SUICIDAL IDEATION DETECTION FROM SOCIAL MEDIA

ÖZAY EZERCELİ

IŞIK UNIVERSITY
AUGUST, 2023

SUICIDAL IDEATION DETECTION FROM SOCIAL MEDIA

ÖZAY EZERCELİ

Işık University, School of Graduate Studies, Computer Science Engineering Master Program, 2023

Submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science in Computer Engineering

IŞIK UNIVERSITY
AUGUST, 2023

IŞIK UNIVERSITY
SCHOOL OF GRADUATE STUDIES
COMPUTER SCIENCE ENGINEERING MASTER PROGRAM

SUICIDAL IDEATION DETECTION FROM SOCIAL MEDIA

ÖZAY EZERCELİ

APPROVED BY:

Asst. Prof. Rahim Dehkharghani (Thesis Supervisor)	Işık University
Asst. Prof. Emine Ekin	Işık University
Asst. Prof. Faik Boray Tek	Istanbul Technical University

APPROVAL DATE: 24./08./2023

SUICIDAL IDEATION DETECTION FROM SOCIAL MEDIA

ABSTRACT

Suicidal ideation is a global cause of life-threatening injury and, most of the time, death. Mental health issues have been rapidly increasing, and most are being avoided without adequate treatment. Due to the developments in social media platforms and the online anonymity that these platforms provide, people would like to interact more with others on social platforms. Social platforms are surveillance tools for mining social content and suicidal tendencies. The current thesis attempts to present a solution to detect depression/suicidal ideation by using state-of-the-art natural language processing (NLP) and deep learning (DL) approaches (BiLSTM, BERT Transformer). Three different novel approaches are proposed for three different datasets of textual content. The SuicideDetection dataset is a publicly available dataset which is a collection from the social platform of Reddit's subreddits ("SuicideWatch", "depression", "bipolar", "offmychest", "anxiety") in Kaggle and the SWMH dataset is a collection from only "SuicideWatch" subreddit. The CEASEv2.0 dataset is another used dataset which is a collection of 4932 suicide notes. The proposed models outperformed the latest models by 2% and 1% F1 scores on SuicideDetection and CEASEv2.0 datasets, respectively. The best models for each dataset have been analyzed and discussed in terms of performance, along with the characteristics of the datasets and limitations in the suicidal ideation classification. This performance can be measured by common metrics such as Accuracy, Precision, Recall, F1-Score, and ROC curve. As its application in the real world, this project can assist psychologists in the early identification of suicidal ideation before the suicidal person harms him/herself. The thesis also demonstrates the potential of employing DL algorithms such as transformers along with the latest word embedding techniques and NLP techniques that could improve the issue of suicidal ideation.

Keywords: Suicidal Ideation Detection, Social Media Content, Word Embedding, Deep Neural Network, BERT Transformers

SOSYAL MEDYA İÇERİĞİNDEN İNTİHAR DÜŞÜNCESİ ALGILAMA

ÖZET

İntihar düşüncesi, yaşamı tehdit eden yaralanmaların ve çoğu zaman ölümün küresel bir nedenidir. Dünyada ruh sağlığı sorunları hızla artmakta ve birçoğu yeterli tedavi görülmeksizin göz ardı edilmektedir. Günümüzdeki sosyal medya platformlarındaki gelişmeler ve bu platformların sağladığı çevrimiçi anonimlik nedeniyle insanlar bu gibi platformlarda başkalarıyla sürekli olarak etkileşim halindedirler. Sosyal media platformları, sosyal içerikler ve intihar eğilimleri madenciliği için gözetim araçları olarak kullanılabilmektedir. Bu tez, son teknoloji doğal dil işleme ve derin öğrenmek yaklaşımlarını (BiLSTM, BERT Transformatörü) kullanarak metinsel içerik üzerinden depresyon/intihar düşüncesini tespit etmek için bir çözüm sunmaya çalışmaktadır. Üç farklı metinsel veri seti için üç farklı yeni yaklaşım önerilmiştir. SuicideDetection veri kümesi, Kaggle'da sunulan halka açık bir veri setidir ve de bu veri seti Reddit'in alt dizinlerinden ("SuicideWatch", "depression", "bipolar", "offmychest", "anxiety") oluşturulan bir koleksiyondur. SWMH veri kümesi sadece "SuicideWatch" alt dizininden toplanan içeriklerle oluşturulmuş bir koleksiyondur. CEASEv2.0 veri kümesi, 4932 intihar notundan oluşan ve kullandığımız bir diğer veri setidir. Önerilen modeller, SuicideDetection ve CEASEv2.0 veri setlerinde sırasıyla %2 ve %1 F1 puanları ile en son modellerden daha iyi performans göstermiştir. Her veri seti için en iyi modeller, veri setlerinin özellikleri ve intihar düşüncesi sınıflandırmasındaki sınırlamalarla birlikte performans açısından analiz edilmiş ve tartışılmıştır. Bu performans, Doğruluk, Kesinlik, Geri Çağırma, F1-Skoru ve ROC eğrisi gibi yaygın ölçütlerle ölçülüp karşılaştırılmıştır. Gerçek dünyadaki uygulaması itibariyle bu proje, intihara meyilli kişi kendisine zarar vermeden önce intihar düşüncesinin erken teşhisinde psikologlara yardımcı olabilir. Ayrıca, intihar düşüncesi sorununu iyileştirebilecek en son kelime gömme teknikleri ve doğal dil işleme teknikleriyle birlikte dönüştürücüler gibi derin öğrenme algoritmalarının kullanılma potansiyelini de göstermektedir.

Anahtar Kelimeler: İntihar Düşüncesi Tespiti, Sosyal Medya İçeriği, Kelime Temsil, Derin Sinir Ağı, BERT Transformatörü

ACKNOWLEDGEMENTS

I would like to express my heartfelt appreciation to my thesis advisor, Asst. Prof. Rahim Dehkharghani, for his invaluable guidance and support throughout the research and writing of this thesis. His patience, insights, and feedback were essential to the completion of this project.

I would also like to thank my all instructors and mentors, especially Asst. Prof. Emine Ekin, for their encouragement and instruction. I am grateful for the knowledge and skills they have imparted to me, which have helped me to become a better researcher and writer.

Finally, I would like to thank my family for their love and support. My parents have always been my biggest motivation and inspiration. Their encouragement and sacrifices have made my education possible. I am also grateful to my sister, for her infinite love and support.

Özay EZERCELİ

To my loved ones…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP: Natural Language Processing

WHO: World Health Organization

LGBTI: Lesbian Gay Bisexual Trans Intersex

US: United States

ML: Machine Learning

DL: Deep Learning

CEASE: Corpus of Emotion Annotated Suicide notes in English

SWMH: Reddit SuicideWatch and Mental Health Collection

LFS: Linear Forward Selection

CRF: Conditional Random Fields

LIWC: Linguistic Inquiry and Word Count

BoW: Bag of Words

UMD: The University of Maryland Reddit Suicidality Dataset

LSTM: Long Short-Term Memory

LR: Logistic Regression

MTL: Multitask Learning

AUC: Area under the ROC Curve

ROC: Receiver Operating Characteristic Curve

GRU: Gated Recurrent Unit

CNN: Convolutional Neural Network

GloVe: Global Vectors for Word Representation

TF-IDF: Term Frequency - Inverse Document Frequency

BERT: Bidirectional Encoder Representations from Transformers

RoBERTa: A Robustly Optimized BERT

NB: Naïve Bayes

kNN: K-Nearest Neighbors

SVM: Support Vector Machine

RF: Random Forest

NLTK: Natural Language Toolkit

HTML: The HyperText Markup Language

ReLu: Rectified Linear Units

MV: Majority Voting

LLM: Large Language Model

AI: Artificial Intelligence

DNN: Deep Neural Network

CHAPTER 1

1. INTRODUCTION

1.1 Suicide Statistics

Suicidal ideation is a severe mental disorder condition that may lead to suicide. It refers to persistent thoughts, fantasies, or preoccupations with self-death and self-harm. Suicide is one of the leading causes of death worldwide, according to the World Health Organization (WHO), claiming the lives of more than 700,000 people each year (World Health Organization, 2021).

Especially among 15-29 year-olds, suicide is the fourth leading cause of death. Suicide rates for this age group are shown in Figure 1.1 globally.

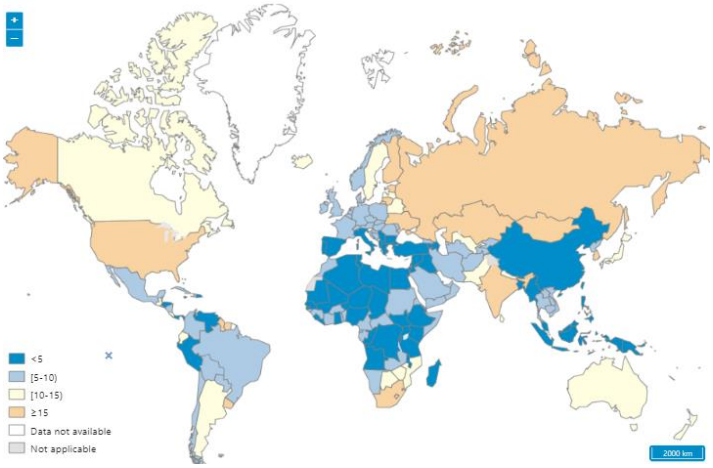

Figure 1.1 Suicide rates between 15-29 age group (per 100.000 population) (Source: https://www.who.int/data/gho/data/themes/mental-health/suicide-rate)

Regardless of any particular parameter, suicide rates globally as shown in Figure 1.2.

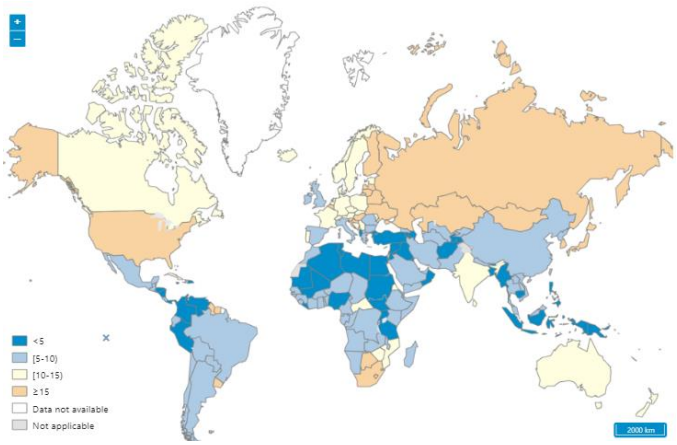

Figure 1.2 Suicide rates (per 100.000 population)
(Source: https://www.who.int/data/gho/data/themes/mental-health/suicide-rates)

The struggle of suicidal ideation among individuals and their visits to the emergency department are highest among people between 14-18 year-olds, and it follows decreasing trend through 55 and over. This situation has pointed out in Figure 1.3.

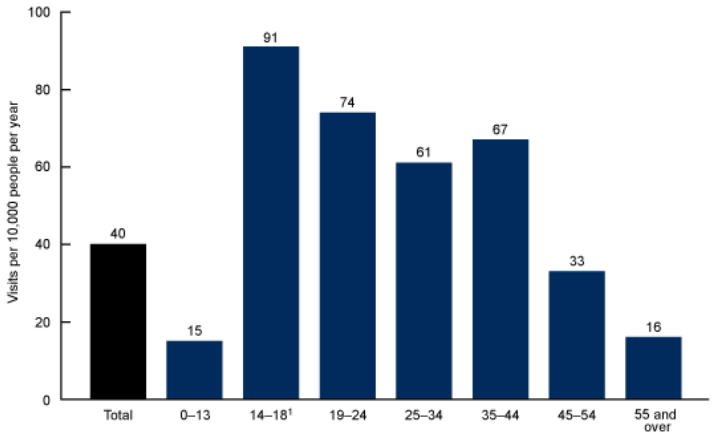

Figure 1.3 Rate of emergency department visits with suicidal ideation, by age group: United States, 2016–2020
(Source: https://www.cdc.gov/nchs/products/databriefs/db463.htm)

1.2 Basis of Suicidal Ideation

There is no precisely one reason to be suicidal. In modern life, various factors such as envitonmental, health and historical can make you suicidal. These factors can be because of experiencing conflict, disaster, violence, abuse, or loss. Besides, suicide rates are highest among vulnerable groups (refugees, LGBTI persons, and prisoners) of people (World Health Organization, 2023).

Especially for people at a young age, any problem at that age might become a trauma if they do not get help. So, it is observed in (Stravynski & Boyer, 2001) that there is a strong association between suicide ideation and loneliness. This can happen in different ways, either mentally (i.e., feeling lonely) or physically (i.e., living alone or being without friends). Sometimes, people might be able to recognize they are suicidal, and most of the time not. In that case suicide prevention can be achieved through various measures at different levels;

- restricting access to means of suicide,
- responsible reporting of suicide in the media,
- fostering socio-emotional skills in adolescents, and
- early identification and management of individuals affected by suicidal behaviors.

Collaboration and coordination among multiple sectors of society, such as health, education, agriculture, business, and media, is crucial for comprehensive and integrated suicide prevention efforts.

1.3 Suicidal Ideation on Social Platforms

Research by Global WebIndex shows that more than half of the global population (60%) now uses social media platforms, and 78% of people older than 18 years in society are using social media (Chaffey, 2023). Most affected people from suicidal ideation use the internet and social platforms the most. Reddit is one of the social platforms that constitute its dominant group of users 18 to 29. This platform is ranked the 9th most popular social media app in the US (Backlinko, 2023). Like Reddit, other social platforms such as Twitter, Facebook, and Instagram provide users to socialize and express their thoughts and feelings by posting text, audio, or video

content. The idea of social anonymity gives the users freedom to tell others anything. This situation puts social platforms into such an essential place in individuals' lives. Even, some people have entirely different identities in social media due to worrying about prejudice or other harmful activities that people can show to them in real life. The problems they are going through in real life direct them to social platforms to talk about themselves and their problems, either in a bad way or a good way.

The Reddit platform has subreddits where people can talk about specific topics. Some of these subreddits are about depression, suicide, anxiety, offmychest, and bipolar which are related to feelings that people can end up committing suicide. Therefore, these platforms can be used as screening tools for early intervention and treatment for potential victims of suicidal ideation, and in this way, we could save lives before they end.

1.4 Suicidal Ideation Detection

Detecting and addressing suicidal ideation in its early stages is crucial to effective treatment.

Traditional approaches to identifying suicidal ideation often rely on self-reporting through clinical interviews or surveys (Abdulsalam & Alhothali, 2022). Despite the high accuracy of these methods, they are limited in their reach, as individuals may be reluctant to reveal their thoughts and feelings in interviews or surveys due to social stigma or fear of being judged.

Identifying suicidal ideation from social media can overcome the limitations of traditional assessment methods. By analyzing publicly shared comments, and interactions of individuals on social media, researchers can learn about their mental and emotional states, including signs of distress and indications of suicidal ideation.

Machine learning (ML) and deep learning (DL) based approaches are gaining popularity in detecting suicidal ideation in social media posts. DL-based algorithms provide promising results in NLP tasks. Among deep neural network (DNN) models, transformer-based algorithms achieve the highest performance; however, the existing research is far from ideal. We attempted to improve the cutting-edge performance achieved in this field by proposing new DNNs and transformer-based models and applying them to three benchmark datasets. Specifically, we proposed three different

approaches and applied them to three different datasets, namely SuicideDetection (Komati, 2021), CEASEv2.0 (Ghosh et al., 2021) and Mental Health Collection (SWMH) (Ji et al., 2021). Two models were designed for binary classification, but the third one accomplished a 5-class classification where the classes are five different states of being suicidals.

The contributions of the current work could be summarized as follows.

- This thesis conducts its experiments on three different datasets, SuicideDetection, CEASEv2.0, and SWMH and provides a comprehensive discussion on those datasets.
- The proposed models achieved state-of-the-art efficiency in two datasets (Suicide Detection, CEASEv2.0), outperforming previous models by 0.02 and 0.01 points in terms of the F1-score.
- This thesis provides a comprehensive comparison on using classic ML and DL techniques in text classification problems for depression and suicidal-ideation detection.

In the remainder, Chapter 2 reports outstanding previous work on suicidal/depression detection. A detailed explanation of the proposed approach is provided in Chapter 3, which is followed by experimental evaluation in Chapter 4. Chapter 5 discusses the results and provides some insights. Chapter 6 concludes thesis and discusses future work.

CHAPTER 2

2. LITERATURE REVIEW

Suicidal ideation is thinking about killing yourself, whether actively or passively. It may involve thoughts about desiring to die, planning to commit suicide, or even attempting it in reality. Suicidal ideation is a severe mental health issue that can result in death if left untreated.

In society, having suicidal thoughts is a severe public health issue. It is essential to identify it early to prevent suicide attempts. Increasing usage of social platforms made it possible to apply automated methods using ML, DL, and transformer-based approaches with NLP algorithms to detect suicidal ideation in text. Following subsections will review cutting-edge studies done with these approaches.

2.1 Machine Learning Based Studies

Suicidal ideation can occur in different manners and in various aspects. Identifying this intention from social media has some limitations, such as the informal language of the posts and the users' demographic characteristics, which makes this text classification task challenging. These challenges are tackled by a group of researchers in the literature using an appropriate feature list. (Shah et al., 2020) proposed a hybrid feature extraction approach that combines Genetic algorithm and Linear Forward Selection (LFS) methods to select the most relevant linguistic and computational features for detecting of suicidal individuals from social platforms and apply the proposed approach to 7098 numbers of Reddit's SuicideWatch subReddit social media posts as their dataset. The article compared various classification methods using different feature sets and concluded that Random Forest performs best.

Identifying the meaning of the context is a crucial part of understanding the ideation (Moulahi et al., 2017). Usually, context composes various parameters that might affect the context, such as sentence length, words starting with capital letters, particular keywords, special stopwords, negative meanings, topic (Grant et al., 2018), abbreviations, and punctuation marks.

In (Moulahi et al., 2017), the authors propose a probabilistic framework that models users' online behaviors as a sequence of psychological states across time and assesses emotional states using Conditional Random Fields (CRF). In addition, some research works (Ramírez-Cifuentes et al., 2020; Guntuku et al., 2017) have investigated statistical differences between textual features (Linguistic Inquiry and Word Count (LIWC) (Chung & Pennebaker , 2012), Bag of Words (BoW) (Qader et al., 2019) or n-grams (Kondrak, 2005) and word embeddings (Lai et al., 2016)) and behavioral characteristics of each risk group for suicidal ideation. They tested statistical and DL-based methods to handle multimodal data for detecting suicidal ideation in Twitter users. The obtained results contribute to our understanding of how the usage of different types (text, visual, relational and behavioral) of data outperforms each model in isolation by 0.08.

2.2 Deep Learning Based Studies

Despite interest in context, advances in DL and attention mechanisms of neural networks in the NLP tasks lead to better representation of the user's context and to capture the connections between sentences (Xu et al., 2020). The authors (Ji et al., 2021) propose a model for effectively encoding relational text in order to detect mental health issues and suicidal ideation. The attention mechanism assigns higher weights to the most important relational features. The datasets used in this study are the UMD Reddit Suicidality Dataset (Shing et al., 2018) which is from 11,129 users post on SuicideWatch, and 11,129 users post from other subreddits. The SuicideWatch and SWMH datasets combine with 54,412 posts in total. The authors report F1 scores of 0.54 and 0.64 for the UMD and SWMH datasets.

The proposed ideas in (Ji et al., 2021) are supported by (Ansari et al., 2023), which proposes a hybrid and ensemble method consisting of Long short-term memory (LSTM) and Logistic regression (LR). This method was tested on three different

datasets: CLPsych 2015 (Coppersmith et al., 2015), Reddit (Pirina & Çöltekin, 2018), and eRisk (Losada & Crestani, 2016). The authors concluded that word embeddings could be one of the driving sources of performance differences among hybrid and DL models. Their best performance was achieved on the Reddit dataset, with 0.77 as the F1 score.

Researchers often do suicidal ideation classifications as binary classification or multiclass classification. Unlike most of them, (Benton et al., 2017) discusses the use of multitask learning (MTL) in a deep learning framework (Zhang & Moldovan, 2019) to estimate the risk of suicide and mental health using a union of multiple Twitter datasets that are manually annotated. The authors compare their MTL model to a well-tuned single-task baseline to predict a potential suicide attempt and the presence of atypical mental health with AUC > 0.8. The final dataset contains 9,611 users in total.

Another study (Ghosh et al., 2021) is one of the first examples of multitask classification for the extension of suicide notes (CEASE dataset (Ghosh et al., 2020)) which is a total of 315 real-life suicide notes. The proposed multitask framework uses GloVe embedding and the Bi-GRU DL layers to perform three different classification tasks such as emotion, sentiment and depression detection. They achieved 0.74 accuracy on the depression detection task.

DNNs have shown high performance in a wide range of tasks, such as NLP, image classification, and speech recognition. In many text classification tasks, DL models outperform classic ML models and data analysis techniques (Janiesch et al., 2021).

In (Tadesse et al., 2019), the authors propose a DNN model that combines LSTM and convolutional neural networks (CNN) that are pre-trained with 300-dimensional Word2vec word embedding vectors for suicidal ideation detection task on Reddit dataset (Ji et al., 2018). The dataset is composed of 3549 suicidal and 3652 non-suicidal posts. Various NLP techniques, such as Term Frequency-Inverse Document Frequency (TF-IDF), BOW, and statistical features, are employed to encode words and extract features from text data. The authors in (Tadesse et al., 2019) observed hopelessness, frustration, anxiety, and loneliness that point to suicidal ideation. The proposed model of combined neural network outperformed other approaches in the literature when applied to the Reddit dataset (Ji et al., 2018) with a 0.93 F1 score.

Another study (Aldhyani et al., 2022) suggested a CNN-BiLSTM based DL model. This model achieved 0.95 suicidal ideation detection accuracy by using textual features. The authors believed that Reddit users who are suicidal may have mental and psychological health issues. Thesis' proposed approach could achieve 0.97 as an F1 score on the same dataset and outperformed pioneering approaches in the literature.

Some users might hide their real identities and feelings and not explicitly express their feelings on social media due to privacy issues. This situation leads to the trade-off between privacy and prevention (Coppersmith et al., 2018) because it prevents reaching a fraction of users that might be at risk; however, those users might express their feelings and intentions implicitly using abstract or sarcastic sentences such as "What a life!", "Ohh I am so lucky!". Such sentences in social media could be misleading, and it is not easy to understand the intention of their writer. However, generally, suicide notes are written shortly, which can be used as clues to detect depressed and in-danger users before suicide. Taking into account these limitations, the authors (Ghosh et al., 2020) provide a collection of English suicide notes, named CEASE, which consist of annotated with 15 emotion labels. The corpus contains 2393 sentences from approximately 205 suicide notes. DL-based ensemble models of CNN, the gated recurrent unit (GRU), and LSTM are employed together to detect emotions in the corpus with an accuracy of about 0.60.

2.3 Transformer Based Studies

Transformers, as recent advances in the NLP field have made a wide variety of tasks more accurate and faster, as it provides a better understanding of sequence data and a shorter training time because it uses larger dimensions for word and sentence embeddings (Vaswani et al., 2017) and processes the whole input in parallel.

Since transformers are open-source libraries, researchers can extend them to build more advanced architectures (Wolf et al., 2020). Pre-trained transformers with domain-specific data can be used for specific tasks. In (Ji et al., 2021), the authors developed two pre-trained masked language models, MentalBERT and MentalRoBERTa, for the mental healthcare research community. The authors used the language representations pre-trained in the target domain to improve mental health detection task performance. They evaluated these models and several variants of pre-

trained language models and could achieve a recall of 0.70 and an F1 score of 0.72 by the MentalRoBERTa model.

In another work (Sawhney et al., 2020), the authors propose a time-aware transformer-based model called STATENet for preliminary screening of suicidal risk on social media. The model jointly learns from the tweet's language and the emotional historical spectrum in a time-sensitive manner to determine users who have suicidal ideation. The authors use Twitter timeline data from the dataset proposed by (Sinha et al., 2019), which contains 34,306 tweets. They report an F1 score of 0.81 for STATENet, more significant than the F1 score of 0.77 reported by the best-performing baseline model. The authors found that suicide risk is not a binary (all-or-nothing) concept but exists on a spectrum. They pointed out that simplifying this spectrum into binary labels, such as "suicidal" and "non-suicidal," could lead to inaccurate and misleading risk assessments.

2.4 Summary

Advances in DL and Transformer-based algorithms lead to better models of suicidal ideation. Models generated using DL and transformer-based techniques outperformed ML-based methods in terms of performance. Table 2.1 summarizes the categorization of methods for suicidal ideation detection systems.

Table 2.1 Review of methodologies for suicidal ideation detection

Approach	References	Methodology	Data
ML Based	(Moulahi et al., 2017)	CRF, DARE framework	Text
	(Guntuku et al., 2017)	RF, LR, NB	Text
	(Pirina & Çöltekin, 2018)	SVM	Text
	(Shah et al., 2020)	LFS, NB, SVM, kNN, RF	Text
	(Ramírez-Cifuentes et al., 2020)	Word embedding, SVM	Text, Image
DL Based	(Coppersmith et al., 2015)	GloVe, BiLSTM, self attention	Text
	(Losada & Crestani, 2016)	Dynamic method	Text
	(Benton et al., 2017)	Multi-task learning, neural networks	Text
	(Shing et al., 2018)	Word embedding, CNN, max pooling	Text
	(Ji et al., 2018)	Word embedding, LSTM, max pooling	Text
	(Tadesse et al., 2019)	Word embedding, LSTM-CNN	Text
	(Ghosh et al., 2020)	CNN, GRU, LSTM	Text
	(Ji et al., 2021)	Relation network, LSTM, attention, lexicon	Text
	(Ghosh et al., 2021)	GLoVe, Bi-GRU, Attention	Text
	(Aldhyani et al., 2022)	CNN-BiLSTM	Text
	(Ansari et al., 2023)	Word embedding, AttentionLSTM, Ensemble	Text
Transformer Based	(Sinha et al., 2019)	BiLSTM, Attention	Text
	(Sawhney et al., 2020)	STATENet (BERT, Time Aware LSTM)	Text
	(Ji et al., 2021)	MentalRoBERTa	Text

CHAPTER 3

3. PROPOSED METHODOLOGY

3.1 Framework of Proposed Methodology

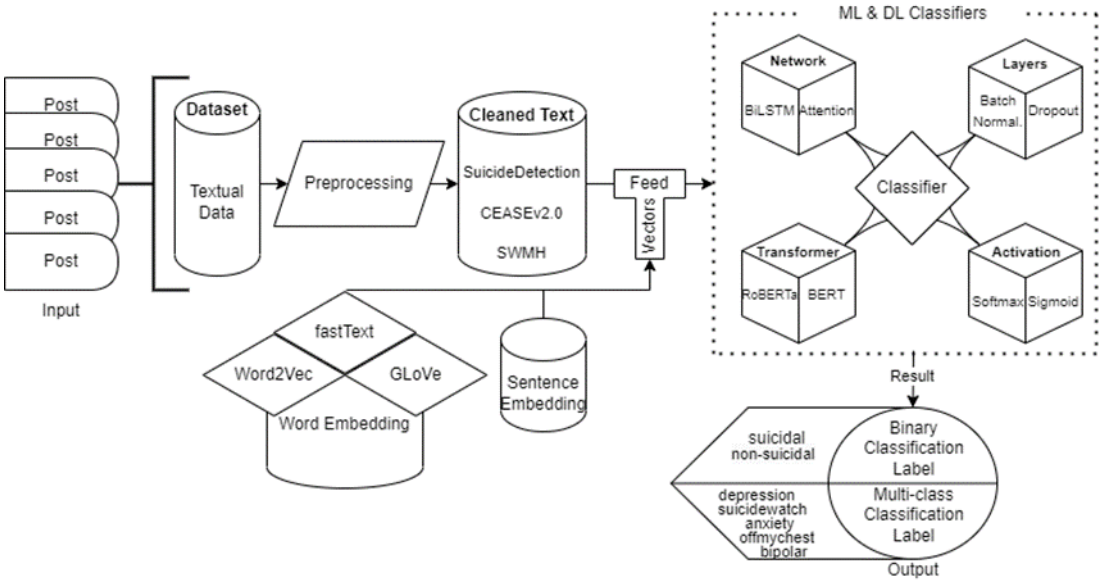

Figure 3.1 Proposed Suicidal Ideation Detection Classifier Framework

Data Availability is essential in choosing a classification model for text classification problems. Current research has used three different datasets with different characteristics. It presents three unique models in this work, which are applied to all three datasets considering their diverse features. Proposed classifier framework for the suggested models is illustrated in Figure 3.1. Firstly, our two datasets had been created from the posts which are taken from the Reddit platform and

other one consists of suicidal notes. These are our datasets and all datasets go through distinct pre-processing stages. The sequence data is then transformed to vectors using several word embeddings such as Word2Vec, fastText, and GloVe or sentence embedding. The models' architecture was built using cutting-edge ML and DL methods. Specifically, the BiLSTM layer, the attention layer (Vaswani et al., 2017), and different types of transformers (BERT, RoBERTa) are used along with the additional dropout and batch normalization layers within the proposed models. Two out of three models result in binary labels (suicidal, non-suicidal), while the third results in multi-class labels (depression, suicidewatch, anxiety, offmychest, bipolar). Therefore, the sigmoid activation function (Narayan, 1997) and loss function of binary cross entropy are used for SuicideDetection and CEASEv2.0 datasets. However, the softmax activation function (Bridle, 1990) and the sparse categorical cross-entropy loss function are used for the SWMH dataset. All steps in Figure 3.1 are explained in detail in the following sections for our proposed models.

3.2 Preprocessing

Textual datasets are mostly unstructured because they are written in natural language by different people. Social media posts often contain images, videos, emoticons, and other multimedia elements. These data should be converted to structured form by applying a preprocessing step before they are given into the model as input. In the preprocessing step, a set of methods are applied to the unstandardized dataset using standard NLP libraries such as NLTK (Loper & Bird, 2002), Beautiful soup (Richardson, 2007), and Neattext (E. Agbe (JCharis), n.d.) to:

- Reduce noise in the data to achieve higher performance,
- Reduce the size of the data for more efficiency,
- Remove irrelevant features for greater consistency.

Some steps in the preprocessing phase are commonly applied to all datasets but some others are dataset-specific. Table 3.1 shows which pre-processing step has been applied to which dataset.

A collected corpus called a contraction dictionary from the social media posts, which is composed of 151 most common contractions and their openings (i.e., "don't" -> "do not", "ain't" -> "is not"). Preprocessing consists of steps for removing HTML

codes, links, new lines, extra whitespace, and special characters. Stopwords are removed except for the keyword "I" because it emphasizes suicidal ideation.

Preprocessing steps are separately applied to datasets according to their nature and writing style; for example, due to spelling errors in the CEASE dataset, spell correction is applied only on this dataset, or due to the short length of sentences in the SWMH dataset, singular to plural form conversion is not applied on this dataset.

Table 3.1 Preprocessing steps for each of the datasets

	Datasets		
Preprocessing	SuicideDetection	CEASEV2.0	SWMH
Remove HTML	✓	✓	✓
Remove new lines	✓	✓	✓
Remove Links	✓	✓	✓
Remove special characters	✓	✓	✓
Replacing Contractions	✓	✓	✓
Remove Stopwords	✓	✓	✓
Words to Singular	✓	✓	-
Spell correction	-	✓	-
Lemmatization	-	-	✓

Each dataset used has different characteristics. In addition, the informal language of the datasets makes it difficult to structure them. Therefore, part of the pre-processing of the datasets is differentiated. In the given Table 3.2, three example sentences are given for each dataset before and after preprocessing. For example, the sentence from the SuicideDetection dataset "It ends tonight.I can't do it anymore. I quit" becomes "It ends tonight.I can't do it anymore.I quit". Stop words such as "can't", "do", "it" are removed and the letter "I" is retained. Punctuation is removed, but case-folding is not applied because people using capital words might be trying to emphasise some part of their feelings or thoughts. In the other given sentence example from the CEASEv2.0 dataset, "but you shattered my dreams" becomes "shattered dream". We changed plural words to singular to increase the rate of common features and also to get rid of redundant features. An example sentence is given for the SWMH dataset, but since this dataset is used for the BERT Transformer, the preprocessing of the BERT Transformer itself would be sufficient to have an optimal feature set.

Table 3.2 Example preprocessings for each dataset

Dataset	Before	After	Label
SuicideDetection	It ends tonight.I can't do it anymore. I quit.	It end tonight I anymore I quit	Suicide
	I'm f*cked assignment is due tomorrow and I haven't even started yet.	I fucked assignment tomorrow I started	Non-suicide
	PLEASE HELP ME I CANT STOP SCREAMING I NEED HELP	PLEASE HELP ME I CANT STOP SCREAMING I NEED HELP	Suicide
CEASEv2.0	but you shattered my dreams.	shattered dream	Depression
	i expect pain very likely to outweigh happiness and satisfaction in my life.	i expect pain likely outweigh happiness satisfaction life	Depression
	i love you completely you will find my body on the lot on the north side of the house.	i love completely body lot north house	Non-depression
SWMH	I just took about 30 anxiety pills. I doubt it will kill me but if I does, great!	take anxiety pills doubt kill great	Self.depression
	I'm gonna do it tonight I've attempted 4 times already. Why not once more?	go tonight attempt times already	Self.suicidewatch
	I fell in love while travelling, and ruined it. I'm never going to see him again and it's wrecking me.	fall love travel ruin never go see wreck	Self.offmychest

3.3 Word Embedding

Word embedding is an NLP technique for representing words as numeric vectors in a high-dimensional spaces (100, 200, 300, ...). It allows words to be compared based on their semantic meaning; e.g., words with similar meanings will have similar vectors. Most popular word embeddings are used, GloVe, fastText, and transformer-based sentence embeddings for training the proposed models.

3.3.1 Word2Vec

The Word2Vec (Mikolov et al., 2013), is designed as a two layers neural network to represent the linguistic contexts of words. It is used to understand the similarity relationship between the words. It uses a large text corpuses like our datasets as input and generates a vector space with hundreds of dimensions (between 100 and 300). The standart dimension for Word2Vec pre-trained model is 300 dimensions. The calculation of the Word2Vec algorithms consist of two models such as skip-gram and CBOW. The skip-gram model generates a word from the words that surround it in a text sequence. In addition, the CBOW model generates a center word based on context words surrounding it. The cost function of Word2Vec can be calculated as in (3.1).

$$J(\theta) = \sum_{c} \sum_{o} \log P\left(w_o | w_c\right) \tag{3.1}$$

where:

$J(\theta)$ is the cost function, which measures how well the model predicts the context words given the center word,

θ are the parameters of the model, which are the weights of the neural network,

c is the index of the center word,

o is the index of the context word,

$P(w_o|w_c)$ is the probability of the context word w_o given the center word w_c. This is calculated using softmax function.

3.3.2 GloVe: Global Vectors for Word Representation

GloVe (Pennington et al., 2014), developed by Standford, as one of the prominent word embedding models, unlike Word2vec (Mikolov et al., 2013), utilizes global statistics (word co-occurrence) along with local statistics (local context information of words) to generate word representations. GloVe enables the identification of meaningful semantic relationships among words. Experiments are started by using the GloVe embeddings first. Equation (3.2) shows the cost function for GloVe word embeddings.

$$J = \sum_{i,j=1}^{V} f(X_{ij})\left(w_i^T \widetilde{w_j} + b_i + \widetilde{b_j} - logX_{ij}\right)^2 \tag{3.2}$$

where:

X_{ij} represents how often word i appears in the context of word j,

w_i is the vector of main word,

$\widetilde{w_j}$ is the vector of the context word,

$b_i, \widetilde{b_j}$ are the main and context words' scalar biases, and

f is the weighting function that helps us to prevent learning only from prevalent word pairs.

3.3.3 FastText

FastText (Bojanowski et al., 2017), is an open-source library for efficient learning of word representations. It includes two models for computing word representations: skip-gram and continuous-bag-of-words (cbow). The skip-gram model acquires knowledge to estimate the neighbors of a target word, while the cbow model forecasts the target word based on its neighbors. Using subword-level information, it builds vectors for unknown words. It uses cosine similarity between the vectors. Equation (3.3) shows how cosine similarity is calculated as the dot product of the vectors normalized by their size.

$$cosine\_similarity(u, v) = \frac{u \cdot v}{||u|| \cdot ||v||} \tag{3.3}$$

3.3.4 Transformer Based Sentence Embedding

SentenceTransformers (Reimers & Gurevych, 2019) are state-of-the-art sentence embeddings that use siamese and triplet network structures. In (Reimers & Gurevych, 2019), the authors concluded that Sentence-BERT (SBERT) uses siamese and triplet network structures is 46.8k faster than cosine similarity to find the most similar pair. Transformers have their preprocessing step because, unlike other ML & DL algorithms, transformer models such as BERT, RoBERTa have their input type. While Word2Vec and fastText are at the word granularity level, sentence embeddings work at the sentence granularity level. A set of pre-trained models has been fine-tuned for various tasks. The embedding models that we used and experimented for the SuicideDetection and CEASEv2.0 datasets can be found in Table 4.1 and Table 4.3. In these tables, each column represents a distinct set of parameters for a model of that dataset.

3.4 Activation Functions

Activation functions are critical components of neural networks. They are used to add nonlinearity to the network, letting it to learn more complex correlations between input and output variables.

Many different activation functions are available, each with its advantages and disadvantages. Therefore, we must use them according to the model's task, where we can utilize them most to get better results. We have used the most common and most helpful activation functions in our models, such as relu, tanh, sigmoid, and softmax. The details about each of these functions are given in the following subsections.

3.4.1 ReLU

Rectified Linear Units (ReLU) (Agarap, 2018) output the input value if it positive; else, the result is 0 as it is seen in Figure 3.2. It is a very popular activation function, as it is computationally efficient and can help to prevent the vanishing gradient problem. The calculation as in equation (3.4), and we have used it along with the soft plus function, a smoothed version of ReLU, and its calculation is shown in equation (3.5).

$$f(x) = max(0, x) \tag{3.4}$$

$$f(x) = log(1 + e^{x}) \tag{3.5}$$

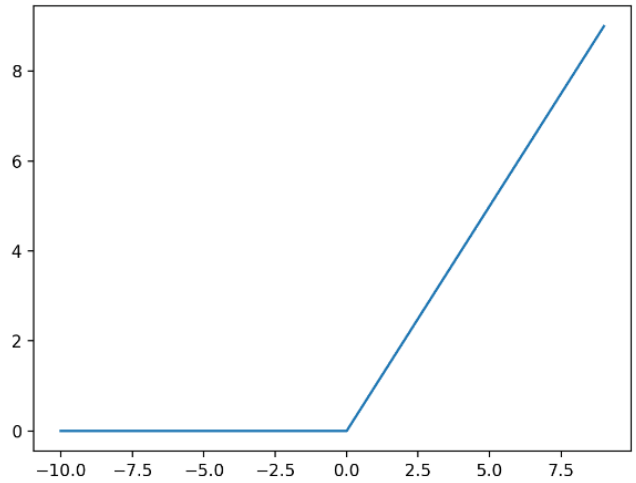

Figure 3.2 ReLU graph

3.4.2 Hyperbolic Tangent (Tanh)

Tanh (Namin et al., 2009) function is similar to the sigmoid function, but it has a range of [-1, 1]. This makes it more suitable for regression problems, as it can represent both positive and negative values. Calculation of tanh function is as shown in equation (3.6) and it's graph as in Figure 3.3.

$$f(x) = (e^{x} - e^{-x})/(e^{x} + e^{-x}) \tag{3.6}$$

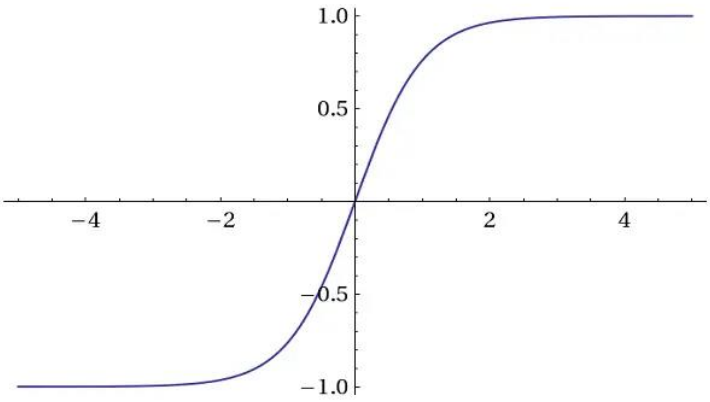

Figure 3.3 Tanh graph

3.4.3 Sigmoid

The sigmoid function (Narayan, 1997) is a S-shaped function that has a range of [0, 1]. It is often used in classification problems, as it can be understood as the likelihood of a particular class. We used this activation function in the binary classification task for suicidal ideation detection for two of the datasets. The formula and the graph of this function shown as in equation (3.7) and in Figure 3.4.

$$f(x) = 1/(1 + e^{-x}) \tag{3.7}$$

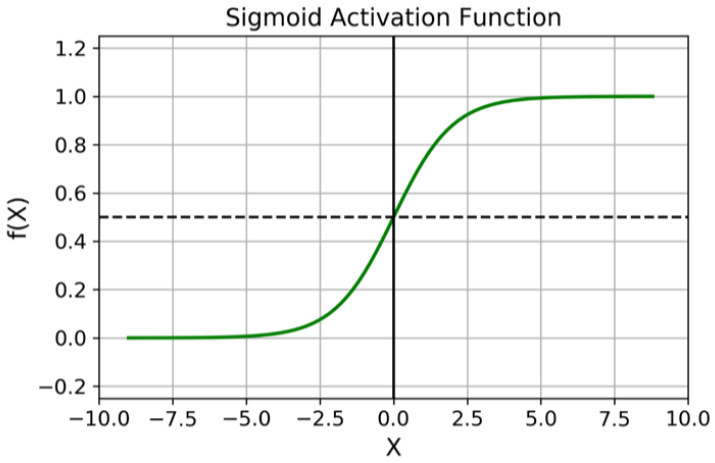

Figure 3.4 Sigmoid graph

3.4.4 Softmax

The softmax function (Bridle, 1990) is a non-linear function that is often used in the output layer of neural networks for multi-class classification problems as it converts a vector of real numbers to a vector of probabilities where the summation of these probabilities is 1. Calculation formula is given in equation (3.8). We have used this activation function for SWMH dataset which has 5 different class labels.

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{3.8}$$

3.5 Loss Functions

A loss function measures how well a model predicts the ground truth labels. It guides training of the model by minimizing the loss function. Their essential nuances are shown in Table 3.3.

Table 3.3 Differences of Binary cross entropy and Sparse categorical cross entropy

Feature	**Binary cross entropy**	**Sparse categorical crossentropy**
Number of classes	2	More than 2
Representation of ground truth labels	Single value (0 or 1)	Vector of integers
Use cases	Binary classification problems	Multi-class classification problems

Three different classification models have been built where two of them accomplish a binary classification and the third one does a multi-class classification. In the training phase, we used “binary-crossentropy” which is calculated in equation (3.9) for the binary classification, and “SparseCategoricalCrossentropy” which is calculated in equation (3.10) for the multi-class classification. The estimated labels for the SuicideDetection and CEASEv2.0 datasets are either ``suicidal'' or ``non-suicidal'' but for the SWMH dataset it is one of the five classes: ``depression'', ``suicidewatch'', ``anxiety'', ``offmychest'', and ``bipolar''. These class labels are also shown in Figure 3.1.

$$BCE(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{3.9}$$

$$SCCE(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{3.10}$$

where:

BCE is Binary Crossentropy,

$SCCE$ is Sparse Categorical Crossentropy,

N is the output size,

y is the true label value and

$\hat{y}$ is the predicted label value.

3.6 Callback Functions

Callbacks are utilities in ML to prevent overfitting. They can be used from Keras library. We used Early Stopping, Reducing the LR, and Checkpointer methods in our models. We explained these callback functions in the following section.

3.6.1 Early Stopping

Early stopping (Team, n.d.) stops training after a specific number of attempts if there is no further advancement in validation accuracy. Therefore, there would be a good use of time for training. It takes three parameters, such as monitor, patience and mode. Monitor is for metric to be checked if overfitted. The patience is the number of epoch that is allowed to train without improving the given metric parameter. Mode parameter decides whether the metric is monitored while increasing or decreasing. In our models, our metric was validation accuracy, patience could be between 3-8 and mode was max.

3.6.2 Reduce Learning Rate

Reduce Learning Rate (Team, n.d.) callback function can be also used by Keras library. Reducing the LR improves the accuracy of a model by allowing it to adjust the learning rate during training. It has three parameters, such as monitor, patience and factor. Monitor and patience is used as we explained for early stopping. The factor parameter specifies the reducing rate of learning rate, default is 0.1.

3.6.3 Model Checkpoint

Model checkpoint (Team, n.d.) callback function saves the model weights after each epoch by watching the model performance to save the best model if there is any interruption or crush. Thus, we could have pre-trained models for further usage, or even fine-tuning can be applied. It can take parameters such as file path for the place of saving the model, monitor for choosing the best weights, and save_best_only for whether saving all model weights or just best weights.

3.7 Classification

In the architecture of the suggested models, various latest word and sentence embedding layers and neural network layers have been used. Hyperparameter optimization is applied to each layer and general parameters of the model. Different optimizers such as Root Mean Square Propagation (RMSprop), Adam (Kingma & Ba, 2014), and Nadam (Dozat, 2016) have been used. The BERT model also uses an unmodified Adam optimizer.

In all datasets, the Adam optimizer was more successful in finding global minima for the current problem since it combines two gradient descent methodologies: momentum and RMSprop.

Relu activation function is used, primarily along with the soft plus function, as the activation function of the dense layers. Tanh activation function is used for BiLSTM layers with the relu function for the dense layers usually obtains promising results. Activation of the last dense layers differs in the proposed models as two models use the sigmoid function for the binary classification while the third one use the softmax function for multi-class classification.

Overfitting is a trap that trained models may fall into it. In this trap, while a model performs very well when applied to the train set, it shows a deficient performance on the test set (unseen data). To prevent this issue and generalize the proposed model, the dropout method is used with different dropout rates and batch normalization layers. The dropout layers randomly drop neurons from the network during training. In contrast, the batch normalization layers normalize the input data to each layer to reduce the internal covariate shift. For the SuicideDetection dataset, the dropout rate between 0.3-0.7 was used, where a 0.3 dropout rate achieved the best results. For the CEASEv2.0 dataset, this rate was between 0.18-0.25, and the 0.25 dropout rate achieved the best results. This difference is due to the dataset size; the SuicideDetection dataset is 46.4k times larger than the CEASEv2.0 dataset.

3.7.1 BiLSTM Networks

Proposed models for the SuicideDetection and CESEv2.0 datasets use BiLSTM (Graves & Schmidhuber, 2005) network. BiLSTM layers can capture both the forward and backward directions of the given context as shown in Figure 3.5, allowing for a better understanding of the sequence as a whole. Two layers of BiLSTM are used; for the first layer, a parameter called "return_sequences" is used as true because return_sequences returns the output of the hidden state for each time step of input, and the output of each time step is sent to the second BiLSTM layer.

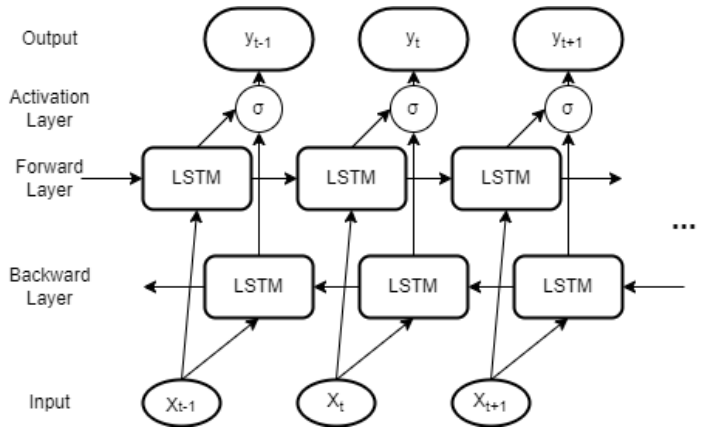

Figure 3.5 BiLSTM Network Structure

3.7.2 BERT Transformer

BERT Transformer (Devlin et al., 2018) for the SWMH dataset is used. There exists an attention mechanism in each block of the transformers. Attention mechanism allows the extraction of only the most related feature from a given input, which reduces the computational complexity of a model (Cohen & Maunsell, 2009).

Modern DNN based algorithms like transformers provide higher efficiency than traditional ML algorithms because they process substantial amounts of data in parallel and can automatically extract features from data without relying on hand-crafted features.

The most promising models have been trained with different input shape sizes. This value depends on the maximum text length of each document because as word plays its role as a feature and using all words as features results in better performance.

Different pre-processing steps may result in different input shapes. The best model for the SuicideDetection dataset has 188 input shape and is trained for 20 epochs with 128 batch sizes.

CEASEv2.0 dataset has the smallest data among other datasets and the best model for this dataset has 43 input shape and is trained for 10 epochs with 16 batch size. The model does not have batch normalization, because it does not have enough data to fall into the overfitting trap.

The model trained for the SWMH dataset has 256 input shape and is trained for 2 epochs. The total value of the trainable parameters is 108.706.565. The BERT itself has 768 dimensions. The summary of the SWMH model is given in Figure 3.6.

```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_ids (InputLayer)         [(None, 256)]        0           []

 attention_mask (InputLayer)    [(None, 256)]        0           []

 bert (TFBertMainLayer)         TFBaseModelOutputWi  108310272   ['input_ids[0][0]',
                                thPoolingAndCrossAt               'attention_mask[0][0]']
                                tentions(last_hidde
                                n_state=(None, 256,
                                 768),
                                 pooler_output=(Non
                                e, 768),
                                 past_key_values=No
                                ne, hidden_states=N
                                one, attentions=Non
                                e, cross_attentions
                                =None)

 intermediate_layer (Dense)     (None, 512)          393728      ['bert[0][1]']

 output_layer (Dense)           (None, 5)            2565        ['intermediate_layer[0][0]']

==================================================================================================
Total params: 108,706,565
Trainable params: 108,706,565
Non-trainable params: 0
__________________________________________________________________________________________________
```

Figure 3.6 SWMH Model Summary

CHAPTER 4

4. EXPERIMENTAL EVALUATIONS

4.1 Datasets

In this section, a comprehensive evaluation of the proposed method is provided. The details of the datasets, the evaluation metrics, baseline and settings, the results and comparison, and followed by a discussion, are given in the following subsections.

4.1.1 Suicide Detection Dataset

The Suicide Detection Dataset (Komati, 2021) is a collection of user posts on Reddit ("Suicide Watch" subreddit) that are publicly available on Kaggle ("Suicide and Depression Detection," 2021). The dataset consists of 232,074 posts on 'SuicideWatch' from December 16, 2008, to January 2, 2021. It has 116,037 suicide posts and 116,037 non-suicidal posts. The SuicideWatch subreddit refers to a monitoring procedure designed to prevent suicide attempts by individuals for those who display suicidal warning signals since they can be at risk for intentional self-harm. The dataset has mostly larger sentences and the overview of the dataset can be seen in Figure 4.1.

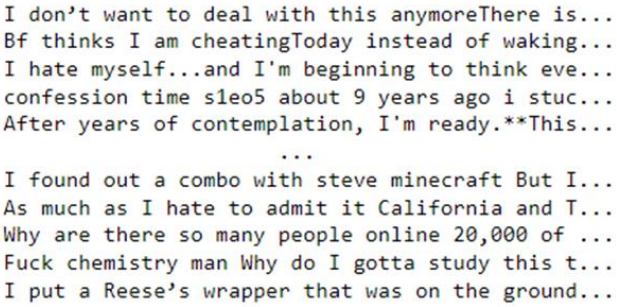

Figure 4.1 Overview of SuicideDetection Dataset

4.1.2 CEASEv2.0 Dataset

The CEASEv2.0 dataset (Ghosh et al., 2021) is the extended version of CEASE dataset (Ghosh et al., 2020) which is annotated with 15 fine-grained emotions at the sentence level of suicide notes in English, comprising 2393 sentences (from 205 suicide notes). With the addition of 2539 sentences to the base version, the dataset became to consist of 4932 sentences collected from a totally collected 325 suicide notes. This dataset mostly has shorter sentences than the rest of the datasets and the overview of the dataset can be seen in Figure 4.2.

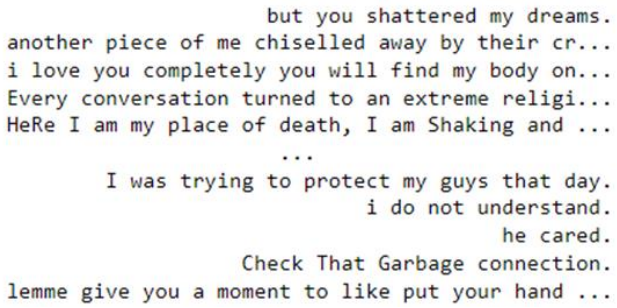

Figure 4.2 Overview of CEASEv2.0 Dataset

CEASEv2.0 is publicly available for research purposes. It is the most complex/challenging dataset among these three datasets because of the smaller data size and an unbalanced rate of labels.

4.1.3 SWMH Dataset

The SuicideWatch and Mental Health Collection called the SWMH dataset (Ji et al., 2021), is the collection of 54,412 posts specific to the subreddits of depression, suicidewatch, anxiety, offmychest and bipolar using the Reddit API. Unlike the other datasets that are used for binary text classification, this dataset is used for multi-class classification. This dataset is a multiclass dataset and has mix of shorter and longer sentences. The overview of the dataset can be seen in Figure 4.3.

```
Suicidal, but won't do it. Just need someone t...
A lie. A lie ruined my life A lie.\n\n\nA lie ...
Suicidal Thoughts and Venting I would first li...
Guided Meditation Disclaimer:  I am in no way ...
Effexor day 1 is going badly Good news! So the...
                              ...
Anxiety, Depressed, Obsessed. Torturing mental...
I could really use... Someone to talk to . I'm...
Just wanted to share a moment. I met this girl...
I wish I had someone. I'm tired of not having ...
Losing hair and on the brink Hi SW,\n\nI don't...
```

Figure 4.3 Overview of SWMH Dataset

4.2 Evaluation Metrics

In the experiments, efficiency of proposed methods are calculated according to the AUC score and the F1 score through the performance comparison. The AUC defines the model's general performance on a binary classification (positive and negative classes) task, as in (4.1). The example confusion matrix for binary classification is shown in Figure 4.4. F1 score measures the harmonic mean of precision and recall as in (4.4). The calculation of precision and recall is mentioned in (4.2) and (4.3), respectively. The AUC and the F1 scores provide a better overall measure of model performance than recall and precision alone. Since the SWMH dataset is evaluated for multi-class classification, by binarizing the output whether using One-vs-Rest or One-vs-One, the AUC scores of each class can be calculated. AUC score result can be examined depending on the value; 0.5-0.7 shows poor discrimination, 0.7-0.8 shows acceptable discrimination, 0.8-0.9 shows excellent discrimination and >0.9 shows outstanding discrimination.

$$AUC = \frac{1}{n}\sum_{i=1}^{n}(TP_i + 0.5 \times FN_i) \tag{4.1}$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{4.2}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{4.3}$$

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{4.4}$$

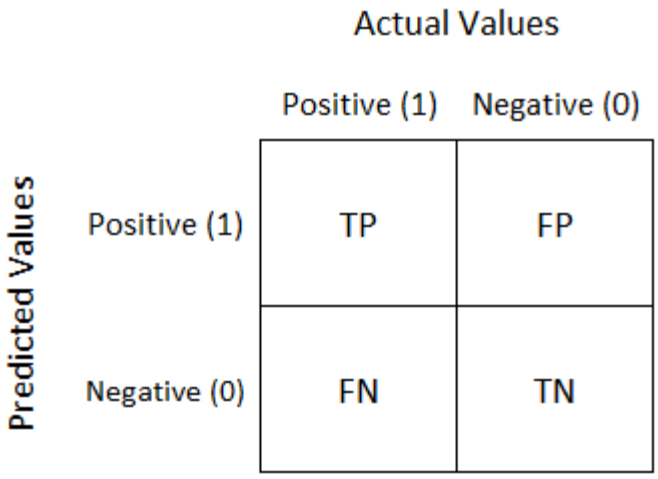

Figure 4.4 Confusion Matrix for Binary Classification

4.3 Results and Comparison

We have developed three distinct models, each trained and tested on separate datasets. We improved the performance of the state-of-the-art models on two of the three datasets. In the following subsections, results for each dataset and detailed comparisons of the models are given.

4.3.1 SuicideDetection

The SuicideDetection dataset is experimented with ML and DL algorithms, the comparison of experiments can be seen in Table 4.2 and Table 4.1. The most effective model for the Suicide Detection dataset is built with deep learning algorithms. The created model has ten layers with different parameter values, as in Table 4.1. The order of layers is shown in Figure 4.6 and as follows: Embedding + Dropout (0.3) + 2BiLSTM (128 units) + Dense (32 units) + Dropout (0.3) + Dense (16 units) + Dropout (0.3) + Batch Normalization + Dense (1 unit). FastText (crawl-300d-2M-subword.bin) is used for the weights of the embedding layer. The dataset's size is needed for a deep-layered model to get greater performance.

The model achieved a 0.97 F1 score from training for 20 epochs with 128 batch sizes as in Figure 4.5. Our suggested model outperformed the most advanced model's performance of 0.95 accuracy (Aldhyani et al., 2022) based on CNN-BiLSTM along with Word2Vec. Besides the F1 score, the AUC score of our model is 0.996. Additionally, our validation loss is 0.08, as in Figure 4.5, while (Aldhyani et al., 2022) has its model's validation loss of 0.15.

Table 4.1 Details of Experimented DL models for SuicideDetection dataset.

Layer & Parameters	**Models for SuicideDetection Dataset**								
Data Split	%80 Training - %20 Test								
Input Shape	75	100	100	191	191	188	200	184	184
Embedding	GloVe	GloVe	GloVe	GloVe	fastText	fastText	BERT	Sentence	Sentence
Embedding Model		glove.840B.300d			crawl-300d-2M-subword		bert-base-uncased	all-MiniLM-L6-v2	all-mpnet-base-v2
Embedding Size	300	300	300	300	300	300	768	384	768
Network	BiLSTM	BiLSTM	2BiLSTM	2BiLSTM	2BiLSTM	2BiLSTM		2BiLSTM	2BiLSTM
Activation	relu	relu	relu	relu	relu	relu		relu	relu
Dropout Rate	0.5	0.5	0.5	0.7	0.5	0.3		0.3	0.3
Flatten	✓	✓	✓	✓	✓	-	✓	✓	✓
Batch Normalization	-	-	-	✓	✓	✓		✓	✓
Epoch	10	5	5	15	10	20	2	20	20
Batch	256	256	128	256	128	128	64	128	128
Optimizer	RMSprop	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Callbacks									
Early Stopping	✓	✓	✓	✓	✓	✓		✓	✓
Reduce LR	✓	✓	✓	✓	✓	✓		✓	✓
Checkpointer	✓	✓	✓	✓	✓	✓		✓	✓
F1-score (%)	96	96	96	97	97	**97.42**	96.38	94.69	94.77
AUC Score (%)	-	-	-	-	96.7	**99.6**	-	-	

Table 4.2 Comparison and evaluation of ML models for SuicideDetection dataset

ML Algorithms	Score (%)			
	Accuracy	Precision	Recall	F1 score
Multinomial Naive Bayes	56.32	16.42	78.52	27.16
Random Forest	81.71	82.23	81.14	81.68
Linear Regression	70.78	71.53	70.78	70.88
Support Vector Machine	93.21	93.31	93.21	**93.21**
Decision Tree	82.97	83.94	82.97	83

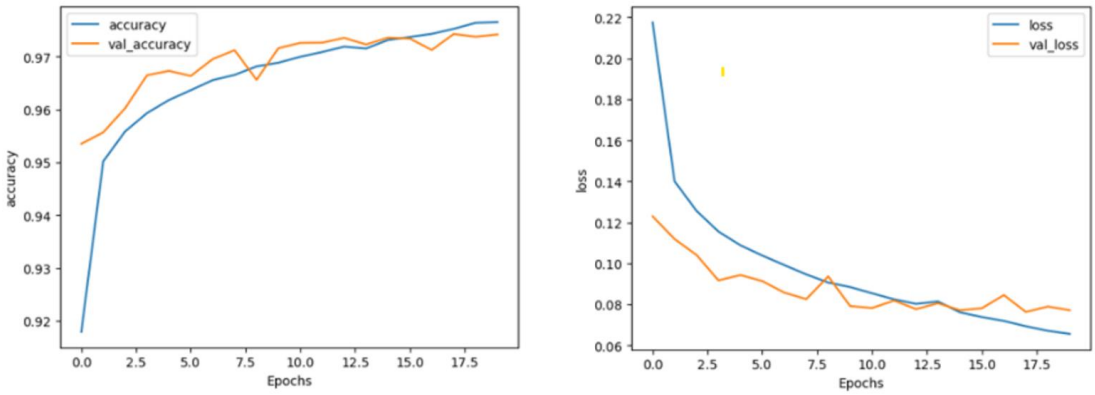

Figure 4.5 Accuracy & Loss Graph of Best Proposed Model for SuicideDetection dataset

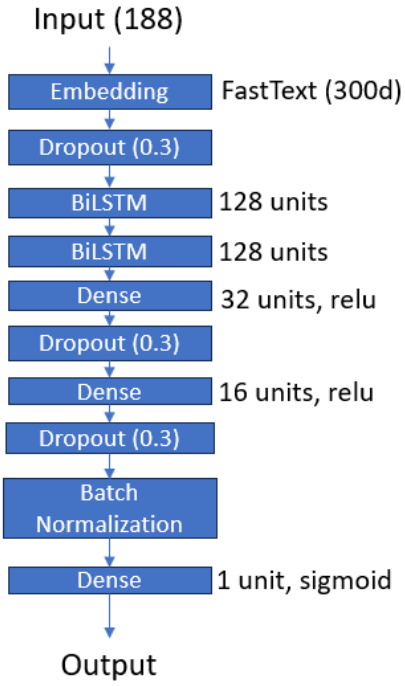

Figure 4.6 Proposed Model Architecture for SuicideDetection Dataset

4.3.2 CEASEv2.0

CEASE dataset is the most complex/challenging dataset among the three datasets due to less data size and an unbalanced label rate. It has features that make creating a better model harder. In the preprocessing part, words are corrected depending on the sentence context to decrease the disadvantage of dataset size. The created model has seven layers with different parameter values as in Table 4.3, and the order of layers is shown in Figure 4.8 and as follows: Embedding + BiLSTM (32 units) + Dense (8 units) + Dropout (0.25) + Dense (4 units) + Dropout (0.25) + Batch Normalization + Dense (1 unit). Depending on the size of the dataset, a model has been built that is shallower than the others. Training of the model was for 10 epochs with 16 batch sizes, and additional bias initializer is used in the model due to the unbalanced rate of the labels. Based on performance metrics, our model has achieved a 0.75 F1 score, as in Figure 4.7, and a 0.70 AUC score, as in Table 4.3. The score achieved for our model outperformed the best model of (Ghosh et al., 2021), which has a 0.7435 F1 score with GLoVe + Bi-GRU approach.

Table 4.3 Details of Experimented DL models for CEASEv2.0 dataset.

Layer & Parameters	Models for CEASEv2.0 Dataset	
Data Split	%80 Training - %20 Test	
Input Shape	44	43
Embedding	Sentence	fastText
Embedding Model	all-mpnet-base-v2	crawl-300d-2M-subword
Embedding Size	768	300
Network	2BiLSTM	BiLSTM
Activation	softplus	relu
Dropout Rate	0.18	0.25
Flatten	-	-
Batch Normalization	✓	✓
Epoch	42	10
Batch	265	32
Optimizer	Nadam	Adam
Callbacks		
Early Stopping	✓	✓
Reduce LR	✓	✓
Checkpointer	-	-
F1-score (%)	68	**75**
AUC Score (%)	-	**70**

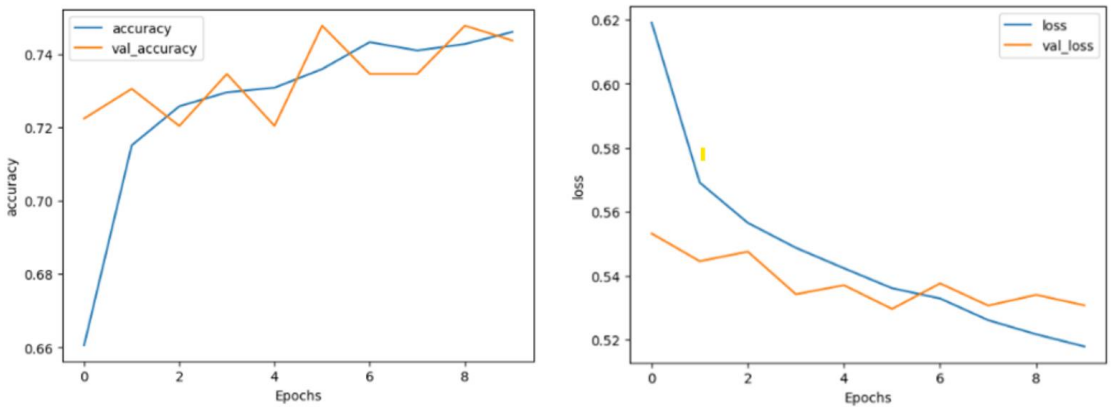

Figure 4.7 Accuracy & Loss Graph of Best Proposed Model for CEASEv2.0 dataset

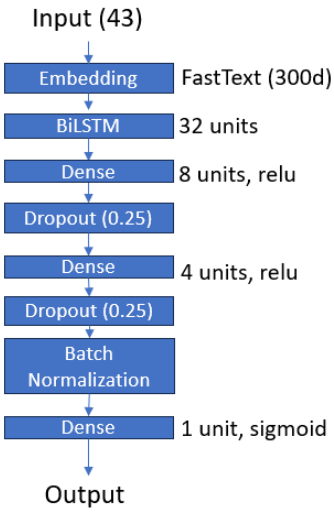

Figure 4.8 Proposed Model Architecture for CEASEv2.0 Dataset

4.3.3 SWMH

SWMH dataset was tested and compared with ML algorithms in Table 4.4 and Deep Learning algorithms in Table 4.5. The proposed model architecture for SWMH dataset is shown in Figure 4.9. The architecture consists of attention mask layer, base BERT with 12 blocks, Dense layer (512 units) and last dense layer (5 units). Each block of BERT has structure as follows: multi-head attention + add & norm + feed forward + add & norm. Add represents the residual connections and norm means layer normalization. BERT also uses dropout in this layer. The best F1 score is achieved using a BERT transformer trained for 2 epochs with batch size 16, and the model performed a 0.68 F1 score. The model outperformed the model in (Ji et al., 2021), which is built with RN and achieved a 0.64 F1 score. Best accuracy is 0.72 F1 score using MentalRoBERTa model as in (Ji et al., 2021). That is one of the transformer models that was fine-tuned using Reddit posts about mental health. That transformer is more domain-specific, and because of this reason, training the model by using it gives better results.

Table 4.4 Comparison and evaluation of ML models for SWMH dataset

ML Algorithm/Features	F1-Score (%)				
	One-hot Encoding	TF-IDF	TF-IDF+n-grams	TF-IDF + char ngram	Pad sentences (fastText word embedding)
Multinomial Naive Bayes	62	58	52	57	-
Linear Regression	62	**67**	53	65	-
K-Nearest Neighbors	45	52	25	49	-
Random Forest	58	62	50	59	-
Stochastic Descent + early stopping	60	66	52	65	21
Gradient Boosting + early stopping	60(Count Vectors)	61	46	62	35
XGBoost Classifier	66 (Count Vectors)	66	53	65	40

Table 4.5 Comparison and evaluation of DL models for SWMH dataset

	FastText Word Embedding	
DL Algorithms	F1-score (%)	AUC Score (%)
Shallow Network	66	86.8
DL(Just Embedding)	64	86.3
DL(Emb.+2 Dense)	63	86.4
RNN	55	-
CNN	60	-
LSTM	65	86.5
CNN+LSTM	64	87.3
CNN+GRU	63	85.8
RNN+GRU	64	85.3
2BiLSTM	62	84.2
Bidirectional GRU	64	85.0
RCNN	66	87.5
RCNN-v2	66	**87.8**
RCNN-v3	65	86.5
BERT	**68**	-

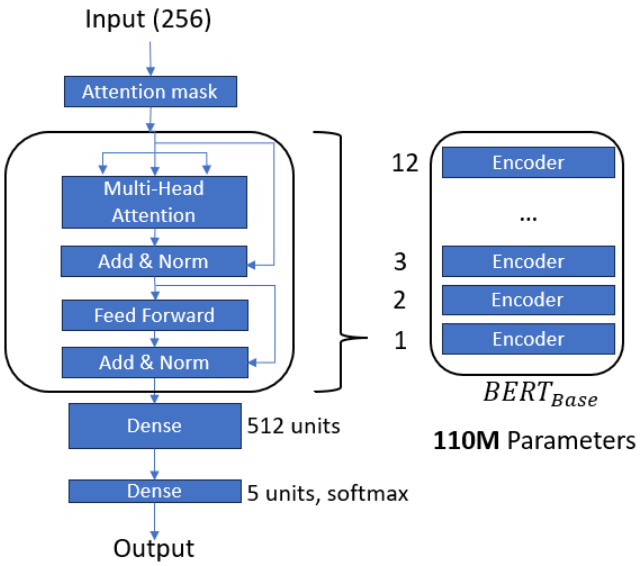

Figure 4.9 Proposed Model Architecture for SWMH Dataset

4.4 Summary

Table 4.6 Best proposed models for each of the dataset

Dataset	Size	Description	Best Method (Approach)	Results
Suicide Detection	232.074 documents	Reddit data (%50-%50)	FastText Word Embedding+2Layer BiLSTM (same with also attention layer)	0.97 F1-score 0.996 AUC score
CEASE-v2.0	4.932 suicide notes	%64 non-depresision, %36 depression	FastText Word Embedding + BiLSTM + word correction + initial bias	0.75 F1-score 0.70 AUC score
SWMH[1]	54.410 documents	Reddit data (multi-class) Balanced-dataset	BERT	0.68 f1-score

[1] SWMH dataset has five different class labels.

We have designed and tested different models, the details of which are listed in Table 4.1, Table 4.3, Table 4.4 and Table 4.5. Finally, we selected the most efficient model for each dataset which is shown with details in Table 4.6.

CHAPTER 5

5. DISCUSSION

A discussion of data and obtained results, as well as the limitations of the proposed methods are given in these sections.

5.1 Data Analysis

Data is the first-factor affecting model performance, as NLP applications are generally data-driven. SuicideDetection dataset has a balanced rate of suicidal & non-suicidal data. The SWMH dataset has five different classes, and the class weight is calculated using the compute_class_weight method in the Sklearn library, as shown in (5.1). Class weights labels are presented in Figure 5.1. The ratio of the dataset is balanced (ratio > 0.1) with the value of 0.408 (depression label=0.5805) / (bipolar label=1.4234)) as calculated in (5.2). Also, the distribution of class rates of SWMH can be seen in Figure 5.2.

$$w_j = \frac{n\_samples}{n\_classes * n\_samples_j} \tag{5.1}$$

where:

w_j is the weight for each class (j represents the class),

$n\_samples$ represents total number of rows in the dataset,

$n\_classes$ represents the total number of unique classes in the target, and

$n\_samples_j$ represents the total number of rows of the respective class.

$$ratio = \frac{np.\min(df.label.value\_counts())}{np.\max(f.label.value\_counts())} \quad (5.2)$$

```
Class weight: 1.1389    class: self.Anxiety
Class weight: 1.0688    class: self.SuicideWatch
Class weight: 1.4234    class: self.bipolar
Class weight: 0.5805    class: self.depression
Class weight: 1.3139    class: self.offmychest
```

Figure 5.1 Weight rates of each label on SWMH

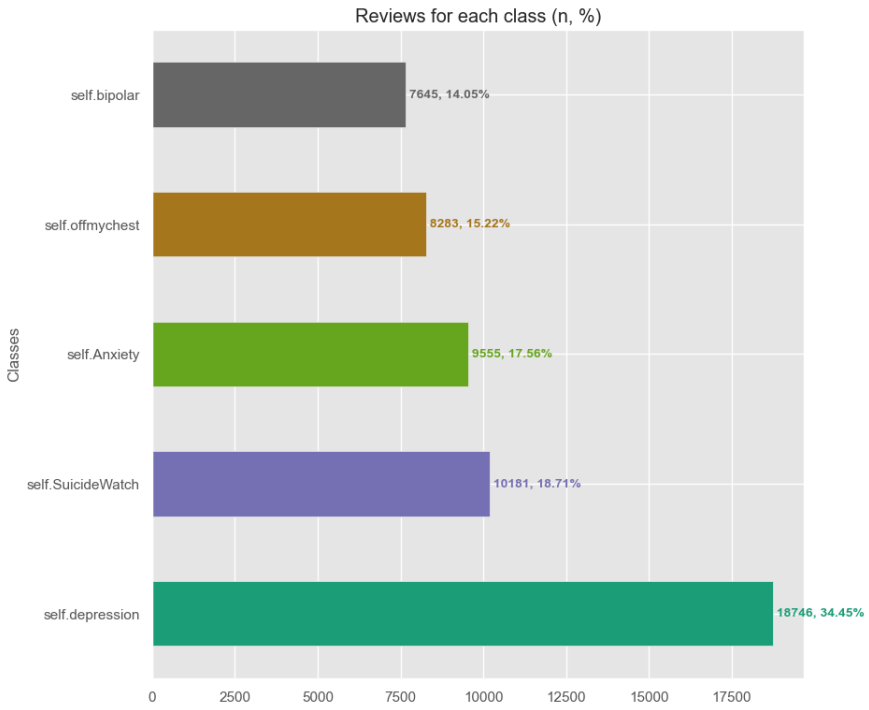

Figure 5.2 Distribution rates of each class label for SWMH

The CEASEv2.0 dataset is an imbalanced dataset, with 64% non-depression data and 36% depression data. Due to this imbalance, it is necessary to take extra steps, such as initial bias settings, when training the dataset.

The training process may vary depending on the size of the data. In the case of a larger dataset, it is necessary to use a deeper network to improve pattern recognition and achieve optimal feature extraction. This is because a larger dataset can contain more complex patterns and features that a shallower network may not be able to recognize.

As previously mentioned, traditional ML techniques that utilize hand-crafted features can be combined with DL approaches. However, it is often necessary to incorporate more than just hand-crafted features in order to gain a complete understanding of the context and patterns within a given dataset.

In Table 4.1 and Table 4.5, the performance of DL algorithms, and ML algorithms with handcrafted features are compared for the SuicideDetection and SWMH datasets. While ML algorithms have been used for a long time, advanced NLP techniques using DL algorithms have the ability to extract features more efficiently. Additionally, the use of word embeddings enhances the encoding process and contributes significantly to better model performance. Together, these factors explain why proposed models outperform others (Aldhyani et al., 2022; Ghosh et al., 2021) for the SuicideDetection and CEASEv2.0 datasets.

The model for the SuicideDetection dataset has ten layers, while the BERT Transformer model for SWMH has twelve encoder blocks. The shallowest model is for CEASEv2.0 has seven layers.

The words 'feel', 'like', 'want', 'think', 'life', 'really', 'time', friend', 'help', 'day', 'live', and 'end' are some of the most common words for these datasets. These keywords could give a general idea of individuals about themselves and their intentions. The 50 most frequent words in the datasets are compared, and it is seen that shared words are related to self-harm, hopelessness, sadness, and fear. The source of these feelings may bring suicidal ideation out of people. Most common used words of datasets are visualized as WordClouds, shown in Figure 5.3, Figure 5.4 and Figure 5.5.

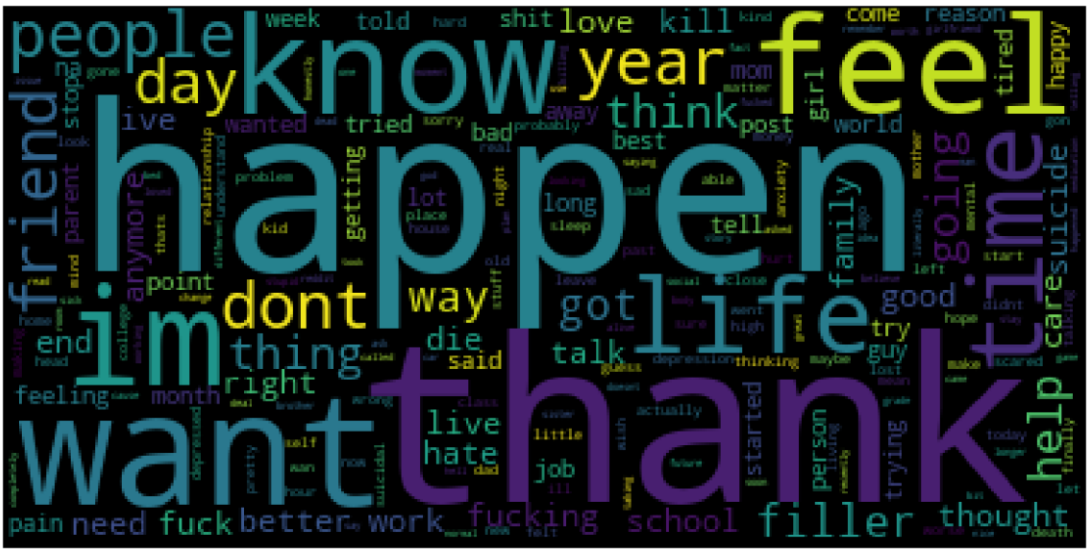

Figure 5.3 WordCloud of SuicideDetection dataset

Figure 5.4 WordCloud of CEASEv2.0 dataset

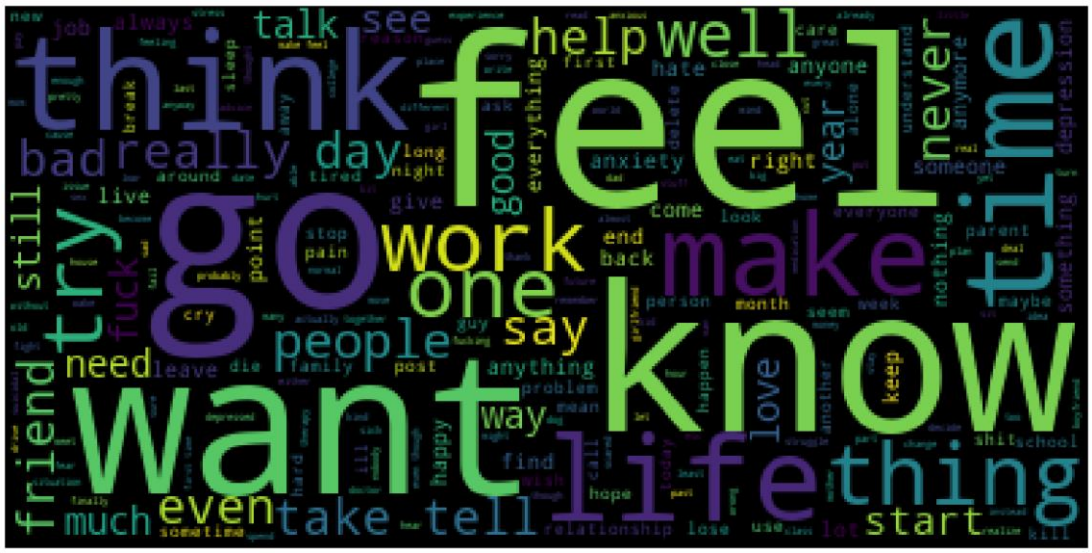

Figure 5.5 WordCloud of SWMH dataset

We have used three different embedding techniques. Word2Vec and FastText are based on word embeddings and sentence embedding is transformer-based sentence embedding. In Table 5.1, similarity values for different word pairs are shown according to Word2vec, FastText, and Sentence embedding. For instance; the word pair suicide-happy has similarity values of 0.20, 0.14, and 0.32 respectively from Word2vec, FastText, and Sentence Embedding. We expect to get a small value since semantically the relationship between this word pair should be far from each other. FastText gives similar results for the word pair of suicide-forgive. We can conclude from the values in Table 5.1 and model performances that FastText is more stable and its values are more valuable.

Table 5.1 Similarity review of different embedding techniques

Word Pairs	Word2Vec	FastText	Sentence Embedding
	glove.840B.300d.pkl	crawl-300d-2M-subword.bin	all-mpnet-base-v2
(suicide-happy)	0.20	**0.14**	0.32
(family-love)	0.44	0.39	0.41
(suicide-forgive)	0.26	**0.21**	0.26
(understand-forgive)	0.52	0.50	0.34
(suicide-death)	0.67	0.65	0.75
(suicide-life)	0.39	0.41	0.57
(love-life)	0.59	0.44	0.48
(end-suicide)	0.32	0.30	0.27
(time-suicide)	0.22	0.18	0.31

Individuals who are struggling with suicidal thoughts may tend to seek ways to express themselves and connect with others, while those without such thoughts may not have the same motivation to seek support or share their struggles. Suicidal individuals' struggle can also be understood from Figure 5.6 and Figure 5.7. Figures show that suicidal texts have a higher value of word count, sentence count, average sentence length, and unique word count than non-suicidal texts.

```
word_count          43.125460 word_count          138.657230
sent_count           2.384041 sent_count            6.984608
avg_sentlength      23.239315 avg_sentlength       46.962059
unique_word_count   30.619035 unique_word_count    87.921921
```

Figure 5.6 Sentence Features of SuicideDetection Dataset (Non-suicidal - Suicidal)

```
word_count          7.978964 word_count          11.055524
sent_count          0.987441 sent_count           1.012593
avg_sentlength      8.070517 avg_sentlength      10.979870
unique_word_count   7.442700 unique_word_count   10.137951
```

Figure 5.7 Sentence Features of CEASEv2.0 Dataset (Non-suicidal – Suicidal)

5.2 Limitations

Suicidal ideation detection through social media has several limitations. We used datasets from the Reddit subreddit platform and suicidal notes. First, social media content is limited regarding information about an individual's mental health and may not always accurately reflect a person's suicidal thoughts or intentions. Furthermore, self-reported data from social media may be subject to biases and may not accurately represent an individual's true feelings or experiences. Even individuals may write sarcastic sentences that can ultimately cause misunderstanding.

Additionally, the demographic structure of the users, such as race, gender, age, and socioeconomic status, may not be easily inferred from social media data, which can limit the model's applicability to specific populations.

Moreover, the temporal aspect of posts on social platforms can make to detect the immediate risk of suicide harder, as posts may be made hours or days before the attempt.

Two models perform binary classification, while one model performs multi-class classification. Therefore, another challenge is the severity of suicidal ideation due to the risk of binary classification, where the model can only classify suicidal or non-suicidal categories without indicating the level of danger.

Lastly, there are ethical implications of using DL models to detect suicidal ideation, such as ensuring data privacy.

Despite these challenges, proposed models based on DL & Transformers can still play a valuable role in identifying patterns and trends in suicidal ideation detection on social platforms, but better to combine with state-of-the-art approaches and expertized knowledge to guarantee precise and prompt interventions.

CHAPTER 6

6. CONCLUSION AND FUTURE WORK

6.1 Conclusion

This study aimed to develop three suicidal ideation detection models for three diverse datasets, two from the Reddit platform and one from suicidal notes using ML, DL, and Transformer algorithms. Motivation comes from the urgent need to detect and prevent self-harm among people, particularly those who express suicidal thoughts online or in written notes. This task of NLP applications needs to perform better than other applications due to the direct relationship between the result and human life.

Proposed models of this study outperformed the state-of-the-art in two out of three datasets. Our obtained performance in the Suicide Detection and CEASE-v2.0 datasets in terms of F1 score are 0.97 (0.9742 accuracy) and 0.75, outperforming the state-of-the-art accuracy of 0.95 and 0.74 F1 scores, respectively. However, our model did not achieve state-of-the-art performance on the SWMH dataset, with an F1-score of 0.68 ranking the second after 0.72 F-score of (Ji et al., 2021). Depending on the class-labels of the dataset, Our models accomplished binary and multi-class classification.

The findings underscore the importance of adopting a multifaceted approach when identifying suicidal ideation. Both transformer and DL models, such as BERT and BiLSTM, yielded encouraging results in our experiments. These advanced techniques enable the processing complex patterns and relationships within the textual data.

The current research approves the the potential of NLP, ML, and DL methods to be used for suicide prevention. These techniques can be specifically used by psychologists to detect those social media users who have the potential of self-harm.

6.2 Future Work

As NLP continues advancing, we can expect further advancements in word embedding techniques, leading to even more refined and nuanced language representations.

The current study focused solely on Reddit data, and future work could expand the application and have a dataset consisting of a combination of textual, visual, video, or audio content from other social platforms to predict suicidal ideation.

Although there are privacy concerns and limitations to using the data of individuals, including other personal differences such as age, gender, location, and interests could improve predictions.

Most prior studies on detecting suicidal ideation relied on analyzing data from the textual context for many individuals. However, exploring and analyzing the individuals followed over time could increase the understanding of the nature of suicidal ideation and the various impact of personal characteristics.

The proposed models are trained for the dataset collected in English. Data from different languages and cultures could be combined to increase the models' generalizability.

A large language model (LLM) (Zhao et al., 2023) is a type of artificial intelligence (AI) system that is trained on a large corpus of text data to generate natural language text. The size and diversity of the training data could be expanded and focused on collecting and curating larger and more diverse datasets of suicidal language to improve the accuracy and reliability of LLM-based detection models. Also, incorporating domain knowledge from psychology, sociology, and psychiatry experts into LLMs could increase performance in detecting suicidal ideation. Furthermore, LLM-based detection models could be deployed in real-world applications.

REFERENCES

Abdulsalam, A., & Alhothali, A. (2022). Suicidal Ideation Detection on Social Media: A Review of Machine Learning Methods. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2201.10515

Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (ReLU). *arXiv (Cornell University)*. Retrieved from https://arxiv.org/pdf/1803.08375.pdf

Aldhyani, T. H. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. a. T. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International Journal of Environmental Research and Public Health*, *19*(19), 12635. https://doi.org/10.3390/ijerph191912635

Ansari, L., Ji, S., Chen, Q., & Wang, Z. (2023). Ensemble Hybrid Learning Methods for Automated Depression Detection. *IEEE Transactions on Computational Social Systems*, *10*(1), 211–219. https://doi.org/10.1109/tcss.2022.3154442

Backlinko. (2023, March 27). Backlinko. Retrieved from https://backlinko.com/reddit-users

Benton, A., Mitchell, M., & Hovy, D. (2017). Multi-Task Learning for Mental Health using Social Media Text. *arXiv (Cornell University)*. Retrieved from https://arxiv.org/pdf/1712.03538.pdf

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Bridle, J. S. (1990). Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In *Springer eBooks* (pp. 227–236). https://doi.org/10.1007/978-3-642-76153-9_28

Chaffey, D. (2023, June 7). Global social media statistics research summary 2023 [June 2023]. Retrieved from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Chung, C. K., & Pennebaker, J. W. (2012a). Linguistic Inquiry and Word Count (LIWC). In *IGI Global eBooks* (pp. 206–229). https://doi.org/10.4018/978-1-60960-741-8.ch012

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*(12), 1594–1600. https://doi.org/10.1038/nn.2439

Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2015). CLPsych 2015 Shared Task: Depression and PTSD on Twitter. *Association for Computational Linguistics*, 31–39. https://doi.org/10.3115/v1/w15-1204

Coppersmith, G., Leary, R. B., Crutchley, P., & Fine, A. B. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights*, *10*, 117822261879286. https://doi.org/10.1177/1178222618792860

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv (Cornell University)*. Retrieved from https://arxiv.org/pdf/1810.04805v2

Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. *International Conference on Learning Representations*, 1–4. Retrieved from https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ

E. Agbe (JCharis), J. (n.d.). Retrieved from https://jcharis.github.io/neattext/

Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2020). CEASE, a Corpus of Emotion Annotated Suicide notes in English. *Language Resources and Evaluation*, 1618–1626. Retrieved from http://dblp.uni-trier.de/db/conf/lrec/lrec2020.html#GhoshEB20

Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2021). A Multitask Framework to Detect Depression, Sentiment and Multi-label Emotion from Suicide Notes. *Cognitive Computation*, *14*(1), 110–129. https://doi.org/10.1007/s12559-021-09828-7

Grant, R. N., Kucher, D., Leon, A. M., Gemmell, J., Raicu, D., & Fodeh, S. J. (2018). Automatic extraction of informal topics from online suicidal ideation. *BMC Bioinformatics*, *19*(S8). https://doi.org/10.1186/s12859-018-2197-z

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Ji, S., Li, X., Huang, Z., & Wang, Z. (2021). Suicidal ideation and mental disorder detection with attentive relation networks. *Neural Computing and Applications*, *34*(13), 10309–10319. https://doi.org/10.1007/s00521-021-06208-y

Ji, S., Yu, C. P., Fung, S., Pan, S., & Long, G. (2018). Supervised learning for suicidal ideation detection in online user content. *Complexity*, *2018*, 1–10. https://doi.org/10.1155/2018/6157249

Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2021). MentalBERT: Publicly available Pretrained Language Models for Mental Healthcare. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2110.15621

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1412.6980

Komati, N. (2021). Retrieved from https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch.

Kondrak, G. (2005). N-Gram similarity and distance. In *Lecture Notes in Computer Science* (pp. 115–126). Springer Science+Business Media. https://doi.org/10.1007/11575832_13

Lai, S., Liu, K., He, S., & Zhao, J. (2016). How to generate a good word embedding. *IEEE Intelligent Systems*, *31*(6), 5–14. https://doi.org/10.1109/mis.2016.45

Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. *arXiv (Cornell University)*. Retrieved from https://arxiv.org/pdf/cs/0205028

Losada, D. E., & Crestani, F. (2016). A test collection for research on depression and language use. In *Lecture Notes in Computer Science* (pp. 28–39). https://doi.org/10.1007/978-3-319-44564-9_3

Mikolov, T., Chen, K., Corrado, G., & Dean, J. M. (2013). Efficient estimation of word representations in vector space. *arXiv (Cornell University)*. Retrieved from http://export.arxiv.org/pdf/1301.3781

Moulahi, B., Azé, J., & Bringay, S. (2017). DARE to Care: A Context-Aware Framework to Track Suicidal Ideation on Social Media. In *Springer eBooks* (pp. 346–353). https://doi.org/10.1007/978-3-319-68786-5_28

Namin, A. H., Leboeuf, K., Muscedere, R., Wu, H., & Ahmadi, M. (2009). Efficient hardware implementation of the hyperbolic tangent sigmoid function. *2009 IEEE International Symposium on Circuits and Systems*. https://doi.org/10.1109/iscas.2009.5118213

Narayan, S. (1997). The generalized sigmoid activation function: Competitive supervised learning. *Information Sciences*, *99*(1–2), 69–82. https://doi.org/10.1016/s0020-0255(96)00200-9

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Association for Computational Linguistics*, 1532–1543. https://doi.org/10.3115/v1/d14-1162

Pirina, I., & Çöltekin, Ç. (2018). Identifying Depression on Reddit: The Effect of Training Data. *Association for Computational Linguistics*, 9–12. https://doi.org/10.18653/v1/w18-5903

Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges. *2019 International Engineering Conference (IEC)*, 200–204. https://doi.org/10.1109/iec47844.2019.8950616

Ramírez-Cifuentes, D., Freire, A., Baeza-Yates, R., Puntí, J., Bravo, P. M., Velazquez, D. A., Gonfaus, J. M., Gonzàlez, J. (2020). Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *Journal of Medical Internet Research*, *22*(7), e17758. https://doi.org/10.2196/17758

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. https://doi.org/10.18653/v1/d19-1410

Richardson, L. (2007). Beautiful soup documentation. April.

Sawhney, R. S., Joshi, H., Gandhi, S., & Shah, R. R. (2020). A Time-Aware Transformer Based Model for Suicide Ideation Detection on Social Media. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7685–7697. https://doi.org/10.18653/v1/2020.emnlp-main.619

Shah, F. M., Haque, F., Nur, R. U., Jahan, S., & Mamud, Z. (2020). A Hybridized Feature Extraction Approach To Suicidal Ideation Detection From Social Media Post. *2020 IEEE Region 10 Symposium (TENSYMP)*. https://doi.org/10.1109/tensymp50017.2020.9230733

Shing, H., Nair, S., Zirikly, A., Friedenberg, M., Daumé, H., & Resnik, P. (2018). Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 25–36. https://doi.org/10.18653/v1/w18-0603

Sinha, P. P., Mishra, R., Sawhney, R. S., Mahata, D., Shah, R. R., & Liu, H. (2019). #suicidal - A Multipronged Approach to Identify and Explore Suicidal Ideation in Twitter. *The 28th ACM International Conference*, 941–950. https://doi.org/10.1145/3357384.3358060

Stravynski, A., & Boyer, R. (2001). Loneliness in Relation to Suicide Ideation and Parasuicide: A Population-Wide Study. *Suicide and Life Threatening Behavior*, *31*(1), 32–40. https://doi.org/10.1521/suli.31.1.32.21312

Suicide and depression detection. (2021, May 19). Retrieved from https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. *Algorithms*, *13*(1), 7. https://doi.org/10.3390/a13010007

Team, K. (n.d.). Keras documentation: EarlyStopping. Retrieved from https://keras.io/api/callbacks/early_stopping/

Team, K. (n.d.-b). Keras documentation: ModelCheckpoint. Retrieved from https://keras.io/api/callbacks/model_checkpoint/

Team, K. (n.d.-b). Keras documentation: ReduceLROnPlateau. Retrieved from https://keras.io/api/callbacks/reduce_lr_on_plateau/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is All you Need. *arXiv (Cornell University)*, *30*, 5998–6008. Retrieved from https://arxiv.org/pdf/1706.03762v5

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. V., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., Alexander, R. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

World Health Organization (WHO) (2021). Retrieved from https://www.who.int/news-room/fact-sheets/detail/suicide.

World Health Organization (WHO) (2023). Retrieved from https://www.who.int/data/gho/data/themes/mental-health/suicide-rates.

Xu, S., Shijia, E., & Xiang, Y. (2020). Enhanced attentive convolutional neural networks for sentence pair modeling. *Expert Systems With Applications*, *151*, 113384. https://doi.org/10.1016/j.eswa.2020.113384

Zhang, L., & Moldovan, D. (2019). Multi-Task learning for semantic relatedness and textual entailment. *Journal of Software Engineering and Applications*. https://doi.org/10.4236/jsea.2019.126012

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J., Wen, J. (2023). A survey of large language models. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2303.1822

RESUME

Publications

[1] Ezerceli, Ö., & Eskil, M. T. (2022, November). Convolutional Neural Network (CNN) Algorithm Based Facial Emotion Recognition (FER) System for FER-2013 Dataset. In 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-6). IEEE.

[2] Gümüşçekiçci, G., Ezerceli, Ö., & Tek, F. B. (2020, September). Web service translating content into Turkish Sign Language. In 2020 5th International Conference on Computer Science and Engineering (UBMK) (pp. 355-259). IEEE.