Article

A Privacy-Oriented Approach for Depression Signs Detection Based on Speech Analysis

Federica Vitale ¹ , Bruno Carbonaro ¹ , Gennaro Cordasco ² , Anna Esposito ² , Stefano Marrone ^{1,*} ,
Gennaro Raimo ² and Laura Verde ¹

¹ Dipartimento di Matematica e Fisica, Università della Campania “Luigi Vanvitelli”, 81100 Caserta, Italy; federica.vitale@unicampania.it (F.V.); bruno.carbonaro@unicampania.it (B.C.); laura.verde@unicampania.it (L.V.)

² Dipartimento di Psicologia, Università della Campania “Luigi Vanvitelli”, 81100 Caserta, Italy; gennaro.cordasco@unicampania.it (G.C.); anna.esposito@unicampania.it (A.E.); gennaro.raimo@unicampania.it (G.R.)

* Correspondence: stefano.marrone@unicampania.it; Tel.: +39-0823-27-5101

Abstract: Currently, AI-based assistive technologies, particularly those involving sensitive data, such as systems for detecting mental illness and emotional disorders, are full of confidentiality, integrity, and security compromises. In the aforesaid context, this work proposes an algorithm for detecting depressive states based on only three never utilized speech markers. This reduced number of markers offers a valuable protection of personal (sensitive) data by not allowing for the retrieval of the speaker’s identity. The proposed speech markers are derived from the analysis of pitch variations measured in speech data obtained through a tale reading task performed by typical and depressed subjects. A sample of 22 subjects (11 depressed and 11 healthy, according to both psychiatric diagnosis and BDI classification) were involved. The reading wave files were listened to and split into a sequence of intervals, each lasting two seconds. For each subject’s reading and each reading interval, the average pitch, the pitch variation (T), the average pitch variation (A), and the inversion percentage (also called the oscillation percentage O) were automatically computed. The values of the triplet (T_i , A_i , O_i) for the i -th subject provide, all together, a 100% correct discrimination between the speech produced by typical and depressed individuals, while requiring a very low computational cost and offering a valuable protection of personal data.

Keywords: speech recognition; human–computer interaction; statistical parametric speech analysis

Citation: Vitale, F.; Carbonaro, B.; Cordasco, G.; Esposito, A.; Marrone, S.; Raimo, G.; Verde, L. A Privacy-Oriented Approach for Depression Signs Detection Based on Speech Analysis. *Electronics* **2021**, *10*, 2986. <https://doi.org/10.3390/electronics10232986>

Academic Editors: Manohar Das, Taeshik Shon and Juan M. Corchado

Received: 31 August 2021

Accepted: 25 November 2021

Published: 1 December 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, the detection of depressive states is made through psychiatric examinations requiring the administration of psychiatric tests (such as the well-known Beck Depression Inventory (BDI) [1] test) and semi-structured interviews (such as Structured Clinical Interview-II (SCID-II) (<https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5>, last accessed on 31 August 2021) [2]). These procedures are extremely time-consuming, require a high level of expertise to reliably interpret the interviews’ outputs, and can be affected by clinicians’ theoretical orientations and an overestimation of patient’s progresses. In the last years, however, an ever-growing interest has been turned toward the possibility of detecting depression through automatic methods by making use of the interaction between human beings and computer-aided devices via a suitable definition of measurable behavioural depression features [3–10]. Some authors analyse speaking rates and silences in read and spontaneous speech (see [3,5,8–10] for further details), whereas others consider handwriting, drawing, and other behavioral features (see [4,6,8], also for more complete references). Nevertheless, speech remains the easiest source of behavioral data to be collected. Thus, identifying additional speech features that are able to more accurately discriminate between typical and depressed subjects and require low computational

costs is a goal worthy to be pursued. Notwithstanding the promises of Artificial Intelligence (AI) and Machine Learning (ML), technologies exploiting behavioural features can be used unethically, enabling possible discriminating usages of them and compromising the privacy of citizens. To protect citizens from these threats, the EU law on data protection and privacy in the European Union and European Economic Area (General Data Protection Regulation (GDPR) (https://ec.europa.eu/info/law/law-topic/data-protection_en (accessed on 31 August 2021) requires that the AI and ML methods and architectures developed by the scientific and industrial communities be endowed with advanced smart privacy-preserving procedures.

Hence, the main objective of this research is to provide a theoretical and practical framework for the detection of depressive signs based on two main pillars: privacy protection and accurate performance, to be executed on mobile devices. Accordingly, this paper proposes three more markers (described in Section 5), whose effectiveness will be demonstrated here.

The main differences between this paper and other approaches are: (1) the usage of a lightweight approach, based on the accurate choice of speech features rather than ML approaches; (2) the definition of an architecture for the continuous improvement of the classification method. Both the features and the architecture are privacy-oriented due to their capability of providing new data and computing on them without exposing personal information.

This research work is framed into the AutoNomous DiscoverY Of depressive Disorder Signs (ANDROIDS) project, whose objective is to investigate on AI-based tools and methodologies for the early multimodal detection of depressive signs.

The paper is structured as follows: Section 2 frames the paper in the technological and societal concern of privacy preservation in AI. Section 3 reports background information on the ANDROIDS project, as well as reviewing a brief state-of-art on the topic of the paper. Section 4 presents the approach followed in this paper, as well as a hw/sw architecture supporting it. Section 5 gives details on the data sampling and the detection of the used features. Section 6 describes the results of the data analysis task and describes the classification model. Section 7 discusses the results, and Section 8 ends the paper and draws future research tasks.

2. Motivation

Cloud-centric architectures and Deep Neural Network (DNN) applications have raised the necessity to move from the massive exploitation of acquired data to more sustainable computational paradigms. Such paradigms would be able to improve both performance and privacy, according to legislative regulations, such as the European Commission's GDPR [11] and Consumer Privacy Bill of Rights in the US [12].

Some privacy issues are solved by the introduction of blockchain technologies, since they can regulate data exchange during transactions among multiple stakeholders. However, blockchain introduces a computational overload that is not negligible for its actual application. As an example, in [13,14], a blockchain-based service infrastructure is proposed that matches European GDPR and privacy-related issues in the Internet of Vehicles (IoV) context.

On the other hand, DNNs perform better if they are fed with a great amount of data, opening privacy issues. In this way, a trade-off between performance and privacy does not allow for DNNs to scale. However, the problem of privacy for users and the lack of enthusiasm for sharing their personal data (e.g., the recordings of driving behaviours) represent obstacles for using this valuable data.

Mobile Edge Computing (MEC) [15] has provided an effective mitigation to scalability and performance levels degradation, implementing collaborative schemas for cooperative distributed ML models training. Nonetheless, data privacy is still an issue in MEC systems, since both processing at edge servers involves the transmission of potentially sensitive personal data. To date, privacy-sensitive users are still discouraged from exposing their

data and contributing to model training, since they fear being controlled by servers or the risk of violating increasingly stringent privacy laws.

3. Background and Related Works

This section is devoted to framing the present work into the ANDROIDS project (in Section 3.1) and then in the international scientific panorama (in Section 3.2).

3.1. The ANDROIDS Project

ANDROIDS investigates the features of human interactions in order to model cognitive and emotional processes. The aim is to study the fundamentals and the means to develop a Clinical Decision Support System (CDSS) in the field of depression detection and care. The pillars of the project are: (1) combining diverse sources of information (e.g., audio, handwriting, video); (2) defining detection methods and supporting software architectures able to guarantee the privacy of the patients in sharing their data with computer-based tools; (3) implementing multi-sensor data fusions [16]; (4) pursuing performances of detectors to realize scalable tools. The final aim is to train ML algorithms to generate a tool that can support the physicians and psychologists in the detection of signs of depressive disorders and to deepen further investigations.

The presented work focuses on the second and the third pillars. Starting from the non-spontaneous tales of both clinically diagnosed depressive patients and non-depressed people while reading a pre-defined tale, this research concentrates on analysing non-verbal features.

3.2. Related Works

There are several papers in the scientific literature focused on the prediction of depressive signs. The main approaches are summarized according to the nature of used data:

- Speech: para-verbals features—e.g., speed, silences, pauses [3,5,8–10]—as well as non-verbal features [17–20] in read and spontaneous speeches;
- Handwriting and drawing [4,6,8] mainly focusing on the shape of the drawn lines;
- Video analysis: face expressions [21], eye movements [22];
- Content of written and spoken words [23–26];
- Electroencephalogram (EEG) [27–31];
- Multimodality: more than a source of data is used and combined to improve detection performance [32,33].

Focusing on speech, it has been traditionally used to detect physical injuries and diseases. As an example, the project Vox4Health investigated the definition and design of an m-health system able to detect physical throat issues [34,35]. In [36], a smartphone records a voice signal of a client and sends it to a cloud server, where the signal is analyzed with DNNs. In [37], a feature selection method in terms of the wrapper approach is used to combine the single features ranked by using the Fisher discrimination ratio, where voice pathology detection experiments were carried out using Random Forest.

While the detection of some of these physical injuries is based on protocols accepted by the medical scientific community [38], the level of study of psychological pathological and pre-pathological states has not produced such protocols yet. Hence, a final selection of which features should be included in automatic classification tasks is still an open research topic.

Some research papers frame the privacy concern when speech is analyzed for clinical diagnosis [39].

Other highly cited papers dealing with the problem of privacy in smart healthcare are [40–42]. In particular, ref. [40] focuses on the usage of mobile phones in participating in distributed collection and analysis healthcare systems. Another important work in scientific literature is [43], where the authors also surveyed the main approaches and architecture to cope with this problem. The approach followed in [43] is based on data anonymization (i.e.,

the subjects' features do not allow for their identification). GDPR described two different ways to identify individuals:

- Directly, from their name, address, postcode, telephone number, photograph or image, or some other unique personal characteristic;
- Indirectly, from certain information linked together with other sources of information, including their place of work, job title, salary, their postcode, or even the fact that they have a particular diagnosis or condition.

The combination of Fully Homomorphic Encryption (FHE) and Neural Networks (NNs) is becoming a research trend that is attracting a growing interest in the scientific and industrial community. Some research papers dealing with this promising and not yet much researched topic are: [44], which is focused on the application of FHE to speech analysis [45], which proposes a test bed for evaluating the effectiveness of FHE on specific algorithms and computing functions; Ref. [46], where the authors assess different de-identification methods in speech emotional analysis.

All things considered, many of the approaches presented suffer from performance issues. To implement a system able to run on users' mobile terminals, some authors propose the usage of typing pattern analysis [47]. Other speech-based analysis methods, which involve the automatic transcription of speech using cloud-based resources, present a possible loss of privacy [48].

4. Methodology and Supporting Architecture

This section is devoted to describing the approach and presenting a supporting hardware and software architecture. Figure 1 depicts the overall architecture supporting the proposed approach.

Figure 1. A supporting architecture.

Under the privacy aspect, there are three actors: the final user, the privacy authority and the medical structure. In the first domain, patients use the detection tool (i.e., the early detector) that runs on their computing device. The application can record their voice and apply the criteria reported in Section 6 on the device itself. Patients can receive feedback from the application and, in case of the detection of depressive signs, they may contact their personal doctor in order to have a more profound analysis. In our vision, the feedback should be given by a message appearing on the screen or by sending a secure link via email.

After being given informed consent, the patient can contact a privacy authority that collects voice samples, anonymizes, and stores them in a repository (patients participate in this study on a voluntary basis on the base of the GDPR regulation).

This repository can be queried by different medical structures in order to create—using an Audio Processing Tool (APT) and the learning tool—the detection model. The detection model contains the ranges of the H, D, and M areas, which, respectively, represent healthy subjects, depressed subjects, and an overlapping zone of uncertainty.

The updated detection model can then be sent to the final user, who uses the detection model to classify himself/herself by applying the method described in Section 5. The final user can also considerably improve the detection mechanism by sharing speech data. This cyclic mechanism guarantees the capability of improving the detection model on the basis of fresh data continuously added in the repository.

Figure 2 provides an overview of the method on which the detection model is based.

Figure 2. Detection model computing.

The speech from a set of patients—both healthy and affected by depression—are acquired and split to improve the number of considered samples. Marker values are extracted from these samples to define those regions of values related to the different classifications. Finally, a posteriori analysis is used to quantify the uncertainty of classification for each region in order to estimate their accuracy.

This method is also applicable to data acquired to refine the detection model, as described in this section. In fact, when new patients are added to the dataset, they are examined by medic personnel who certify their clinical state.

In order to guarantee reaching the two objectives of this work, privacy and performance, the feature definition phase is critical. In fact, the markers used for the classification should be as few and easy-to-compute as possible in order to improve the performance (especially mobile devices). On the other hand, a preprocessing stage that hinders the indirect identification of users becomes necessary. Section 5 presents the details on which the detection model is computed.

It is important to underline that privacy is preserved in this architecture, since sensitive data are not sent to medical structures, which are ruled by private people and can also be regulated by different data privacy acts. Otherwise, data are simply sent to the privacy authority, which could be a public organization, and could expose its data handling services according to a trust chain supported by the adherence to international privacy regulations and standards.

Furthermore, the markers on which the algorithm is based are required to prevent revealing the identity of patients. As a last observation, it is important to highlight that the reports that the automated recognition software (based on the detection model) performs are presented to the final users, suggesting them to consult a medical professional in the case of depression suspects. Furthermore, in this case, privacy is preserved.

5. Feature Selection

To identify some features that could be able to characterize the speech of depressed and typical subjects (at least in sufficiently standardized conditions) a database consisting of 11 depressed subjects and 11 healthy subjects was used. The 11 Italian depressed patients were recruited with the help of psychiatrists at the Department of Mental Health and the General Hospital in Caserta, the Institute for Mental Health and the General Hospital in Santa Maria Capua Vetere, the Centre for psychological Listening in Aversa (Italy), and in a private psychiatric office. They were diagnosed as depressed and were under treatment (treatments are different and personalized on the basis of specific medical and other needs).

Both the control and the depressed participants were administered the Beck Depressive Inventory Second Edition (BDI-II). The data collection was conducted under the approval of the ethical committee of the Department of Psychology, Università della Campania “Luigi Vanvitelli”, code n. 25/2017.

Although the size of the dataset appears too small to lead to sufficiently general conclusions, it has been chosen for two main reasons: (a) from a purely methodological viewpoint, before widening the scope of our study, just a pilot field of investigation was needed as a guide toward a plausible hypothesis about the features that could be considered relevant to discriminating the speech of depressed people from that of healthy people; (b) from a practical viewpoint, the selected items were the ones about whom the information (with special regard to their BDI classes) was more complete. On the other hand, these samples are equally distributed with respect to gender (three men and eight women), and at least similarly distributed with respect to age.

Both healthy people and depressed patients were subjected to two experiments. The first experiment, called “Diary” was essentially an interview about their daily life in the last week; the other, called “Tale”, was the reading of a short tale by Aesop, *The North Wind and the Sun*. However, only the sound files of these readings (i.e., the “Tale” experiment) were taken into consideration, as they allowed for standard speech conditions, not affected by momentary moods depending on some possible experiences lived by the subjects.

Data have been extracted from the files by using the well-known Praat tool [49]. Each entire file was split in time intervals of 2 s and, for each interval, the average pitch and the average intensity of the voice were computed.

The average pitch was preferred to intensity, since the latter can strongly (much more than pitch, however) depend on the device used and on the conditions of experiment. It is worth noting that the analysis of pitch, like that of intensity, also captures at least partly the information delivered by the analysis of empty pauses [3], as, in the time intervals where pauses occur, average pitches present sudden cut-off of values.

Then, denoting by S the combination of the two samples and labelling each element of S with its ID number i , two classes of data were obtained: (1) the number of intervals associated with i , defined by Equation (1), where ℓ_i is the length of its speech (in seconds) for each $i \in S$, and (2) the sequence $\{p_{ij}\}_{1 \leq j \leq n_i}$ of the average pitches on the different intervals, where j is the index of the interval.

$$n_i = \begin{cases} \lfloor \ell_i/2 \rfloor & \text{if } (\ell_i/2) - \lfloor \ell_i/2 \rfloor \leq 0.5 \\ \lfloor \ell_i/2 \rfloor + 1 & \text{if } (\ell_i/2) - \lfloor \ell_i/2 \rfloor > 0.5 \end{cases} \quad (1)$$

where $\lfloor \cdot \rfloor$ denotes, as usual, the integer part.

Next, as the first good marker candidate, the *total variation* of pitch over the speech of i -th subject (for any $i \in S$) was chosen (see Equation (2) below). This marker can give some information about the uniformity and the “smoothness” of speech. The greater T_i is, the less smooth and uniform the speech.

$$T_i = \sum_{j=2}^{n_i} |p_{ij} - p_{i,j-1}| \quad (2)$$

The measure above is affected by the speed of the speech of each subject; in fact, although it is quite plausible that a subject who is speaking quickly also has a less smooth and uniform voice, two subjects with the same smoothness sound different if one of them speaks faster than the other. Thus, the second marker to be used in order to correct and sharpen the results given by T_i is the *average variation* of pitch over the whole reading (see Equation (3)).

$$A_i = \frac{T_i}{n_i} = \frac{\sum_{j=2}^{n_i} |p_{ij} - p_{i,j-1}|}{n_i} \quad \forall i \in S \quad (3)$$

On the other hand, a pitch variation between two subsequent intervals can be due to two different causes: (1) the reading expression, dictated by the meaning of what one is reading, and (2) the psychologic stability (or instability) of the reader. This remark has suggested considering the ratio as the third marker (see Equation (4)).

$$O_i = \frac{1}{2n_i} \sum_{j=2}^{n_i-1} \left(1 - \frac{(p_{ij} - p_{i,j-1})(p_{i,j+1} - p_{ij})}{|p_{ij} - p_{i,j-1}| |p_{i,j+1} - p_{ij}|} \right) \quad (4)$$

This is the percentage of *inversions* (also called *oscillations*) of the sign of two subsequent differences $p_{ij} - p_{i,j-1}$ and $p_{i,j+1} - p_{ij}$ over the whole reading for any $i \in S$. The percentage was preferred to the *number* of inversions, since the latter also depends on the length (hence, on the speed) of reading. This last marker can also give some information on the stability and uniformity of speech.

Therefore, we transformed our speech dataset into a set of 22 values T_i , 22 values A_i , and 22 values O_i recorded in suitable tables. Next, we computed the position and dispersion indexes for each variable and assigned a suitable confidence interval to each of them. For each subject i , the position of the triplet (T_i, A_i, O_i) , inside or outside the Cartesian product of these confidence intervals, is the main tool to discriminate whether the subject is normal or depressed.

6. Results

6.1. The Total Variation

The first step of the analysis was to evaluate, for any $i \in S$, the absolute values of the $n_i - 1$ pitch differences $p_{ij} - p_{i,j-1}$ and their sum. The results of this first step are listed in Table 1 for healthy subjects and in Table 2 for depressed subjects.

Table 1. Total variations (T_i) in non-depressed subjects.

Subject	Sex	Age	BDI	T_i
ID_39	F	60	1	379.75
ID_44	M	55	2	452.26
ID_6	M	49	5	228.98
ID_32	F	48	5	330.87
ID_10	F	46	4	364.33
ID_35	F	36	0	485.38
ID_33	F	36	9	361.04
ID_36	F	39	4	306.70
ID_42	M	33	2	281.21
ID_40	F	27	8	342.29
ID_30	F	27	10	364.87

The mean value of the data T_i and related standard deviation, as computed from Table 1, are reported in Equation (5) as a reference interval for the classification of data. One finds that the values of T_i for seven items in Table 1, namely, items ID_39, ID_32, ID_10, ID_33, ID_36, ID_40, and ID_30, belong to I_T , so that I_T turns out to be the confidence interval for T at the confidence level 64%.

$$I_T \equiv [\bar{T} - \sigma_T, \bar{T} + \sigma_T] \equiv [285.79, 422.88] \quad (5)$$

As observed in Table 2, the average value of depressed subjects is 435.69 ± 272.19 . Furthermore, nine items are *outside* I_T . More precisely, five items have a value $T_i < 285.79$ and four items have a value $T_i > 422.88$. Accordingly, 82% of the sample of depressed subjects are correctly classified *outside* I_T .

Table 2. Total variations (T_i) in depressed subjects.

Subject	Sex	Age	BDI	T_i
ID_4	F	63	27	869.59
ID_3	M	60	33	202.87
ID_17	M	53	20	1005.01
ID_15	F	49	20	289.03
ID_13	F	48	16	280.75
ID_12	F	46	31	541.34
ID_8	F	35	36	598.51
ID_9	F	33	38	351.97
ID_18	M	29	25	115.88
ID_14	F	27	19	279.31
ID_19	F	27	25	258.32

6.2. The Average Variation

The results of the average variations of pitch for the subjects in the sample S are now collected in Tables 3 and 4, the former for healthy subjects and the latter for depressed ones.

Table 3. Average variations (A_i) in non depressed subjects.

Subject	Sex	Age	BDI	A_i
ID_39	F	60	1	17.26
ID_44	M	55	2	16.75
ID_6	M	49	5	11.45
ID_32	F	48	5	17.41
ID_10	F	46	4	20.24
ID_35	F	36	0	24.27
ID_33	F	36	9	17.19
ID_36	F	39	4	19.17
ID_42	M	33	2	13.39
ID_40	F	27	8	20.13
ID_30	F	27	10	20.27

Table 4. Average variations (A_i) in depressed subjects.

Subject	Sex	Age	BDI	A_i
ID_4	F	63	27	24.16
ID_3	M	60	33	8.82
ID_17	M	53	20	35.89
ID_15	F	49	20	12.57
ID_13	F	48	16	13.37
ID_12	F	46	31	23.54
ID_8	F	35	36	31.50
ID_9	F	33	38	16.76
ID_18	M	29	25	5.27
ID_14	F	27	19	15.52
ID_19	F	27	25	13.60

The mean value and standard deviation of A_i , as computed from Table 3, are in Equation (6), as a reference interval for the classification of data. One finds that the values of A_i for eight items in Table 3 belong to I_A , so that I_A turns out to be the confidence interval for A at the confidence level 73%.

$$I_A \equiv [\bar{A} - \sigma_A, \bar{A} + \sigma_A] \equiv [14.62, 21.30] \quad (6)$$

As observed in Table 4, we find that their average value is 18.27 and their standard deviation is 9.02. The most relevant factor to note is that nine items in Table 4 turn out to be *outside* interval I_A . More precisely, five items have a value $A_i < 14.62$ and four items have a value $A_i > 21.30$. Again, 82% of the sample of depressed subjects are correctly classified *outside* I_A .

6.3. The Inversion Percentage

The last marker is the inversion percentage. The values O_i of this marker are listed in Tables 5 and 6 for healthy and depressed subjects, respectively.

Table 5. Inversion percentages (O_i) in non-depressed subjects.

Subject	Sex	Age	BDI	O_i
ID_39	F	60	1	43.48%
ID_44	M	55	2	71.43%
ID_6	M	49	5	66.67%
ID_32	F	48	5	55.00%
ID_10	F	46	4	84.21%
ID_35	F	36	0	61.90%
ID_33	F	36	9	59.09%
ID_36	F	39	4	47.06%
ID_42	M	33	2	68.18%
ID_40	F	27	8	61.11%
ID_30	F	27	10	78.95%

Table 6. Inversion percentages (O_i) in depressed subjects.

Subject	Sex	Age	BDI	O_i
ID_4	F	63	27	72.97%
ID_3	M	60	33	62.50%
ID_17	M	53	20	65.52%
ID_15	F	49	20	58.33%
ID_13	F	48	16	59.09%
ID_12	F	46	31	50.00%
ID_8	F	35	36	55.00%
ID_9	F	33	38	77.27%
ID_18	M	29	25	56.52%
ID_14	F	27	19	78.95%
ID_19	F	27	25	60.00%

The average value and standard deviation of O_i , as computed from Table 5, are reported in Equation (7) as a reference interval for the classification of data. One finds that seven items in Table 5 belong to I_O , so that I_O is, in turn, the confidence interval for O at the confidence level 64%.

$$I_O \equiv [\bar{O} - \sigma_O, \bar{O} + \sigma_O] \equiv [51.56, 75.18] \quad (7)$$

Finally, observing the values O_i in Table 6, we find that their average value is 63.29% and that their standard deviation is 8.97%. Only three items in Table 6 are *outside* interval I_O ($1 < 51.56\%$ and $2 > 75.18$). With this marker, only 27% of depressed subjects are correctly classified. However, the important contribution of the inversion percentage is that it *correctly places* subject ID_9, who had instead escaped the classification provided by total and average variations.

7. Discussion

The main contribution of the work is finding a quick and easy-to-compute method for classifying depressed and non-depressed subjects from speech non-verbal analysis. Three markers have been found to accomplish this goal.

The choice of the markers was mainly due to perception. As a matter of fact, human ears seem to be able to perceive the psychological conditions of a subject who reads a tale: at least some suggestions about these conditions can be obtained from the empty pauses [3], from the reading speed [50], from the voice tones, and from their variations [51]. This suggests that further attention should be placed on pitch and, above all, on its variations during the reading; hence, the chosen marker, which is why they were chosen as the main markers in this study.

On the other hand, all of this information is mainly qualitative, and depends on numerous interfering psychological states, so that one must take into account several “modes” of depression, which could produce (in principle with the same probability) a too fast or a too slow reading, too high or too low tones, some “flatness” of the pitch, or some high-frequency oscillation of pitch (on a purely perceptive ground, all of these features have actually been detected in all of the depressed subjects of the sample) [50,51]. None of the above described markers could have been able, by themselves, to give a sharp discrimination between typical and depressed subjects. They were to be used as “filters” selecting, one after another, the classes of subjects. Furthermore, as a matter of fact, the joint use of these three markers has led to a correct discrimination of 100% of depressed subjects (and, of course, 100% of healthy subjects).

In fact, the main conclusion that seems to be drawn from the results reported in the previous Section can be expressed in terms of probability as follows. Let us first consider the four (random) variables

$$B : i \in S \rightarrow B_i \in \{0, \dots, 63\}$$

(where B_i is the BDI score assigned to $i \in S$. The BDI score spans from 0 to 63 according to its definition [1]),

$$T : i \in S \rightarrow T_i \in [0, +\infty),$$

$$A : i \in S \rightarrow A_i \in [0, +\infty),$$

and

$$O : i \in S \rightarrow O_i \in [0, 100].$$

We may now envisage S as the space of probability $\{0, \dots, 63\} \times [0, +\infty) \times [0, +\infty) \times [0, 100]$.

We denote by N the set $\{0, \dots, 15\} \times [0, +\infty) \times [0, +\infty) \times [0, 100]$ and by $P \equiv N^c$ the set $\{16, \dots, 63\} \times [0, +\infty) \times [0, +\infty) \times [0, 100]$, i.e., the events $N \equiv$ “healthy subject according to BDI test” and $P \equiv$ “depressed subject according to BDI test”, respectively (with no information on the values of T , A and O). Thus, the results described in the previous Section can be epitomized by the relations

$$\text{Prob}((I_T \times I_A \times I_O)^c | P) = 1 \quad (8)$$

and

$$\text{Prob}((I_T^c \times I_A^c \times I_O^c)^c | N) = 1, \quad (9)$$

or, equivalently (and perhaps more expressively),

$$\text{Prob}(I_T \times I_A \times I_O | P) = 0 \quad (10)$$

and

$$\text{Prob}(I_T^c \times I_A^c \times I_O^c | N) = 0, \quad (11)$$

where the symbol “Prob” means “probability” (Figure 3 gives a visual synthetic picture of the situation described by Relations (10) and (11)).

Figure 3. Region H: Only healthy subjects; Region D: Only depressed subjects; Region M: Overlapping zone.

Two remarks now spontaneously arise in connection with these results. First, the limits of the confidence intervals I_T , I_A , and I_O depend on the subset N of the sample S . This means that, if we change our sample, we find different limiting values to discriminate between healthy and depressed subjects. In addition, the sets excluded by Relations (10) and (11) are not complementary, i.e., their combination is not the whole S , so that there is an overlapping zone M , namely, the set $(I_T \times I_A \times I_O)^c \cap (I_T^c \times I_A^c \times I_O^c)^c$, where, in principle, both depressed and typical subjects are somehow mixed. In connection with the first remark, it is to be expected that, taking ever larger samples, the averages of the variables T , A , and O will tend to fluctuate each in a sufficiently small neighborhood of a stable value, and, also, that their standard deviations will describe sets of values whose upper bounds will cluster around a stable value.

One can observe, on the one hand, that no medical diagnosis is free of a greater or smaller degree of uncertainty (i.e., in most cases, a diagnosis requires further examinations to be validated or rejected), and, on the other hand, the proposed markers seeming to work as filters likely suggests that it could be possible to find a suitable function $F(T, A, O)$ leading to a sharp distinction between a properly defined healthy zone and a depressed zone, with no overlapping.

From a diagnostic viewpoint, however, it will be of particular interest to give the explicit expressions of the “inverse” probabilities:

1. $\text{Prob}(N | I_T \times I_A \times I_O)$ —the probability that a subject whose three measured markers are *all* in I_T , I_A , and I_O , respectively, is a typical subject;
2. $\text{Prob}(P | I_T^c \times I_A^c \times I_O^c)$ —the probability that a subject whose three measured markers are *all* outside I_T , I_A , and I_O , respectively, is depressed;
3. $\text{Prob}(P | M)$ —the probability that a subject for whom one or two measured markers, *but not all three*, are outside I_T , I_A , and I_O , respectively, is depressed.

In fact, these probabilities express the plausibility of a diagnosis given the values of the markers. Without these probabilities, our analysis would have been severely incomplete.

Now, since $M = (I_T \times I_A \times I_O)^c \setminus (I_T^c \times I_A^c \times I_O^c)$, we have

$$M = [(I_T^c \times I_A \times I_O) \cup (I_T \times I_A^c \times I_O) \cup (I_T \times I_A \times I_O^c)] \cup [(I_T^c \times I_A^c \times I_O) \cup (I_T^c \times I_A \times I_O^c) \cup (I_T \times I_A^c \times I_O^c)].$$

and, applying the Bayes formula, we obtain

$$\text{Prob}(N | I_T \times I_A \times I_O) = \frac{\text{Prob}(I_T \times I_A \times I_O | N) \mathbf{P}(N)}{\text{Prob}(I_T \times I_A \times I_O | N) \text{Prob}(N) + \text{Prob}(I_T \times I_A \times I_O | P) \text{Prob}(P)}$$

where $\text{Prob}(N) = \text{Prob}(P) = 1/2$ (11/22) e $\text{Prob}(I_T \times I_A \times I_O | P) = 0$, so that

$$\text{Prob}(N | I_T \times I_A \times I_O) = \frac{\text{Prob}(I_T \times I_A \times I_O | N) \text{Prob}(N)}{\text{Prob}(I_T \times I_A \times I_O | N) \text{Prob}(N)} = 1.$$

and

$$\text{Prob}(P | I_T \times I_A \times I_O) = 1 - \text{Prob}(N | I_T \times I_A \times I_O) = 0.$$

Next,

$$\text{Prob}(N | I_T^c \times I_A^c \times I_O^c) = \frac{\text{Prob}(I_T^c \times I_A^c \times I_O^c | N) \text{Prob}(N)}{\text{Prob}(I_T^c \times I_A^c \times I_O^c | N) \text{Prob}(N) + \text{Prob}(I_T^c \times I_A^c \times I_O^c | P) \text{Prob}(D)}$$

where $\text{Prob}(I_T^c \times I_A^c \times I_O^c | N) = 0$, so that

$$\text{Prob}(N | I_T^c \times I_A^c \times I_O^c) = 0$$

and

$$\text{Prob}(P | I_T^c \times I_A^c \times I_O^c) = 1 - \text{Prob}(N | I_T^c \times I_A^c \times I_O^c) = 1.$$

Now

$$\text{Prob}(N | M) = \frac{\text{Prob}(M | N) \text{Prob}(N)}{\text{Prob}(M)},$$

that is

$$\text{Prob}(N | M) = \frac{\text{Prob}(M | N) \text{Prob}(N)}{\text{Prob}(M | N) \text{Prob}(N) + \text{Prob}(M | P) \text{Prob}(P)} = \frac{\text{Prob}(M | N)}{\text{Prob}(M | N) + \text{Prob}(M | P)}.$$

Now, if we check the measures listed in the Tables reported in the previous Sections, we find that

$$\text{Prob}(M | N) = 8/11 \approx 73\%$$

and

$$\text{Prob}(M | P) = 10/11 \approx 90\%$$

so that

$$\text{Prob}(N | M) = 4/9 \approx 44\% \quad (12)$$

and, as a consequence,

$$\text{Prob}(P | M) = 5/9 \approx 56\%. \quad (13)$$

In conclusion, when the triplet of the values of the markers is in the H region, then we can be sure that the subject examined is healthy, and when it belongs to the D region, we have the same certainty that the subject is depressed. When the triplet of the values of markers is in the M region, as often happens in medical investigations, the results are not conclusive, but the probability that the subject examined is depressed is perceptibly higher than that he is healthy. As a matter of fact, according to Relations (12) and (13), the probability that a subject is healthy, given that the triplet of the values of markers is in the

region M , is 44%, whereas the probability that a subject is depressed, given that the triplet of the values of markers is in the region M , is 56%.

These mathematical considerations provide a method to check about the depressive and healthy status of new acquired speech samples. Specifically, by locating the speech samples of users in the proper space depicted in Figure 3, the proposed tool can determine their classification in the healthy/depressed category.

These results are limited to the rather small sample we have considered. Nevertheless, it still seems plausible that a significantly larger set of samples could lead not only to a better definition of the H and D regions, but also to more significant probability differences in the M region.

8. Conclusions and Future Works

According to the above discussion, it is clear that the present work is just a first step in a promising direction, and is a preliminary study toward an automatic procedure to discriminate between typical and depressed subjects through speech analysis while preserving their anonymity and privacy.

The preliminary nature of the study is due to the limited number of samples analyzed in this paper. In these samples, the results show the potential of the proposed algorithm; hence, the authors aim to run this approach on larger datasets (e.g., the DAIC-WOZ database [52]) in the near future. In addition to this, ML approaches tend to over-perform in laboratory settings as compared to real-life application. To overcome this, the architecture presented in this paper incorporates a continuous improvement of classification accuracy.

Regardless, the results obtained and presented here seem to point out an interesting and promising pathway to deepen and extend the research on the markers proposed in the present work. They deserve to be examined more deeply, as they seem to work effectively for the desired discrimination because they actually allow for identifying a “healthy region” and a “depressed region”, though, at this preliminary stage, it is dependent on the selected sample.

Needless to say, a crucial step in the future development of the research about markers is a Bayesian analysis of inverse conditional probabilities to the ones considered in Relations (10) and (11). Such an analysis will be the first step of the future research on the effectiveness of the variables T , A , and O as markers of depression.

Another important future study will be conducted on the impact of the quality of speech on recognition performance. Since mobile devices are not likely to assure a very high quality of recordings, and since background noise can be frequent in such recordings, this study will give a firm contribution toward defining the means of adopting the approach in mobile settings. To this aim, the samples considered in this paper will be distorted with growing levels of noise and with typical background noises, such as cafeterias, streets, etc.

Additionally, other ML approaches able to generate new detectors from the dataset generated by the APT will be explored. The performance obtained using ML techniques, such as Random Forests, Decision Trees, or Support Vector Machines, will be evaluated and compared with the results achieved by the proposed approach. Finally, to preserve the privacy of subjects and to improve the protection of their sensitive data, anonymization methods will be explored.

Author Contributions: Conceptualization, F.V. and B.C.; Formal analysis, F.V.; Methodology, F.V. and S.M.; Writing, all the authors; data curation, G.R. and L.V.; writing–review and editing, G.C. and S.M.; supervision, A.E. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results was mainly supported by the project ANDROIDS, funded by the program V:ALERE 2019 Università della Campania “Luigi Vanvitelli”, D.R. 906 del 4/10/2019, prot. n. 157264,17/10/2019.

Institutional Review Board Statement: The Ethical Board of Università della Campania “Luigi Vanvitelli” authorized this research and approved the clinical protocol on which basis data has been acquired and organized.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Participation was on a voluntary basis. People also agreed to the management of their data for research aims and in the respect of to the GDPR.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANDROIDS	AutoNomous DiscoverY Of depressive Disorder Signs
APT	Audio Processing Tool
BDI	Beck Depression Inventory
CDSS	Clinical Decision Support System
DNN	Deep Neural Network
DTs	Decision Trees
EEG	Electroencephalogram
FHE	Fully Homomorphic Encryption
GDPR	General Data Protection Regulation
IoV	Internet of Vehicles
MEC	Mobile Edge Computing
ML	Machine Learning
NNs	Neural Networks
RFs	Random Forests
SCID	Structured Clinical Interview
SMVs	Support Vector Machines

References

- Steer, R.; Rissmiller, D.; Beck, A. Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behav. Res. Ther.* **2000**, *38*, 311–318. [\[CrossRef\]](#)
- First, M.; Spitzer, R.; Gibbon, M.; Williams, J. Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II). In *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*; American Psychiatric Association: Washington, DC, USA, 1997.
- Esposito, A.; Esposito, A.; Likforman-Sulem, L.; Maldonato, M.; Vinciarelli, A. On the significance of speech pauses in depressive disorders: Results on read and spontaneous narratives. *Smart Innov. Syst. Technol.* **2016**, *48*, 73–82. [\[CrossRef\]](#)
- Likforman-Sulem, L.; Esposito, A.; Faundez-Zanuy, M.; Clemencon, S.; Cordasco, G. EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 273–284. [\[CrossRef\]](#)
- Scibelli, F.; Roffo, G.; Tayarani, M.; Bartoli, L.; De Mattia, G.; Esposito, A.; Vinciarelli, A. Depression Speaks: Automatic Discrimination between Depressed and Non-Depressed Speakers Based on Nonverbal Speech Features. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 6842–6846. [\[CrossRef\]](#)
- Cordasco, G.; Scibelli, F.; Faundez-Zanuy, M.; Likforman-Sulem, L.; Esposito, A. Handwriting and drawing features for detecting negative moods. *Smart Innov. Syst. Technol.* **2019**, *103*, 73–86. [\[CrossRef\]](#)
- Esposito, A.; Callejas, Z.; Hemmje, M.; Fuchs, M.; Maldonato, M.; Cordasco, G. Intelligent Advanced User Interfaces for Monitoring Mental Health Wellbeing. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 12585 LNCS; Springer: Cham, Switzerland, 2021; pp. 83–95. [\[CrossRef\]](#)
- Esposito, A.; Raimo, G.; Maldonato, M.; Vogel, C.; Conson, M.; Cordasco, G. Behavioral sentiment analysis of depressive states. In Proceedings of the 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Mariehamn, Finland, 23–25 September 2020; pp. 209–214. [\[CrossRef\]](#)
- Aloshban, N.; Esposito, A.; Vinciarelli, A. Detecting Depression in Less Than 10 Seconds: Impact of Speaking Time on Depression Detection Sensitivity. In Proceedings of the 2020 International Conference on Multimodal Interaction, Utrecht, The Netherlands, 25–29 October 2020; pp. 79–87. [\[CrossRef\]](#)
- Tao, F.; Esposito, A.; Vinciarelli, A. Spotting the traces of depression in read speech: An approach based on computational paralinguistics and social signal processing. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020; pp. 1828–1832. [\[CrossRef\]](#)
- Custers, B.; Sears, A.M.; Dechesne, F.; Georgieva, I.; Tani, T.; van der Hof, S. *EU Personal Data Protection in Policy and Practice*; Springer: Berlin/Heidelberg, Germany, 2019.
- Gaff, B.M.; Sussman, H.E.; Geetter, J. Privacy and big data. *Computer* **2014**, *47*, 7–9. [\[CrossRef\]](#)

13. Campanile, L.; Iacono, M.; Marulli, F.; Mastroianni, M. Privacy Regulations Challenges on Data-centric and IoT Systems: A Case Study for Smart Vehicles. In Proceedings of the 5th International Conference on Internet of Things, Big Data and Security (IoTBDS), Prague, Czech Republic, 7–9 May 2020; pp. 507–518.
14. Campanile, L.; Iacono, M.; Levis, A.H.; Marulli, F.; Mastroianni, M. Privacy Regulations, Smart Roads, Blockchain, and Liability Insurance: Putting Technologies to Work. *IEEE Secur. Priv.* **2020**, *19*, 34–43. [\[CrossRef\]](#)
15. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.C.; Yang, Q.; Niyato, D.; Miao, C. Federated learning in mobile edge networks: A comprehensive survey. *arXiv* **2019**, arXiv:1909.11875.
16. Khaleghi, B.; Khamis, A.; Karray, F.; Razavi, S. Multisensor data fusion: A review of the state-of-the-art. *Inf. Fusion* **2013**, *14*, 28–44. [\[CrossRef\]](#)
17. Kaya, H.; Fedotov, D.; Dresvyanskiy, D.; Doyran, M.; Mamontov, D.; Markitantov, M.; Salah, A.; Kavcar, E.; Karpov, A.; Salah, A. Predicting depression and emotions in the cross-roads of cultures, para-linguistics, and non-linguistics. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop, Nice, France, 21–25 October 2019; pp. 27–35. [\[CrossRef\]](#)
18. Nilsonne, A. Acoustic analysis of speech variables during depression and after improvement. *Acta Psychiatr. Scand.* **1987**, *76*, 235–245. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Nilsonne, A.; Sundberg, J.; Ternström, S.; Askenfelt, A. Measuring the rate of change in voice fundamental frequency in fluent speech during mental depression. *J. Acoust. Soc. Am.* **1988**, *83*, 716–728. [\[CrossRef\]](#)
20. Talavera, J.; Saiz-Ruiz, J.; Garcia-Toro, M. Quantitative measurement of depression through speech analysis. *Eur. Psychiatry* **1994**, *9*, 185–193. [\[CrossRef\]](#)
21. Singh, J.; Goyal, G. Decoding depressive disorder using computer vision. *Multimed. Tools Appl.* **2021**, *80*, 8189–8212. [\[CrossRef\]](#)
22. Alghowinem, S.; Goecke, R.; Wagner, M.; Parker, G.; Breakspear, M. Eye movement analysis for depression detection. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 4220–4224. [\[CrossRef\]](#)
23. Trotzek, M.; Koitka, S.; Friedrich, C. Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 588–601. [\[CrossRef\]](#)
24. Huang, Z.; Epps, J.; Joachim, D.; Sethu, V. Natural Language Processing Methods for Acoustic and Landmark Event-Based Features in Speech-Based Depression Detection. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 435–448. [\[CrossRef\]](#)
25. Polignano, M.; De Gemmis, M.; Narducci, F.; Semeraro, G. Do you feel blue? Detection of negative feeling from social media. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 10640 LNAI; Springer: Cham, Switzerland, 2017; pp. 321–333. [\[CrossRef\]](#)
26. Yazdavar, A.; Al-Olimat, H.; Ebrahimi, M.; Bajaj, G.; Banerjee, T.; Thirunarayan, K.; Pathak, J.; Sheth, A. Semi-Supervised approach to monitoring clinical depressive symptoms in social media. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, Australia, 31 July–3 August 2017; pp. 1191–1198. [\[CrossRef\]](#)
27. Hinrikus, H.; Suhhova, A.; Bachmann, M.; Adamsoo, K.; Vohma, U.; Lass, J.; Tuulik, V. Electroencephalographic spectral asymmetry index for detection of depression. *Med. Biol. Eng. Comput.* **2009**, *47*, 1291–1299. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Subhani, A.; Mumtaz, W.; Saad, M.; Kamel, N.; Malik, A. Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access* **2017**, *5*, 13545–13556. [\[CrossRef\]](#)
29. Li, X.; Hu, B.; Sun, S.; Cai, H. EEG-based mild depressive detection using feature selection methods and classifiers. *Comput. Methods Programs Biomed.* **2016**, *136*, 151–161. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Cai, H.; Han, J.; Chen, Y.; Sha, X.; Wang, Z.; Hu, B.; Yang, J.; Feng, L.; Ding, Z.; Chen, Y.; et al. A Pervasive Approach to EEG-Based Depression Detection. *Complexity* **2018**, *2018*, 5238028. [\[CrossRef\]](#)
31. Sharma, M.; Achuth, P.; Deb, D.; Puthankattil, S.; Acharya, U. An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals. *Cogn. Syst. Res.* **2018**, *52*, 508–520. [\[CrossRef\]](#)
32. Alghowinem, S.; Goecke, R.; Wagner, M.; Epps, J.; Hyett, M.; Parker, G.; Breakspear, M. Multimodal Depression Detection: Fusion Analysis of Paralinguistic, Head Pose and Eye Gaze Behaviors. *IEEE Trans. Affect. Comput.* **2018**, *9*, 478–490. [\[CrossRef\]](#)
33. Pampouchidou, A.; Simantiraki, O.; Vazakopoulou, C.M.; Chatzaki, C.; Padiaditis, M.; Maridaki, A.; Marias, K.; Simos, P.; Yang, F.; Meriaudeau, F.; et al. Facial geometry and speech analysis for depression detection. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 1433–1436. [\[CrossRef\]](#)
34. Cesari, U.; De Pietro, G.; Marciano, E.; Niri, C.; Sannino, G.; Verde, L. Voice Disorder Detection via an m-Health System: Design and Results of a Clinical Study to Evaluate Vox4Health. *BioMed Res. Int.* **2018**, *2018*, 8193694. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Verde, L.; De Pietro, G.; Alrashoud, M.; Ghoneim, A.; Al-Mutib, K.; Sannino, G. Leveraging artificial intelligence to improve voice disorder identification through the use of a reliable mobile app. *IEEE Access* **2019**, *7*, 124048–124054. [\[CrossRef\]](#)
36. Alhussein, M.; Muhammad, G. Automatic Voice Pathology Monitoring Using Parallel Deep Models for Smart Healthcare. *IEEE Access* **2019**, *7*, 46474–46479. [\[CrossRef\]](#)
37. Wu, Y.; Zhou, C.; Fan, Z.; Wu, D.; Zhang, X.; Tao, Z. Investigation and Evaluation of Glottal Flow Waveform for Voice Pathology Detection. *IEEE Access* **2021**, *9*, 30–44. [\[CrossRef\]](#)
38. Verde, L.; De Pietro, G.; Alrashoud, M.; Ghoneim, A.; Al-Mutib, K.; Sannino, G. Dysphonia Detection Index (DDI): A New Multi-Parametric Marker to Evaluate Voice Quality. *IEEE Access* **2019**, *7*, 55689–55697. [\[CrossRef\]](#)

39. Trancoso, I.; Correia, J.; Teixeira, F.; Raj, B.; Abad, A. Analysing Speech for Clinical Applications. In *Statistical Language and Speech Processing*; Dutoit, T., Martín-Vide, C., Pironkov, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–6.
40. Boulos, M.; Wheeler, S.; Tavares, C.; Jones, R. How smartphones are changing the face of mobile and participatory healthcare: An overview, with example from eCAALYX. *BioMed. Eng. Online* **2011**, *10*, 1–14. [[CrossRef](#)]
41. Baker, S.; Xiang, W.; Atkinson, I. Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities. *IEEE Access* **2017**, *5*, 26521–26544. [[CrossRef](#)]
42. Zhang, K.; Ni, J.; Yang, K.; Liang, X.; Ren, J.; Shen, X. Security and Privacy in Smart City Applications: Challenges and Solutions. *IEEE Commun. Mag.* **2017**, *55*, 122–129. [[CrossRef](#)]
43. Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S.; Wang, G. Security and Privacy in the Medical Internet of Things: A Review. *Secur. Commun. Netw.* **2018**, *2018*, 5978636. [[CrossRef](#)]
44. Teixeira, F.; Abad, A.; Trancoso, I. Patient privacy in paralinguistic tasks. In Proceedings of the Interspeech, Hyderabad, India, 2–6 September 2018; pp. 3428–3432. [[CrossRef](#)]
45. Marrone, S.; Tortora, A.; Bellini, E.; Maione, A.; Raimondo, M. Development of a testbed for fully homomorphic encryption solutions. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021.
46. Lopez-Otero, P.; Docio-Fernandez, L. Analysis of gender and identity issues in depression detection on de-identified speech. *Comput. Speech Lang.* **2021**, *65*, 101118. [[CrossRef](#)]
47. Cao, B.; Zheng, L.; Zhang, C.; Yu, P.; Piscitello, A.; Zulueta, J.; Ajilore, O.; Ryan, K.; Leow, A. DeepMood: Modeling mobile phone typing dynamics for mood detection. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; Volume F129685, pp. 747–755. [[CrossRef](#)]
48. Lopez-Otero, P.; Docio-Fernandez, L.; Abad, A.; Garcia-Mateo, C. Depression detection using automatic transcriptions of de-identified speech. In Proceedings of the Interspeech 2017, 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, 20–24 August 2017; pp. 3157–3161. [[CrossRef](#)]
49. Boersma, P.; Weenink, D. Praat: Doing Phonetics by Computer [Computer Program], Version 6.1.51; 2021. Available online: <https://www.praat.org> (accessed on 31 August 2021).
50. Hamade, N.; Hodge, W.; Rakibuz-Zaman, M.; Malvankar-Mehta, M. The Effects of low-vision rehabilitation on reading speed and depression in age related macular degeneration: A meta-analysis. *PLoS ONE* **2016**, *11*, e0159254. [[CrossRef](#)] [[PubMed](#)]
51. Breznitz, Z. Verbal indicators of depression. *J. Gen. Psychol.* **1992**, *119*, 351–363. [[CrossRef](#)] [[PubMed](#)]
52. Gratch, J.; Artstein, R.; Lucas, G.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; De Vault, D.; Marsella, S.; et al. The distress analysis interview corpus of human and computer interviews. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 26–31 May 2014; pp. 3123–3128.