Deep Learning and Natural Language Processing for Suicidal Ideation Using Instagram Posts

MSc Research Project
Data Analytics

Vrushali Darade
Student ID: x21123764

School of Computing
National College of Ireland

Supervisor: Athanasios Staikopoulos

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vrushali Darade
Student ID:	x21123764
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Athanasios Staikopoulos
Submission Due Date:	15/12/2022
Project Title:	Deep Learning and Natural Language Processing for Suicidal Ideation Using Instagram Posts
Word Count:	7034
Page Count:	19

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:	Vrushali Bhanudas Darade
Date:	31st January 2023

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):	

Deep Learning and Natural Language Processing for Suicidal Ideation Using Instagram Posts

Vrushali Darade
x21123764

Abstract

Suicide is a global problem [1]. Every year, more than 8 million people die due to suicide, which means one person commits suicide every 40 seconds [2]. Identification of suicide ideation is the first step in saving a life. Traditionally, suicidal tendencies are identified using a questionnaire dataset, which the victims mostly provide. As data is provided by victims, its authenticity is dependent on various aspects, such as the purpose of collection, time of collection, who is collecting, whether is it anonymous and so on. Hence it will be difficult to exactly understand the emotions of the victims. The research aim of this project is to identify suicidal tendencies by analyzing Instagram images and captions, especially for the age group between 15 to 30. Analysing Instagram posts using Deep learning concepts such as image and text processing will help in understanding the emotion of the person who posted it, such as whether the person is happy, sad, depressed, or lonely. This information will be beneficial for consolers, psychologists, school and college administrators, and others who will provide appropriate treatment to the suffering person. To achieve the research goal, the Deep Learning algorithm VGG16 will be applied to image processing, as it has shown the best accuracy results in categorisation and Long-short term memory, along with CNN for text processing. The evaluation demonstrates high accuracy with VGG16, LSTM, and hybrid CNN+LSTM resulting in an excellent accuracy of 87% and 98% respectively. This demonstrates that images and captions posted on Instagram social media can be successfully processed for the identification of suicidal ideation. As the model produced excellent results, it can be extended to analyse various social media platforms such as Twitter, Facebook, Snapchat, and so on.

Keywords Deep learning, Natural Language processing, VGG16, LSTM, Instagram Post, Suicidal Ideation

1 Introduction

Suicide is the third leading cause of death, especially among teenagers [3]. At the same time, it is not an exaggeration to say that teenagers' new lifestyle is social media as they post everything on social media platforms. Such as where they go, what they eat, their thoughts, pictures, and captions. This post isn't just for fun; users usually share their

[1] https://www.who.int/health-topics/suicide#tab̄tab_1
[2] https://worldpopulationreview.com/country-rankings/suicide-rate-by-country
[3] https://www.who.int/news-room/fact-sheets/detail/suicide

emotions with a larger audience through pictures and captions. Several studies have found that people, particularly young people, who are contemplating or planning suicide use social media (Fahey et al.; n.d.). This data contains a large number of hidden aspects of data analysis in terms of location details, objects, text, video, audio and image categorisation and classification. This research will examine hidden information in order to detect suicidal tendencies. Suicide is a critical issue that has prompted many researchers in most fields; however, in terms of machine learning, it is restricted to datasets available in the form of questions and answers as it is provided by victims. The dataset and methodologies used in this research paper make it unique. It will use a combination of Instagram image and caption data for processing with deep learning methodologies. This helps in categorizing the images and captions as suicidal or non-suicidal behaviour. VGG 16 will be used for image classification (Ferreira et al.; n.d.) (Jeong and Cho; 2017), as it has proven to be one of the best. VGG16 is a popular image categorization algorithm that has achieved 92.7% accuracy in some types of classification [4]. Natural language processing is used for caption categorisation, which categorizes words that suggest signs of depression, sadness, or loneliness, such as "feeling low," "sad," "fear", "angry", "anxiety" and so on. The categorization of Instagram posts will benefit the early detection of depression or suicidal thoughts. Once identified, the victim can receive special attention and treatment, which can help to prevent suicide attempts. This classification will be useful for psychologists, counsellors, colleges, schools, and offices because it will provide a basic understanding of victims and non-victims.
The question of the research paper is **"Deep learning and natural language processing: How successful are they in classifying Instagram posts to identify suicidal tendencies?"**. With the advent of technology and the internet, various social media platforms have emerged, each with its own distinct identity. Few have only text posts, such as Quora and Reddit, while others have a mix of pictures and text, such as Facebook, Twitter, and Instagram. This paper will concentrate on processing the Instagram dataset. As ideation of suicidal tendencies is successfully classified in this paper with VGG 16 and NLP, this can be applied to social media platforms with text and image data as a combination.
Section 2 of the paper begins with a literature review, which provides information about the history of similar research as well as the methodologies used for image and social media data classification and its limitations. Section 3 describes the methodologies, and Section 4 describes the design approach used to achieve the goal. Section 5 goes into detail about how models are built and implemented for this specific goal. The following section is evaluation, which assists in clearly stating the achievement of the goal. Finally, section 7 concludes the research with a conclusion and future research scope.

2 Related Work

Suicidal tendency identification has become the most important research area because it is a worldwide concern (*Suicide*; 2021). Suicidal tendency ideation has been researched in domains, such as psychology, academics, medicine, and machine learning. All the research are limited in terms of data and traditional machine-learning approach. This section will provide a summary of the work done in terms of machine learning and data analysis to detect suicidal ideation.

[4]https://paperswithcode.com/sota/image-classification-on-imagenet

2.1 Suicidal Ideation with traditional machine learning

The research (Ji et al.; 2020) conducted an extensive exploration of the two major methods for Suicidal Ideation Detection (SID), namely, machine learning and clinical methods. They thoroughly reviewed and classified the methods and data domain for SID, as illustrated in Figure 1. The conclusion clearly explains the limitations of previous research in terms of data efficiency, quality of the data and lack of understanding of the intention behind the suicide. They provided suggestions for future work in terms of applying deep-learning methodologies to identify suicidal tendencies. In research (Jordan et al.; n.d.) suicide ideation is treated as a dependent variable when determining the variables self-esteem, depression, hopelessness, worrying and sleep disturbance. The research data is cross-sectional data collected via a questionnaire from 32 primary care practices. Two methods were used: pattern recognition and identification of three blocks of input variables that will be used for classification methods. For prediction, they used classification trees, Support Vector Machines, and neural networks. They have successfully shared the symptoms that lead to suicidal ideation. They have also provided scope for future work by including other factors such as past suicide attempts, functional impairment, stressful life events or physical illness that can lead to suicidal tendencies. The research (van Mens et al.; n.d.) uses six different machine-learning techniques to identify suicidal tendencies in young adults. The study's main goal was to investigate the role of machine learning when applied to longitudinal data with 15 different psychological risk and protective factors. They develop an Integrated Motivational-volitional (IVM) model to explain the complex interplay of variables that leads to suicidal behaviour and attempts. Using the Confusion matrix and AUC, they conclude that Random Forest is best for predicting suicidal ideation and Gradient Boosting is best for predicting suicidal attempts, with AUCs of 0.83 and 0.80, respectively. Another research (Macalli et al.; n.d.) on a longitudinal population-based study for identifying suicidal thoughts among college students. They created a risk algorithm that predicts the main contributing factors to suicidal thoughts and behaviour. The Random Forest model is implemented for statistical analysis with aggregation of decision trees with data as 3:1 for implementation and testing. The analysis is carried out in two steps: the first includes 70 predictors, and the second includes childhood adversity. Out-of-bag, AUC, sensitivity, and positive prediction are used to evaluate the model. The first evaluation results show that girls and boys have the same predictor values contributing to suicidal thoughts, which are baseline thoughts, self-esteem, trait

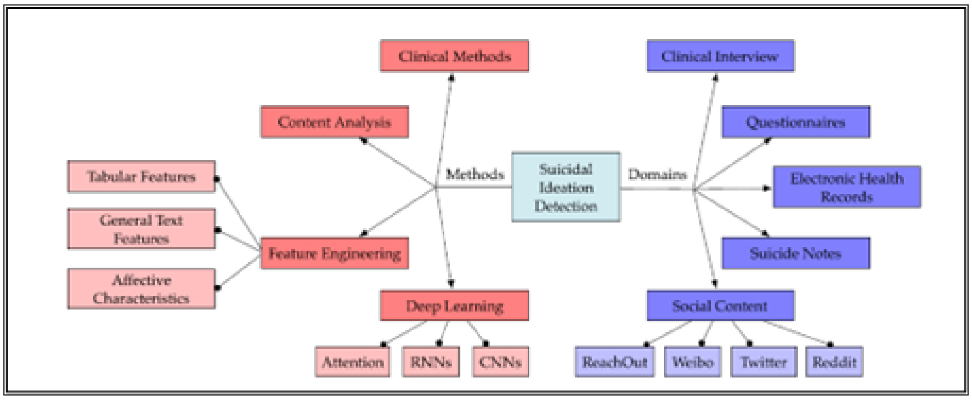

Figure 1: Review of Machine Learning methods and Application

anxiety, and depression. In the second evaluation, both boys and girls had predictors such as suicidal thoughts, self-esteem, and trait anxiety, but one contributor differed between the sexes: depression symptoms and perceived stress, respectively. The research (Chadha and Kaushik; n.d.) used tweeter data to apply machine learning algorithms to tweets to assess suicidal risk. Their focus and differentiating factor were keywords for identifying and forecasting suicidal behaviour. Data is collected using the Tweeter API, and tweets are chosen based on 34 predetermined keywords. Data is pre-processed and converted into probabilistic values so that machine learning and ensemble learning algorithms can use it. They conclude that Logistic regression provides better accuracy with a lower recall rate after experimenting with various algorithms in both machine learning and ensemble learning. They proposed using deep learning on datasets such as clinical data, written blogs, and shared photos and videos from Snapchat, Instagram, and Facebook in the future. In a similar vein to the previous two papers, research (Rajesh Kumar et al.; n.d.) used tweeter data for prediction in three stages. To begin, they created the dataset for pattern extraction, then implemented sentimental analysis and classification methods for better prediction, and finally validate the proposed methods. Vader sentimental analysis is used exclusively for data extraction. Tweets that have been pre-processed are separated by 62 predefined keywords and an n-gram. The formula is set up to categorize tweets as suicide-positive, suicide-negative, or suicide-neutral. They concluded that the Random Forest algorithm outperforms the others.

2.2 CNN/NLP for Suicidal Ideation

DeprNet, a unique architecture based on CNN, is developed for the classification of depressed and normal people using an electroencephalogram (EEG) dataset collected from 33 subjects as well as a Patient Health Questionnaire 9 score in research (Seal et al.; n.d.). They achieved a high accuracy of 0.914 by overcoming previous work's limitations, such as manual and automatic techniques for investigating depression condition. Following network training, data points are forwarded to the network or obtained as an activation map. 19 response vector values are used to identify the channels that suggest depression. Using the Heatmap visualization, a formula is created to assist in determining the response variable values for the right and left electrodes. The findings of this study are more promising because they achieved several goals, including identifying the presence of depression, having a high accuracy of 0.9937 for record-wise split, and believing that depression affects both hemispheres of the brain differently. Using a combination of CNN and machine learning techniques to build a suicidal ideation detection system using publicly available data is researched in (Aldhyani et al.; n.d.). They used the Reddit dataset, which was pre-processed for text filtering using the Natural Language Toolkit (NLTK). Following that, TF-IDF and Word2Vec were used to vectorize words and sentences for suicidal and non-suicidal classification. Classification is carried out using XGBoost and a hybrid CNN-bidirectional LSTM DL algorithm. They added the BiLSTM layer to handle the situation of dealing with past data information and advancing the required one by having two hidden layers. Using a confusion matrix, they concluded that CNN-BiLSTM outperforms XGBoost. Furthermore, the LIWC lexicon aids in highlighting the higher score for the suicidal post by utilizing various symptoms such as depression, decreased attention, mind-thinking, and so on. Different from the traditional machine learning methods (Macalli et al.; n.d.) (Chadha and Kaushik; n.d.) (Rajesh Kumar et al.; n.d.), in research (Roy et al.; n.d.), they developed an algorithm called "Suicide Artificial Intelligence Prediction Heuristic (SAIPH)" to achieve two major goals using tweeter dataset. The first is to classify Suicidal ideation cases from controls, and the second is to generate a method to assess when the individual is most likely to commit suicide. They used a combination of neural networks and the random forest method to accomplish this. As a large amount of data is collected from tweeters, various scanning techniques are used to identify relevant tweets. The text tweets are converted into a score between 0 and 1 using a neural network, and then the cases are predicted using random forest on neural network-derived data. The user's timeline is evaluated in the second phase to identify the pattern. Various evaluation models are used across different scores, such as sensitivity, specificity, and positive prediction values. They also performed a temporary analysis for reword data analysis by restricting the tweeter data to 120 days, demonstrating that the peak occurs within 5 to 120 days using logistic regression. The research (Cusick et al.; n.d.) uses a different dataset with a similar approach. They analysed unstructured clinical notes using a combination of statistical machine learning models and a deep learning model called text classification convolution neural network. Annotation of training and validation notes was done using a rule-based NLP approach, while testing was done manually. Cohen's kappa statistic is used to perform the evaluation. They used NegEx as a foundation for developing a rule-based NLP algorithm based on two lexicons, one with suicide keys and the other without. The evaluation is done on both the document and the patient level. McNemar's test of homogeneity is then used to validate the prediction. Finally, they conducted the analysis to determine the accuracy of the implemented method. Classifiers are implemented with the vector score for clinical text using bag-of-words, bi-gram, and TF-IDF for statical NLP, and word2vec for CNN. The results show that CNN outperforms the other models without the need for manually annotated datasets, which saves time and money. The research (Weng et al.; n.d.) conducted ground-breaking research by implementing an Autoencoder and machine learning model for suicidal ideation. The dataset's uniqueness lies in the brain images, data preparation, and model implementation. The 3 T MRI system is used to collect brain image data, followed by the use of a 3D autoencoder for feature extraction. Supervised machine learning and XGB is used to distinguish between HCs and depression patients with rest. One challenge of data imbalance is overcome by dividing non-ideation subjects in half and then forming train, validation, and test sets with a 4:1:1 ratio. They claim that the results are the best because they were able to successfully find the pattern of structure across multiple brain locations classifying suicidal and non-suicidal ideation. Another research (McMullen et al.; n.d.) implemented the Random Forest, logistic regression, and XGBoost machine learning algorithms with the combination of bootstrapping, as in previous research work, but they used z-score for comparing AUROC matrices. They used an equation to compare the AUROC curves using variables such as area and standard error for models 1 and 2, while r represents the correlation between A1 and A2. According to the AUROC curve analysis, SCS has a high predictability of near-term suicide risk. This research (Malhotra and Jindal; n.d.) provided in-depth knowledge about various techniques for detecting suicidal tendencies, mental illness, and depression. They reviewed 96 research papers covering the state of the art in terms of technologies, datasets, and evaluation metrics in this area, which provides an idea about the domain and why more research was required. They observe that, in the last four to five years, deep learning has grown in popularity for analyzing text, images, sentiments and emotions. The study also assisted in understanding the research topic chosen and the trend of publication on the subject. Aside from the commonly used evaluation metrics, they discovered others, such as ERDE (Early Risk Detection Error), Average Hit Rate (AHR), and Average Difference in Overall Depression Level (ADODL). Furthermore, their research clearly states that the combined dataset of text and image has not yet been thoroughly researched, as well as social media platforms such as Facebook, Instagram, and others, as shown in the Figure 2 below.

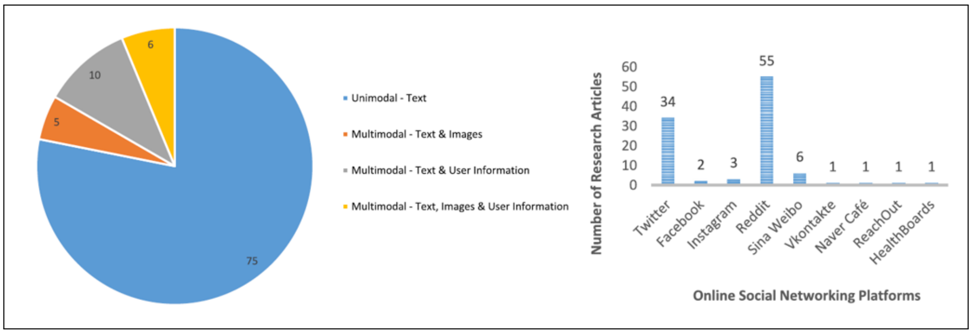

Figure 2: Comparison of implementation techniques and social media dataset usage

They concluded the work by suggesting future implementation steps describe in Figure 3, along with various techniques for each one.

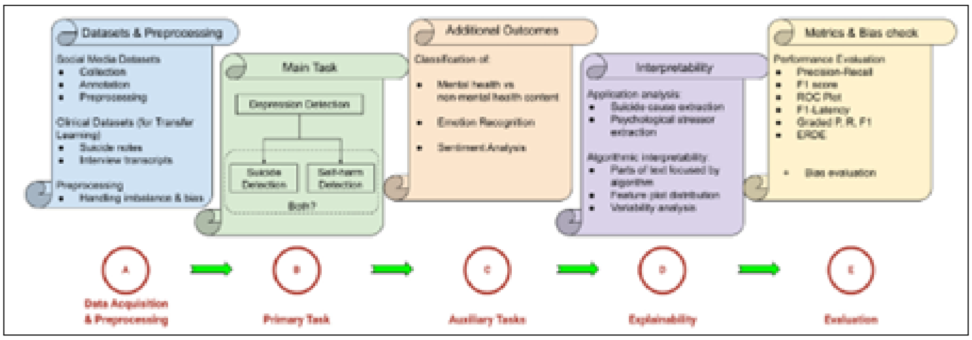

Figure 3: Suggestions for CNN implementation steps

Unlike the others, the research (Zhang et al.; n.d.) proposed a transformed-based model for extracting text data using a transformer encoder and BiLSTM. The dataset is divided into three subsets: suicide notes, last statements, and neutral posts. LIWC is used to analyse data in terms of words, topics, and linguistic features. Following data analysis, the TransformerRNN model is implemented with five components: GloVe input embedding, transformer encoder, BiLSTM, max-pooling, and finally classification. They used Precision, recall, and F1-Score for evaluation, as did others. Additionally, weighted average metrics were used to display performance. In comparison to traditional machine learning models, the model's performance with F1-score is the best, with 93.3%. Based on this result, they believe it will be useful for identifying suicidal notes. In research (Haque et al.; n.d.) use the transformer model as they believe in language dominance. Beginning with the BiLSTM, they expanded the implementation to include the BERT, ALBERT, ROBERTa, and XLNET transformer models. A three-tier architecture is used, with a data layer, an embedding layer, and a classification layer. Using the same evaluation methods as others, the researchers concluded that ROBERTa outperformed the other models. The research (?) used a combination of Word2vec and CNN for sentiment analysis, inspired by many deep learning methodologies. They claim to be the first to use the 7-layer architecture for sentiment analysis. To improve the performance of the implemented model, PReLU, Normalization, and Dropout are used. The model's performance is evaluated using a dataset of movie reviews. Word2vec is used to assign vector values, which are then fed into the CNN, which has three convolution layers and three pooling layers. They claim that by varying train and test evolution, this method of model implementation outperforms classification algorithms. Furthermore, when compared to RNN and Matrix-vector RNN, pre-trained and fine-tuned CNN performs better. In research (Trotzek et al.; n.d.), they evaluated convolution neural networks based on different word embeddings using posts and comments from reddit.com. The classification is compared to linguistic metadata at the user level. Text-based metadata features are extracted from each message's text and title. They used the GloVe and fastText models to implement word embedding. The word embedding is validated by training it on the Reddit comment dataset. Rectified Linear Units (CReLU) were also concatenated for layer activation. They concluded that the correct prediction of a small number of positive samples has an effect on the EDRDE score and that the best results can be obtained by minimizing false positives.

2.3 Image and text processing for social media

In research Dadiz and Marcos (n.d.), a classification model is implemented by processing images with various facial expressions. The image dataset was created by cropping a video's facial expressions and extracting a local binary pattern for each frame. The Viola and Jones method (Viola and Jones; n.d.) is used for key feature extraction because their focus was on extracting frames with full faces. Images were trained using an accumulated histogram, with Obadiah's expression labelled as depressed and the rest of the poppy, Spike, and prudence labelled as non-depressed. They conclude a) face analysis focuses on the eyes, nose, mouth, cheeks and head temple. b) 81% accuracy c) PCA of ULBP features demonstrated improved accuracy. The research (Amanatidis et al.; n.d.) uses image classification transfer learning and text processing to identify COVID-19 vaccine stakeholders via an Instagram post image and text processing. They downloaded the images, videos, and captions from Instagram using an open-source tool. After tuning the images according to the model requirements, the VGG16, Inception V3, and ResNet50 are implemented. WorldCloud is used to get a quick inside look at the post. They were able to determine which of the three companies was the most active in image and text processing. The research (Xian et al.; n.d.) conducted an Instagram analysis to distinguish between self-injury and non-self-injury posts. A content crawler is created to collect data with specific hashtags and save it to the cloud. The two major challenges with image datasets are the variety of objects available on post and the cost and time required to process large, labelled data. They overcame this by implementing Adversarial Complementary Learning (ACoL) methods. ACoL results are compared to VGG16 and ResNet50. The evaluation is carried out for quantitative analysis to show the trend of the NSSI hashtag and another for determining image classification accuracy. The quantitative analysis of the images, captions and 68 different hashtags revealed a significant increase in the hashtags. When it came to image classification, the ACoL model outperformed the VGG16 and ResNet50.

3 Methodology

To ensure the success of any research project, a clear strategy should be identified and implemented. This research project follows the five-step process outlined in the Figure 4 below for implementation, referring to the CRISM-DM process. Each step is followed separately for image and caption processing and analysis.

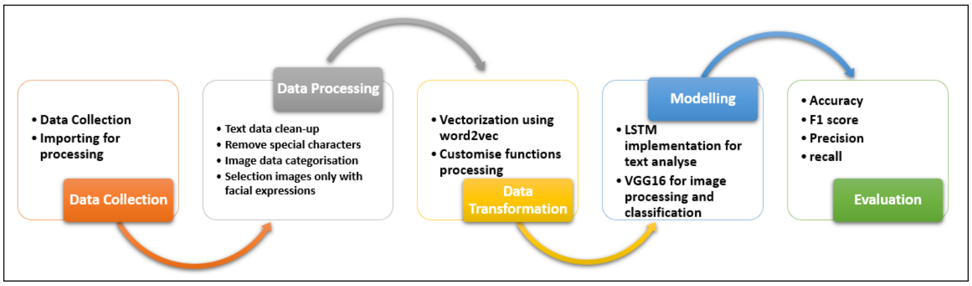

Figure 4: Methodology for implementation

Data Collection: Kaggle is used to collect the dataset for the model implementation [5] [6]. The dataset consists of Instagram images as well as captions posted on Instagram. The captions are available in.csv format. Another dataset contains, face images categorised into happy and sad. The data is uploaded to Google Drive for further processing and model implementation, using the functionality of the Python package.

Data Processing: Text and image data processing are handled independently. The caption dataset cleanup is performed by removing special characters, emojis, hashtags, and contraction words. After cleaning up the text data, the final dataset contains 11545 unique rows for model implementation. The image dataset, on the other hand, was cleaned up by removing images that did not include a face and were not in .png format. The images are then divided into four subcategories based on their facial expressions: angry, sad, fearful, and neutral.

Data Transformation: Data transformation is required because it enables achieving accurate results, maintaining data quality, and reducing processing time by filtering out garbage data. Various techniques for text data transformation are used in this study such as contraction, hashtag, and emoji removal. Customise functions are created using python libraries for achieving text data transformation. On other hand, data augmentation is performed for image transformation. This helps in reshaping, resizing, zooming, and constructing the data.

Modelling: The next critical stage is modelling; models are built on pre-processed data to achieve the desired results. The models used in this study help in understanding how images and captions posted on Instagram can be used to detect suicidal tendencies. Text analysis and image classification are accomplished through natural language processing and pre-trained VGG16 respectively. The split of 60:20:20 is used for text data splitting for data training, testing, and validation, while the split of images is used for the data source. Word2vec is used to assign vectors to words, and a customised list is declared to convert contraction words into simple words to convert text data into the same format. This enabled the implementation of an LSTM and hybrid CNN+LSTM model for text

[5] https://www.kaggle.com/datasets/prithvijaunjale/instagram-images-with-captions
[6] https://www.kaggle.com/datasets/muhammadusmansaeed/depression

analysis with a focus on suicidal keywords.
Evaluation: This stage facilitates in the evaluation of the models that are implemented. While evaluating the models, well-known evaluation methods such as accuracy, precision, F1 score, and recall are used. While visualization is used to plot and comprehend various aspects of outcome such as loss, epochs, and accuracy.

4 Design Specification

This section is more critical for conceptualizing the architecture of implemented models. Both natural language processing and deep learning use three-layered architectures, with each layer subdivided into multiple layers to carry out specific tasks for the next layer. The architecture is depicted in the Figure 5 below.

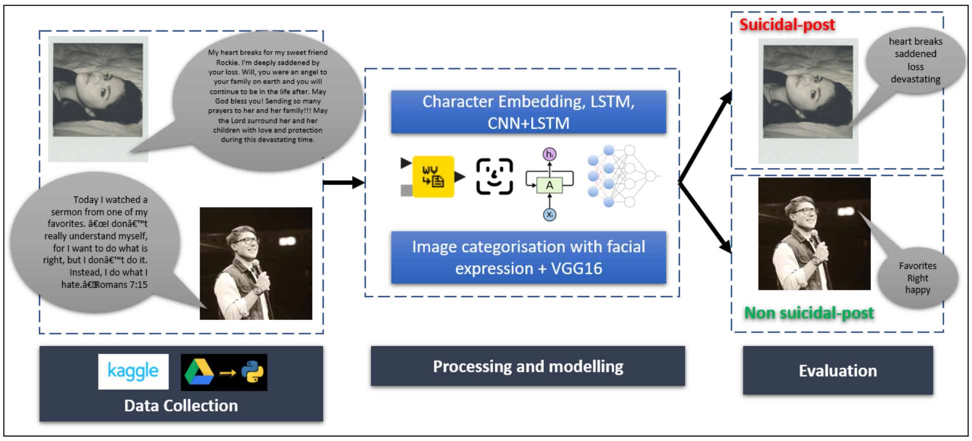

Figure 5: Framework Architecture

The initial layers are concerned with the relevant data for modelling. This layer collects and exposes data on Google Drive, followed by data cleaning. The actual model implementation follows, with LSTM and hybrid CNN+LSTM used for text analysis and VGG16 used for image processing. The model is then evaluated and analysed.

5 Implementation

This research used three layered architectures, as described in the previous section. The subsections that follow provide detailed implementations of both the LSTM model for text analysis and the VGG16 model for image processing.

5.1 Environmental Setup

The implementation necessitates a lengthy processing time due to the large image and text dataset, which totals approximately 4GB of data. The hardware used has 16GB RAM, a 458GB hard drive, and a GPU NVIDIA T4 Tensor Core GPUs. Google Colabs is used for implementation. Data is uploaded to Google Drive using the Kaggle package, which makes use of the processing power of the Google platform to handle large amounts of data. Python has become the most popular data analytic platform due to its large number of libraries and packages that allow for multiple model implementations and evaluations (Stančin and Jović; 2019). In addition to standard processing packages such as Numpy, pandas, and re, libraries used in this research include Ftfy, nltk, gensim.models for text analysis and deepface, cv2 for image processing, as well as keras,sklearn.metrics, and tensorflow. Ftfy aids in the handling of bad Unicode characters in text data, while nltk is a Python package that provides various libraries for unstructured data such as captions, comments, and text analysis. In Python, gensim.models allow you to assign vectors to words that will be used as input for models. DeepFace is a framework for face recognition and attributes inspired by Keras and TensorFlow. CV2 is an open-source library for image processing, object recognition in videos, and other tasks. The Figure 6 below shows some of the packages and libraries that were used in this study.

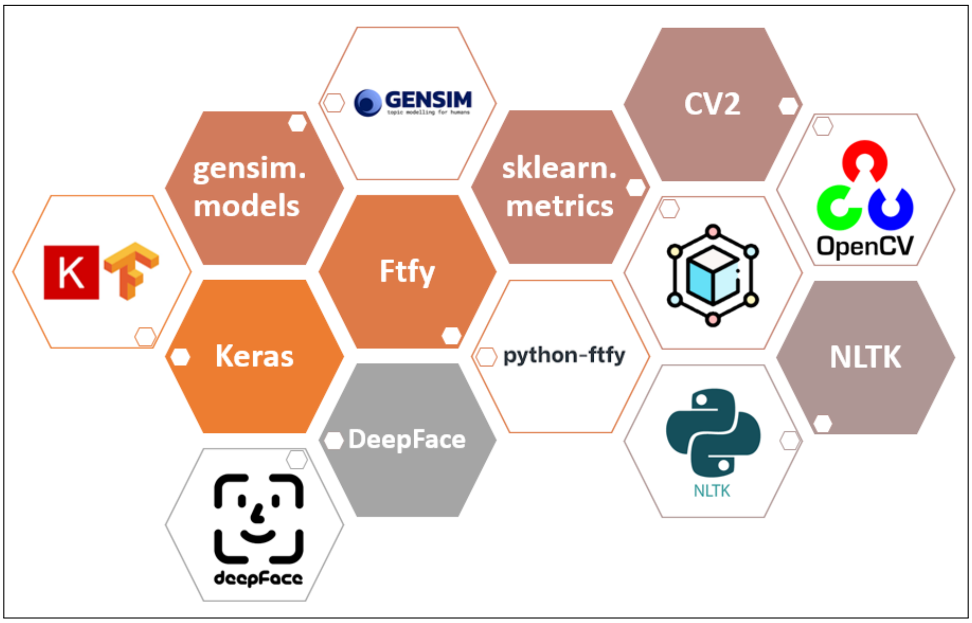

Figure 6: Packages and Libraries used

5.2 Data Transformation

Data Transformation:
DataFrame is used to read the .csv file before beginning the data transformation for "Captions". In the first step, data is subdivided into two DataFrame as suicidal and non-suicidal categories based on common suicidal keyword (Chadha and Kaushik; n.d.). As the model must understand the input provided, word embedding is used to assign the vectors to text. Word2vec, a popular method for vectorizing captions, is implemented using pre-trained GoogleNews-vectors-negative300. Since the captions are unstructured data, they contain numerous extra texts, emojis, and special characters which are handled by implementing customized functions. A contraction list is declared, and each one is assigned its simple form. A customised caption_cleanup function is also included for removing emojis, hashtags, and special characters. Furthermore, ftfy.fix text handles Unicode. As captions contain additional stopwords that are irrelevant to this research paper, they are removed using nltk libraries such as stopwords and word tokenize. The PorterStemmer is then used to remove stemming words and convert each word back to

its original format. As the data is cleaned up, there will be a total of 13891 captions available for model implementation. Various Tokenizer methods are used to convert a list into a 2D array. The wordCloud Figure 7 Figure 8 below help to clearly visualize the captions with suicidal and non-suicidal categories on the used classified data.

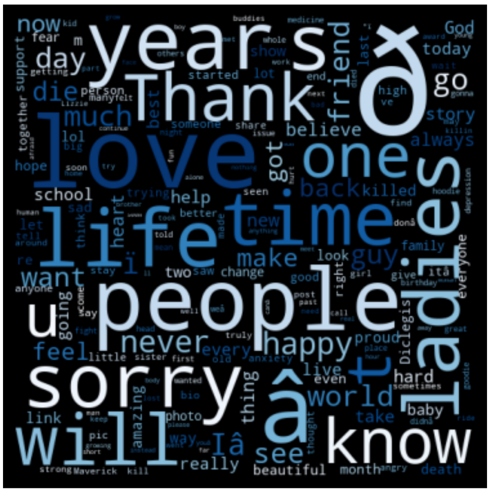

Figure 7: Suicidal Caption keywords

Figure 8: Non-Suicidal Caption keywords

Image Data Transformation:
It is critical to identify the images that will be useful for implementing the model from the massive, downloaded dataset. Images with faces are identified first because they will enable in identifying emotions and then data is classified as suicidal or non-suicidal. The cv2 is used to load haarcascade_frontalface_default.xml for face detection, which is then iterated through the image folder for categorizing the face images and others. Following that, the face images are classified into four subfolders: "Angry","Sad","Fear" for suicidal tendencies, and "Neutral" for non-suicidal tendencies. DeepFace framework, which makes use of the common features and texture of the face, enabled subfolder categorisation (Wang and Deng; n.d.). Another dataset has been loaded for training and validation. Each has a subfolder with images of happy and sad faces that will be used for training and validating the model. The training dataset contains 12,102 unique images, while the validation dataset contains 2964 unique images. It is critical to resize the input images for VGG16 to learn the model quickly. All of the images have been converted to 224*224 pixels. Data augmentation is accomplished through rescaling, zooming, changing the rotation, and horizontal shifts. Labels are also assigned to the images as they are loaded into the model. Another keras library, flow_from_directory, is used to augment the images learned on the training dataset. Furthermore, all of the images are labelled with OneHotEncoder, which outperforms state-of-the-art methods (Karthiga et al.; 2021).

5.3 Long-Short Term Memory:

After the data has been pre-processed and transformed, the Long-Short Term Memory model is used to perform text training and classification. The LSTM network is a type of recurrent neural network that performs well in terms of memory cell vectors. Memory cells replace the updates to the hidden layers. This is extremely beneficial when analyzing text data because it is better at detecting long-term dependencies in data. The LSTM architecture is simple, with input gates, a forget gate, and an output gate. The input gate value in this research paper is 11546, calculated with an embedding matrix, which is the minimum of word index and maximin captions available. The embedding value for the forget gate is 300, and the maximum caption value for the output gate is 140. Since the model has a binary output, the loss is "binary_crossentropy", the density is 1, and the activation is "Sigmoid". According to (Zhang et al.; 2018), a CNN layer improves model performance along with LSTM. A hybrid model shows a better result for improving text classification accuracy. CNN+LSTM is achieved by adding a CNN layer before the LSTM. CNN will extract the text's local features first, and then use the memory cells characteristic of LSTM by implementing the LSTM layers. CNN is implemented with a 'relu' activation, a filter size of 32, and a karnel size of 3.

5.4 VGG16:

The VGG16 model is used to identify suicidal and non-suicidal images in image processing. VGG16 has already been trained on a massive ImageNet dataset, and it has demonstrated excellent image classification and recognition accuracy (Sikha and Bharath; n.d.). The first step in implementing the VGG16 mode is to pre-train it. As we have chosen categories for classification, the VGG16 model is pre-trained to remove the last layer of the model. During pre-training, the input size is given as image size plus [3], indicating that the input images are RGB format, the weights are given as 'imagenet' due to the huge number of images, which provides the standard measure, and include_top is set to false to remove the last layer. The layers of VGG16 are then set to False for trainability because they have already been trained on a large dataset. This will also save computational power because the images will not be trained repeatedly, allowing the model to perform better. For dense layers, the layer is flattened, and the input dataset is appended. The model is compliant with "binary_crossentropy" as output, and "adam" as optimiser.

6 Evaluation

In this research, implemented models are evaluated to comprehend the achievement of the stated goal. This will assist in comprehending how image and text classification using transfer learning and NLP successfully categorize Instagram posts as suicidal or non-suicidal. Epochs, dropouts, and various layers in the model are major factors used and fine-tuned to achieve accuracy.

6.1 Text processing evaluation:

6.1.1 Experiment 1: Text classification with LSTM

The initial model has 128 LSTM layers and 5 epochs. When we observe the model accuracy Figure 9 Figure 10 below, we can clearly see that the obtained accuracy is 98.63%, while the loss decreases from 0.1 to 0.03. The fact that the accuracy loss is decreasing while the validation loss is increasing indicates that the training dataset is easier to predict than the validation dataset.

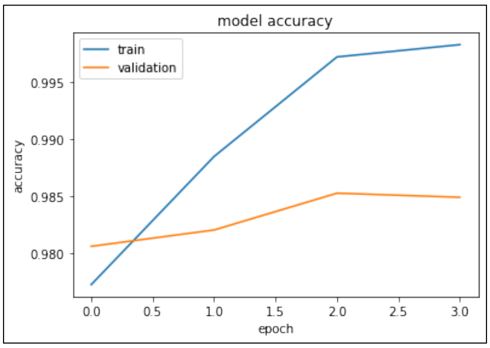

Figure 9: LSTM Model Accuracy

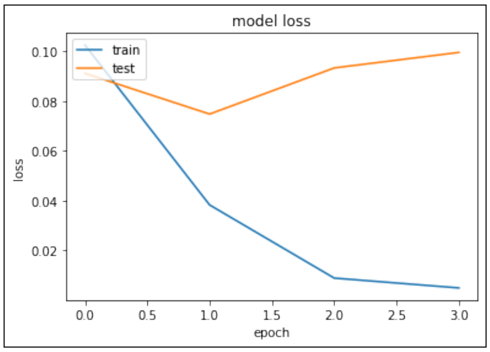

Figure 10: LSTM Model Loss

```
Epoch 1/5
261/261 - 10s - loss: 0.1024 - accuracy: 0.9772 - val_loss: 0.0910 - val_accuracy: 0.9806 - 10s/epoch - 38ms/step
Epoch 2/5
261/261 - 3s - loss: 0.0382 - accuracy: 0.9885 - val_loss: 0.0747 - val_accuracy: 0.9820 - 3s/epoch - 12ms/step
Epoch 3/5
261/261 - 3s - loss: 0.0089 - accuracy: 0.9972 - val_loss: 0.0932 - val_accuracy: 0.9852 - 3s/epoch - 12ms/step
Epoch 4/5
261/261 - 3s - loss: 0.0049 - accuracy: 0.9983 - val_loss: 0.0995 - val_accuracy: 0.9849 - 3s/epoch - 11ms/step
```

Figure 11: Model execution details

6.1.2 Experiment 2: Text classification with CNN+LSTM

Another experiment is being conducted using the hybrid text analysis model CNN+LSTM. With this model, a CNN layer with Conv1D, MaxPooling1D, and dropout is added before the LSTM layer. The number of LSTM layers was increased to 300 while the training and validation datasets remained unchanged. The Figure 12 Figure 13 below shows that accuracy loss decreases from 0.1 to 0.03 while validation loss decreases from the second epoch and increases at the last epoch. The model had a 98.27% accuracy.

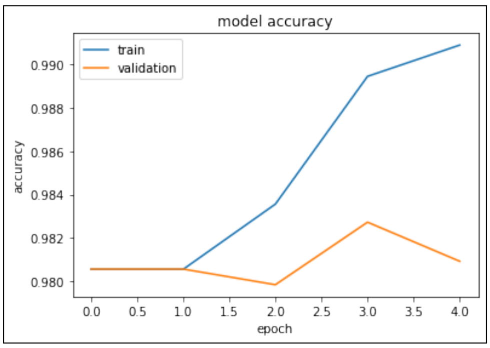

Figure 12: Hybrid CNN$_LSTMModelAccuracy$

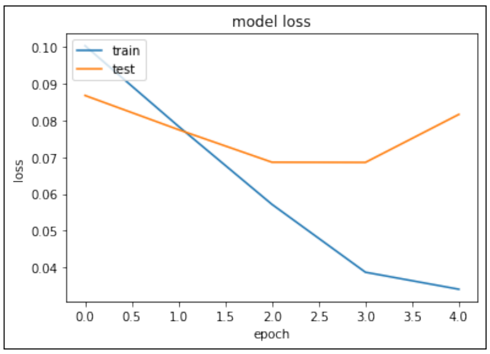

Figure 13: Hybrid CNN$_LSTMModelLoss$

6.2 Image processing evaluation:

6.2.1 Experiment 1: Effectiveness of VGG16 model for image classification

VGG16 model is triggered with 5 epochs with steps_per_epoch set as per the training_Set containing 12102 and validation_steps as test_sets containing 2964 images. The model took more than an hour for running successfully. In below Figure 14 Figure 15, gives clear results for achieving model accuracy and loss. Figure 16 shows that achieved accuracy is 85% with loss is dropping from 0.5 to 0.3

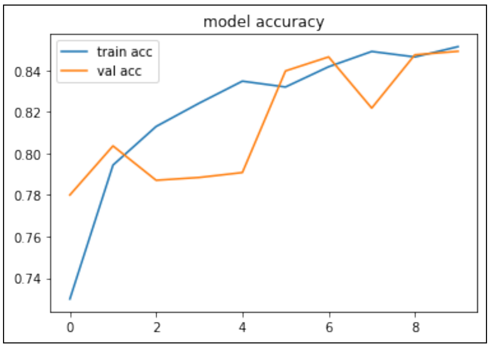

Figure 14: VGG16 Model Accuracy with 10 Epochs

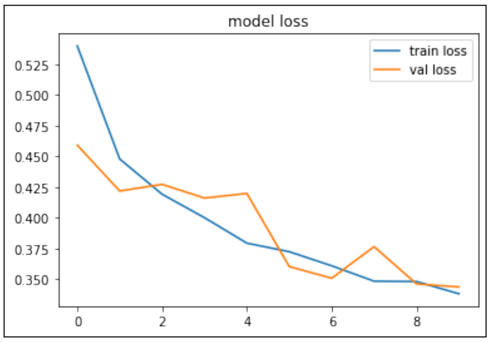

Figure 15: VGG16 Model Loss with 10 Epochs

```
Epoch 1/10
757/757 [==============================] - 2947s 4s/step - loss: 0.5399 - accuracy: 0.7300 - val_loss: 0.4592 - val_accuracy: 0.7800
Epoch 2/10
757/757 [==============================] - 132s 174ms/step - loss: 0.4479 - accuracy: 0.7944 - val_loss: 0.4219 - val_accuracy: 0.8036
Epoch 3/10
757/757 [==============================] - 136s 179ms/step - loss: 0.4193 - accuracy: 0.8130 - val_loss: 0.4273 - val_accuracy: 0.7871
Epoch 4/10
757/757 [==============================] - 132s 174ms/step - loss: 0.4002 - accuracy: 0.8242 - val_loss: 0.4161 - val_accuracy: 0.7885
Epoch 5/10
757/757 [==============================] - 131s 173ms/step - loss: 0.3795 - accuracy: 0.8348 - val_loss: 0.4199 - val_accuracy: 0.7908
Epoch 6/10
757/757 [==============================] - 131s 174ms/step - loss: 0.3725 - accuracy: 0.8320 - val_loss: 0.3604 - val_accuracy: 0.8397
Epoch 7/10
757/757 [==============================] - 131s 173ms/step - loss: 0.3611 - accuracy: 0.8418 - val_loss: 0.3509 - val_accuracy: 0.8465
Epoch 8/10
757/757 [==============================] - 132s 174ms/step - loss: 0.3485 - accuracy: 0.8491 - val_loss: 0.3767 - val_accuracy: 0.8219
Epoch 9/10
757/757 [==============================] - 132s 174ms/step - loss: 0.3483 - accuracy: 0.8465 - val_loss: 0.3463 - val_accuracy: 0.8475
Epoch 10/10
757/757 [==============================] - 132s 174ms/step - loss: 0.3383 - accuracy: 0.8514 - val_loss: 0.3439 - val_accuracy: 0.8492
```

Figure 16: VGG16 Model execution details

6.2.2 Effectiveness of VGG16 with increased epochs model for image classification

The optimal epoch number is determined with the classification accuracy of the model (Thenmozhi and Reddy; n.d.). The accuracy increased with the number of epochs for evaluation 1, as shown in Figure 16. As a result, epochs are increased from 10 to 20 in the second evaluation. When compared to evaluation 1, the model took a longer time but provided greater accuracy. The Figure 17 Figure 18 below shows a consistent drop in loss from 0.5 to 0.3 while maintaining an accuracy of 87

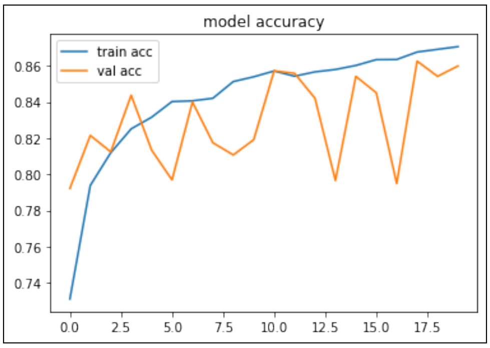

Figure 17: VGG16 Model Accuracy with 20 Epochs

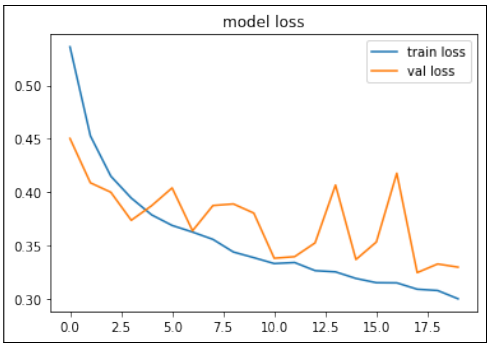

Figure 18: VGG16 Model Loss with 20 Epochs

6.3 Discussion:

This research performed an analysis of text and images posted on Instagram. The result achieved is excellent and can be compared with other research.The research has evaluated proposed classification models, which compared in the below Figure 19.

Model	Dataset	Evaluation	Accuracy
VGG16	Image	Epoch 10	85%
VGG16	Image	Epoch 20	87%
LSTM	Text	Epoch 5	98%
CNN+LSTM	Text	Epoch 5	98%

Figure 19: Various model comparison

The first evaluations of VGG16 have shown the accuracy of 85% and 87% with epochs increased, which infers that the model is successful for the identification of suicidal and non-suicidal images. On the other hand, different model implementations compared to (Amanatidis et al.; n.d.), text classification with LSTM and Hybrid CNN+LSTM have achieved excellent accuracy of 98% which claims that both models are equally good for text classification contract with (Zhang et al.; 2018). Below Figure 20 Figure 21 shows the classification of Instagram images into suicidal and non-suicidal categories.

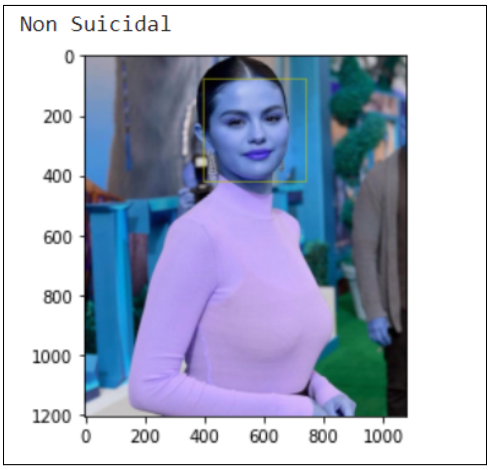

Figure 20: Non-suicidal post

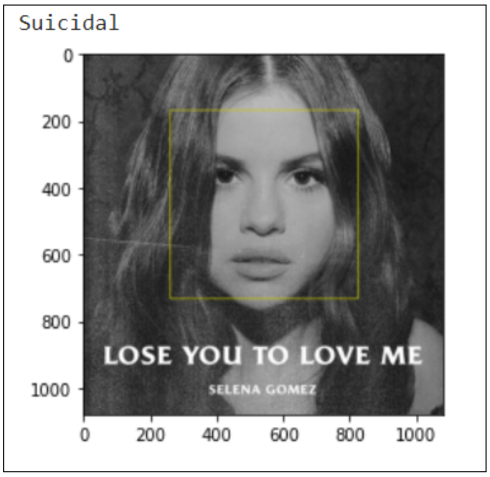

Figure 21: Suicidal post

7 Conclusion and Future Work

The objective of this research is to categorize Instagram posts as suicidal or non-suicidal. People, particularly young people, who are thinking about or considering suicide approach social media for emotional sharing, information, and support. Deep learning and natural language processing models for image and text classification were proposed in this study. The results show a successful classification of suicidal and non-suicidal Instagram posts with the highest accuracy. This information can be used by counsellors, doctors, psychologists social workers and the government to assist the identified victim in saving his or her life.

Future work on this research could include improving data quality for text datasets. Changing the epochs per model can improve image classification accuracy. It will also be important to comprehend images with no facial expression, as people nowadays post text and quotes in form of images.

Acknowledgement

I would like to express my heartfelt gratitude to my supervisor, Professor Athanasios Staikopoulos, for his unwavering support, guidance, and encouragement in completing the research successfully. Sincere thanks to NCI and faculty members for providing end-to-end support during the Data Analytics course. Last but not least, I would like to take this opportunity to thank my family and friends for their encouragement and support in helping me achieve my goal.

References

Aldhyani, T. H. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H. and Ahmed, Z. A. T. (n.d.). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models, **19**(19): 12635. Number: 19 Publisher: Multidisciplinary

Digital Publishing Institute.
URL: *https://www.mdpi.com/1660-4601/19/19/12635*

Amanatidis, D., Mylona, I., Kamenidou, I. E., Mamalis, S. and Stavrianea, A. (n.d.). Mining textual and imagery instagram data during the COVID-19 pandemic, **11**(9): 4281. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/2076-3417/11/9/4281*

Chadha, A. and Kaushik, B. (n.d.). A survey on prediction of suicidal ideation using machine and ensemble learning, **64**(11): 1617–1632.
URL: *https://doi.org/10.1093/comjnl/bxz120*

Cusick, M., Adekkanattu, P., Campion, T. R., Sholle, E. T., Myers, A., Banerjee, S., Alexopoulos, G., Wang, Y. and Pathak, J. (n.d.). Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation, **136**: 95–102.
URL: *https://www.sciencedirect.com/science/article/pii/S0022395621000637*

Dadiz, B. G. and Marcos, N. (n.d.). Analysis of depression based on facial cues on a captured motion picture, *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*, pp. 49–54.

Fahey, R. A., Boo, J. and Ueda, M. (n.d.). Covariance in diurnal patterns of suicide-related expressions on twitter and recorded suicide deaths, **253**: 112960.
URL: *https://www.sciencedirect.com/science/article/pii/S0277953620301799*

Ferreira, L. A., De Rizzo Meneghetti, D., Lopes, M. and Santos, P. E. (n.d.). CAPTION: Caption analysis with proposed terms, image of objects, and natural language processing, **3**(5): 390.
URL: *https://doi.org/10.1007/s42979-022-01322-7*

Haque, F., Nur, R. U., Jahan, S. A., Mahmud, Z. and Shah, F. M. (n.d.). A transformer based approach to detect suicidal ideation using pre-trained language models, *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pp. 1–5.

Jeong, N. and Cho, S. (2017). Instagram image classification with deep learning, *Journal of Internet Computing and Services* **18**: 61–67.

Ji, S., Pan, S., Li, X., Cambria, E., Long, G. and Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications, *IEEE Transactions on Computational Social Systems* **8**(1): 214–226.

Jordan, P., Shedden-Mora, M. C. and Löwe, B. (n.d.). Predicting suicidal ideation in primary care: An approach to identify easily assessable key variables, **51**: 106–111.
URL: *https://www.sciencedirect.com/science/article/pii/S0163834317303432*

Karthiga, R., Usha, G., Raju, N. and Narasimhan, K. (2021). Transfer learning based breast cancer classification using one-hot encoding technique, *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 115–120.

Macalli, M., Navarro, M., Orri, M., Tournier, M., Thiébaut, R., Côté, S. M. and Tzourio, C. (n.d.). A machine learning approach for predicting suicidal thoughts and behaviours among college students, **11**(1): 11363. Number: 1 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41598-021-90728-z*

Malhotra, A. and Jindal, R. (n.d.). Deep learning techniques for suicide and depression detection from online social media: A scoping review, **130**: 109713.
URL: *https://www.sciencedirect.com/science/article/pii/S1568494622007621*

McMullen, L., Parghi, N., Rogers, M. L., Yao, H., Bloch-Elkouby, S. and Galynker, I. (n.d.). The role of suicide ideation in assessing near-term suicide risk: A machine learning approach, **304**: 114118.
URL: *https://www.sciencedirect.com/science/article/pii/S0165178121004157*

Rajesh Kumar, E., Rama Rao, K., Nayak, S. R. and Chandra, R. (n.d.). Suicidal ideation prediction in twitter data using machine learning techniques, **23**(1): 117–125. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09720502.2020.1721674.
URL: *https://doi.org/10.1080/09720502.2020.1721674*

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W. and Kaminsky, Z. A. (n.d.). A machine learning approach predicts future risk to suicidal ideation from social media data, **3**(1): 1–12. Number: 1 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41746-020-0287-6*

Seal, A., Bajpai, R., Agnihotri, J., Yazidi, A., Herrera-Viedma, E. and Krejcar, O. (n.d.). DeprNet: A deep convolution neural network framework for detecting depression using EEG, **70**: 1–13. Conference Name: IEEE Transactions on Instrumentation and Measurement.

Sikha, O. K. and Bharath, B. (n.d.). VGG16-random fourier hybrid model for masked face recognition, **26**(22): 12795–12810.
URL: *https://doi.org/10.1007/s00500-022-07289-0*

Stančin, I. and Jović, A. (2019). An overview and comparison of free python libraries for data mining and big data analysis, *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 977–982.

Suicide (2021).
URL: *https://www.who.int/news-room/fact-sheets/detail/suicide*

Thenmozhi, K. and Reddy, U. S. (n.d.). Crop pest classification based on deep convolutional neural network and transfer learning, **164**: 104906.
URL: *https://www.sciencedirect.com/science/article/pii/S0168169919310695*

Trotzek, M., Koitka, S. and Friedrich, C. M. (n.d.). Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences, **32**(3): 588–601. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

van Mens, K., de Schepper, C., Wijnen, B., Koldijk, S. J., Schnack, H., de Looff, P., Lokkerbol, J., Wetherall, K., Cleare, S., C O'Connor, R. and de Beurs, D. (n.d.). Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study, **271**: 169–177.
URL: *https://www.sciencedirect.com/science/article/pii/S0165032719321937*

Viola, P. and Jones, M. J. (n.d.). Robust real-time face detection, **57**(2): 137–154.
URL: *https://doi.org/10.1023/B:VISI.0000013087.49260.fb*

Wang, M. and Deng, W. (n.d.). Deep face recognition: A survey, **429**: 215–244.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231220316945*

Weng, J.-C., Lin, T.-Y., Tsai, Y.-H., Cheok, M. T., Chang, Y.-P. E. and Chen, V. C.-H. (n.d.). An autoencoder and machine learning model to predict suicidal ideation with brain structural imaging, **9**(3): 658. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/2077-0383/9/3/658*

Xian, L., Vickers, S. D., Giordano, A. L., Lee, J., Kim, I. K. and Ramaswamy, L. (n.d.). #selfharm on instagram: Quantitative analysis and classification of non-suicidal self-injury, *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 61–70.

Zhang, J., Li, Y., Tian, J. and Li, T. (2018). Lstm-cnn hybrid model for text classification, *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 1675–1680.

Zhang, T., Schoene, A. M. and Ananiadou, S. (n.d.). Automatic identification of suicide notes with a transformer-based deep learning model, **25**: 100422.
URL: *https://www.sciencedirect.com/science/article/pii/S2214782921000622*