A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning

by

Md Tahmidur Rahman Ullas
15101038
Mariyam Begom
15101003
Anamika Ahmed
15101052
Raihan Sultana
15101139

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2019

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

<table>
<tr><td>Md Tahmidur Rahman Ullas
15101038</td><td>Raihan Sultana
15101139</td></tr>
<tr><td>Mariyam Begom
15101003</td><td>Anamika Ahmed
15101052</td></tr>
</table>

Approval

The thesis titled "A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning" submitted by

1. Md Tahmidur Rahman Ullas (15101038)

2. Raihan Sultana (15101139)

3. Mariyam Begom (15101003)

4. Anamika Ahmed (15101052)

Of Spring, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 8, 2019.

Examining Committee:

Supervisor:
(Member)

_____

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

_____

Dr. Jia Uddin
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

_____

Md. Abdul Mottalib, PhD
Professor Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Depression, a major depressive disorder and anxiety are common medical illness which cause several symptoms that affect the way a person feels, thinks, and the way he/she acts. These disorders are not only hard to endure, but are also risk factors for heart disease, panic attacks, dementia, and thus causing severe distress and functional impairment.Overall, more than 50% of the general population in middle and high-income countries suffers from at least one of these mental disorder at some point in their lives. In order to detect these disorders at an early age, we have proposed a model that uses a standard psychological assessment and machine learning algorithms to diagnose the different levels of such mental disorders. In our proposed model we used five different types of AI algorithms: Convolutional neural network, Support vector machine, Linear discriminant analysis, K Nearest Neighbor Classifier and Linear Regression on the two datasets of anxiety and depression. These algorithms are used to find the severity level of anxiety and depression, a patient is going through. In this paper we compared the results of the five algorithms on the two datasets separately on the basis of different measurement metrics. The proposed model achieves the highest accuracy of 96% for anxiety and 96.8% for depression using the CNN algorithm while its results were compared with the other five algorithms we have used for our model.

Keywords: Machine Learning; Depression; Anxiety; Support vector machine; Convolutional neural network ; Linear Regression ; K Nearest Neighbor Classifier ; Linear discriminant analysis

Dedication

We would like to dedicate this thesis to our loving parents, families and friends. . .

Acknowledgement

Firstly we would like to thank the Almighty for enabling us to initiate our research, to put our best efforts and successfully complete it.

Secondly, we offer our genuine and heartiest appreciation to our regarded Supervisor Dr. Md. Ashraful Alam for his contribution, direction and support in leading the research and preparation of the report. His involvement, inclusion and supervision have inspired us and acted as a huge incentive all through our research. Our thesis would not have been possible without his guidance and motivation.

We would also like to acknowledge the help we got from various assets over the internet; especially from fellow researchers' work. Our research would not have been possible without these helps.

Lastly but not the least, we would like to thank our parents, for giving us birth at first place and supporting us spiritually through out our life. We would love to dedicate this thesis to them.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

According to World health report around 450 million people suffer from mental health problems which shows how mental disorders are the leading cause of disability and poor health condition worldwide [3]. On the other hand depressive disorders are the fourth leading cause of global disease burden and are expected to be second leading by 2020 [2]. Depression a very common yet a very serious mental illness that negatively affects how a person feel, what kind of thoughts come in their mind and how they react. It causes feelings of sadness and loss of interest in activities once enjoyed. It affects a person's ability to concentrate in work, studies and daily works. One in 15 adults are affected by depression in a given year[10] which shows how serious the illness is. Anxiety disorder is another major psychological issue that affects nearly 30 percent of adults at some point of their life[10]. It involves excessive fear and anxiety. Due to depression, anxiety and stress a person might end up hurting themselves and also have suicidal thought. These are the two major issues on which we focused in our work.

We live in a society where due to social stigma and embarrassment people are still uncomfortable talking about their mental health issues. Moreover most of the time people suffer through such illness yet do not know about it since they cannot recognize the symptoms. A large proportion of people do not have access to mental health care, and others who have access often hesitate to take mental health checkups. Due to their hesitation of taking mental health checkups, they never know whether they are having such disorders and hence are left untreated. So in our paper we propose a system that would diagnose the above two common psychological disorders: depression and anxiety using machine learning algorithms. It will help to find the severity level of the disorder on the patient.

After having discussions with counselors at Brac University, counselling unit, we have come across with the information that most of the psychologist follow a standard scale for initial diagnosis of common mental illness. Such standard scale questionnaires for depression and anxiety were collected from department of Psychology, of University of Dhaka, which was filled up by people having psychological disorder and also by people that have healthy mental condition. Then such filled questionnaires were collected from archive of Dhaka Medical College and Rangpur Medical College. Using these files our dataset was prepared. Then we applied five machine learning techniques Convolutional Neural Network, Support Vector Machines, Lin-

ear Discriminant Analysis, K Nearest Neighbor and linear Regression algorithm on the preprocessed dataset. The performance metrics of each algorithm was compared to find the optimal algorithm to find the severity of the anxiety or depression.

1.2 Objectives

Our aim is to prepare a model which can predict the severity level of depression and anxiety of the patients that can be helpful for the psychologist and counsellor at the initial step of their treatment. This will help them to decide the treatment procedure suitable for the patients. Later, comparative analysis is to be carried out by comparing different performance metric of different machine learning algorithms we will be using in this paper to find the optimal algorithm for our proposed model to detect the severity of the disorders.

1.3 Literature Review

We have gone through several papers and journals in order to understand the contribution of machine learning in mental illness diagnosis of patients of different age group. The research on this topic started since 1980's. Roland H.C. Yap, David M. Clarke described in their paper an expert system, MILP (Monash Interview for Liaison Psychiatry) that can diagnose mental disorder based on DSM-III-R, DSM-IV and ICD-10 using constraint-based reasoning. This system was developed using Constraint Logic Programming (CLP) language. However it was a very straight forward process[1]. On the other hand Rozita Masri, Hajar Mat Jani used three artificial intelligence reasoning techniques, rule-based reasoning, fuzzy logic and fuzzy-genetic algorithm to diagnose and treat the mental patients. Their proposed model also suggested suitable treatment plans to the users[7]. Moreover in order to diagnose adult depression by neoro-fuzzy approach, Subhagata Chattopadhyay developed a hybrid system using Mamdani's Fuzzy logic controller on a Feed Forward Multilayer Neural Net. This algorithm could successfully diagnose depression with an accuracy of approximately 95.5%.

Many research aimed particularly on diagnosing mental illness in children. M.R. Sumathi, B. Poorna used machine learning techniques, Bayesian Networks and Fuzzy clustering to predict common mental health problems in children, like Attention problem, Anxiety problem, Development delay, Attention Deficit Hyperactivity Disorder, Pervasive Development Disorder, etc. They compared their proposed framework which is Fuzzy Clustering based Bayesian Framework with Expert knowledge based Bayesian network (EKBN) and Data based Bayesian network (DBN) and came to the conclusion that their framework was more accurate [30]. On the other hand, Anjume S, Amandeep K, et al. also worked to find the best algorithm to diagnose common mental disorders in children. They used Feature Selection algorithm on the attribute sets of their dataset to extract more relevant attributes. Then they compared the accuracy level of the three machine learning techniques, Multilayer Perceptron, LAD Tree and Multiclass Classifier using three measures that are Kappa Statistics, ROC Area and Accuracy. It was proved that Multiclass Classifier produced more accurate results on the attributes that were selected[18].

On the other hand many research included image processing to study the nonverbal

behavior of patients to identify psychological state of a user more precisely [17]. David DeVault, Ron Artstein, et al. introduced a prototype of a virtual human interviewer, SimSensei Kiosk that could assess depression, anxiety or post-traumatic disorder (PTSD) by analyzing a patient's gesture. It can sense real-time audiovisual behavior signals from patient. It was mainly designed to make patients more comfortable sharing their problems face-to-face with the interviewer and the user's vocal parameter, facial expression and body gestures helps the system to diagnose the problem more accurately. Their design also have the capability to engage the user in 15-20 minutes interaction. This also helps to monitor the user's mental health condition by comparing the result after a month or a weak [13]. On the other hand, Stefan Scherer, Giota Stratou, et al. also investigated in detecting these psychological disorders. In their investigation, they found a new dataset, Distress Assessment Interview Corpus (DAIC) which included 167 interaction between an interviewer and a paid participant and found out that it showed a correlation between non-verbal behavior and a specific psychological disorder. Some of the behavior descriptors they analyzed were vertical head gaze, vertical eye gaze, smile intensity and smile duration. They also analyzed manual behavior annotation like hands self-adaptors and legs fidgeting [12].

Other studies included disorder detection from social media data like facebook, twitter and instragram. For example, Guangyao Shen, Jia Jia, et al. proposed a multimodal depressive dictionary learning model that can detect depressed people on twitter. They analyzed twitter dataset which showed difference in behavior in depressed and non-depressed user [21]. In another paper, Janith Weerasinghe, Kediel Morales, et al. stated that patients having mental illness can be differentiated from normal people by their language pattern. They also suggested a model that could detect mental illness in twitter. They used feature analysis by creating feature vectors and used word unigrams and word clusters and used feature importance measures and statistical methods in order to detect important features. Their main aim was to diagnose mental illness by analyzing language pattern [31].

1.4 Contribution

In our paper we worked on how to predict the two common phenomena: depression and anxiety more accurately. In the above subsection, we came across various approaches to predict various kind of physiological conditions. Some used attributes of the patients as datasets that was further used for training purpose to predict specific illness, whereas some used image processing for analyzing depression from the facial expressions and some also used social media data to analyze disorders from behavior of users. However we chose a different approach. We worked with the standard structured questionnaires for depression and anxiety.

After discussing with some licensed psychologists, we came across the fact that, many licensed counsellors and psychologist uses some tools in the initial state to find out the illness and the severity of the illness the patient is going through. They follow separate standard structured questionnaires for depression and anxiety, known as 'depression scale' and 'anxiety scale' respectively. The depression scale developed by Uddin and Rahman (2005) was made to detect depression, for the use of people in Bangladesh, whereas Bangla anxiety measuring scale developed by Farah Deeba and Roquia Begom (2004) [4] was used for measuring the severity of

anxiety of patients. Using the set of filled questionnaires of depression and anxiety that was collected from archive of two reputed hospitals of our country, dataset was prepared. Thus, this data was used for the training our model.

The two questionnaires were relevant to the climate and culture of Bangladesh. Since it was separately prepared for the people of our country, it contained certain lifestyle related questions that was connected to the way of life of people of our country.

1.5   Thesis Orientation

Our paper consists of four chapters. Chapter one consists of the introduction, which highlights the motivation behind the thesis and the literature review that states all the previous contributions in papers and journals that are relevant to our thesis. Section two illustrates the proposed model of our system. It consists of three subsections. The first subsection describes the system architecture of the system; the second subsection describes the dataset collection and preparation; the third subsection describes the algorithms applied in our paper elaborately. Chapter three mentions about how we implemented the algorithms to find the most accurate results and comparative analysis of the results found. It also states the comparison between the algorithms. Finally chapter four concludes the thesis and states our limitations and future plan.

Chapter 2

Proposed Model

2.1   Architecture of the Proposed Model

Our proposed model consists of a user interface where the patient will be able to interact with the system efficiently and more comfortably. As the patient starts using our system, he/she encounters with an interface displaying some mental disorder cases: Anxiety and Depression. In each case, there will be a set of questionnaires which the patient has to answer. The questionnaire we used is the standard psychological evaluation, collected from Department of Psychology, University of Dhaka.
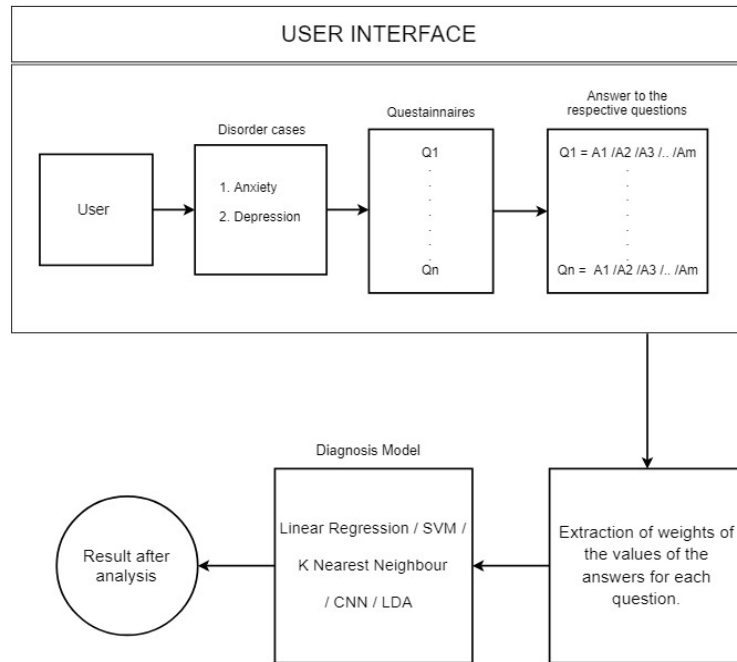
Figure 2.1: System Architecture Design.

According to the Fig. 2.1, the patient comes across an interface provided by our model. When the patient chooses a case, he/she has to answer every question accordingly. There will be a set of options provided in the answering section, where the patient has to choose the most appropriate option based on his/her condition or suffering. For every question Q1, Q2,..., Qn , there will be a set of answers, A1/A2/A3/../Am where each option, i.e. A1/A2/A3/../Am has corresponding weighted values. Later these weighted values of the answers for each question are extracted and calculated by the system separately. This way the patient has to answer the questions for the other cases available in our system. These data are then given to the diagnosis model. Then the diagnosis model gives the severity level of the following mental disorder case for the following patient. We have trained our model using five different machine learning algorithm such as Linear Regression, K nearest neighbor, Support Vector Machines, Linear Discriminant Analysis and Convolutional Neural Network algorithm. Each algorithm has its own specialty and gives results effectively depending on the cases. Then by comparative analysis we have found the optimal algorithm for our model.

2.2 Data Collection and Preparation

The most difficult and time consuming part of our thesis paper was collecting data. We have tried finding the datasets on different websites that consists of publish datasets, but could not find any that goes best with our purpose. Hence we had to make the dataset of our own. For this purpose, we have collected separate standard scale questionnaires for the two common psychological disorder, depression and anxiety as mentioned in Chapter 1. The hard copy of the questionnaires were collected from department of Psychology, University of Dhaka. One of the core reason for using the depression scale in Bangla and Bangla Anxiety Measuring scale is that it is relevant to the climate and culture of Bangladesh. These both questionnaires contains certain lifestyle related questions that are connected to the way of life of the people of this country.

Among the two questionnaires, the first one was the depression scale developed by Rahman and Uddin (2005). It contains 30 questions and each of them have five options with weight 1, 2, 3, 4 and 5 respectively. This weight is used to calculate the score which is further used to measure the severity of depression of the patient. Higher score indicates higher level of depression. The levels of severity of depression was divided based on the score of the answers as shown in the Table 2.1.

Table 2.1: Severity level of Depression scale.

Severity of Depression	Corresponding range of score
Minimum	30-100
Mild	101-114
Moderate	115-123
Severe	124-150

The second questionnaires, Bangla anxiety measuring scale developed by Farah Deeba and Roquia Begom (2004) [4] consists 35 questions. Each question contained five options with weight 0, 1, 2, 3 and 4 respectively. Again, the weight is used to calculate the total score. Four levels of severity of anxiety was divided as shown is the Table 2.2 based on the total score calculated.

Table 2.2: Severity level of Bangla Anxiety Measuring scale.

Severity of Anxiety	Corresponding range of score
Mild	54 and less
Moderate	55-66
Severe	67-77
Profound	78-135

Using the two questionnaires we collected the data for our model. We are extremely thankful to Dr. Sufia Khatun, Head Of Nursing (Retired) and Dr. Ahad Shah, Psychologist, Rangpur Medical College and Hospital for their contribution in our data collection process. They have provided us with the filled up questionnaire of their patients anonymously. The age of the patients who filled the questionnaires were between 15-35 years, and were both male and female. Among them some were

students, some were homemakers and others were corporate professionals. Using it we have recorded the data in separate csv files. Later we have preprocessed the data by removing anomalies from each file respectively. For this firstly, result from each column was summed up to check if it lied within the range given in the questionnaire forms. The column containing noise was removed entirely. Next, we checked each column for missing values, and removed the entire column if any missing values were found. Thirdly we removed columns containing duplicate values. This way the data was preprocessed.

Later, the well-known library of python, pandas was used to read the csv file to generate data frame from it. Using the data frame that was generated, and assuming the prediction value to be in Y column (dependent variable) and other variables in X column (independent variable), training and testing data was separated using k fold cross validation.

2.3 Algorithm

Linear regression

Regression analysis is a technique for analyzing multiple factors and is one of the most widely techniques used to find the relationship between two variables [8] [6]. It gives useful results from the conceptually logical process of using the equation which expresses the relation between a variable of interest (the response) and a set of related predictor variables. If a set of random data (x1, y1) T, (x2, y2) T, (x n, y n) T for two numerical variables X and Y, where X is a cause of Y. In this linear regression analysis, the distribution of the random data appears as a straight line in X, Y space when X and Y are perfectly related linearly. In this linear regression a straight line is formed in X,Y space when X and Y are perfectly related linearly and thus captures a relationship between the two variables. This line function can be shown as [9].

$$y = ax + b \tag{2.1}$$

In our system, the questionnaires are considered as the variable X and the output is considered as Y. The straight line formed gives the trend of the data. The lesser the distance between the data and the line, the more is the accuracy of the data.

Linear discriminant analysis algorithm

Linear Discriminant Analysis (LDA) is like a pre-processing step for machine learning and pattern classification application which is a common technique for dimensionality reduction problems and for feature extraction [22]. It is a classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes' rule.

In order to maximize the ratio of the between-class distance to the within-class distance, classical LDA projects the data onto a lower-dimensional vector space. This results in achieving maximum discrimination. The optimal projection (transformation) can be readily computed by applying the eigen decomposition on the scatter matrices [5]. The model fits a Gaussian density to each class, assuming that all classes share the same covariance matrix [5]. Three procedures are carried

out to achieve this goal. The first one is the distance between the means of different classes is calculated which is called the between-class variance or between-class matrix. Second, the distance between the mean and the samples of each class is calculated which is called the within-class variance or within-class matrix. Finally, the lower dimensional space is constructed which maximizes the between class variance and minimizes the within class variance [22].

Convolutional Neural Network

Convolutional Neural Network is one of the main categories in neural network that does image recognition, image classification, object detection and identify many other aspects of visual data [16].To make natural language processable to analog and handwritten documentations and to digitize texts , convolutional neural network performs optical character recognition (OCR). CNN can be also applied on sounds and to text analytics as well as graph data (using graph convolutional networks). Computers recognize images as an array of pixels depending on the resolution of the image. Convolutional layer is the first layer to extract features from an input image. When CNN takes images as an input, it classifies them to categories. The image inputs are ingested and processed in tensors, where tensors are matrices with additional dimensions (Eg. 3D, 4D etc.). Convolutional networks perceive images as three dimensional objects because images have red-blue-green (RGB) encoding. A convolutional network take in such images as three separate strata of color stacked one on top of the other. Technically, deep learning CNN is modelled to train and test, each input image that will pass through a series of convolution layers with filters which are also known as Kernels, Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. As the images passes through each layer, they are mathematically expressed as multi dimensional matrices and are described in terms of input and output volumes as shown in Fig. 2.2. Images in every layer will be expressed in different multi dimensional matrices. As discussed previously, images have RGB encoding, thus, for each pixel of an image, the R , G and B will have intensities that will be expressed by a number. These numbers are the elements of the multi dimensional matrices which together form the volume of the image. These pixels are then divided into square patches and passed through a filter , instead of considering the entire image pixel at once. The filters mentioned here known as kernels, which are also square matrices , smaller than the image itself but are of the same size compared to the square patches of the image pixel. Here, kernels help in finding the patterns in the pixels. Here, in each step, we get a dot product, and the results of that dot product in a third matrix known as an activation map.
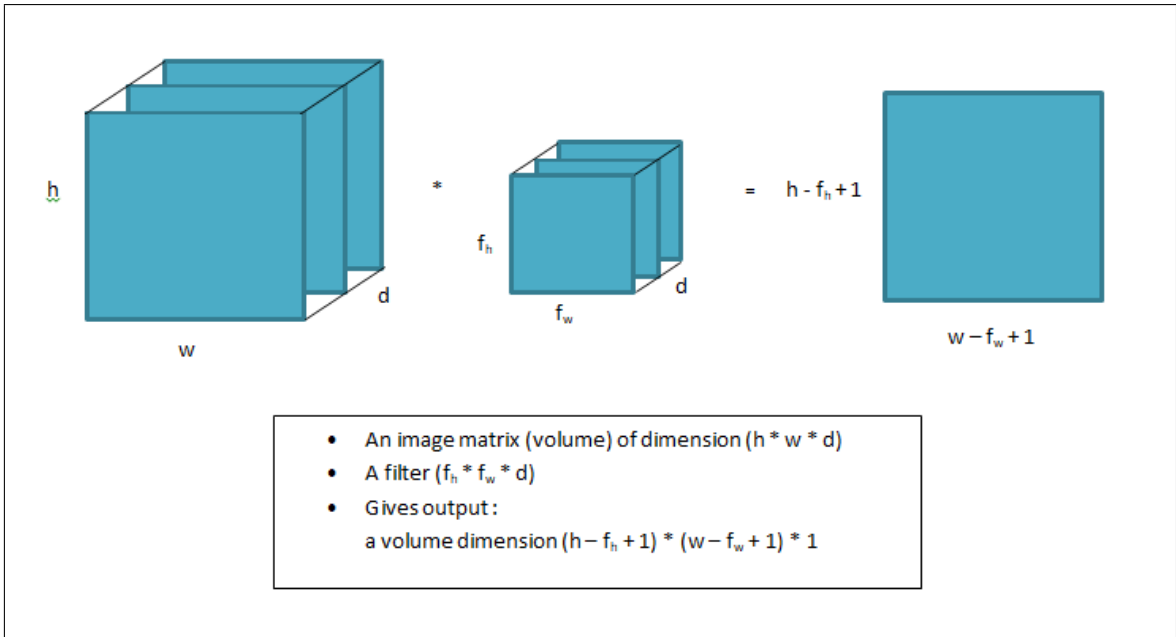
Figure 2.2: Image matrix multiplies kernel or filter matrix.

After performing convolution on the image, ReLU is applied for the activation to the matrix. ReLU stands for Rectified Linear Unit for a non-linear operation whose output is represented as

$$f(x) = max(0, x) \qquad (2.2)$$

ReLU is mainly used to introduce non linearity in the ConvNet. This helps ConvNet to learn non-negative linear values as shown in Fig. 2.3.
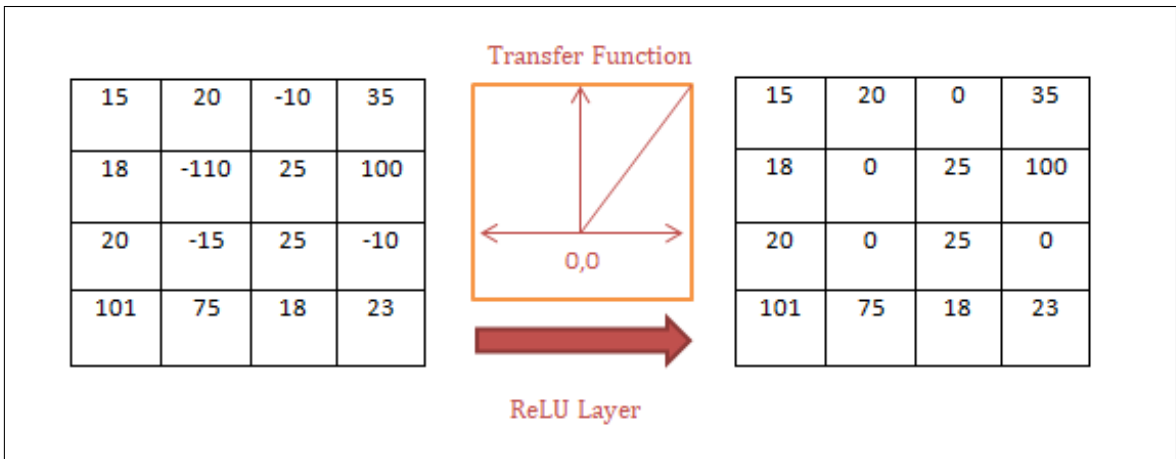
Figure 2.3: ReLU Operation.

The next layer in a convolutional network has three names: max pooling, downsampling and subsampling. Pooling is done to reduce the number of parameters of images which are large in size. The activation maps are fed into the downsampling layer and this process is applied on one patch at a time just like convolution. Here, the max pooling takes the patch matrix with largest value and places it next to the largest value of the other patch matrix, discarding the rest of the information in the activation maps. Later, convolutional layers are added till satisfied, resulting

in flattening of matrix into vector and then are fed into connected layer like neural network. The features are then combined together with the fully connected layers to form a model. Finally the outputs are classified using activation function such as softmax or sigmoid as per the Fig. 2.4 [25].
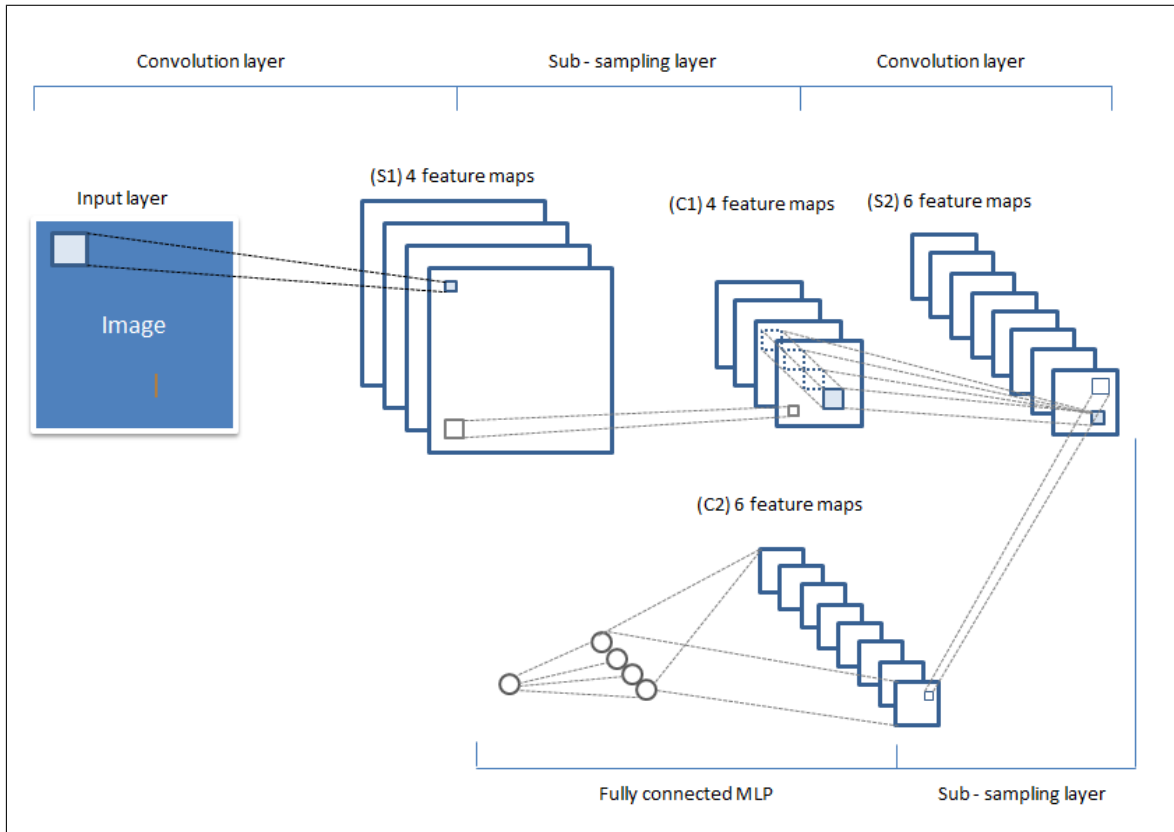
Figure 2.4: Sequence of transformations involved in a typical convolutional network.

K nearest neighbor

K Nearest Neighbour (KNN) is a supervised machine learning algorithm. Among all machine learning algorithms, it is one of the simple, effective and easy to implement algorithm. KNN is used to solve both classification and regression problem. It is a non- parametric and a lazy learning algorithm [11]. By non-parametric, it means that it does not make any assumptions on the underlying data distribution. The model structure is determined from the data. As we know in real world, most of the data do not follow the theoretical assumptions made such that in linear regression algorithm. Hence KNN algorithm is a good choice for classification study where there is less idea about the distribution of data. On the other hand it is known as a lazy algorithm because it does not use any training data point for doing generalization. It also means that the algorithm has no explicit training phase or has a small amount of training phase, due to which the training phase is quite fast. The training data is kept for testing phase.

Now how do a KNN algorithm work? The algorithm is basically based on feature similarity. By feature similarity it means that how it classifies a given data is determined by how closely out of sample feature resemble our training set.

Figure 2.5: Example of KNN classification.

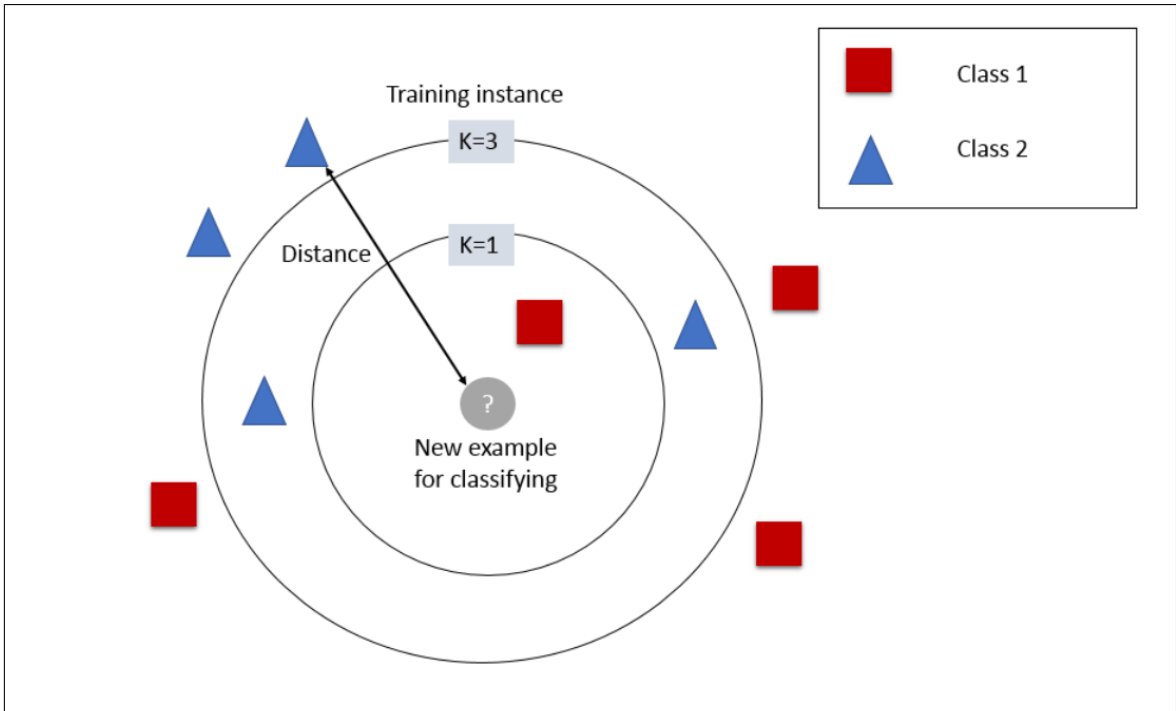As we can see in Fig. 2.5, the test sample must be classified to either the first class (class 1) or to the second class (class 2). Hence based on the feature similarity, when K=3 the sample will be assigned to class 2 since there are more number of triangles than the square inside the circle whereas when k=5, it is classified to the class 1 since there are more number of squares compared to number of triangle outside the outer circle.

As discussed above, KNN algorithms can be used for both classification and regression problems. However in our paper we are applying it for solving classification problem. Therefore a sample is classified depending on majority vote given by its neighbor, and the sample is assigned to the class that is most common among its k nearest neighbor as explained above [19] [11].

However in case of classification, An appropriate value of K is to be chosen in order to apply the algorithm. The success of classification is dependent on this value. In our paper using parameter tuning we found the value of k, which gives the best result. Then distance between data points are measured using distance measures such as Euclidean distance, Hamming distance, Manhattan distance and Minkowski distance. Using the distance, K closest neighbors are found. Then points are classified by majority vote of its k neighbor. Each object votes for their respective class. Then the class with most of the vote is taken as the prediction [11].

Support Vector Machines

Support Vector Machines are set of related supervised learning methods used for classification and regression. In this algorithm, the data are plotted as n-dimensional spaces where n represents the number of feature we use. The value of each feature represents the value of a particular coordinate. SVM takes the data as inputs and presents the output using a line which separates the data of the particular number

of classes. The classification is done by finding the hyperplane which is done by differentiating the two classes very well. Hyperplanes are the decision boundaries that help classify the data points. Many possible hyperplanes can be chosen to separate two classes of data points. Moreover, the dimension of the hyperplane depends upon the number of features as shown in the Fig. 2.6. SVM computes the hyperplane that maximizes the margin between the class boundary and the training examples when the data are linearly separable. This hyperplane is also known as the optimal hyperplane.The margin of the classifier is maximized using the support vectors, which are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane as shown in Fig. 2.7 [23].
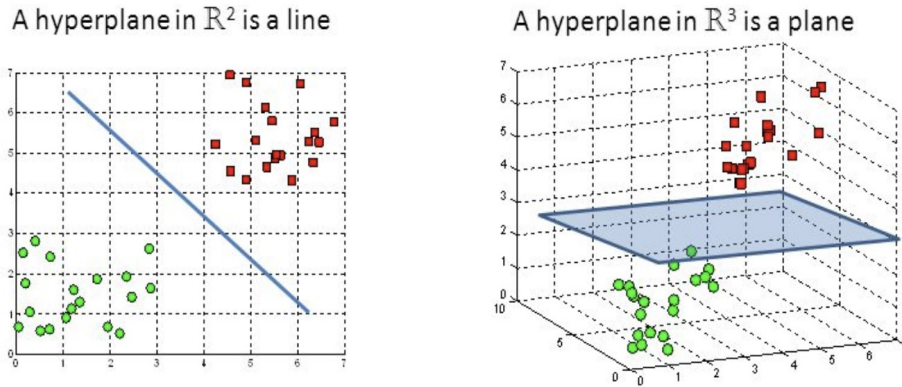
Figure 2.6: Hyperplanes in 2D and 3D feature space.
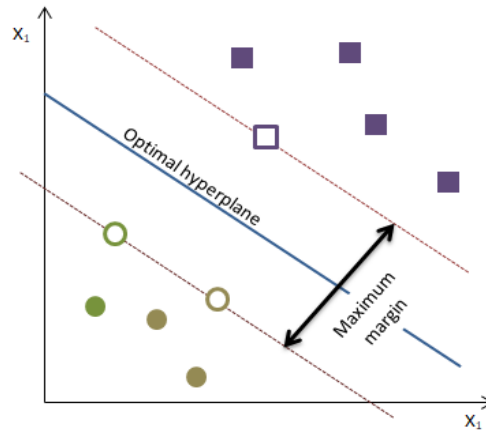
Figure 2.7: Optimal Hyperplane using the SVM algorithm.

However there can be data that are not linearly separable, the examples are mapped to a high dimensional space in a way that a separating hyperplane can be computed. This is done by adding one more dimension and call it as z-axis. Therefore, we get an equation ,

$$z = x^2 + y^2 \tag{2.3}$$

where z basically is the square of the distance of the point from the origin. Thus, the classes can be now separated linearly as shown in Fig. 2.8. The name of the mechanism that defines this mapping process of converting low dimensional input space to high dimensional input space is the kernel function. The algorithm gives good performance in the domain of Microarray data and thus, becomes a powerful classifier [26].
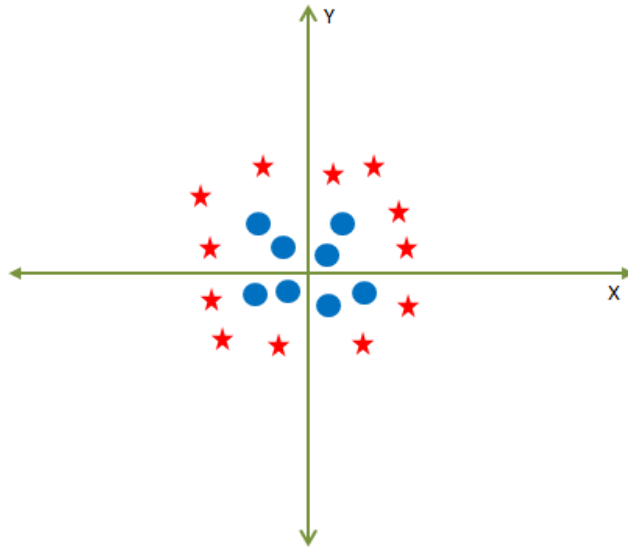
(a) Non linear data.
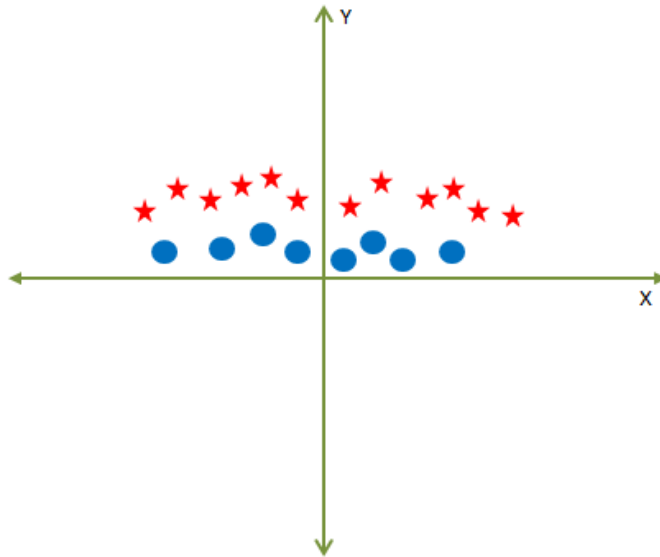
(b) Linear data.

Figure 2.8: Non linear data converted to linear data after applying kernel.

Chapter 3

Implementation and Result

As we have discussed earlier, we have applied the five machine learning algorithms: CNN, SVM, LDA, KNN and linear Regression on the two dataset separately. Then, we have used K-fold cross validation for training and testing. On the other hand we have done hyperparameter optimization using Grid Search strategy for finding the best hyperparameter values for our dataset. After applying the algorithms we have compared them using different performance metrics: accuracy, precision and recall.

3.1 K-fold Cross Validation

In machine learning, the dataset is divided into training set and testing set. The training set is used for training the model.Usually 20% of the total data is used for the training data and 60% of the total data is used as the training set. The testing data is used for evaluating the model. Sometimes another set of data, known as validation data set is used that provides unbiased evaluation of the model's fitness. When the error on validation data set is high, then we can say that we have an overfitting model. The dataset used for validation data is usually 10-20% of the total data. However this is not an efficient technique. One of the reason is that due to variability between training and testing set, the model can give less error for training set and higher error rate for testing set. Another reason is that when we train on less observation, the model will not give a good result and will overestimate the test error rate for the model to fit on the dataset [24] [27].

Therefore cross validation is used that splits the data into subsets, training the data into subsets and use other subset to evaluate the model's performance [24]. The variability between testing and training data set is reduced by applying multiple rounds of cross validation using different subsets of the same data. the validation result is then combined to find an estimate of the model's predictive performance. It is an efficient way to estimate a model's performance.

Out of different types of cross validation techniques, we have used k fold cross validation in our paper. In this technique dataset is randomly splitted into k groups or folds of approximately equal size. The first fold is used for testing and k-1 fold is used for training the model. For k times, the process is repeated and every time for validation purpose different group of data points or different folds are used [27]. One of the advantage of k fold cross validation is that every data points are tested exactly once and also used for training k-1 times. However if the value of k is low, the model becomes cheaper and more biased. On the other side, with larger value

of k, the model becomes expensive but less biased. Moreover with larger value of k the model may suffer from large variability. Keeping everything in mind, the most common value of k used is 10. Hence in our paper, we have used the value of k as 10.

3.2 Performance Metrics

3.2.1 Accuracy

In classification problems, the total number of predictions made by the system out of all kinds predictions made is known as Accuracy (A_γ).

$$A_\gamma = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.1}$$

Where TP is the true positive value, TN is the true negative value, FP is the false positive value and FN is the false negative value. The numerator consists of the correct predictions (true positive and true negative) and denominator consists of all kind of predictions made by the algorithm (true and false predictions) [15].

3.2.2 Precision

The measurement metric that is used to measure the positive patterns that are correctly predicted from the total predictions in the positive class (true positive and false positive) is known as Precision (P_n).

$$P_\gamma = \frac{TP}{TP + FP} \tag{3.2}$$

The value of 1 states that it is the best score of precision whereas the value of 1 states that it is the worst score of precision [15].

3.2.3 Recall

The metric that is used to measure the fraction of positive pattern that are correctly classified is known as Recall(R_n).

$$R_\gamma = \frac{TP}{TP + FN} \tag{3.3}$$

When we get the value of recall as 1, we can say that it gives the most accurate result, however is we get the value of recall as 0, we can say that we get the worst result [15].

3.3 Results Obtained

3.3.1 Linear Regression

The optimal parameter values of linear regression for our dataset was found using parameter optimization using Grid Search Algorithm. From Fig. 3.1 we can see, that the parameter min tolerance was set to 0.1. It specifies the minimum tolerance

for eliminating collinear features. The parameter ridge is set to 2. It specifies the ridge parameter for using in ridge regression [28]. For depression dataset, the error metric, root mean squared error(RMSE) is 0.693. The RMSE is defined as the sample standard deviation of difference between the observed value and the predicted value [14] [29]. On the other hand absolute error is 0.615. The squared correlation is 0.498. We know that as squared correlation is closer to 1, we can say that the model's prediction mirror true outcome, tightly. Whereas when the squared correlation is closer to 0, then we can say that either our model do not mirror true outcome or it only mirrors slightly [14]. Since our squared correlation is closer to 0, it indicates that the algorithm do not mirror true prediction properly. On the other hand the prediction average is found to be 0.985 as shown is Fig. 3.1.

```
Parameter set:

Performance:
PerformanceVector [
*****root_mean_squared_error: 0.693 +/- 0.003 (micro average: 0.693 +/- 0.000)
-----absolute_error: 0.615 +/- 0.003 (micro average: 0.615 +/- 0.318)
-----squared_correlation: 0.498 +/- 0.007 (micro average: 0.498)
-----prediction_average: 0.980 +/- 0.011 (micro average: 0.980 +/- 0.978)
]
Linear Regression.min_tolerance = 0.1
Linear Regression.ridge = 2
```

Figure 3.1: Hyperparameter values and performance of Linear Regression for Depression dataset.

As we can see in Fig. 3.2, the root mean squared error found for anxiety dataset is 0.693, which is high since it is closer to 1. However the prediction average is found to be high, which is 0.980.

```
PerformanceVector:
root_mean_squared_error: 0.693 +/- 0.003 (micro average: 0.693 +/- 0.000)
prediction_average: 0.980 +/- 0.017 (micro average: 0.980 +/- 0.978)
```

Figure 3.2: Performance vector of Linear Regression for Anxiety dataset.

3.3.2   Linear Discriminant Analysis

LDA has given 93% accuracy for anxiety and 87.9% for depression on the datasets after parameter tuning has been done. In the case of anxiety, the parameters : solver, shrinkage, priors, ncomponents, store covariance were set to their default values. The default solver is 'svd' (Singular value decomposition) which is used for both classification and transformation without depending on the calculation of the

covariance matrix. This has a great advantage on situations having larger number of features. However we have set Tol as 0.0001. Tol is a threshold used for rank estimation in SVD solver. For depression, the parameter, solver is set to lsqr , which can be combined with shrinkage and is an essential algorithm that works efficiently for classification. Shrinkage is assigned as 0.999 and Tol has been again set to 0.0001. Other parameters like priors, ncomponents, store covariance are assigned to their default values. Thus, these parameters were tuned accordingly to give the most accurate result on our dataset. As per the Fig. 3.3, for anxiety we have got 93.7% of precision and 92.9% of recall. For depression, the precision is 93.6% and recall is 87.9%, according to the Fig. 3.4.

```
Evaluating LDA
Parameters:
    Solver: svd
    Shirkage: None
    Priors: None
    NComponents: None
    Store Covariance: False
    Tol: 0.0001
    =================================================================================
LDA: Accuracy -> 0.929697 | Standard Deviation -> 0.009692
LDA: Precision Weighted -> 0.937877 | Standard Deviation -> 0.007705
LDA: Recall Weighted -> 0.929697 | Standard Deviation -> 0.009692
LDA: F1 Weighted -> 0.923683 | Standard Deviation -> 0.011820
    =================================================================================
```

Figure 3.3: Hyper parameter values used in LDA algorithm for Anxiety dataset.

```
Evaluating LDA for Depression analysis
Parameters:
    Solver: lsqr (default was 'svd')
    Shirkage: 0.9999999999999999 (default was 'auto')
    Priors: None
    NComponents: None
    Store Covariance: False
    Tol: 0.0001
    =================================================================================
LDA: Accuracy -> 0.879167 | Standard Deviation -> 0.073243
LDA: Precision Weighted -> 0.936956 | Standard Deviation -> 0.033783
LDA: Recall Weighted -> 0.879167 | Standard Deviation -> 0.073243
LDA: F1 Weighted -> 0.890763 | Standard Deviation -> 0.061647
    =================================================================================
```

Figure 3.4: Hyper parameter values used in LDA algorithm for Depression dataset.

3.3.3 Convolutional Neural Network

CNN achieved nearly 96% accuracy for anxiety and 96.8% for depression. For both depression and anxiety, the epoch was set to 45 whereas its default value is 100. We have reduced the epoch size to 45 because for 100 epoch size, model was overfitting. As a result, increasing the epoch may give better accuracy but there can be changes that will make the model overfitting. Epoch is the number of iteration the program runs and tries to minimize the error. The higher the epoch, the higher the accuracy. Every time an epoch is run, it compares the previous validation result with original result. Thus, if any error is found, it tries to minimize the error by upgrading the layer function. The validation , testing and training accuracies of both depression and anxiety are shown in the Fig. 3.5 and Fig. 3.6.
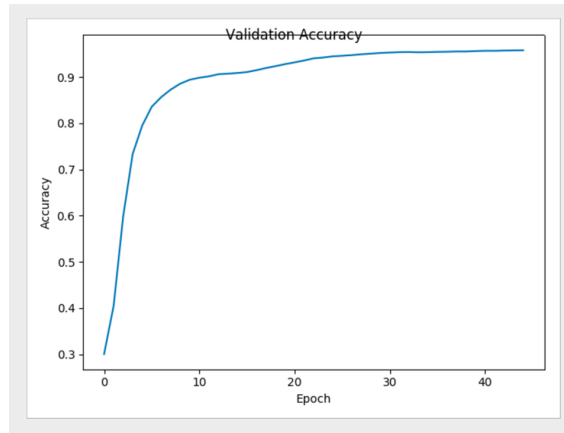
(a) Validation accuracy.
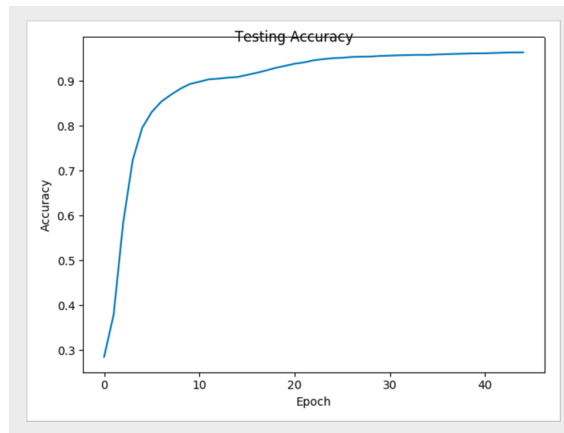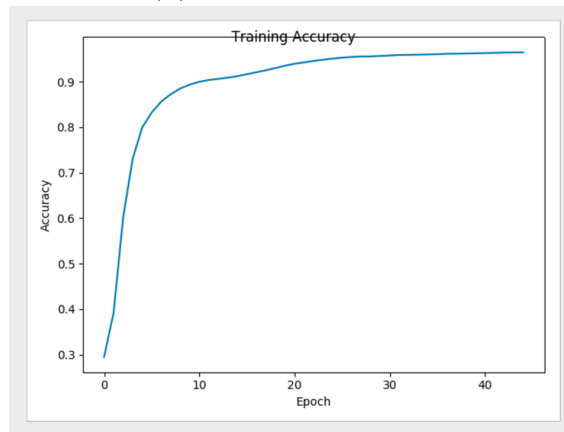
(b) Testing accuracy.

(c) Training accuracy.

Figure 3.5: Validation, testing and training accuracy for Anxiety.
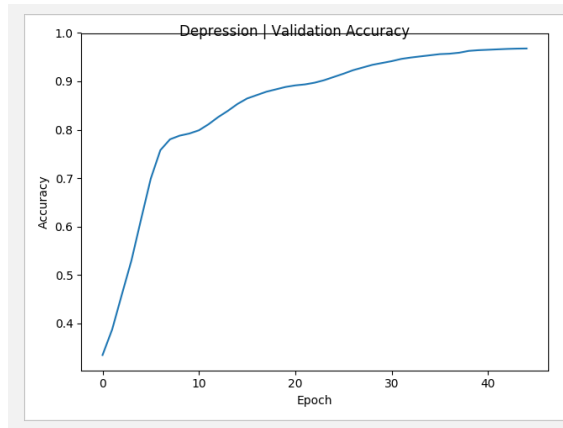
(a) Validation accuracy.
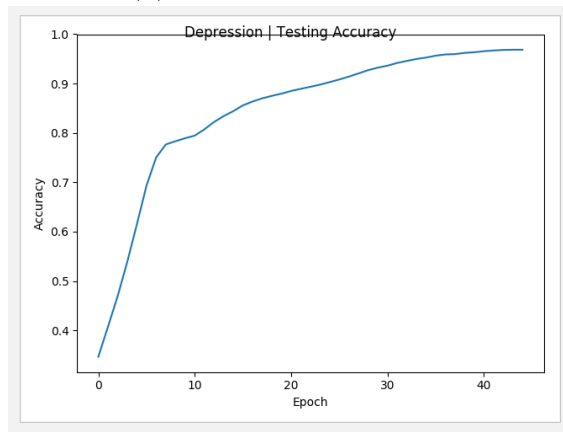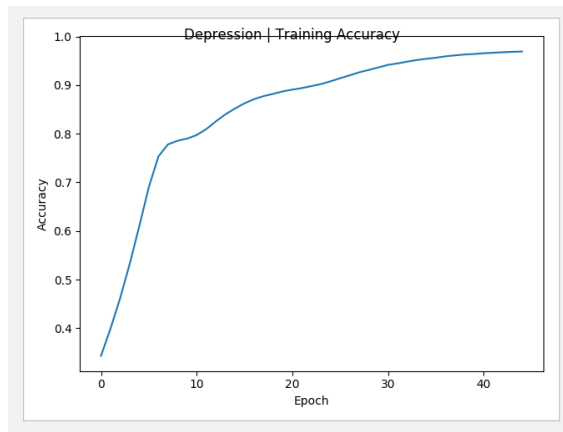
(b) Testing accuracy.

(c) Training accuracy.

Figure 3.6: Validation, testing and training accuracy for Depression.

Moreover, the early stop parameter was set to 3. The default value of early stop is 5. Early stopping is a technique used for controlling the overfitting in machine learning models and it is done by stopping training before the convergence of the weights. The remaining parameters used on our dataset are set to their default values, shown in Fig. 3.7.

```
"training": {
    "batch_size": 128,
    "bucketing_field": null,
    "decay": false,
    "decay_rate": 0.96,
    "decay_steps": 10000,
    "dropout_rate": 0.0,
    "early_stop": 3,
    "epochs": 45,
    "gradient_clipping": null,
    "increase_batch_size_on_plateau": 0,
    "increase_batch_size_on_plateau_max": 512,
    "increase_batch_size_on_plateau_patience": 5,
    "increase_batch_size_on_plateau_rate": 2,
    "learning_rate": 0.001,
    "learning_rate_warmup_epochs": 5,
    "optimizer": {
        "beta1": 0.9,
        "beta2": 0.999,
        "epsilon": 1e-08,
        "type": "adam"
    },
    "reduce_learning_rate_on_plateau": 0,
    "reduce_learning_rate_on_plateau_patience": 5,
    "reduce_learning_rate_on_plateau_rate": 0.5,
    "regularization_lambda": 0,
    "regularizer": "l2",
    "staircase": false,
    "validation_field": "combined",
    "validation_measure": "loss"
}
```

Figure 3.7: Hyper parameter values used in CNN algorithm for both Depression and Anxiety.

3.3.4 K Nearest Neighbor Algorithm

We have found the optimal parameter values for KNN algorithm by parameter tuning using Grid Search algorithm. The values are shown in the Fig. 3.8 . K is the number of neighbor. The range of value of K is between 1 to 133 [19]. Here we used the value as 1. Mixed measure is another important parameter for our dataset. It is the measure used to find the distance between the data points. There are four type of distance measure, Euclidean distance, Hamming distance, Manhattan distance and Minkowski distance [20]. For our dataset the best one is Euclidean distance. The Euclidean distance between two particular data point is the length of line segment that connects them. The third important parameter is kernel type. The best kernel type for our dataset was found to be radial.

```
k-NN.k  = 1
k-NN.mixed_measure      = MixedEuclideanDistance
k-NN.kernel_type        = radial
```

Figure 3.8: Hyper parameter values used in KNN algorithm for Depression dataset.

Fig. 3.9 shows the confusion matrix of KNN model on the depression dataset. From

the confusion matrix we can find that the model predicted 316 minimum, 120 mild, 76 severe and 38 moderate level of severity of depression in the tested data, whereas the true result are 370 minimum, 91 mild, 59 severe and 30 moderate level. Hence as shown in Fig. 3.10, we found three performance metrics: accuracy, precision and recall. We obtained an accuracy of 81.82%, recall of 75.73% and precision of 68.63%.


```
ConfusionMatrix:
True:      Minimum Mild      Severe    Moderate
Minimum:           314     2         0         0
Mild:      56      63      0         1
Severe: 0          8       56        12
Moderate:          0       18        3         17
```

Figure 3.9: Confusion matrix of KNN for Depression dataset.


```
Performance:
PerformanceVector [
*****accuracy: 81.82% +/- 3.93% (micro average: 81.82%)
ConfusionMatrix:
True:    Minimum Mild     Severe   Moderate
Minimum:          314     2        0        0
Mild:    56       63      0        1
Severe: 0         8       56       12
Moderate:         0       18       3        17
-----weighted_mean_recall: 75.73% +/- 11.91% (micro average: 76.42%), weights: 1, 1, 1, 1
ConfusionMatrix:
True:    Minimum Mild     Severe   Moderate
Minimum:          314     2        0        0
Mild:    56       63      0        1
Severe: 0         8       56       12
Moderate:         0       18       3        17
-----weighted_mean_precision: 68.74% +/- 11.51% (micro average: 67.57%), weights: 1, 1, 1, 1
ConfusionMatrix:
True:    Minimum Mild     Severe   Moderate
Minimum:          314     2        0        0
Mild:    56       63      0        1
Severe: 0         8       56       12
Moderate:         0       18       3        17
]
```

Figure 3.10: Hyper parameter values used in KNN algorithm for Depression dataset.

The optimal parameter values were found using parameter optimization as shown in the Fig. 3.11. Unlike the depression dataset the best value for the anxiety dataset was found to be 21. The the mixed measure was found to be same for both the dataset. Another important parameter for anxiety dataset was nominal measure. The best value found for this parameter is dice similarity.

Fig. 3.12 shows confusion matrix of KNN for anxiety dataset. From the confusion matrix we can find that our model predicted 14930 mild, 13388 severe, 3229 moderate and 2169 profound level of severity of anxiety whereas the true predictions were 14740 mild, 8594 severe, 9307 moderate and 2359 profound level. Moreover in Fig. 3.13, we can see that, for anxiety dataset the KNN algorithm can predict the severity level with accuracy of 70.96%, recall of 63.14% and precision of 67.04%.

```
k-NN.k   = 21
k-NN.mixed_measure       = MixedEuclideanDistance
k-NN.nominal_measure     = DiceSimilarity
```

Figure 3.11: Hyper parameter values used in KNN algorithm for Anxiety dataset.

```
ConfusionMatrix:
True:     Mild      Severe    Moderate        Profound
Mild:     13879     0         1051      0
Severe:   96        7473      5819      1284
Moderate:           765       54        2410      0
Profound:           0         1067      27        1075
```

Figure 3.12: Confusion matrix of KNN for Anxiety dataset.

```
Performance:
PerformanceVector [
*****accuracy: 70.96% +/- 0.90% (micro average: 70.96%)
ConfusionMatrix:
True:    Mild    Severe  Moderate        Profound
Mild:    13879   0       1051    0
Severe: 96       7473    5819    1284
Moderate:        765     54      2410    0
Profound:        0       1067    27      1075
-----weighted_mean_recall: 63.14% +/- 1.00% (micro average: 63.14%), weights: 1, 1, 1, 1
ConfusionMatrix:
True:    Mild    Severe  Moderate        Profound
Mild:    13879   0       1051    0
Severe: 96       7473    5819    1284
Moderate:        765     54      2410    0
Profound:        0       1067    27      1075
-----weighted_mean_precision: 67.04% +/- 1.36% (micro average: 67.02%), weights: 1, 1, 1, 1
ConfusionMatrix:
True:    Mild    Severe  Moderate        Profound
Mild:    13879   0       1051    0
Severe: 96       7473    5819    1284
Moderate:        765     54      2410    0
Profound:        0       1067    27      1075
]
```

Figure 3.13: Confusion matrix of KNN for Anxiety dataset.

3.3.5 Support Vector Machine

SVM gave 95% accuracy for anxiety and 95.8% for depression. Amongst the parameters in SVM, gamma and c are the effective parameters of the Radial Basis Function (RBF) kernel SVM. The gamma parameter shows how far the influence of the training example reaches. High values are usually signified as 'close' and low values are signified as 'far'. Whereas, c behaves as the regularization parameter in SVM. It controls the tradeoff by correcting the classification of training examples against the maximisation of the decision function's margin. For depression, the gamma parameters was set to 'poly' with the polynomial kernel with degree 4 as shown in Fig. 3.15 .In the case of anxiety the parameters shown in Fig. 3.14, were set to their default value excluding the cache size parameter. Cache size specifies the size of the kernel cache (in MB). We have considered 200MB for both anxiety and depression in the parameter tuning process in our dataset. Thus , these hyper-parameters have led to better accuracy for both depression and anxiety datasets.

As we can see from the Fig. 3.14, the precision is 96.2% and recall is also 96.2% , in the case of anxiety. In the case of depression, the precision is 97.02% and recall is also 95.83% , as shown in the Fig. 3.15.

```
        Evaluating SVM
        Parameters:
            Penalty parameter C of the error term: 1.0
            Kernel: poly
            Degree: 3
            Gamma: auto_deprecated
            Independent term in kernel function: 0.0
            Shrinking: True
            Probability: False
            Tolerance for stopping criterion: 0.001
            Cache Size: 200
            Max Iteration: -1
            Decision Function Shape: ovr
            Random State: None
        ==================================================================================================
        SVM: Accuracy -> 0.962446 | Standard Deviation -> 0.007008
        SVM: Precision Weighted -> 0.962739 | Standard Deviation -> 0.006912
        SVM: Recall Weighted -> 0.962446 | Standard Deviation -> 0.007008
        SVM: F1 Weighted -> 0.962434 | Standard Deviation -> 0.006982
[b]0.7 ==================================================================================================
```

Figure 3.14: Hyper parameter values used in SVM algorithm for Anxiety dataset.

```
Evaluating SVM                                                                        .
Parameters:
    Penalty parameter C of the error term: 1.0
    Kernel: poly (default was 'rbf')
    Degree: 4 (default was 3)
    Gamma: auto_deprecated
    Independent term in kernel function: 0.0
    Shrinking: True
    Probability: False
    Tolerance for stopping criterion: 0.001
    Cache Size: 200
    Max Iteration: −1
    Decision Function Shape: ovr
    Random State: None
===================================================================================
SVM: Accuracy −> 0.958333 | Standard Deviation −> 0.045644
SVM: Precision Weighted −> 0.970278 | Standard Deviation −> 0.032680
SVM: Recall Weighted −> 0.958333 | Standard Deviation −> 0.045644
SVM: F1 Weighted −> 0.958505 | Standard Deviation −> 0.044854
===================================================================================
```

Figure 3.15: Hyper parameter values used in SVM algorithm for Depression dataset.

3.4   Result Comparison

From the above results, we can clearly see that linear regression do not give a good prediction. As we can see in Fig. 3.1 and in Fig. 3.2 the rmse for depression and anxiety is not very high nor very low. We can say that the data points are not very close to the the predicted value of the model, which means it gives an average prediction. This is because linear regression is best for linear models, where the data follows a specific trend. However our data do not follow any specific trend hence the model cannot generate a specific relation between the data and thus do not give a proper prediction.

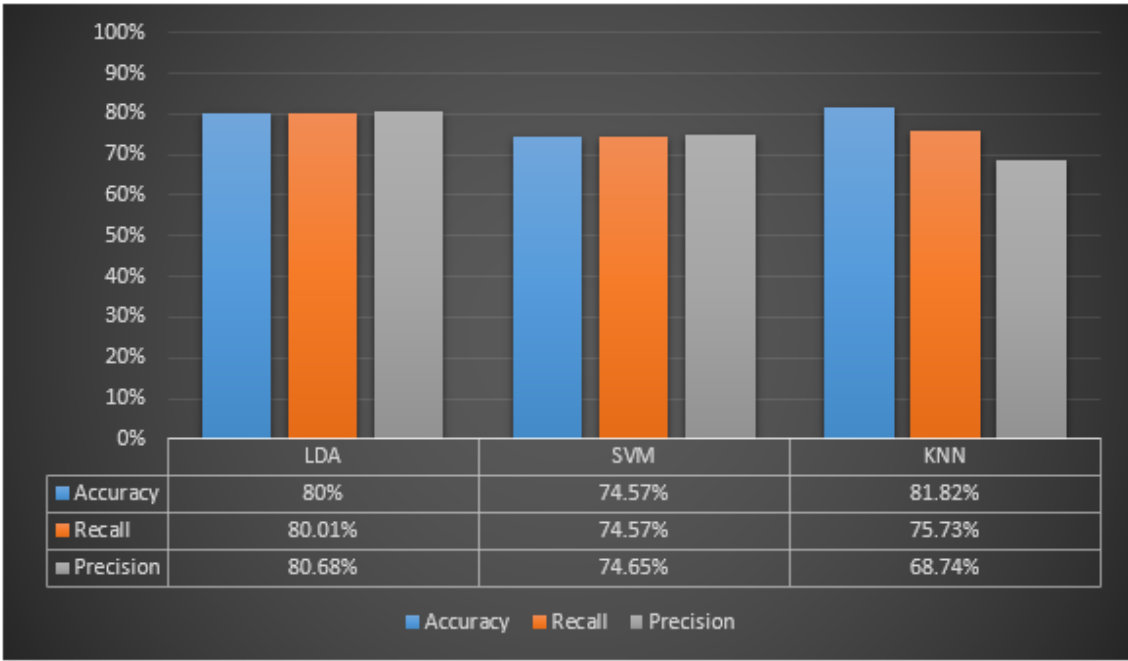|	LDA	SVM	KNN
Accuracy	80%	74.57%	81.82%
Recall	80.01%	74.57%	75.73%
Precision	80.68%	74.65%	68.74%

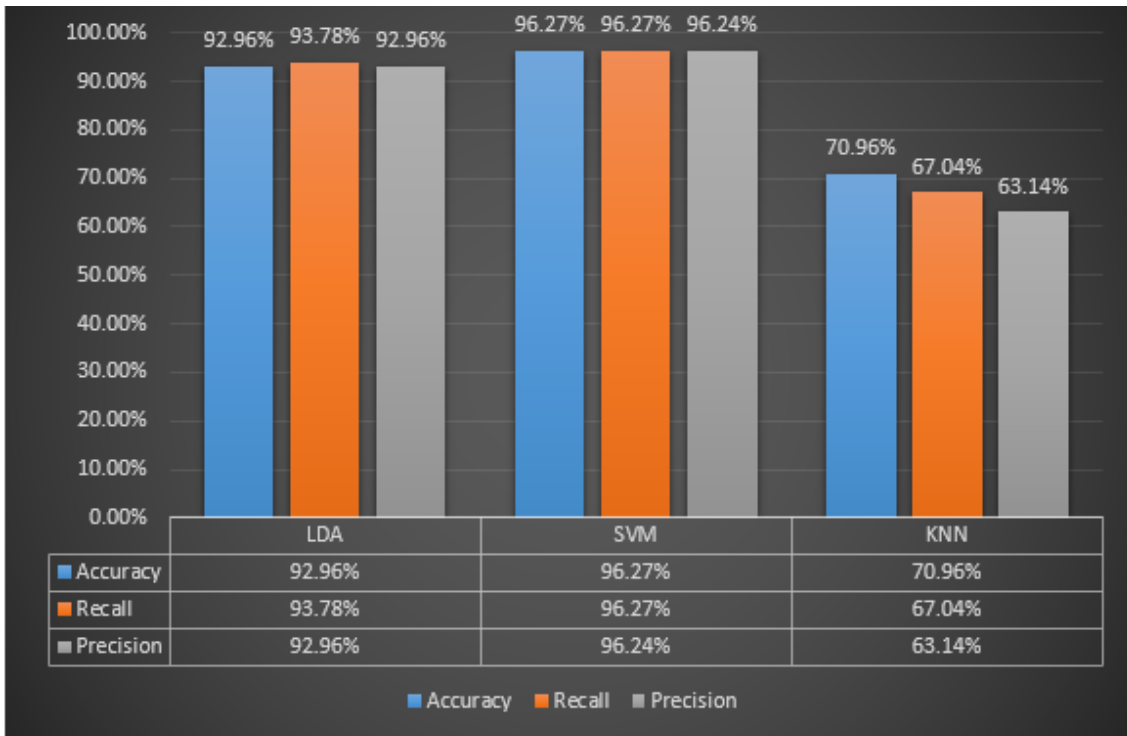Figure 3.16: Accuracy, Precision and Recall comparison of Depression.

Figure 3.17: Accuracy, Precision and Recall comparison of Anxiety.

From the above result we can say that the accuracy of all the algorithms are very close. However we got the highest accuracy for CNN, which is 96.8% for depression and 96% for anxiety, On the other hand we got the lowest accuracy for KNN, which is 81.82% for both depression and anxiety. Fig. 3.16 and Fig. 3.17 shows accuracy, precision and recall comparison of the three algorithms: LDA, KNN and SVM for depression and anxiety respectively, amongst which the highest precision and recall was found in SVM and lowest was found in KNN. From this comparison we can say that KNN is not appropriate for our model compared to the other three models. This is because KNN works well for models that have less features while classifying. The main reason behind this is that in high dimensional space the distance between all neighbors becomes very similar. Thus, the notion for nearest and far neighbor becomes blurred. Since we have used a huge number of features, KNN could not give a good result.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

Mental or psychological disorder is a common yet less treated area in health sector. Anxiety and depression are more or less common phenomena for anyone at some point of their life. Young generation is facing these psychological disorders yet they are trying to cope with it rather than seeking help as psychological treatment which is still a taboo in this era of globalization. The person who is going through these kind of disorders are continuously in risk for hypertension, cardiac arrest, brain anomaly etc. According to WHO (World Health Organization) people are being more prone to psychological disorders in recent years. Most of the patients of mental illness don't come out and seek help , also the treatment is quite a bit costly for general mass. To eradicate these phenomena we have proposed a system which can predict the severity of depression and anxiety using supervised machine learning algorithms. The system was trained with two different datasets using five different types of algorithms. The results of these algorithms were compared and thus, CNN proved to be the best method for detecting both depression with accuracy of 96% and anxiety with accuracy of 96.8% . Our model can help in assisting the psychologists and counsellors for assisting the follow ups of the patients, making their treatment more efficient and easy.

4.2 Limitations and Future Work

Though our model predicts depression and anxiety with good accuracy, there were some limitations too. Since we have a a long list of questions in the questionnaire section, patient might feel reluctant to answer all the questions. Hence they might not concentrate properly while answering the questions and might fill up the wrong answers, which might result in false result. Moreover patients who usually come to counsellors might not be mentally stable due to the problems they are going through, hence while answering the questions, they might panic and feel their condition is severe even though it's not the case. According to psychologists, patience not always reveal their true mental state and try to hide many things due to which they might not answer all the questions honestly. Due to these reasons, result pattern will not be identified properly resulting in wrong results.
Our future plan for the system is to make it execute precisely and accurately and be able to diagnose psychological illness at an early stage with higher success rate. We

are planing to make our system more accurate by filtering out the irrelevant features and taking the most relevant features using suitable algorithms. This would help in narrowing down the results and give better accuracy. Moreover in future we are looking to launch this system as an assistant for the psychologists to monitor post therapy improvements, having the potential to be an open platform for analyzing the patients. It will help in easing the post psych-therapy monitoring process. As soon as the accuracy of this system is reliable towards the psychologists, we will start working on the system update. Furthermore, the system can be made more efficient and accurate through the responses of the future users of the system using unique suitable algorithms developed to meet cross referencing between disorders to compute the psychological stability of a person. A series of questionnaires made by specialist psychologists will be prepared and would be used to get more data that would help in relating depression and anxiety with stronger reasons. According to the present accomplishment of our work, the future of this system looks very much achievable.

Bibliography

[1] R. Yap and D. M. Clarke, "An expert system for psychiatric diagnosis using the dsm-iii-r, dsm-iv and icd-10 classifications.", in *Proceedings of the AMIA Annual Fall Symposium*, American Medical Informatics Association, 1996, p. 229.

[2] *World health report*, https://www.who.int/whr/2001/media_centre/press_release/en/, 2001.

[3] W. H. Organization *et al.*, "Investing in mental health", 2003.

[4] F. Deeba and R. Begum, "Development of an anxiety scale for bangladesh population", *Bangladesh Psychological Studies*, vol. 14, pp. 39–54, 2004.

[5] P. M. C. Guerreiro, "Linear discriminant analysis algorithms", *Unpublished master's thesis). Technical University of Lisbon, Portugal*, 2008.

[6] A. Schneider, G. Hommel, and M. Blettner, "Linear regression analysis: Part 14 of a series on evaluation of scientific publications", *Deutsches Ärzteblatt International*, vol. 107, no. 44, p. 776, 2010.

[7] R. Y. Masri and H. M. Jani, "Employing artificial intelligence techniques in mental health diagnostic expert system", in *2012 International Conference on Computer & Information Science (ICCIS)*, IEEE, vol. 1, 2012, pp. 495–499.

[8] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2012, vol. 821.

[9] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 329.

[10] *American psychiatric association*, https://www.psychiatry.org/, 2013.

[11] S. B. Imandoust and M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background", *International Journal of Engineering Research and Applications*, vol. 3, no. 5, pp. 605–610, 2013.

[12] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency, "Automatic behavior descriptors for psychological disorder analysis", in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, 2013, pp. 1–8.

[13] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support", in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[14]   M. Dobbali, *Metrics to understand regression models in plain english: Part 1*, https://towardsdatascience.com/metrics-to-understand-regression-models-in-plain-english-part-1-c902b2f4156f, 2014.

[15]   M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations", *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, p. 1, 2015.

[16]   H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4293–4302.

[17]   A. Rizzo, R. Shilling, E. Forbell, S. Scherer, J. Gratch, and L.-P. Morency, "Autonomous virtual human agents for healthcare information support and clinical interviewing", 2016.

[18]   S. Anjume, K. Amandeep, A. Aijaz, and F. Kulsum, "Performance analysis of machine learning techniques to predict mental health disorders in children", *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, no. 5, 2017.

[19]   A. Bronshtein, *A quick introduction to k-nearest neighbors algorithm*, https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7, 2017.

[20]   V. Prasath, H. A. A. Alfeilat, O. Lasassmeh, and A. Hassanat, "Distance and similarity measures effect on the performance of k-nearest neighbor classifier-a review", *arXiv preprint arXiv:1708.04321*, 2017.

[21]   G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution.", in *IJCAI*, 2017, pp. 3838–3844.

[22]   A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial", *AI communications*, vol. 30, no. 2, pp. 169–190, 2017.

[23]   R. Gandhi, *Support vector machine—introduction to machine learning algorithms*, https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47, 2018.

[24]   R. Khandelwal, *K fold and other cross-validation techniques*, https://medium.com/datadriveninvestor/k-fold-and-other-cross-validation-techniques-6c03a2563f1e, 2018.

[25]   Prabhu, *Understanding of convolutional neural network (cnn)—deep learning*, https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148, May 2018.

[26]   R. Pupale, *Support vector machines(svm)—an overview*, https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989, 2018.

[27]   S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning", *arXiv preprint arXiv:1811.12808*, 2018.

[28]   *Assumptions of linear regression*, https://www.statisticssolutions.com/assumptions-of-linear-regression/, 2019.

[29]   *Tutorial: Understanding regression error metrics in python*, https://www.dataquest.io/blog/understanding-regression-error-metrics/, 2019.

[30]   M. Sumathi and B. Poorna, "Prediction of mental health problems among children using fuzzy clustered bayesian model",

[31]   J. Weerasinghe, K. Morales, and R. Greenstadt, "Analyzing machine learning models that predict mental illnesses from social media text", *studies*, vol. 3, p. 5,