Suicidal Prediction Using Video, Audio, And Text Analysis

Article *in* Cyber-Physical Systems · September 2022

DOI: 10.1109/ICPS55917.2022.00015

Gagan Chordia
VIT University

Jayakumar Kaliappan
VIT University

Suicidal Prediction Using Video, Audio, And Text Analysis

Gagan Chordia
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
chordiagagan@gmail.com

Yash Mehta
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
yash.mehta2019@vitstudent.ac.in

S Nayan
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
nayans.2019@vitstudent.ac.in

Harshit Soni
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
harshit.soni2019@vitstudent.ac.in

Aayush Gupta
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
aayush.gupta2019@vitstudent.ac.in

Jayakumar K
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
jayakumar.k@vit.ac.in

***Abstract*— In today's society, suicide has evolved into a legitimate social medical issue. Suicidal ideation refers to people's plans to commit suicide. Suicide ideas and attempts can be triggered by a variety of factors, including long-term exposure to negative emotions or life events. Early detection of suicidal ideation is possibly the most effective method of suicide prevention of all. People can share their sufferings and sentiments in real-time thanks to developments in internet communication and interpersonal connection benefits, which provides a source for suicide thought identification. Suicide has a variety of causes, and suicidal variables vary from person to person.**

Keywords*— *Deep Learning, Natural Language Processing, Speech Recognition, Sentiment Detection

I. Introduction

Every year, 703,000 people commit suicide, according to the World Health Organization, with many more attempting suicide. Every suicide is a tragedy that affects entire families, towns, and countries, as well as the survivors. Suicide affects people of all ages, and in 2019, it was the world's fourth highest cause of death among 15–29-year-olds.

Suicide is a severe public health issue, but it can be avoided by implementing timely, evidence-based, and frequently low-cost treatments. For effective national responses, a comprehensive multi-sectoral suicide prevention strategy is required.

Suicide prevention efforts, such as the installation of safety barriers at the Golden Gate Bridge, have been hampered by engineering practise; however, technological advancements offer new possibilities for suicide prevention and the potential to reduce the number of suicide deaths.

Making a model that recognises sentiments based on facial and text expressions is not a novel concept, but putting it into practise in a way that can provide solutions to real-world problems is. Many researchers, including Lee S et al[4], Reddy P et al[16], and others, have proposed such innovations. For example, Lee S et al[4] used only facial patterns as a variable, whereas Ververidis, D., & Kotropoulos, C.[15], and Reddy P et al[16] used only vocal patterns as a classifier. As a result, taking into account all variables becomes a major reason to classify such a tense situation.

II. Literature Review

Larsen M, et al[1] provide an overview about the current technologies developed which are doing research in the field of suicide prevention. Machine learning algorithms collect data from twitter and label them as strongly concerning, possibly concerning or safe to ignore. Shen, D., et al[2] have used Feed-Forward Neural Networks, Fine-Tuning Deep Models for Target Tasks and Convolutional Neural Networks. Author developed an algorithm to handle images from different scanning protocols. Patel, K., et al[3] used algorithms for face detection are Viola-Jones Face Detection, Principal Component Analysis and Linear Discriminant Analysis. Local Binary Pattern, DCT, Scale Invariant feature transform and histogram of oriented gradients are used for Feature Extraction. Machine learning algorithms used for Emotion classification are CNN, SVM, RNN, Artificial Neural Network and Deep Belief Network. Lee, S., et al[4] used a random forest, ensemble tree-based classification algorithm, and it works on images from a 3D camera. The information about human features such as face, hands, legs etc is obtained from OpenNI. The data is sent to random forest classifier to classify it into suicidal or non-suicidal. Ker, J., et al[5] have used different supervised and unsupervised models. Supervised Models- convolutional neural networks, transfer learning with CNN and recurrent neural networks. Unsupervised Learning Models- autoencoders, restricted boltzmann machines and deep belief net-works and generative adversarial networks.

Chen, J., et al[6] have attached a texture prediction module to traditional CNN. This calculates inner products from multiple convolution layers. This inner product is used to obtain affective vectors. Bernert, R. A., et al[7] used supervised learning techniques including ensemble learning

978-1-6654-7022-3/22/$31.00 ©2022 IEEE
DOI 10.1109/ICPS55917.2022.00015

methods, random forests, decision trees, naive Bayes classification, SVM logistic/least squares regression. Onie, S., et al[8] conducted a search for reviewing patterns using various different keywords “suicide,” “cctv,” and “video” on PubMed, Inspec, and Web of Science. Malhotra, A., & Jindal, R[9] proposed an algorithm which would be able to: (i) process films in real time, (ii) recognise people and objects, and (iii) classify videos into one of two categories: self-harming / suicidal or not. The videos are processed with three successive modules, keeping the above tasks in mind.

Shatte, A. B., et al[10] covered the domain and were found in eight health and information technology research databases. Two reviewers evaluated the articles, and data on the article's mental health application, machine learning technique, data type, and study findings were extracted. T. M. Fonseka, et al.[11] conducted a literature search in the OVID Medline, EMBASE, and PsycINFO databases to aid them in developing artificial intelligence to predict suicide risk. Ji, S., et al[12] used face-to-face interaction, as well as spoken and audio information. Suicide notes were written messages left by persons before they take their own lives. They're frequently written in letters and posted on websites, and they're usually captured in audio or video.

Abdel-Hamid, et al[13] recognized speech to evaluate the effectiveness of CNN in two ways, one TIMIT and the second one was using CNN. The speech was analysed and feature vectors were generated to apply different deep learning algorithms. Graves, A., et al[14] followed the idea of RNN on TIMIT Corpus on Phoneme recognition by applying Fourier transformations and then they were mapped to the respective class based on the scores. Ververidis, D., & Kotropoulos, C.[15], used the utterance of a speech as a measure to evaluate the performance and cross-validation of emotion. It was further divided into 2 categories as a sequence of short time prosody and their statistics. Reddy, P. P., et al[16] had a base idea to detect emotion based on speech notes and to label them as normal or abnormal. Then to follow a top-down approach to optimise and remove any sort of disturbance, formulate it to get the accuracy. Juang, B. H., & Rabiner, L. R.[17], proposed a model based on statistical methods to evaluate, estimate, decode and recognize speech apart from the traditional method by modelling, sizing, training and recognizing the sequences.

The use of natural language processing, hierarchical and non-hierarchical machine learning techniques to predict the complexity of texts used in iSTART, a reading comprehension tutoring system, was proposed by Balyan, R., et al[18]. The classification is based on linguistic characteristics like the FKGL, syntactic complexity score, uncommon or rare words, lexicons, word familiarity, and acquisition age. BERT, according to Devlin, J., et al[19], only uses the blocks of the Transformer's encoders in a novel way and does not use the decoder stack. Transformer models benefit from BERT's bidirectional attention. BERT introduces unsupervised embedding scenarios, as well as pre-training models with unlabeled text. During the attention learning process, this forces their model to think and train harder. This makes it easier for their model to understand how these are constructed, and they can apply this knowledge to downstream tasks without having to pretrain each time. Vajjala, S.[20] provided a brief overview of recent research on Automatic Readability Assessment (ARA). This survey's data was gathered from textbooks, ARA research-trained authors, web content materials, and crowd-sourcing experiments and user studies. They gathered evaluation and validation objectives from other papers and used them to judge models. Mukherjee, P., et al[21] used lexical chains, which are groups of semantically related words in a document, to predict the complexity score proposed in the paper. The classification system is based on the fact that more complex sentences have a greater number of synonymous chains, whereas simple sentences have fewer but longer chains. Nadeem, F., & Ostendorf, M.[22], estimated the complexity level of the text passages using four different models. Sequential RNN comes first, followed by hierarchical RNN, hierarchical RNN multi-head attention, and finally bidirectional hierarchical RNN. GRUs are used by all of them.

Qiu, X., et al. [23] proposed a deep learning model with syntactic and semantic dense embeddings based on linguistic features like percentage of conjunctions and average parse tree height. To deal with feature relationships, the authors create a correlation graph among features and use it to learn their embeddings, so that similar features are represented by similar embeddings. V. Santucci et al. [24] present an automatic classification system for determining the level of complexity of an Italian text from a linguistic standpoint. Measuring the complexity of any task presented using a supervised learning classification problem by deriving a dataset of texts purposefully produced by linguistic experts for second using twitter scraping of any person's thought. Ma, Y., et al[25] proposed a model that uses an SVM model to implement a modified version of a standard ranking algorithm. To begin, a levelling classification is created by ranking the books and then assigning a level to each one based on the results of the ranking. To train the model and generate a prediction for a new book, features are extracted from the training set. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) deep learning models were combined with the word2vec model to produce word embeddings by Sawhney, R., et al[26].

Du, J., et al[27] used a multi-step pipeline to collect suicide-related tweets from Twitter streaming data. The authors then applied the concept of Recurrent Neural Networks (RNN) to improve the transfer learning strategy. L. Zheng et al.[28] applied machine learning technology and deep neural network algorithms to data from an electronic health record (EHR) and interpreted the results. Cusick, M., et al[29] had a basic idea to create a large manually annotated dataset using weakly supervised machine learning methods and then apply various machine and learning algorithms such as logistic classification, SVM, CNN, and the Naive Bayes classifier. Using data from many suicide notes, Last Statements, and neutral posts from mental health domain experts and healthcare centres, Zhang, T., et al. proposed a transformer-based model called TransformerRNN and a Bi-directional LSTM (Bi-LSTM), which includes an input embedding, transformer encoder, Bi-LSTM, a max-pooling layer, and finally a classification layer.

III. Methodology

A. Proposed Architecture

First, we will be collecting audio, video, text, and images of a person. Then we will process this collected data to give input and predict the suicidal tendencies of a person. Collected data will be preprocessed using standard techniques of computer vision and natural language processing. Then important features will be extracted for input in the pre-trained model for suicide prediction. [Fig 1]

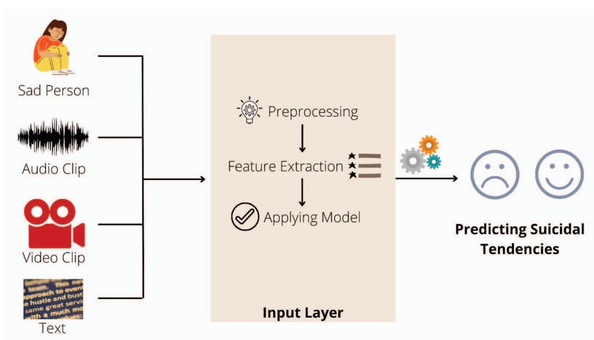

Fig1. Proposed Architecture Diagram

B. Detailed Algorithm

Collecting Audio, Video, Images, and text of a person to analyse and preprocess the data. In the preprocessing part, the audio is converted into text and the text is tokenized using the nltk package. The tokenized text is then matched with a dataset containing words with suicidal tendencies. Cosine similarity is used as a metric to check how similar the text is to the dataset. If the cosine similarity is greater than 0.1. The same algorithm can be used to analyse the text part also. For images, we use the deepface library to analyse the attributes of the face. The library analyzes different attributes like age, gender, dominant_emotion, dominant_race. We can use these attributes to find the dominant emotion and using the emotion we can find whether the person is sad, angry, etc. The dominant emotion can be used if a person is suicidal or not. Videos are groups of frames or images, hence we can use the same algorithm used for images and videos to find the dominant algorithm.

C. Module Description

1. Image Recognition

1.1 Data Collection

The collection of data can be done through data-rich websites like Kaggle. However, this can also be achieved using personal cameras. Thereby, creating a unique dataset.

1.2 Data Preprocessing

All the grey scale content in the images are converted to RGB scale to detect different features (like, emotions, facial patterns and structure), and to normalise all the values for the respective features.

1.3 Feature Detection & Extraction

Deep-face being a powerful library can detect and analyse emotions as well as other physical features from the sample image.

1.4 Analysing Facial Attributes and Landmarks

Several tags can be used such as “Sad”, “Happy”, “Fear”, “Angry”, “Disgust”, “Surprise”, “Neutral”, as shown in Table 1.

Table 1: Emotion analysis table

Analysis of Emotion	Analysis Score
Angry	2.274358831346035
Disgust	0.0027545289412956
Fear	9.727241843938828
Happy	2.5050161056583e-08
Neutral	0.2127377316355705
Sad	87.46248483657837
Surprise	0.3204204607754946

1.5 Predicting Dominant Motion

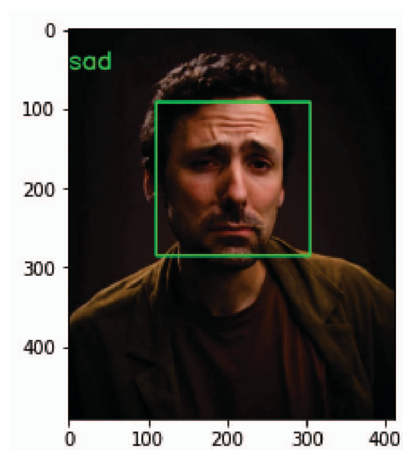

Fig 2. Emotion Detection

A motion is detected whenever there is a change in the pixels of the image. Predicting a dominant motion involves a series of actions. A “Hi” motion of a hand is completely different from choking the neck. Deep-face library is incredibly efficient and accurate in predicting such gestures.

2. *Video Recognition*

2.1 Data Collection

Data can be collected by using data-rich platforms such as Kaggle. Personal cameras, on the other hand, can be used to accomplish this. As a result, a unique dataset is created.

2.2 Data Preprocessing

Data is preprocessed by converting each video frame to an array of images and going through the same processes as image preprocessing as mentioned above.

2.3 Analysing Facial Attributes frame by frame

Deep-face being a powerful library can detect and analyse emotions as well as other physical features from the sample image.

2.4 Detecting Emotions based on Attributes

When there is a change in the pixels of an image, motion is recognized. A set of acts are required to predict a dominant motion. A hand motion that says "Hi" is not the same as choking the neck. In predicting such gestures, the deep-face library is highly efficient and accurate. Similarly, a frown face indicates a sad emotion.

3. *Speech Recognition for Suicide Prevention*

3.1 Converting Speech to text-

So using the microphone of various devices to detect the audio and converting it to speech using Zero Crossing Rate or Entropy of Energy Autocorrelation and then getting it in for Speech Recognition.

3.2 Preprocessing the converted speech to analyse further-

In order to correct the textual data of the model structure, we make pre-processing text. The first step in NLP projects. Some of the pre-processing steps are: Deleting similar punctuation marks. ,! $ () *% @ , Releases URLs, Deleting stop words ,Bottom bag, Making tokens, Stemming, Lemmatization, Applying tokenization & Removing Stopwords.

3.3 Sentiment Model-

Using various algorithms like Naive Bayes, Linear SVM, VADER to show the accuracy of different algorithms used for sentiment analysis. Polarity generated ranges from 0 to 1 according to the audio frequency and is quoted based on cosine similarity.

4. *Text Recognition for Suicide Prevention*

4.1 Collecting dataset of suicide prone words and sentences

Clinical methods based on interactions between social workers or professionals and targeted individuals, as well as machine learning techniques using feature engineering or in-depth automated discovery learning based on online social content, are currently used to detect suicidal ideation.

4.2 Feature Engineering

Text refinement and modification are the first steps in the processing process. Different measures are used to measure length. Determine whether the text is favourable or negative using emotional analysis. Named Business Identification: tags text with preset categories including people's names, organisations' names, and places' names. Find the most essential n-grams by looking at their frequency. Convert names to numbers using name vectors. Remove essential topics from the corpus as a model topic.

4.3 Generating keywords & Word Cloud for Each emotion.

The word cloud is generated by the keywords of the student's emotions. Related colours and moods: Angry (Red), Happy (Green), Warm (Orange), Worried (Brown), Depressed (Dark Blue), Boring (Magenta), Odd (Light Blue).

4.4 Calculating cosine similarity between dataset and data collected.

$$\mathrm{Cos}(x, y) = x \cdot y \,/\, ||x|| * ||y||$$

where, x.y = product dot of the vectors ‘x’ and ‘y’.

||x|| and ||y|| = length of the two vectors ‘x’ and ‘y’.

||x|| * ||y|| = cross product of the two vectors ‘x’ and ‘y’.

Table 2: Similarity and Polarity of text

Measure	Output (0 - 1)
Similarity	0.8142
Polarity	0

4.5 Predicting suicidal tendency

Using all the methods we applied, we will predict using all the measures to detect suicide.

IV. Results & Conclusion

In this paper, we’ve described how we can help one to prevent committing suicide by analyzing their natural features to detect the possibility for the same. We’ve taken into account the main features of speech and communication like video, audio and text messages to predict whether a person is a victim or not.

Table 3: Comparison between some methodologies

Methodology	Prediction based on Video Frames	Prediction based on Audio	Prediction based on Text
Lee S, et al [4]	✓	✗	✗
Ververidis, D., & Kotropoulos, C. [15]	✗	✓	✗
Shawney R, et al [26]	✗	✗	✓
Our method	✓	✓	✓

We've observed a better result with use of multiple features when compared to one attribute being used to detect a possible suicide. We've attained a result of ~94% when all the features were taken into account. In addition, we are proposing an innovative method of using deep-face algorithms over Artificial Neural Networks(ANNs).

Future works include developing an Application Program Interface(API), to integrate it with applications that are used by the users on a daily basis such that the API can send an SOS signal for help to organisations that are invested in the field.

Acknowledgment

We are grateful to Prof. Jayakumar for guiding us throughout the research, helping us to attain effective results and motivating us throughout the work.

References

[1] Larsen, M. E., Cummins, N., Boonstra, T. W., O'Dea, B., Tighe, J., Nicholas, J., ... & Christensen, H. (2015, August). The use of technology in suicide prevention. In *2015 37th annual international conference of the IEEE engineering in Medicine and biology society (EMBC)* (pp. 7316-7319). IEEE.

[2] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*, 221-248.

[3] Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N., & Alazab, M. (2020). Facial sentiment analysis using AI techniques: state-of-the-art, taxonomies, and challenges. *IEEE Access*, *8*, 90495-90519.

[4] Lee, S., Kim, H., Lee, S., Kim, Y., Lee, D., Ju, J., & Myung, H. (2014). Detection of a suicide by hanging based on a 3-D image analysis. *IEEE sensors journal*, *14*(9), 2934-2935.

[5] Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. *Ieee Access*, *6*, 9375-9389.

[6] Chen, J., Mao, Q., & Xue, L. (2020). Visual sentiment analysis with active learning. *IEEE Access*, *8*, 185899-185908.

[7] Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *International journal of environmental research and public health*, *17*(16), 5929.

[8] Onie, S., Li, X., Liang, M., Sowmya, A., & Larsen, M. E. (2021). The use of closed-circuit television and video in suicide prevention: narrative review and future directions. *JMIR mental health*, *8*(5), e27663.

[9] Malhotra, A., & Jindal, R. (2020). Multimodal deep learning based framework for detecting depression and suicidal behaviour by affective analysis of social media posts. *EAI Endorsed Transactions on Pervasive Health and Technology*, *6*(21), e1.

[10] Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, *49*(9), 1426-1448.

[11] Fonseka, T. M., Bhat, V., & Kennedy, S. H. (2019). The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian & New Zealand Journal of Psychiatry*, *53*(10), 954-964.

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

[13] Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, *22*(10), 1533-1545.

[14] Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.

[15] Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication*, *48*(9), 1162-1181.

[16] Reddy, P. P., Suresh, C., Rao, V. K., Chandana, K. S., Sowkya, S., & Akhila, R. (2021). Vocal Analysis to Predict Suicide Tendency. In *Proceedings of International Conference on Advances in Computer Engineering and Communication Systems* (pp. 481-488). Springer, Singapore.

[17] Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, *33*(3), 251-272.

[18] Balyan, R., McCarthy, K. S., & McNamara, D. S. (2020). Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education*, *30*(3), 337-370.

[19] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[20] Vajjala, S. (2021). Trends, limitations and open challenges in automatic readability assessment research. *arXiv preprint arXiv:2105.00973*.

[21] Mukherjee, P., Leroy, G., & Kauchak, D. (2018). Using lexical chains to identify text difficulty: a corpus statistics and classification study. *IEEE journal of biomedical and health informatics*, *23*(5), 2164-2173.

[22] Nadeem, F., & Ostendorf, M. (2018, June). Estimating linguistic complexity for science texts. In *Proceedings of the*

thirteenth workshop on innovative use of NLP for building educational applications (pp. 45-55).

[23] Qiu, X., Chen, Y., Chen, H., Nie, J. Y., Shen, Y., & Lu, D. (2021). Learning Syntactic Dense Embedding with Correlation Graph for Automatic Readability Assessment. *arXiv preprint arXiv:2107.04268*.

[24] Santucci, V., Santarelli, F., Forti, L., & Spina, S. (2020). Automatic classification of text complexity. *Applied Sciences*, *10*(20), 7285.

[25] Ma, Y., Fosler-Lussier, E., & Lofthus, R. (2012, June). Ranking-based readability assessment for early primary children's literature. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 548-552).

[26] Sawhney, R., Manchanda, P., Mathur, P., Shah, R., & Singh, R. (2018, October). Exploring and learning suicidal ideation connotations on social media with deep learning. In *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 167-175).

[27] Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., & Xu, H. (2018). Extracting psychiatric stressors for suicide from social media using deep learning. *BMC medical informatics and decision making*, *18*(2), 77-87.

[28] Zheng, L., Wang, O., Hao, S., Ye, C., Liu, M., Xia, M., ... & Ling, X. B. (2020). Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Translational psychiatry*, *10*(1), 1-10.

[29] Cusick, M., Adekkanattu, P., Campion Jr, T. R., Sholle, E. T., Myers, A., Banerjee, S., ... & Pathak, J. (2021). Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. *Journal of psychiatric research*, *136*, 95-102.

[30] Zhang, T., Schoene, A. M., & Ananiadou, S. (2021). Automatic identification of suicide notes with a transformer-based deep learning model. *Internet interventions*, *25*, 100422.