Survey on Depression Detection

A Project and Thesis Report

Submitted in fulfillment of the requirements for the degree of

**Bachelor of Science in Computer Science and Engineering**

By

Md.Ali Akbar 201711047028

Supervised by

Dr. Uzzal Kumar Prodhan Associate Professor

Department of Computer Science and Engineering



BANGLADESH UNIVERSITY DHAKA, BANGLADESH SEPTEMBER 2020

# DECLARATION

I, hereby, declare that the work presented in this report is the outcome of the investigation performed by me under the supervision of [**Khan Md. Hasib** Lecturer] Department of Computer Science and Engineering, Bangladesh University, Dhaka, Bangladesh. The work was spread over one final year courses, CSE 400 : Project & Thesis I , CSE 400-2: Project & Thesis II and CSE 400-3: Project & Thesis III, in accordance with the course curriculum of Department for the Bachelor of Science in Computer Science and Engineering program.

... ... ... ... .... ... ... ... ... ... ... ...

Md.Ali Akbar

Countersign by

... .... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

Khan Md. Hasib, Lecturer & Coordinator

# Acknowledgements

I would like to express my sincere gratitude to my research supervisor **Dr. Uzzal Kumar Prodhan**, Associate Professor, Department of Computer Science & Engineering, Bangladesh University for his patient guidance, encouragement, pushing us to the limits and his excellent critiques of this research work.

I would like to thank Associate Professor **Md. Sadiq Iqbal**, Head (in charge) of the department of Computer Science & Engineering of Bangladesh University and all our teachers, lab assistance for their moral supports and helpful discussion.

Finally I wish to thank my parents for their support and encouragement throughout my study.

# ABSTRACT

Social networks enable users to connect with each other by exchanging knowledge, sending updates, creating friends, and holding conversations that produce large quantities of data every day, typically referred to as user-generated content. These data are available in a number of ways, such as photos, text, videos, connections, and others, representing user actions, including their mental state. It is difficult but promising to automatically diagnose mental health issues from data that are short, incomplete, and often incorrectly phrased. However, attempts are being made to automatically learn trends using statistical models for such user-generated content. Several forms of studies have been carried out over the past few years on the topic of depression detection including machine and deep learning approach. However, in order to clarify the actual state of deployment, machine learning and deep learning approaches to address the problems of depression diagnosis in this study paper covered some similar experiments on the construction of some single,multi-modal and hybrid model for detecting depression. This survey paper also provides a comparative comparison of the classifier algorithms, the datasets used and some other experimental setups, as well as a consideration of the selection stage.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| **DECLARATION** | 2 |
| **ACKNOWLEDGMENTS** | 3 |
| **ABSTRACT** | 4 |
| **TABLE OF CONTENTS** | 5 |
| **LIST OF ABBREVIATION** | 7 |

|  |  |
| --- | --- |
| **Chapter 1: Introduction** |  |
| 1.1 Why depression detection is important?... ..... .... ... ... ... ... ... | 9 |
| 1.2 Depression detection procedure... .... ...... ..... ..... ..... ..... .... .... | 10 |
| 1.3 .How do we collect papers?. ..... ..... ..... .... ..... ... .... .... .... | 11 |
| 1.4 Contribution of this paper . .... .... ...... .... ....... | 11 |
| 1.5.Overview?. ..... ...... ....... . | 12 |

|  |  |
| --- | --- |
| **Chapter 2:** Algorithm for depression detection |  |
| 2.1 Machine learning algorithm for depression detection....... .... | 14 |
| 2.1.1 Decision Tree | 14 |
| 2.1.2 Naïve Bayes... .... .... .... ..... .... .... .... .. | 14 |
| 2.1.3 Logistic Regression... ... .... .... .... ..... ..... ... | 15 |
| 2.1.4 Support vector machine | 15 |
| 2.1.5 Random Forest... .... | 15 |
| 2.1.6 K-Means clustering .... .... . | 15 |
| 2.1.7 Gradient Boosting.... .... ..... .... ..... .... | 16 |
| 2.2 Deep learning algorithm for depression detection | 18 |
| 2.2.1 Convolution neural network | 18 |
| 2.2.2 Long short term memory | 18 |
| 2.2.3 Re-current neural network | 18 |
| 2.2.4 Radio basis fusion networks | 19 |
| 2.2.5 Multilayer perceptron | 19 |
| 2.2.6 Deep belief networks | 19 |
| 2.2.7 Generative Adversial Networks | 20 |

|  |  |
| --- | --- |
| **Chapter 3: Methodologies on Related Works** |  |
| 3.1 Reviews on Related Research Papers on depression detection using machine learning..... ...... ....... ..... ...... | 22 |
| 3.2 Reviews on Related Research Papers on depression detection using machine learning | 24 |

|  |  |
| --- | --- |
| **Chapter 4: Existing Dataset** |  |
| 4.1 Used Dataset.... .... ..... ..... .... .... ..... ..... ..... .. | 27 |

|  |  |
| --- | --- |
| **Chapter 5 : Discussion** |  |
| 5.1 Discussion. ... ..... ..... ..... ..... ...... ...... ...... ...... ...... | 29 |
| 5.2 Problem Statement. .... ..... .... ..... ....... ..... ....... | 30 |

|  |  |
| --- | --- |
| **Chapter 6: Conclusion and future work** |  |
| 6.1 Conclusion... .... .... .... .... ..... ..... ..... ...... ..... ...... .... | 31 |
| 6.2 Future work. ..... . ...... ..... ..... ..... ..... ..... ..... .. | 31 |
|  |  |

**LIST OF FIGURES**

|  |  |
| --- | --- |
| 1.1 Existing Dataset. ... ... ..... ..... ..... ..... ..... ... ... . | 27 |

**LIST OF TABLES**

|  |  |
| --- | --- |
| Table 2.1: Machine learning algorithm for depression detection | 17 |
| Table 2.1: Deep learning algorithm for depression detection | 20 |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **AI** | **Artificial Intelligence** |
| **DT** | **Decision Tree** |
| **RF** | **Random Forest** |
| **LR** | **Logistic Regression** |
| **KNN** | **K Nearest Neighbor** |
| **NLP** | **Natural Language Processing** |
| **NER** | **Named Entity Recognition** |
| **POS** | **Parts of Speech** |
| **SVM** | **Support Vector Machine** |
| **SVC** | **Support Vector Classifier** |

**Chapter 1 Introduction**

Depression is a long term disease or illness that most teenager nowadays are facing. In the good old days, depression was only seen in war ridden people and soldiers who have survived war. In the current world, depression has stood to be a major issue in the economic, social and most importantly in the health sectors. Any person deemed to be depressed is considered a liability and people often issue immediate treatments to these people. The reason being, depression has become the leading cause of death and is slowly rising above the ranks in being the top disease with zero to no cure. The way people communicate and interact with each other has vastly become dependent on the internet and its access to social media. Applications such as Facebook, Twitter, Instagram and alike not only host the written and multimedia content but also offer their users to express their feelings, emotions and sentiments about a topic, subject or issue online. In its entirety, this is great for users of social networking sites to freely contribute to a cause, topic or any subject and in turn helping others. Then there are also the advantage of doctors and people working in various health sectors to gather insight about the conditions of a person’s mental state through their reaction and responses to a certain topic or issue . On the other hand, these social networking sites also makes one vulnerable to emotional attacks from various predators and people with ill intent. These malicious activities are known to be the root cause of depression. Globally more than 264 million people of all ages suffer from depression. Moreover, over 50 of the time doctors are unable to detect depression since hiding the symptoms is part of depression.

**1.1 Why Depression Detection is important?**

Depression is one of the most significantly underdiagnosed and undertreated forms of mental illness out there. One of the primary causes of it being so hard to detect is the absence of any visual symptoms. It has a profound effect on the patient’s life as well as the people around them. However due to a lack of knowledge in the field depressive patients are often unaware of their condition themselves. Over the years medical practitioners have identified several degrees of treatments that can treat a patient suffering from depression. As studies into depression and its causes are advancing, it is widely agreed that remission of the disease is directly related to the stage at which it is detected. Depressive people often exhibit negative behavioral patterns that can compound the negative effects of the disease and make treatment difficult. The Institute of Medicine Committee on the Prevention of Mental Disorders has labeled depression as the most preventable disorder. Studies on depression among young people have seen that age plays a vital role during treatment . Depression is often linked to other underlying mental illnesses and if left untreated it can cause the patient to trigger other latent mental illnesses. The risk of a person being vulnerable to depression has increased to 20 in the last 30 years, a number which is double if the person has been raised by depressed parents. There are significant differences in symptoms and patterns of depression between a young person and an adult . Unlike most teenagers’ adults have been shown to be treatment resistant in certain severe cases. While most medical specialists will attempt to set the disease to remission until symptoms disappear, treatment of depression is still experimental to a degree and new approaches are being made every day . Most antidepressant therapies are a combination of psychotherapy and medication. As most antidepressants are controlled substances healthcare professionals has to consider numerous factors such as interaction with other drugs and previous drug use history, both of which are commonly not the strongest factors among teenagers \cite{kellough2008time}. Due to the burdensome nature of the disease, early detection and appropriate treatment can prevent most cases from spiraling and can improve a lot of lives.

Depression is now deemed as the leading cause of death and most cases are related to suicide . The gravity of the effects of depression are beyond comprehension. The fact that someone’s words can lead a person towards self-destruction is a very sensitive matter which most of us do not take heed to. However, with the prowess of machine learning , deep learning and its detection techniques, we will be able to expedite the detection process and prevent such suicidality before they can take place .

**1.2 Depression Detection Procedure**

The detection of depression in such cases often stands to be a challenge amongst experts. Reading people’s minds is not possible yet but analyzing their emotions through texts, comments, and responses or even using some methods for utilizing audio cues for depression diagnosis

might give us an insight of their state of mind.

Through the applications of machine learning techniques, using a deep learning model, it can potentially offer some unique features that can assist in examining the unique patterns hidden in these online communications and process them to reveal the mental state among the users of social networks. The analysis of such will reveal a number of different things starting from happiness, sadness, anxiety, depression, distress, etc. The parameters of this calculation will be a steady search of keywords followed by their percentage appearances in all of their responses. These various linguistic cues are examined thoroughly through a different classifier such as Decision Tree, SVM, KNN or Ensemble . The different linguistic cues may include emotion cause events keywords such as positive emotions (for instance, “happy”, “love”, “nice”, “good”), negative emotions (for instance, “worthless”, “loser”, “hurt”, “pain”, “ugly”, “nasty”), sadness (for instance, “worry”, “crying”, “grief”, “sad”), anger (for instance, “stop”, “shit”, “hate”, “kill”, “annoyed”) and anxiety (for instance, “worried”, “fearful”) \cite{xezonaki2020affective}. While these may be some of the factors of the deep learning techniques that can be used in the detection, other factors are also taken into account for a cleaner analysis.

Some of hybrid model classify or estimating depressed people from audio , video ,image or text information .

**1.3 How do we collect the Papers?**

We collected more than 80 papers for this survey. Most of them year wise downloaded from Google scholar. We also adopted Sci-Hub as an important tool for downloading paper with URL or DOI . The major keywords we use for downloading paper are: Machine learning, Deep learning , Depression, Detection, social, Algorithm.

**1.4 Contribution of this survey**

The main objective of this paper is to thoroughly review literature on depression detection technique using machine learning and deep learning algorithm.With our current thesis, we are examining the spectrum of social media-based depression interventions . We have discussed machine-learning and deep learning methods that can be used measures to diagnose people suffering from depression .

The contribution of this survey at a glance:

* Systematic review for depression detection based on machine learning and deep learning
* Methodological analysis for Machine learning and deep learning algorithms
* Discuss challenges, open issues and future work

In summarize, Section 2 consists methodological analysis for machine learning and deep learning separately .Finally section 3 about discussion section 4 conclusion and future work.

To understand the challenges that a class imbalance imposes, let's consider two common ways we'll train a model: tree-based logical rules developed according to some splitting criterion, and parameterized models updated by gradient descent.

**1.5 Overview of thesis**

Our thesis book consists of 6 chapters which is related to the objectives of our proposed model. It gives a brief overview of our research. The topics discussed throughout the chapters are listed below:

**Chapter 1 Introduction:** This chapter gives a brief introduction about the importance of product review, the thesis problems and objectives.

**Chapter 2 Review on related Study:** In this chapter the previous works are described elaborately. Some important topics described are, Data mining, Sentiment Analysis, major kinds of classification algorithms used like Naïve Bayesian, Support Vector machine, logistic and linear regression , gradient boosting, Decision tree as machine learning algorithm. Deep learning algorithms are also described in this chapter.

**Chapter 3 Reviews of Related Research Works:** This chapter shows some related works regarding our topics. In addition we analyzed the importance and flows of some related researches.

**Chapter 4 Existing Dataset:** The datasets those were used in the experiments, the approaches to process the datasets, pull based active learner, preprocessing, feature extraction and experiment cases are described elaborately here.

**Chapter 5 Discussion:** It provides details about the experiment sets, experiment results. It also gives comparison of results with others’ result.

**Chapter 6 Conclusion and Future work**: This Chapter consists of Conclusion and future work .

# Chapter 2

# Algorithm for depression detection

In this chapter the basic concepts and methodologies of necessary topics used in our research paper are described elaborately. It includes sentiment analysis overview, methodologies, and discussion on machine learning algorithm, different approaches of machine learning and the classifiers used.

**2.1 Machine learning algorithm for depression detection**

Social media channels, such as Facebook, Twitter, and Instagram, have altered our world forever. People are now increasingly connected than ever and reveal a sort of digital persona. Although social media certainly has several remarkable features, the demerits are undeniable as well. Recent studies have indicated a correlation between high usage of social media sites and increased depression. The present study aims to exploit machine learning techniques for detecting a probable depressed Twitter user based on both, his/her network behavior and tweets. For this purpose, we trained and tested classifiers to distinguish whether a user is depressed or not using features extracted from his/her activities in the network and tweets. The results showed that the more features are used, the higher are the accuracy and F-measure scores in detecting depressed users.. This study’s main contribution is the exploration part of the features and its impact on detecting the depression level.

**2.1.1 Decision Trees (DT)**

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance-event outcomes, resource costs, and utility. Take a look at the image to get a sense of how it looks like. From a business decision point of view, a decision tree is the minimum number of yes/no questions that one has to ask, to assess the probability of making a correct decision, most of the time. As a method, it allows you to approach the problem in a structured and systematic way to arrive at a logical conclusion.

**2.1.2 Naive Bayes Classification (NB)**

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes’ theorem with strong (naive) independence assumptions between the features. The featured image is the equation — with P (A|B) is posterior probability, P(B|A) is likelihood, P(A) is class prior probability, and P(B) is predictor prior probability.

Some of real world examples are:

* To mark an email as spam or not spam
* Classify a news article about technology, politics, or sports
* Check a piece of text expressing positive emotions, or negative emotions?
* Used for face recognition software

**2.1.3 Logistic Regression (LR)**

Logistic regression is a powerful statistical way of modeling a binomial outcome with one or more explanatory variables. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution.

In general, regressions can be used in real-world applications such as:

* Credit Scoring
* Measuring the success rates of marketing campaigns
* Predicting the revenues of a certain product
* Is there going to be an earthquake on a particular day?

**2.1.4 Support Vector Machines (SVM)**

SVM is binary classification algorithm. Given a set of points of 2 types in N dimensional place, SVM generates a (N — 1) dimensional hyperplane to separate those points into 2 groups. Say you have some points of 2 types in a paper which are linearly separable. SVM will find a straight line which separates those points into 2 types and situated as far as possible from all those points. In terms of scale, some of the biggest problems that have been solved using SVMs (with suitably modified implementations) are display advertising, human splice site recognition, image-based gender detection, large-scale image classification.

**2.1.5 Random Forest (RF)**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

**2.1.6 K-means clustering**

K-Means clustering is an unsupervised learning algorithm. There is no labeled data for this clustering, unlike in supervised learning. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects belonging to another cluster.

The term ‘K’ is a number. You need to tell the system how many clusters you need to create. For example, K = 2 refers to two clusters. There is a way of finding out what is the best or optimum value of K for a given data.

**2.1.7 Gradient boosting**

Gradient boosting is a type of machine learning **boosting**. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

* If a small change in the prediction for a case causes a large drop in error, then next target outcome of the case is a high value. Predictions from the new model that are close to its targets will reduce the error.
* If a small change in the prediction for a case causes no change in error, then next target outcome of the case is zero. Changing this prediction does not decrease the error.

The name *gradient boosting*arises because target outcomes for each case are set based on the gradient of the error with respect to the prediction. Each new model takes a step in the direction that minimizes prediction error, in the space of possible predictions for each training case.

Table 2.1: Machine learning algorithm for depression detection

|  |  |
| --- | --- |
| **Algorithm** | **Problem Tackle** |
| Random Forest (RF) | To labelled data by using learned patterns |
| K-Nearest Neighbors (KNN) | Classify data without training |
| Decision Tree (DT) | Predict class or value by a decision rule |
| Support Vector Machine (SVM) | Finding possible output boundary |
| Naive Bayes (NB) | Labeling low amounts of data |
| K-means clustering | Partition data into k-clusters |
| Gradient Boosting | Make prediction for large amount of data |
| Logistic Regression | Used for binary classification task |

2.2 Deep learning algorithm for depression detection

A sampling method is a procedure for selecting sample members from a population. Three common sampling methods are: simple random sampling, stratified sampling, and cluster sampling.

### 2.2.1 Convolutional Neural Networks (CNNs)

CNN's, also known as ConvNets, consist of multiple layers and are mainly used for image processing and object detection. Yann LeCun developed the first CNN in 1988 when it was called LeNet. It was used for recognizing characters like ZIP codes and digits.

CNN's are widely used to identify satellite images, process medical images, forecast time series, and detect anomalies.

### 2.2.2 Long Short Term Memory Networks (LSTMs)

LSTMs are a type of Recurrent Neural Network (RNN) that can learn and memorize long-term dependencies. Recalling p2.2.ast information for long periods is the default behavior.

LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structure where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development.

### 2.2.3 Recurrent Neural Networks (RNNs)

RNNs have connections that form directed cycles, which allow the outputs from the LSTM to be fed as inputs to the current phase.

The output from the LSTM becomes an input to the current phase and can memorize previous inputs due to its internal memory. RNNs are commonly used for image captioning, time-series analysis, natural-language processing, handwriting recognition, and machine translation.

### 2.2.4 Radial Basis Function Networks (RBFNs)

RBFNs are special types of feedforward neural networks that use radial basis functions as activation functions. They have an input layer, a hidden layer, and an output layer and are mostly used for classification, regression, and time-series prediction.

RBFNs perform classification by measuring the input's similarity to examples from the training set. RBFNs have an input vector that feeds to the input layer. They have a layer of RBF neurons. The function finds the weighted sum of the inputs, and the output layer has one node per category or class of data.

### 2.2.5 Multilayer Perceptrons (MLPs)

MLPs belong to the class of feedforward neural networks with multiple layers of perceptrons that have activation functions. MLPs consist of an input layer and an output layer that are fully connected. They have the same number of input and output layers but may have multiple hidden layers and can be used to build speech-recognition, image-recognition, and machine-translation software.

MLPs feed the data to the input layer of the network. The layers of neurons connect in a graph so that the signal passes in one direction. This algorithm compute the input with the weights that exist between the input layer and the hidden layers.

### 2.2.6 Deep Belief Networks (DBNs)

DBNs are generative models that consist of multiple layers of stochastic, latent variables. The latent variables have binary values and are often called hidden units.

DBNs are a stack of Boltzmann Machines with connections between the layers, and each RBM layer communicates with both the previous and subsequent layers. Deep Belief Networks (DBNs) are used for image-recognition, video-recognition, and motion-capture data.

### 2.2.7 Generative Adversarial Networks (GANs)

GANs are generative deep learning algorithms that create new data instances that resemble the training data. GAN has two components: a generator, which learns to generate fake data, and a discriminator, which learns from that false information.

The usage of GANs has increased over a period of time. They can be used to improve astronomical images and simulate gravitational lensing for dark-matter research. Video game developers use GANs to upscale low-resolution, 2D textures in old video games by recreating them in 4K or higher resolutions via image training.

GANs help generate realistic images and cartoon characters, create photographs of human faces, and render 3D objects

Table 2.1: Deep learning algorithm for depression detection

|  |  |
| --- | --- |
| **Algorithm** | **Problem Tackle** |
| Convolutional Neural Networks (CNNs) | Image classification, NLP task |
| Recurrent Neural Networks (RNNs) | To analyse sequential data,Speech processing |
| Long Short Term Memory (LSTMs) | Classify time series data |
| Artificial Neural Network (ANN) | Classify critical statistical standards |
| RvNN (Recursive Neural Network) | Use a tree structure which is capable to learn deep structured erudition |
| Deep Belief Networks (DBFs) | Create connection between layer |
| Generative Adversarial Networks (GANs) | Game framework & image |

# Chapter 3

**Methodologies on Related Works**

In this chapter reviews on different research papers to our topic has been discussed briefly. We tried to analyze research papers related to depression detection with machine learning and deep learning

### 3.1 Reviews on Related Research Papers on depression detection using machine learning

### of mining depression based on TML often use characteristics i.e. numeric vectors that have been analyzed and extracted manually from users to represent the expected object (a user, a tweet, a posted image, etc.).The groundbreaking work in this field of research was proposed by Choudhury et al. To conduct a user-oriented detection of depression, they investigated possible user actions. They found useful signs for characterizing depression by measuring behavioral features on Twitter users related to social interaction, emotion, vocabulary, and linguistic types. While their qualified classifiers did not achieve high performance in classification, they provided a thorough feature engineering analysis method and a simple modeling approach as a pioneering job in this area. Wang et al. undertook additional research using Twitter and Weibo info. This research applied a sentiment analysis approach and introduced man-made rules by using vocabulary to quantify depressive tweet tendencies compared to the work of that rendered a more detailed feature analysis. Their work has shown that text-based features play a key role in the diagnosis of online depression.Cacheda et al. explore various methods by using machine learning algorithm. Deshpande et al. suggested a natural language processing (NLP)-based representation learning method to model Twitter text information. They used the Bag of Words (BOW) algorithm, unlike the previously mentioned work , to represent the tweet text as a sparse vector, enabling the classifier to learn latent features automatically. An F1-Score of 0.8329 was achieved by their qualified Naive Bayes (NB) classifier, whereas the Support Vector Machine (SVM) classifier only achieved an F1-Score of 0.7973.In zhao et al. Here buit a system for sentiment analysis for monitoring real time sentiment of Chinese tweets in Weibo called MoodLens capable of 95 emoticons

### are mapped into four categories of sentiments as angry,

### disgusting, joyful, and sad using Naive Bayes classifier.but in sun et al. here only

### text is considered from the transcript in different depressive levels and extracted feature from text and applied Random Forest method.Another Support Vector Machine based classification with several feature selection techniques like speaking behaviour, eye activity, and head pose were proposed in alghowinem et al.

### Sharma et al. proposed a model with XGBoost machine learning algorithm For identifying important biomarkers for

### depression and predicting the depression cases on different sample

### balanced.

### Pampouchidou et al. proposed a depression detection technique based on facial expression geometry and speech, by interpreting non-verbal manifestations of depression.On the basis on euclidean and cityblock distance, and the time window used,extract the landmark motion features was systematically

### manipulated here.

### yates et al.explored the self harm risk assessment of utilizing the online social network to recognize and examine major depression problems in humans. They quantified behavioral credits via their web-based social networking postings. Identifying/identifying? For or from? Social interaction, thought, dialect and semantic forms, sense of self-system, and antidepressant meds. A promising method for public health, based on the usage of Twitter presence in the manufacture of statistical models for the potential effect of childbirth on the behaviour and temperament of new mothers. Using Twitter messages, 376 mothers were assessed postpartum improvements along the way. Measurement of social interaction, emotion, casual culture, and phonetic type.

### In senoussaoui et al.shows combination of audio and video based model where i-vector based representation used for short term audio features and two other models based on the LGBP- TOP video features are used.

### .

### Nadeem et al

### investigated Twitter as a matter of fact. Gradually studied as tools for identifying psychological well-being, like depression and suicidality in the community, and shown to identify the degree of depression/anxiety in suicide-related tweets. On the other side. Zhang et al.

### have demonstrated that if persons with a high risk of suicide can be detected by online microblogging, it is feasible to upgrade a complex intervention framework to save their lives.

### For effective early depression detection, Framework is proposed based on text classification over social media streams named SS3 which can deal with early risk detection (ERD) problems.

### Many researchers have shown that using user-created/generated content (UGC) reliably will help assess human psychological wellbeing levels. M. Mr. Aldarwish \cite{aldarwish2017predicting} has discussed the fact that the usage of Social Network Sites (SNS) is expanding these days, especially in the more youthful eras. Accessibility of the SNS

### Clients to share their desires, emotions and propose regular schedules. Nowadays, a variety of scholars are utilizing online groups to explore psychological well-being problems. T. Nguyen et al. used computer learning and predictive methods to distinguish online communications between depression and temperament management classes, psycholinguistic techniques, and Substance topics omitted from posts produced by people in these classes. From our point of view, this study is the first effort to do so. Join the depression issue in Facebook or twitter user comments when neither of the above leaving work used Facebook user comments to diagnose depression in katchapakirin, tsugawa et al. A novel classifier proposed, named Inverse Boosting Pruning Trees with a strong classification ability detection on 7862 Twitter users are shown in tong et al. for identifying social networks' users proposed framework is promising for comprehensively evaluating the classification capability of the IBPT( Inverse Boosting Pruning Trees).

### 3.2 Reviews on Related Research Papers on depression detection using deep learning

### 

### In chlasta et al. this paper, a novel approach were proposed to detect depression from speech using CNN and multipart interactive training. Wang et al. proposed an approach for Early risk detection of depression from social media using CNN .

### Many experiments have undertaken the sensing of distress utilizing different variations of speech characteristics such as Mel Frequency Cepstral Coefficients (MFCCs), formants and voice efficiency are shown in rejaibi and ittichaichareon et al. Early research models of machine learning were mainly built on Support Vector Machines and Gaussian Mixture Models prior to creation of DN Network models focused on network architectures including Convolutionary Neural Networks (CNNs) and Long Short Term Memory (LSTM) networks. This models have been trained with one depression database.In zhang et al. CNN and BiGRU combined together for classify depressed people.

### Among the speech features used, the ACFs collect details that can discriminate between depressed speech and non-depressed speech with the multi-scale structure of the distinctions between time-series signals. This method was primarily tested using acoustic features such as formers and MFCCs as a surrogate for underlying articulatory synchronization. In our previous studies xezonaki et al. we have shown that ACFs are extracted from direct joint speech Features known as Vocal Tract Variables (TVs) are most successful in classifying undepressed expression. mumtaz et al. proposed an EEG(electroencephalographic)based deep learning framework consisting two different deep learning architectures that utilized one dimensional convolutional neural network(1DCNN) and 1DCNN with long short-term memory (LSTM) architecture which can automatically differenciate depressed or not depressed people.CNN-based raw signal modelling techniques based on voice source related information were proposed in dubagunta et al. A combined regional CNN and LSTM model were proposed to take a sentence as a region and the sentence were divided by various portion so that the useful affective information in

### each region can be extracted in wang et al. In xu et al. Here bidirectional long short term memory (BiLSTM) used to capture the context information effectively, and able to gain higher precision, recall, and F1 score.

### To enhance the model predictability here use multi-task learning with a multi-channel convolutional neural network use additional inputs which helps in detecting mental disorders gamaarachchige et al.Build a MIL predictive model using Long Short-Term Memory (LSTM) and gated recurrent unit (GRU) to detect

### depression from social network users with an accuracy of 74.65\%

### .In chao et al. depression scale are predicted by multi task sequence learning.Here encode the dynamic temporal information of the abnormal audio visual behavior by long short memory recurrent neural network (LSTM-RNN).

### In li et al. proposed an electroencephalography (EEG) based novel approach for recognizing mild depression.CNN and functional connectivity gives this model a classification accuracy of 80.74\%.

### statistical descriptors on gaze, pose; low level audio features

### and head pose and text features were also extracted as feature. Gaussian Mixture Model (GMM)

### clustering and Fisher vector approach were used on the visual data and classify them by using Support

### Vector Machine (SVM) and neural networks dham et al.Proposed a Deep Learning based approach called “Sentiment and

### Semantic Based Emotion Detector (SS-BED)" Considering both

### sentiment and semantic based features for more accurately predicting user emotions from their utterances Chatterjee et al.

### Lam et al. proposed transformer based topic modeling and deep 1D convolutional neural network (CNN) using acoustic feature modeling.

### Focusing on three thing a hybrid architecture were propsed in yang et al. Firstly,for estimating the Patient Health Questionnaire depression scale (PHQ-8) they designed A Deep Convolutional Neural Network (DCNN) and Deep Neural Network (DNN) based audio-visual multi-modal depression recognition model,secondly for inferring the physical and mental conditions of the individual from the transcripts of the interview A Paragraph Vector (PV) and Support Vector Machine (SVM) based model were built and lastly A Random Forest (RF) model for depression classification.

### DK-LSTM is proposed li et al. to identify people with depression which is a novel design based on LSTM deep learning algorithm .In bahera et al Co-LSTM is proposed which is an ensemble model combining CNN and LSTM aimed at examining big social data from any particular domain.

### In terms of the depression classification process, help vector machines, logistic regression and multi-layer neural networks are widely used models (Cohn et al. 2009; Nasir et al. 2016; Gong and Poellabauer 2017; Cai et al. 2018, please) .Deep learning frameworks such as convolutionary neural networks (CNN) (Yang et al. 2017; Rodrigues Makiuchi et al. 2019) ,long-short-term memory (LSTM) networks (Orabi et al. 2018) or their mixture have been built over recent years. CNN-LSTM (Ma et al. 2016) is being explored for the diagnosis of depression. These models are leveraged to automatically learn semantically rich features from text, audio, or visual data pampouchidou et al.However, these studies train models utilizing static multimedia data, without considering temporal data.

# Chapter 4

**Existing Dataset**

**4.1 Used dataset**

Dataset are used for various tasks like Classification, Regression, Function learning, Clustering. In this paper datasets are reviewed for classification purpose. Figure 4 shows that DAIC-WoZ dataset are the most used dataset in our research. We found 2 papers from 2020 and 1 papers from 2018 where DAIC-WoZ dataset were used. The dataset contains audio and video recordings as well as the transcripts of clinical interviews .DAIC-WoZ dataset contains 189 subjects in total. Dataset are split into train, test, and development set with 107, 35, 47 subjects respectively. CLPsych dataset was developed from public twitter users status between 2008 and 2013 with the help of Twitter application programming interface (API) which were updated in the form of a statement of diagnosis and combining them. In this dataset, up to 3000 of their most recent public tweets were included and they were different from each other. This dataset consists of 477 depressed users, 396 PTSD (an anxiety disorder caused by very stressful, frightening or distressing events) users and 873 control users.

AVEC 2016 dataset is used for audio and video features. From audio features, 5 formant features and 74 prosodic and voice quality features. AVEC 2016 provides histogram of oriented gradients (HOG) features, eye gaze features, and head pose features when OpenFace Framework were applied on the video features.

The AVEC-2014 dataset is an audio-visual depression corpus that contains 300 videos of subjects (one subject per recording) recorded by a webcam and a microphone. This dataset includes 84 subjects, in three recordings 18 subjects were appeared, in two recordings 31 subjects were appeared and remaining 34 subjects were appeared in one recording. 20 minutes to 50 minutes were The duration of each recording with an average duration of 25 minutes. So 240 hours is The total duration of all clips .The age of the subject ranging from 18 to 63 years with the average of 31.5 years.

AVEC2013 depression database contains 340 video clips from 292 subjects which is a sub-challenge used a subset of the audio-visual depressive language corpus (AVid-Corpus). This videos were recorded by a webcam or microphone. Each video clip consisting one subjects without no constrains. 25 minutes is the average length of each video clip. Subject age ranging from 18 to 63 with an average of 31.5 years.

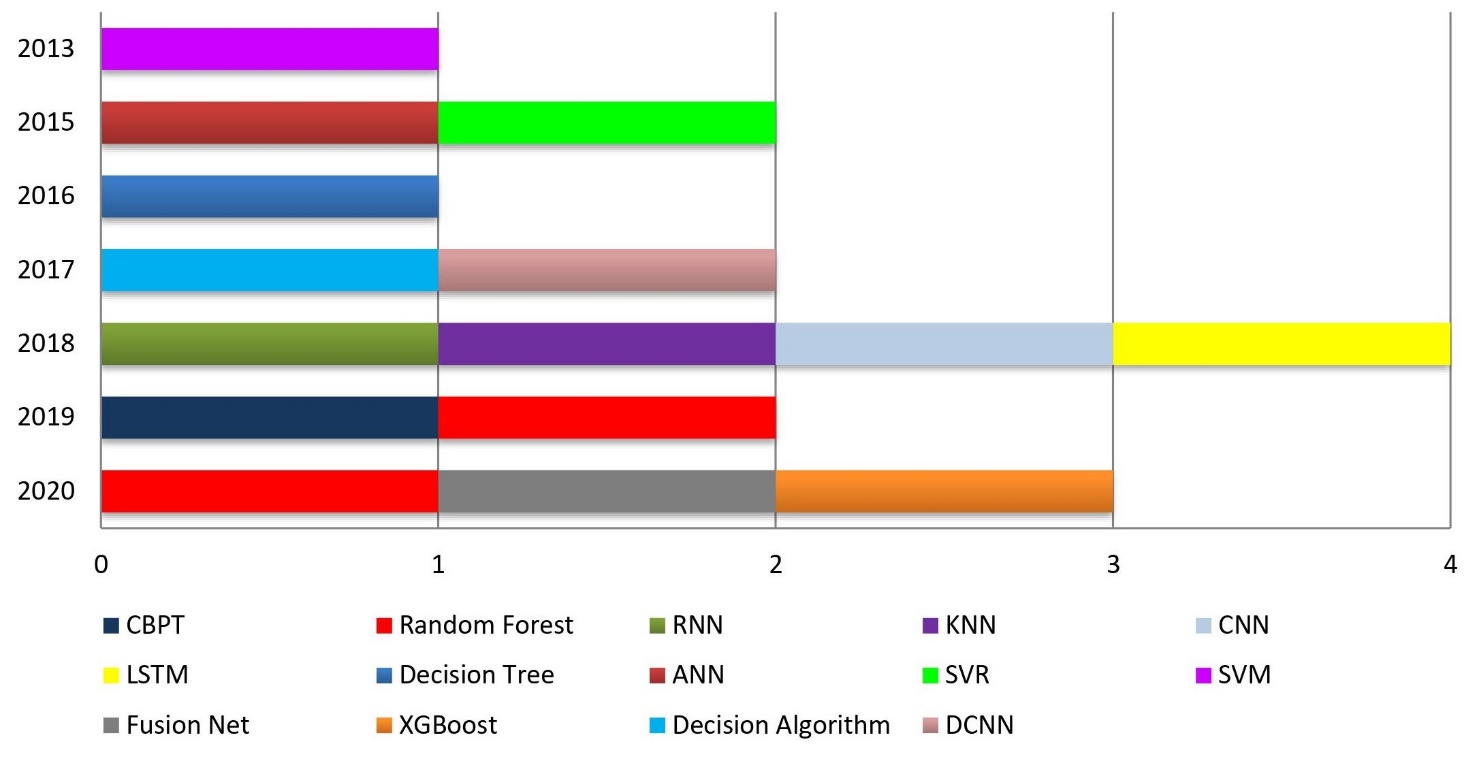
DASS-21 dataset , consists on 21 question for the Depression, Anxiety and Stress Scale questionnaire. Each of the scales of Stress, Anxiety and Depression 7 questions are required.

GPC dataset which is General Psychotherapy Corpus is a clinical approach consisting over 1,300 transcribed therapy sessions.

The Reddit Self-reported Depression Diagnosis (RSDD) dataset which contains some post of reddit users with two groups . In depression group, there are there are 9210 users and 107,274 users in the control group.

In our research ,we find the DAIC-WoZ and CLPsych 2015 are mostly used for depression detection which are graphically represented by figure 4 . Avec2013,Avec2014 and Avec2016 were used for many depression detection sub-challenge.

Figure: 4.1



# Chapter 5

**Discussion**

Depression detection is a highly application oriented research topic . Machine learning and deep learning both algorithm build upon several method for detecting depression using supervised semi-supevised or unsupervised learning algorithm .

some study exploit machine learning techniques for detecting depression with a rich, diverse, and

discriminating feature set from social media text and

behavioral trends of different users.A unique model were built to categorize a human as depressed or not depressed by considering Acoustic based feature.The persons response to the PHQ-8 Questionnaire are uploaded to the server .The depression classification model(DCM) designed with SVM decide the person depressed or not. More classification algorithm can be applied for gaining better accuracy.A new approach for detecting depression designed on facial expression to differentiate between

depressed and normal individuals, using machine learning

classification algorithm. Random forest classification algorithm gives an accuracy of 80.1%.

visual-based nonverbal

behavior analysis for automatic depression diagnosis is a comparatively good depression detection technique .Deal with background noise become a good challenge for facial dynamic analysis to diagnose depression.

More machine learning models can be applied to find out a more dependable way to measure feature impacts and degree of depression.

Deep Learning is a growing sector for depression detection.

Handling the word vector

classification task and the statistical feature classification task , a multitask learning model was proposed with DNN classifier and FusionNet .This method attempts to overcome low classification performance.

Here text-based word features are extracted with XLNet.Depression can be detected from clinical interviews using hierarchiacal attention network external affective information and external features are integrated into the attention mechanism which helps to improve the model performance.Dialogue interaction may be a challenge for this model .Another challenge faced by deep learning methods is big social data handling without losing valuable information needed for classification.

These models classify or clusters data from labelled or unlabelled dataset. The dataset were described in the existing dataset portion.

**5.2 Problem Statements**

There are some significant scientific challenges both machine learning and deep learning algorithm for depression detection.

* **Feature selection:** Reducing irrelevant or partially relavant features is the main objective of feature extraction. Due to unstructured big data feature can to be extracted properly by existing model.
* **Unstructured and unlabeled data:** The accuracy may be reduced by unstructured or unlabelled data , and causes data loss.
* **Unexpected data:** Irrelevant or unexpected data causes problem to classify data label which decrease model accuracy.
* **Evaluation:** Accuracy, Precision, Recall, and F1-scores are used for evluation measures.The confusion matrix use to compare predicted value with the actual target values by the machine learning model.

# Chapter 6 Conclusion

**6.1 Conclusion**

Social media diagnosis of mental illness may be considered a difficult activity, largely because of the ambiguous nature of mental illnesses. This research field has begun to grow in recent years with the continuous rise in the prevalence of social media sites that have become an integral part of people's lives. These sites have been made to reflect the personal life of the users on several levels through this intimate relationship between social media platforms and their users. In such a setting, a wealth of knowledge concerning one's existence is provided to researchers. Many researchers used traditional machine learning algorithms and deep learning algorithms to detect depression in various criteria. This survey paper provided a direct comparison of these papers and a reasonable view of this area, but this research cannot provide an in-depth review of those papers.

**6.2 Future Works**

In future studies, the following points could be helpful.

* Device efficiency is a crucial factor in this. During the training process, the elimination of unnecessary and irrelevant features improves system performance.
* In the classification methods, the selection stage will play a crucial role in the future.
* For output assessment, the use of ensemble classifiers is more feasible than both single classifier and hybrid classifier.
* More accurate feature selection must be considered
* The size of the dataset shuold be expanded.
* To handle huge data Deep learning should shifting to unsupervised and semi-supervised learning

There are several places for potential work that are clear. Applying newly designed techniques to a broader variety of data sets, comparing findings with many complementary success metrics and disclosing statistical proof can help to identify optimal deep learning approaches as well as conventional machine learning methods for future class imbalance implementations. Experimenting with modern hybrid and cluster-based machine learning approaches along with deep learning approaches to address depression level in the context of big data that can prove useful for the future of big data analytics

=0=