persist

Real World Data to aid in the transition from cancer patient to cancer survivor – Experiences from PERSIST

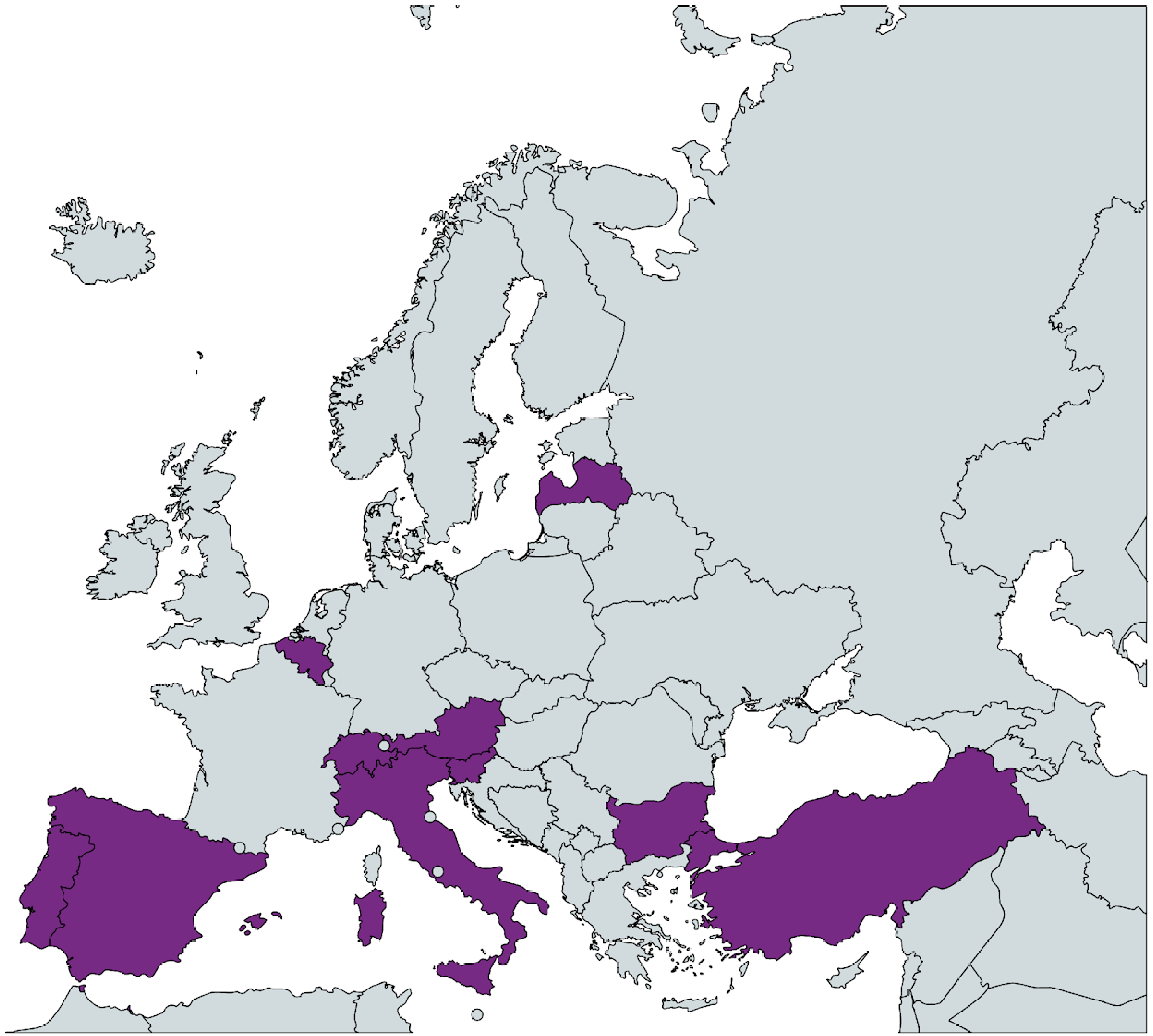
dr. Izidor Mlakar, University of Maribor

Project PERSIST

Grant No.: 875406
Timeframe: 01.01.2020 – 28.02.2023
Project coordinator: GRADIANT
Research & Innovation Action (TRL3-5)

- Fundacion Centro Tecnoloxico De Telecomunicacions de Galicia (GRAD)
- Servizo Galego de Saude (SERGAS)
- It Corporate Solutions Spain SL (DXC)
- EMODA Software (EMO)
- Univerza v Mariboru (UM)
- Univerzitetni klinicni center Maribor (UKCM)
- Haute Ecole Specialisee De Suisse Occidentale (HES-SO)
- Latvijas Universitate (UL)
- Cyberethics Lab SRLS (CEL)
- Centre Hospitalier Universitaire de Liege (CHU)
- Symptoma GMBH (SYMP)
- Rubynanomed, Unipessoal LDA (RUBY)
- National Patients Organisation (NPO)

Created with mapchart.net ©



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

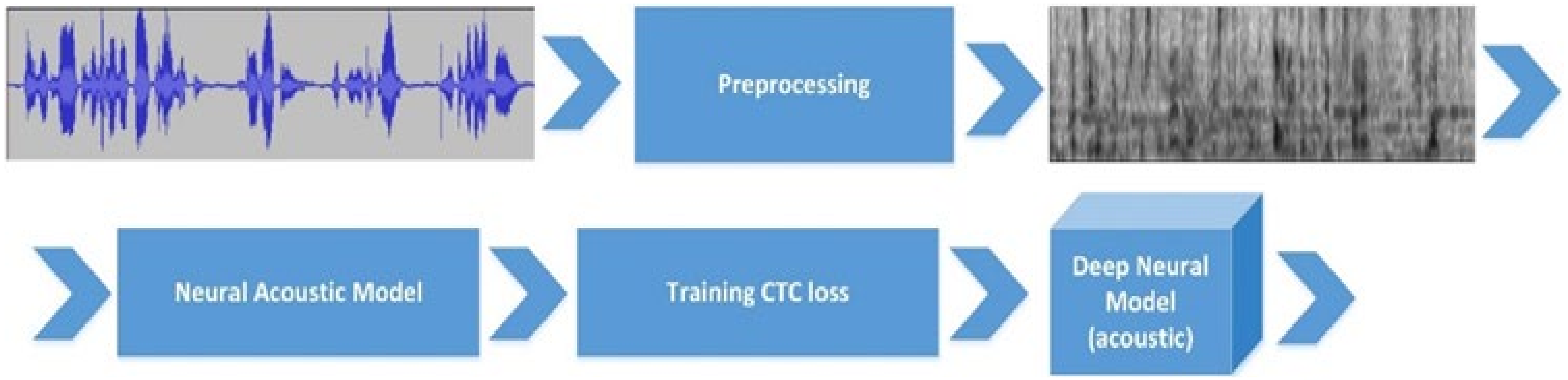
PERSIST Ecosystem

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Symmetric Model of Interaction on input

Automated Speech Recognition (ASR) - SPREAD

➔ ASR system SPREAD is built from the following boxes
- end-to-end Connectionist Temporal Classification-based deep neural model.



Training data	Testing data	Training time	Model size	Platform
		SLOVENIAN		
1869.14 hours	478 hours	152 days	2.6 GB	HPC GPU 4xA6000
Batch WER	**Test WER**			
1.5%	**0.7%**			
		LATVIAN		
782.65 hours	197.08 hours	93 days	2.6 GB	HPC GPU 2xRTX8000
Batch WER	**Test WER**			
2.03%	0.35%			
		ENGLISH		
1272.87 hours	319.97 hours	81 days	2.6 GB	HPC GPU 8xA100
Batch WER	**Test WER**			
0.7%	2.92%			
		SPANISH		
1406.84 hours	364.79 hours	35 days	2.6 GB	HPC GPU 8xA100
Batch WER	**Test WER**			
2.2%	5.5%			
		RUSIAN		
2796.00 hours	709.42 hours	145 days	2.6 GB	HPC GPU 6xA100
Batch WER	**Test WER**			
9.1%	2.7%			
		FRENCH		
1272.48 hours	335.49 hours	185 days	2.6 GB	HPC GPU 4xV100
Batch WER	**Test WER**			
5.3%	7.6%			

Language parameters for testing and training of the SPREAD Acoustic Model

- A spell checker to mitigate the issue of WER and precision in wild (**WER does not account for the variables that impact speech recognition**).

- A a language model to further mitigate the issue of WER

Slovenian language
Dictionary size: 215.851 unique words. (a list of fully correct words in a given language)

Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
15.1 GB	90461818	23202569	56min	3.51%	5.6GB	HPC 2xRTX8000

Latvian language
Dictionary size: 384.448 unique words (a list of fully correct words in a given language)

Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
654 MB	5478552	3210779	63min	4.75%	600MB	HPC 2xRTX8000

English language
Dictionary size: 826.491 unique words (a list of fully correct words in a given language)

Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
2 GB	64733542	72410719	71min	2.6%	1.5 GB	HPC 2xRTX8000

ES language
Dictionary size: 636.598 unique words (a list of fully correct words in a given language)

Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
1.253 GB	10752825	879921	71min	5.62%	1.369 GB	HPC 2xRTX8000

RU language
Dictionary size: 5.074.140 unique words (a list of fully correct words in a given language)

Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
1.435 GB	8566970	1807880	71min	2.20%	2.23 GB	HPC 2xRTX8000

FR language
Dictionary size: 742.308 unique words (a list of fully correct words in a given language)

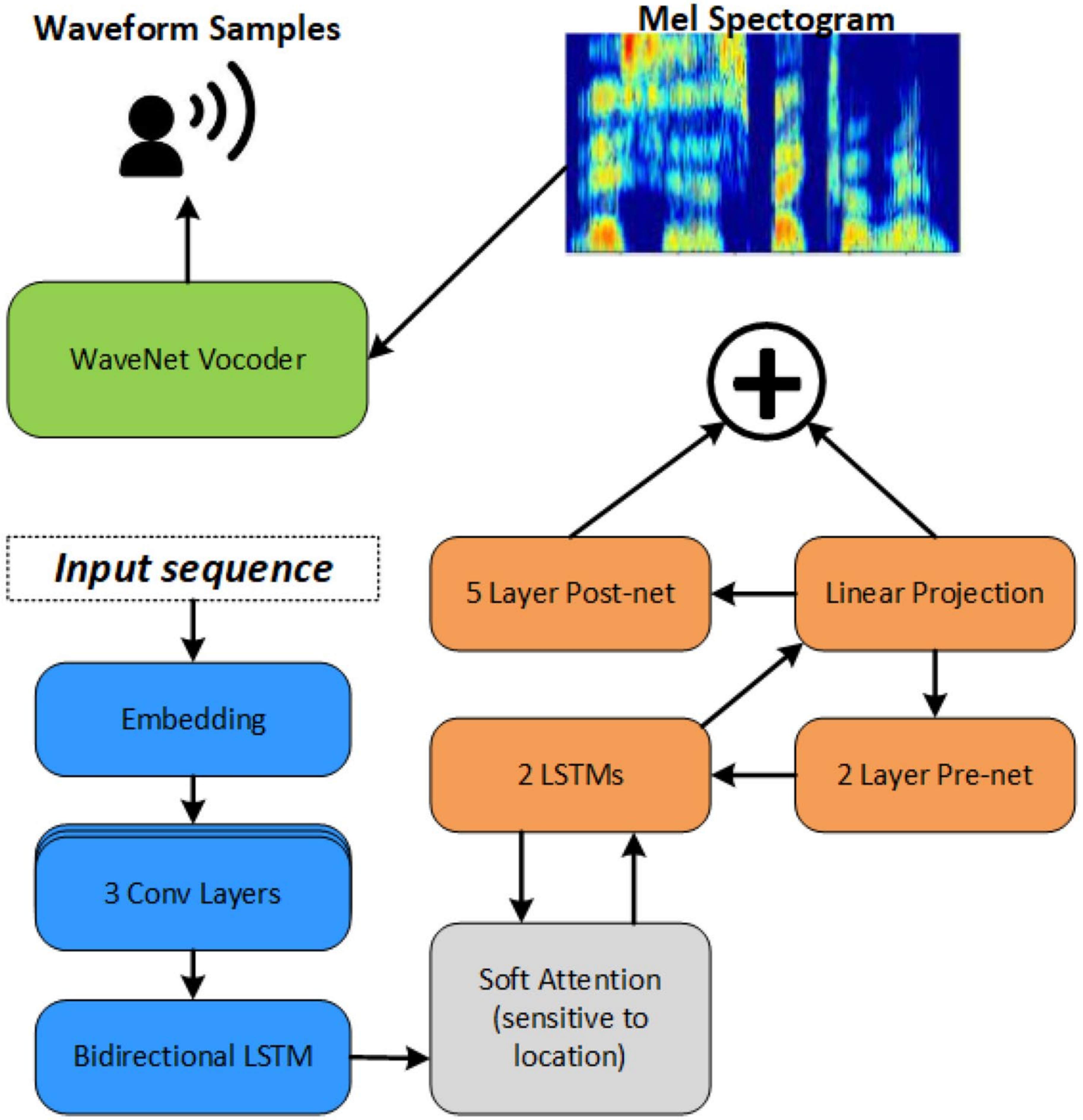
Dataset size	N sentences	N unique words	Training time	ErrRate	Model size	Platform
6.588 GB	82382142	16308542	71min	4.32%	948 MB	HPC 2xRTX8000

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Symmetric Model of Interaction on output

PLATOS Speech Synthesis (Text-to-Speech (TTS))

➔ TTS models used the Tacotron 2 architecture to generate mel spectrograms.



Encoder	Decoder	Attention	Mel-Post Network
Kernel size = 5	RNN dim = 1024	RNN dim = 1024	Embedding dim = 512
N convolutions = 3	Pre-net dim=256	Dim = 128	Kernel size = 5
Embedding dim= 512	Attention dropout = 0.1	N filters = 32	N convolutions = 5
	Dropout = 0.1	Kernel size = 31	
	Steps = 1000		
	Gate threshold = 0.5		

Tacatron 2 configuration

System	Min	Max	Mean	Median	Q1	Q3
Original recordings	60.0	100.0	93.67	100.0	91.0	100.0
TTS-3	6.0	100.0	51.35	53.0	35.0	68.0
TTS-4	0.0	65.0	26.44	25.0	15.0	33.0

Blizzard 2013 subset dataset: The MUSHRA Evaluation test for TTS-3 (Tacotron+Waveglow) and TTS-4 (Tacotron+Griffin lim)

System	Min	Max	Mean	Median	Q1	Q3
recordings	50.0	100.0	92.08	95.0	89.75	100.0
TTS-2	5.0	100	64.07	70.5	43.0	86.75
TTS-3	25.0	100.0	74.41	79.0	63.5	90.0
TTS-4	0.0	100.0	28.80	25.5	12.25	45.0

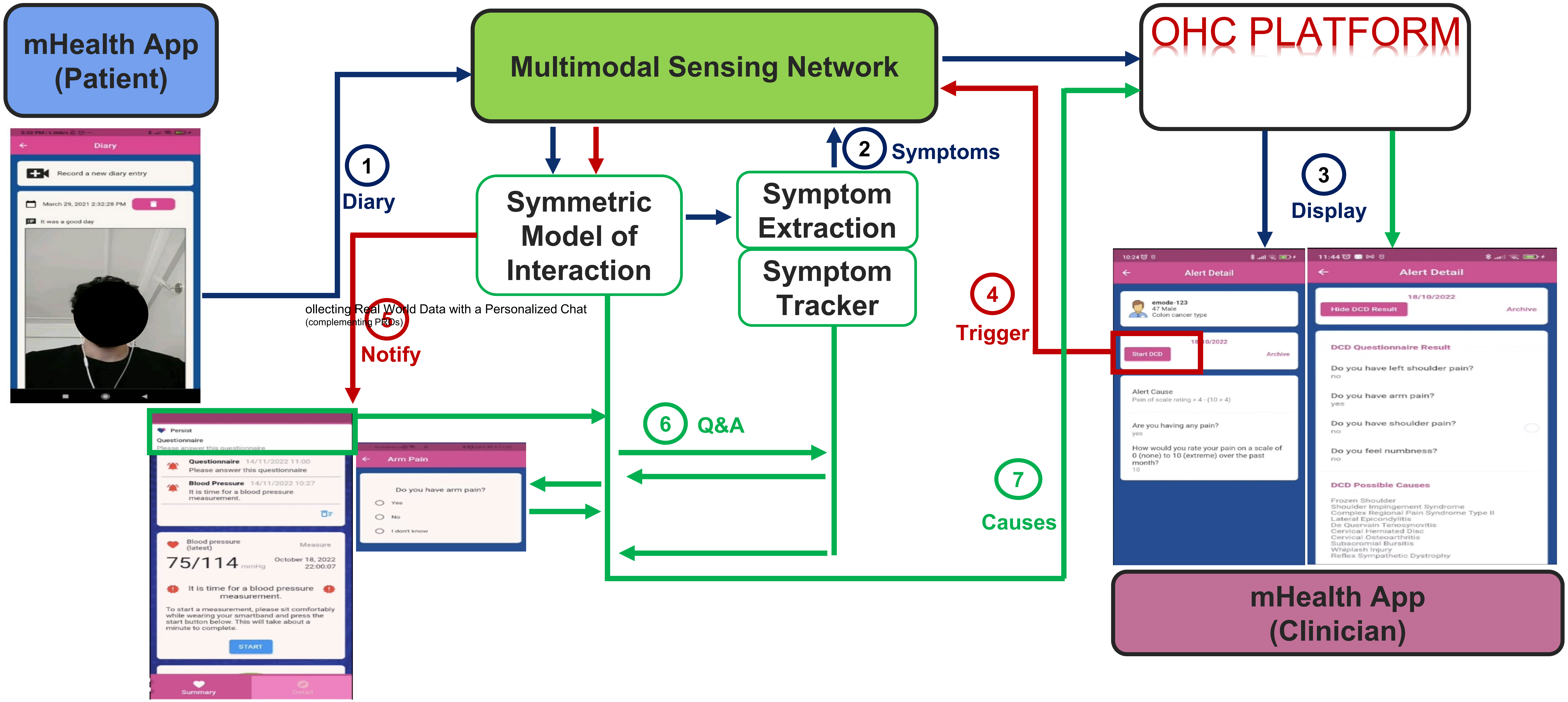
LJSpeech: MUSHRA Evaluation test for TTS-2 (WaveRNN), TTS-3 (Tacotron+Waveglow) and TTS-4 (Tacotron+Griffin lim)

For all PERSIST languages, we developed 2 TTS systems: TTS-3 (Tacotron+Waveglow) and TTS-4 (Tacotron+Griffin lim). For the English language we also developed a third system: TTS-2 (Tactron + WaveRNN).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Collecting Real World Data with a Personalized Chat

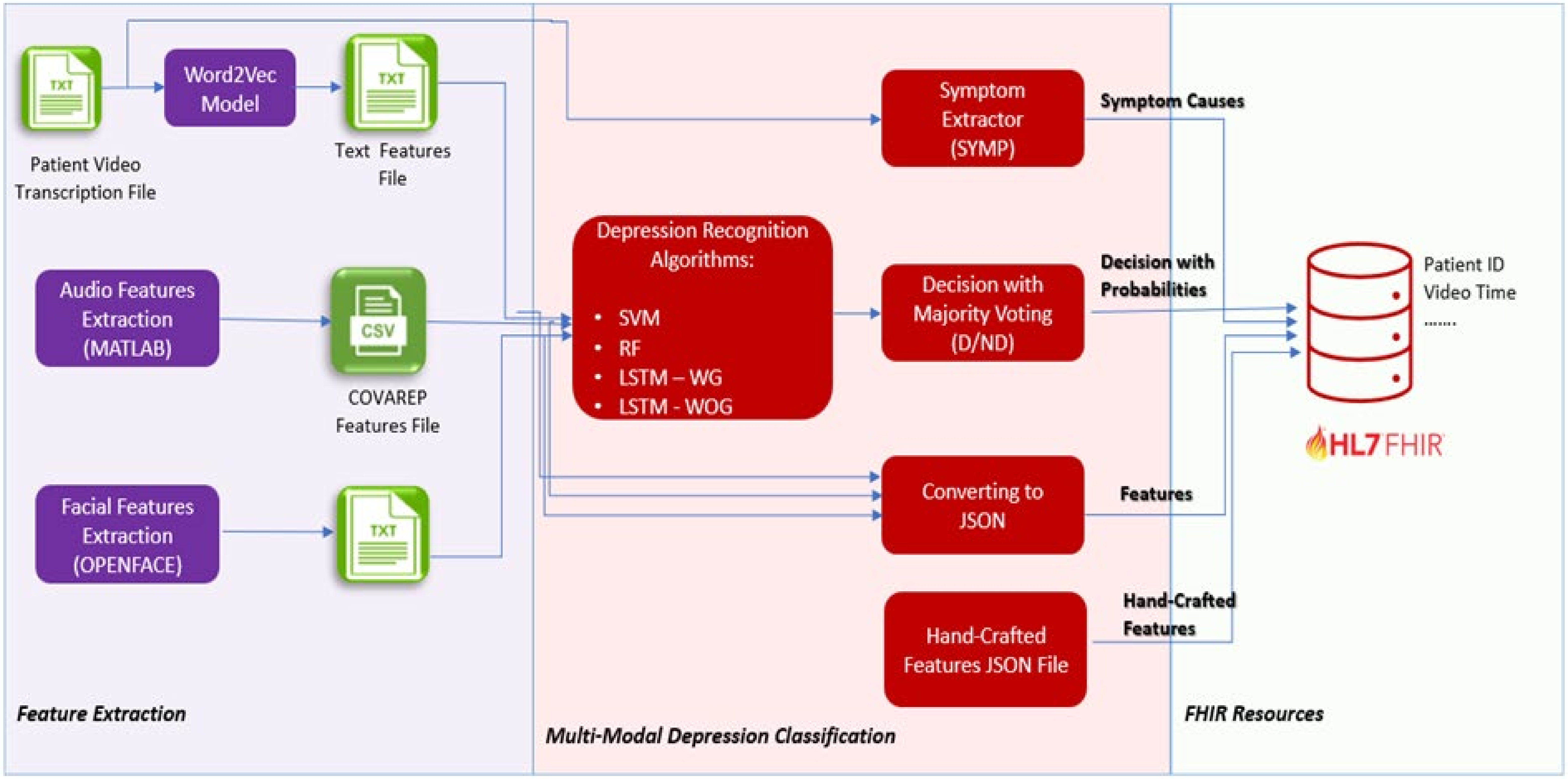
(complementing PROs)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Symptom Extraction from Diaries

- ➔ In the flow, we extract audio from the patient video. That audio is then sent to the automatic speech recognition engine to get the transcription text.
- ➔ The results are stored on the FHIR server.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Evaluating the feature extraction pipeline

Analysing language complexity using the Linguistic Feature Extraction Pipeline

Observable cues	SymptomMedia recordings			DAIC-WOZ recordings		
	Depressive disorder	*Without*	*r*	*Depressive disorder*	*Without*	*r*
	M (SD)	M (SD)		M (SD)	M (SD)	
A6.1.1: Sentence length (avg = 24.9)*	8.08 (2.66)	7.70 (1.80)	0.09	9.16 (2.84)	9.92 (3.25)	-0.13
A6.1.2: Sentence complexity (> 1.5 for high)*	1.32 (0.18)	1.28 (0.11)	0.14	1.22 (0.09)	1.21 (0.09)	0.08
A6.2.1: Lexical diversity (> 0.7 for high)*	0.13 (0.01)	0.14 (0.01)	-0.59	0.14 (0.00)	0.15 (0.01)	-0.41
A6.2.2: Lexical sophistication (> 16.25 for high)*	2.77 (0.73)	2.47 (0.53)	0.24	3.29 (1.02)	3.51 (1.18)	-0.11
A6.2.3: Lexical density (> 65 for high)*	36.72 (4.79)	36.06 (3.02)	0.09	35.99 (2.58)	35.49 (2.24)	0.11

Notes. M = mean, SD = standard deviation, *r* = effect size of the difference between recordings of depressive disorder and those without within each dataset. * = The values indicate expected result on general population (for more information see Appendix A.1)

Prosodic and voice quality cues related to Engagement and Quality of verbal communication using the Speech Feature Extraction Pipeline

Observable cues	SymptomMedia recordings			DAIC-WOZ recordings		
	Depressive disorder	*Without*	*r*	*Depressive disorder*	*Without*	*r*
	M (SD)	M (SD)		M (SD)	M (SD)	
B2: Engagement in verbal communication						
Pitch (Hz)	150.51757 (35.16333)	175.06171 (45.55554)	-0.31	161.71342 (38.62093)	149.71620 (37.78544)	0.16
Local jitter (%)	0.02236 (0.00356)	0.02071 (0.00216)	0.30	0.01450 (0.00350)	0.01660 (0.00540)	-0.24
Intensity (Db)	40.55071 (4.29640)	36.28843 (7.91860)	0.37	49.46458 (4.97600)	49.22040 (4.37559)	0.03
B5: Decreased voice quality						
Local absolute jitter (sec)	0.00017 (0.00006)	0.00013 (0.00004)	0.35	0.00010 (0.00005)	0.00013 (0.00006)	-0.24
Local shimmer (%)	0.12571 (0.01542)	0.13114 (0.01123)	-0.21	0.07208 (0.01574)	0.07927 (0.01738)	-0.22
Local dB shimmer (dB)	1.19400 (0.12285)	1.23007 (0.09239)	-0.18	0.67625 (0.16521)	0.74633 (0.17961)	-0.21
Harmonics-to-noise ratio (HNR)	10.71571 (2.15857)	10.35021 (1.22188)	0.12	13.83333 (2.28963)	12.80000 (2.27408)	0.23

Notes. M = mean, SD = standard deviation, *r* = effect size of the difference between recordings of depressive disorder and those without within each dataset.

Visual cues related to facial emotional expressivity using the Visual Extraction Pipeline

Observable cues	SymptomMedia recordings			DAIC-WOZ recordings		
	Depressive disorder	*Without*	*r*	*Depressive disorder*	*Without*	*r*
	M (SD)	M (SD)		M (SD)	M (SD)	
C1: Occurrence of facial expressions (Frame (%))						
Emotion - Surprise	1.23 (1.36)	1.77 (1.31)	-0.20	1.31 (1.85)	0.55 (0.62)	0.36
Emotion - Anger	0.053 (0.08)	0.049 (0.07)	0.03	1.52 (3.96)	0.82 (1.15)	0.16
Emotion - Fear	2.18 (1.62)	2.84 (2.10)	-0.19	12.33 (9.55)	10.71 (7.06)	0.11
Emotion - Happiness	5.80 (9.33)	10.67 (17.40)	-0.20	11.66 (9.62)	9.34 (6.39)	0.17
Emotion - Sadness	8.12 (6.38)	10.82 (5.76)	-0.23	20.62 (6.91)	18.11 (9.41)	0.16
Emotion - Disgust	25.00 (17.51)	23.46 (24.84)	0.04	20.49 (14.60)	15.71 (8.68)	0.24
C2: Intensity of facial expressions (Intensity (0 – 5 point))						
Emotion - Surprise	1.79 (0.64)	1.81 (0.22)	-0.03	1.81 (0.35)	1.79 (0.22)	0.04
Emotion - Anger	0.67 (0.81)	0.87 (0.69)	-0.14	0.86 (0.36)	0.78 (0.26)	0.15
Emotion - Fear	1.42 (0.24)	1.55 (0.20)	-0.31	1.50 (0.31)	1.36 (0.12)	0.37
Emotion - Happiness	1.36 (0.44)	1.37 (0.46)	-0.01	1.45 (0.19)	1.37 (0.16)	0.23
Emotion - Sadness	1.61 (0.18)	1.66 (0.18)	-0.14	1.77 (0.35)	1.58 (0.20)	0.39
Emotion - Disgust	0.88 (0.13)	0.97 (0.18)	-0.30	0.83 (0.12)	0.78 (0.13)	0.23
C3: Occurrence and emotional variability						
Positive emotions (Frame (%))	5.80 (9.33)	10.67 (17.40)	-0.20	11.66 (9.62)	9.34 (6.39)	0.17
Negative emotions (Frame (%))	36.57 (18.77)	38.93 (30.05)	-0.05	56.26 (27.00)	45.90 (20.93)	0.24
Positive emotions (Intensity (0 - 5 point))	1.36 (0.44)	1.37 (0.46)	-0.01	1.45 (0.19)	1.37 (0.16)	0.24
Negative emotions (Intensity (0 - 5 point))	1.14 (0.29)	1.26 (0.23)	-0.24	1.36 (0.16)	1.26 (0.11)	0.39
Total number of emotion variability	4.43 (0.94)	4.93 (1.00)	-0.26	4.33 (0.49)	4.27 (0.46)	0.08

Notes. M = mean, SD = standard deviation, *r* = effect size of the difference between recordings of depressive disorder and those without within each dataset.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Multi-Modal Feature Extraction and Depression Classification

Evaluation of AI Algorithms

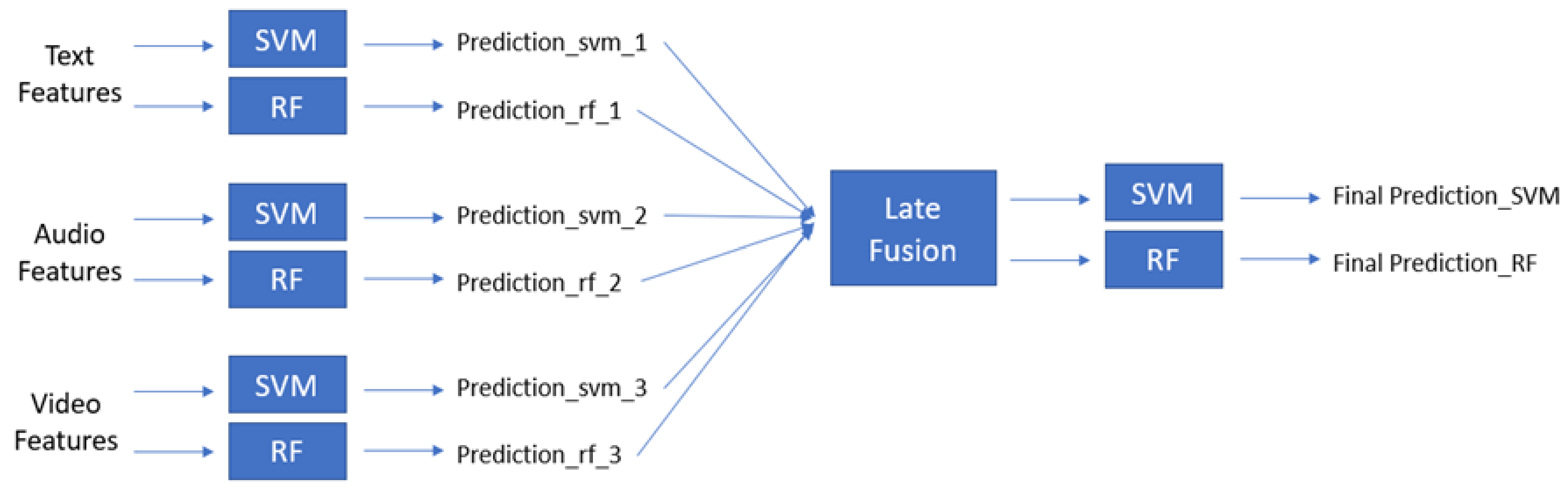
Depression classification was tested with two different multi-modal datasets:

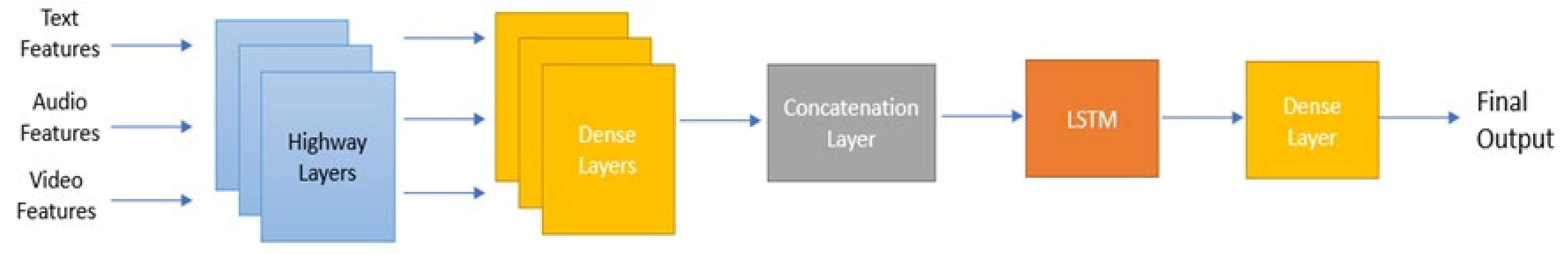
- ➔ (Distress Analysis Interview Corpus Wizard-of-Oz (DAIC-WOZ) dataset as a reference dataset for multi-modal depression classification;
- ➔ SymptomMedia dataset as a test database)

and different algorithms:

- ➔ SVM late fusion,
- ➔ RF late fusion,
- ➔ LSTM (with gating),
- ➔ LSTM (without gating)

	DAIC-WOZ	SymptomMedia
Testing Set	47	28
Development Set	35	-
Training Set	107	-





	DAIC-WOZ DB Dataset						SymptomMedia Dataset					
Modality	F1-Score	Recall	Precision	MSE	RMSE	MAE	F1-Score	Recall	Precision	MSE	RMSE	MAE
T+A	0.38	0.39	0.55	0.61	0.78	0.61	0.46	0.46	0.46	0.53	0.73	0.53
T+V	0.42	0.41	0.55	0.59	0.76	0.59	0.42	0.43	0.42	0.57	0.75	0.57
A+V	0.56	0.55	0.6	0.45	0.67	0.45	0.37	0.39	0.38	0.6	0.77	0.6
T+A+V	0.46	0.45	0.61	0.54	0.73	0.54	0.4	0.43	0.41	0.57	0.75	0.57

Performance and error metrics for SVM+RF with SVM Late Fusion

	DAIC-WOZ DB Dataset						SymptomMedia Dataset					
Modality	F1-Score	Recall	Precision	MSE	RMSE	MAE	F1-Score	Recall	Precision	MSE	RMSE	MAE
T+A	0.38	0.39	0.55	0.61	0.78	0.61	0.46	0.46	0.46	0.53	0.73	0.53
T+V	0.42	0.41	0.55	0.59	0.76	0.59	0.42	0.43	0.42	0.57	0.75	0.57
A+V	0.46	0.45	0.61	0.54	0.73	0.54	0.3	0.36	0.29	0.64	0.8	0.64
T+A+V	0.44	0.43	0.59	0.56	0.75	0.56	0.4	0.43	0.41	0.57	0.75	0.57

Performance and error metrics for SVM+RF with RF Late Fusion

	DAIC-WOZ DB Dataset						SymptomMedia Dataset					
Modality	F1-Score	Recall	Precision	MSE	RMSE	MAE	F1-Score	Recall	Precision	MSE	RMSE	MAE
T+A	0.53	0.57	0.51	0.27	0.52	0.48	0.56	0.57	0.58	0.37	0.61	0.47
T+V	0.52	0.55	0.5	0.29	0.54	0.49	0.45	0.46	0.46	0.36	0.6	0.48
T+A+V	0.53	0.55	0.52	0.35	0.59	0.51	0.56	0.57	0.58	0.37	0.6	0.47

Performance and error metrics for LSTM without Gating

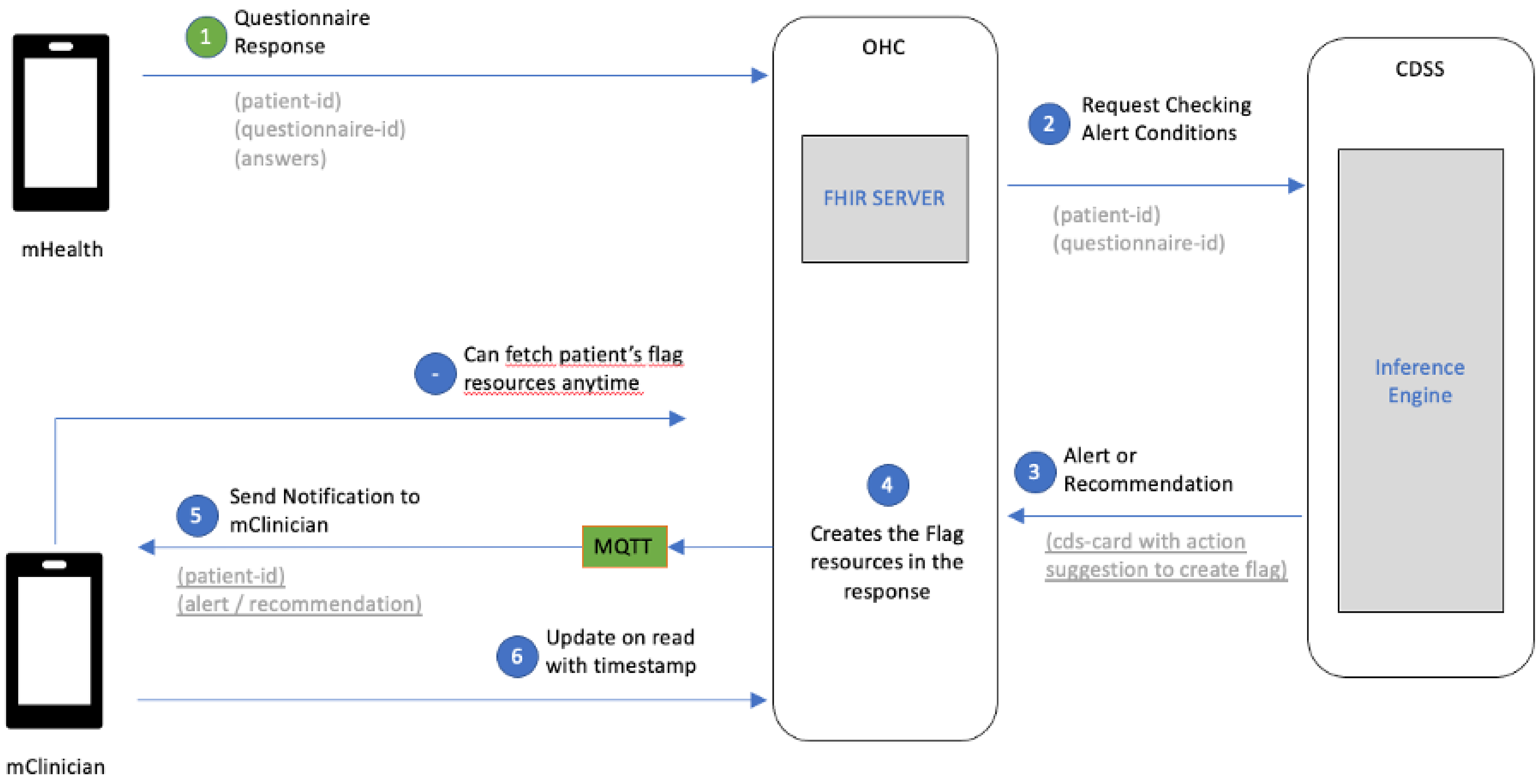
	DAIC-WOZ DB Dataset						SymptomMedia Dataset					
Modality	F1-Score	Recall	Precision	MSE	RMSE	MAE	F1-Score	Recall	Precision	MSE	RMSE	MAE
T+A	0.45	0.43	0.57	0.33	0.58	0.48	0.48	0.5	0.5	0.35	0.59	0.44
T+V	0.42	0.41	0.55	0.35	0.59	0.49	0.5	0.5	0.5	0.38	0.61	0.47
T+A+V	0.64	0.66	0.64	0.32	0.57	0.45	0.48	0.5	0.5	0.34	0.58	0.44

Performance and error metrics for LSTM with Gating

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

CDSS Inference Engine: Use of RealWorld Data in Practice

Alert Mechanism flow in PERSIST



- Questionnaire responses arrives at FHIR server
- Inside Workbench requests inference automatically via CDS Hooks
- Response of inference engine in CDS Hooks response structure
- Auto-generates Flag resource upon inference response
- MQTT notification is sent to clinician app

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

PERSIST Multicenter Clinical Trial

mHealthApp will collect **objective markers** (vital signs) and **subjective markers** (PREMs/PROMs and symptoms of depression).

The clinical decision support system will enable oncologist to personalize treatment and care plans/follow-up for efficient management of patients.

Hypothesis: Performing a comparison at the beginning and at the end of the intervention, **participants will significantly increase their self-efficacy following the personalized intervention supported by the mHealthApp**.

Subjects: 80 Breast Cancer Survivors, 80 Colorectal Cancer Survivors, Two subgroups (**chemotherapy** and **non-chemotherapy**). **At least 33%** of patients that have had **chemotherapy**.

Design: A single-case experimental prospective, cross-over deisgn,, 6 months follow up

Public Study Protocol ISRCTN97617326

Latvia	**UL (University of Latvia)**	Riga East Clinical University Hospital (Latvian Oncology Center)
Belgium	**CHU (Centre Hospitalier Universitaire De Liège)**	Centre Hospitalier Universitaire De Liege
Slovenia	**UMC (University Medical Centre Maribor)**	University Medical Centre Maribor
Spain	**SERGAS (Servizo galego de saude)**	Complejo Hospitalario Universitario de Ourense

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

PERSIST Multicenter Clinical Trial, Recruitment & Droput

CLINICAL PARTNER	RECRUITED PATIENTS	MEAN AGE	BREAST CANCER	COLORECTAL CANCER	MALE	FEMALE
UL	46	54	24	22	7	39
UKCM	40	57	20	20	11	29
CHU	41	55	21	20	7	34
SERGAS	39	56	20	19	12	27
TOTAL	**166**	**55**	**85**	**81**	**37**	**129**

Droput: 26 patients (16%, mostly at the begging), for most of them a substitute was recruited

Some Reasons:

- Due to unforeseen circumstances, i.e. move to the countryside where wifi is not available.
- Anxiety and re-living the diagnosis and treatment phase
- Technical issues and complexity of use (i.e. never used similar tech. before)
- Burdened by diary recording and activates, and technology disturbing their lifestyle (e.g. sleep) and their rythm
- Recurrence of cancer and (new) medical issues unrelated to the trial

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Activation levels of patients (PAM questionnaire)

	Score at recruitment	Score at last follow-up
N	78	78
Mean	65,10	65,71
Median	63,10	63,10
Std. Deviation	14,605	16,063
Minimum	38	37
Maximum	100	100
Percentiles 25	53,20	52,65
50	63,10	63,10
70	75,00	77,70

Level	Recruitment (N=75)	Last follow-up (N=75)	P value
Level 1 n (%)	5 (6,4)	6 (7,7)	1,000
Level 2 n (%)	15 (19,2)	16 (20,5)	1,000
Level 3 n (%)	33 (42,3)	28 (35,9)	0,486
Level 4 n (%)	25 (32,1)	28 (35,9)	0,648

The PAM Score is an interval-level scale from 0-100 that correlates with one of the four levels of patient activation. PAM levels 1 and 2 indicate lower patient activation, while PAM levels 3 and 4 indicate higher patient activation

Conclusion:

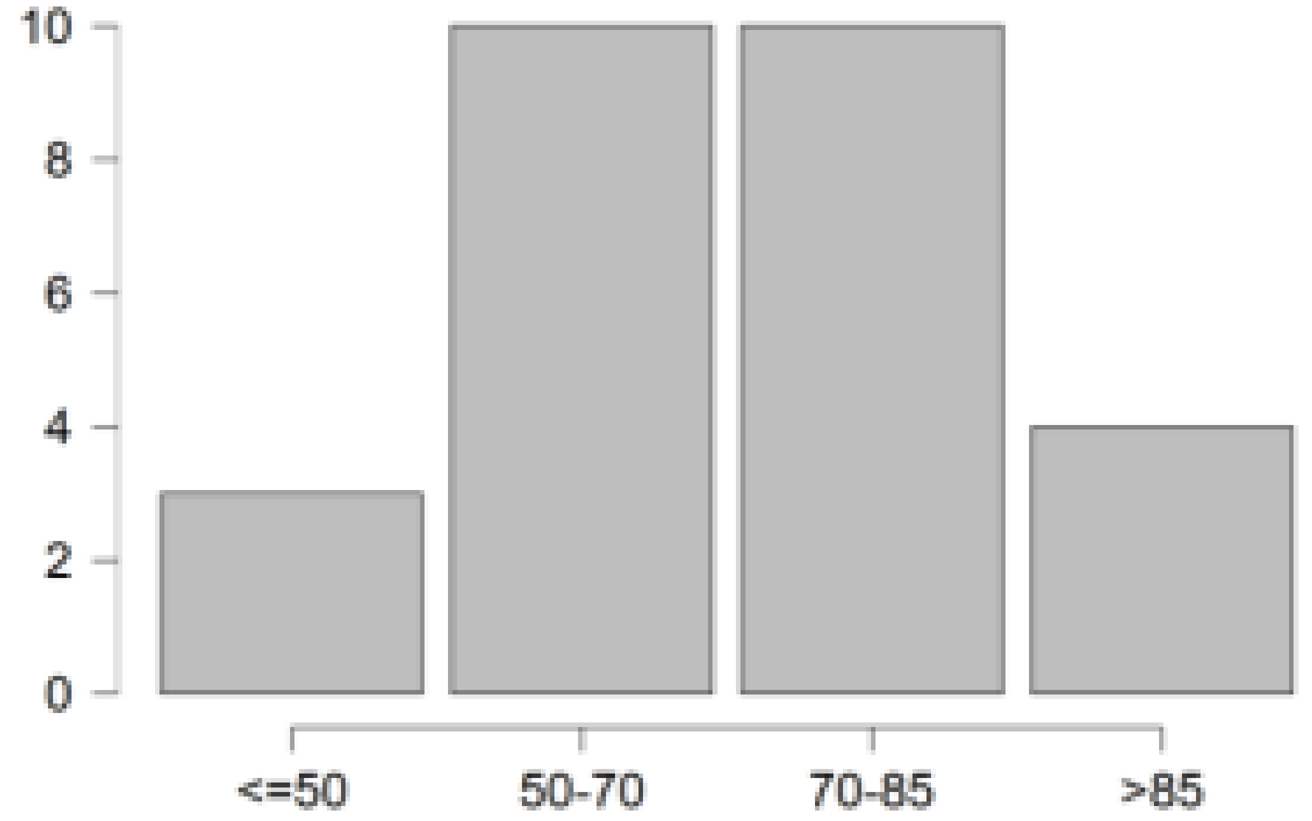
- most patients level 3 or 4 of activation at both recruitment and last follow-up.
- taking action and gaining control over their condition
- no statistically significant differences were found in the activation levels
- However, results suggest that the mHealth app under development may be able to support patients in improving their selfmanagement skills

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

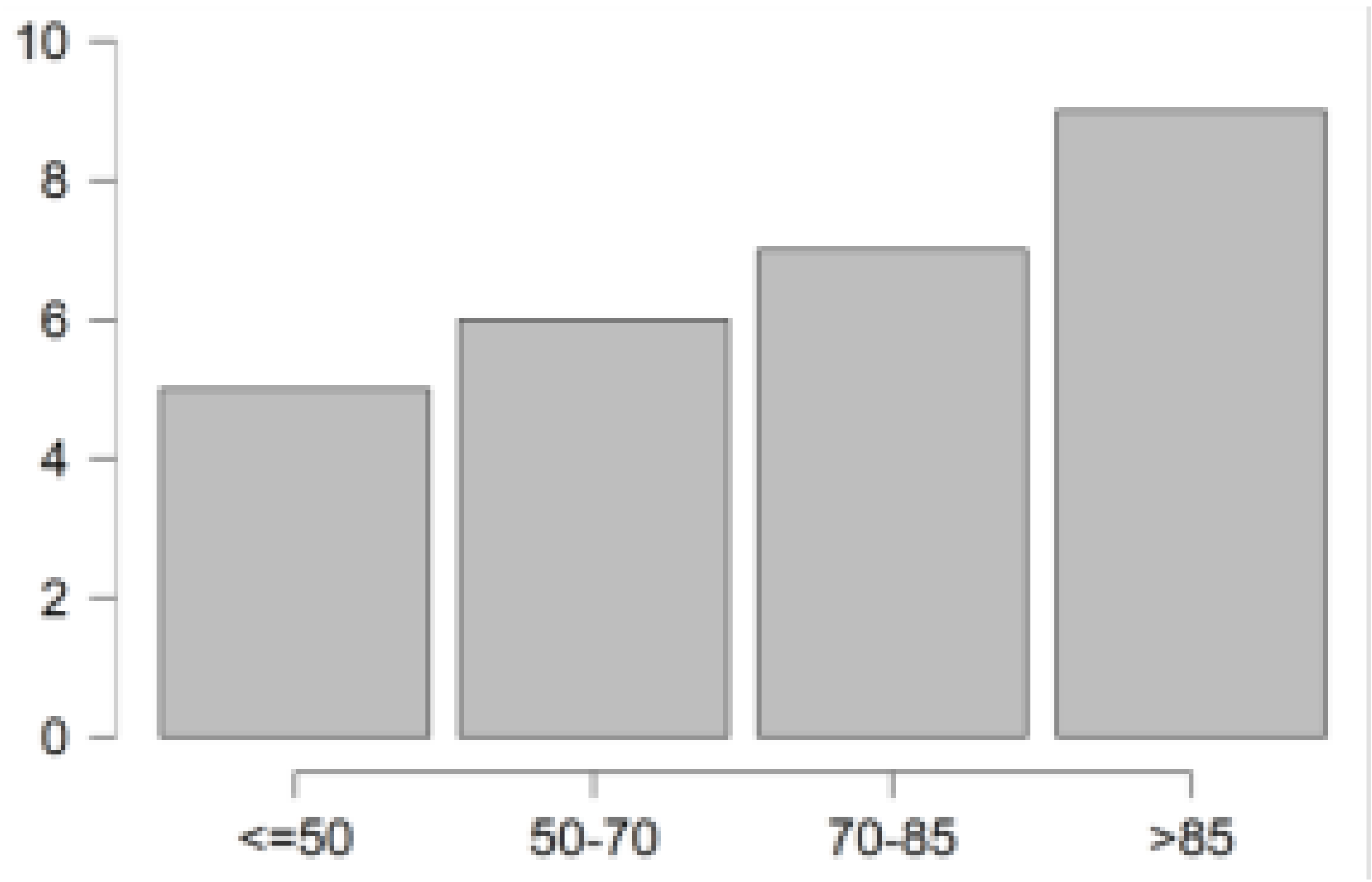
User acceptance of mHealthApp (SUS) from patients

Level	Definition
<=50	Not easy to use
50-70	Experiencing usability issues
70-85	Acceptable to good
>85	Excellent usability

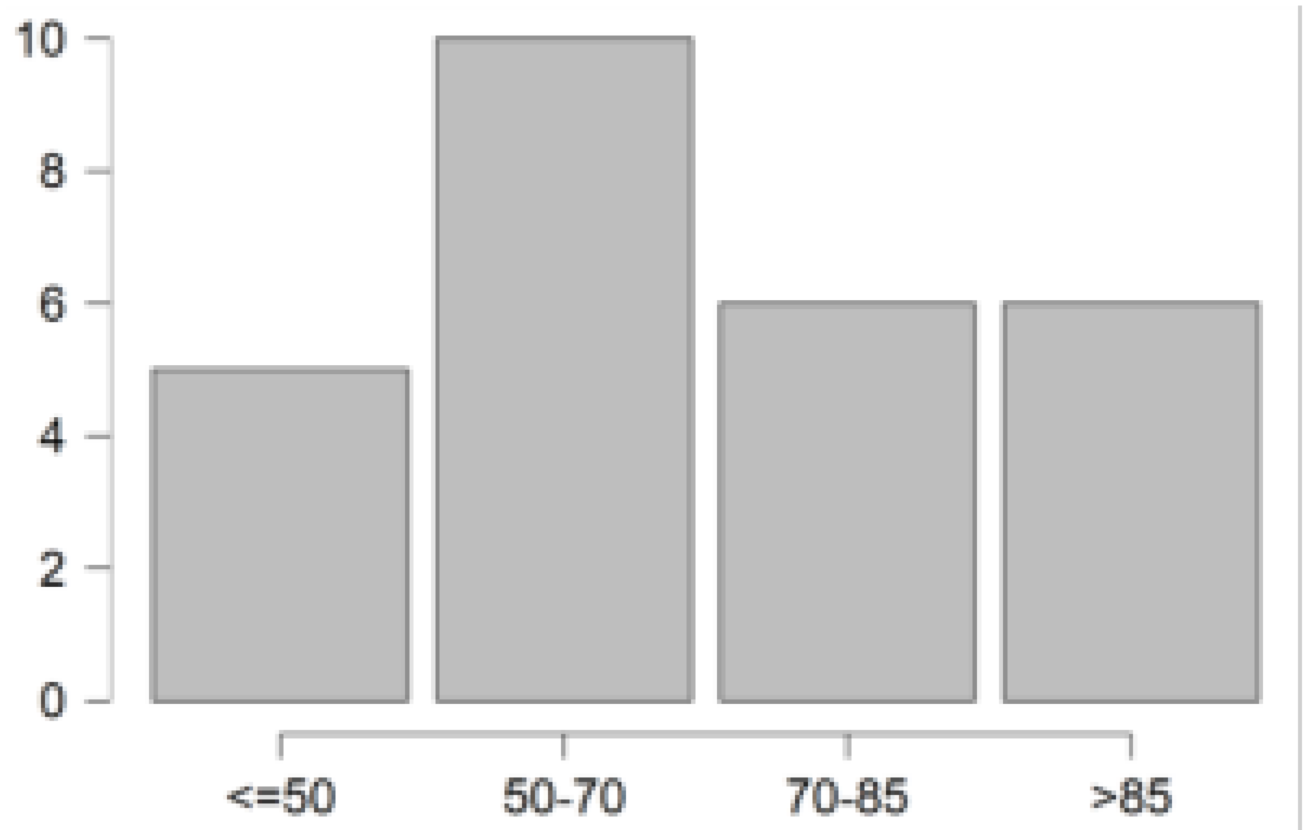
Table 13 The definition of system usability level



The sum score of the points acquired in all 10 questions in the beginning



The score group in the middle of the study

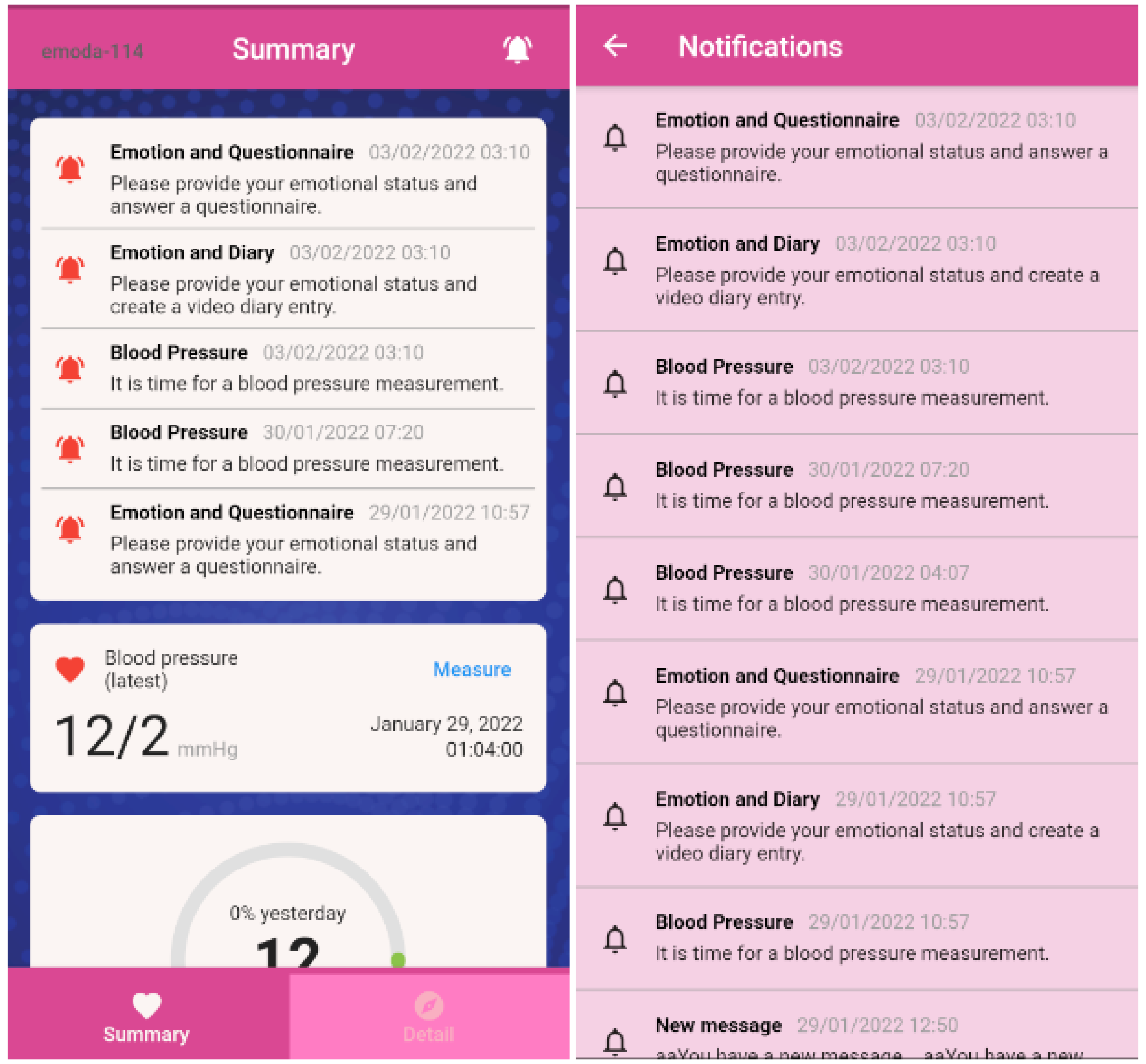


The score group at the end of the study

Despite some negative feedback about the virtual agent introduction, 44,44% of patients still evaluated the usability of the app as good or excellent.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

PERSIST Leasons Learnt



- Patients need feedback and must not be left alone
- Co-creation is much appreciated, and talking with patients discussing their issues related to trial keeps them engaged and makes the effort they put in worthwhile
- Refining features based on their remarks
- Providing feedback also on other participants in other clinical sites
- Considering patient specific context and allow for some minor deviations (i.e. during holidays, manual data inserts, etc.)
- Organize as many workshops which connect patients with technical partners, not only clinicians

1- Let Patients use their own devices in their own way (e.g. report data once a day; for example, smartphone has step counter, manually measure blood pressure in other devices) even if some issues related to data credibility occur

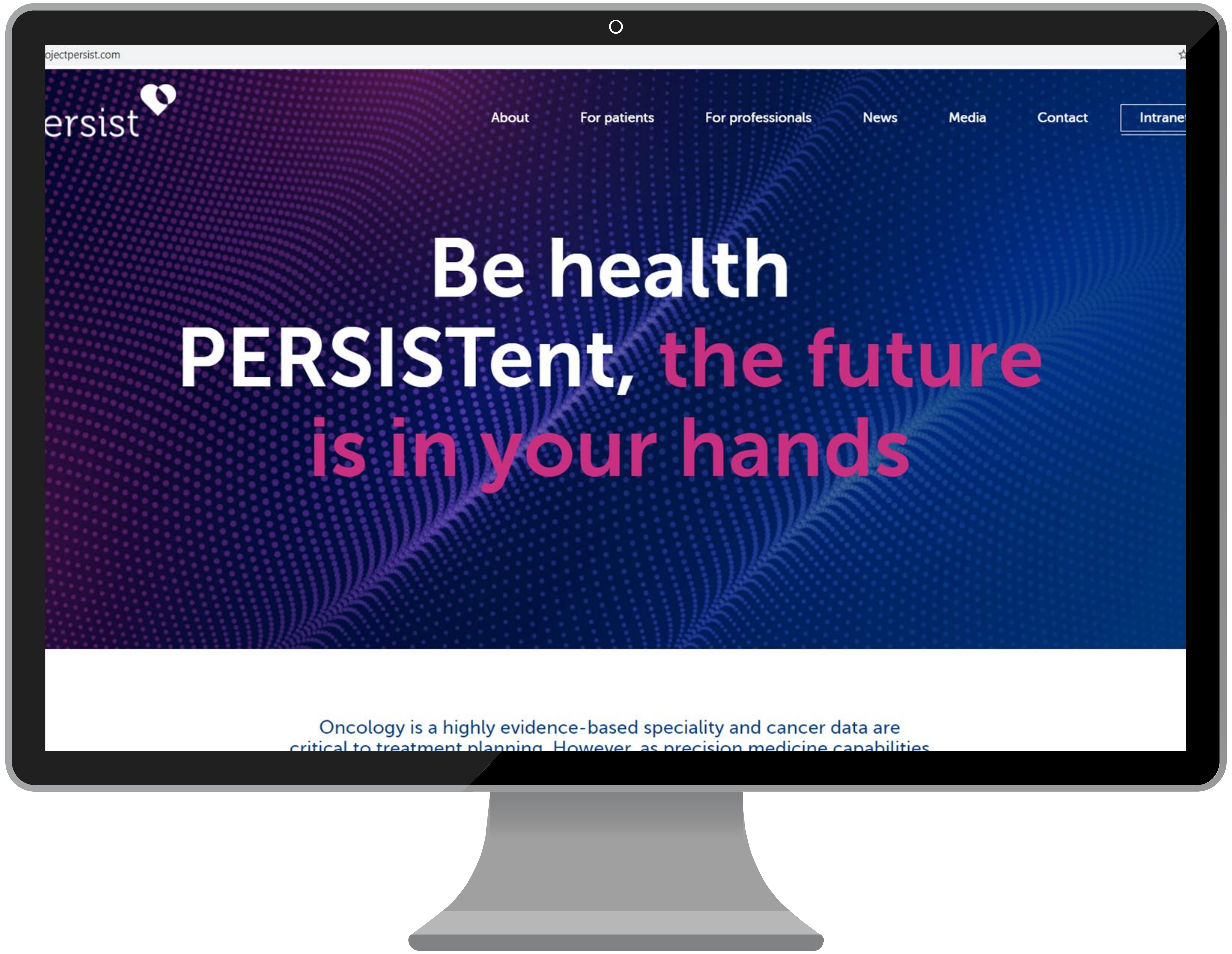
2 – Include patients in the study design, especially parts related to data collection. Most patients feel that what we ask them with standardized questionaries, does not really address issues they experience or their needs they have

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

PERSIST Progress and Updates

Channels and online presence

- ✓ Website: https://projectpersist.com/
- ✓ Facebook: https://www.facebook.com/PERSIST.H2020/
- ✓ Twitter: https://twitter.com/PERSIST_H2020
- ✓ LinkedIn: https://www.linkedin.com/company/persist-oncology/
- ✓ YouTube: https://www.youtube.com/channel/UCQgPynxNv1TlfNcg4vXNssA/



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

persist