Developing an AI-driven tool to analyse expressed emotion in parental speech: key results and stakeholder perspectives

Zoë Firth[1], Nicholas Cummins[1], Bahman Mirheidari[2], Tassadaq Hussain[1], Christine Aicardi[3], Stephen Oram[4], Alice Wickersham[1], Helen L. Fisher[1,5], Johnny Downs[1,6]

1. Institute of Psychiatry, Psychology & Neuroscience, King's College London; 2. School of Computer Science, University of Sheffield; 3. Department of Global Health & Social Medicine, King's College London; 4. Independent Applied Science Fiction Practitioner, London, UK; 5. ESRC Centre for Society and Mental Health, King's College London; 6. South London and Maudsley NHS Foundation Trust, London, UK

Introduction

'Expressed emotion' (EE): the type and intensity of emotions expressed by a caregiver about their child[1]

Index of 'the emotional climate of the **parent-child relationship**'[2]

Parental EE during childhood predicts children's later mental health outcomes[2], indicating it could be used to target early mental health intervention & prevention for children and families

Problem: resource-intensity of coding EE means that measuring EE in clinical settings has limited feasibility

Research question: Can we build an AI-driven tool that can analyse expressed emotion in parents' speech as effectively as humans?

Methods

Cohort data

E-Risk Longitudinal Twin Study
- *Twin births 1994-5, England & Wales*
- *1,116 families; 2,232 twins (same-sex)*
- *Covers full socioeconomic spectrum*
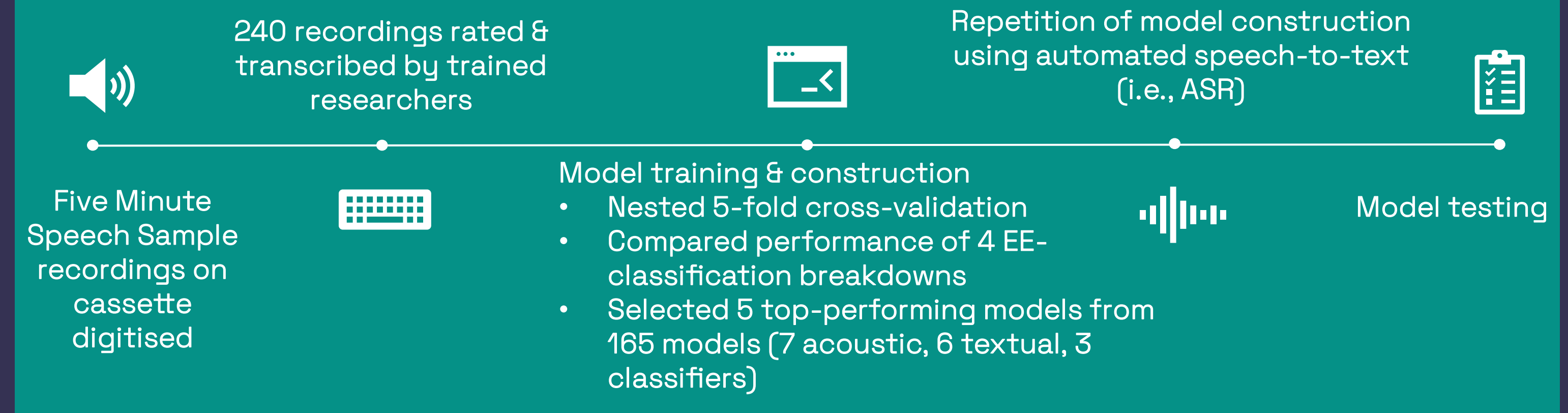- *Data collection at ages 5, 7, 10, 12, 18 years (with >90% retention)*

Speech data

240 Five Minute Speech Sample (FMSS) recordings
- *Unstructured interviews with mothers*
- *Recorded when twins were 10 years old*
- *'Tell me about [child], what are they like?'*
- *Analysed files 3.5-8.5 minutes long*

EE negativity rated by independent researchers from 0 (no negativity) to 5 (overtly hostile)[3]

Model construction

240 recordings rated & transcribed by trained researchers

Repetition of model construction using automated speech-to-text (i.e., ASR)

Five Minute Speech Sample recordings on cassette digitised

Model training & construction
- Nested 5-fold cross-validation
- Compared performance of 4 EE-classification breakdowns
- Selected 5 top-performing models from 165 models (7 acoustic, 6 textual, 3 classifiers)

Model testing

Results & Discussion

Demographics

N	Age	Gender	Zygosity
240	10	57.5% female	60% monozygotic

Socio-economic status (ACORN)

1 (high)	2	3	4	5 (low)
55	11	71	32	65

Acoustic features

Classifier	Features	Accuracy	Sensitivity (high)	Specificity (high)
RF	eGeMAPS	0.700	0.881	0.338
LR	Spec	0.692	0.875	0.325
LR	AVEC13	0.608	0.581	0.663
RF	IS9	0.717	0.956	0.238
LR	IS13	0.600	0.588	0.625

Textual features

Classifier	Features	Accuracy	Sensitivity (high)	Specificity (high)
LR	TFIDF	0.663	0.794	0.400
Linear SVM	TFIDF	0.658	0.806	0.363
Linear SVM	RoBERTa	0.575	0.644	0.438
LR	RoBERTa	0.575	0.656	0.413
Linear SVM	ALBERT	0.567	0.606	0.488

Fusion [acoustic & textual]

Classifier	Features	Accuracy	Sensitivity (high)	Specificity (high)
Linear SVM	eGeMAPS + TFIDF	0.654	0.675	0.613
LR	Spec + TFIDF	0.654	0.675	0.613
Linear SVM	Spec + DistilBERT	0.650	0.681	0.588
LR	Spec + Doc2Vec	0.646	0.663	0.613
Linear SVM	eGeMAPS + TFIDF	0.646	0.675	0.588

Score classification

Low-EE / High-EE

N: 40, 40, 80, 50, 28, 2
EE rating [gold standard]: 0, 1, 2, 3, 4, 5

	Low-EE [AI]	High-EE [AI]
Low-EE (0/1) [gold standard]	49	31
High-EE (2/3/4/5) [gold standard]	52	108

LR = Logistic Regression, RF = Random Forest, SVM = Support Vector Machines
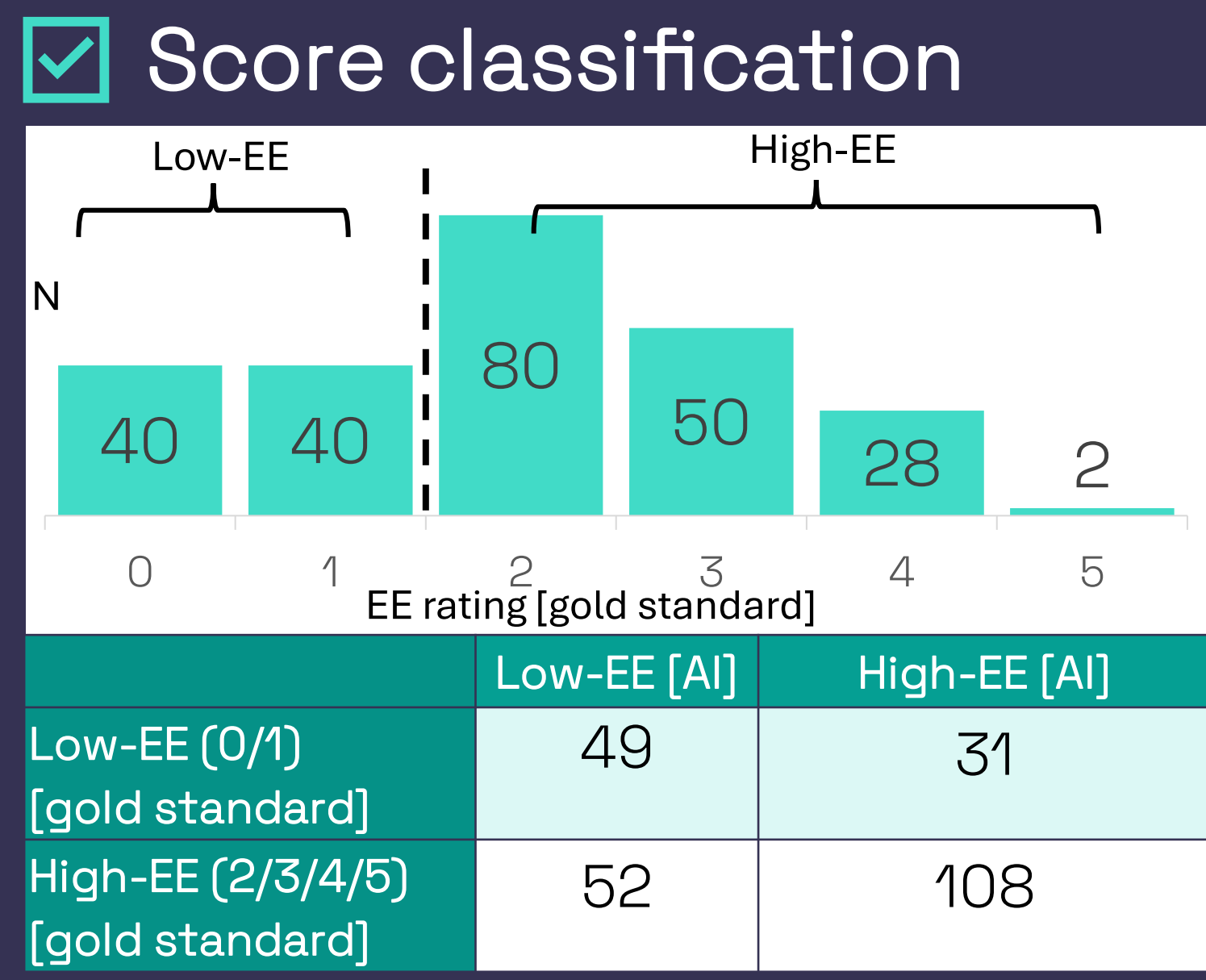
- Superior performance based on binary classification of EE scores (0/1 vs. 2/3/4/5)
- Acoustic features performed stronger than textual ones
- Textual did not contribute as much as expected to performance of fusion model
- Despite lower accuracy, fusion model has more stable metrics than acoustic

Automatic Speech Recognition

- Fine-tuning of open-source models, compared on WER (Word Error Rate)
- Whisper-large-v2 had best WER (19.52% before fine-tuning, 10.53% after fine-tuning)
  Accuracy: 0.654 **Sensitivity:** 0.675 **Specificity:** 0.613
- Automatic speech recognition may be viable pathway over manual transcription, despite low-quality data

Engaging stakeholders using science fiction

- Needed public perspectives on our research, but abstraction & complexity may affect how people engage with AI methods[4]
- We worked with Stephen Oram[5], a near-future science fiction author, to write 2 stories about possible futures of our project
- Stories developed after consultation with clinicians, young people, parents, and researchers on the project
- Stories presented by author at 9 further events (e.g., literature festivals, seminars)

EXTRACTING HUMANITY And Other Stories — Stephen Oram
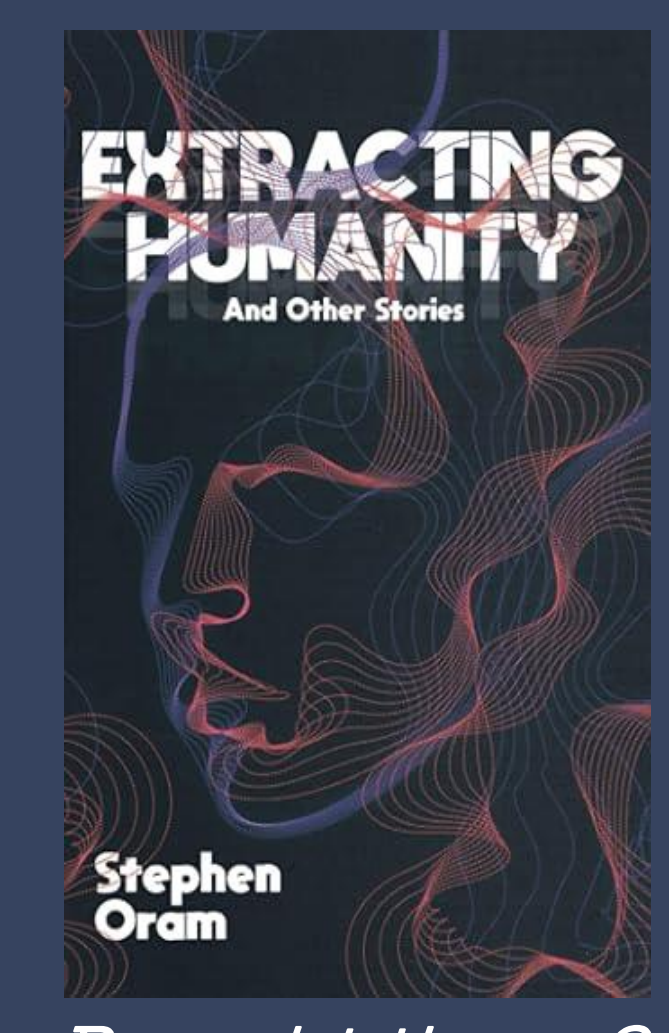
'A Mother's Nightmare'

'Standard Deviations'

Brought these 2 stories to separate workshops with young people & parents

Organisation	Attendees	#
Young Carers' Group (Carers' Hub Lambeth)	Young carers (aged 13-19)	8
Adolescent Mental Health Advisory Group (NIHR Maudsley BRC)	Young people with mental illness (aged 13-18)	4
Autistica	Parents of autistic children	6

Parents
- Concerns over ability of one FMSS to capture variability & complexity of parenting
- Need for continuous, dynamic models of consent
- Key context: timely access to assessment and treatment
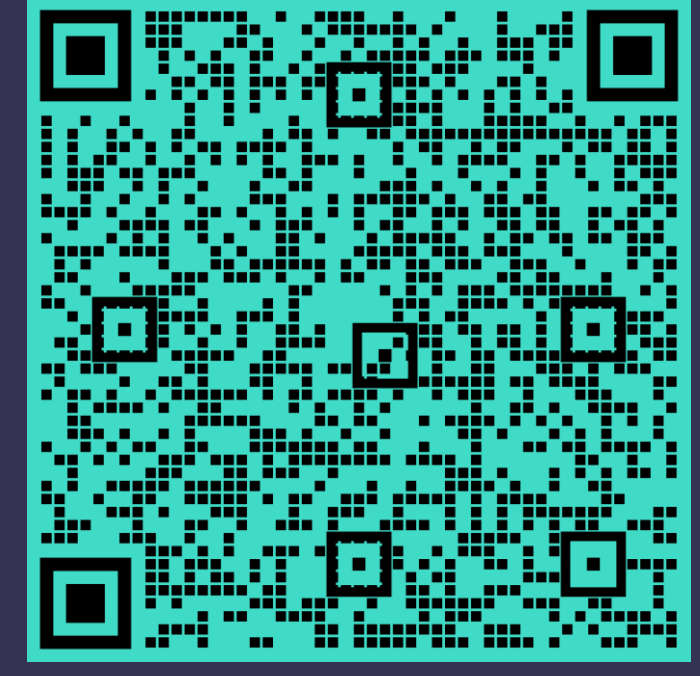- Spontaneous suggestions for applications of AI in healthcare

Young people
- Concerns over accuracy of parents' accounts
- Preference for clinician over AI-driven tool, given expertise
- Predicting mental health outcomes: is it useful for the child? Self-fulfilling prophecy effect?

Key themes

Conclusions

- Automated methods show promise for analysis of 'expressed emotion' negativity
- Both textual and acoustic features contribute to classification, with fusion models showing strongest performance
- Variety of public attitudes to use of AI-driven technology in healthcare, with a focus on implementation & accessibility
- Creative methods such as storytelling are productive ways of engaging the public in science
- Future work: gather and analyse FMSS data from multiple sites at scale