measurements in detection of depression,” in *Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care, HealthMedia '19*, (New York, NY, USA), p. 9–15, Association for Computing Machinery, 2019.
- [26] T. Nguyen, B. O’Dea, and M. e. a. Larsen, “Using linguistic and topic analysis to classify sub-groups of online depression communities,” *Multimed Tools Appl*, no. 76, p. 10653–10676, 2017.
- [27] I. Fatima, H. Mukhtar, H. F. Ahmad, and K. Rajpoot, “Analysis of user-generated content from online social communities to characterise and predict depression degree,” *Journal of Information Science*, vol. 44, no. 5, pp. 683–695, 2018.
- [28] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan, J. Pathak, and A. Sheth, “Semi-supervised approach to monitoring clinical depressive symptoms in social media,” in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17*, (New York, NY, USA), p. 1191–1198, Association for Computing Machinery, 2017.
- [29] D. Galiatsatos, G. Konstantopoulou, G. Anastassopoulos, M. Nerantzaki, K. Assimakopoulos, and D. Lymberopoulos, “Classification of the most significant psychological symptoms in mental patients with depression using bayesian network,” in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS), EANN '15*, (New York, NY, USA), Association for Computing Machinery, 2015.
- [30] B. Ojeme and A. Mbogho, “Selecting learning algorithms for simultaneous identification of depression and comorbid disorders,” *Procedia Computer Science*, vol. 96, pp. 1294–1303, 2016. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 20th International Conference KES-2016.
- [31] H. Yang and P. A. Bath, “Automatic prediction of depression in older age,” in *Proceedings of the Third International Conference on Medical and Health Informatics 2019, ICMHI 2019*, (New York, NY, USA), p. 36–44, Association for Computing Machinery, 2019.
- [32] T. C. Panagiotakopoulos, D. P. Lyras, M. Livaditis, K. N. Sgarbas, G. C. Anastassopoulos, and D. K. Lymberopoulos, “A contextual data mining approach toward assisting the treatment of anxiety disorders,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 567–581, 2010.
- [33] N. Rastogi, F. Keshtkar, and M. S. Miah, “A multi-modal human robot interaction framework based on cognitive behavioral therapy model,” in *Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era, H3 '18*, (New York, NY, USA), Association for Computing Machinery, 2018.
- [34] A. Thieme, D. Belgrave, and G. Doherty, “Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems,” *ACM Trans. Comput.-Hum. Interact.*, vol. 27, Aug. 2020.
- [35] T. Hirsch, C. Soma, K. Merced, P. Kuo, A. Dembe, D. D. Caperton, D. C. Atkins, and Z. E. Imel., “‘it’s hard to argue with a computer’: Investigating psychotherapists’ attitudes towards automated evaluation,” *Proceedings of the 2018 Designing Interactive Systems Conference (DIS'18)*, no. 10, p. 559–571, 2018.
- [36] P. Paredes, R. Gilad-Bachrach, M. Czerwinski, A. Roseway, K. Rowan, and J. Hernandez, “Popterapy: Coping with stress through pop-culture,” *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'14)*, pp. 109–117, 2014.
- [37] C. Su, Z. Xu, and J. e. a. Pathak, “Deep learning in mental health outcome research: a scoping review,” *Transl Psychiatry*, no. 10 116, 2020.
- [38] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinformatics*, no. 19, pp. 1236 – 1246, 2017.
- [39] D. Durstewitz, G. Koppe, and A. Meyer-Lindenberg, “Deep neural networks in psychiatry,” *Mol. Psychiatry*, no. 24, p. 1583–1598, 2019.
- [40] S. Vieira, W. Pinaya, and A. Mechelli, “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications,” *Neurosci. Biobehav. Rev.*, no. 74, pp. 58–75, 2017.
- [41] T. Hirsch, K. Merced, S. Narayanan, Z. E. Imel, and D. C. Atkins., “Designing contestability: Interaction design, machine learning, and mental health,” *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS'17)*, 2017.
- [42] CDC, “Anxiety and Depression Symptom Occurrence.” (<https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm> Last Accessed: 06/12/2021).
- [43] John Hopkins University & Medicine, “Coronavirus Resource Center: NY State COVID-19 Data.” (<https://coronavirus.jhu.edu/region/us/new-york> Last Accessed: 06/12/2021).
- [44] J. M. Banda, R. Tekumalla, and et al., “A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration,” Aug. 2021.
- [45] S. Peterson and N. Toribio, “Nonresponse bias report for the 2020 household pulse survey,” *United States Census Bureau*, 2020.
- [46] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit (NLTK)*. O’Reilly Media, Inc., 2009.