

# Diagnosis Application of MINI and BERT

Subjects: **Others** | **Health Care Sciences & Services**

Contributor: Ridha Mezzi

Researchers propose a mental health diagnosis application for Arabic-speaking patients using both The MINI-International Neuropsychiatric Interview (MINI) and the supervised machine learning BERT model to equip the psychiatry department of the Military Hospital of Instruction of Tunis with a rapid and intelligent tool handling the high number of patients treated every day.

mental health

psychiatry

MINI

intent recognition

BERT model

natural language processing

machine learning

## 1. Background

### 1.1. Mental Health Issues

At some point, every person feels disturbed, anxious, or even downhearted, leading to significant mental health issues. According to the ICD-11 <sup>[1]</sup> as a reference for mental health illnesses, the identification of a mental health illness is often related to mental problems disturbing thinking ability, relation with others, and day-to-day behaviors.

Several mental issues have been studied. Most of these studies focused mainly on schizophrenia, stress disorder, depression, bipolar disorder, obsessive-compulsive disorder, and others <sup>[2]</sup>.

Accordingly, mental issues are global problems and equal opportunity issues. They equally affect the young and the old, male or female, every race, ethnic group, and different education and income levels.

There has been considerable talk about how new settings generate a depression pandemic <sup>[3]</sup>. More than 200 million people of different age categories are suffering depression according to WHO (World Health Organization) <sup>[4]</sup>. In addition, up to 20% of children and teenagers worldwide suffer from mental health issues. Moreover, they are believed to impact one out of every four people at some point in their lives. As a result, issues such as alcohol and substance addiction, abuse, and gender-based violence emerge and influence mental health. Thus, failing to treat mental health has ramifications for entire civilizations. These disturbing statistics indicate the widespread incidence of mental illness. The good news is that they are frequently treatable.

Although symptoms of mental illness might manifest themselves physically in the form of stomach pain and back pain in some people, other symptoms do not include physical pain; among these symptoms are the following:

- Feeling down for awhile.
- Severe mood fluctuations.
- Avoiding contact with family and friends.
- Decreasing energy or difficulty sleeping.
- Feeling enraged, aggressive, or violent regularly.
- Having hallucinations, hearing unreal voices, or feeling paranoid.
- Having thoughts about ending their lives or death.

Hence, consulting a therapist and sticking to a regular treatment plan that may involve medication can help people with mental illnesses feel better and reduce their symptoms.

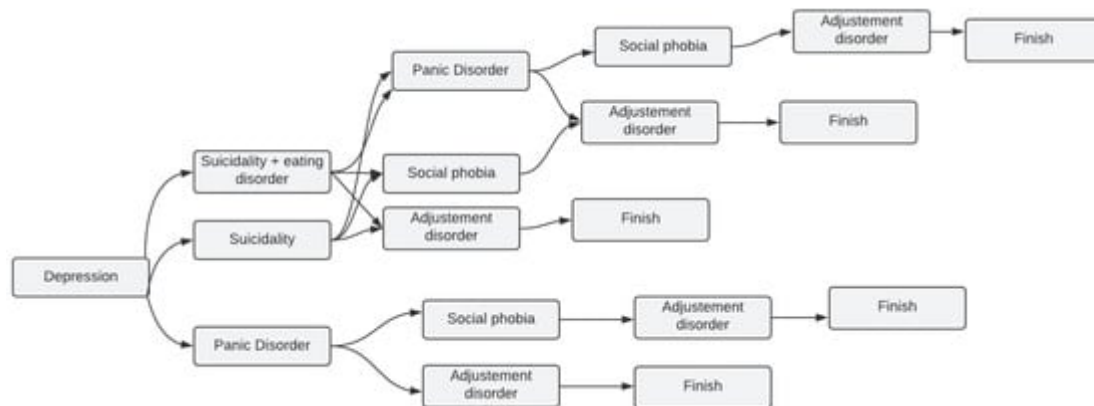
## 1.2. The MINI-International Neuropsychiatric Interview (MINI) International Neuropsychiatric Interview

Performing a psychiatric interview necessitates a set of questions to ask the patient. “MINI PLUS” (MINI International Neuropsychiatric Interview) [\[5\]](#) is a commonly used psychiatric structured diagnostic interview instrument that requires “yes” or “no” responses, and it is divided into modules where each module has a set of questions to ask the patient. The questions are labeled with letters that relate to diagnostic groups. For instance, module 1 checks if the patient has a depressive episode using questions A1a, A1b, A2a, A2b, etc. (e.g., question A1a: Have you been consistently depressed or down, most of the day, nearly every day, over the past two weeks?). According to the patient's answers, there is a specific diagnosis at the end of each module. It is considered a gold standard for AI because it is very structured, and all its questions are known. Researchers can easily predict all of their answers even if they are other than “yes” and “no”. Accordingly, it made it easy to build current dataset to test and train the machine learning models. The interview works on many modules such as depression, suicidality, panic disorders, social phobia, etc. However, in current case, trying to make the interview shorter than a real one, researchers tested the five most frequent and common disorders or modules according to the WHO [\[6\]](#), which are depression, suicidality, panic disorder, social phobia, and adjustment disorder and tried to implement them in current application.

### Different Scenarios While Taking the Test with MINI

In current application, researchers did not take the MINI by letter. Instead, researchers took the modules researchers needed to work with and tried to implement them as researchers could see fit with the majority of cases researchers are dealing with, which led to some common scenarios with only a few disorders among the five

mentioned in [Section 2.2](#). After all, it is rare that a person has all the diseases at once. **Figure 1** depicts all the possible scenarios a patient might go through while taking the application test from start to finish.



**Figure 1.** Most common scenarios of depression using MINI in current system.

The following is an example of how the scenario might unfold when using the MINI modules in current system:

- According to the patient's answers, while taking the depression test, it turned out that the patient needed suicidality and eating disorder test, the trial went to the next module, which is, in current case, suicidality and eating disorder.
- In some other cases, if the patient from the beginning of the test is not identified as having the issue related to the tested module (one of the five modules), the test jumps to start the next one. For instance, if the patient from the beginning does not have panic disorder by answering the first two questions of the module with “No” or “Not”, according to the MINI, the test jumps to the social phobia test.

### 1.3. AI and ML in Healthcare and Mental Health

While many different fields of society are willing to embrace the potential of AI, caution remains deep in medical areas, among which are psychiatry and mental health, proven by recent headlines in the news such as “Warnings of a Dark Side to AI in Health Care” [\[7\]](#). Psychiatry is a promising area for the use of AI, though, despite the claimed worries, AI implementations in the medical field are progressively expanding. Thus, researchers are compelled to apprehend its present and future applications in mental health and work intelligently with AI as it enters the clinical mainstream.

### 1.4. NLP with BERT

NLP is the automated manipulation of natural language by software, such as speech and text. It has been studied for more than 50 years, and it sprang from the discipline of linguistics as computers became more prevalent. Most works use convolutional neural networks (CNN) and recurrent neural networks (RNN) to achieve NLP functions [\[8\]](#). A novel architecture [\[9\]](#) has advanced existing classification tasks by using deep layers that are commonly used in computer vision to perform text processing. They concluded that adding more depth to the model would improve its

accuracy. It was the first time deep convolutional networks have been used in NLP, and it has provided insight into how it can help with other activities. Opinion mining, also known as sentiment analysis, is another used field. It is a primary method for analyzing results. For text preprocessing, NLP strategies are checked, and opinion mining methods are studied for various scenarios [10]. Human language can be learned, understood, and generated using NLP techniques. Speaking conversation networks and social media mining are examples of real-world applications [11]. As the purpose of this paper is a mental health diagnosis system for Arabic-speaking patients, researchers will deal with Arabic text (in Tunisian dialect “Darija”) from right to left, which makes this model the perfect choice to achieve the ultimate results. After all, researchers are dealing with a medical condition where high-quality results cannot be less important.

BERT is a multilingual transformer-based ML technique for NLP pre-training developed by Google [12]. It has sparked debate in the ML field by showing cutting-edge findings in a wide range of NLP tasks, such as question answering, natural language inference, and others. The transformer’s bidirectional training to language modeling is the cornerstone of BERT’s technological breakthrough, which contains two distinct mechanisms: an encoder that reads text input and a decoder that generates a task prediction. Only the encoder technique is required because BERT’s objective is to create a language model. In contrast, previous research has focused on text sequences from the left to the right or the right to the left (directional models). However, the transformer encoder scans the complete word sequence in one go. Accordingly, it is considered bidirectional, although it is more appropriate to be described as nondirectional. This feature enables the model to divine the context of a word from its surrounds (on the left and the right of the word) and gain a superior understanding of language context and flow more than single-direction language models.

## 2. Mental Health Proposed Diagnosis System for Arabic-Speaking Patients

### 2.1. Purpose and Global Architecture

current goal was to develop a system that records patient responses to questions during a medical interview (e.g., “I am desperate, and I have no hope in life”). The proposed system provides a detailed psychological diagnosis report of the patient’s condition (e.g., major depressive episode, moderate depressive episode, suicidal, etc.).

**Figure 2** provides a global idea of how the system works. It is divided into two layers: (1) in the visualization layer, the patient interacts with the system via a graphical user interface, and (2) in the processing layer, all his interactions as responses to questions are stored in a database and processed by intent recognition module to generate the final result describing his medical state.

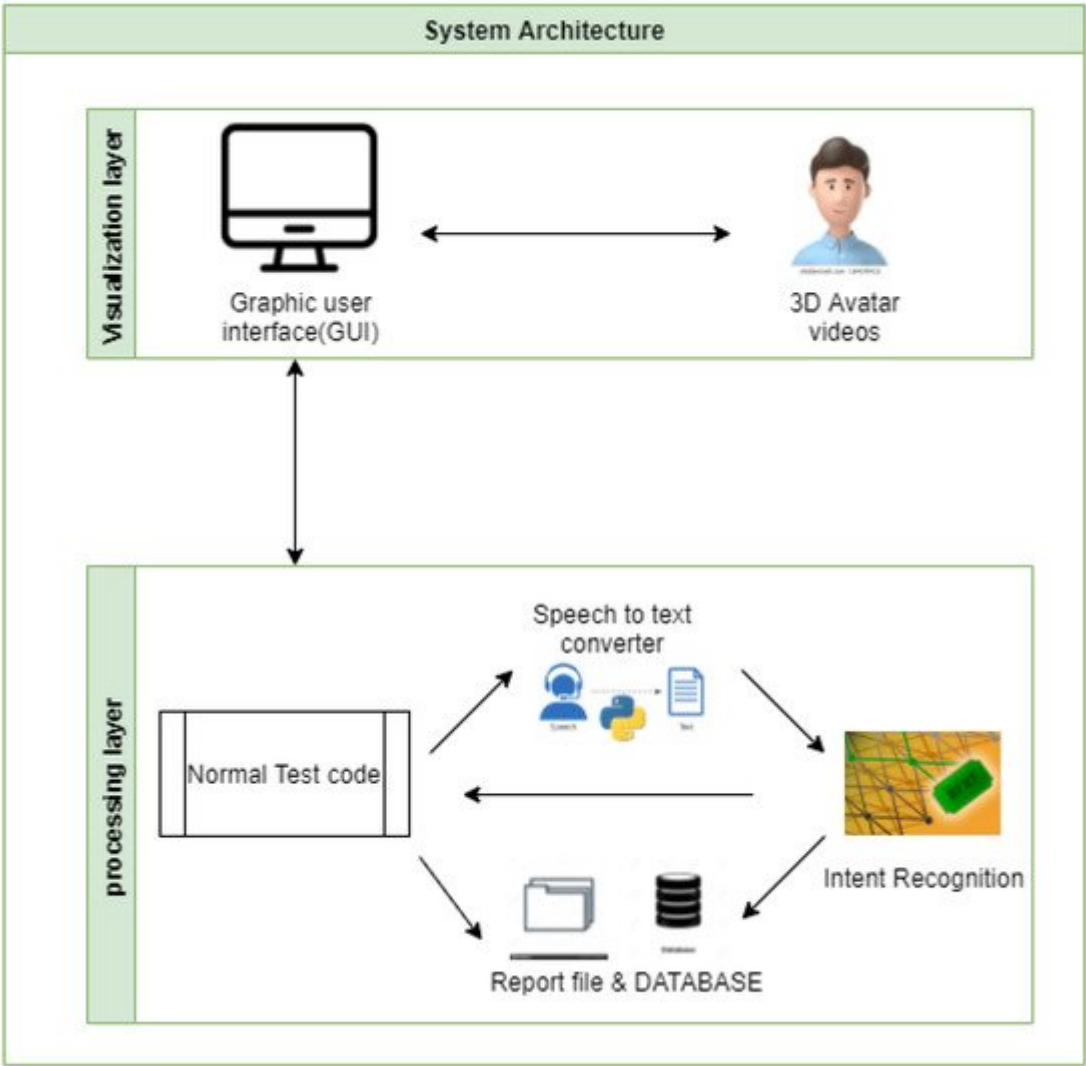


Figure 2. System’s global architecture.

2.2. The Input Data and System Patient Interaction

To make the system more realistic, researchers chose to simulate a real-life psychiatric interview where a 3D human avatar, as depicted in **Figure 3**, plays the doctor and asks the patient the psychiatric questions according to the MINI in its Tunisian Arabic version. The patient, in return, interacts with the avatar by answering the questions vocally.

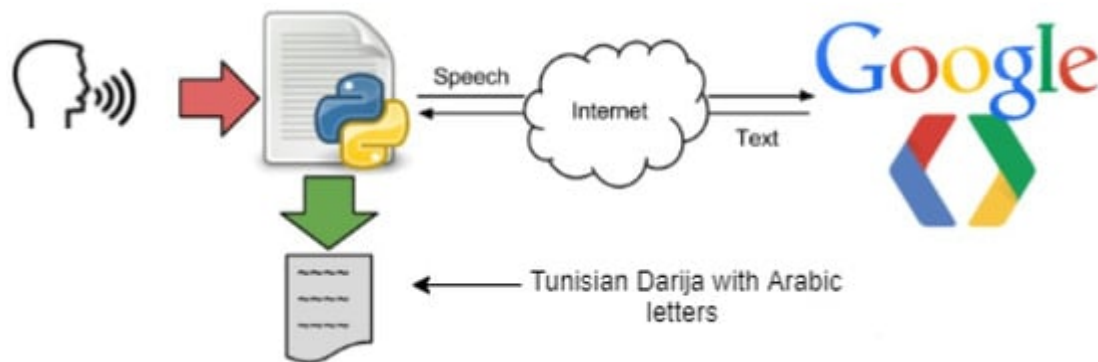


**Figure 3.** 3D human avatar–patient interaction.

However, the BERT model deals with text and not speech. Thus, to convert the speech to a text, researchers used a method called speech recognition which refers to automatic recognition of human speech.

Speech recognition is one of the critical tasks in human–computer interaction. Some well-known systems using speech recognition are Alexa and Siri. In current case, researchers are using the Google Speech-to-Text API with synchronous recognition request, which is the simplest method to perform recognition on speech audio data. It can process up to one minute of speech audio data sent in a synchronous request, and after Speech-to-Text API processes and recognizes all of the audio, it returns the converted text response. It is capable of identifying more than 80 languages to keep up with the global user base, and benchmarks assess its accuracy as 84% [13]. However, in current case, researchers are dealing with Tunisian “Darija” speech, which is a novelty with Google Speech API, although it offers a Tunisian Arabic option which is in reality different from Tunisia “Darija”, although there are many common words and similarities. The process works by giving the API a speech in Tunisian Darija that it converts and returns back as a text written in Arabic letters, as depicted in **Figure 4**. Several tests were conducted for the Tunisian dialect (Darija), and 80% of conversion accuracy was achieved in these tests. Researchers had to deal with some limits, such as the complexity of the Tunisian “Darija” (different accents, different languages included in it other than standard Arabic, such as French, Amazigh, Turkish, Maltese, Italian, etc.), which made it very difficult to convert the audio data into a text in a specific language. Current closest option was to convert the speech to text with Arabic letters, although the previously mentioned issues led to some errors such as missing some letters or failing to separate between words due to the tricky pronunciation of the dialect, which researchers had to take into account while building current dataset. Another limit researchers had to deal with is blocking the synchronous request, which means that Speech-to-Text must return a response before processing the subsequent request. After testing its performance, researchers found that it processes audio fast (30 s of audio in 15 s on average). In cases of poor audio quality, current recognition request can take significantly

longer, which is a problem researchers had to deal with when applying the API in current application by reducing noise in the room and enhancing the quality of the microphone.



**Figure 4.** Speech-to-Text process.

### 2.3. Dataset

Researchers used the MINI in its Tunisian Arabic version to prepare current dataset. researchers could not use all of its modules, so researchers ended up using the five most important ones recommended by the psychiatrists of the military hospital of Tunisia: depression, suicidality, adjustment disorder, panic disorder, and social phobia. researchers built the dataset by taking the questions of each module and anticipating the answers (e.g., Are you regularly depressed or down, most of the time, nearly every day, over the past two weeks?). This question is depression-related, so researchers anticipated all of its possible answers (e.g., yes, no, I am depressed, I have been consistently depressed, etc.), and researchers associated each answer with an intent (e.g., yes—depressed, no—other, I am depressed—depressed, I have been consistently depressed—depressed, etc.).

This process was challenging in the Tunisian Darija because many answers could have two intents depending on the situation and the nature of the question. For instance, two questions may have the same answer and mean completely two different things. Thus, researchers risk the repetition of the same answers many times in the same dataset, which may cause overfitting and wrong intent recognition.

Accordingly, to avoid this problem, researchers used five separate BERT models with a separate dataset for each module instead of one dataset with all the modules. In addition, researchers ensured that each dataset has unique answers without repetition and with one intent, which justifies the slight imbalance in the dataset between the “intent” class and the “other” class. The “nothing” class includes the misses of the Google Speech-to-Text API, after testing it several times with bad quality audio data (a lot of background noise, bad microphone, etc.), which are added in all the datasets just in case the Speech-to-Text failed under some conditions. It is more of a sign to tell the application user that something wrong in the audio and has to be fixed.

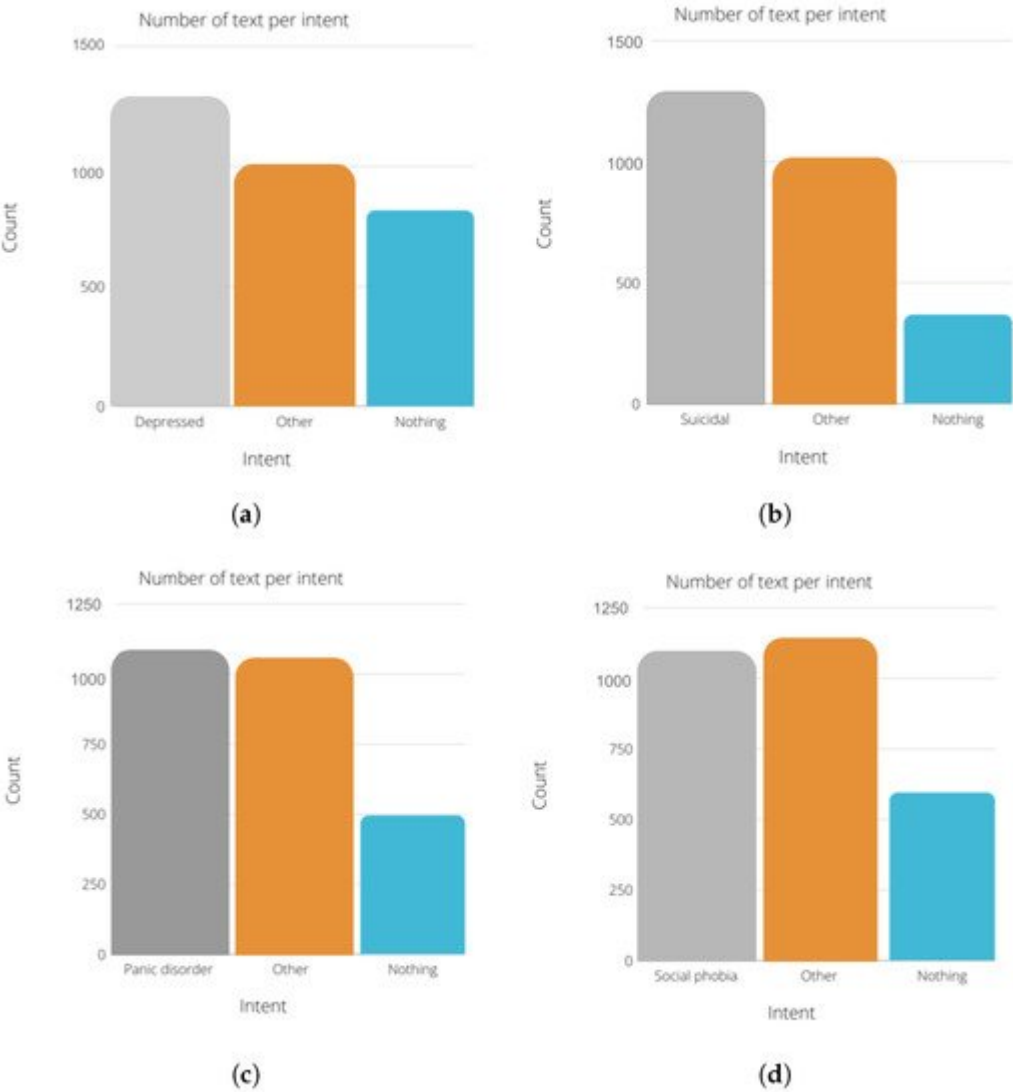
Although this justifiable unbalance in datasets may affect the accuracy of the models, especially on the level of the “nothing” class, by providing high-quality audio data and not letting the Google Speech-to-Text API make unwanted errors, researchers were able to overcome this issue so that the models would not have to deal with the “nothing”

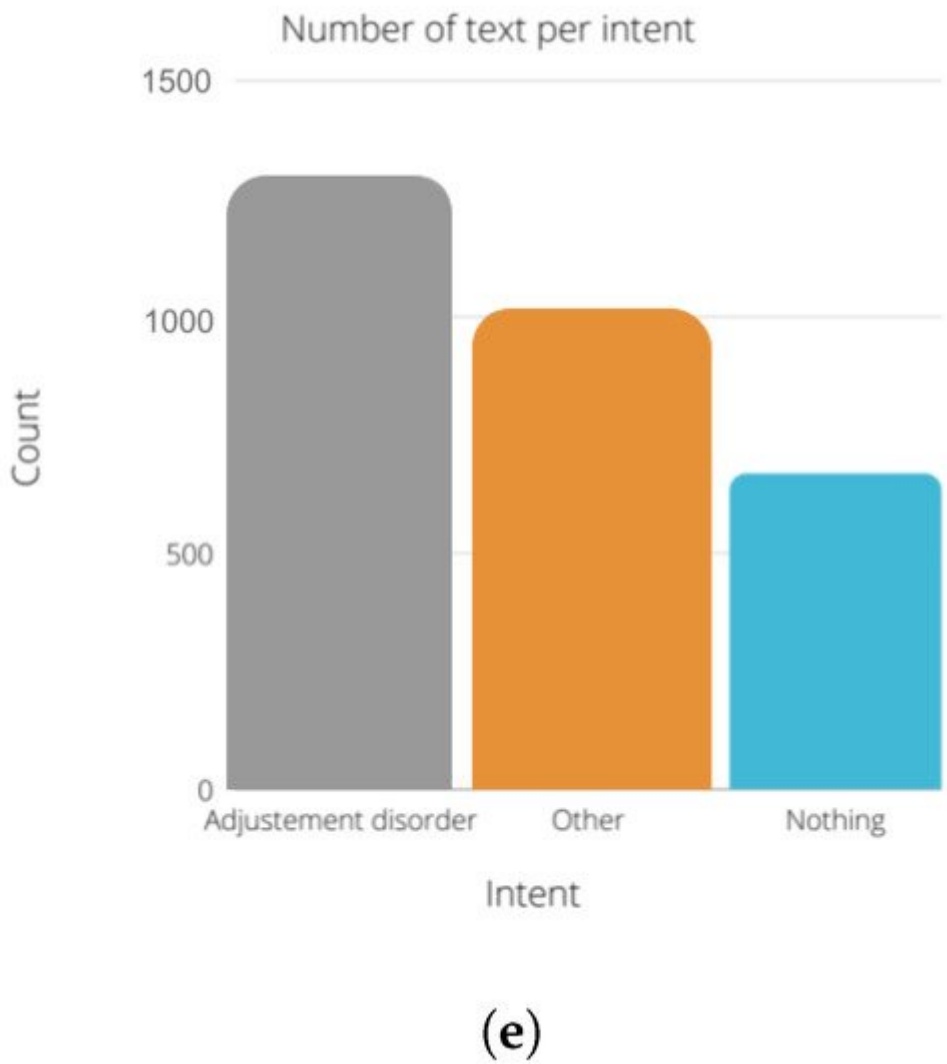
class only under very bad conditions, which will be a written warning when the user starts the application. **Figure 5** depicts an example of instances in this dataset.

1	text	intent	1	text	intent
2	الموت ارحم	SUICIDAL	2	Death is better	SUICIDAL
3	ما عندي علاش باش نعيش	SUICIDAL	3	I have nothing left to live for	SUICIDAL
4	ما عاد يقيد فيها شيء	SUICIDAL	4	I have no other solution	SUICIDAL
5	ما عادش عندي حل	SUICIDAL	5	I don't have a choice	SUICIDAL
6	عمري ما خمنت في الموت	Other	6	I have never thought of death	Other
7	مخمنش في الموت	Other	7	I did not think of death	Other
8	مخمنش	Other	8	No! that's insane	Other
9	معتدش شيء هذا	Other	9	it never crossed my mind	Other
10	لا عمري ما خمنت في الشيء هذا	Other	10	I can't commit a suicide	Other

**Figure5.** A sample from the suicidality dataset in Tunisian Darija and in English.

**Figure 6** describes the number of text per intent in each module. researchers chose to use the same amount of “nothing” text for all the modules. The number of instances for the other and the diagnosis state (depressed, suicidal, etc.) are between 1000 and 1500 instances for each module.





**Figure 6.** Class distribution of the datasets. (a) Depression dataset. (b) Suicidality dataset. (c) Panic disorder dataset. (d) Social phobia dataset. (e) Adjustment disorder dataset.

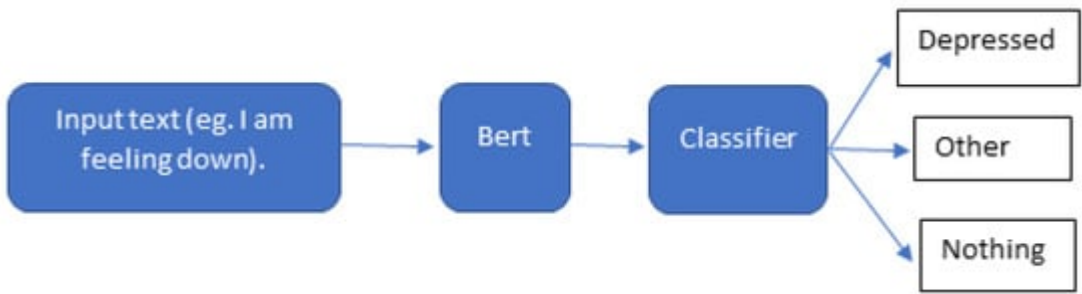
Data Split

researchers split current datasets into the training set, validation set, and test set as follows:

- The training set has to include a diverse collection of inputs so that the model can be trained in all settings and predict any unseen data samples.
- Separately from the training set, the validation set is used to make the validation process, which helps us tune the model's hyperparameters and configurations accordingly and prevent overfitting.

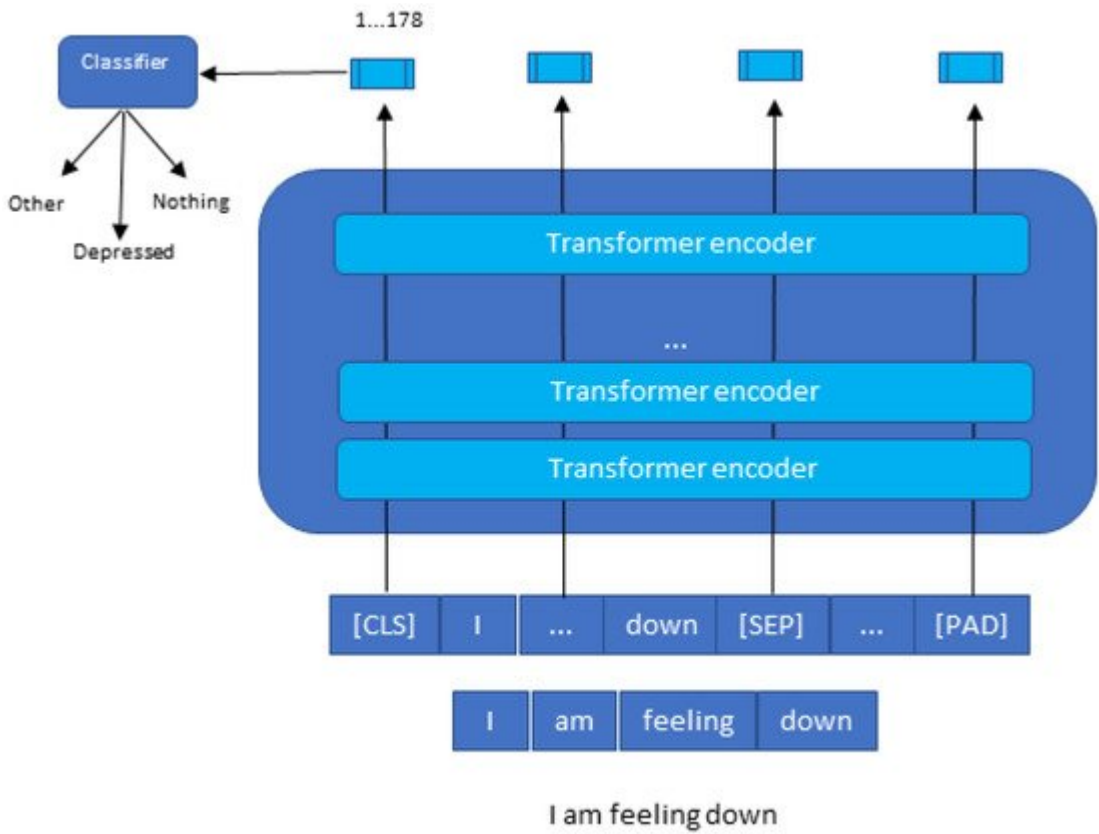
3. Intent Recognition and Text Classification with BERT

Current main goal from using the BERT model is to classify the speech of the patient among three classes which are the diagnosis class (depressed, suicidal, etc.), the other class, and the nothing class, as depicted in **Figure 7**.



**Figure 7.** Text classification with BERT.

To understand clearly how the classification with BERT works, **Figure 8** explains in detail how the process is carried out.



**Figure 8.** Text classification with BERT in detail.

The BERT model expects a sequence of tokens (words) in the input. In each series of tokens, there are two specific tokens that BERT would expect as an input: CLS, which is the first token of every sequence and stands for classification token, and SEP, which makes BERT recognize which token belongs to which sequence. This token is

essential for a next sentence prediction task or question-answering task. If researchers only have one series, this token will be appended to the end of the sequence.

For example, if researchers have a text consisting of the following short sentence: “I am feeling down”, first, this sentence has to be transformed into a sequence of tokens (words). As a result, researchers call this process “tokenization”, as is shown at the bottom of **Figure 8**. Then, researchers have to reformat that sequence of tokens by adding CLS and SEP tokens before using them as an input to the BERT model. It is crucial to consider that the maximum size of tokens that the BERT model can take is 512. If they are less, researchers can use padding to fill the unused token slots with PAD token. If they are longer, then a truncation has to be performed.

Once the input is successful, the BERT model will output an embedding vector of 768 in each of the tokens. These vectors can be used as an input for different NLP applications, such as the classification where researchers focus current attention on the embedding vector output from the special CLS token. This means the use of the embedding vector of size 768 from CLS token as an input for current classifier, and it will output a vector of size for the number of classes in current classification task.

## References

1. Reed, G.M.; Roberts, M.C.; Keeley, J.; Hooppell, C.; Matsumoto, C.; Sharan, P.; Robles, R.; Carvalho, H.; Wu, C.; Gureje, O.; et al. Mental health professionals' natural taxonomies of mental disorders: Implications for the clinical utility of the ICD-11 and the DSM-5. *J. Clin. Psychol.* 2013, 69, 1191–1212.
2. Mental Health; Harvard-Health-Publishing: Boston, MA, USA, 2021.
3. Hidaka, B.H. Depression as a disease of modernity: Explanations for increasing prevalence. *J. Affect. Disord.* 2012, 140, 205–214.
4. World-Health-Organization. Improving the Mental and Brain Health of Children and Adolescents; WHO: Geneva, Switzerland, 2020.
5. Sheehan, D.V.; Lecrubier, Y.; Sheehan, K.H.; Amorim, P.; Janavs, J.; Weiller, E.; Hergueta, T.; Baker, R.; Dunbar, G.C. The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 1998, 59, 22–33.
6. WHO. Depression and Other Common Mental Disorders: Global Health Estimates; Technical Report; World Health Organization: Geneva, Switzerland, 2017.
7. Metz, C.; Smith, C.S. Warnings of a Dark Side to A.I. in Health Care. *The New York Times*, 21 March 2019.

8. Al-Sarem, M.; Saeed, F.; Alsaeedi, A.; Boulila, W.; Al-Hadhrani, T. Ensemble methods for instance-based arabic language authorship attribution. *IEEE Access* 2020, 8, 17331–17345.
9. Conneau, A.; Schwenk, H.; Barrault, L.; Lecun, Y. Very Deep Convolutional Networks for Natural Language Processing. *arXiv* 2016, arXiv:1606.01781.
10. Sun, S.; Luo, C.; Chen, J. A review of natural language processing techniques for opinion mining systems. *Inf. Fusion* 2017, 36, 10–25.
11. Hirschberg, J.; Manning, C.D. Advances in natural language processing. *Science* 2015, 349, 261–266.
12. Devlin, J.; Chang, M.W. Open sourcing BERT: State-of-the-art pre-training for natural language processing. *Google AI Blog* 2018. Available online: <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html> (accessed on 30 March 2021).
13. Statista. Speech-to-Text Transcript Accuracy Rate among Leading Companies Worldwide in 2020. 2020. Available online: <https://www.statista.com/statistics/1133833/speech-to-text-transcript-accuracy-rate-among-leading-companies/> (accessed on 23 May 2021).

---

Retrieved from <https://encyclopedia.pub/entry/history/show/44952>