Dementia Detection Using LSTM and GRU

Neha Shivhare[1], Shanti Rathod[2], M. R. Khan[3]
[1, 2]*Electronics and Telecommunication Department, Dr. C. V. Raman University, Kota, Bilaspur C. G., India.*
[3]*Electronics and Communication department, GEC Jagdalpur C.G., India.*

Abstract:
Neuro-degenerative infections, like dementia, can affect discourse, language, and the ability of correspondence. A new report to work on the precision of dementia identification examined the utilization of conversation analysis (CA) of meetings between patients and nervous system specialists to recognize reformist neuro-degenerative (ND) memory issues patients and those with (non-reformist) FMD (Functional Memory Disorder). In any case, manual CA is expensive for routine clinical use and hard proportional. In this work, we present an early dementia discovery framework utilizing discourse acknowledgment and examination dependent on NLP method and acoustic component handling strategy apply on various element extraction and learning using LSTM (Long Short-Term Memory) and GRU which strikingly catches the transient provisions and long haul conditions from authentic information to demonstrate the abilities of grouping models over a feed-forward neural organization in estimating discourse investigation related issues. Dementia dataset is taken where the audio file is considered for speech recognition analysis on basis of that data is generated and it is predefined given in dementia data databank. That audio file is converted to text based on speech analysis. Using LSTM and GRU gives efficient results.

Keywords: Dementia Detection, Long Short-Term Memory, GRU, Speech Analysis, features extraction.

*Corresponding author Email: neharai10@gmail.com

1. Introduction

The most common sort of dementia is Alzheimer's Disease (AD) among older people with considerable and extended financial expenses for society. This is portrayed by reformist and irreversible intellectual diminishing, loss of memory, misguided thinking and language, likewise other intellectual disabilities and conduct indications, which are eventually seriously enough to keep a person from working together, social, or privately-owned company. Patients have dynamically genuine handicaps as the sickness creates and ultimately become completely dependent. Early and exact AD analysis is exceptionally useful in making arrangements for the future, in the early treatment of sickness indications, and their families. According to current criteria, when dementia exists and other possible causes have been excluded, diagnosis is expressed in various degrees of certainty or with probable AD, but unambiguous diagnosis requires that a typical autopsy of AD be pathologically modified in the brain tissues be demonstrated [1-3]. Wordy memory debilitation is a clinical trademark and first sign of AD. Other intellectual inadequacies are for the most part currently obvious in your language, direction, leader measures, constructional skills&perceptual capacities all through the clinical show. Irritability, apathy, anxiety, depression, hallucinations, delusions, disinherits ion, aggressive motor activity, and changes in eating or sleeping habits are associated with psychological or behavioral symptoms [4-5]. Every one of these indications lead to an inadequacy in regular living exercises of family, society, and business as the ailment propels from gentle to direct and severe.Alzheimer's sickness (AD) is analyzed clinically and requires affirmation of a reformist dementia condition and prohibition of other potential reasons for dementia through clinical history and assessment, complete blood workup tests, and mind imaging investigation tests likeCT (Computed Tomography) or MRI (Magnetic Resonance Imaging). In this specific situation, it would be vital to foster non-obtrusive finding procedures for early recognition and order of variouskinds of dementia, particularly as there is no requirement for gifted staff or research center gear so everyone can apply these methods in quiet's routine climate without obstructing or modifying patient after legitimate preparing [6-7]. The capability of ESA (Emotional Speech Analysis) is: Emotions are intellectual cycles that are unequivocally associated with Human Mind Architecture, including dynamic, recollections, or consideration, additionally that arise when needed to live in changing and generally unusual world in clever regular or fake frameworks. [8-9].

The remaining paper is organized as follows: Section II. Literature Review work III. Section The proposed LSTM GRU model algorithms are presented in Section III followed by the Problem Statement. Section IV. Presents and discusses the results of the model simulation. Finally, in section V, conclusions & suggestions are given for future work.

2. Literature Review

Haider et al. [10] **i**ncluded comprehensive research with a computational paralinguistic perspective on predictive usefulness of simply acoustic characteristics automatically derived from sponge speech for Alzheimer's disease recognition. A balanced sample of spontaneous speech dataset of DementiaBank, with patients matched by age &gender, evaluated efficiency of different state-of-the-art paralinguistic feature sets for Alzheimer's detection. The evaluated feature sets included the expanded eGeMAPS(Geneva Minimalist acoustic parameter set), emobase feature set, ComParE 2013 set, & new multi-Resolution Cochleagram Features (MRCG). They offered a new approach for feature extraction in Alzheimer's dementia recognition using Active Data Representation (ADR). The results demonstrate which categorization models depend only on acoustic speech features collected by their ADR technique may attain accuracy equivalent to models using higher-level language characteristics. Results analysis indicates that all feature sets provide non-feature set information. They show that while eGeMAPS offers a little higher precision than other feature sets (71.34 percent), a hard fusion of feature sets increases precision to 78.70 percent.

PARO's objectives and functions have been explained by Shibata et al. [11]. Secondly, since numerous observational studies on the treatment of the elderly with dementia in PARO are carried out, several typical examples are mentioned as well as some interesting special cases. These instances include depression recovery, agitation decrease, and speech disruption recovery. Finally, this article examines why PARO can improve older people's emotions and behavior without using a pharmacological method.Lopez-de-Ipina et al. [12]examined the possibility of using intelligent algorithms to enhance both early detections of AD and severity of noninvasive analytical procedures on suspect patients. The study is built upon traditional and novel speech properties, emotional response automatic analysis (ERAA); ET (Emotional Temperature) & Higuchi FD (Fractal Dimension). In addition to the noninvasive, this approach offers the excellent benefit of being low-cost & free of side impacts. It is a pre-clinical study for future diagnostic techniques and biomarkers to be validated. For defining features orientated to the early diagnosis of AD, ERAA demonstrated quite good and promising outcomes.

The implementation of the Voice XML Dialog Manager (VoiceON) has been explained by Firouzian et al. [13]and suggested for voice facilitated dialogues. The major objective of the proposed system is senior citizens who suffer from mild & moderate dementia. Human aspects of the multi-mode interface are examined and numerous situations are assessed in an experiment with senior citizens.Chien et al. [14]proposed a new representation of the feature sequence and used a recurrent neural network to perform classification in this article. To confirm their approach, an experiment has been carried out using 150 speakers, with a rating of 0.95, which may surpass existing state-of-the-art methods in terms of a region under the receiver operating feature curve.

Automated technique for evaluating the language and speech characteristics from 92 subjects' neuropsychological recording in the Framingham Heart Study has been done by Alhanai et al. [15]. To categorize and choose highly predictive characteristics, a total of 265 features in an elastic-net regularized binomial regression model were employed. Comparing the performance of the population model in the larger study cohort with 6,258 subjects (0.79 AUC), the system with incorporated text and audio functionality (0.92 AUC) performed best, with the True Positive Rate of 29 percent (0 percent False Positive Rate) and a good model (Test HosmerLemeshow>0.05). They also showed that lowering pitch and jitter, shorter speech segment, and response phrases are positively linked to cognitive impairment.

3. Proposed Work

3.1.1Problem Identification

Whereas automated interaction analysis is a relatively new area of study & technology that was not used for differential memory diagnosis, important research was undertaken using ML methods to detect dementia symptoms in patients' speech & language. Manual CA for frequent clinical use is expensive & difficult to scale. Automatic classification is a new and difficult field of study that has produced some promising results. Automatic Speech Analysis of Conversations using data analytics (like Machine Learning and Deep Learning) is a critical task with NLP because it generally uses Text annotation, correction, and cleaning are done on the second features data frame which is complex. As with the descriptive studies, the primary issues were with patient selection.

3.1.2Proposed Methodology

Dementia dataset is taken where the audio file is considered for speech recognition ,this audio file is converted to text based on speech analysis.

Then creating the data frame with the help of the plan ACQ library. Identifying the number of Utterances from the data frame. Then we identify participants for defining the part of speech and calculated the number of words. Then identify the tagged word using the part of speech. From participant’s information, extracting Reg-Ex and encoding them UTF-8. After all the processing labeling the dataset as which file or information lies in what section is it belongs to Control or Dementia. Shuffle both label data into one data frame, using the spacy audio model “en_core_web_sm” to create tagged dialogues. Finally, creating a dictionary with information like interjection, pronoun, noun, proper noun, etc. Regular expression and formed a final data frame with 9 features like: label, sentence, text, pos_text, pos_, pos_text_complete, pos_complete, new_text, text_for_pos. We are getting some things sentence and clean data as well as for dementia label. Tokenize the dataset, generating the full interview feature file then performing sentiment analysis on the data frame. Then load the data for the training.

3.2 Data preprocessing

Creating the data frame with the help of the PylangACQ library.PylangACQ is a Python library for language acquisition research.Easy access to other TalkBank datasets Intuitive Python data structures for flexible data access and manipulation. Standard developmental measures readily available: TTR (Type-Token Ratio), MLU (Mean length of utterance), &IPSyn (index of productive syntax)

This library is used for the audio processing of the control label's Files, identifying the number of Utterances from the data frame.

How GRU Works: Let's now examine how these gates work. It uses a two-step approach to identify Hidden state H_t in GRU. 1st stage is to build what is called the hidden state of candidates. As demonstrated below.

Candidate Hidden State

The input and hidden state from the previous timestamp t-1 are taken into account, multiplied by the reset output rt. The result is the value of the candidate's hidden state, which afterward sent the complete information to the tanh function.

$$\widehat{H_t} = \tanh\left(X_t * U_g + (r_t \circ H_{t-1}) * W_g\right) \qquad \text{(A)}$$

The most significant aspect of this equation is how we use a value of reset gate to regulate how much preceding hidden state may impact candidate state.

If the r_t value is equal to 1, whole H_t-1 data from the preceding hidden state will be taken into account. Also, if the rt value is zero, information from the hidden state is disregarded altogether.

Hidden state: The current hidden state $H_t$is utilized to produce a candidate state once we obtain a candidate state. This is where the update gate enters the image. Rather than utilizing a separate gate, such as in LSTM in GRU, we utilize a single update gate to manage both historic information that is Ht-1 and new information coming from the candidate state.

$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \widehat{H}_t \qquad \text{(B)}$$

Now imagine that a value of u_t about 0 will disappear, such that the first term of the equation does not include much information about the previously hidden state. The second part becomes, on the other hand, almost one which implies that the concealed state solely consists of information from the candidate state at the present stamp.

$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \widehat{H}_t \qquad \text{(C)}$$

Also, if u_t's value on the 2nd term is 0, the currently hidden state H_t is dependent on the 1st term, i. e., information from the hidden state at preceding timestamp t-1.

$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \widehat{H}_t \qquad \text{(D)}$$

We may thus infer that in this equation the value of ut is extremely crucial and might be between 0 and 1.

LSTM Model

It is a specific type of RNN that may develop long-term data dependencies. This is because the recurring module of the model is engaging with each other in a combination of four layers.

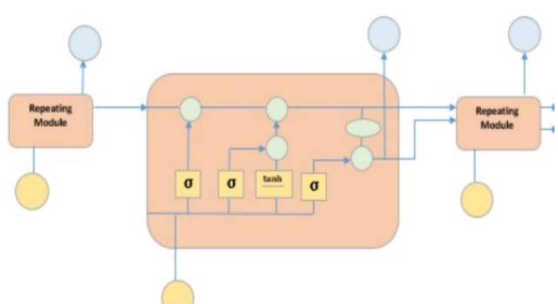


Fig 1: LSTM model

The accompanying image shows four layers of a neural network in yellow boxes, wise point operators in green circles, yellow circle input, and blue cell state. The cell state of an LSTM module and the 3 gates enable you to collect, unlearn or keep selectively information from each unit. The LSTM cell state allows the information to flow across the units without changing the interactions. Each unit contains a cell state with input, output & forget gate to add or delete data. Forget gate chooses information for which it employs a sigmoid function from preceding cell state must be forgotten. With the 'sigmoid' and 'tanh' multiplication operations, the input gate regulates information flow into the current cell state. Finally, the output gate determines information be sent to the next hidden state.

3.2.1 Add Model layer

Input layer: input sentence to this framework.

Embedding layer: map every word into a lower dimension vector.

GRU and LSTM layers: Gated recurrent units (GRUs) are the gate mechanism in recurrent Neural networks, invented by Kyunghyun Cho et al. [17] in 2014. GRU has a forgotten gate, but fewer parameters than LSTM, because it doesn't have a gate to output.

Activation: To utilize activation function. Default: hyperbolic tangent (tanh). No activation (i.e. "linear" activation: a(x) = x) if you pass None is applied.

Dense: The term refers to which neurons in a network surface are fully linked (dense).A dense surface is a fully interconnected surface, indicating that all neurons in 1 layer are coupled to those in the next.

3.2.2 Proposed Algorithm

Step1: Collect the Dementia Dataset of speech audio files (.cha).

Step2: Audio file is considered for speech recognition and text data generation.

Step3: After that Creating a data frame with the help of the pylangACQ library.

Step4: First, we process the control label's files, identifying the number of Utterance from the data frame.

Step5: Then we identify participants for defining the part of speech and calculated the no. of the word then identify the tagged word.

Step6: Then repeating the same process for the Dementia label.

Step7: From participants' information, extracting RegEx and encoding them UTF-8.

Step8: Now Encode the tokenized data using UTF-8 that consists of a tokenized list and tokenized id that creates a data frame using text, level, and id in dementia list.

Step9: Shuffle both label data into one data frame, using the spacy audio model "en_core_web_sm" to create tagged dialogues.

Step10: Replacing special tags with regular expression and formed a final dataframe with 9 features like: label, sentence, text, pos_text, pos, pos_text_complete, pos_complete, new_text, text_for_pos.

Step11: We are getting some things sentence and clean data as well as for dementia label.

Step12: Tokenize the dataset, generating the full interview feature file then performing sentiment analysis on the data frame.

Step13: Then loading the data for the training.

Figure 2 shows the flow chart of proposed algorithm. This shows step by step process for the same.

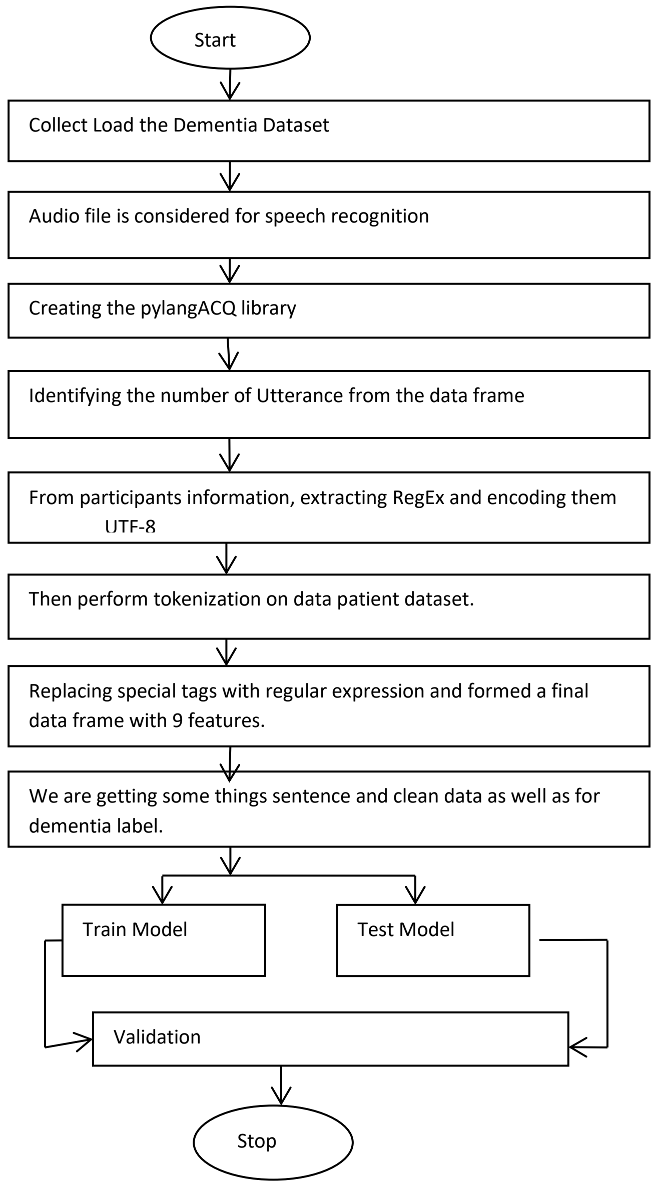


Fig.2: Flow Chart of Proposed Work

4. Results and Discussion

This research has been conducted out using Python programming language and the platform is a Jupyter notebook (version 6.3.1). here, we have used the Dementia dataset. Experiment. The description of such dataset and achieved results of the proposed model has given below.

4.1 Dataset description

The survey uses the biggest freely available database of transcripts and Dementia bank (Boller and Becker,2005),

Fig. 3: The explanation task of a Boston cookie theft. All activities in the picture were asked from participants [16]

Patient interviews with audio recordings (and control). 1) Patient was requested to carry out a variety of tasks; for example, patients were shown an image in the description of 'Boston Cookie Theft' (See Figure 3). The "recall test" also includes recalling the features of a story that was previously told to the patients. Each Dementia bank transcript has an automatic syntactic morph analysis, like standard part-of-speech tagging, tense descriptions, & repetition markers. 2) Note that these characteristics are generic linguistic properties that are automatically extracted and are not AD-specific. To utilize as data samples, we have broken every transcript into separate statements. Note that we have also removed utterances without POS tags. This balance lowered the number of details but made sure that the models were compared with tagged and non-tagged parameters.

The transcripts and audio files have been acquired under the greater protocol of the University of Pittsburgh School of Medicine, overseen by the Alzheimer & Related Dementia studies. NIH grants AG005133 and AG003705 to the University of Pittsburgh were funded in the first gathering of the Dementia Bank data. Elderly controls, individuals with likely and potential Alzheimer's disease were included. Longitudinally, the data were collected annually [16].

4.2 Performance Matrix

1. Accuracy

Accuracy is a measure of how many predictions your model has done for the whole test dataset. The following formulation measures it:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Recall

Recall – or the true positive rate – is the measure of how many true positive values of all the positive ones in the data set are expected. Sensitivity is also occasionally termed. The following formula collects the measurement:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

3. Precision

Precision is a measure of the accuracy of a positive forecast. In other terms, it signifies how positive you can be if a result is projected as good. The following formula is used for calculation:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

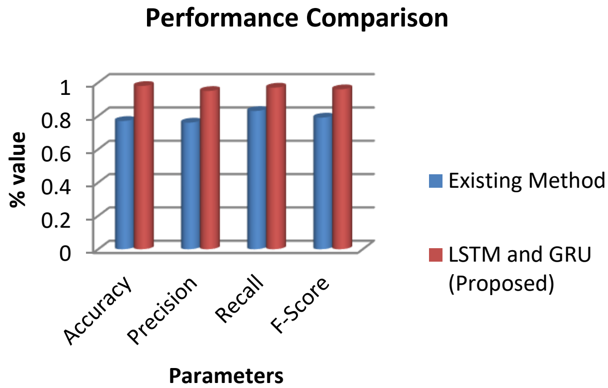
4. F-score

The F1-score is the F-score most frequently utilized. It is a combination of accuracy and memory, that is, its harmonic significance. The following formula allows you to compute the F1-score:

$$\text{F1} = 2.\frac{Precision.recall}{Precision+recall} \quad (4)$$

4.3 The outcome of the Model

The comparison of several classification methods is represented in Fig 3. It represents overall performance comparison output in contrast to several existing methods and LSTM and GRU Method like Accuracy, Precision, Recall, and F-Score.This Comparison is with existing method like SVM.



Parameter	Existing Method (LSVM)	LSTM and GRU (Proposed)
Accuracy	0.77	0.98
Precision	0.76	0.95
Recall	0.83	0.97
F-Score	0.79	0.96

Fig.4: Performance parameters for Existing Method and Proposed LSTM and GRU

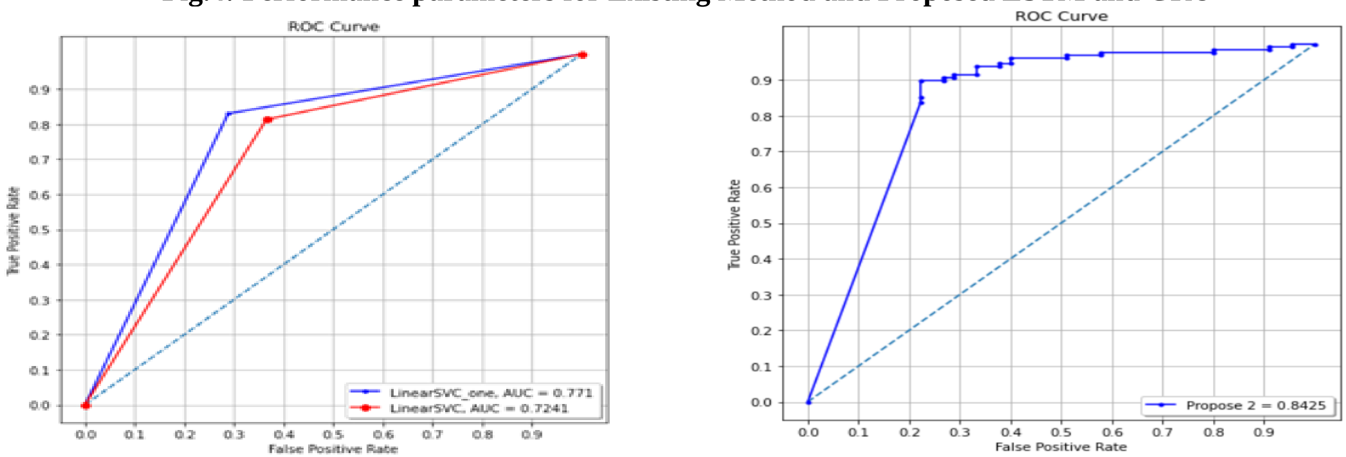


Fig 5: ROC Graph

5. Conclusion

According to prior research, dementia impairs speech in the following ways: syntactic, semantic, info, & acoustic impairment. They used a hybrid technique to increase automatic AD prediction by comparatively a small targeted speech dataset without expert-defined linguistic features. On Cookie-Theft image explanation test of Pitt corpus, they assessed recently built pre-trained transformer-based language frameworks that they supplemented by augmentation approaches. Using sentence level LSTM and GRU the accuracy (98%), precision (95%), recall (97%), and F1 scores of (96%) were obtained which enhanced state-of-the-art outcomes. There are pre-trained language frameworks accessible in a variety of languages. As a result, a technique in this study may be tested in languages other than English. Furthermore, using multilingual versions of these frameworks, information of AD prediction in one language may be transferred to another language when an adequately big dataset is not available. Future work will include expanding the size of our dementia dataset, developing additional LSTM and GRU for dementia detection, and replacing manual annotation with automatic speech divarication and identification.

Conflict of interest:

The authors declare that they have no known competing financialinterests or personal relationships that could have appearedto influence the work reported in this paper.

References:

[1]. McKahnn G, Drachman D, Folstein M.(1984)"Clinical diagnosis of AD: report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease" Neurology. 34, 939-44. https://doi.org/10.1212/WNL.34.7.939

[2]. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC.(2011)"The diagnosis of dementia due to Alzheimer's disease: Recommendations from the NIAA Association workgroups on diagnostic guidelines for AD" Alzheimers Dement. 7(3) 263-269. https://doi.org/10.1016/j.jalz.2011.03.005
[3]. Van de Pole, L.A., et al., (2005) "The effects of age and Alzheimer's disease on hippocampal volumes, a MRI study" Alzheimer's and Dementia.1(1) 51. https://doi.org/10.1016/j.jalz.2005.06.205
[4]. Morris JC, (1991)"The Clinical Dementia Rating (CDR) : current version and scoring rules"Young. 411588-1592.
[5]. American Psychiatric Association.(2000)"Diagnostic and Statistical Manual of Mental disorders" 4th Edition Text Revision. Washington D C.
[6]. Faundez-Zanuy M, Hussain A, Mekyska J, Sesa-Nogueras E, Monte-Moreno E, Esposito A, Chetouani M, Garre-Olmo J, Abel A, Smekal Z, Lopez-de-Ipiña K(2013)."Biometric Applications Related to Human Beings: There Is Life beyond Security" Cognitive Computation 5(1) 136-51. DOI 10.1007/s12559-012-9169-9. https://doi.org/10.1007/s12559-012-9169-9
[7]. Lopez-de-Ipiña K, Alonso JB, Solé-Casals J, Barroso N, Faundez-Zanuy M, Ecay-Torres M, Travieso CM, Ezeiza A, Estanga A(2012)"Alzheimer Disease Diagnosis based on Automatic Spontaneous Speech Analysis"IJCCI 2012 - Proceedings of the 4th International Joint Conference on Computational Intelligence, 698-705. https://doi.org/10.1007/978-3-642-35395-6_55
[8]. M. L. Knapp.(1980) E ssentials of nonverbal communication. H olt, Rinehart & Winston.
[9]. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG.(2001)"Emotion Recognition in Human-Computer Interaction". IEEE Signal Processing Maga zine, 18(1) 32- 80. https://doi.org/10.1109/79.911197
[10]. Haider, F., De La Fuente, S., & Luz, S.(2020) "An Assessment of Paralinguistic Acoustic Features for Detection of Alzheimer's Dementia in Spontaneous Speech" IEEE Journal of Selected Topics in Signal Processing,14(2) 272-281. https://doi.org/10.1109/JSTSP.2019.2955022
[11]. Shibata, T. (2012)"Therapeutic Seal Robot as Biofeedback Medical Device: Qualitative and Quantitative Evaluations of Robot Therapy in Dementia Care"Proceedings of the IEEE, 100(8) 2527-2538. https://doi.org/10.1109/JPROC.2012.2200559
[12]. Lopez-de-Ipina, K., Alonso, J. B., Travieso, C. M., Egiraun, H., Ecay, M., Ezeiza, A., … Martinez-Lage, P. (2013). "Automatic analysis of emotional response based on non-linear speech modeling oriented to Alzheimer disease diagnosis" 2013 IEEE 17th International Conference on Intelligent Engineering Systems (INES)San Jose, Costa Rica. https://doi.org/10.1109/INES.2013.6632783
[13]. Firouzian, A., Pulli, P., Pleva, M., Juhar, J., & Ondas, S. (2017). "Speech interface dialog with smart glasses" 2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)Stary Smokovec, Slovakia. https://doi.org/10.1109/ICETA.2017.8102483
[14]. Chien, Y.-W., Hong, S.-Y., Cheah, W.-T., Fu, L.-C., & Chang, Y.-L. (2018). "An Assessment System for Alzheimer's Disease Based on Speech Using a Novel Feature Sequence Design and Recurrent Neural Network" 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Miyazaki, Japan. doi:10.1109/smc.2018.00557. https://doi.org/10.1109/SMC.2018.00557
[15]. Alhanai, T., Au, R., & Glass, J. (2017). "Spoken language biomarkers for detecting cognitive impairment" 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) Okinawa, Japan. https://doi.org/10.1109/ASRU.2017.8268965
[16]. https://dementia.talkbank.org/access/English/Pitt.html.
[17]. Kyunghyun Cho ,Junyoung Chung, Caglar Gulcehre, Yoshua Behgio(2014)"Empirical of Gated Recurrent Neural Network on Sequence Modelling" Dec 2014,NISP 2014,Deep Learning and Representation learning workshop.