Exploring the Impact of Social Media on Mental Health and Well-being: A Multi-dimensional Analysis

MSc Research Project
MSc Data Analytics

Vasit Ali
Student ID: x22144170

School of Computing
National College of Ireland

Supervisor: Abid Yaqoob

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vasit Ali
Student ID:	x22144170
Programme:	Masters in Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Abid Yaqoob
Submission Due Date:	31/01/2024
Project Title:	Exploring the Impact of Social Media on Mental Health and Well-being: A Multi-dimensional Analysis
Word Count:	7705
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Vasit Ali
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Exploring the Impact of Social Media on Mental Health and Well-being: A Multi-dimensional Analysis

Vasit Ali
x22144170

Abstract

This research uses advanced machine learning techniques to analyse the impact of demographic and various mental illness related variables on the diagnosis of depression, a critical issue in mental health. The dataset consists of responses from diverse demographics, focusing on factors such as birth year, state, gender, student status, and international student status. The methodology encompasses thorough data preprocessing, including label encoding and correlation analysis, followed by an exploratory data analysis utilizing histograms and pie charts for visual insights. Four machine learning models - Decision Tree, Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) – are implemented and evaluated. The models performance is assessed using accuracy, precision, recall, and F1-score metrics. This multi-model approach provides a holistic view of the data, revealing intricate patterns and associations between demographics and depression diagnosis. The findings indicate correlations between certain demographic characteristics and the likelihood of a depression diagnosis. This study not only contributes to the academic understanding of mental health factors but also has practical implications for healthcare policy and individualized interventions. It highlights the critical role machine learning in deciphering complex health-related data, offering a foundation for future research in this area. This research demonstrates the fusion of data science and mental health, opening avenues for more targeted and effective mental health strategies. By leveraging machine learning, it underscores the potential for data-driven approaches in enhancing our understanding and management of mental health issues.

1 Introduction

1.1 Research Background

Social media platforms have completely changed the way individuals exchange information, communicate, and develop relationships. Social media has already raised the questions about its impact on mental health and general well-being, even though it presents never-before-seen possibilities for communication and community development. The way that digital communication is developing needs an in-depth investigation of the social and psychological effects of heavy social media use Twenge and Campbell (2018).The widespread usage of social media in today's digital environment has emerged as a key characteristic of modern interaction and communication. As the number of people using virtual spaces is getting increased day by day, it also a reason of concern about how social media affects people's psychological, emotional, and social health. The main aim of this research is to understand the complexity in relationship between social media use and wellbeing, as well as it's a step to suggest ways to individuals to reduce any potential negative effects in their life.

1.2 Motivation

The motivation behind this research lies in necessity to examine the effects of social media usage on users' mental health and social connections. The research is propelled by the escalating prevalence of online platforms and the consequential impact on individuals and society. As the number of users engaging with social media continues to rise, concerns have grown regarding potential consequences for mental well-being and interpersonal relationships. This investigation aims to delve into the psychological impacts, such as anxiety and depression, associated with prolonged social media use, enabling the development of targeted interventions and support systems. Additionally, it seeks to understand how online interactions influence the quality of relationships, evaluating factors like communication patterns and the sense of social support. The research also aims to identify strategies, including the development of guidelines, educational programs, and collaborative efforts with platforms, to mitigate adverse effects and promote healthier social media use. Ultimately, the goal is to strike a balance that allows individuals to harness the benefits of online interactions while minimizing potential negative consequences, contributing to a more informed and resilient digital society.

1.3 Research Question

"How does social media usage impact users' psychological, social, and emotional well-being, and what measures can individuals and society implement to mitigate potential negative effects?" This research investigates the diverse impacts of social media usage on users' psychological, social, and emotional well-being. It aims to understand the intricate dynamics involved, exploring effects on self-esteem, social interactions, and emotional experiences. Additionally, the study seeks to propose measures, both at the individual and societal levels, to mitigate potential negative consequences. By adopting a mixed methods approach and emphasizing ethical considerations, the research aims to provide valuable insights into fostering a healthier social media environment and promoting user well-being in the digital age.

1.4 Research Objectives

The following objectives have been set to address to this question:

- **Analysis of Negative Effects:** To find and measure any possible harm that social media use may do to a user's psychological, social, or emotional health, apply statistical analysis Twigg et al. (2020).

- **Mitigation Measures:** Provide evidence-based measures and strategies to individuals and society to lessen and to prevent them from the harmful effects of social media usage.

1.5 Structure of the report

The report structure includes Section 1, presenting the study's background, motivation, research questions, and outline. The Section 2 reviews relevant literature, setting the context for this study. In the Section 3, the process of data selection, exploration, preparation, and model training is detailed, along with evaluation strategies. The Section 4 describes the modelling and evaluation techniques used. The Section 5 covers the tools used, data handling, exploratory data analysis, hyperparameter tuning, model evaluation, visualization techniques, and comparative analysis of models. The Section 6 involves detailed case studies assessing model performance. The discussion section offers an in-depth analysis of findings considering existing literature. The Section 7 summarizes key findings and proposes directions for future research. Finally, a comprehensive list of References is included to credit sources and provide further reading.

2 Related Work

In this research, we explore the multi-dimensional impact of social media on mental health and well-being, prompted by escalating concerns over the excessive use of social media and its potential effects on individuals' psychological, social, and emotional states. The aim is to unravel how prolonged social media engagement may give rise to issues such as addiction, social comparison, and cyberbullying, while also identifying effective strategies to promote healthy online habits De Choudhury and De (2014). The study follows a structured approach, commencing with an introduction that underlines the ubiquity of social media in daily life and the associated challenges. The literature review, comprising three key domains, explores the psychological effects of social media on adolescents and young adults, harnesses machine learning for mental health analysis from social media studies, and scrutinizes social media addiction with a focus on psychological patterns and behavioural outcomes. By synthesizing insights from these diverse perspectives, the research aims to contribute new dimensions to the existing body of knowledge, providing a nuanced understanding of the complex interplay between social media and mental health Andreassen (2015). This collective analysis intends to empower individuals and policymakers with tools to cultivate a supportive and mentally healthy digital environment, emphasizing positive online habits. The motivation behind this study lies in the imperative to comprehend the impact of increasing social media usage on mental health and, subsequently, devise strategies to safeguard individual well-being, potentially informing policy frameworks to create a healthier social media environment for all.

2.1 Digital Age Dilemmas: The Psychological Effects of social media on Adolescents and Young Adults

The researchers in Abirami (2022) investigates the effects of digital media on the psychological and lifestyle changes in young individuals. With digital media offering diverse features such as communication, image and video sharing, and fast information access, it has become crucial for all age groups. The aim is to understand the impact on the well-being of the youth. A validated questionnaire with 15 questions was administered, garnering 350 responses. The data underwent tabulation, graphical representation, and Chi-square analysis. Demographically, (58.2% were males, and 58.8% were students. Common activities leading to excessive usage were watching series/movies (42.2%) and keeping updated (34.5%). 61.6% felt online content affected their mood, and 75.4% deemed likes on posts important. Procrastination was noted in 62.1%, while 57.5% reported positive lifestyle changes. Exposure to digital media widened viewpoints for 77.65%. The study contributes insights into the nuanced relationship young individuals have with digital media. It encourages a holistic understanding of the psychological and lifestyle impacts, acknowledging both positive and negative aspects. Future research could delve deeper into specific content types, age groups, and cultural nuances to refine recommendations for a healthier digital engagement which our research covers it. While the study provides valuable insights, our project, focusing specifically on the impact of social media on psychological, social, and emotional well-being, introduces a more nuanced and targeted research question. The refined focus on social media allows for a deeper exploration of its influence on users' mental health, social connections, and emotional states.

The researchers in Luca et al. (n.d.) investigates the medical risks associated with social media addiction among 15-24 years old, with focused attention on social media's impact on well-being. Their study focuses on medical ramifications, contributes to the growing concern surrounding excessive social media use, Aligning with existing literature on mental health implications. Their target's narrow demography supports earlier studies that identify age-related differences in social media usage. In our research, while acknowledging the significance of medical perspectives extends the examination to target psychological, social, and emotional well-being. Both the studies share a common goal of public awareness and preventive measures, it's just that our project widens the study to provide a better understanding of the multifaceted effects of social media usage on individual life.

The study Satyaninrum et al. (2023) by highly influences our research into broader implications of social media usage. Our project mirrors the concerns raised by this study, mainly regarding the detrimental effects of intense social media engagement among adolescents. Much like the study, we acknowledge the negative outcomes tied to increased social media use, for example drop in self-esteem and a surge in dissatisfaction with their physical appearance. The study's revelations about heightened depression and anxiety levels in heavy social media users are also a part of our investigation, which aims to thoroughly examine the impact on psychological, social, and emotional well-being. While our study extends beyond the specific location of West Java, the shared findings emphasize the universal nature of the issue. As we explore mitigation and preventative techniques, our research is in line with the larger goal of enhancing mental health vulnerability in the digital age.

The researchers in Chukwuere and Chukwuere (2023) synthesizes existing research on the multifaceted relationship between social media usage and individuals' psychological, social and well-being. The project comprehensively explores both positive and negative impacts of social media, emphasizing the need for a nuanced understanding of its effects. Drawing on multidisciplinary perspectives, including psychology, sociology, and data science, the study identifies common themes, such as the influence on body image and self-esteem. Apart from common themes we've some distinct merits as well. Firstly, our research extends the investigation beyond body image and self-esteem, encompassing a broader spectrum. Additionally, our incorporation of machine learning models to analyse survey data showcases a forward-looking and innovative methodology, enhancing the depth and rigor of our research. Ultimately, the project's scope, actionable recommendations, and methodological innovation collectively position it as a valuable contribution to the evolving discourse on social media and mental health.

2.2 Harnessing Machine Learning and Data Science in Mental Health Analysis: Insights from Social Media Studies

The researchers in Kim et al. (2021) analysed over 1.2 million tweets and other data from Weibo, Instagram, Reddit, and Facebook employing logistic regression and sentiment analysis to identify tweets and posts related to mental health stigma or seeking help, with an AUC of 0.83 while our project involves mainly analysing survey responses but at the same time focuses on a broad analysis of mental health factors. This study Kim et al. (2021) Employs a variety of techniques, including logistic regression, sentiment analysis, CNN models, linear regression, tensor techniques, deep-learning models (LSTM), SVM, and more. While both projects share commonalities in terms of mental health analysis and prediction, they differ in terms of data sources, specific applications, techniques, and platforms. The study Kim et al. (2021) highlights the diverse approaches and applications of machine learning in monitoring mental health across various social media platforms while our project's unique contribution lies in its use of survey data for a comprehensive analysis of mental health factors.

The research Boy (2023) presents a comprehensive analysis utilizing data mining and processing frameworks to examine the time series data in conjunction with social media feeds and well-being surveys. Focusing on mental health and well-being during the COVID-19 pandemic, the study establishes a significant correlation between Google Trends data and the public's mental health status. Comparatively, our research employs a range of machine learning models, including decision trees, random forests, support vector machines, and neural networks, to analyse a dataset related to mental health and depression. By utilizing features such as demographic information and student status, your models aim to predict whether individuals have been diagnosed with depression. Both studies, however, share the common goal of leveraging data-driven approaches to gain valuable insights into mental health, with our research focusing on prediction and classification based on demographic features. Integrating elements from both approaches could potentially enhance the understanding and application of open-source intelligence in mental health research.

The researchers in Khasnis et al. (2022) conducts a comprehensive exploration of sentiment analysis in the context of the COVID-19 pandemic using machine learning algorithms to categorize tweets into fear or panic sentiments. With a dataset consist of over 900,000 tweets collected from Twitter during the early months of the pandemic, the study used Naïve Bayes and Logistic Regression algorithms, achieving significant accuracies for brief tweets. Visualizations, including fear curves and geographical maps, contribute to an understanding of sentiment dynamics over time and across regions. The limitations of the current research include a confined temporal scope (limited to February 2020 to March 2020), potentially hindering a comprehensive understanding of sentiment evolution throughout the entire pandemic. The report mentions the potential for algorithmic improvements without providing specific details. Addressing these limitations could enhance the study's overall depth and applicability.

The researchers in Chadha et al. (2022) emphasizes the significance of addressing mental health issues, particularly the identification of individuals at risk of suicide using social media data. The study aims to distinguish between suicidal and non-suicidal posts collected from social media platforms. Data is gathered from the "Suicide Watch" subreddit, comprising 10,000 user posts. Natural Language Processing (NLP) techniques are applied, including tokenization, stop word removal, and elimination of non-alphabetic characters. Support Vector Machine (SVM), Logistic Regression, and AdaBoost are employed for classification. Performance metrics include accuracy, precision, and recall. Results indicate that SVM has the highest precision (80.72%), while Logistic Regression achieves the best accuracy (80.75%) and recall (77.81%). For the limitations, the reliance on social media data, particularly from Reddit, may introduce biases as users vary in their posting habits and the platforms may not be fully representative of the general population. Overall, while the study provides valuable insights, these limitations emphasize the need for cautious interpretation and the consideration of broader contextual factors in understanding mental health through social media analysis.

The research Illahi et al. (2022) discusses the impact of social media on human interactions and the potential negative consequences, including stress. It highlights the role of machine learning, particularly natural language processing, in detecting mental health issues. The study explores classical machine learning techniques and ensemble approaches such as boosting, bagging, and voting. The results are compared with baseline models and show that the proposed models outperform classical and non transformer based deep learning models. The study acknowledges limitations such as a small dataset and a monolithic data source and suggests future work to extend the research to a more diverse and larger dataset, incorporating deep learning techniques like BERT.

2.3 Social Media Addiction: Psychological Patterns and Behavioural Outcomes

In this study Priya and Prakash (2023) prevalence, economic burden, and consequences of underdiagnosis are highlighted. The diverse data sources, including electronic health records and social media, are discussed. The global burden of depression and the potential of predictive models in aiding diagnosis are underscored. Concepts such as sentiment analysis and emotion detection are introduced as tools for comprehending user emotions and identifying signs of depression within the vast pool of social media data. Research gaps are identified, including the need for more sophisticated models and methodologies. It encourages researchers to explore new avenues for improving accuracy and efficiency in depression detection through social media data.

The research in Tadesse et al. (2019) employs Natural Language Processing (NLP) techniques and machine learning (ML) approaches to analyse Reddit users' posts and detect factors which indicates depression. The goal is to enhance the efficiency of depression detection through innovative methodologies. The study utilizes a dataset from Reddit, consists of posts from depressed and non-depressed users. The framework involves data pre-processing, feature extraction, and the application of various text classification techniques, namely Logistic Regression, Support Vector Machine, Random Forest, Adaptive Boosting, and Multilayer Perceptron. The features extracted include linguistic dimensions, psychological processes, personal concerns, N-gram probabilities, and topics identified using Latent Dirichlet Allocation (LDA). The MLP classifier, in conjunction with LIWC, LDA, and bigrams, stands out as the most effective in detecting depression. As per limitations the reliance on data from Reddit raises concerns about generalizability to other online platforms and cultural contexts. The accuracy of user self-diagnosis in specific subreddits introduces potential labelling inaccuracies. These limitations highlight the need for cautious interpretation and suggest avenues for future research improvement.

The researchers in Reece et al. (2017) focuses on developing computational models for predicting the emergence of depression and Post-Traumatic Stress Disorder (PTSD)

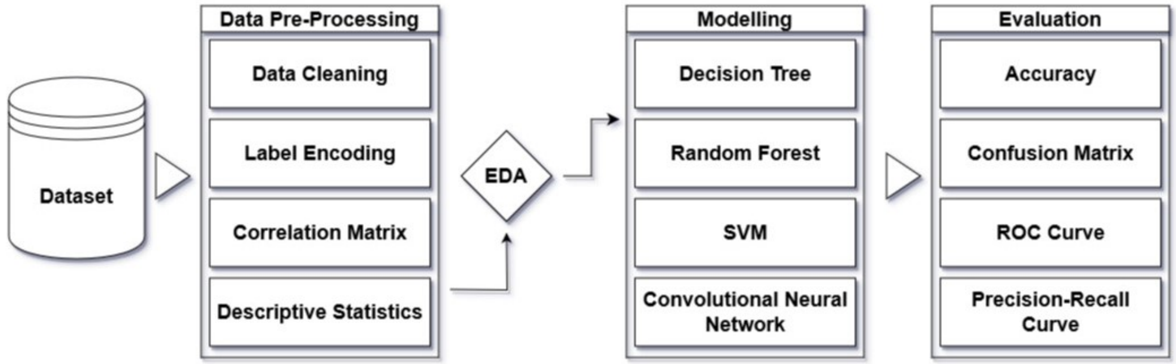

Figure 1: Project Workflow

in Twitter users. Data from 204 individuals (105 depressed, 99 healthy) were collected, including Twitter content and details of depression history. Predictive features related to affect, linguistic style, and context were extracted from participant tweets (279,951 in total), and models were built using supervised learning algorithms. The importance of social media in providing early-warning indicators for health conditions is emphasized, and the research contributes improved methods for predicting and tracking depression and PTSD on Twitter. The study provides valuable insights into leveraging social media data for early mental health screening and emphasizes the potential societal impact of such predictive approaches. The study relies on Twitter data, which may not fully capture individuals' mental health statuses or experiences. The use of social media content for mental health prediction assumes that users express their genuine feelings and that these expressions are reflective of their mental health. However, social media posts may be curated or filtered to present a particular image, and users may not always disclose their mental health struggles openly.

2.4 Critical Reflections and Insights from Reviewed Studies on Social Media and Mental Health

Despite the valuable insights provided by the literature reviewed, several limitations and key takeaways emerge. The studies, while contributing to an understanding of the impact of social media on mental health, exhibit limitations such as narrow demographic focus, reliance on specific social media platforms, potential biases in user behaviour, and limited generalizability. Additionally, the methodologies employed, including machine learning and data analysis, vary across studies, affecting the comparability of results. The need for nuanced context-aware approaches to comprehend the complex interplay between social media usage and mental well-being becomes evident. Despite these limitations, the literature underscores the urgency of addressing the escalating concerns over excessive social media use and its multifaceted effects on psychological, social, and emotional states. The collective analysis encourages a holistic perspective, emphasizing the positive and negative impacts, and sets the stage for more comprehensive, diverse, and contextually sensitive research to inform interventions and strategies for promoting a healthier digital environment.

3 Methodology

3.1 Data Selection

For this study, data was sourced from a comprehensive survey examining the impact of social media on mental health. The dataset[1], part of the ‘ Social Media and Mental Health’ project, includes detailed responses from a diverse range of variables including demographics, primarily young adults. Key aspects such as survey response times, engagement levels, and a wide range of mental health-related questions were analysed. The selection prioritized data that could provide insights into the relationship between social media usage and various mental health outcomes Bryman (2016).

3.2 Data Exploration

The data exploration step including assessing response patterns, analysing demographic information, and understanding the distribution of key variables. The survey contained diverse questions on mental health and social media usage. Analysis focused on identifying trends, outliers, and patterns in the data. This comprehensive exploration aimed to ensure a robust understanding of the dataset's structure and contents, laying the groundwork for subsequent statistical analysis.

1. Data Composition Analysis: The dataset was thoroughly examined to understand its structure, including the number of entries, range of responses, and types of questions. This involved identifying the nature of each column - numeric, categorical, Boolean, etc.

2. Demographic Assessment: Special attention was given to demographic data, particularly the age distribution of respondents as shown in Figure 2 and Figure 3, to ascertain the primary audience of the survey and its relevance to the study's focus on young adults Trochim and Donnelly (2001).

3. Engagement Metrics Evaluation: Metrics such as response duration and timing of clicks were analysed to gauge respondent engagement and identify potential outliers or anomalies in response patterns.

3.2.1 General Overview

1. Entries and Columns: The dataset contains 581 entries (responses) across 129 columns (questions or data points).

2. Data Types: The columns include various data types:

3. Numeric: 44 numeric columns, including integers and floats.

4. Categorical/String: 84 object-type (string) columns.

5. Boolean: 1 Boolean column.

6. Memory Usage: Approximately 581.7 KB.

[1] `https://www.openicpsr.org/openicpsr/project/175582/version/V1/view?path=/openicpsr/175582/fcr:versions/V1/Data/Input/MentalHealthSurvey/Mental_Health_Survey_Feb_20_22.csv&type=file`

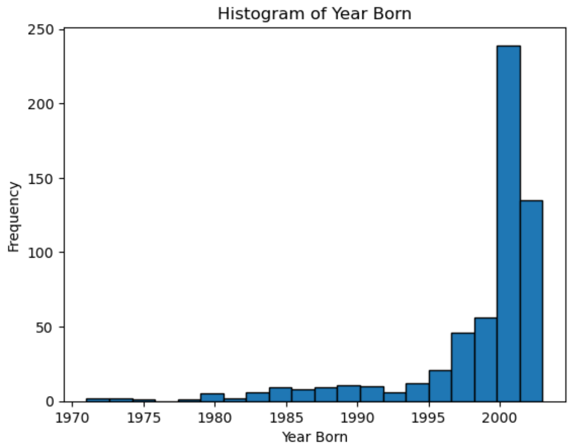

Figure 2: Histogram of Year Born

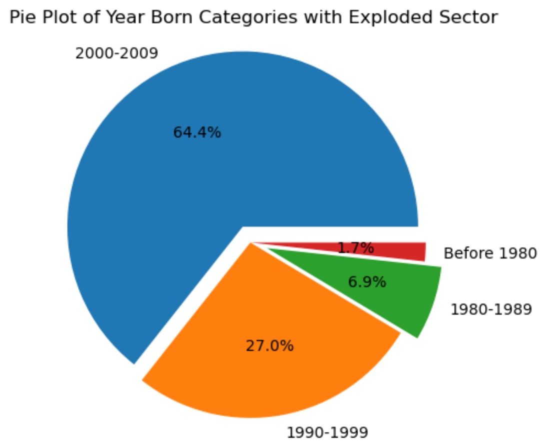

Figure 3: Pie Chart of Year Born

3.2.2 Descriptive Statistics for Numeric Columns

1. Progress: All responses show 100% progress, indicating completion.

2. Duration: The duration of responses varies significantly, with a mean of approximately 279 seconds. The maximum duration is 10,693 seconds, indicating some very lengthy responses.

3. Year Born: Respondents' birth years range from 1971 to 2003, with a mean birth year of 1998. This suggests a predominantly young adult demographic.

4. Timing Metrics: Various timing-related columns (e.g., 'Timing - First Click', 'Timing - Last Click') show a wide range of values, indicating different engagement levels.

3.3 Data Preparation

The data preparation phase was primarily conducted using Jupyter notebook, a powerful tool for data cleaning and analysis. Initially, the raw data was imported from an Excel sheet. This step involved handling various data types and ensuring compatibility with analytical tools. Additionally, the transformation of categorical data into a format suitable for statistical analysis was performed. This preparation phase was crucial in ensuring the accuracy and reliability of the dataset for further analysis.

3.3.1 Data Cleaning

Each step of this process was carefully executed to ensure the resulting dataset was clean, relevant, and structured for effective analysis Osborne (2013).

1. Initial Dataframe Creation: The raw data from the Excel file was loaded into a pandas DataFrame for manipulation and analysis.

2. Irrelevant Column Removal: Columns such as 'Start Date', 'End Date', 'Response Type', etc., were identified as not pertinent to the analysis and thus dropped.

3. Timing Metrics Removal: Columns related to response timing ('Timing - First Click', 'Timing - Last Click', etc.) were also removed, as they were not relevant to the study's objectives.
4. Dropping Specific Health-Related Columns: Certain health-related columns, not directly relevant to mental health or the study's focus, were dropped.
5. Removal of Introductory and Feedback Columns: Columns containing introductory text and feedback questions were removed to streamline the dataset.
6. Label Encoding: Categorical variables, such as 'State' and various health-related questions, were encoded into numerical form using Label Encoding for easier analysis.
7. Correlation Analysis: A correlation matrix heatmap was generated to identify relationships between variables. This step also involved further column removal based on insights from the correlation analysis.

3.3.2 Feature Engineering

Below approach in feature engineering focused on preparing the data in a way that maximizes its relevance and utility for the specific analyses and models used in the study.

1. Label Encoding of Categorical Variables: Categorical data such as state of residence, general health status, and various mental health-related responses were converted into a numerical format using Label Encoding. This was crucial for enabling subsequent statistical analyses and machine learning models, which often require numerical inputs Kotsiantis et al. (2007).
2. Correlation Analysis for Feature Selection: A detailed correlation matrix heatmap was generated, offering insights into the relationships between variables. This step was instrumental in identifying highly correlated features. Understanding these correlations helped in making informed decisions about feature selection, ensuring that the features included in the model provided unique and relevant information.
3. Refinement of Feature Set: Based on the correlation analysis, certain features were dropped to avoid redundancy and potential multicollinearity in the model. This step was essential in refining the feature set to include only those variables that contribute the most meaningful information for the analysis James et al. (2013).

3.3.3 Exploratory Data Analysis (EDA)

In addition to Feature Engineering, Exploratory Data Analysis (EDA) was conducted to gain deeper insights into the dataset. This phase consists visual and statistical examination of the data to uncover underlying structures, detect anomalies, and test hypotheses. Key EDA steps included:

1. Visualizations: Utilizing tools like matplotlib and seaborn for graphical representation of data, including distribution plots and bar charts to understand the frequency and relationship of various variables as can be seen in Figure4 and Figure5.

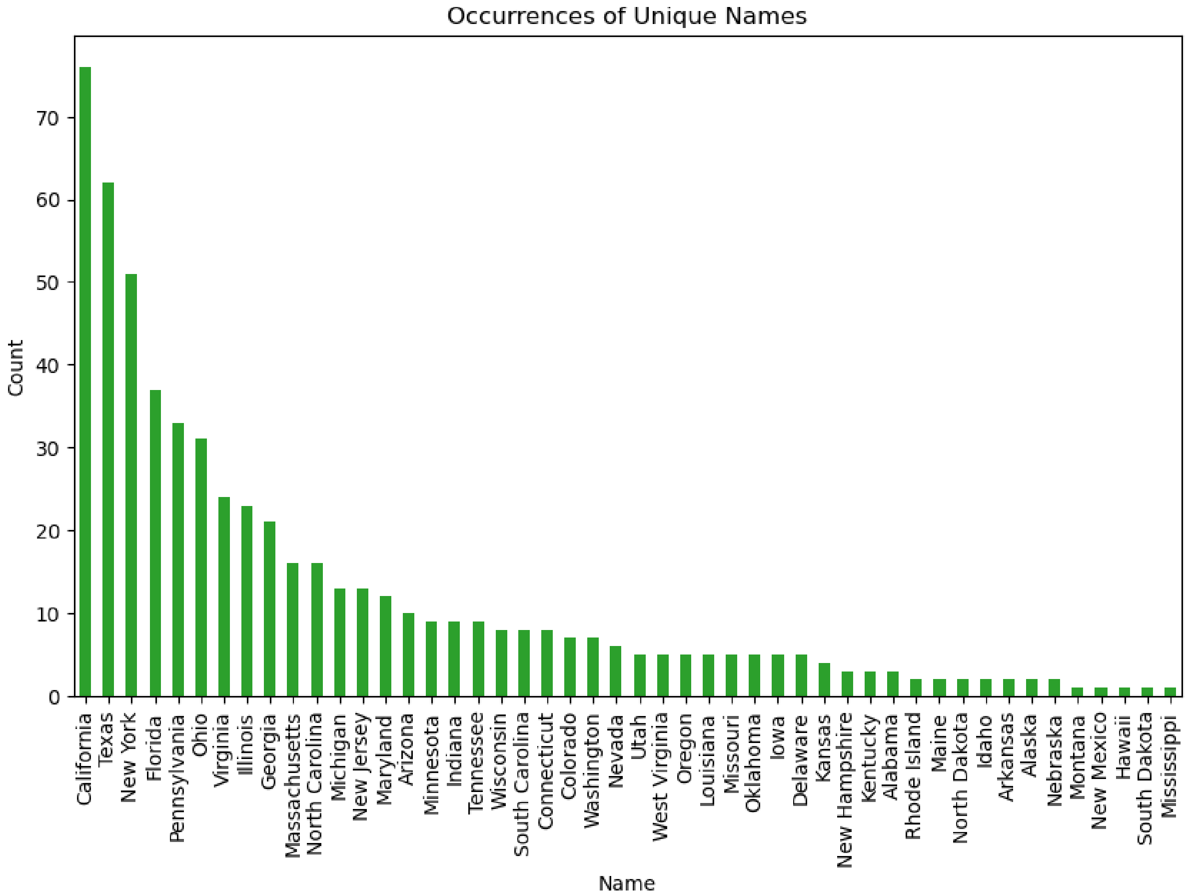

Figure 4: Region wise data

2. Statistical Summaries: Generating descriptive statistics like mean, median, standard deviation to grasp the central tendency and dispersion of data.

3. Identifying Patterns and Anomalies: Careful analysis to spot any patterns, trends, or outliers in the dataset that could influence subsequent analyses.

4. Relationship Exploration: Investigating relationships between different variables using scatter plots and correlation matrices, aiding in understanding the interdependencies among variables.

3.3.4 Data Splitting

I've split my dataset into training and testing subsets, using a common practice of an 80/20 split. This method ensures that most of the data (80%) is used for training the model, allowing it to learn from a large and diverse set of examples. The remaining 20% is reserved for testing, which provides an unbiased evaluation of the model's performance on unseen data. This approach helps in validating the effectiveness of the model and assessing its generalizability to new data.

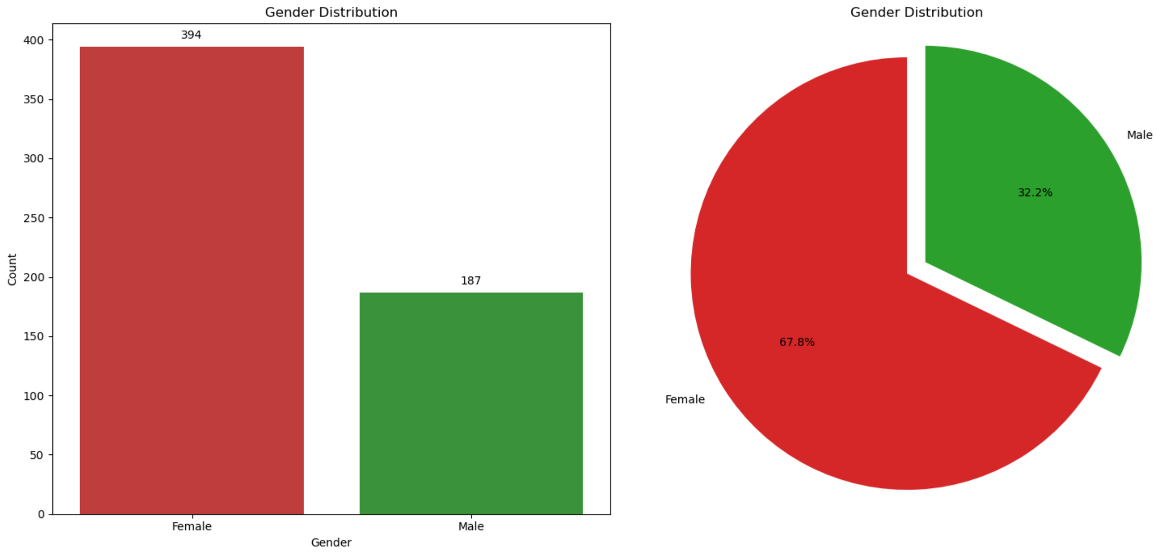

Figure 5: Gender based data

3.4 Model Training

In this section, I've employed various models including Decision Trees, Random Forests, Support Vector Machine (SVMs) and Convolutional neural network (CNNs) to analyse demographic and mental health data. Each model was trained on subsets to predict depression diagnoses. The CNN model utilized data standardized with a scaler, enhancing its performance. After that I've evaluated these models using accuracy scores and classification reports. Furthermore, I've utilized visualizations like confusion matrices, ROC curves, precision-recall curves, and features importance plots to comprehensively understand each model's performance and the factors influencing their predictions. This approach allowed for a detailed comparison of the models' effectiveness in diagnosing depression based on the selected features.

3.5 Model Evaluation and Presentation

The models were thoroughly evaluated for their predictive accuracy in identifying depression. This evaluation involved a detailed analysis using confusion matrices, ROC curves, and precision-recall curves to understand each model's true positive and false positive rates. The feature importance analysis in the Random Forest model provided critical insights into the variables that most significantly influenced the predictions. The comparative accuracy of each model, including Decision Trees, Random Forests, SVMs, and CNNs, was visualized, highlighting their respective strengths and effectiveness in mental health diagnosis. This section distinctively emphasized the practical implications and reliability of the models in a real-world context.

4 Proposed System Model

The System's specification of the study is structured to ensure a comprehensive and rigorous analysis of the relationship between demographic characteristics, mental health symptoms, and the likelihood of depression diagnosis. The specification encompasses both the methodology for data modelling and the techniques for model evaluation.

4.1 Modelling Techniques

The study utilized various statistical and machine learning techniques, including Decision Trees, Random Forests, Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs). These models were selected for their relevance and efficacy in analysing and predicting mental health conditions.

4.1.1 Decision Tree

Decision Trees are a form of supervised learning that models' decisions and their possible consequences, including chance event outcomes, resource costs, and utility. The Decision Tree Classifier was used to unravel the decision-making process behind the prediction of depression. Its tree-like model of decisions and their possible outcomes made it possible to clearly see which features (variables) were most influential. This model is valued for its simplicity, ease of interpretation, and visualization. It's particularly beneficial in breaking down a complex decision-making process into more understandable parts.

4.1.2 Random Forest

Random Forest is an ensemble learning method. It constructs a multitude of decision trees at training time and outputs the class that is the mode of the classes (classification) of the individual trees. The Random Forest Classifier was applied to improve prediction accuracy and combat the overfitting problem that can be prevalent in decision trees. By using multiple decision trees, it reduces the risk of overfitting while still providing interpretable results. Effective for large datasets and capable of handling thousands of input variables without variable deletion. It gives estimates of what variables are important in the classification.

4.1.3 Support Vector Machine (SVM)

SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. SVMs were employed for their efficacy in high-dimensional spaces, such as complex mental health datasets. Its ability to perform linear and non-linear classification (using the kernel trick) makes it adept at handling the nuanced patterns in mental health data. Offers high accuracy and works well with clear margin of separation and with high dimensional space.

4.1.4 Convolutional Neural Network (CNN)

CNNs are deep learning algorithms which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. In the context of mental health data, CNNs were utilized to detect complex patterns and relationships within the data that might not be immediately apparent. These patterns could be indicative of underlying mental health conditions. Reduces the number of parameters to learn and the amount of computation performed in the network.

4.2 Evaluation Metrics

- **Accuracy Measurement:** The primary metric for model evaluation was accuracy, which provided a straightforward measure of overall model performance.

- **Classification Report:** This included precision, recall, and F1-score for each class, offering a more nuanced view of model performance beyond simple accuracy.

- **Confusion Matrix:** This tool helped visualize the model's performance, particularly the types of errors made (false positives and false negatives).

- **ROC Curve and AUC:** The Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) were used to evaluate the models' ability to distinguish between classes.

- **Precision-Recall Curve:** This was particularly important given the imbalanced nature of medical datasets, where the focus is often on the positive (minority) class.

- **Feature Importance Analysis:** For tree-based models like Random Forest, this analysis was crucial to identify which features most strongly influenced the prediction of depression.

- **Model Comparison and Visualization:** The performance of different models was compared and visualized to determine the most effective approach for this specific mental health assessment.

5 Implementation

The development of a machine learning model requires meticulously planned and executed procedures. Ensuring the model's functionality, deployability, and real-world applicability demands careful development and implementation at each stage.

5.1 Environmental Setup

This research primarily leverages Python, a programming language known for its extensive library of tools encompassing modelling, analysis, and visualization. This makes Python an ideal choice for programming, analysing data, and interpreting results. Additionally, various Python libraries such as pandas, scikit-learn, Keras, Matplotlib, and Seaborn were used for data handling, machine learning model development, and data visualization.

5.2 Data Selection

The dataset used in this research is based on a survey conducted to understand the prevalence of depression. The dataset, initially consisting of various demographic and health-related questions, was processed to focus on specific features relevant to the study. The target variable is the diagnosis of depression among participants. The processed dataset includes numerous rows and relevant columns, with the degree of depression diagnosis being the target column.

5.3 Hyperparameter Tuning

Hyperparameter tuning was crucial for optimizing the performance of each model:

- **Decision Tree Classifier:** Parameters like the maximum depth of the tree were tuned.
- **Random Forest Classifier:** Parameters like the number of estimators, maximum depth, and minimum samples split were optimized using techniques like GridSearchCV.
- **Support Vector Machine (SVM):** The linear kernel was chosen based on its effectiveness for the given dataset.
- **Convolutional Neural Network (CNN):** The model included dense layers with different units and used activation functions like ReLU and sigmoid. The model was compiled with a binary cross-entropy loss function and the Adam optimizer.

5.4 Model Evaluation and Visualization

- **Performance Metrics:** The models were evaluated using accuracy scores and classification reports to understand their effectiveness in predicting depression diagnosis.
- **Visualization Techniques:** Confusion matrices, ROC curves, and precision recall curves were plotted for each model to provide a comprehensive view of their performance.
- **Feature Importance Analysis:** For models like the Random Forest, feature importance plots were generated to identify the most significant predictors of depression.

5.5 Comparative Analysis

- **Model Comparison:** A comparative analysis was conducted by visualizing the accuracies of different models, thereby identifying the most effective model.
- **Learning Curves:** The learning curves for models were plotted to assess how the model's performance evolved with the increasing amount of training data.

6 Evaluation

The evaluation phase in the machine learning pipeline for this research plays a crucial role in measuring the effectiveness of the implemented models. This step ensures the models perform as intended, making it possible to gauge their applicability in real-world scenarios. To achieve this, we employed various strategies and techniques:

- **Class Imbalance Problem:**

- **Issue:** In mental health datasets, particularly those concerning depression diagnosis, there is often a class imbalance problem. This imbalance can significantly impact model learning, leading to biases where models overly Favor the majority class and neglect the minority class He and Garcia (2009).
- **Solution:** To mitigate this, we implemented strategies like oversampling of the minority class and class weighting. These techniques help in balancing the distribution of classes, ensuring that the models do not become biased towards the more prevalent class and can learn effectively from both classes.

- **Train and Test Split:**
  - **Approach:** For evaluating the performance of the models and ensuring they generalize effectively to new, unseen data, the dataset was partitioned into training and testing subsets. Utilizing the sklearn package, the split was executed with an 80:20 ratio. This approach allocated 80% of the data for model training, while reserving 20% for testing purposes James et al. (2013).
  - **Rationale:** This split ratio was chosen to provide a substantial amount of data for training the models, ensuring they learn the underlying patterns comprehensively. Meanwhile, the 20% test set offers a sufficient and unbiased sample for evaluating the models' performance. This balance helps in achieving a reliable estimation of how the models would perform in real-world scenarios, effectively balancing between learning capability and evaluation accuracy.

6.1 Case Study 1: Demographic Information

The evaluation of different classification models on the dataset reveals a range of performances, with each model having its unique strengths and weaknesses. The Decision Tree Classifier, which achieved an accuracy of 58% as we can see in Table 1, showed a relatively balanced precision between the two classes (0 and 1), but with a modest recall, especially for class 1. This indicates that while the model is moderately good at identifying the positive class, it tends to miss a significant number of actual positive cases. The Random Forest Classifier, with a slightly higher accuracy of 62%, demonstrated improved performance in identifying the negative class (0), as indicated by its higher precision and recall in this category. However, its ability to correctly identify the positive class (1) was limited, shown by the lower precision and recall for this class. This suggests that while the Random Forest model is better at predicting the negative class, it struggles with the positive class, potentially due to class imbalance or other complexities in the data. The Support Vector Machine (SVM) Classifier also achieved an accuracy of 62%, mirroring the Random Forest's proficiency in predicting the negative class but falling short in the positive class prediction. Its precision for the positive class was notably lower, indicating a higher rate of false positives.

The Convolutional Neural Network (CNN) showcased the highest accuracy among the models at approximately 65.81% as shown in Figure6. It demonstrated a significant strength in identifying the negative class with high precision and recall. However, like the other models, it showed limitations in accurately identifying the positive class as shown in Figure7, as evidenced by its lower recall in this category.

Table 1: CNN - Classification Report

	Precision	Recall	F1-score	Support
Negative	0.67	0.93	0.78	76
Positive	0.55	0.15	0.23	41
Accuracy	0.66			Total: 117
Macro average	0.61	0.54	0.51	Total: 117
Weighted average	0.63	0.66	0.59	Total: 117

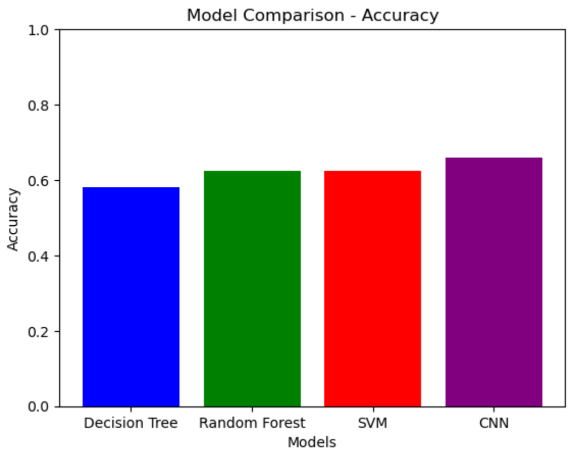

Figure 6: Model Accuracy Case Study 1

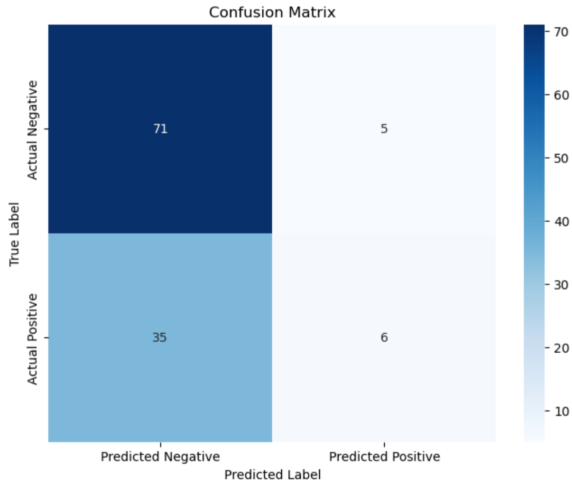

Figure 7: Confusion Matrix CNN - Case Study 1

6.2 Case Study 2: General Health and Mental Health Assessment

The evaluation metrics for the four classifiers exhibit distinct performances, with each model displaying its unique strengths in predicting depression diagnosis. The Decision Tree Classifier, with an accuracy of 59% as we can see in Table 2, shows moderate performance. Its precision and recall for the negative class (0) are reasonably good, indicating effectiveness in identifying true negatives, but it is less effective for the positive class (1), as seen by the lower recall and precision. This suggests a challenge in correctly identifying positive cases without many false positives. The Random Forest Classifier outperforms the Decision Tree with an accuracy of 67%. It demonstrates a notable improvement in predicting the negative class, with higher precision and recall, indicating its strength in correctly identifying negatives and reducing false positives. However, its performance on the positive class is still limited, showing a moderate precision but a low recall, which means it struggles to identify a significant portion of actual positive cases. The Support Vector Machine (SVM) Classifier shows an accuracy of 65%, but its performance is skewed. It has high precision for the negative class but zero precision and recall for the positive class, indicating it fails to correctly identify any true positive cases. This could be a result of model overfitting to the negative class or an inherent limitation in the SVM model's capacity to handle this specific dataset's complexity. The Convolutional Neural Network (CNN) demonstrates the best overall performance with an accuracy of 72% as shown in Figure8. It shows a balanced and superior performance in both precision and recall across both classes, especially in the negative class. While its recall for the positive class is lower, it still maintains a respectable precision, suggesting it can identify positive cases more accurately than the other models, albeit missing some true positives as shown in Figure9. In summary, while the Random Forest and SVM classifiers show a strong bias towards predicting the negative class, the CNN provides a more balanced approach, showing proficiency in both classes. The Decision Tree, though relatively weaker, still offers some balance between the two classes. Each model has its merits, but the CNN's ability to better handle the complexities of the dataset makes it the most effective for this task.

Table 2: CNN - Classification Report

	Precision	Recall	F1-score	Support
Negative	0.71	0.93	0.82	76
Positive	0.79	0.27	0.40	41
Accuracy	0.72			Total: 117
Macro average	0.75	0.61	0.61	Total: 117
Weighted average	0.74	0.72	0.67	Total: 117

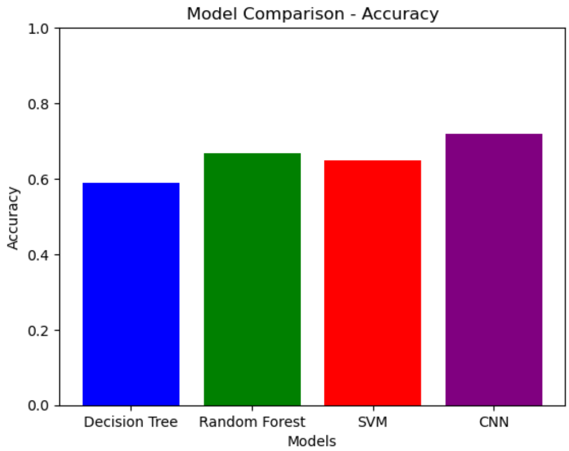

Figure 8: Model Accuracy Case Study 2

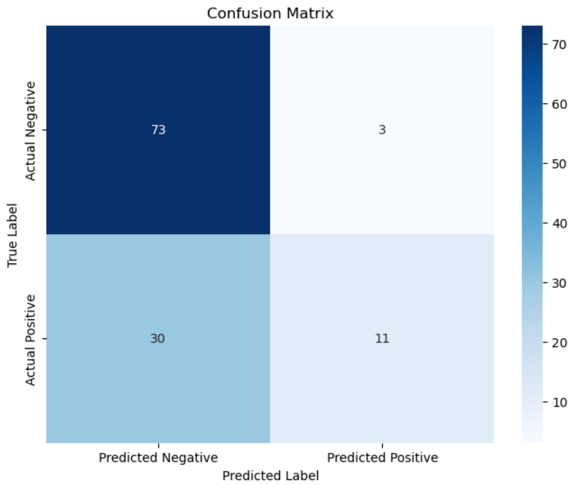

Figure 9: Confusion Matrix CNN - Case Study 2

6.3 Case Study 3: Medical and Health Conditions Analysis

In this analysis, we observe varied performances across four different classifiers - Decision Tree, Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) - when applied to a dataset for predicting depression diagnoses. The Decision Tree Classifier shows a balanced performance with an accuracy of 72%. It achieves a precision of 0.79 and a recall of 0.78 for the negative class as we can see in Table 3, indicating its effectiveness in identifying true negatives. For the positive class, it maintains a moderate precision and recall, suggesting it can accurately identify positive cases, though with some room for improvement. The Random Forest Classifier emerges as the top performer with an accuracy of 77% as shown in Figure10. It demonstrates strong precision (0.79) and recall (0.88) for the negative class and respectable scores for the positive class. This indicates its efficiency in correctly identifying both negative and positive cases, albeit slightly better in handling negative ones as we can see in Figure11. The SVM Classifier, with an accuracy of 68%, shows a significant bias towards the negative class. While it achieves high precision in identifying the negative class, its recall for the positive class is notably low. This imbalance suggests that the SVM model is more inclined to predict negatives correctly while struggling to identify a significant portion of the actual positive cases. The CNN model also shows an accuracy of 68%. It performs well in identifying the negative class with a high recall but is less effective in accurately identifying the positive class, as indicated by its lower recall for positives. This suggests that while the model is good at avoiding false positives, it tends to miss several actual positive cases. In conclusion, the Random Forest Classifier demonstrates the most balanced and effective performance among the four models, followed by the Decision Tree, which also shows a fair balance between precision and recall for both classes. The SVM and CNN, while effective in certain aspects, display a notable imbalance, particularly in their ability to identify positive cases accurately.

Table 3: Random Forest Classifier - Classification Report

	Precision	Recall	F1-score	Support
Negative	0.79	0.88	0.83	76
Positive	0.72	0.56	0.63	41
Accuracy		0.77		Total: 117
Macro average	0.75	0.72	0.73	Total: 117
Weighted average	0.76	0.77	0.76	Total: 117

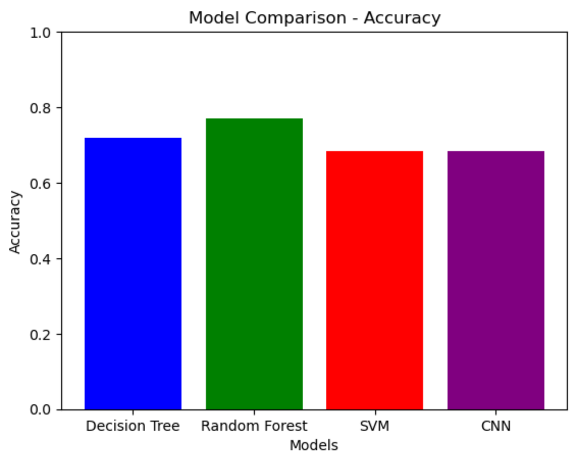

Figure 10: Model Accuracy Case Study 3

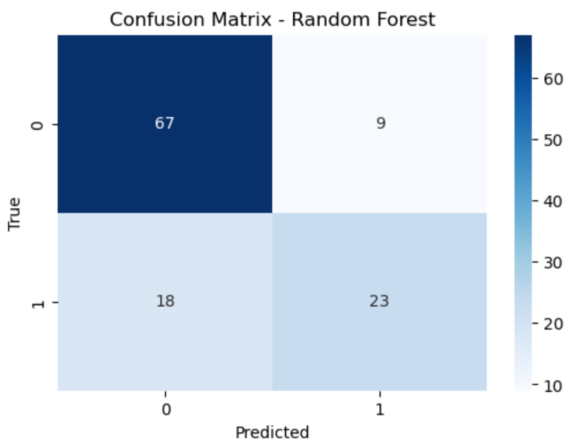

Figure 11: Confusion Matrix Random Forest - Case Study 3

6.4 Case Study 4: Depression Diagnosis and Other Health Related Variables

In this analysis of four classifiers - Decision Tree, Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) - notable distinctions are observed in their performance metrics when predicting depression diagnoses. The Decision Tree

Classifier achieves a high accuracy of 87%. It demonstrates commendable precision and recall for both negative (0.89 precision and 0.92 recall) and positive classes (0.84 precision and 0.78 recall) as we can see in Table 4, indicating its robustness in classifying both classes with relatively high reliability. The Random Forest Classifier outperforms with an impressive accuracy of 91% as shown in Figure12. It shows very high precision and recall scores for both classes, nearly equal for the negative (0.93 for both) and positive classes (0.88 for both). This suggests an excellent balance in correctly identifying true negatives and positives, underlining its strong predictive capability as we can see in Figure13. The SVM Classifier also registers an accuracy of 91%, with a slightly different pattern. It excels in precision for the negative class (0.99) but has a lower recall (0.87), indicating a strong tendency to correctly label negative cases but missing some positive cases. Conversely, for the positive class, it shows high recall (0.98) but lower precision (0.80), suggesting it captures most positive cases but with some false positives. The CNN model, with an accuracy of 89%, exhibits balanced performance. It achieves high precision and recall for the negative class (0.92 and 0.91 respectively) and slightly lower yet commendable scores for the positive class (0.83 precision and 0.85 recall). This indicates its effective capability in distinguishing both classes with a slight leaning towards the negative class. Overall, the Random Forest and SVM classifiers showcase the highest accuracy and a balanced approach towards both classes, with the Random Forest slightly edging out in overall balanced performance. The Decision Tree also displays strong performance, particularly in handling negative cases, while the CNN provides a balanced approach with a slight inclination towards accurately predicting negative cases.

Table 4: Random Forest Classifier - Classification Report

	Precision	Recall	F1-score	Support
Negative	0.93	0.93	0.93	76
Positive	0.88	0.88	0.88	41
Accuracy	0.91			Total: 117
Macro average	0.91	0.91	0.91	Total: 117
Weighted average	0.91	0.91	0.91	Total: 117

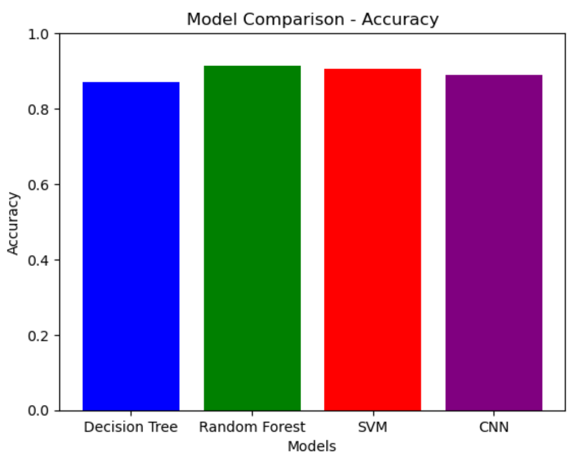

Figure 12: Model Accuracy Case Study 4

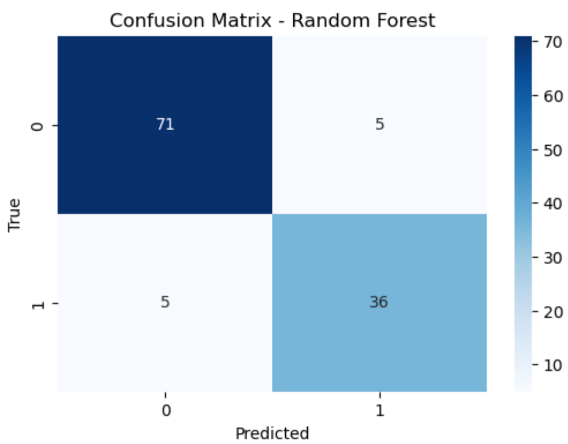

Figure 13: Confusion Matrix Random Forest - Case Study 4

6.5 Discussion

In this research, we utilized four machine learning models to predict depression diagnoses, namely Decision Tree, Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN). Each model offered unique insights into the complexities of mental health analysis. The Decision Tree Classifier achieved an 87% accuracy but demonstrated a tendency to overfit, especially for the positive class. This aligns with its known limitations in handling complex datasets. The Random Forest Classifier stood out with a 91% accuracy, effectively reducing overfitting through its ensemble approach, but faced challenges in computational intensity and interpretability. The SVM, also scoring a 91% accuracy, excelled in high-dimensional spaces, though it showed limitations in recalling the positive class, suggesting a need for refined hyperparameter tuning. The CNN, with an 89% accuracy, balanced performance well but required extensive data and careful tuning to prevent overfitting. Improvements for each model include pruning and boosting for the Decision Tree, feature selection for the Random Forest, varied kernel types for the SVM, and meticulous tuning for the CNN. These results contribute to the understanding of machine learning in mental health prediction, emphasizing the importance of selecting appropriate models based on project specifics and balancing accuracy with interpretability and computational demands.

7 Conclusion and Future Work

This study rigorously evaluated four machine learning models: Decision Tree, Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN), in the context of predicting depression diagnoses. The investigation revealed distinct performance characteristics for each model, providing critical insights for mental health analysis. The Decision Tree showcased high accuracy but was limited by overfitting, suggesting a need for improved modelling techniques. Random Forest emerged as a robust performer, achieving high accuracy but facing challenges in interpretability and computational intensity. The SVM displayed potential with high-dimensional data but required meticulous parameter tuning for balanced precision and recall. Meanwhile, the CNN demonstrated impressive pattern recognition capabilities, although it required extensive datasets and faced complexity issues.

These findings align with existing literature, emphasizing the strengths and limitations of these models in handling complex mental health data. The research contributes significantly to the application of machine learning in healthcare, particularly in the nuanced area of mental health diagnosis and prediction.

Future research directions include advancing model performance through innovative techniques and feature engineering, expanding and diversifying datasets for broader applicability, and enhancing model interpretability, especially for complex models like Random Forest and CNN. Implementing these models in real-world clinical settings and integrating them into healthcare systems are crucial steps for practical application. Additionally, addressing ethical and privacy considerations in AI and healthcare is paramount. This study forms a foundational step towards developing more accurate, efficient, and ethical AI tools for mental health diagnosis, offering substantial contributions to AI in healthcare.

References

Abirami, S. (2022). Psychological effects and changes in the lifestyle of young individuals due to exposure to digital media, *International Journal of Clinicopathological Correlation* **6**(2): 4–8.

Andreassen, C. S. (2015). Online social network site addiction: A comprehensive review, *Current addiction reports* **2**(2): 175–184.

Boy, F. (2023). Google trends™ dynamics for the guidance of open-source intelligence: Augmentation of social-media and survey surveillance of population mental health, *2023 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 1–9.

Bryman, A. (2016). *Social research methods*, Oxford university press.

Chadha, A., Gupta, A. and Kumar, Y. (2022). Suicidal ideation detection on social media: A machine learning approach, *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pp. 685–688.

Chukwuere, G. and Chukwuere, J. (2023). The difficulties posed by digital technology: Understanding the psychological consequences of social media use on young adults' body image and self-esteem, *Academic Journal of Interdisciplinary Studies* **12**: 379.

De Choudhury, M. and De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity, *Proceedings of the international AAAI conference on web and social media*, Vol. 8, pp. 71–80.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data, *IEEE Transactions on Knowledge and Data Engineering* **21**(9): 1263–1284.

Illahi, M., Siddiqui, I. F., Ali, Q. and Alvi, F. A. (2022). Ensemble machine learning approach for stress detection in social media texts, *Quaid-E-Awam University Research Journal of Engineering, Science & Technology, Nawabshah.* **20**(02): 123–128.

James, G., Witten, D., Hastie, T., Tibshirani, R. et al. (2013). *An introduction to statistical learning*, Vol. 112, Springer.

Khasnis, N. S., Sen, S. and Khasnis, S. S. (2022). A machine learning approach for sentiment analysis to nurture mental health amidst covid-19, *Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence*, DSMLAI '21', Association for Computing Machinery, New York, NY, USA, p. 284–289.
URL: *https://doi.org/10.1145/3484824.3484877*

Kim, J., Lee, D., Park, E. et al. (2021). Machine learning for mental health in social media: Bibliometric study, *Journal of Medical Internet Research* **23**(3): e24870.

Kotsiantis, S. B., Zaharakis, I., Pintelas, P. et al. (2007). Supervised machine learning: A review of classification techniques, *Emerging artificial intelligence applications in computer engineering* **160**(1): 3–24.

Luca, L., Ciubara, A. B., Antohe, M. E., Peterson, I. and Ciubara, A. (n.d.). Social media addiction in adolescents and young adults-psychoeducational aspects.

Osborne, J. (2013). Is data cleaning and the testing of assumptions relevant in the 21st century?, *Frontiers in Psychology* **4**.
URL: *https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00370*

Priya, P. L. and Prakash, R. V. (2023). A broad survey on detection of depression in societal platforms using machine learning model for the public health care system, *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 1643–1649.

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M. and Langer, E. J. (2017). Forecasting the onset and course of mental illness with twitter data, *Scientific reports* **7**(1): 13006.

Satyaninrum, I. R., Rumondor, P., Kurniawati, H., Aziz, A. M. et al. (2023). Promoting mental health in the digital age: Exploring the effects of social media use on psyhcological well-being, *West Science Interdisciplinary Studies* **1**(06): 248–256.

Tadesse, M. M., Lin, H., Xu, B. and Yang, L. (2019). Detection of depression-related posts in reddit social media forum, *IEEE Access* **7**: 44883–44893.

Trochim, W. M. and Donnelly, J. P. (2001). *Research methods knowledge base*, Vol. 2, Atomic Dog Pub. Macmillan Publishing Company, New York.

Twenge, J. M. and Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study, *Preventive medicine reports* **12**: 271–283.

Twigg, L., Duncan, C. and Weich, S. (2020). Is social media use associated with children's well-being? results from the uk household longitudinal study, *Journal of Adolescence* **80**: 73–83.