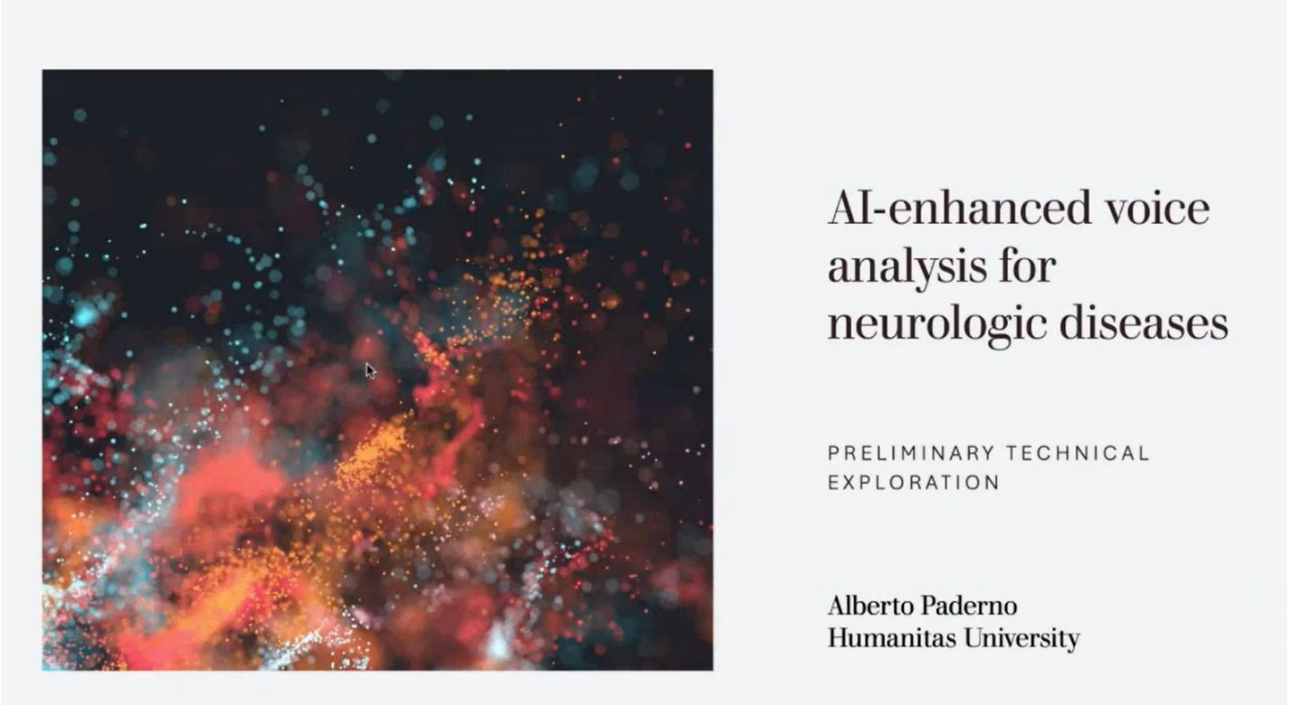
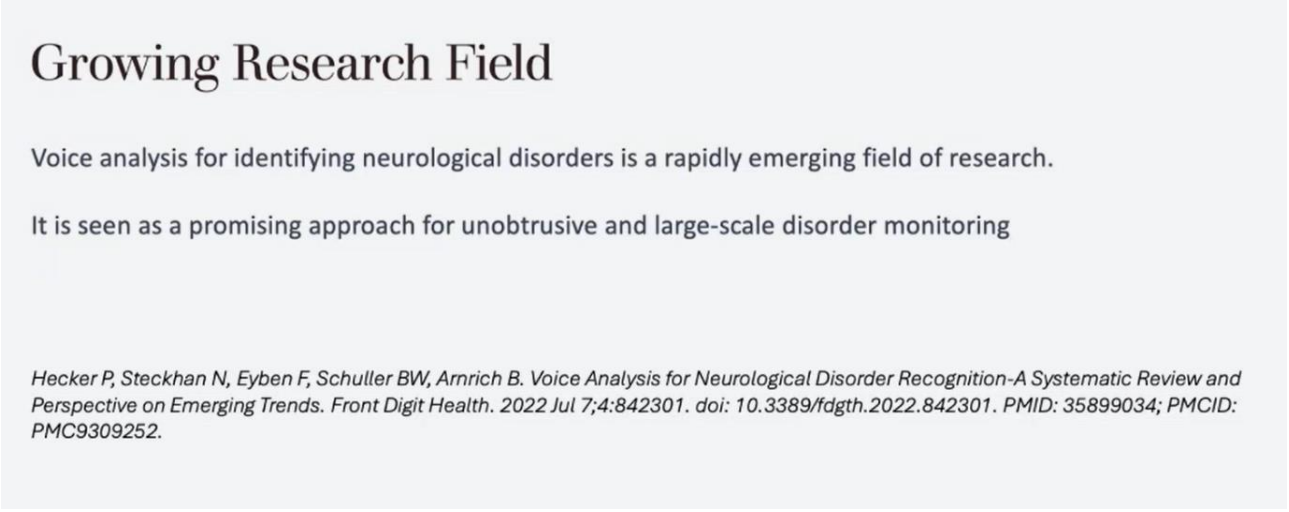
PRESENTATION

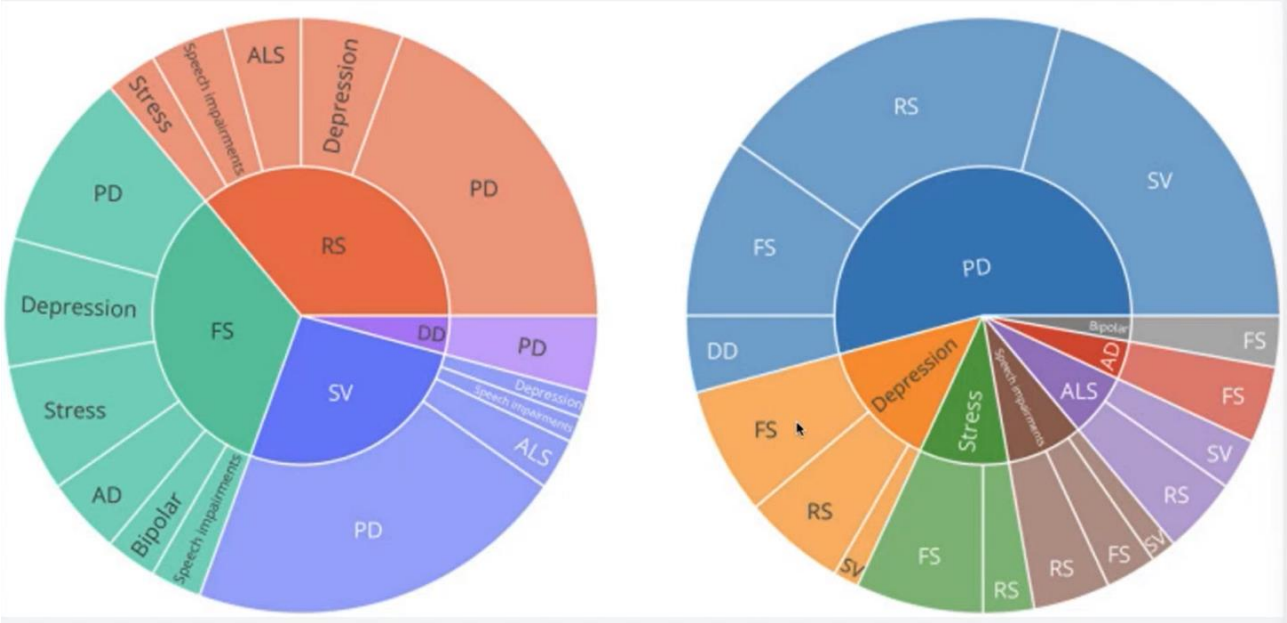
Alberto:



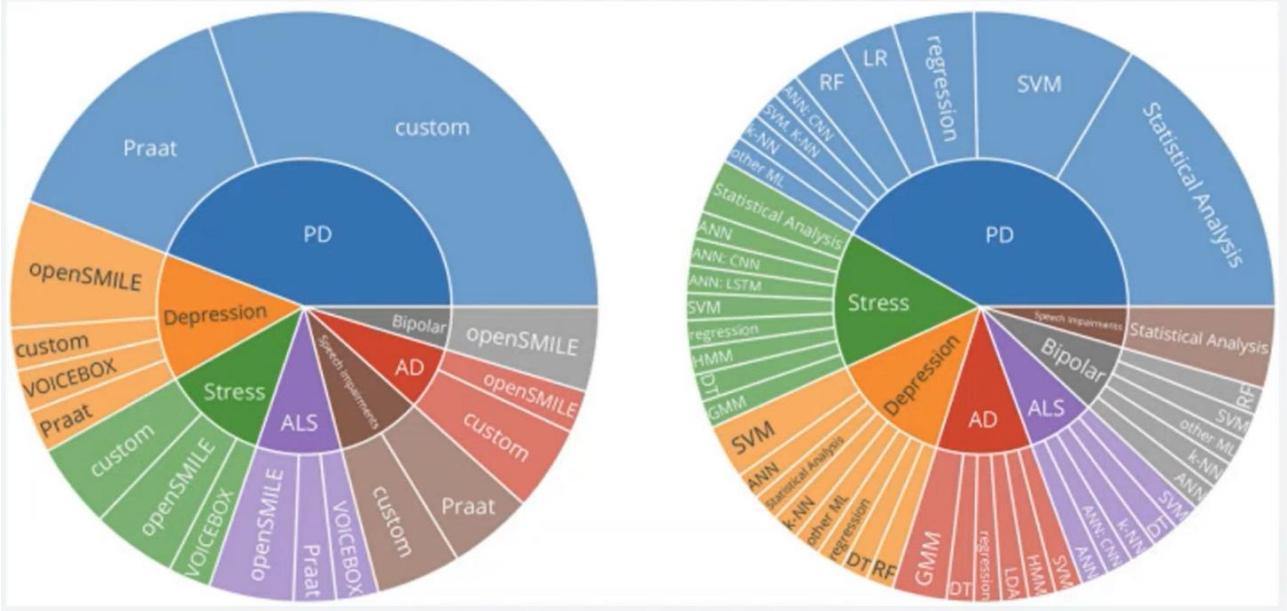
Of course. So it's a very informal presentation where I wanted to frame the topic of A I and force analysis, especially focusing on neurology, neurological diseases. But it it can be applicable to many type of fields. And it's a markedly technical exploration and in a sense that I wanted to understand what's the best approach in terms of A Ienhanced analysis of these diseases.



So when uh we talk about this type of analysis, it's pretty clear when looking in the literature that it's a rapidly emerging field of research and it's seen as particularly promo uh promising and it's pretty similar to other biomarkers fields. So uh people are trying to find noninvasive biomarkers like liquid biopsy and cancer, for example, in order to get a diagnosis without invasive procedures for for the patient.



And that's a uh pretty uh comprehensive evaluation of the field. You can see here that it's uh evaluating the type of source data on the left and the type of disease on the right that has been evaluated with voice biomarkers. And in particular on the left, you can see that there's a pretty uh homogeneous distribution between uh recorded speech RS free speech FS and other types of biomarkers. And this has been applied to different neurologic diseases, especially uh when, when talking about voice analysis. And if we go to the right, we can see that Parkinson's disease. A PD is the main uh domain where people have applied this type of analysis. It's particularly interesting because uh different modalities of evaluation have been applied to, to Parkinson's disease and it's more than 50% of uh the type of applications. Uh And that's why I I would try to focus on that. Uh with my, with my talk

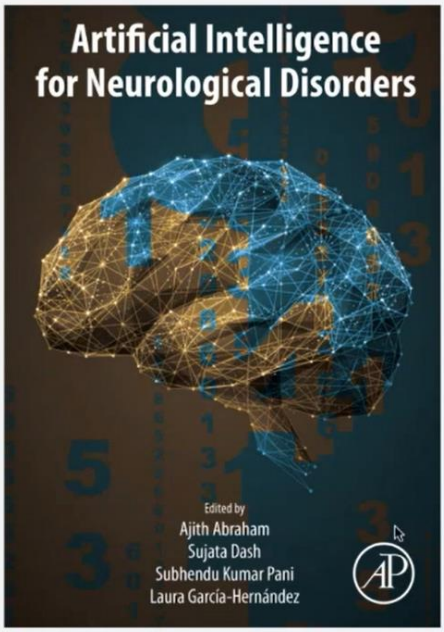


just to give a view of uh of the the entire topic. So let's go to the type of evaluation. And I don't want you to look at all the types of evaluations here. I just want you to get an idea of the heterogeneity and the variability of the types of evaluation both on the left when considering the type of software that has

been used for the analysis and on the right when considering the type of statistical analys. So we are in a field where uh there's a huge variability and we will see where is the problem in this presentation? I will mostly focus on uh the current issues and how to solve them and how to find a solution uh to, to this type of issues.

Noninvasive screening methods for speech analysis

Neurological disorders are increasing worldwide, thus creating a public health crisis. Early diagnosis is key to detect and prevent these diseases from manifesting. Some of the available procedures to assess or evaluate these disorders involve speech analysis. As described by Lopez-Poveda (2014), ASSA involves testing fluency using Praat software (Rocco et al., 2018), which has the potential to extract voice and unvoiced parts of a sound. It measures the duration, time, frequency, and quality of signals. The brightness of the sound is measured using the spectral centroid (McKechnie et al., 2018). ERA evaluates three characteristics in speech: acoustic characteristics, voice quality, and duration (Schwartz & Pell, 2012). Acoustic characteristics measure the pitch, intensity, standard deviation, and root mean square amplitude of intensity. Voice characteristics measure noise-to-harmonic ratio and harmonic-to-voice ratio. The duration characteristics measure the degree of voice breaks and unvoiced frame. The prosodic and paralinguistic feature sets obtained from the temporal domain of a speech signal are measured using emotional temperature analysis (Schwartz & Pell, 2012).



And it's a field that uh people are saying it's particularly interesting. But if we go and see uh in, in this uh recent book on Artificial Intelligence for Neurological Disorders. It's a 300 pages book and that's the Holy Subchapter on speech analysis. So, I mean, the field has a lot of room for evolution and we will need to understand what's the uh the best approach to identify this type of evolution.

An Overview of the Literature

- Vocal impairments are a common symptom in Parkinson's disease (PD) patients, often presenting as decreased voice loudness, monopitch, and improper consonant articulation.
- Smartphone technology has expanded diagnostic capabilities, enabling the assessment of symptoms like dexterity and gait, and has been found to reliably identify speech problems.
- Acoustic measurements from spontaneous speech are effective in distinguishing between PD patients and controls.
- AI and machine learning techniques are successful in identifying PD using extracted speech data, with some studies reporting high classification accuracy.
- Multi-classifier frameworks, autonomous feature extraction systems, and ensemble learning have been employed to improve PD classification accuracy.

Just uh some bullet points to give you an overall view of the literature. I've done a comprehensive literature research. Uh but uh the different topics are really varied and the different techniques that have been employed in this field are extremely varied. And we can see that um people have tried to identify by the different types of vocal impairments in Parkinson's disease. For example, I'm using Parkinson's disease, for example, as a case study, but these can be applied to many other diseases

and people are trying are trying to uh find a way to use smart phone uh to provide an adequate diagnosis. And uh there has been research showing that acoustic measurements are accurate to distinguish between patients. Parkinson's disease patients and control and even to stage the degree of disease of these patients. And uh we will see that there are different technical approaches in terms of the analysis of voice and Parkinson's disease and neurologic diseases.

An Overview of the Literature

- Limitations of current studies include small cohort sizes, which affect the generalizability of results, and the potential loss of vital information when summarizing voice examinations.
- Identity confounding, where voice samples from the same subjects appear in both training and testing data, may lead to overestimated model performance.
- Deep learning techniques have been suggested to improve model performance, including data normalization, feature selection, and avoiding data leakage.
- Large-scale studies, like the Parkinson's Speech Initiative, aim to distinguish PD individuals from controls using phone-quality voice in non-acoustically controlled environments.
- Ensemble methods, convolutional neural networks, and other machine learning techniques have achieved high accuracy in detecting PD from speech samples.
- Privacy-sensitive methods for classifying PD have been developed, utilizing passively recorded voice calls and language-aware training of classifiers.

But of course, there are also limitations and confounding factors. Uh Some major limitations are the fact that current studies are usually uh small in quart size and have a huge general disability issue. Uh it's difficult to generalize the results and that's not useful when you're not just writing a paper, but your uh main outcome would be to get a, a device or an application that is actually uh usable in the clinical practice. And uh people are starting to test deep learning techniques in this setting. Uh But this is still uh a field in its infancy and we will see that it's different. And when we are speaking about uh text analysis or image analysis, we are really far behind when considering voice. And there are large scale studies like the Parkinson's speech initiative that is not being responsive in the last years. Uh that are uh is trying to uh use phone quality voice uh in order to diagnose patients with Parkinson's disease. But it's a difficult uh issue. It's particularly complex because we don't want to diagnose Parkinson. When it it's already advanced, we want to achieve an early diagnosis. And so the classification is particularly uh complicated.

An Overview of the Literature

- Deep learning (DL) is increasingly used for processing complex voice signal issues in Parkinson's disease (PD) research.
- The process involves converting speech into feature vectors or tensors for DL models, taking into account variations due to the speaker's native language.
- Vocal biomarkers from these features are used for monitoring PD symptoms and severity.
- Multiple DL architectures, like 1D CNN models, have been tested to detect PD with accuracies around 87%.
- Sparse kernel transfer learning has been proposed to improve performance, with reported accuracies of up to 86.7%.
- Deep dual-side learning models with weighted fusion mechanisms have achieved high accuracies, up to 98.4%.
- Traditional CNNs have been utilized on the Max Little dataset with an accuracy of 93.10%.

We will see that uh neural networks or even con convolutional neural networks are being used. So, uh what people are doing is using recurrent uh neural networks to use and analyze the voice as a signal as a temporal segment signal or a convolutional neural networks to analyze this pet program as an image. And so this has also proved to be particularly effective.

An Overview of the Literature

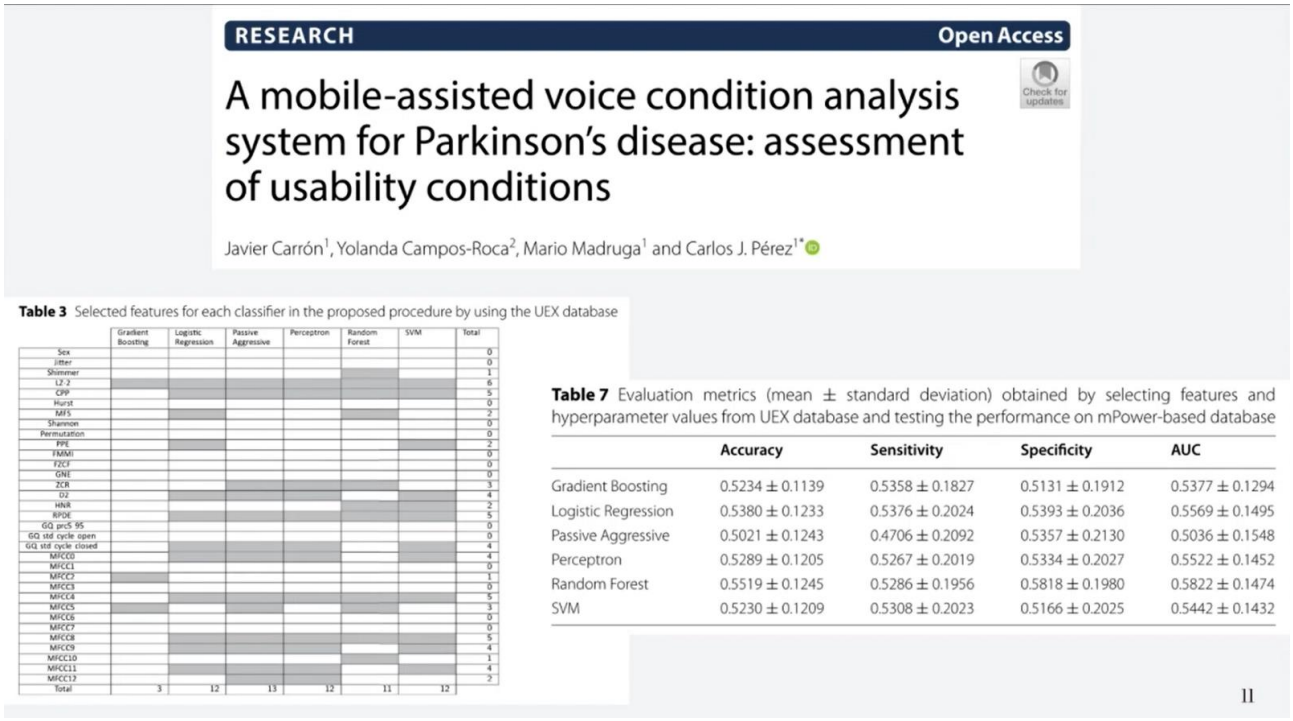
- The use of pre-trained architectures like ResNet18 with audio as image input has shown accuracies of 97.1%.
- Ensemble methods with bagged and boosted weak learners can outperform basic voting and stacking, with an accuracy of 94.12%.
- Deep neural networks have been suggested for predicting PD severity with 82% accuracy, improved by feature selection techniques.
- DNN and LSTM network-based models have shown promise in predicting PD from speech samples with up to 97.1% accuracy.
- Bidirectional LSTM models have been proposed for capturing time-series dynamics in speech, indicating the potential for PD identification.
- Attention-based LSTM models have been developed for specific datasets, like Indian languages, with improved accuracies of 92.5%.

But uh this is just a general overview of the different techniques I will just point out really recent articles to show you uh which are the current approaches and what's the problem here.

Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method

The dataset was **balanced** by the synthetic minority oversampling technique (**SMOTE**) and features were arranged according to their contribution to the target characteristic by the **recursive feature elimination** (RFE) algorithm. We applied two algorithms, t-distributed stochastic neighbour embedding (**t-SNE**) and principal component analysis (**PCA**), to reduce the dimensions of the dataset. Both t-SNE and PCA finally fed the resulting features into the classifiers support-vector machine (**SVM**), K-nearest neighbours (**KNN**), decision tree (**DT**), random forest (**RF**), and multilayer perception (**MLP**). Experimental results proved that the proposed techniques were superior to existing studies in which RF with the t-SNE algorithm yielded an accuracy of 97%, precision of 96.50%, recall of 94%, and F1-score of 95%. In addition, MLP with the PCA algorithm yielded an accuracy of 98%, precision of 97.66%, recall of 96%, and F1-score of 96.66%

So this is a very recent article that shows uh pretty good results. But you can see uh on the B text that the current approach is to manually increase uh the uh the sample size uh particularly considering the disease group with these modes, which is the sen sensitive minority over sampling technique. And then there is a feature selection, a recursive feature elimination and the selection of different machine learning algorithms that only work with numbers. So what you do is to try to find different features inside the voice of the patient and from these features. Uh And you use these features as numbers and you try to analyze these numbers using simple machine learning uh techniques. And this is an approach that people in image analysis were doing uh in 2008-2010 but have stopped doing since 2012. We are, we are really far behind when considering voice.



RESEARCH Open Access

A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions

Javier Carrón[1], Yolanda Campos-Roca[2], Mario Madruga[1] and Carlos J. Pérez[1*]

Table 3 Selected features for each classifier in the proposed procedure by using the UEX database

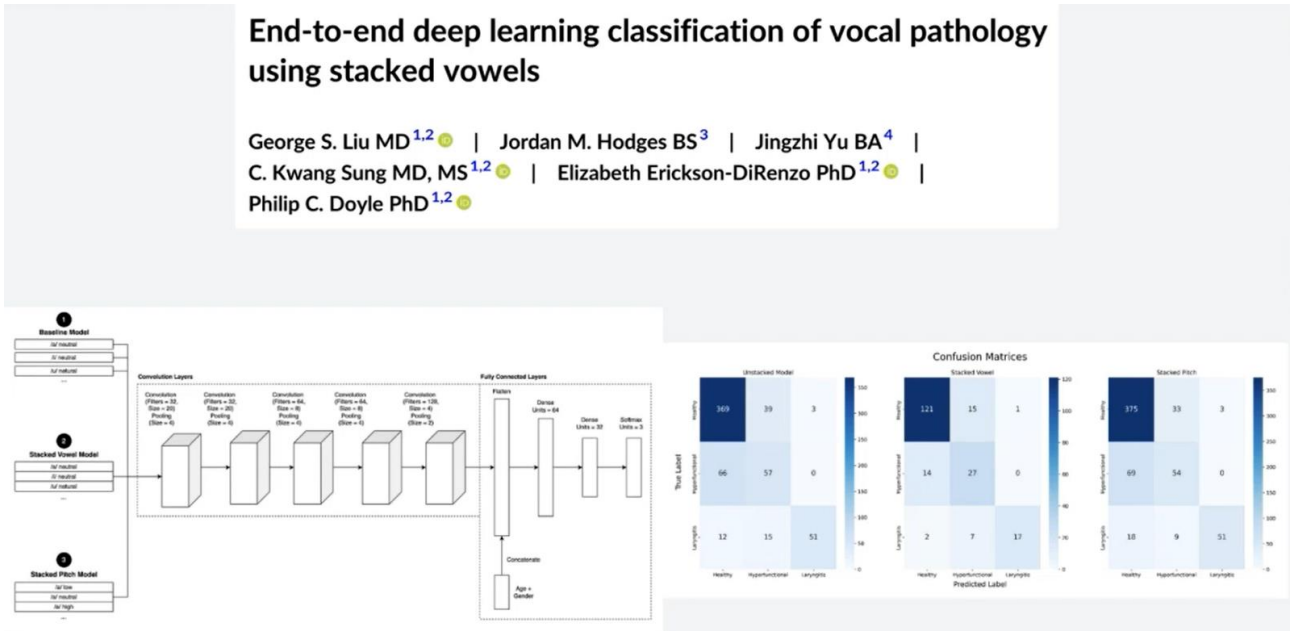
	Gradient Boosting	Logistic Regression	Passive Aggressive	Perceptron	Random Forest	SVM	Total
Sex							0
Jitter							0
Shimmer							1
LZ-2							6
CPP							5
Hurst							0
MFS							2
Shannon							0
Permutation							0
PPE							2
FMMI							0
F2CF							0
GNE							0
ZCR							3
D2							4
HNR							2
RPDE							5
GQ prc5 95							0
GQ std cycle open							0
GQ std cycle closed							4
MFCC0							4
MFCC1							0
MFCC2							1
MFCC3							0
MFCC4							5
MFCC5							3
MFCC6							0
MFCC7							0
MFCC8							5
MFCC9							4
MFCC10							1
MFCC11							4
MFCC12							2
Total	3	12	13	12	11	12	

Table 7 Evaluation metrics (mean ± standard deviation) obtained by selecting features and hyperparameter values from UEX database and testing the performance on mPower-based database

	Accuracy	Sensitivity	Specificity	AUC
Gradient Boosting	0.5234 ± 0.1139	0.5358 ± 0.1827	0.5131 ± 0.1912	0.5377 ± 0.1294
Logistic Regression	0.5380 ± 0.1233	0.5376 ± 0.2024	0.5393 ± 0.2036	0.5569 ± 0.1495
Passive Aggressive	0.5021 ± 0.1243	0.4706 ± 0.2092	0.5357 ± 0.2130	0.5036 ± 0.1548
Perceptron	0.5289 ± 0.1205	0.5267 ± 0.2019	0.5334 ± 0.2027	0.5522 ± 0.1452
Random Forest	0.5519 ± 0.1245	0.5286 ± 0.1956	0.5818 ± 0.1980	0.5822 ± 0.1474
SVM	0.5230 ± 0.1209	0.5308 ± 0.2023	0.5166 ± 0.2025	0.5442 ± 0.1432

11

And another thing that I wanted to show you is that people are also trying to show that there's the possibility to um to analyze the voice of patients with the mobile. But uh what I want to focus on in this article is not the uh typical tables but this table here, you can see that it shows the accuracy and all the metrics when using a different data set uh from the data set used in the study. And you can see that the accuracy is approximately 50 55% which is really, really low. And it's not applicable for a diagnostic uh setting. And it's not even considered as a test. So we can see that there's a huge issue in terms of uh applicability to different data sets. And that's what we will need to do.



End-to-end deep learning classification of vocal pathology using stacked vowels

George S. Liu MD[1,2] | Jordan M. Hodges BS[3] | Jingzhi Yu BA[4] | C. Kwang Sung MD, MS[1,2] | Elizabeth Erickson-DiRenzo PhD[1,2] | Philip C. Doyle PhD[1,2]

And here another pretty recent article which I think is particularly useful because they show that by using stacked vowels, it's uh the result is more effective than by just using a single letter or uh by using

a stopped pit uh with a single letter. So what we will need to do is to have an heterogeneous data set to have the right way to analyze this type of data set. So the right input which is heterogeneous input and then use a technique that is able to manage heterogeneous data and give consistent results.

Article

Artificial Intelligence-Based Voice Assessment of Patients with Parkinson's Disease Off and On Treatment: Machine vs. Deep-Learning Comparison

Giovanni Costantini [1], Valerio Cesarini [1,*], Pietro Di Leo [1], Federica Amato [2], Antonio Suppa [3,4], Francesco Asci [3,4], Antonio Pisani [5,6], Alessandra Calculli [5,6] and Giovanni Saggio [1]

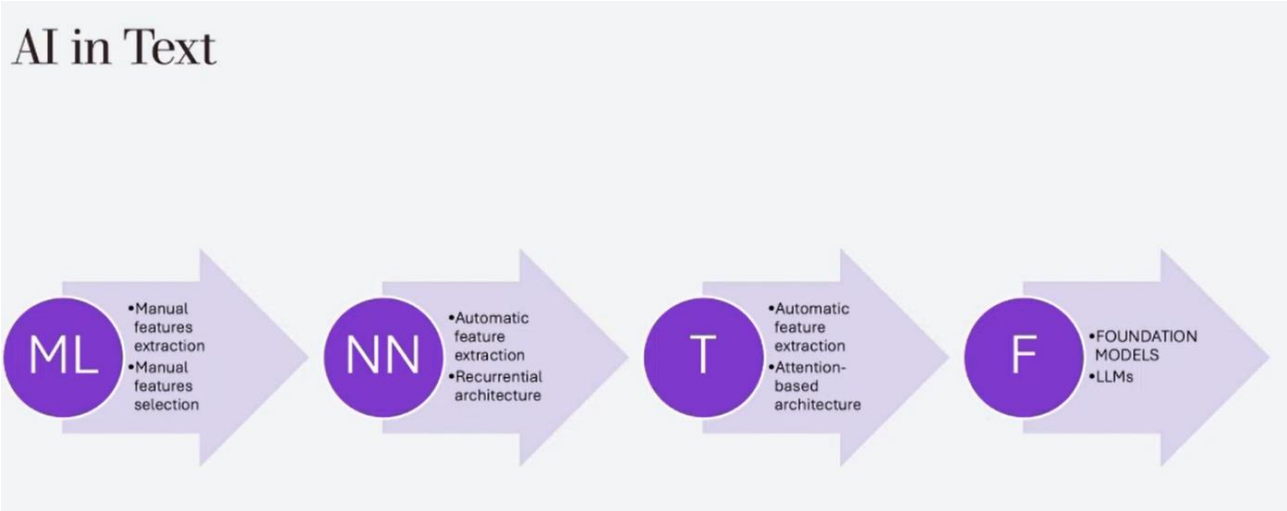
Comparison	Model	Acc	PPV	NPV	Sen	Spec	AUC	F1 Score
1. Mid-Advanced PD vs. HC	KNN	0.80 ± 0.01	0.79 ± 0.03	0.80 ± 0.02	0.80 ± 0.01	0.79 ± 0.02	0.87 ± 0.04	0.80 ± 0.03
	CNN	0.82 ± 0.07	0.87 ± 0.05	0.78 ± 0.06	0.75 ± 0.04	0.87 ± 0.05	0.83 ± 0.05	0.79 ± 0.05
2. Early PD vs. HC	SVM	0.83 ± 0.02	0.81 ± 0.01	0.83 ± 0.01	0.83 ± 0.03	0.82 ± 0.02	0.88 ± 0.05	0.82 ± 0.01
	CNN	0.70 ± 0.06	0.72 ± 0.04	0.75 ± 0.03	0.73 ± 0.04	0.66 ± 0.07	0.73 ± 0.02	0.70 ± 0.03
3. Mid-Advanced PD vs. Early PD	KNN	0.85 ± 0.02	0.77 ± 0.05	0.86 ± 0.02	0.83 ± 0.02	0.81 ± 0.03	0.91 ± 0.06	0.80 ± 0.04
	CNN	0.74 ± 0.09	0.75 ± 0.05	0.76 ± 0.06	0.69 ± 0.08	0.75 ± 0.07	0.75 ± 0.05	0.68 ± 0.05
4. Mid-Advanced PD ON vs. OFF L-dopa	KNN	0.79 ± 0.01	0.71 ± 0.02	0.87 ± 0.05	0.84 ± 0.01	0.75 ± 0.02	0.82 ± 0.03	0.77 ± 0.03
	CNN	0.53 ± 0.08	0.53 ± 0.06	0.57 ± 0.08	0.69 ± 0.05	0.37 ± 0.08	0.58 ± 0.05	0.65 ± 0.06

And here is one of the uh first comparisons of conventional machine learning techniques with deep learning. And we will see what's the main difference here. The main difference is that with deep learning or even with transformer based algorithms, you're not doing feature extraction.

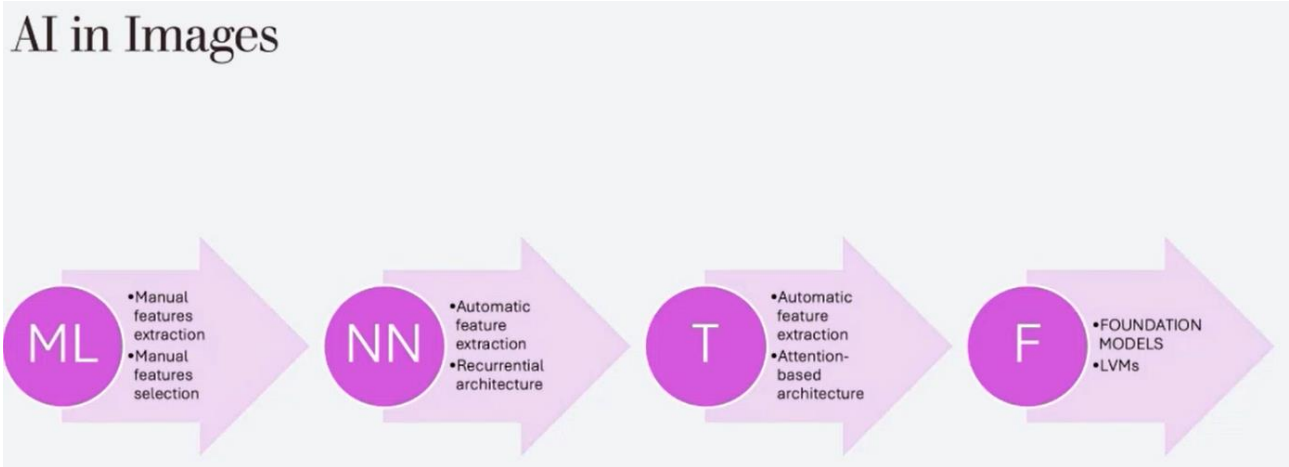
The importance of multimodality

Recent studies	Sample size (PD/HC)	Number of repetitions	Out clinic	Modalities considered							Accuracy	Ensemble[1] improvement
				Voice	Gait	Balance	Dexterity	Rest tremor	Postural tremor	Others		
Current study	37/35	4, 883*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	0.973/0.971 (Sens/Spec) 0.987 (Accuracy) 0.993 (AUC)	Yes
Prince *et al.* (2018) [2]	949/866	NA	Yes	No	No	No	Yes	No	No	No	0.65 (Accuracy)	No
Zhan *et al.* (2018) [3]	129/0	6,148	Yes	Yes	Yes	Yes	Yes	No	No	Reaction Time	0.81 (Pearson correlation)	Yes
Prince *et al.* (2018) [4]	312/236	48,892	Yes	No	No	No	Yes	No	No	Memory	NA	No
Bot *et al* (2016). [5]	1087/5581	78,887	Yes	Yes	Yes	No	Yes	No	No	Memory	NA	No
Zhan *et al.* (2016) [6]	121/105	1,600	Yes	Yes	Yes	Yes	Yes	No	No	Reaction Time	0.693/0.727 (Sens/Spec)	Yes
Neto *et al.* (2017) [7]	23/23	NA	Yes	Yes	Yes	No	Yes	No	No	No	0.5-0.6 (AUC)	No
Arora *et al.* (2015) [8]	10/10	18	Yes	Yes	Yes	Yes	Yes	No	No	Reaction Time	0.962/0.969 (Sens/Spec)	No
Lee *et al.* (2016) [9]	57/87	432	No	No	No	No	Yes	No	No	No	0.92 (AUC)	No
Arroyo-Gallego *et al.* (2017) [10]	21/23	51	No	No	No	No	Yes	No	No	No	0.810/0.810 (Sens/Spec)	No
Kassavetis *et al.* (2016) [11]	14/0	14	No	No	No	No	Yes	No	No	No	NA	No
Printy *et al.* (2014) [12]	18/0	54	No	No	No	No	Yes	No	No	No	NA	No

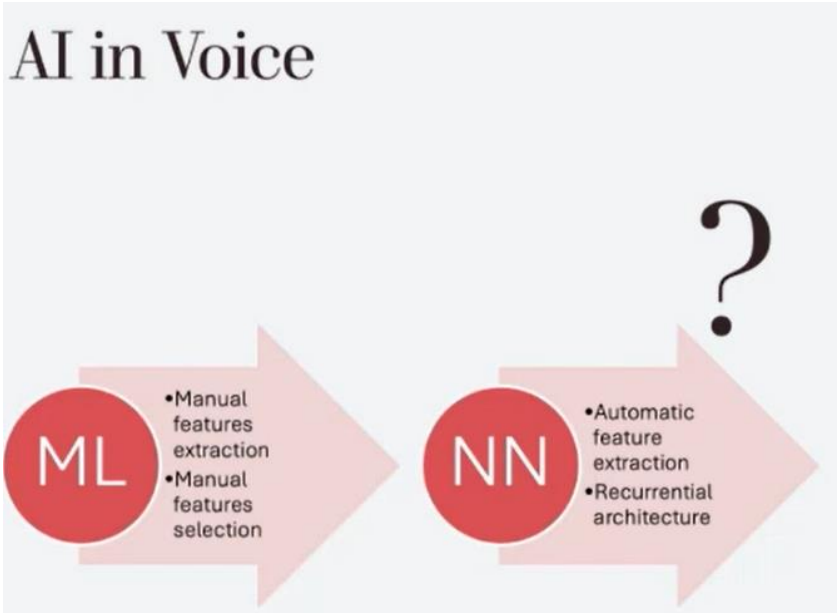
And another important concept is to think about multimodality. So we will need to do voice analysis. But in the most type of applications, people are trying to ST different types of evaluation that are even over voice itself in order to improve the diagnostic accuracy of the model.



So just to give you a comparison with after different fields, what we have seen in uh in text. And this is uh this is typical with ChatGPT. For example, there we are having continuous updates. Yesterday was the last time that the algorithm has been updated. And what people in the research field of AI and TE were doing wars initially, machine learning with manual extraction of features and man manual feature selection. Then in 2012, with Alex net, people started using neural networks with automatic feature extraction and a record architecture. So here the clinician doesn't need to select feature. The feature selection is automatic and it's even more automatic when they started using transformers because this is a different architecture in terms of uh models in deep learning that allow to use automatic attention that is applied to the text, to the image, to the uh voice. And now what people are doing is having a foundation model. So a big model that is able to manage different types of inputs.



And this is pretty similar when considering image, there is been the same type of evolution from manual feature extraction and machine learning to deep learning to transform it and foundation models.



So uh what I'm guessing is that we will need to probably envision the potential evolution even if AI in voice. And even considering that we might consider uh starting to analyze voice with neural networks. Of course transform it which is a new architecture or even build a foundation model for voice.

And how would it work? Uh The potential approach would be to collect a data set for pathological and normal voice and perform with deep learning or transformers a large scale super self supervised learning. What is self supervised learning, self supervised learning is a way to pass in the input without any annotation. And what the algorithm does is to remove part of the information and to ask another part of the algorithm to feel the part of the information that has been removed. In this way, the algorithm builds an understanding of the input and is able to uh summarize the input in a mathematical model. And that's a foundation model. So that from voice, it's possible to have numbers that represents the voice and with these numbers which are voice embeddings, it's possible to uh perform classification tasks with different potential outcomes and different potential algorithms. But the heavy lifting of the field is done by the vision uh the voice foundation model which gives a correct representation of the voice. So uh it's a particularly technical uh presentation. I I know and I'm absolutely open to questions, but I wanted to see uh what's missing right now in the literature and what we can do with a large scale effort in order to get something meaningful for the clinical setting. Thank you. Thank you very much.

DISCUSSION

Mieke:

Alberto is nice, very nice presentation. Very nice overview. Um Does anyone has a question for Alberto? Anyone of the audience? OK. Not yet. So what I wanted to ask to you is um you, you um you just go back to slide 18 (last slide). Can you share slide 18 again? Yes, this one. So you have here the Voice Foundation model. Uh And you talked about um and that is very important that features are transformed to numbers because we are dealing with an algorithm with a mathematical algorithm. This deals with numbers. So we must represent features as numbers. And this is quite easily done in the acoustics. But um as you recall the uh my presentation in our first very first biomarker committee meeting, I was aiming or my ambition was aiming towards representing the multidimensionality of voice into numbers. So not only the acoustic dimension but also the aerodynamics, which is not hard to do, but also the self assessment, the perception and so on. How do you see that working?

Alberto:

Uh I think that's a pretty interesting question and that's basically uh how foundation models are used right now. Uh The main uh the main outcome when using a foundation model is that you are basically able to transform the input to embeddings. And what's interesting there is that when you have embeddings, I mean, uh I I will go a bit more technical here. Embeddings are hyper dimensional vectors. So there are uh these vectors can be represented as a point in a hyper dimensional space. And that is located according to the characteristics of the input. And what's interesting uh about embeddings is that the embedding space is the same regardless of the inputs. So you can transform embeddings from voice, you can transform embeddings from images, you can transform embeddings for any type of input and it will go directly in the same uh hyper dimensional space. So that's uh how multimodal models are built. So for example, when you're uh thinking about models that are able to process the word cat and are able to process the image of a cat, uh They will have foundation models that will create embeddings for the image and for the world and the image and the world will be really close in the latent space, which is this hyper dimensional uh concept. So I think that when thinking about voice, we should think about voice, yeah, about H dynamics but also as all the other components, articulation uh and the intonation, use of voice in general and try to extract these characteristics. And then if we will have other inputs, multimodal inputs, it will also be possible to get embeddings from these inputs and include them in a multimodal model. So what we want, we, I think that what we would want to do is to be able to work with different inputs. And this is the best way to work with different inputs. OK. But um we still need to quantify a qualitative feature. Uh No, actually, that's the advantage of foundation models. Uh People are trying to uh use this type of approach because it's human naive. What you do is to create a representation of the input and use this representation of the input in order to uh apply a classifier to the representation. What's the issue right now with the current field is that we are selecting uh different type of inputs. Uh Pit frequency Jitter, shimmer, et cetera. But uh it has been shown at least in text and images that when you start selecting features, you're reducing the diagnostic accuracy because the human brain is able to select just a small set of features. While when you do it automatically with a deep learning algorithms, it's possible to avoid selecting, avoid bias and go directly to the result. But it it depends on the different approaches that you want to, to have.

Mieke:

But, but um what I wanted, um what I'm curious about is how would you deal with a self assessment like a voice hand appendix or a perception like the Rabai? That is what I mean with. We still have to quantify a certain quality or not.

Alberto:

Uh It's possible to uh add different inputs to the algorithm and it's pretty useful to do it. I would basically just quantify even subjective in indexes and add them to the algorithm when you have a foundation model, you can add all the data, which is categorical data, which is uh NPL. So language which is we can be image et cetera and you just pack all the data together. I think that the uh subjective perception could increase the diagnostic accuracy uh in some certain settings. But what people are seeing right now is that uh when using inadequate representation of inadequate input, you get really good results. So even just with voice itself, it would be possible to get really good results. But uh that's why I'm trying to propose a foundation models, a model in this term in order to be able to add any kind of input. But in these terms, uh the selection of input is also essential. Of course.

Mieke:

Yes. And that is something we can discuss in the next committee meeting after the C MC of Ramona. Yeah.

Alberto:

Yeah.

Mieke:

OK. But any anyone else because I am talking the whole time and Alberto is talking the whole time. But please, shoot.

Mette:

As far as I know there has been no prospective randomized control trials about this. Exactly, nothing, nothing. So there's nothing that we can use in our clinics. Exactly. Not yet. Not yet. We are elaborating on that matter. But the next question is of course, how is your, how, how, what is your, your perspective to make this?

Alberto:

Uh So I uh I would say that uh we will need to think in terms of frameworks, if we want to think about uh certain applications uh in the clinics, I mean, starting from the last component in order to get application in the clinics, you need approval from the EMA and FDA. Uh because right now diagnostic uh artificial intelligence algorithms are considered as a type two device, a class two device. And so it needs significant regulatory scrutiny. And starting from there, we can uh understand that we will need to have something really solid, which is probably focused on something very specific because that's something that DEA is focused on. And potentially it's using the core of a foundation model in order to get a small setting, uh a small diagnosis or something diagnostic in a small setting. And in order to get that, uh fortunately, it's not needed to have randomized clinical trials. Uh the EMA and the FDA are quite open for that. They are usually open to, to preliminary approval or 5 10-K approval for the FDA and and it's a bit easier than with drugs, for example, but there is no application yet in, in this. So I

mean that it's a long path to get approval for a clinical device because it's a clinical device when speaking about artificial intelligence. And I think that the only way to do it is to identify a specific clinical issue, uh a specific clinical objective and ensure that we are able to deliver an algorithm with sufficient uh diagnostic capability.

Mieke:

OK. Other, other questions? OK.

Neveen:

Can I ask a question? Uh When I was reviewing the literature, I noticed that some studies uh uh determined that the, the, the voice analysis data were based on uh vowels, some other uh were based on a certain sentence they utter, you know that. So that's how we, we are going to, to, to add as an input uh can differ from uh different languages that differ from different countries. Does it influence the, the, the, the, the output? Uh does the AI interfere with the, would it influence, what's the, the AI interpretation of these data? Different languages, different power, a vowel or uh content?

Alberto:

Yeah, that's a really interesting uh topic and it's a pretty uh complex issue because people are trying to use for example, simple vowels in order to take away the complexity of different languages. But by using vowels, it's not as easy to uh perform an analysis of articulation, which is particularly interesting and useful in many neurological disorders. Uh So the issue here, I think that with a very large foundation model, it would be possible to even take into account the differences in different languages and still get an adequate classification even with sentences. But I think that we are not yet at a scale that would allow that. So we will need to decide whether to go with simple inputs like stacked vowels, for example, or which is more complex than, than a simple power and simpler than a sentence, which is language dependent and try to get some results from there. But I mean, we will need to discuss about that. Uh I think that language is a pretty uh important issue when dealing with that and sentences, uh each language as a different structure as a different structure, even in terms of articulation. And if we are not able of getting a huge model, it's difficult that the model can take into account these differences and take away the variability.

Mieke:

Yes, I, I agree with that. I think that we really should think about the difference of voice related biomarkers or speech related biomarkers. So articulation language, this is speech. And actually, if we are continuing a thorough voice related biomarker model, then we have to stick to vowels. But the question mark is this possible? So we have to discuss that in our next meeting.