

Computational Audio Modelling for Robot-Assisted Assessment of Children’s Mental Wellbeing

Nida Itrat Abbasi^{1,†}, Micol Spitale^{1,†}, Joanna Anderson², Tamsin Ford², Peter B. Jones² and Hatice Gunes¹

¹ Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Ave, Cambridge, UK, CB3 0FD.

² Department of Psychiatry, University of Cambridge, Herchel Smith Bldg, Robinson Way, Cambridge, UK, CB2 0SZ.

† Equal Contribution, Corresponding authors. [nia22, ms2871, jpa44, tjf52, pbj21, hg410]@cam.ac.uk.

Abstract. Robots endowed with the capability of assessing the mental wellbeing of children have a great potential to promote their mental health. However, very few works have explored the computational modeling of children’s mental wellbeing, which remains an open research challenge. This paper presents the first attempt to computationally assess children’s wellbeing during child-robot interactions via audio analysis. We collected a novel dataset of 26 children (8-13 y.o.) who interacted with a Nao robot to perform a verbal picture-based task. Data was collected by audio-video recording of the experiment session. The Short Mood and Feelings Questionnaire (SMFQ) was used to label the participants into two groups: (1) “higher wellbeing” (child SMFQ score \leq SMFQ median), and (2) “lower wellbeing” (child SMFQ score $>$ SMFQ median). We extracted audio features from these HRI interactions and trained and compared the performances of eight classical machine learning techniques across three cross-validation approaches: (1) 10 repetitions of 5-fold, (2) leave-one-child-out, and (3) leave-one-picture-out. We have also computed and analysed the sentiment of the audio transcriptions using the ROBERTa model. Our experimental results show that: (i) speech features are reliable for assessing children’s mental wellbeing, but they may not be sufficient on their own, and (ii) verbal information, specifically the sentiment that a picture elicited in children, may impact the children’s responses.

Keywords: children, wellbeing assessment, affective computing, robots, speech features

1 Introduction

Affective robots are defined as “robots that can recognize human emotions and show affective behaviors” [26]. Past works have largely explored the use of affective robotics, either with virtual or physical agents, to promote the mental well-

being of people, such as improving communication skills during therapy for children with autism [28] and reducing the feeling of loneliness in elderly people [6]. In addition, with the recent advances in affective computing [24], computer scientists investigated how to automatically assess human affect, specifically concerning promoting mental wellbeing in adults (e.g., recognizing depression and anxiety) using speech markers [16]. However, to the best of our knowledge, none of them explored how to automatically assess the mental wellbeing of children in a robot-aided interaction.

This paper presents the first step towards the automatic robot-assisted assessment of children’s wellbeing from speech using the child-robot interaction (cHRI) dataset introduced in [1]. To assess children’s mental wellbeing, this work focuses on speech features because they have been shown to be good indicators to recognize depression in people [10]. We ran multiple experiments with eight classical machine learning techniques (e.g., logistic regression, decision tree), and we cross-validated them using three approaches: i) 10 repetitions of 5-fold, ii) leave-one-child-out, and iii) leave-one-picture-out cross-validations. These experiments were conducted specifically on a picture task (Task 3 introduced in [1], inspired by the Children Apperception Test (CAT) [4]) because this was one of the tasks in the study with free-flowing conversation. These evaluations enabled us to understand how speech markers can be informative in assessing children’s wellbeing computationally. In our earlier paper [1], we introduced the experimental design of the cHRI study to evaluate mental wellbeing in children. We also compared the robotised administration of psychological questionnaires with the established standards of self-report and parent-report modes of questionnaire administration. Our results show that robotised mode of test administration is more effective in identifying cases with wellbeing related concerns in children. Differently from this previous work [1], in this paper, we undertake the following: (1) the implementation of computational models to assess children’s mental wellbeing in a robot-aided interaction, utilising speech features (not yet explored in the literature); and (2) the interpretation of these models in light of the validation methods utilised that can inform future research on automatic assessment of children’s wellbeing.

2 Background

Children’s psychological distress can impact their mental wellbeing, negatively influencing the academic outcome and relationships. Socially Assistive Robots (SARs) have been effectively used in children to provide companionship [9], clinical related support [25], and academic assistance [22]. In the last decade, many health-related initiatives (“ALIZ-E” project for diabetes management [5], “DREAM” project for providing robot therapy for Autism [14]) have been set up where child-robot interaction has shown promising results. SARs have also been instrumental in enabling children to be more open about their “true feelings” [13] and have been employed successfully not just to interview children regarding abuse, violence and bullying but also to change their perspectives on

some of their formerly carried out misdemeanours [7, 13, 23]. Thus, child-robot interaction can help make children talk about their thoughts and feelings without fear of being intimidated or that they are doing something wrong. To this end, in this work, we have utilised SARs to automatically identify children with wellbeing-related concerns using speech.

With the advance of machine learning, many computational models have been used to learn representations from speech data. Specifically, past works have investigated the use of speech signals for recognizing mental wellbeing disorders, such as depression and anxiety of humans because from a clinical point of view, speech markers usually inform the diagnosis of distress (e.g., duration of the speech, speech tone, pitch) and also the speech signals are very easy to record via non-intrusive devices [27]. In fact, in [10], the authors reviewed the state of the art of speech analysis to assess depression and suicide risk. They highlighted the importance of identifying and using speech markers in automatic model design that are interpretable from a clinical standpoint. Analogously, researchers in [19] conducted a literature review on speech analysis to assess psychiatric disorders automatically (e.g., depression, bipolar, anxiety). They presented a set of limitations to overcome in this field, and they suggested that comprehensive transdiagnostic and longitudinal studies are needed to further advance in the automatic assessment of those disorders. In [32], the authors investigated the relationship between emotion and depression affected speech. Their results showed that speech-based emotional information contributes to the classification of depressed individuals. Also, previous work [21] investigated how noise and reverberation affected depression detection from speech. Finally, the authors of [3] focused on the cross-cultural and cross-linguistic characteristics and how those aspects played a role in depressed speech using verbal biomarkers.

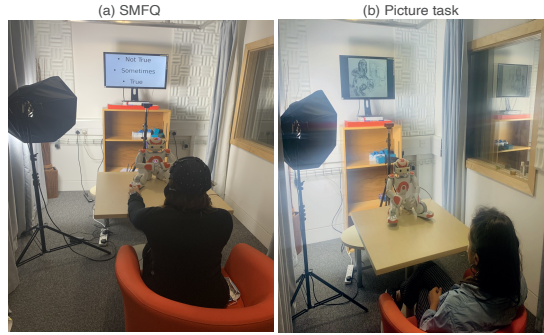
3 Methodology

This section discusses the methodology followed during the study: participants recruited, protocol, experiment tasks and data annotation.

Participants: We collected a dataset of 26 children between 8-13 years old (mean age= 9.6 y.o., SD = 1.5. y.o., 19 females and 7 males) - who were interacting with a Nao robot as shown in Figure 1. We recruited the participants via school advertising and/or through contacts in the research team in the Cambridgeshire area, United Kingdom. We did not include children with neurological and psychological disorders declared by their guardians.

Protocol: We recorded the sessions using the Jabra disc microphone placed on the table where the robot was seated and two cameras. The robot followed a pre-programmed script and asked children to perform a picture task inspired by the Child Apperception Test (CAT). We opted for this specific task because it enables the assessment of the personality and wellbeing traits of children [4], and

Fig. 1: Experimental setup with the Nao robot and the display screen while performing (a) the SMFQ test and (b) the picture task (images from the actual cHRI sessions are not displayed to protect the children’s privacy).



variations of this task have been used previously in HRI [8].

Experimental Tasks: The task consisted of 3 images which correspond to card 7 (a tiger with claws and fangs is seen jumping towards a monkey, Picture 1), card 9 (a rabbit seats on a bed and looks through an open door of a dark room, Picture 2) and card 10 (baby dog lying on another bigger dog, both exhibiting minimum expressions, in the background of a bathroom, Picture 3) respectively from the CAT [4] as they were most related to our goal (task duration: 5 - 10 mins). Children were asked to tell a story related to the pictures displayed on a computer screen placed behind the robot (as shown in Figure 1). Specifically, the robot asked the following questions for each of the displayed images: (1) Picture 1: “What do you see in this picture?”, “What do you think happened before in this picture?”, “What do you think happened after in this picture?”; (2) Picture 2: “What do you think is happening here? Do you notice anything unusual about the picture?”, “Do you think it was something that happened for real, or is this made-up?”; (3) Picture 3: “What do you see in this picture?”, “What do you think happened before in this picture?”, “What do you think happened after in this picture?”. Prior to the picture-based task, the robot administered the Short Mood and Feelings Questionnaire (SMFQ) [30] asking children to verbally answer the questionnaire statements, choosing among the options available displayed on the screen (i.e., “True”, “Sometimes”, or “Not true”).

Data annotation: In our previous work [1], we divided the participants in the study into three groups (lower tertile, medium tertile and higher tertile). However, in this work, we decided to split the participants into two groups (“lower wellbeing” and “higher wellbeing”) based on the median of the SMFQ score because our preliminary analysis - which is out of the scope of this paper - showed no differences in speech features of the three clusters. Note that the SMFQ is typically used for assessing wellbeing over the last two weeks and not for detecting momentary changes during or after a task. Therefore, we used the SMFQ for labelling the population based on their general wellbeing, prior to the task. The resulting groups are (1) low quantile category (child’s SMFQ score \leq median SMFQ score) that we labeled as “higher wellbeing”, and (2) high quantile

category (child’s SMFQ score > median SMFQ score), that we labeled as “lower wellbeing”. For example, if a child has scored 2 (below 3, SMFQ median) in the SMFQ test, we assigned them to the “higher wellbeing” group; while if a child scored 17 (above 3, SMFQ median), we assigned them to the “lower wellbeing” group. This resulted in the following grouping: 14 participants belonged to the “higher wellbeing” category, and 12 belonged to the “lower wellbeing” category.

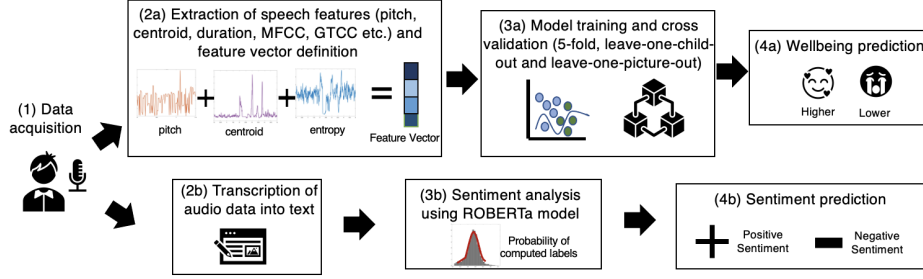


Fig. 2: The experimental methodology consists of (1) dataset acquisition during the cHRI sessions, where audio was recorded using a Jabra mic. (2a) Speech features were extracted and combined as feature vectors that were the input of the (3a) classification algorithms across 3 cross-validation approaches in parallel for (4a) prediction of wellbeing. (2b) Sentiment analysis was also performed on the transcribed text (from the recorded audio files) (3b) using the ROBERTa model for extraction of the sentiment label and the probability of attaining the label for (4b) sentiment prediction.

4 Computational Assessment of Children’s Wellbeing

This section describes the modeling approach used to automatically assess children’s wellbeing from picture task-based interactions with the robot.

Speech Feature Extraction: As children performed the task with the robot via speech-based interaction, we decided to focus only on auditory signals. In addition, speech features [27] (e.g., pitch, speaking rate) have been identified as promising non-verbal cues to recognize depression in patients. In [10], authors reported abnormalities in speech features of patients with depression. For example, monotony in the pitch and toneless voice are often associated with a depressed voice. Thus, our paper focuses only on the speech features of children. For each child’s response clip to each picture task (see Figure 2), we extracted clip-level audio features using a state-of-the-art Matlab audio toolbox ³. Specifically, we extracted 91 audio features, including interpretable features such as the duration of children’s speech and pitch, and lower-level auditory features, such as MFCC, GTCC and spectral centroid. We first removed constant and null

³ <https://uk.mathworks.com/help/audio/ref/audiofeatureextractor.html>

features to prepare the speech features for the machine learning models. Then, we decided to condense the temporal information of each child’s response clip to each picture task into statistical descriptors as in [20], computing a fixed-length vector for each feature of each clip that consists of mean, median, standard deviation, minimum, maximum, and auto-correlation with 1-second lag (resulting in a feature vector with size $26 \times 91 \times 6$).

Model Training and Evaluation for Wellbeing Assessment: We define the problem of assessing children’s wellbeing as a binary classification problem that predicts the “lower wellbeing” and “higher wellbeing” of children based on the audio-clip-level features extracted. We selected a set of classical machine learning techniques, namely logistic regression, linear SVM, decision tree, random forest, AdaBoost, XGBoost, Radial Basis Function (RBF) kernel SVM, and bagging, using scikit-learn python library ⁴ and we trained those models with the dataset collected to assess children’s wellbeing. To validate the models, we exploited three different cross-validation approaches: 1) 10 repetitions of 5-fold, 2) leave-one-child-out, and 3) leave-one-picture-out cross-validations. First, we implemented a 5-fold stratified cross-validation repeated 10 times, resulting in 50 folds. We repeated the 5-fold cross-validation to improve the model results since different dataset splits can result in very different model estimations. We opted for stratified cross-validation to guarantee the same numbers of “higher wellbeing” and “lower wellbeing” labeled data in each fold. Second, we cross-validated our models, leaving one child out to ensure subject-independent predictions, commonly used in [11, 12]. Our dataset contains multiple observations of the same child (for each child, we collected 3 data points, one for each picture task) and using the same subject in both training and test sets can affect the model’s generalization capabilities. This results in 26 folds (the number of children involved in the dataset). Third, our dataset contains 26 observations of the same picture, thus we evaluated our models also exploiting a leave-one-picture-out approach to investigate the effect introduced by the different pictures (3-fold cross-validation). All three validation approaches have been optimized by tuning their hyper-parameters during training with the Optuna framework [2]. To evaluate the models, we computed the classifiers’ accuracy, recall, and precision and then we compared those evaluation metrics to select the best model for each cross-validation approach. Once we obtained the best-performing model, we ran a feature importance analysis to understand which features contributed most to the model.

Sentiment and Speech Feature Analyses: To extract verbal information, we manually transcribed the children’s speech. We ran a sentiment analysis of the transcribed text, exploiting the ROBERTa model [18]. We extracted the predicted sentiment label (positive and negative) and the probability of attaining the computed label. For the categorical predicted labels (positive and negative), we ran Chi square tests to evaluate the differences between population groups

⁴ <https://scikit-learn.org/stable/>

(“higher wellbeing” and “lower wellbeing”) and between pictures (Picture 1 vs. Picture 2 vs. Picture 3) for each population group. In order to investigate group-level changes (“higher wellbeing” vs. “lower wellbeing”) in speech features, we conducted a Wilcoxon rank sum test between the probability of prediction of the negative sentiment and the top 25 most discriminative speech features. Further, we conducted a Friedman’s test to understand the effect of pictures across the “higher wellbeing” and “lower wellbeing” groups. Post-hoc analysis Wilcoxon sign rank tests were conducted across pictures (Picture 1 vs. Picture 2 vs. Picture 3) for the predicted probability of negative sentiment and the top 25 most discriminative features.

5 Results & Discussion

This section reports the results obtained from the training of the models described in Section 4, including the comparison of the models explored and the statistical analyses conducted to interpret the models’ results.

Model Predictions: The models that performed best are respectively the random basis function SVM with a mean accuracy of 83% for the 5-fold (recall=75%, precision= 87%, and F1=79%) and 90% for the leave-one-picture-out validations (recall=83%, precision= 94%, and F1=88%), and the decision tree with a mean accuracy of 70% (recall=36%, precision= 46%, and F1=39%) for leave-one-child-out case. For the sake of clarity, we report in Table 1 the performance of the models in terms of accuracy with the three validation techniques. Our results show that speech features can be promising in investigating children’s mental wellbeing-related concerns. This suggests that speech features may be used for the automatic assessment of children’s wellbeing. As expected, the model validated using the leave-one-child-out has a lower accuracy with respect to others because the training set and the test set contain different subject data (i.e., if the data of a child is included in the training set, this child’s data are not included in the test set), making the model subject-independent. It is also worth noting that the precision of the leave-one-child-out decision tree model is 46%, resulting in a very low performance (i.e., a high number of False positives) that cannot be yet used to assess child wellbeing in cHRI settings, given the importance of the model prediction’s precision. To understand and interpret these results, we looked into the folds of the leave-one-child-out and leave-one-picture-out cross-validations. The folds are depicted in Figure 3. As shown in Figure 3(a), the model accuracy varied between 70% and 80% across the folds, except for children 1 and 2, who behave very differently with respect to the other data points of the sample, resulting in respectively 100% and around 60% of model’s accuracy. Analogously, Figure 3(b) shows that fold 2, corresponding to Picture 2, results in around 92% of accuracy while the other two folds that represent respectively Picture 1 and Picture 3 have around 90% of accuracy.

Analysis: From the sentiment analysis, our results suggest that, in both “lower wellbeing” and “higher wellbeing” groups, all the pictures elicited a more neg-

Models	5-fold	Leave-One-Child-Out	Leave-One-Picture-Out
Logistic regression	0.74	0.56	0.79
Linear SVM	0.73	0.59	0.82
Random Forest Tree	0.68	0.41	0.75
Bagging	0.75	0.50	0.82
XGBoost	0.73	0.51	0.81
AdaBoost	0.72	0.50	0.77
Decision tree	0.70	0.70	0.60
RBf SVM	0.83	0.55	0.90

Table 1: Mean accuracy between folds of the models implemented with the three cross-validation approaches.

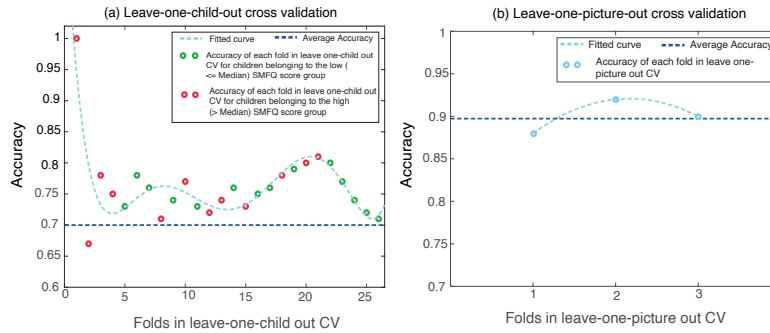


Fig. 3: Visualisation of the best performing classifiers. (a) Leave-one-child-out Decision Tree mean accuracy (x-axis is the fold indicating the child whose data was left out as the test set). (b) Leave-one-picture-out Radial Basis Function Kernel SVM mean accuracy (x-axis is the fold indicating the picture whose data was left out as the test set).

ative sentiment than a positive one in children, as depicted in Figure 4. This result is in line with the CAT test goal that aims to trigger specific feelings in children, such as loneliness in Picture 2. Specifically, Picture 2 shows a higher negative sentiment in the “lower wellbeing” group than in the “higher wellbeing” group. However, when we conducted a Chi-square test between pictures for the two groups, we did not find the results significantly different. For this reason, instead of only using the labels of the predictions (i.e., positive and negative), we decided to analyse the probability of those predictions. Thus, we conducted a Friedman’s test between the picture conditions (Picture 1, Picture 2 and Picture 3) of the “higher wellbeing” and the “lower wellbeing” groups. Within the “higher wellbeing” group, we found that there was a significant difference between the probability of the negative sentiment between the three pictures ($X_F^2(2) = 7, p = 0.03$). We then ran a post-hoc Wilcoxon signed rank tests analysis with a Bonferroni adjusted alpha level of $0.017(0.05/3)$, and we found that the probability of negative sentiment was not significantly different between the conditions (Picture 1, Picture 2, Picture 3). For the group with “lower wellbeing”, another Friedman’s test showed that there was a significant difference between the probability of negative sentiment between the three pic-

tures ($X_F^2(2) = 7.17, p = 0.03$). Post-hoc Wilcoxon signed rank tests with a Bonferroni adjusted alpha level of 0.017(0.05/3) showed that there was no statistically significant difference in terms of the probability of negative sentiment between the conditions (Picture 1, Picture 2 and Picture 3).

Alongside the sentiment analysis, we also conducted a feature importance

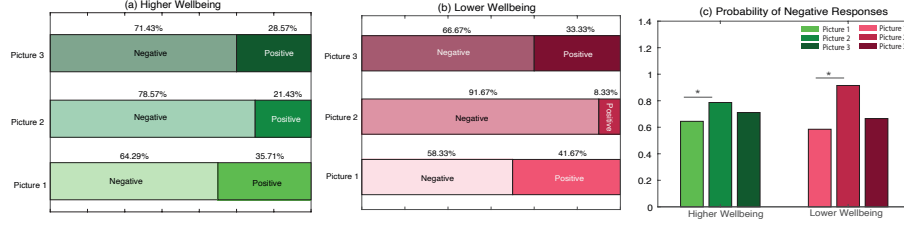


Fig. 4: Sentiment analysis was performed using ROBERTa model to determine the sentiment labels (positive, negative) for “higher wellbeing” (a) and “lower wellbeing” (b) groups. (c) Negative sentiment prediction probability compared between the pictures and the groups (“higher wellbeing” and “lower wellbeing”). * $p < 0.05$ uncorrected.

analysis to compute the top 25 features from a 5-fold cross-validation analysis of the RBF SVM model. We observed that the feature that contributed the most to the model prediction was the duration of the children’s speech, followed by the Delta GTCC. To further investigate the contribution of the most important feature, we conducted a Friedman’s test between the conditions (Picture 1, Picture 2 and Picture 3) of the “higher wellbeing” and the “lower wellbeing” groups for the speech duration. Within the “higher wellbeing” group, we found that a Friedman’s test showed that there was a significant difference in the duration of the speech feature between the three conditions (Picture 1, Picture 2 and Picture 3) ($X_F^2(2) = 19, p < 0.001$). We then ran the post-hoc Wilcoxon signed rank tests with a Bonferroni adjusted alpha level of 0.017(0.05/3) and we found a statistical significant difference of the speech duration ($W = 105, p < 0.001$ corrected) between Picture 1 ($Mdn = 51.5$) and Picture 2 ($Mdn = 18.5$). Post-hoc Wilcoxon signed rank tests with a Bonferroni adjusted alpha level of 0.017(0.05/3) also resulted in a statistically significant difference of the duration of the speech ($W = 12, p = 0.02$ corrected) between Picture 2 ($Mdn = 18.5$) and Picture 3 ($Mdn = 35$). For the “lower wellbeing” population, another Friedman’s test showed that there was a significant difference between the three conditions (Picture 1, Picture 2 and Picture 3) ($X_F^2(2) = 10.085, p = 0.006$). Analogously, we found that the post-hoc Wilcoxon signed rank tests with a Bonferroni adjusted alpha level of 0.017(0.05/3) showed a statistically significant difference in terms of speech duration ($W = 66, p = 0.003$ corrected) between Picture 1 ($Mdn = 49.5$) and Picture 2 ($Mdn = 30$). The other feature tests (e.g., GTCC, Delta GTCC) that resulted in being significant are not reported because they are out of the scope of this paper that only focused on interpretable features.

Discussion: Our results show that the mean accuracy of the models in leave-one-child-out and leave-one-picture-out evaluations varied across folds. Specifically,

we observe that for the leave-one-child-out evaluation, the mean accuracy varied between 70% and 80%, meaning that a set of additional child-specific features have a key role in the model learning. The model should consider child-specific characteristics that, for instance, can emerge from non-auditory behaviors (e.g., visual cues) or personal characteristics (e.g., personality). Such additional factors can further inform the model, resulting in child-specific model training. Past works highlighted that one of the main challenges for automatically assessing mental wellbeing is the need for personalized models [33]. Children can behave very differently (e.g., talk more, use different tones), negatively impacting the models’ training and performance, as shown in our findings. For example, the authors in [29] implemented personalized modeling of depression using multiple modalities of data. To qualitatively interpret our results, we watched the videos of children corresponding to the non-average points in the model (child 1 and child 2, see Figure 3(a)). We observed that child 1 was very talkative and used a wide spectrum of intonations to describe the pictures, while child 2 didn’t speak that long, and their pitch was monotonous. Such behavioral differences could be attributed to other factors, such as the children’s personalities. Child 1 appears to be more extroverted and talkative than child 2. Within the psychology literature, many studies [15, 17] have reported a strong relationship between personality and mental health as well as how personality attributes impact the detection of depression in adults [16]. However, to the best of our knowledge, none of them have yet explored the influence of personality on the detection of mental disorders in children. From our sentiment analysis, our results show that experiment stimuli (Picture 1, Picture 2 and Picture 3) have an impact on the responses of children. From leave-one-picture-out validation, we found that the mean accuracy was higher for Picture 2. This could be due to the sentiment of children’s speech. As shown in Figure 4, Picture 2 was the most negative perceived picture, triggering negative feelings in children. The picture conveys a sentiment of loneliness that elicited stronger negative reactions in children⁵.

6 Conclusion

This paper presents the audio analysis of a cHRI dataset for robot-assisted assessment of children’s mental wellbeing. Our results showed that speech features are reliable indicators for assessing children’s mental wellbeing, but they are insufficient for accurate and precise prediction. This suggests that personalised modeling can be the most suitable approach for this assessment task. Limitations of this work include an imbalanced population sample (more girls than boys), novelty effect [31] and the conservative approach in categorisation of the population groups (“lower wellbeing” and “higher wellbeing”) using the median SMFQ score. In future works, we aim to address these limitations to gain further insights into speech-based modeling of robot-assisted assessment of children’s mental wellbeing.

⁵ <https://childmind.org/article/help-children-manage-fears/>

Open access statement: For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Data access statement: Overall statistical analysis of research data underpinning this publication is available in the text of this publication. Additional raw data related to this publication cannot be openly released; the raw data contains transcripts of interviews, but none of the interviewees consented to data sharing.

Acknowledgments: This work was supported by the University of Cambridge’s OHMC Small Equipment Funding. N. I. Abbasi is supported by the W.D. Armstrong Trust PhD Studentship and the Cambridge Trusts. M. Spitale and H. Gunes are supported by the EPSRC project ARoEQ under grant ref. EP/R030782/1. All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014, particularly T. Ford) and NIHR Applied Research Collaboration East of England (P. Jones, J. Anderson). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

1. Abbasi, N.I., et al: Can robots help in the evaluation of mental wellbeing in children? an empirical study. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 1459–1466. IEEE (2022)
2. Akiba, T., et al: Optuna: A next-generation hyperparameter optimization framework. In: Proc. of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
3. Alghowinem, S., et al: Cross-cultural depression recognition from vocal biomarkers. In: Interspeech. pp. 1943–1947 (2016)
4. Bellak, L., Bellak, S.S.: Children’s apperception test. (1949)
5. Belpaeme, T., et al.: Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction* **1**(2) (2012)
6. Bemelmans, R., et al: Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association* **13**(2), 114–120 (2012)
7. Bethel, C.L., et al: Using robots to interview children about bullying: Lessons learned from an exploratory study. In: RO-MAN 2016. pp. 712–717. IEEE (2016)
8. Bremner, P., et al: Personality perception of robot avatar tele-operators. In: HRI 2016. pp. 141–148. IEEE (2016)
9. Crossman, M.K., et al: The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Professional Psychology: Research and Practice* **49**(1), 48 (2018)
10. Cummins, N., et al: A review of depression and suicide risk assessment using speech analysis. *Speech communication* **71**, 10–49 (2015)
11. Fernandez, R., Picard, R.W.: Modeling drivers’ speech under stress. *Speech communication* **40**(1-2), 145–159 (2003)
12. Gholamiangonabadi, D., et al: Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access* **8**, 133982–133994 (2020)
13. Godoi, D., et al: Proteger: A social robotics system to support child psychological evaluation. In: 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian

- Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE). pp. 1–6. IEEE (2020)
14. Gómez Esteban, P., et al.: A multilayer reactive system for robots interacting with children with autism. arXiv e-prints pp. arXiv-1606 (2016)
 15. Hettema, J.M., et al: A population-based twin study of the relationship between neuroticism and internalizing disorders. *American journal of Psychiatry* **163**(5), 857–864 (2006)
 16. Jaiswal, S., et al: Automatic prediction of depression and anxiety from behaviour and personality attributes. In: *ACII 2019*. pp. 1–7. IEEE (2019)
 17. Klein, D.N., et al: Personality and depression: explanatory models and review of the evidence. *Annual review of clinical psychology* **7**, 269 (2011)
 18. Liao, W., et al: An improved aspect-category sentiment analysis model for text sentiment analysis based on roberta. *Applied Intelligence* **51**(6), 3522–3533 (2021)
 19. Low, D.M., et al: Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology* **5**(1), 96–116 (2020)
 20. Mathur, L., et al: Modeling user empathy elicited by a robot storyteller. In: *ACII 2021*. pp. 1–8. IEEE (2021)
 21. Mitra, V., et al: Noise and reverberation effects on depression detection from speech. In: *ICASSP 2016*. pp. 5795–5799. IEEE (2016)
 22. Papadopoulos, I., et al: A systematic review of the literature regarding socially assistive robots in pre-tertiary education. *Computers & Education* **155**, 103924 (2020)
 23. Peter, J., et al: Can social robots affect children’s prosocial behavior? an experimental study on prosocial robot models. *Computers in Human Behavior* **120**, 106712 (2021)
 24. Poria, S., et al: A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* **37**, 98–125 (2017)
 25. Raigoso, D., et al: A survey on socially assistive robotics: Clinicians’ and patients’ perception of a social robot within gait rehabilitation therapies. *Brain sciences* **11**(6), 738 (2021)
 26. Rhim, J., et al: Investigating positive psychology principles in affective robotics. In: *ACII 2019*. pp. 1–7. IEEE (2019)
 27. Ringeval, F., et al: Avec 2019 workshop and challenge: state-of-mind, detecting depression with ai, and cross-cultural affect recognition. In: *Proceedings of the 9th International on Audio/visual Emotion Challenge and Workshop*. pp. 3–12 (2019)
 28. Scassellati, B., et al: Robots for use in autism research. *Annual review of biomedical engineering* **14**(1), 275–294 (2012)
 29. Shah, R.V., et al: Personalized machine learning of depressed mood using wearables. *Translational psychiatry* **11**(1), 1–18 (2021)
 30. Sharp, C., et al.: The short mood and feelings questionnaire (smfq): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of abnormal child psychology* **34**(3), 365–377 (2006)
 31. Smedegaard, C.V.: Reframing the role of novelty within social hri: from noise to information. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. pp. 411–420. IEEE (2019)
 32. Stasak, B., et al: An investigation of emotional speech in depression classification. In: *Interspeech*. pp. 485–489 (2016)
 33. Xu, X., et al: Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **5**(1), 1–27 (2021)