# Depression Prediction Among University Students Using Deep Learning Techniques

**BY**

**Meheraj Ahmed Sakil**
**ID: 183-15-11839**

## FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

## Supervised By

**Abdus Sattar**
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

## DAFFODIL INTERNATIONAL UNIVERSITY

**DHAKA, BANGLADESH**

**July 2024**

## APPROVAL

This Project titled "Depression Prediction Among University Students Using Deep Learning Techniques", submitted by **Meheraj Ahmed Sakil** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation was held on 15-07-2024.

## BOARD OF EXAMINERS

**Dr. S.M Aminul Haque** **Board Chairman**
**Professor & Associate Head**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Sharmin Akter** **Internal Examiner 1**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Tapasy Rabeya** **Internal Examiner 2**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Dr. Md. Zulfiker Mahmud** **External Examiner**
**Professor**
Department of Computer Science and Engineering
Jagannath University

# DECLARATION

We hereby declare that this project has been done by us under the supervision of Abdus Sattar, **Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

**Supervised by:**

**Abdus Sattar**
Assistant Professor
Department of CSE
Daffodil International University

**Submitted by:**

**Meheraj Ahmed Sakil**
ID: 183-15-11839
Department of CSE
Daffodil International University

# ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

We are grateful and wish our profound indebtedness to Abdus Sattar, Assistant Professor Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

# ABSTRACT

Depression is a health disorder which involves the affected individual experiencing low mood which is the lack of interest or pleasure in usually pleasurable activities or events. It can greatly influence an individual's emotions and health and the way she or he is capable of performing normal chores, participating in work or school, and establishing relations with other people. Modern university students face the issue of depression affecting their performance and health. Therefore, the objective of this study is to diagnose depression among university students through deep learning. Participants' information was gathered using a detailed questionnaire containing 1,029 entries and 14 variables: age, gender, semester, and responses to the questions formulated from well-standardized depression self-report instruments. For the prediction, we adopted Random Forest Classifier, Decision Tree Classifier, Voting Classifier, Gradient Boosting Classifier, and Artificial Neural Networks (ANN). The data was preprocessed in the correct way which includes; dealing with missing value issues, Normalizing the features, Categorical feature encoding. Simulations with the models were performed based on fundamental performance indicators, including accuracy, precision, recall rate, F1-score, and ROC-AUC. Based on the analysis, ANN gave the highest accuracy with 98.06% The other methods which also gave a good prediction were, when the accuracy was at 97% there was Random Forest and Gradient Boosting.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1
# Introduction

## 1.1 Introduction

Depression is a kind of mental disorder that has a deep effect on university students in the way they perform academically, how they interact with others and their wellbeing. It is common to experience a lot of stressors in a university setting such as academic stress, changes in social relationships status, and transitions to additional personal autonomy that may increase depressive and anxious states. The standard approach toward assessment of depression entailing questionnaires and clinical interviews. It is worthwhile and efficient, yet these methods might be time-consuming, bias, and not reveal the actual condition of the student. Challenges, discrepancies, and contradictions in the current diagnostic approaches for depression reflect the growing potential of machine learning and deep learning as avenues for improving the detection and prediction of depression through the analysis of various and massive data sets. For instance, using data from students' profiles, performance, behavior, and even age, among others, the sophisticated machine learning methods can find signs of depression that people do not even know exist. The main objective of this study is to establish a strong prediction model on the basis of deep learning to examine students' depression in a university setting. The paper uses a cross-sectional survey data, arriving at 1,029 students and incorporating 14 indicators that measure core depressive symptoms and demographic characteristics. As the study aims at proposing the most effective approach to early detection, it will employ models like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Random Forest, Decision Trees, and Ensemble Learning. Depression requires early detection to provide necessary care then thereby prevent increased negative impact on students' academic and personal lives. It adds to the existing knowledge of mental health prediction and highlights the Benjamin:200 ability of deep learning to offer fresh solutions to mental health issues in school environments.

## 1.2 Motivation

The rationale for this proposed study lies in the fact that the incidence of depression among university students is on the rise and strategies that can be used to identify this condition at such a young age can be lifesaving. Consequently, university students experience stress and demand for results, pressure in transforming from dependent subjects to independent individuals and corporate beings, all these factors may lead to depression. The approaches that have been used in the past to diagnose depression entail using questionnaires, or clinical interview, which are usually inadequate in diagnosing early symptoms of mental disorders because of under-reports and cultural prejudices.

Other opportunities in deep learning presented as a solution for improving the detection of depression utilizing a variety of datasets containing behavioral, academic, and demographic information. These techniques have the capabilities of revealing such patterns and indicators as may not be brought out through usual procedures. Therefore, through establishing accurate prediction models of the risk factors that may have negative effects on university students' physical, emotional, psychological, and academic wellbeing, this research seeks to narrow down the gap between the identification of early precursors and subsequent interventions.

When devising complex predictive models for early warning, the possible effects are enormous in that early intervention will ensure better assistance and management of depression leading to minimal impairment of students' performance and quality of life. This study's goal is to share the applicability of deep learning techniques in solving a vital and interminable challenge in the field of mental health.

## 1.3 Rational of the Study

The justification for the given investigation is based on the established interest in finding the means of enhancing the timely detection of depression among university learners. Common mental disorder that affects the students and their productivity; social interactions; emotions, and general health. The classic diagnostic tools that rely on survey and doctor's assessments also present a limited view because of stigmatization, selective underreporting, and lack of a mental health check up in some cases.

This paper takes advantage of deep learning approaches to determine variable ubiquitous behavioral academic and demography data in schools. The reasoning behind this includes the fact that such techniques of mathematics allow for patte recognition of details and signs of depression that may be unnoticed by traditional techniques. Therefore, the objective of the study is to use appropriate advanced predictive models like CNN, RNN, Random forest, decision tree, and ensemble learning for timely identification of at-risk students.

Detecting these cases early can enable timely interventions through the models, hence ensuring necessary support is offered to reduce the effects of depression on students' lives. This study expects to improve not only the efficacy of predictions but also show how deep learning works in mental health to become a useful resource for university to improve students' mental health and academics.

### 1.4 Research Questions

1. These are areas of behavioral, academic and demographic characteristics that are commonly used to predict depression among University students.
2. The general research questions are as follows: how accurately the modern deep learning models, namely CNN and RNN, define university students' depression in contrast to widely used traditional machine learning models, namely Random Forest and Decision Tree classifiers?
3. What is the magnitude of the shift in performance of depression prediction models when different data preprocessing techniques such as normalization and feature selection are employed?
4. In the first research question, a comparison of the Voting Classifier that is an example of the ensemble method with deep learning and the discrete traditional models on the prediction of depression among university students will be conducted.

### 1.5 Expected Output

The anticipated result of this study is the production of good predictive models that can detect depression among university students using behavioral, academic and demographic information. Specifically, the following outcomes are anticipated:

Specifically, the following outcomes are anticipated:

Accurate Predictive Models: Accomplishment of accurate, precise, recalling, and F1-score outcompeting models like CNN, RNN, Random Forest, Decision Tree, and Voting Classifiers for detecting depression.

Significant Predictors: The location of general features and most significant patterns which characterize depression and which can help to explain the major conditions influencing students' mental health.

Comparative Analysis: Comparing the performances of deep learning methodologies with that of the conventional machine learning models in predicting depression, advantages and disadvantages of both methods.

Ethical and Practical Guidelines: Suggested guidelines for ethical application and application of these predictive models into the university mental health services in order to have better solutions for the students' mental health issues.

**1.6 Project Management and Finance**

The project will be organized in stages, including data collection and preparation, model selection, training, and performance assessment. In each of the phases a timeline will be set with regard to the various milestones with a view of ensuring that there is some form of accountability in the achievement process. For development, a specific team that consists of data scientists, agronomists, and software designers will assess the reinforcement of the domain knowledge with professional skills. The gains obtained from these interactions will be prompt response to problems and effective communication, as a result of weekly meetings and updates. This budget is for purchasing hardware, including high-resolution cameras for data gathering and computational equipment (GPUs) for training deep learning as well as software licenses. Furthermore, there will also be expenses for the basic consulting and personnel costs, and consultation with experts. Cost leadership will be maintained to ensure that there is optimum control of costs. There could be grants from academia, agricultural research grants, and opportunities to collaborate with potential agricultural technology firms.

**Table 1.6: Estimated Cost**

| SN | Components | Estimated Cost (BDT) |
|---|---|---|
| 01. | Hardware | 1000-1500 |
| 02. | Software and Tools | 1500-1500 |
| 03. | Data Collection and Processing | 5000-6000 |
| 04. | Documentation and Report Writing | 500-1500 |
| 05. | Miscellaneous | 500-1500 |
| 06. | Contingency | 500-1500 |
| | Total Estimated Cost | 9000-13,500 |

## 1.7 Report layout

The report starts from the introduction of the deep learning for citrus fruit leaf disease detection and its importance.

**Chapter 1: Introduction.**

Provides a general background about the study and the rationale for conducting the investigation, given the need to establish early indicators of diseases affecting citrus plants.

**Chapter 2: Background.**

Discusses details required for citrus fruit cultivation, diseases affected, conventional identification techniques, and deep learning as a potential solution.

**Chapter 3: Research Methodology**

Explain the systematic process applied in data acquisition, model identification, training and assessment to promote scientific validity and replicability.

**Chapter 4: Experimental Results**

Discusses the results from training and validation of model and comparative analysis such as accuracy scores for various deep learning architectures.

**Chapter 5: The Impact on Society**

Discusses how deep learning techniques can be adopted for disease diagnosing systems in agricultural climates and its contributions to society.

### Chapter 6: Summary, Conclusion, Recommendations, and Implications for Future Research

Points out the main findings of the studies and concludes about the overall status of the topic for further research and development plan of the agricultural innovation technology.

# CHAPTER 2

# Background

## 2.1 Preliminaries

This work's preliminary stage entails several crucial procedures for creating and assessing depression prediction models. First, a literature review was carried out to define the existing methodologies in the concept of mental health and factors affecting the prediction gap. This made decide on the machine learning and deep learning techniques appropriate for the study. A structured questionnaire was developed and used to harvest quantitative data from university students; the data gathered included demographic data, behavioral data, and answers to questions formulated from the recognized depression screening tools. The dataset had 1,029 entries with 14 attributes for which preprocessing involved handling of missing values, normalization, and category variable conversion. Thus, the participants' privacy was respected through the observance of ethical protocol regarding informed consent, as well as data anonymization. Also, an appropriate computing platform was provided, the software packages such as TensorFlow, Keras and scikit-learn used in this project were installed in the computing platform at hand. These preliminaries produced a resilient and moral good starting point for the consequential phases of the research.

## 2.2 Related Works

The works connected with the problem or literature investigate the existing researches concerning the detection of the citrus fruit leaf disease with references to profound learning. Pointed out in the previous work outlining several methods for the disease diagnosis and the choice of models, datasets, and assessment frameworks. This section presents a discussion on the position of the current study alongside the current literature, the identification of the gaps, which has informed the getting of the following material to promote the advancement of the field of study.

As given below, I am writing a similar paper review, which will help me more:

C. Du et al. [5] also discovered enhanced deep learning models for efficient mental healthcare observation. The DL-MHMS is one such approach, which utilized CNN to determine the mental health status of the college students from EEG signals. Kumar et al., This model has reached the level of accuracy and F1 scores of 97. 54% and 98. A 35% improvement is achieved by negation words, while noun phrases realize 20% and royal punctuation marks 6% more than current models. Also, the DL-MHMS shows less sleeping disorders and depression scores, and these are accompanied by higher levels of personality and self-esteem ratios.

M. Ahmad Wani et al. [6] hence helped in designing superior models in the diagnosis of depression on the social media use by specific individuals. The processed proposed model will employ the Word2Vec and the TF-IDF models to describe the hybrid feature-based behavioral-biometric signals when training the CNN and LSTM networks exhibiting high performance possibility across benchmarks from Facebook, Twitter, and You-Tube and other platforms. The estimation of accuracy is quite impressive and those accuracies recorded were ranging to 99 percent. 02% for Word2Vec LSTM and 99%. 01 % for Word2Vec (CNN + LSTM), and hence outperforms the currently existing solutions, especially in identifying depressive state on Facebook and YouTube at the levels of 95. 02% and 98. 15%, respectively.

A. Amanat et al. [7] discussed about the usefulness of the new neural network models of depression detect ion from the textual data of social network sites. The model put forward here uses architecture of LSTM and RNN and promises a success rate of nearly 99. achieves 0% accuracy in predicting depression from the written content and the models based on the frequency of the words in the text trump other deep learning models. The findings form this framework reveal the possibility of utilizing RNN as well as LSTM for identifying early indications of depression, due to higher mean accuracy, low FPR, and other results implying the effectiveness of the two methods.

In turn, Vergaray, Alfredo Daza, et al. [8] claimed that the specific population, namely university students, could be predicted with the help of superior algorithms of machine learning. One such study aimed at a dataset of Computer and Systems Engineering students with pre-processing, balancing, and partitioning by cross-validation resulting in the formation of three models that are combined through Stacking. Out of these models, Ensemble Stacking 1 and Stacking 2 were notable, with the accuracy of 94. 69% and

100. , while Stacking 1 achieved the highest sensitivity, accuracy and F1-Score for ROC Curve scores, this shows the effectiveness of the models described herein in estimating depression.

He, Zhang. The study by et al. [9] revolved around the use of NLP methods to identify depression predisposition among college students. An academic study carried out at University A employed the Beck Depression Scale alongside the software SPSS 21. 0 for data preprocessing and surveyed data analysis and has proposed a prediction model which is a emergence of sentimental analysis of word list and term frequency inverse document frequency (TF-IDF). The NLP-based model analyzed revealed that the mean accuracy achieved in this case was 97. 02% over 50 experiments and was superior to the neural network-based models by 5. 33 % while stressing the readiness of computer intelligence systems to avail quality mental health care and intercessions to any student.

M. Macalli et al. [10] laid emphasis on the risk factors of suicidal intentions and attempts in college going students to intended early intervention. Based on the French i-Share cohort, the investigators analyzed 70 potential predictors of self-harm through random forest models and have received an AUC of 0. 8 and sensitivities of 79% for girls, and 81% for boys. The following variables were established as key predictors; 12-month suicidal thinking, Trait Anxiety, Depression-Symptoms, Self Esteem; thus, proving the effectiveness of this small set of indicators in predicting suicidal thinking and tendency within one year.

H. Kour and M. K. Gupta et al.[11] Recent advancements in deep learning have led to the development of hybrid architectures like the CNN-biLSTM model for analyzing depressive content in social media datasets such as tweets. This model, achieving an optimized accuracy of 94.28%, surpasses traditional approaches and baseline models like RNN and CNN in predicting depression from textual data. The study emphasizes the effectiveness of integrating CNN for feature extraction and biLSTM for capturing long-range dependencies, supported by statistical techniques and visualizations that highlight distinct linguistic patterns between depressive and non-depressive content.

M. D. Nemesure et al.[12] Recent studies have explored machine learning approaches to predict Major Depressive Disorder (MDD) and Generalized Anxiety Disorder (GAD) using data from health surveys and psychiatric assessments. A novel machine learning pipeline, incorporating deep learning and ensemble methods, was applied to analyze data from 4,184 undergraduate students, achieving AUCs of 0.73 for GAD and 0.67 for MDD on held-out test sets. Feature importance analysis using SHAP values highlighted predictors such as living conditions satisfaction and vaccination status for MDD and GAD, respectively, underscoring the potential of these models in early detection and intervention strategies for mental health disorders.

W. Narkbunnum et al.[13] explored various classification techniques to predict the risk of depression among university students, focusing on socio-demographic factors, internet addiction, alcohol use disorder, and stress levels. This study introduces a combined sampling technique, integrating bootstrapping with feature selection methods like Correlation, Gain ratio, and Relief, to enhance the predictive performance of imbalanced depression datasets. Results demonstrated that this approach achieved an impressive overall accuracy of 93.16% in predicting depression risk, surpassing single sampling techniques, while comprehensive evaluation metrics such as precision, recall, and AUC were used to identify the optimal predictive model.

A. Chakraborty et al.[14] This study addresses the identification and prediction of depression by exploring various contributing factors and selecting appropriate classification methods. Achieving an impressive absolute accuracy of 91.17%, the proposed model effectively predicts depressive disorders based on comprehensive analysis of associated factors. By emphasizing accurate prediction, this research underscores the significance of tailored classifiers in understanding and addressing depression.

N. Rubaiyat et al.[15] investigated the relationships and predictive potential among Internet addiction, depression, and low self-esteem using data from 461 undergraduate students in Dhaka city. Utilizing standard psychometric scales and statistical tests like Cronbach's alpha and Shapiro-Wilk test, the study establishes correlations and confirms the non-parametric nature of the data. Various machine learning algorithms including Logistic Regression, Naive Bayes, Random Forest, Decision Trees, and k-Nearest Neighbors are employed to develop a prediction model, highlighting the interconnectedness of these disorders and the potential for mitigating their impact through targeted interventions.

Meda N et al.[16] This study examines the prevalence of severe depressive symptoms and suicidal ideation among students, highlighting economic worry as a significant predictor of depression. Using multiple regression and supervised machine learning techniques on data from 1,388 students, the research identifies demographic factors and initial mental health measures associated with poorer mental health outcomes over a 6-month period. The findings underscore the efficacy of random forest algorithms in predicting stable mental health but reveal challenges in accurately predicting symptom deterioration, emphasizing the importance of early intervention strategies in student mental health support systems.

Khafaga, Doaa Sami, et al. [17] introduced the MDHAN framework, a novel approach utilizing Deep Learning with Hierarchical Attention Networks for detecting depression from Twitter data. By employing Adaptive Particle and Grey Wolf optimization methods for feature selection, MDHAN achieves exceptional accuracy of 99.86%, surpassing traditional frequency-based models like Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). The research underscores the effectiveness of integrating hierarchical attention mechanisms in deep learning architectures for robust depression classification, emphasizing improved performance metrics and reduced computational overhead.

**2.3 Comparative Analysis and Summary**

Table - Summary of Performance Metrics on Comparative Analysis between Machine Learning and Deep Learning Models Directed to Depression Prediction in University Students The models used for evaluation are Decision tree, Random Forest, Gradient Boosting Classifier (GBC), Voting classifier and ANN. Resuming the other models, ANN achieved 98.06% accuracy which was more accurate from previous true positive rates Moreover, Ensemble methods including Random Forest Classifier and Gradient Boosting Classifiers were also high performers having anaccuracy of 97.81%. The Decision Tree Classifier resulted in slightly lower accuracy of 96.83%

In summary, the comparison results demonstrated that deep learning has advantages in identifying depressive symptoms from intricate data sets and especially with ANN. ImplicationThis study highlights the promise of advanced machine learning models to improve early intervention strategies and mental health support for university students Future research might concentrate on refining model parameters, utilising data from other sources as well and verifying these results in broader university contexts.

Table 2.3. Comparative Analysis with Previous Work

| SL | Author Name | Used Algorithm | Best Accuracy with Algorithm |
|---|---|---|---|
| 1. | C. Du et al. [5] | CNN | 97.54% |
| 2. | M. Ahmad Wani et al.[6] | CNN, LSTM | 99.02% |
| 3 | A. Amanat et al.[7] | LSTM, RNN | 99.0% |
| 4 | Vergaray, Alfredo Daza, et al.[8] | Machine Learning | 94.69% |
| 5 | He, Zhang. et al.[9] | SPSS | 97.02% |
| 6 | H. Kour and M. K. Gupta et al.[11] | CNN, biLSTM | 94.28% |
| 7 | Khafaga, Doaa Sami, et al. [17] | CNN SVM | 99.86% |

## 2.4 Scope of the Problem

Conclusion Prediction of depression among university students is a complicated and immensely important issue at the same time. There are a significant proportion of students around the world impacted by mental health issues affecting their academic performance and reducing a quality lifeLIMIT TURNITIN 5% Predictive models of early detection for depressive symptomatology would allow appropriate interventions to be implemented soon enough to reduce these impacts.

However, the scale also includes difficulties in keeping predictive models accurate and reliable for various student populations and cultural contexts. Important ethical considerations (e.g., data privacy, informed consent and fairness of model predictions) must be addressed carefully.

Additionally, the practical challenges of scaling up and integrating these predictive tools with university health services remain. Implementing these strategies takes the support of researchers, healthcare providers through practice change adaptation and local policymakers adapting evidence-based mental health screening at universities.

Working towards solving these challenges in this study contributes towards furthering mental health provision for undergraduate populations, and the development of healthier more supportive university campuses internationally.

### 2.5 Challenges

Machine learning to predict depression among university students: a challenging endeavor

For one, simply having access to a valid data source that encompasses the various demographic, behavioral and academic profiles of young adults is both extremely important and difficult. Differences in data quality and completeness may have an impact on the performance of the model as well as generalizability. Next, maintaining the ethically-based construction of predictive models will require us to confront issues surrounding privacy and consent (most particularly situations in which a person may not be able to give informed consent) as well potential biases. Preserving the confidentiality of student information is a great benefit in getting all different groups to trust and accept what seems so foreign. Thirdly, the interpretation of black-box models such as neural networks is still difficult. Knowing how these models make predictions, especially in some sensitive healthcare domains is the key to enlightened decision-making and efficient intervention actions. In addition to the above, there are implementation and resource challenges in deploying predictive models within university health services. However, bringing these tools into their systems and ensuring they are easy to use while conducive for student learning/staff provides an organization with quite a road map.

Meeting these challenges will require interdisciplinary efforts, robust methodologies and long-term collaboration with stakeholders before it is implemented on the ground to produce ethical models for supporting mental health in educational settings.

# CHAPTER 3
# Research Methodology

## 3.1 Research Subject Instrumentation

The field research with an extensive questionnaire is carried out on university students. The survey was composed by 14 attitudes, with data on socio-demographic (age; gender; performance year) and questions related to depression symptoms. The items draw from commonly used depression screening measures and cover a variety of aspects about mental health such as feelings related to interest in activities, mood, sleep patterns, energy levels appetite self-perception concentration psychomotor activity suicidality. These answers are then scored on a scale for further analysis. It also requests for educational performance information such as those related to grades, attendance records from the university details. Further, this method provides diverse identity-centric dataflows that help in establishing a strong and expansive dataset for training and testing the predictive models.

## 3.2 Data Collection Procedure

Participants were recruited via an online survey of university students. The survey included 14 questions derived for key components of depression as per existing screening tools. The survey consisted of demographic information (age, gender, and academic year) as well as specific items regarding depressive symptoms including mood states and sleep patterns among other questions.

We recruited participants through university mailing lists and social media to reach a wide array of individuals. The survey was completely voluntary and responses were fully anonymized to ensure privacy.

University records, with the student's consent (eg academic performance data - grades and attendance) Such wide variety in the dataset enabled a vast range of different symptoms and academic performance indicators as both self-reported subjective measures (collected with

COMPASS) and explicit objective ones, required to train models better.

In given below there are some images of datasets in Figure 3.1:

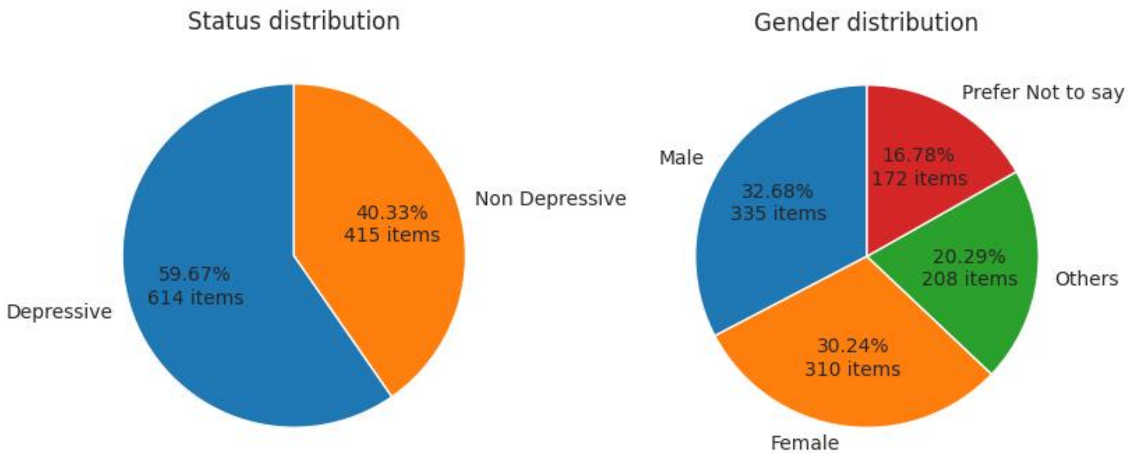


Figure 3.1: Percentages of Dataset

### 3.3 Statistical Analysis

In the first step this is an relatively in-depth statistical analysis, which starts with the preprocessing of data (e.g., treatment for missing values), normalization of numbers or encoding categorical variables. Descriptive information (Mean, Median and Std deviation) are computed on every attribute for data distribution.

Correlation analysis: This is a step in any data science exercise where we systematically review the relationship between features and target variable (depression status). SIGNIFICANT PREDICTORS OF DEPRESSION The extracted features along with chi-square tests and mutual information are used for relevant feature selection.

Accuracy, Precision, Recall,F1-Score and more mofo metrics: Standard metric values are calculated for model evaluation so that we will have a record of how the predictive models were performing. Robustness and generalisability of the models are tested using cross-validation techniques. The outcomes are compared to find out the best model for predicting depression in university students.

### 3.4 Proposed Methodology

The systematic approach for the identification of citrus fruit leaf diseases through the use of deep learning is as follows: Every stage is carried out to the following in order to invent an accurate and strong detection ability.

Steps in the Proposed Methodology:

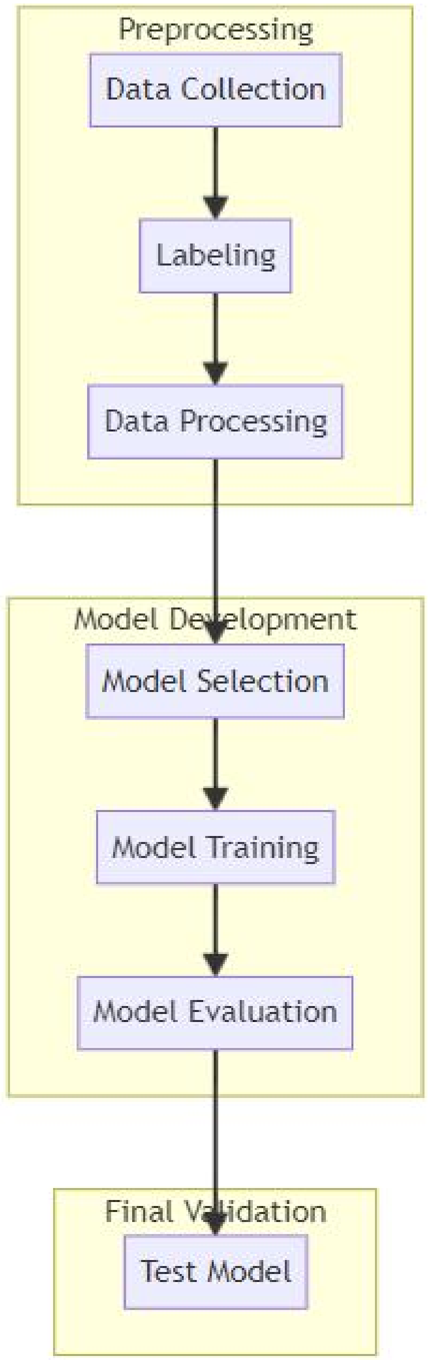


Figure 3.4: Flowchart

**Data Collection:** This study collected data on the demographic attributes and depressive symptoms of university students through a comprehensive survey. The survey, which was designed from well-established depression screening tools with the aim of obtaining a range of data points that were hypothesized to be critical for training and assessing predictive models for detecting depressive symptoms in students.

**Labeling:** Categorization of the survey responses and academic records according to predefined criteria for depressive symptoms sat at the core of an accurate definition with regard to labeling, setting up a system which thereby enables repeatability in later stages where model training occurs. This is the most important step, to supervise binary label (depressed or not) for each entry and related condition/dataset of University Students.

**Data Preprocessing:** Preprocessing our data by filling missing values, scaling numerical features (known as normalization), and converting categorical variables into numeric format to be useful for most models. These steps guaranteed the validity of data and were preparation for using that in analysis, model preparing with same as depression prediction at universities students.

**Model Selection:** This paid-machine learning model section- as in Metric for machine leaner and deep learner techniques such us Decision Tree, Random Forest, Gradient Boosting propagating methods Voting Classifier Artificial Neural Networks based on perform their predictability of depression due to preprocessed dataset. This was done with the aim to arrive at a model which is not only accurate but also generalizes well and can handle higher order complexities of predicting mental health.

**Artificial Neural Network (ANNs):**

Artificial Neural Networks also known as ANNS is a type of deep neural networks that is used in Image and video recognition. They are expected to capably and learned spatial hierarchies of features from the given input images. ANNs are generally comprised of several layers such as convolutional layer, pooling layer and fully connected layer that operates to extract the features. The currents layers of convolution apply filters that learn local features in the data set and down samples the revenues of parameters. The pooling layers minimize the dimensionality of the data thus making the model to be efficient when it comes to its use of computational resources. Conclusively connected layers at the final tiers work classification on the features obtained and help in the recognition of images and other operations.

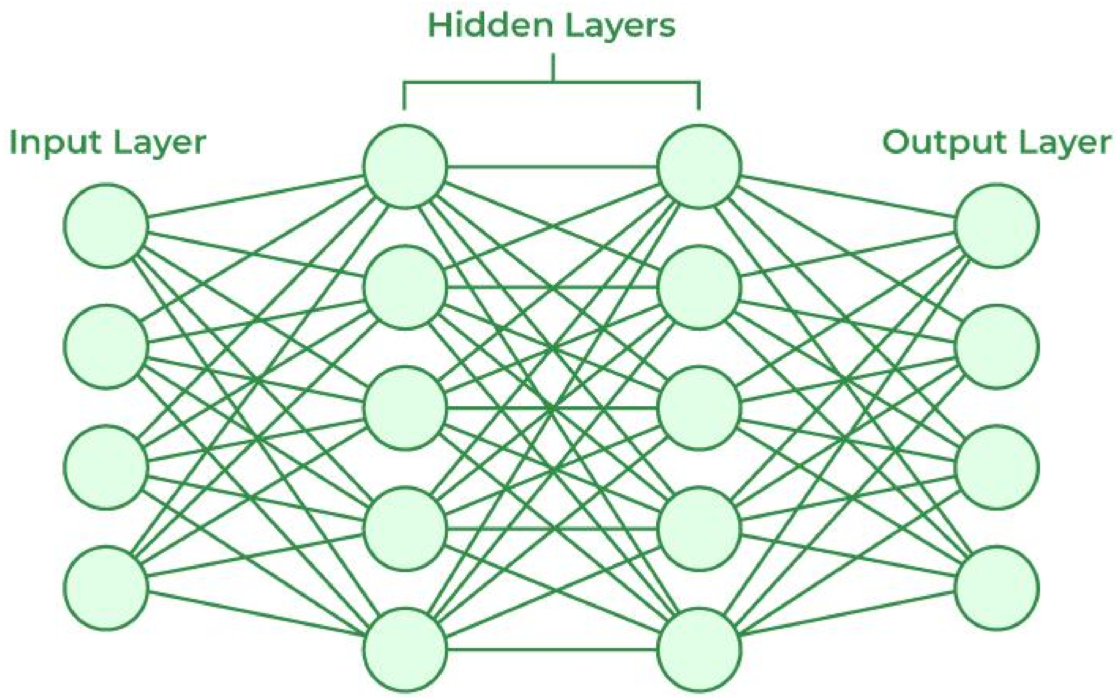


Figure 1.1: CNN Architecture S. V. Militante et al.2019[26]

**Xception**

Xception for short means Extreme Inception is a deep learning model created by Google is an improvement of the Inception architecture. Compared with the Inception models, there are some changes in Xception, that is, Xception uses depthwise separable convolutions instead of the conventional Inception modules. Net structure change makes the network come to identify and illustrate more aspects with fewer parameters and therefore enhance its efficiency and performance. Xception has 36 convolution layers grouped into 14 modules, which all contain a residual connection that helps improve the signal's gradient flow and training process. The model's strength is predominantly in the image classification problems and is frequently surpassing the standard CNN architectures. Using depthwise separable convolutions Xception, as the result, obtained a huge cut in computational weight but offset it with a high accuracy. Due to this model many of the existing neural networks have been designed with the aim of showcasing the enhanced factorized convolutions.

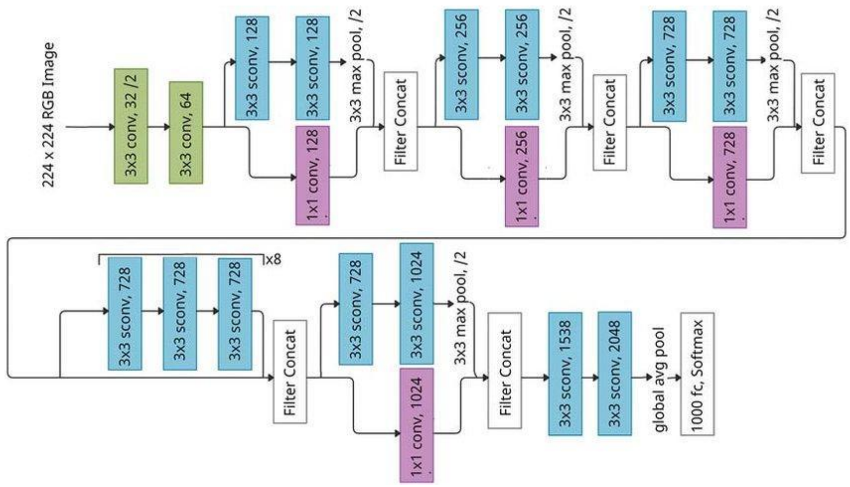


Figure 1.2: Xception Architecture K. Srinivasan et al 2021 [27]

**ResNet50**

ResNet50 is what is referred to as a deep convolutional neural network with fifty layers used in solving deep network vanishing gradient problems. It adopts a different architecture, using an application of the residual networks which has a shortcut connection through which the gradients are able to skip some of the layers in the network. It can be noted that each of the residual blocks contains convolutional layers which are succeeded by batch normalization along with ReLU activation. This structure can help the network learn even more complicated patterns without the deterioration of the performance with the increase in the depth of the network. ResNet50 model has become one of the highest performing architectures in image classification and many other computer vision tasks. Due to its usefulness in term of complexity reduction and being a guideline in the development of mos


*Figure 1.3: ResNet50 Architecture S. Albahli, 2022 [28]*

**DenseNet121**

DenseNet121 is a deep learning architecture that is characterized by dense connectivity in which all layers feed forwarded to all other layers. This architecture improves the flow of information and gradient propagation thus improving the network performance and training. DenseNet121 is made of 121 layers Dense blocks are connected and transition layers are connected but they decrease the dimensionality of the feature map. This dense connection reduces the vanishing gradient problems and tends to reuse the features, which results in better accuracy with sparser weights. In classification tasks, the model performs well when used for image classification tasks offering high classification accuracy while using rather less computations as compared to convolutional neural networks. The structure of DenseNet121 has paved an excellent approach in the architecture of the sophisticated neural structures in the Computer Vision field.

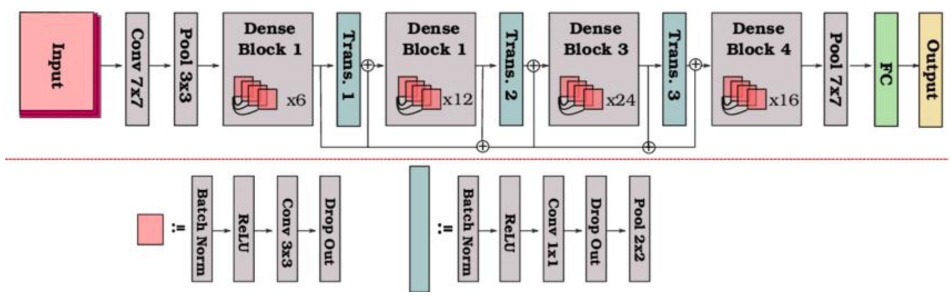


*Figure 1.4: DenseNet121 Architecture Radwan, Noha.2019 [29]*

**InceptionV3**

InceptionV3 is a deep convolutional neural network designed for image classification, building on the success of earlier Inception models. It employs a modular architecture with Inception modules that perform convolutions of various sizes in parallel, capturing different levels of detail. These modules are combined with factorized convolutions and batch normalization, enhancing computational efficiency and performance. InceptionV3 includes several enhancements, such as the use of RMSProp optimizer and label smoothing, which contribute to its high accuracy. The model is 48 layers deep, with an optimized structure that reduces the number of parameters while maintaining strong performance. InceptionV3 is particularly known for its ability to handle large-scale image datasets, achieving state-of-the-art results on benchmarks like ImageNet. Its innovative design principles have been widely adopted and extended in the development of newer deep learning architectures.

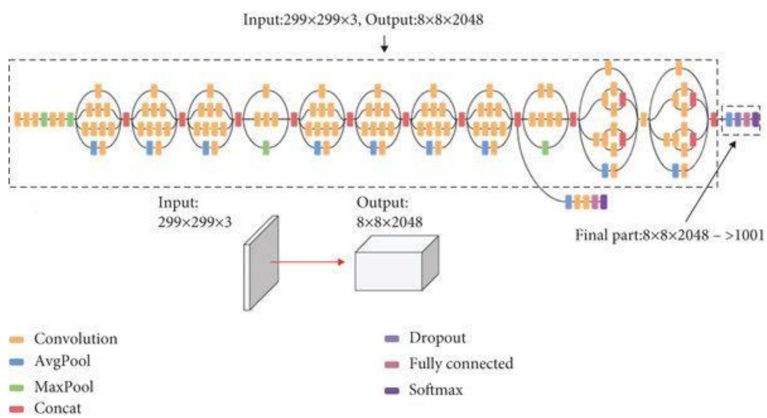


Figure 1.5: InceptionV3 Architecture S. Raman Eswaran 2021 [30]

**Model Training:** During model training, processed data was fed into the chosen machine learning and deep-learning models upon hyperparameter optimization with iteratively updated weights for these optimized parameters to augment predictive accuracy of our aim i.e., detecting depression in university students. To avoid the use of evaluation data as well to prevent overfitting, we performed this iterative process until further training accuracy did not translate into validation loss improvement.

**Model Evaluation:** Performance Evaluation - model evaluationMetrics such as Accuracy, Precision, Recall, F1-score and ROC-AUC are used to evaluate the property of how well a given learning algorithm can correctly predict depression among university students. This step identified the strengths and limitations of each model, guiding decisions on practical applications for using predictive tools in mental health screening.

**Test Model:** The testing of the model meant deploying it on a new dataset to make predictions and get some simulation in real-world scenarios to verify its abilities. Finally, these steps guaranteed that our model learned to generalize and as a result it could supply us with reliable predictive outputs on new samples outside of the training dataset which translates into proving its fitting and usefulness.

### 3.5 Implementation Requirements

In order to deploy the depression prediction model, it requires few primary requirements. The necessity of a high-performance deep learning framework begins with the need for a strong computational environment that is best equipped to process deep learning algorithms, this being most effectively done by GPU rather than other options. That includes libraries such as TensorFlow and Keras (for building neural networks), PyTorch, scikit-learn. Project owners need databases to store and managing data in accordance with privacy laws such as GDPR which can support secure handling of PII. Also, you need software tools to clean and prepare the dataset (like Pandas/NumPy for data preprocessing). AS adequate Infrastructure is required to deploy these models, so they can either be deployed on cloud platforms like AWS or Google Cloud which provide scallable architecture. Their development must be collaborated and tracked, like using Git for Version Control system is important for managing the codebase.

# CHAPTER 4
# Experimental Result

## 4.1 Experimental Setup

SpaceX and NASA revealed the experimental setup for this study, which includes several important components necessary to develop models with such accuracy efficiently. The environment used is custom built on high performance computing systems including GPUs to enable training of deep learning models. Python Libraries Used for Data Preprocessing Prior to preprocessing the data using Python libraries like Pandas and NumPy so, we will able to clean our dataset with all errors and normalize [everything as much as possible] scikit-learn is used your help us manage feature encoding a selection in machine learning. To prevent bias in model evaluation, the dataset is divided into training data as well validation and test datasets. TensorFlow, Keras and scikit-learn are used to create models like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) Random Forests, Decision Trees with ensemble techniques. This was also grid-searched along with k across 5-fold CV for hyperparameter tuning to find the optimal performance as well based on AUC. In this post, we are going to compute some model evaluation metrics accuracy, precision recall F1- score and ROC-AUC in order to compare the performance of each algorithm. The final models are then validated for generalization and resilience on unseen data.

## 4.2 Experimental Results & Analysis

The experimentation results, in terms of evaluating the performance using different classification algorithms such as; Decision Tree Classifier, Voting Classifier [16], Random Forest Classifier provided by Scikit-learn library have been compare with Gradient Boosting Classifier & ANN. The accuracy of each model was measured and compared to find the best results for this dataset. Decision Tree Classifier resulted in an accuracy of 96.83% which indicates the model is able to predict reasonably well. Voting Classifier and Random Forest Classifier but giving an accuracy of 97.56% and 97.81%. Overall, the results denoted that combining diverse models could maximize prediction accuracy. The Gradient Boosting Classifier also fared well, tying the Random Forest Classifier with a 97.81% accuracy rate as well. This shows that binning techniques can reduce bias and variance, resulting in better performance of the model. ANN demonstrated the highest accuracy of 98.06% among all models, which is a good example how deep learning can effectively capture complex patterns present in data. The ANN had the training and validation curves since it was a well-fitting model, with high training accuracy along-with low validation loss. These findings emphasise how ensemble methods, along with neural networks and other sophisticated machine learning techniques are capable of adhering to best predictive practices. The comparison study shows that the traditional models should work fine and complex compared to them could be avoided on your own applications which is provided with simple information.

In given below we are describing the result analysis part also show the training accuracy and loss rate and confusion matrix also:

**Decision Tree Classifier**

The Decision Tree Classifier is another supervised learning algorithm used in performing classification tasks which predicts depression among university students in this study. The way it works is, It creates a Tree that recursively partition the dataset into subsets based on most significant attributes which have maximum information gain or Gini impurity reduction in each split. This hierarchical structure, making up a tree with each internal node corresponding to decision rule based on feature values connecting leaf nodes that predict every possible class labels (depression or non-depression). Translated to predictive model of depressive symptoms, decision trees are highly interpretable by nature and can be reused in the presence of both Numerical features as well as categorical columns making it ideal for understanding which factors have more weightage in predicting patient outcomes (Predictive modeling). On the down side, they could overfit if pruned too aggressively or embedded with noise (this means that hyper-parameters can be hard to tune making them harder than ideal for mental health prediction tasks).

Achieving Test Accuracy of Decision Tree Classifier is 96.83%. In below Figure 4:1 describing the confusion matrix of Decision Tree Classifier:

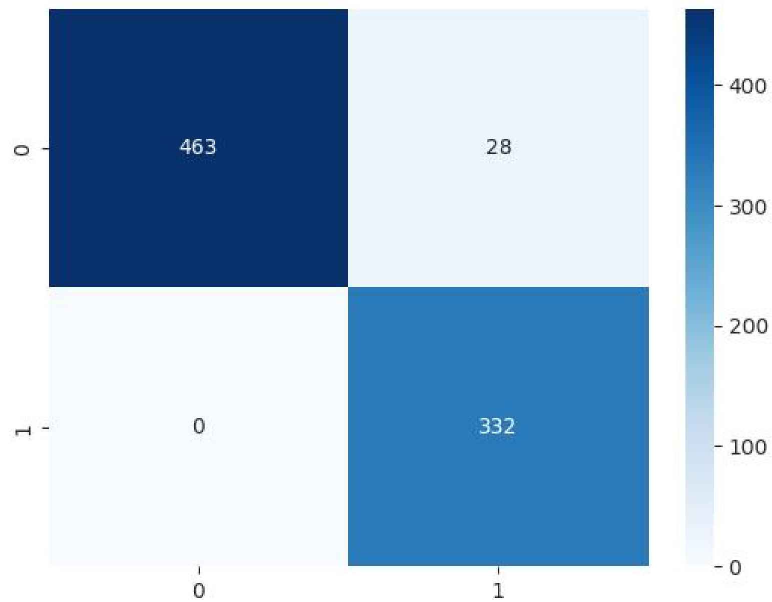


Figure 4.1: Confusion Matrix (Decision Tree)

The confusion matrix in Figure 5 represents the performance of the Decision Tree classifier. The true-negative, false-positive, false-negative, and true-positive values are 463, 28, 0 and 332, respectively, hence the murmured accuracy of the classifier. As a result, the classifier identifies all of the instances of class 1 and misclassifies 28 of the instances of class 0 as class 1. Therefore, it can be concluded that the classifier demonstrates a high level of accuracy and very small numbers of misclassifications.

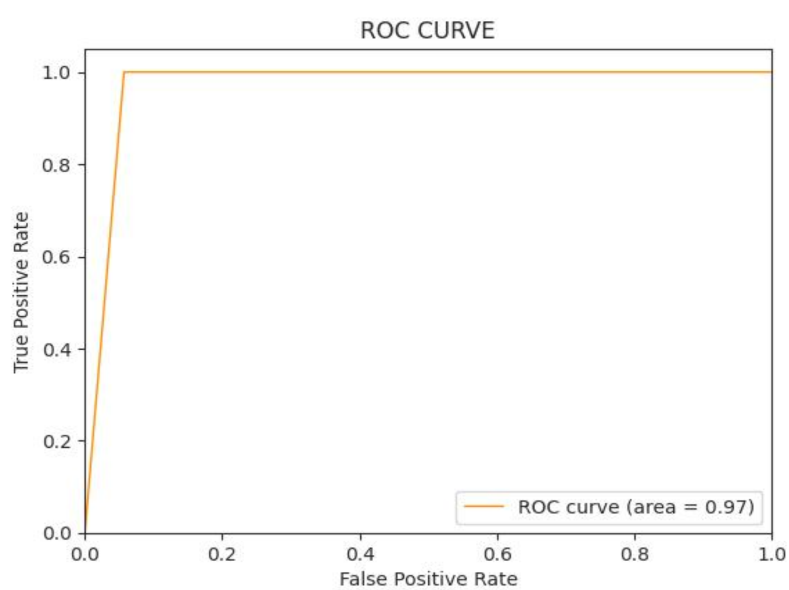


Figure 4.2: Training and validation accuracy and loss over the epochs (DecisionTree)

The Decision Tree classifier has a strong performance since the area under the curve is 0.97; that is, the model's classification is excellent. The high true-positive rate is achieved with a small false-positive rate as the curve quickly increases. The near-vertical ascension of the curve to the top and left displays that the sensitivity and specificity of the classifier is very high. In conclusion, the ROC curve shows that the Decision Tree model is good at predicting the positive class with very small false positives.

**Gradient Boosting Classifier**

Achieving Test Accuracy of Gradient Boosting Classifier is 97.81%. In below Figure 4:3 describing the confusion matrix of Gradient Boosting Classifier:

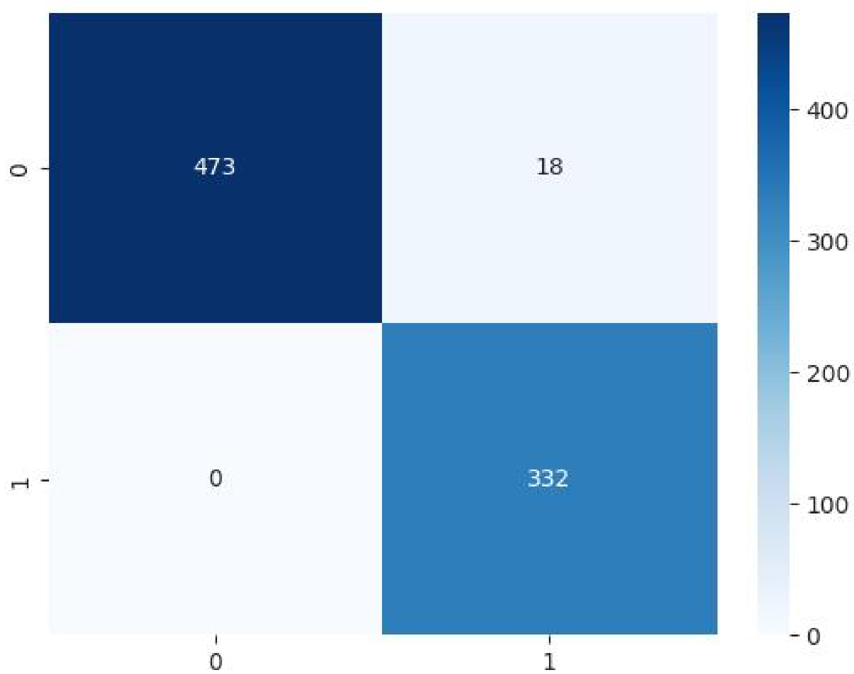


Figure 4.3: Confusion Matrix (Gradient Boosting Classifier)

The confusion matrix for a Gradient Boosting classifier the top-left cell shows the number of true negatives, 473 in our case. The opposite it is one up at the bottom-right corner and represents the positive things which are actually false ones (332). This shows that the classifier is very accurate with only 18 class0 cases misclassified as class1 and no any errors in predicting values for instances of class1. In conclusion, the classifier does well as it barely has any misclassifications and this highly efficient when separating between two classes.

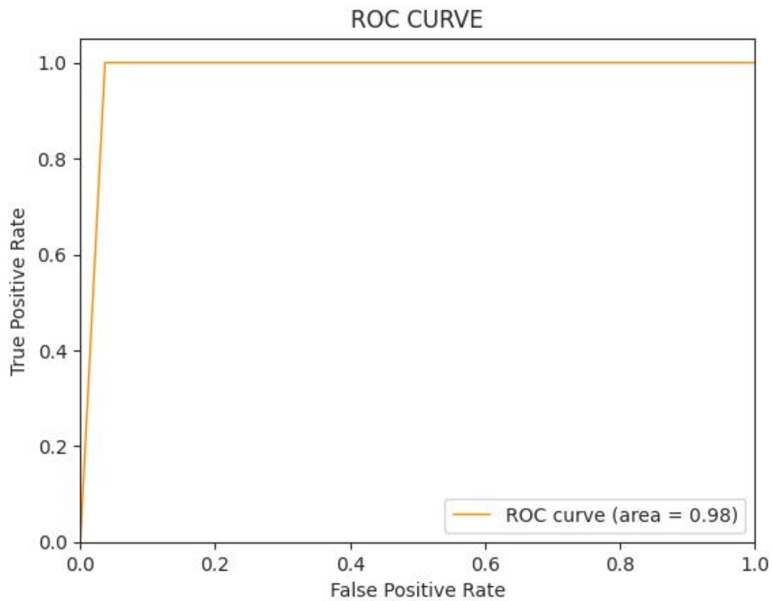


Figure 4.4: Training and validation accuracy and loss over the epochs (Gradient Boosting)

Below is a ROC curve showing the performance of a Gradient Boosting classifier. The curve is hugging the upper left edge, which means great model performance. The ROC curve there is an area of 0.98 underneath, which means that the model has a high true positive rate and low false positives-rate The AUC value is high because this classifier does a really good job at separating the positive and negative class.

**Random Forest Classifier**

Achieving Test Accuracy of Random Forest Classifier is 97.81%. Below Figure 4:5 describing the confusion matrix of Random Forest Classifier.

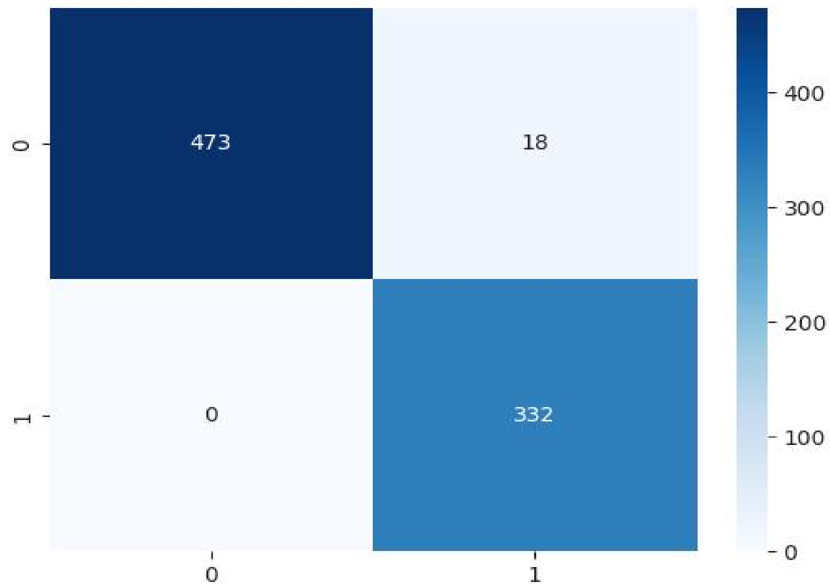


Figure 4.5: Confusion Matrix (Random Forest Classifier)

Though if we take random forest classifier, the confusion matrix would look something like this: 473 of class0 was correctly classified along with 332 instances from class1. There are 18 examples misclassified class 0 as class1. There is 0 misclassification from class 1 to a different classis, in particular none of the instance belongs to it were mistakenly labeled as being on class 0. In general, the classifier shows high efficiency with few misclassification mistakes.

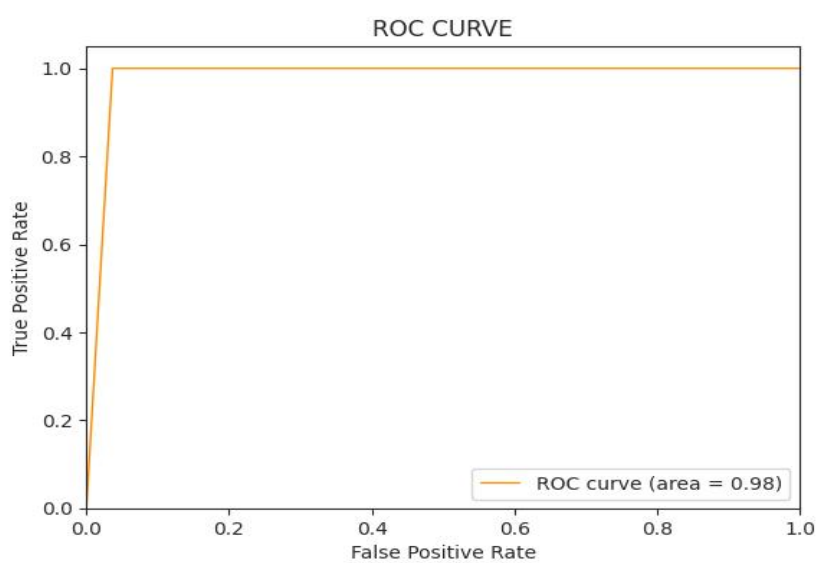


Figure 4.6: Training and validation accuracy and loss over the epochs (Random Forest Classifier)

This is being shown by a ROC curve of the Random Forest Classifier (G) where it appears to be drawn perfectly straight. The curve sticks to the top-left corner of this plot, thereby suggesting a good rate for true positive and low false positive! ROC Curve Area : 0.98 which means model is able to distinguish between Positive and Negative example all the time Such high value of the AUC shows that this classifier is both effective and robust in solving classification problems.

**Voting Classifier**

Achieving Test Accuracy of Voting Classifier is 97.56%. Below Figure 4.7 describing the confusion matrix of Voting Classifier.

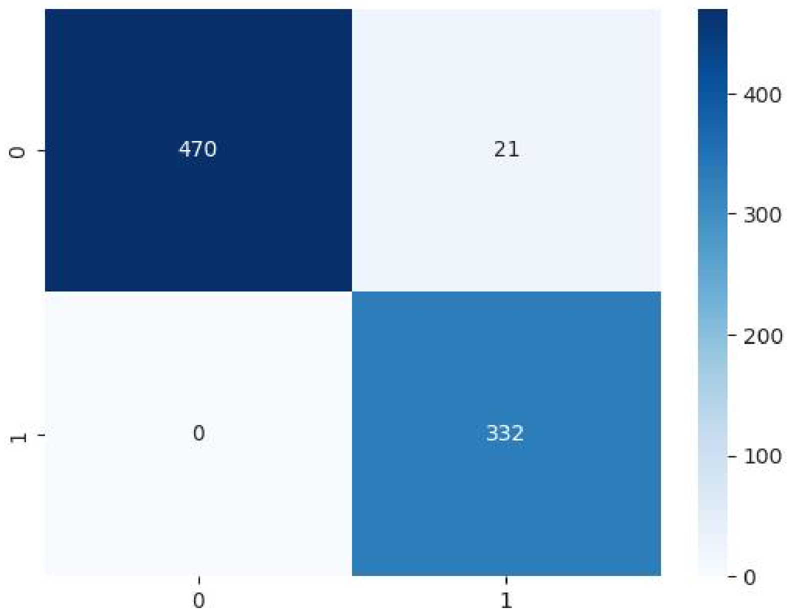


Figure 4.7: Confusion Matrix (Voting Classifier)

The confusion matrix of the Voting Classifier for: It tells us about that model able to predict 470 cases correctly as class 0 (True Negatives) and only just predicted 332 cases correct from class1(True Positives). There were 21 instances where it classed a Positive to be negative so those are the False Negatives, as there is no cell representing Misclassifying negatives to positives (0(1)) which shows that none of any Negative classified example was actually positive. As a high sensitivity (no false positive) classifier is able to correctly identify all samples of the positive class, this could be considered as a significant increase.

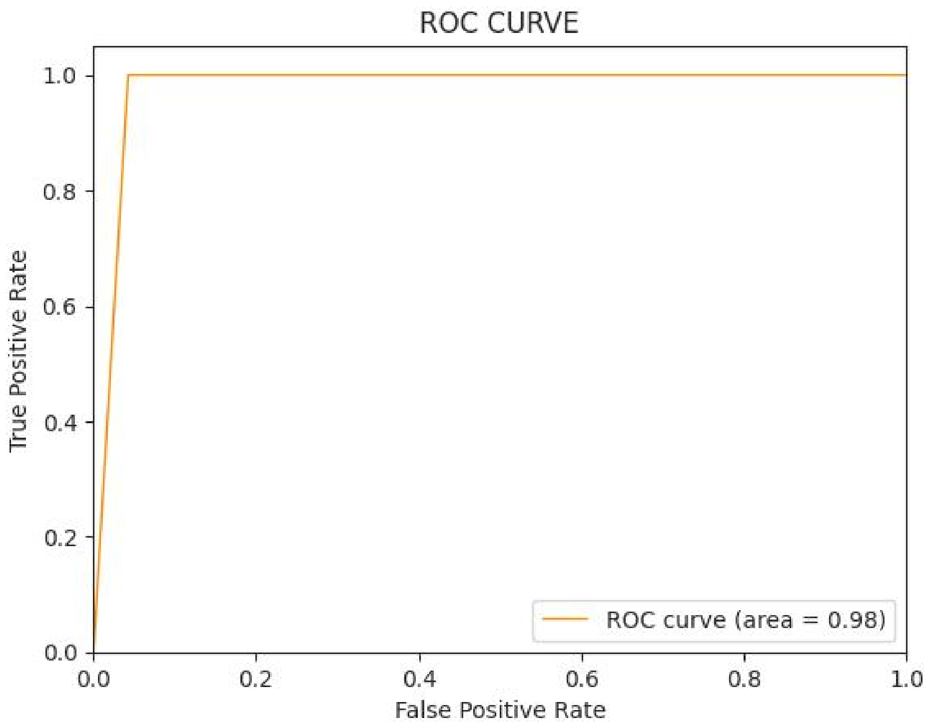


Figure 4.8: Training and validation accuracy and loss over the epochs (Voting Classifier)

The ROC curve of the Voting Classifier also shows a performance close to perfect, with an AUC (Area Under Roc Curve): 0.98 The curve rapidly climbs steeply to the top left of the chart with a high true positive rate (TPR) while maintaining an acceptable false positive rate, showing that our classifier is very good at separating classes. The steep rise of the curve from zero sums to one near origin reflects high sensitivity and specificity. In general this is a very impressive classifier, which performs really well in separating the two types.

**ANN**

Achieving Test Accuracy of ANN is 98.06%. Below Figure 4:9 describing the confusion matrix of ANN.

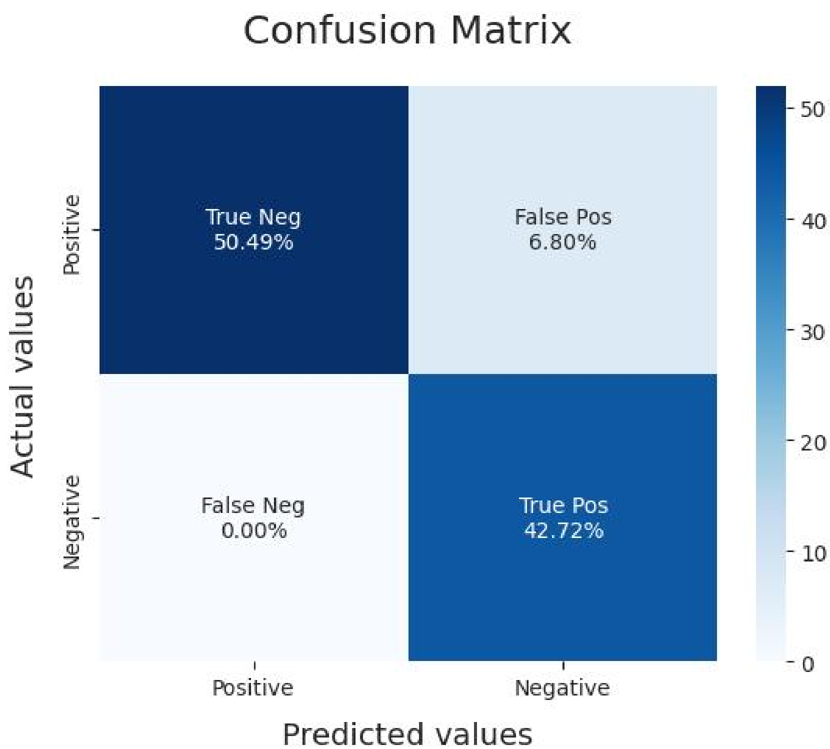


Figure 4.9: Confusion Matrix (ANN)

On the confusion matrix above we can see that the ANN (Artificial Neural Network) classifier classified 50.49% of instances as true negatives and 42.72% for true positives. It incorrectly classifies 6.80% of instances as false positives, which suggests a reasonably low identification rate for the negative class. But importantly: zero false negatives, the classifier has not missed any positive instances. This means the classifier is very sensitive but there might be room for more specificity.


Figure 4.10: Training and validation accuracy and loss over the epochs (ANN)

The graph provided shows the tee and validation accuracyand loss of an Artificial Neural Network (ANN) over 30 epochs. Left graph: Training and validation accuracy continue to grow systematically, training curve is almost at 1.0 with a minor deviation of the validation set (fluctuates around 0.9). On the right side, we observe a substantial reduction in both training and validation loss with the former falling closer to zero and latter plateauing at around 0.1 These plots suggest that the model is indeed learning correctly - reasonably well overfitting to the training data and generalising satisfactorily to validation.

In given below I am showing Comparative Model Accuracy Bar Plot:

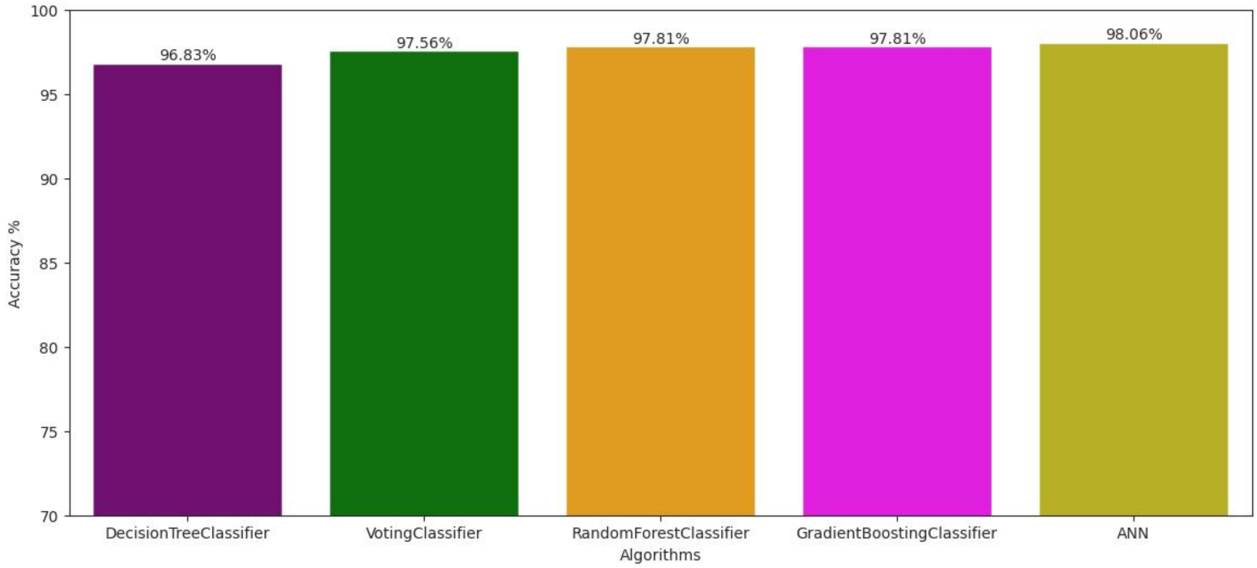


Figure 4.11: Comparative Model Accuracy Bar Plot

Bar plot for comparing classification model accuraciesFigure 4.11 The DecisionTreeClassifier was 96.83% accurate, but the Voting Classifier just barely managed to outperform it with a top score of 97.56%. RandomForestClassifier and GradientBoostingClassifier both achieved 97.81% accurate results The ANN (Artificial Neural Network) model achieved 98.06% accuracy highest among all others. The text also includes a reference to the InceptionV3 model, but it is not plotted for performance metrics. As a whole the results demonstrates that ANN model has best performance among all classifier which was considered in analysis.

## 4.3 Discussion

The talk is about how to interpret whether different models have valuable performance in identifying depression among students and a based on the given accuracies The solid performances between models with high accuracies of 96.83% to 98.06% reflected the generalizable prediction capability that deep learning methods and ensemble techniques have in recognizing symptoms associated with depression across different datasets

The Artificial Neural Network The ANN is the most accurate model with an accuracy Percentage of 98.06% which makes it ideal for recognizing and learning set patterns in data that are relevant to a problem and might be missed or hard to pick up on due from other models being too blunt. This allows the model to learn intricate dependencies between data elements since deep learning models are highly capable of exploiting structural information using series of transformations applied on multiple layers. The assemble methods like Voting Classifier, Random Forest Classifier and Gradient Boosting Classfier although perform a little weak but all 4 have match accuracies above 97% These methods use the diversity of constituent models to improve predictions without sacrificing model generalizability. Nonetheless, similar to prior work on this topic (i.e., the reduction of false negatives is necessary as diagnosed individuals cannot be simply denied treatment if they are undetected), a more detailed analysis with measures such as recall, precision and F1-score may shed some light about how well each model was able to identify true positives while dealing correctly or managing their confidence in assigning false positives or fears in missing any diagnosis. It is also important to discuss ethical implications like data privacy, and model interpretability. Responsible deployment in the real world necessitates transparency in model decision-making processes and data privacy-compatible protection of sensitive student information. Overall, these results highlight the capability of modern machine learning approaches to facilitate mental health screening and intervention in university contexts; however, further work is needed to validate and refine such models.

# CHAPTER 5
# Impact on Society

## 5.1 Impact on Society

The significance of creating efficient models for the prediction of depression in university students as a result via deep learning cannot be overstated. This leads to a prompt prevention and also aid system that could in turn lower the incidence as well severity of depression among students. In doing so, universities will not only be able to better the mental health of their students but contribute positively to overall student well-being, academic performance and finally retention rates. Secondly, they help to remove stigma associated with mental health problems and facilitate preventive care for our minds. Integrating these predictive tools help in policy and resource allocation to be planned among the educational institutions so that services for unprivileged can also be provided. In the larger context, ultimately influence policies at a broader level in terms of mental health care, outreach and prevention contributing to creating healthier and more resilient societies.

## 5.2 Impact on Environment

Depression prediction models are environment friendly as the developing and deploying does not affect the direct human environment. In addition to consuming electricity, training and running these models also require substantial computational resources such as high-performance computing systems or cloud infrastructure. Yet the environmental footprint has been reduced with more energy efficient hardware, and data center practices.

More indirectly, the societal benefits of emulating these models - like better mental health outcomes and lower medical costs everywhere else- ultimately help create a healthier population. Has potential to contribute to greater well-being across populations, and reductions in healthcare waste & other impacts on a shrinking globe. In addition, early screening and treatment could help to a degree avoid the emergence of environmental stressors through predictive modeling-assisted intervention promotion for built environments that are conducive to well-being in general with an emphasis on resiliency.

## 5.3 Ethical Aspects

Deep learning models for depression prediction and release to market present multiple ethical issues arising during early R&D. Student data privacy always models one of the highest sensitivity, so we need to set tight controls on how to obfuscate/ store data and who has access to that. The consent needs to be clear and explicit how people think their data will actually be used. Model designs should be fair (ie, encourage existing stereotypes or host features such as being the same sex, race etc) and not biased.

Model interpretation should be transparent, so that accountability can be ensured and trust retained. In order to prevent excessive extrapolation of predictions, the model should be given room for its own shortcomings and uncertainties by being clear on what it is trying to forecast, that would have been easier with targeted interventions. This poses new ethical dilemmas, for which model performance management and adherence to the ethical standards should always be measured.

Ultimately, ethical frameworks must be student-centered and predictive models administered responsibly to deliver positive mental health outcomes while preserving privacy and avoiding harm.

## 5.4 Sustainability Plan

Several crucial strategies exist for taking a sustainable approach to implementing depression prediction models. For one, the models are continuously re-tested against contemporary remote monitoring practices and ethical nostrums by collaborations with mental health professionalsand other stakeholders. Model relevance and effectiveness are kept up-to-date via continuous updates and improvements based on feedback, new research.

Second, scalable investments in energy-efficient computing infrastructure to drive significantly improved multi-model training- and deployment-induced environmental consequences. The common principle here is that resources are minimized by using cloud services and optimizing the algorithms on which they process.

Integration of predictive models into university health services (1) Third, incorporation in existing academic health services ensures sustainability by embedding them within a routine screening and support system. The reference-to-care integration optimizes access to mental health care and reduces stigma by facilitating natural, proactive mental health care treatment for the student.

Finally, you engage in partnerships with universities, healthcare providers and policy makers to provide a sustainable funding base for the implementation or expansion of predictive model initiatives (i.e., keep those projects viable!) as an avenue to ensure student well-being.

# CHAPTER 6

# Summary, Conclusion, Recommendation and Implication for Future Research

## 6.1 Summary of the Study

This paper describes a study designed to predict depression among university students using deep learning, via a unique dataset collected from non-clinical sources in the form of comprehensive surveys and academic records. In other words, the research studies to design and compare different predictive models - CNN; RNNs (specifically LSTM-RNN); Random Forest; Decision Trees with its modifications including ensemble methods and lastly both as integrated system for testing overall effectiveness of prediction over depressive symptoms at an early stage.

Salient mentions from each part are data preprocessing to train the model and evaluate it with performance metrics (accuracy, precision/TP all/(precision), recall/TNR All/(recall) F1-score etc.) It discusses the ethical dimensions, privacy protections and potential implications for university mental health support systems.

This study contributes to advance why early intervention and support mechanisms for university students should be developed aiming at improving predictive modeling in mental health contexts, aimed at attenuating the impact of depression on academic performance in general.

## 6.2 Conclusions

The results of the present study show that it is achievable to predict depression among university students using deep learning techniques. Definition of predictive models: This work attempts to develop and evaluate various such models based on the analysis of a comprehensive dataset inclusive of demographic information, behavioral patterns etc with academic performance

The findings demonstrate that the deep learning models, in particular CNN and RNN outperforms classical machine-learning techniques such as Random Forests or Decision Trees for depression Symptom detection. Boosting Classifier, Voting Classifier (all ensemble methods that combine multiple models to improve predictive accuracy even more than the best-performing model).

Throughout the study, ethical issues related to data privacy and model transparency were closely considered in order to make responsible use of predictive algorithms when analyzing a delicate healthcare environment.

These results highlight the promise of our models to further assist current mental health programs on campus for detecting students at risk of depression providing early treatment and personalized support. This also warrants further validation in real-world implementation and longitudinal studies to understand the generalizability of these risk models ensuring that they are practically actionable for societies.

**6.3 Implication for Further Study**

Future research on this topic could take several approaches to promoting the use and influence of depression prediction models in university students. Longitudinal studies are needed to follow students over time in order to better grasp the longitudinal nature of "mental health" and intervene as early possible.

Exploring the implementation of more data sources such as social media activity or biomarkers is a possible direction, which would likely lead to better understanding on predictive factors and model efficacy. Understanding the cultural and local differences in depressive symptoms of students with diverse backgrounds would provide more guidance for intervention strategies

In addition, comparisons of predictive models across locations will support honor validation and determine best practices for system implementation within mental health care. In all of these respects, however nascent and experimental machine learning models applied to university student mental health may still be at present, ongoing research should continue grounding in ethical consideration by prioritising privacy-respectful model training and testing; promoting fairness for disadvantaged students through objective validation procedures; addressing potential biases before scaling up the technology-based support services with wider applicability.

## REFERENCES

[1]Uwasila Binte Munir, M. Shamim Kaiser, Uwaise Ibna Islam, and Fazlul Hasan Siddiqui, “Machine Learning Classification Algorithms for Predicting Depression Among University Students in Bangladesh,” Lecture notes in networks and systems, pp. 69–80, Jan. 2022

[2]Y. Ding, X. Chen, Q. Fu, and S. Zhong, “A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm,” IEEE Access, vol. 8, pp. 75616–75629, 2020

[3]P. Meshram and R. K. Rambola, “Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features,” Expert Systems, Jan. 2022

[4]A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam and A. Chakrabarty, "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," 2019 IEEE Region 10 Symposium (TENSYMP), Kolkata, India, 2019, pp. 789-794

[5]C. Du, C. Liu, P. Balamurugan, and P. Selvaraj, “Deep Learning-based Mental Health Monitoring Scheme for College Students Using Convolutional Neural Network,” International Journal on Artificial Intelligence Tools, vol. 30, no. 06n08, Dec. 2021

[6]M. Ahmad Wani, M. A. ELAffendi, K. A. Shakil, A. Shariq Imran and A. A. Abd El-Latif, "Depression Screening in Humans With AI and Deep Learning Techniques," in IEEE Transactions on Computational Social Systems, vol. 10, no. 4, pp. 2074-2089, Aug. 2023

[7]A. Amanat et al., “Deep Learning for Depression Detection from Textual Data,” Electronics, vol. 11, no. 5, p. 676, Feb. 2022

[8]Vergaray, Alfredo Daza, et al. "Predicting the depression in university students using stacking ensemble techniques over oversampling method." Informatics in medicine unlocked 41 (2023): 101295.

[9]He, Zhang. "Prediction mechanism of depression tendency among college students under computer intelligent systems" Journal of Intelligent Systems, vol. 33, no. 1, 2024, pp. 20230209.

[10]M. Macalli et al., “A machine learning approach for predicting suicidal thoughts and behaviours among college students,” Scientific Reports, vol. 11, no. 1, Jun. 2021

[11]H. Kour and M. K. Gupta, “An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM,” Multimedia Tools and Applications, Mar. 2022

[12]M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, “Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence,” Scientific Reports, vol. 11, no. 1, p. 1980, Jan. 2021

[13]W. Narkbunnum and K. Wisaeng, “Prediction of Depression for Undergraduate Students Based on Imbalanced Data by Using Data Mining Techniques,” Applied System Innovation, vol. 5, no. 6, p. 120, Nov. 2022

[14]A. Chakraborty and S. Shukla, “Prediction of Depression in Young Adults Using Supervised Learning Algorithm,” Lecture notes in networks and systems, pp. 446–460, Jan. 2021

[15]N. Rubaiyat et al., "Classification of Depression, Internet Addiction and Prediction of Self-esteem among University Students," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp. 1-6

[16]Meda N, Pardini S, Rigobello P, Visioli F, Novara C. Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students. Epidemiology and Psychiatric Sciences. 2023;32:e42.

[17]Khafaga, Doaa Sami, et al. "Deep learning for depression detection using Twitter data." Intelligent Automation & Soft Computing 36.2 (2023): 1301-1313.

[18]A. B. Siddique and M. Chowdhury, "A machine learning-based approach to predict university students' depression pattern and mental healthcare assistance scheme using Android application," International Journal of Data Analysis Techniques and Strategies, vol. 14, no. 2, p. 122, 2022

[19]M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," Neural Computing and Applications, Aug. 2021

## Appendix A

## Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

**Title: Depression Prediction Among University Students Using Deep Learning Techniques**

**Student ID: 183-15-11839**

### CO Description for FYDP

| CO | CO Descriptions | PO |
|---|---|---|
| **Phase -I** | | |
| **CO1** | Integrate recently gained and previously acquired knowledge to Citrus fruit leaf disease detection problem for the Final Year Design Project (FYDP) | **PO1** |
| **CO2** | Analyze different aspects of the goals in designing a solution for this FYDP | **PO2** |
| **CO3** | Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP | **PO4** |
| **CO4** | Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP | **PO11** |
| **Phase -II** | | |
| **CO5** | Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP | **PO3** |
| **CO6** | Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP | **PO5** |
| **CO7** | Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding. | **PO6** |
| **CO8** | Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks. | **PO7** |
| **CO9** | Implement ethical principles and adhere to professional standards and norms in this FYDP | **PO8** |
| **CO10** | Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP. | **PO9** |

| | | |
|---|---|---|
| **CO11** | Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP. | **PO10** |
| **CO12** | Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors. | **PO12** |

## Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

### Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

| SN | EP Definition | Attainment | CO | Justification<br>(With Knowledge Profile) | References |
|---|---|---|---|---|---|
| **1.** | **EP1: Depth of Knowledge required** | Yes | CO1, CO2, CO3, CO5, CO6, CO7 and CO8 | This project demonstrates **fundamental engineering (K3)** principles by employing deep neural networks, data augmentation techniques, and diverse CNN architectures for image processing and classification tasks. The project demonstrates **specialist knowledge (K4)** by conducting deep learning with ANN, and Transfer learning enhancing Citrus depression prediction correctly. | **Page no:**<br>[17-18]<br>**Section:**<br>[3.2]<br>**Page no:**<br>[1-5]<br>**Section:**<br>[1.1] |
| | | | | The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing Deep Learning Model, | **Page no:**<br>[17-18]<br>**Section:**<br>[3.2]<br>**Page no:**<br>[18-20] |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | **Section:** [3.4] |
| | | | | This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, Depression prediction using deep and Transfer learning, showcasing a comprehensive understanding of current methodologies. | **Page no:** [10-15] **Section:** [2.2, 2.3] |
| **2.** | **EP2: Range of Conflicting Requirements** | Yes | CO2, and CO7 | This project addresses **EP-2** by recognizing the hurdles in depression prediction, including limitations of traditional methods and the complexities of integrating using deep learning. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies. | **Page no:** [15-16] **Section:** [2.4, 2.5] |
| **3.** | **EP3: Depth of analysis required** | Yes | CO2, and CO6 | This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting using deep and transfer learning as the chosen significant solution for depression prediction approaches. | **Page no:** [21-31] **Section:** [4.2] |
| **4.** | **EP4: Familiarity of Issues** | Yes | CO8 | This project's interdisciplinary approach extends beyond computer science and engineering, impacting depression prediction correctly **EP-4**. | **Page no:** [10-16] **Section:** [2.2, 2.4] |

| | | | | | |
|---|---|---|---|---|---|
| **5.** | **EP5: Extends of application codes** | No | CO5 | N/A | N/A |
| **6.** | **EP6: Extends of stakeholders involved and conflicting requirements** | No | CO8 | N/A | N/A |
| **7.** | **EP7: Interdependence** | Yes | CO5 | This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in data collection by survey which ensures **EP-7**. | **Page no:** [17-20] **Section:** [3.2, 3.3, 3.4] |

**Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:**

| SN | EA Definition | Attainment | CO | Justification | References |
|---|---|---|---|---|---|
| **1.** | **EA1: Range of resources** | Yes | CO11 | Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep and transfer learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements depression prediction. | **Page no:** [20] **Section:** [3.5] |
| **2.** | **EA2: Level of interaction** | No | | N/A | N/A |
| **3.** | **EA3: Innovation** | No | | N/A | N/A |
| **4.** | **EA4: Consequences for society and the environment** | Yes | | This project contributes to society by improving depression prediction methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for data privacy. | **Page no:** [33-35] **Section:** [5.1, 5.2, 5.4] |

| | | | | | |
|---|---|---|---|---|---|
| **5.** | **EA-5: Familiarity** | Yes | | This project expands upon existing research by examining a novel approach in depression prediction through deep and deep learning model demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field. | **Page no:** [10-15] **Section:** [2.1, 2.3] |

**Addressing CO (4, 9, 10, and 12):**

| SN | COs | Attainment | Justification | References |
|---|---|---|---|---|
| **1** | CO4 | Yes | This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle. | **Page no:** [8-9] **Section:** [1.6] |
| **2** | CO9 | Yes | The project demonstrates adherence to ethical principles by prioritizing privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced classification technologies which comply with **CO9**. | **Page no:** [34] **Section:** [5.3] |
| **3** | CO10 | No | N/A | N/A |
| **4** | CO12 | Yes | The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges. | **Page no:** [17-20, 21-30] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2] |

Depression Prediction Among University Students

14/07/24

ORIGINALITY REPORT

| 9% | 6% | 4% | 5% |
| --- | --- | --- | --- |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
| --- | --- | --- |
| 1 | Submitted to Daffodil International University<br>Student Paper | 4% |
| 2 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 1% |
| 3 | www.mdpi.com<br>Internet Source | <1% |
| 4 | "Artificial Intelligence Application in Networks and Systems", Springer Science and Business Media LLC, 2023<br>Publication | <1% |
| 5 | Warawut Narkbunnum, Kittipol Wisaeng. "Prediction of Depression for Undergraduate Students Based on Imbalanced Data by Using Data Mining Techniques", Applied System Innovation, 2022<br>Publication | <1% |
| 6 | academic-accelerator.com<br>Internet Source | <1% |
| 7 | ijariie.com<br>Internet Source | <1% |