Running Head: Multimodal Affect Analysis

Multimodal Affect Analysis of Psychodynamic Play Therapy

Sibel Halfon, PhD[1]*
Metehan Doyran, MS[2]
Batıkan Türkmen, MS[3]
Eda Aydın Oktay, MS[4]
Ali Albert Salah, PhD[5]

*Corresponding author

This study was partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) Project No: 215K180. This is the uncorrected author proof. For the definitive version, please check the journal website. Please cite as: S. Halfon, M. Doyran, B. Turkmen, E. Aydin Oktay, A.A. Salah, "Multimodal Affect Analysis of Psychodynamic Play Therapy", *Psychotherapy Research,* vol. 31, no. 3, pp. 402-417, 2021. https://doi.org/10.1080/10503307.2020.1839141

1 Istanbul Bilgi University, Kazım Karabekir Cad. No: 2/13
34060 Eyüp İstanbul,Turkey
Phone: 90 212 311 76 75
Email: Sibel.halfon@bilgi.edu.tr
2 Utrecht University, Dept. Information and Computing Sciences, Princetonplein 5, Utrecht, the Netherlands
Phone: +31 6 477 886 68
Email: m.doyran@uu.nl
3 Bogaziçi University, Computer Engineering Dept., 34342 Bebek/İstanbul Turkey
Phone: +90 531 264 05 50
Email: batikan.turkmen@boun.edu.tr
4 Northeastern University, Khoury College of Computer Science, Boston, MA
phone: +1 512 9989610
email: aydinoktay.e@husky.neu.edu
5 Utrecht University, Dept. Information and Computing Sciences,
Princetonplein 5, Utrecht, the Netherlands
Bogazici University, Computer Engineering Dept., 34342 Bebek, Istanbul, Turkey
phone: 0031 6 82 49 83 58
Email: a.a.salah@uu.nl

Abstract

Objective: We explore state of the art machine learning based tools for automatic facial and linguistic affect analysis to allow easier, faster, and more precise quantification and annotation of children's verbal and non-verbal affective expressions in psychodynamic child psychotherapy. Method: The sample included 53 Turkish children: 41 with internalizing, externalizing and comorbid problems; 12 in the non-clinical range. We collected audio and video recordings of 148 sessions, which were manually transcribed. Independent raters coded children's expressions of pleasure, anger, sadness and anxiety using the Children's Play Therapy Instrument (CPTI). Automatic facial and linguistic affect analysis modalities were adapted, developed, and combined in a system that predicts affect. Statistical regression methods (linear and polynomial regression) and machine learning techniques (deep learning, support vector regression and extreme learning machine) were used for predicting CPTI affect dimensions. Results: Experimental results show significant associations between automated affect predictions and CPTI affect dimensions with small to medium effect sizes. Fusion of facial and linguistic features work best for pleasure predictions; however, for other affect predictions linguistic analyses outperform facial analyses. External validity analyses partially support anger and pleasure predictions. Discussion: The system enables retrieving affective expressions of children, but needs improvement for precision.

KEYWORDS: Multimodal Affect Analysis, Face Analysis, Text Analysis, Psychodynamic Play Therapy

CLINICAL IMPACT STATEMENT: The multimodal approach introduced in this paper uses state of the art machine learning based tools specifically adapted, developed, and combined in a system that predicts verbal and non-verbal affect expressions in psychodynamic play therapy. Initial findings show promising results for retrieving affect expressions of children. Suggestions for improvement and future applications are discussed.

Multimodal Affect Analysis of Psychodynamic Play Therapy

The in-session arousal of emotions, also known as affect experiencing, reflects the degree to which a patient viscerally experiences and then expresses their feelings during therapy (Greenberg & Leone, 2006). There is substantial evidence that emotional arousal is important for the success of many different forms of psychotherapy (Lane et al., 2015). Some psychodynamic treatment models with adults (e.g., Malan, 2001; Luborsky, 1984) and children (Hoffman, et al., 2016; Kernberg & Chazan, 1991) specifically emphasize the central importance of experiencing and expressing feelings in therapy, in particular negative affect associated with patients' symptomatology. The premise of these approaches is that certain defenses or anxieties block the expression of difficult and hard to tolerate feelings, which are avoided and acted out as symptoms. For example, in the case of children with externalizing problems, acting out aggressive behaviors is used in order to avoid feelings of hurt and disappointment (Hoffman et al., 2016). Therapists' facilitation of such affective experience/expression predicts patients' improvement in adult psychodynamic treatments (Diener & Hilsenroth, 2007). However, there have been very few studies in this area in psychodynamic play psychotherapy.

One reason for lack of research in this area is that affective analysis of psychodynamic play therapy sessions is a meticulous process, which requires many passes over the collected data to annotate different markers of affective displays. Moreover, because children don't yet have the symbolic lexicon to verbalize emotions as do adults, there is a need to integrate multiple modalities that take into account verbal and nonverbal indicators of affect for a comprehensive affect assessment. Research in multimedia analysis suggests that automatic tools could be used to help the therapists in these tasks. Recording videos of interactions that are subsequently rated by experienced coders has been a viable alternative to self-report measurements, which are difficult to use with children (Larochette et al., 2006; Zeinstra et al., 2009). Since expert coding of facial expressions via Facial Action Unit coders is an expensive

process (Ekman, Friesen & Hager, 2002), computational alternatives were developed to automatically recognize action units from videos (Littlewort et al., 2011). Recently, progress in machine learning, combined with access to very large face datasets caused marked improvements in the accuracy and robustness of such approaches (Baltrušaitis et al., 2018; Jaiswal & Valstar, 2016). Nonetheless, a clear view of the face in the video is a precondition for facial analysis, and it is not possible to guarantee this during the free play of children, unless many cameras are used simultaneously. This, in turn, increases the setup and running costs of such systems. We propose a multimodal approach that combines facial with language-based affect analysis, which overcomes these limitations to a certain degree.

We will present a state-of-the-art multimodal system to predict affect scores of young children with internalizing, externalizing and comorbid problems in psychodynamic play therapy conducted at an outpatient clinic. We will rely on an automated affective analysis approach for Turkish language that uses dimensions of Valence and Arousal (Aydın Oktay, Balcı & Salah, 2015), and a deep learning based facial expression analysis approach. We have specifically developed and/or adapted these tools for use in play therapy sessions, and combined them with further machine learning for representing the overall affect during sessions. Such tools, after development, can be used for representing and indexing large amounts of accumulated session data, to uncover patterns and trends, to visualize affect dynamics, as well as to measure correlations between the automatically extracted expression streams of the therapist and the patient.

The automated analysis of only two modalities will be limited in comparison to trained expert annotations; even if the modality-specific performance is high, the system may miss some affective indicators that are expressed in other modalities (body, paralinguistics, etc.). We test the performance of our system against affect ratings annotated by reliable outside judges using the affect dimensions of the Children's Play Therapy Instrument (CPTI; Kernberg et al., 1998), an observer-rated psychodynamic tool that assesses children's play activity in the

sessions. We will also be assessing the external validity of our predictions, investigating associations with demographic and presenting problem characteristics.

Associations between Negative Emotionality, Internalizing and Externalizing Problems

Externalizing and internalizing problems have concurrently and longitudinally been related to extreme negative emotions (e.g., Eisenberg et al., 2005). In particular, anger has been related to externalizing and comorbid problems (Eisenberg et al., 2005), whereas dysphoric affect such as fearfulness, anxiety and depression have been linked to internalizing symptoms (e.g., Eisenberg, et al., 2005; Oldehinkel, Hartman et al., 2004; Lengua, 2003). The relation between negative emotions and play characteristics has also been widely studied (see Russ & Niec, 2011 for a review). Overall, the frequency of negative affect observed during a play task is significantly correlated with mental health difficulties. Empirical research mostly coming from developmental psychology shows that children with externalizing problems display more negative affect in their play, particularly aggression (Dunn & Hughes, 2001). Von Klitzing et al. (2000) found that expressing negative and/or aggressive affect in disorganized pretend play predicted externalizing problems. Children with internalizing problems also show higher levels of negative affect and low affective arousal in play (Halfon, et al., 2016). Additionally, in a sample of 322 six year-olds, negative affect in play significantly correlated with both internalizing and externalizing behaviors (Scott et al., 2006).

Most psychodynamic models of treatment aim to help children with internalizing and externalizing problems express disruptive negative emotions in symbolic play in a supportive therapy relationship followed by the interpretations of their underlying meanings (e.g., Hoffman et al., 2016; Kernberg & Chazan, 1991). For example, Regulation-Focused Psychotherapy for Children (RFP-C; Hoffman et al., 2016), a manualized psychodynamic treatment for externalizing problems, encourages the expression of children's negative emotions to increase the children's understanding that disruptive behaviors have meaning in the service of avoiding painful dysphoric affect such as shame, sadness and/or anxiety. Kernberg and Chazan (1991), in their manualized psychodynamic treatment model for children

with conduct disorders, also help children express feelings in a way that is both communicative and safe, followed by more expressive/interpretative techniques aimed at understanding the ways in which negative emotions are avoided.

In fact, expression of affect is a particular feature of psychodynamic therapies and can reliably distinguish it from both cognitive-behavioral (CBT) and interpersonal approaches (Blagys & Hilsenroth, 2000). In adult psychodynamic treatments, bringing troublesome feelings to awareness, and facilitating expression of the patient's negative affect are associated with good outcome (Ablon, et al., 2006; Jones & Pulos, 1993). In child psychodynamic psychotherapy, only Halfon et al. (2019) studied the associations between increase in in-session negative affect expression and outcome. They found that negative emotion expression in treatments where there was an emphasis on mentalization (defined as labeling, understanding and attuning to mental states such as feelings, needs, beliefs and desires; Fonagy et al., 2002) significantly predicted children's affect regulation and positive outcome. In particular, they found that for both children with both internalizing and externalizing problems, it was the expression of dysphoric affect such as fear, sadness and anxiety in the context of therapists' and children's understanding and attunement towards these feeling states that was associated with improvements in affect regulation.

Automatic Analysis of Affect

Given the negative emotionality of children with internalizing, externalizing and comorbid problems and the central role of affect expression in psychodynamic play therapy, there is a need for immediate tools that can effectively assess verbal and nonverbal affect characteristics of children and therapists. To the authors' knowledge, there is presently no such tool that is particularly developed for psychodynamic play therapy.

Affective computing is the subfield of computer science that seeks to develop computer-based approaches for automatic assessment of affect in humans (Picard, 1995). Efforts in this area have mostly focused on the evaluation of facial appearance and dynamics for basic and non-basic displays of affect, as well as paralinguistic analysis from speech (Zeng

et al., 2009). Basic emotions refer to the affect model developed by Ekman (1994), and refer to six classes: happiness, sadness, surprise, fear, anger and disgust. More recently researchers considered non-basic emotion recognition using a variety of alternatives which represents a wider range of emotions, using continuous modelling of affect dimensions such as arousal, valence, and dominance (Gunes & Schuller, 2013). Other modalities from which affect can be sensed include body posture and motion (Kleinsmith & Bianchi-Berthouze, 2013), physiological signals (Calvo & D'Mello, 2010), and language use (Munezero et al., 2014). The earlier approaches have focused on detecting and distinguishing a small number of classes (e.g., six basic facial expressions like surprise, happiness, anger, fear, sadness and disgust), recorded in highly controlled conditions. The field slowly moved towards the detection of continuous and dimensional emotions (Gunes & Pantic, 2010) and from recordings in the lab conditions to detecting emotions 'in the wild' (Kaya et al., 2017a). This is particularly relevant for psychotherapy research, where in-session data is collected in uncontrolled unobtrusive conditions in order not to disrupt the natural flow of the session.

For each modality of computer-based affect analysis, deep learning approaches currently dominate the state-of-the-art. A deep neural network is a complex machine learning model with millions of free parameters, and is typically trained with very large datasets. The trained models can be used later in other application settings. For face and body analysis, a number of open source tools have been made available, such as OpenPose (Cao et al., 2017), and OpenFace (Baltrušaitis et al., 2018). With these tools, it is possible to detect body pose and faces in videos, obtain the locations of facial landmarks, estimate head pose and gaze directions, and to evaluate basic facial expressions. By tracking these modalities over time, it is possible to get a good estimate of expression and pose changes. However, complex emotional displays (such as frustration, disappointment, elation, triumph, shame) are still largely beyond these approaches, as it is difficult to find sufficient labeled data to properly train automatic analysis tools. Under controlled recording conditions, these models can be trained to infer basic expressions of emotion (happy, sad, angry, fearful, disgusted, surprised) with relatively high

reliability, comparable to that of humans (Martin Wegrzyn et al., 2017). For example, 99.6% accuracy on the CK+ facial expression dataset (Lucey et al., 2010) was reported with automatic approaches for this task (Li & Deng, 2020). Moreover, deep neural networks such as AffectNet (Mollahosseini et al., 2017) can do a good job of mapping continuous (i.e. valence and arousal) affect of a facial image. The valence-arousal representation is based on several empirical works on representation of affect that were very influential in the development of computational models of analysis (Mehrabian, 1970; Russell, 1980). Face analysis tools specifically tailored for children are lacking in the affective computing literature. Recently, Khan et al. (2019) have introduced the LIRIS-CSE database for children's spontaneous expression recognition, but this database contains data from 12 children, not sufficient for training customized classifiers.

For language and speech based affect analysis, different sets of tools should be used for each language. Even for paralinguistic affect detection, analysis methods do not generalize well from one language to the other (Kaya et al., 2017b). There are comprehensive tools for affect detection from English texts, and these are useful in a play therapy setting when the speech of the child is transcribed, but for underrepresented languages like Turkish, there are no comprehensive tool sets (especially for affect analysis) based on large corpus studies (Oflazer & Saraçlar, 2018). Current approaches for automatic emotional and affective content analysis from text generally focus on the processing of social media texts (Mohammad, 2016), and less on speech in natural interactions. The most straightforward approaches are based on keyword spotting, which matches a set of detected keywords to a look-up table that contains keywords and their affective values. The basic limitation of this approach is that it is incapable of dealing with negation and with complex sentence structures.

To increase the robustness of the results, it is beneficial to combine multiple modalities such as face and text analyses (D'Mello & Kory, 2015), which can be done at feature level (i.e., concatenating features and using a single model), or at decision level (i.e., using different models for each modality and fusing their decisions with another, typically simpler model). Feature level combinations can learn to model correlations between modalities, but we chose

to fuse at the decision level, which is more suited when modality representations are very different (Kaya et al., 2017a).

Aims of the Study

Our aim is to describe and test the preliminary effectiveness of the multimodal system for computational affect analysis. Previously, the potential of such an automatic approach applied to psychodynamic play therapy was illustrated by measuring the mean squared error between affect scores rated by trained observers and the affect predictions of a face and text analysis based system (Doyran et al., 2019). Moreover, Halfon et al. (2016) applied natural language processing (NLP) techniques to study the affect expression in psychodynamic play therapy, automatically annotating spoken sentences on valence and arousal dimensions. The performance of the affect analysis was tested on longitudinal psychotherapy data, showing good results in line with observer rated assessments. We extend these studies to assess the use of state-of-the-art facial and text analysis systems as a tool for the child psychotherapists. These can serve for affective indexing, search and retrieval of the collected session material, as well as for quantifying some affective indicators, such as facial and verbal expressions of affect.

We work on a dataset that is collected under natural settings during psychotherapy sessions (i.e., a legacy data set which is more difficult to process). Thus, there was much background noise behind the speech data collected from the microphones; therefore, it was not eligible for automatic speech recognition (ASR). Instead we use manual text transcripts of the sessions. Our recordings come from two static cameras resulting in relatively few clear face shots with low resolution, which is known to reduce the quality of automatic assessment. However, the recording setup represents a typical, realistic, in-the-wild setting. This setup benefits from substantial external validity, as it more accurately reflects the reality of conditions with patients in clinics.

We describe the face and text modalities in our system under separate subsections, followed by a short description of the fusion approach. In each modality, we describe the development of the measure and afterwards we present preliminary associations with the affect

scores derived from independent codings of the affect items from the Children's Play Therapy Instrument (Kernberg et al., 1998) from 53 children and 148 sessions in psychodynamic play therapy. Significant associations between our automatic measures and CPTI affect codings illustrate the potential of the automatic approach. Moreover, in order to test the external validity of our automatic predictions, we investigate associations with gender, age and internalizing and externalizing problems.

Methods

Patient Characteristics

This data is comprised of children who were admitted to the Istanbul Bilgi University Psychological Center between Fall 2018 to Fall 2019. This is a subsample of a larger research program that aims to assess baseline predictors and effective treatment factors associated with outcome in psychodynamic child psychotherapy. The larger data collection and detailed outcome analyses have been reported by Halfon (2020). Referrals were made by parents themselves or by mental health, medical, and child welfare professionals. The parents and the children were screened by a licensed clinical psychologist in order to determine whether the patients fit the study protocol inclusion criteria: ages between 4-10 years old, no psychotic symptoms, no significant developmental delays, no significant risk of suicide attempts and no drug abuse. The patients and their parents were extensively informed before commencing therapy about research procedures, and parents provided written informed consent, and the children provided oral assent concerning use of their data, including questionnaires, videotapes and transcripts of sessions for research purposes. This research was approved by Istanbul Bilgi University Ethics Committee.

All the children were born in Turkey and came from relatively homogeneous urban neighborhoods and belonged to low to middle socioeconomic status (SES). Twenty-six percent of the children were 4–5 years old, 28 % were 6–7 years old, 46 % were 8–10 years old (M = 6.98; SD = 2.14). 69 % of the sample was female. They were referred most frequently due to internalizing and externalizing problems such as rule-breaking and aggressive acts (48 %),

followed by anxiety complaints (26 %), school-related problems (19 %) and social problems (7 %). 9 % of the children had internalizing problems, 11 % had externalizing problems and 57 % had comorbid internalizing and externalizing problems according to Child Behavior Checklist (CBCL; Achenbach, 1991) and 23 % were in the non-clinical range.

Therapists

The therapists were 24 clinical psychology master's level clinicians, who were mostly female (95 %) and aged between 23 to 27 years (*M* Age= 23, *SD* = 1.15). Each therapist was educated in the theoretical background of psychodynamic play therapy for two years in theoretical courses. All therapists had one to two years of supervised psychotherapy experience. On average, therapists treated two patients. Each therapist received one hour of individual and three hours of group supervision by licensed psychodynamic supervisors with at least ten years of experience.

Treatment

The standard treatment applied at our psychotherapy center is psychodynamic play therapy. The therapy mainly follows an object-relational framework, working on children's self-other representations and the accompanying mental states such as feelings, needs, wishes and beliefs using children's play as a main source of internal expression (i.e., Verheugt-Pleiter, Zevalkink & Schmeets, 2008). Cases were assigned to therapists on the basis of therapists' availability. The standard treatment plan at the clinic involves once weekly therapy sessions of 50-minutes with the child, along with once a month parent sessions. The treatments are open ended in length and are determined based on progress towards goals, life changes and patients' families' decisions. On average patients receive 40 sessions over a ten-month period. The treatment lengths vary among 53 patients in the current study, with the mean number of sessions for this sample being 36.5 (*SD* = 19.25, range = 12-65).

Assessment Measures

The Child Behavior Checklist (CBCL; Achenbach, 1991) is a widely used method of identifying problematic behaviors in children with two separate versions for ages 1.5-5 and 6-

18. CBCL indicates how true a series of 112 problem behavior items are on a three-point scale (0 = “not true”, 1 = “somewhat true”, and 2 = “very true or often true”). Outcomes can be determined for significant problems for internalizing (e.g., depression, anxiety), externalizing (e.g., aggression, violence), or total problems. This scale has high levels of internal consistency (CBCL 1.5–5 and 6–18: α = 0.97) and one week test–retest reliability (CBCL 1.5–5: $r = 0.90$; CBCL 6–18: $r = 0.94$). The scale has been adapted to Turkish with good internal consistency and test-retest reliability for the internalizing ($\alpha = 0.87$, $r = 0.93$), externalizing ($\alpha = 0.90$, $r = 0.93$) and total problems scales ($\alpha = 0.94$, $r = 0.93$; Erol & Şimşek, 2010). In the current study, all three subscales showed good to high degrees of internal consistency ($\alpha = 0.75, 0.87, 0.92$ for internalizing, externalizing and total problems, respectively).

Observer-rated Affect Measure

Children’s Play Therapy Instrument (CPTI; Kernberg et al., 1998) rates children’s play activity in therapy at different levels such as descriptive, cognitive, affective, social and functional dimensions. Previous studies have shown good inter-rater reliability (Chari et al., 2013; Kernberg et al., 1998). The measure has been found to be sensitive to changes in psychotherapy (Chazan, 2000, 2001; Chazan & Wolf, 2002) and has shown good convergent and predictive validity in relation to associations between play characteristics and behavioral problems (Halfon, 2017) and discriminant validity in differentiating traumatic vs. normal play characteristics (Cohen & Chazan, 2010). Author #1 was trained by Saralea Chazan on the CPTI. Six master’s level research assistants, who received 20-hours of training on the CPTI by Author #1 and rated 10 training sessions prior to the study, rated the sessions. They were independent assessors, who were not associated with the treating clinicians or the cases, and blind to the purposes of the study. During the training, they rated practice videos until their inter-rater reliability reached an intra class correlation (ICC (2,1)) of 0.70. Afterwards, pairs of coders independently coded the sessions with good to excellent ICCs (2,1) ranging from 0.75 to 0.96 ($M = 0.89$; $SD = 0.08$). The two sets of independent ratings were then averaged.

In this study, we only use the affect expression items of the instrument. Affect expressed in play measures how much the child shows the following emotions on a 0-5 Likert scale (5 = Most Characteristic, 0 = No Evidence). Affect is rated either when an affect theme is expressed in the play (e.g., one animal hitting or saying "I hate you" to another animal) or when affect-laden content is referenced (e.g., "This is a gun"), and/or there is non-verbal expression of affect such as facial expressions, postural cues, nuances of language (tone, volume, intonation). In general, combinations of affect expression, affect word, and content themes get higher ratings. (1) *Anger* may include themes of fighting, destruction, harm to another character, or aggressive dialogue, as well as expressions such as "I am mad". (2) *Anxiety* may include scary themes like monsters, ghosts or hiding from others, as well as expressions such as "I am scared". It may also include themes like school anxiety, concerns about punishment, and worry, as well as signs of agitation. (3) *Sadness* may include themes of loneliness and expressions of pain, sadness or crying. *(4) Pleasure* may include general preference statements, indications of having fun as well as expressions of happiness.

Automatic Affect Analysis Tools

While fully automatic analysis of affect from either facial or linguistic cues is error-prone and noisy, combining and pooling results of such analysis over months of sessions can provide the therapist with an overview, simplifying access to stored data and saving time in analysis. We adapt several state-of-the-art tools for our problem, and describe these modalities separately (see Fig.1).

Face analysis. We automatically locate the child's face to extract emotional features. The main difficulties for a computer-based facial affect analysis system are the frequent occlusions and diverse range of body and head movements in the recorded therapy videos. Our recordings come from two static cameras (2MP WDR EXIR Turret Network Camera, Hikvision) positioned at the two opposite corners of the room around 1.5 meters high, resulting in relatively few clear face shots of the child during play because the children play mostly sitting on a chair or on the floor and facing downwards towards the toys. Most extracted faces

are low resolution, which is known to reduce the quality of automatic assessment. However, the recording setup represents a typical, realistic, in-the-wild setting.

We use the OpenPose system (Cao et al., 2017) to detect face and body landmarks in play therapy videos. OpenPose is a multi-stage convolutional neural network with early stages processing visual primitives, and each later stage responding to more complex features. Convolutions give the network flexibility in locating features across the image. We selected this network, as it has been shown to perform well for uncontrolled imaging conditions. The system locates 70 facial and 25 body landmarks (neck, hip, shoulders, etc.), whose trajectories are then smoothed by a sliding window approach and combined with a tracker, which eliminates the effects of noise introduced into the system by frame-by-frame detection. This allows us to locate, detect, and track the people in the videos. By assuming that each video frame should contain one child and one therapist at most and by comparing the distance between the automatically detected hip joint and neck landmarks, we can automatically track the child and the therapist in the room, and extract only the child's facial images.

To represent the facial affect, we use a state-of-the-art deep neural network that was pre-trained on the AffectNet database (Mollahosseini et al., 2017). The network is used to produce valence and arousal scores for faces that have emotional expressions, and has a ResNeXt (Xie et al., 2017) style architecture with 50 layers and 25 million parameters. The technique of pre-training a network on a larger dataset with a related task or setting is known as transfer learning (Torrey & Shavlik, 2010). We use this approach, since we do not have a large set of affect-annotated facial expressions acquired from psychotherapy sessions. The AffectNet database has more than 450,000 manually annotated in-the-wild facial images for continuous valence/arousal scores, which ensures a robust performance for face analysis. AffectNet database is a reliable source because of the large variability in the database, which makes it a benchmark dataset in its field.

Text analysis. Our approach here is based on a semi-automatically prepared resource that used automatic translation on a dictionary of English lemmas (Warriner, Kuperman &

Brysbaert, 2013), followed by manual correction of the entries (Aydın Oktay, et al., 2015). The affect scores are annotated in a five point scale (1-5), and our database contains 15,383 words and phrases in base form with VAD (Valence, Arousal, Dominance) annotations. These are complemented by a list of 72 Turkish words (adverbials, adjectivals, and nominals) that can intensify or diminish the affective attribute of a sentence and a list of 50 interjections. By pooling sentence-level affect scores, the approach is able to provide a single affect value (for valence or arousal) for each session.

Corpus-based affect values of words are not indicative of a specific context of usage, but represent average or most frequent usage context. Certain words that are frequently encountered in play therapy (such as 'father' or 'mother') have positive valence in most affective corpora, but in this context, are typically used in a neutral context. The preparation of a small, context-specific lexicon to override default VAD values improves text-based analysis results. In the experimental setup, we use a development set for preparing this lexicon, in order not to bias the results on the test set positively. The system performs sentence-level affect analysis by computing affect scores for smaller units first (i.e. words and phrases), and then by evaluation of the effect of modifiers and negation (Halfon et al., 2016).

Fusion. When the feature spaces are not similar in terms of dimensionality, combining classifier systems at decision (or score) level can be a better solution, preventing one modality from dominating the results. We fuse the two modalities at the decision level, because the raw feature spaces of video and text modalities are not similar to each other (pixel values vs. letters). The two modality-specific algorithms separately map their input to the valence/arousal space, and the regressors use these to predict affect scores of the children. Finally, for each session, we combine the predictions coming from different modalities by averaging them.

Procedures

Data collection and processing. All psychotherapy sessions were videotaped. One session was randomly chosen from sessions 1-10, 11-20, 21-30, 31-40, 41-50 in each psychotherapeutic process, and sessions from the later phases of treatment were added when

available, with a total of 148 sessions of 53 children. For CPTI ratings, each child's sessions were segmented and the longest play segments were coded by outside observers for affect dimensions. The coding of each session took about 1.5 h. The sessions were also automatically processed for facial and linguistic affect predictions. For text analysis, both the child's and therapist's speech in the sessions were manually transcribed by psychology students. Each transcript took about 3 h of transcription time.

Data analysis. When reporting the results, leave-one-subject-out cross validation is used, which is a technique where all but one participant's data is used for training and the excluded participant's data is used for testing. This is repeated for each participant, and the results are averaged. This approach is computationally expensive, but produces the most reliable estimate of the accuracy and is shown to avoid non-independence bias (Fazli et al., 2009; Esterman et al., 2010).

We approach the problem of predicting affect scores derived from CPTI affect dimensions as a regression problem. The face analysis deep neural network we use was trained with almost half a million annotated face images, which is required to train a generic deep neural network classifier. We use two regressors which require less annotated data, therefore more suitable for our dataset; Support Vector Regressor (SVR) and Extreme Learning Machine (ELM), respectively. To show the differences between these machine learning systems with the statistical regression methods, we also use Linear Regression and Polynomial Regression. An ELM regressor is a neural network with one hidden layer that uses matrix factorization to speed up the training by replacing backpropagation, which is a widely used but slower technique for training neural networks. In backpropagation, the error of the model is expressed mathematically, and the derivative of the error function is used to determine iterative parameter updates to the model. Conversely, ELM uses a formulation of the desired output as a multiplication of a feature matrix and a parameter matrix, and obtains the parameters by a single pseudo-inverse operation. Subsequently, it can be rapidly trained and produces comparable results to deep neural networks (Huang, 2006; 2012). A support vector machine

(SVM) (Suykens & Vandewalle, 1999) is a machine learning algorithm which finds a high-dimensional projection of the original data where different classes become linearly separable. A data point represented as a point in a d-dimensional feature space (i.e., represented by d feature values) is thus represented in a much higher dimensional feature space. The projection is called a kernel transformation, and typically, linear, polynomial, and Gaussian radial basis function (RBF) kernels are used. A toy example is the separation of values [-1(+), 0(-), 1(+)] in a one-dimensional problem, where +/- denote the class. While this problem is not linearly separable, in a 2-d space created by $f(x) = \{x, x^2\}$, it becomes linearly separable. The discriminating boundary (i.e., the line separating the classes) is projected back to the original feature space. A support vector regressor (SVR; Drucker et al., 1997) is a variant of SVM used for regression problems. Both approaches rely on identifying training samples close to the boundary, which are then called support vectors. In our experiments we used SVR with RBF and linear kernels[2].

In order to investigate the external validity of our automatic predictions, we investigate whether the children's age, gender, internalizing and externalizing problems predict the affect scores generated by our automatic system. Because our session level automated affect predictions were nested within children who were nested within therapists, we use Multilevel Modeling (MLM) in our external validity analyses.

Results

Descriptive Statistics

Table 1 shows the descriptive statistics of our best predictions and CPTI affect scores. Our prediction of each affect class mean has less than 5 % error compared to CPTI scores.

Evaluation Results for Different Modality Predictions

Face and text modalities of the children are used for predicting the four CPTI affect classes: Anger, Anxiety, Pleasure and Sadness. Table 2 shows the performance evaluations of

[2] The code developed for the project can be downloaded from the GitHub repository: https://github.com/dmetehan/Multimodal-Affect-Analysis-of-Psychodynamic-Play-Therapy

face modality, text modality and fusion of the two modalities. We use Pearson correlation coefficient (PCC) as our evaluation metric to assess associations between our affect predictions and CPTI affect dimensions. We use leave-one-child-out evaluation, where in each test the regressors are trained with data from 51 children and tested with one child from the test set, excluded from training. The five different regression approaches for mapping valence/arousal to CPTI values are individually tested, and show differences in performance depending on the affective dimension.

Comparing the PCC results across modalities in Table 2 and Figure 2a, overall the best performing results are achieved by SVR with RBF kernel (Anger), SVR with linear kernel (Sadness) and ELM (Anxiety, Pleasure). The significant correlations show small to medium effect. Then we compare the face and text modalities with each other by averaging the results of all the regressors for all CPTI affect classes. Figure 2b shows that from face modality the regressors only learned to predict pleasure. The fusion of the two modalities outperforms both modalities in predicting pleasure. Face modality performed poorly for all other affect classes, and because of that, the fusion approach could not outperform the text modality. Text modality seems to capture details to predict all the affect classes to some extent with small to medium effect. Overall, sadness predictions have the lowest effect size and are the hardest affect to predict. Finally, in Figure 2c, we show the average performance of each regressor. Linear regressors such as linear regression and SVR with linear kernel have lower performance than other regressors. Polynomial regression, SVR with RBF kernel, and ELM show more promising results.

External Validity Analyses

Our data sessions (N = 148) were nested within children (N = 53) who were nested within therapists (N = 24). Therefore, we used a multilevel modeling (MLM) approach with MLwin v3 (Rasbash et al., 2019). Multiple patients were treated by the same therapists, so we investigated the degree of interdependency. In our initial model, we estimated an empty multilevel model (i.e., no explanatory variables) predicting our automated affect predictions

to decompose the therapist-level (Level 3) and child-level (Level 2) and session-level (Level 1) variances for the purpose of computing the intraclass correlations (ICC). The therapist-level ICCs were 0.000 for anger and anxiety, which showed that therapists accounted for none of the variance in these affect dimensions. For sadness and pleasure, the therapist-level ICCs was 0.006 and 0.007 consecutively, indicating that therapists accounted for less than 1 % of the variance in these affect categories. Therefore, the variance was not attributable to differences between therapists. In contrast, the between-patient ICCs were 0.22 ($p < 0.01$) for anger, 0.30 ($p < 0.01$) for anxiety suggesting significant variance at the patient level. However, the patient-level ICCs for sadness and pleasure were 0.000 and 0.04 respectively, showing no significant variance at the patient level. We decided to conduct analyses with two-level models for all affect categories since not all variance was attributable to session-level for anger and anxiety.

Next, we tested multilevel models with maximum likelihood (MLM) estimation to analyze whether children's age, gender, internalizing and externalizing problems predicted our automated affect predictions. All predictors were grand-mean centered. The results are presented in Table 3. Our results showed that boys expressed significantly more anger and less pleasure than girls. Age and internalizing problems negatively predicted anger expression. There were no significant associations between age, gender, problem behaviors, anxiety and sadness.

Discussion

This is the first multimodal system in Turkish for automatic face and language-based affect analysis of children specifically adapted for use in psychodynamic play therapy sessions. For face analysis, we use a state-of-the-art deep learning based approach. For both modalities, machine learning models provide a flexibility in estimation that goes much beyond simpler linear models, as evidenced by our empirical results.

Our findings show that the fusion of face and text based affect analysis best predicts pleasure. For all other affect classes, text based affect analysis outperforms face based affect analysis and can predict significantly anger, anxiety and sadness. Overall, sadness predictions

have the lowest performance scores. The results illustrate that the automatic affect analysis is promising, however, needs further development. In particular, a setup with two static cameras is not sufficient to capture the child's face during play, and this limits the performance of face analysis. We provide specific suggestions for improvement in the next section.

From a clinical perspective, the outperformance of our system on anger and pleasure predictions may be related to our sample characteristics. These two classes of affect are easier to express for children with internalizing, externalizing and comorbid problems. Children with internalizing problems tend to avoid self-related negative emotions such as anxiety (Bizzi et al., 2018). Children with externalizing and comorbid problems may use aggressive affect to protect themselves from the dysregulation that comes from experiencing dysphoric feelings, such as sadness, anxiety and fear (Rice & Hoffman, 2014). Further, it has been suggested that some expressions of anger and aggression may serve to mask felt sadness or other dysphoric feelings (Cole & Zahn-Waxler, 1992). Therefore, it is possible that pleasure and anger are more easily observable with this cohort. Feelings of anxiety and sadness may be expressed more readily towards the later stages of treatment, when children develop a more intact emotion regulation capacity to tolerate these emotions (Hoffman et al., 2016).

Our external validity analyses show that boys expressed more anger and less pleasure than girls. This is consistent with meta-analytic reviews showing gender differences in emotion expression, with girls displaying greater levels of positive emotions than boys, particularly happiness, whereas boys expressing more externalizing emotions such as anger (Chaplin & Aldao, 2013). In our data, anger expression decreased as children became older. Younger children generally express more emotion than their older counterparts and on average mean levels of anger decrease after toddlerhood and into middle childhood possibly because children regulate their emotions better (Liu et al., 2018). Internalizing problems were inversely associated with anger expression. In the literature, anger is generally associated with externalizing problems, whereas more anxiety, fear and sadness are associated with internalizing problems (Eisenberg et al., 2001). Therefore, our external validity analyses

partially support the literature. We have not found significant associations with anxiety and sadness, which suggest the need to develop the sensitivity of our system on these dimensions.

Directions for Improvement

The small to medium effect sizes implicate the need to further develop the system, particularly on facial affect analyses. These results are partly caused by the recording conditions; two cameras are not enough to capture the children's faces for sufficiently many frames during play. The main challenge in the face modality was to capture the face from a frontal and clear view during play sessions with only a few static cameras. The performance can be improved by adding more resources, which will trade off cost vs. accuracy. We also note that a typical recording setup rarely uses more than two cameras. Fine-tuning the network on children's faces would definitely improve our system's performance but the lack of large and annotated children face databases does not allow that. This is mostly caused by privacy concerns of the parents, with fewer parents consenting to camera recordings. There is also the difficulty of annotating children's faces because, compared to adult faces, children's faces have much more expressivity and their expressions change much more rapidly. Moreover, affect analysis of their body motions can be considered as additional modalities to improve accuracy.

The main challenge in language-based analysis was the fact that the interaction was conducted in Turkish, for which analysis tools are still being developed. Since we did not have access to high quality voice sampling, we have not attempted a paralinguistic affect analysis on these data, but it is an additional modality that can be considered in such systems. For the application phase of our tool including a microphone and automatically transcribing the sessions would be an additional feature to consider (Wu et al., 2019).

Since we have aggregated within-session data to produce a single affect prediction per session (per dimension), temporal dynamics were disregarded. It may be argued that temporal analysis from the facial expressions would require data with less gaps, but since the session transcriptions are complete, temporal dynamics can be studied from the text. The language analysis, on the other hand, could be improved by using annotated domain-specific data to train

classifiers. Also, improved affective corpus studies will directly have an impact. However, having a single valence and arousal score for each word is a highly simplified representation. Ideally, each word will have context-dependent sets of scores, but extensive research is needed to establish such representations of text-based affect.

Directions for Future Use

After implementing the suggestions above and improving the precision and explainability of the system, it is possible to use these tools in different applications. Future research can assess emotion regulation, which refers to processes that amplify, maintain or decrease the intensity of emotions (Gratz, Weiss, & Tull, 2015). There is evidence showing that therapists' interventions help children regulate their emotions in psychodynamic child psychotherapy (Halfon & Bulut, 2019; Halfon et al., 2017; Hoffman et al., 2016). With the tools introduced in the paper, it may be possible to map the differences in the children's arousal levels after therapist's interventions.

There is also a wealth of research on this area in close relationships exploring how dyads help each other manage their emotional experiences (e.g., parent-infant dyads, Feldman, 2003; romantic relationships, Butner et al., 2007; social psychology Butler & Randall, 2013). This process, also known as emotion co-regulation, refers to the idea that regulation is a dyadic process not only determined by an individual's own internal emotional state, but also by the emotional states of other people with whom the individual is interacting. This sort of dyadic regulation can happen through a wide range of modalities such as body movement, facial expressions, eye movements, physiological signals, paraverbal behavior, linguistic style and others and is described in terms of synchrony (Kleinbub, 2017). Non-verbal indicators that assess the coupling of therapist's and patient's emotional and bodily responses have been studied in adult psychotherapy, such as coupled patterns in vocal pitch (Imel et al., 2014), head movements (Ramseyer & Tschacher, 2014), and whole body movements (Ramseyer & Tschacher, 2014). However, these have not yet been investigated in psychodynamic child psychotherapy. Alternatively, it may be possible to monitor therapists' countertransference

reactions to the child's emotions, some of which the therapists may not be aware of. The moment-to-moment affective expressions between the therapist and the patient and their temporal dynamic unfolding, coordination as well as the bi-directionality of their affective exchanges would be obvious extensions of this work. Such an analysis of synchrony and rapport will increase the value of the system as a clinical tool.

The tools introduced in the paper can also be used to help clinicians compare affect characteristics of their patients and conduct more detailed investigations of their sessions. By getting affect expression scores of each session, and looking more closely at the intra-session variance in affect expression, it would be possible to detect sessions that differ in their affect distributions, which can allow the clinician to pick up patterns. The clinician can observe the therapists' and patients' valence and arousal scores in the course of the session. It might also be possible to compare whether patients follow an expected course of evolution in affect characteristics. For instance, for children with externalizing problems, one would expect them to start treatment with elevated anger, however as the treatment progresses, the children should be able to express more distressing affects such as anxiety and sadness that had initially been warded off by aggressive affect (Hoffman et al., 2016).

In conclusion, this is the first automatic facial and linguistic affect analysis tool adapted, developed, and combined in a system that predicts affect over longitudinal data in psychodynamic play therapy. Our results are promising and specific suggestions for improvement in future research and applications for use in future settings are provided.

References

Ablon, J. S., Levy, R. A., & Katzenstein, T. (2006). Beyond brand names of psychotherapy: Identifying empirically supported change processes. *Psychotherapy: Theory, Research, Practice, Training, 43(2)*, 216.

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2000). *Mental health practitioners' guide for the Achenbach System of Empirically Based Assessment (ASEBA).* Burlington, VT: University of Vermont, Department of Psychiatry.

Aydın Oktay, E., Balcı, K., & Salah, A. A. (2015). Automatic assessment of dimensional affective content in Turkish multi-party chat messages. In *Proceedings of the International Workshop on Emotion Representations and Modelling for Companion Technologies* (pp. 19-24). ACM.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. In *Proceedings of 13th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 59-66).

Bizzi, F., Ensink, K., Borelli, J., Charpentier-Mora, S., & Cavanna, D. (2019). Attachment and reflective functioning in children with somatic symptom disorders and disruptive behavior disorders. European Child and Adolescent Psychiatry, 28(5), 705–717.

Butler, E. A., & Randall, A. K. (2013). Emotional coregulation in close relationships. *Emotion Review*, 5(2), 202-210.

Butner, J., Diamond, L. M., & Hicks, A. M. (2007). Attachment style and two forms of affect coregulation between romantic partners. *Personal Relationships*, 14(3), 431-455.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18-37.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7291-7299).

Chari, U., Hirisave, U., & Appaji, L. (2013). Exploring play therapy in pediatric oncology: a preliminary endeavour. *The Indian Journal of Pediatrics*, *80*(4), 303-308.

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin,* 139(4), 735.

Chazan, S. E. (2000). Using the children's play therapy instrument (CPTI) to measure the development of play in simultaneous treatment: A case study. *Infant Mental Health Journal, 21*(3), 211–221.

Chazan, S. E. (2001). Toward a nonverbal syntax of play therapy. *Psychoanalytic Inquiry, 21*(3), 394–406.

Chazan, S. E., & Wolf, J. (2002). Using the children's play therapy instrument to measure change in psychotherapy: The conflicted player. *Journal of Infant, Child, and Adolescent Psychotherapy, 2*(3), 73–102.

Cohen, E., Chazan, S., Lerner, M., & Maimon, E. (2010). Posttraumatic play in young children exposed to terrorism: An empirical study. *Infant Mental Health Journal, 31*(2), 159– 181.

Cole, P. M., & Zahn-Waxler, C. (1992). Emotional dysregulation in disruptive behavior disorders. In *D. Cicchetti & S. L. Toth (Eds.), Rochester Symposium on Developmental Psychopathology: Vol. 4. Developmental perspectives on depression* (pp. 173-210). Rochester, W: Univeisity of Rochester Press.

D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*, *47*(3), 43.

Diener, M. J., Hilsenroth, M. J., & Weinberger, J. (2007). Therapist affect focus and patient outcomes in psychodynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry, 164(6)*, 936-941.

Doyran, M., Türkmen, B., Oktay, E. A., Halfon, S., & Salah, A. A. (2019). Video and Text-Based Affect Analysis of Children in Play Therapy. *2019 International Conference on Multimodal Interaction* (pp. 26-34).

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. *In Advances in neural information processing systems*

Dunn, J., & Hughes, C. (2001). "I got some swords and you're dead!": Violent fantasy, antisocial behavior, friendship, and moral sensibility in young children. *Child development, 72*(2), 491-505.

Eisenberg, N., Sadovsky, A., Spinrad, T., Fabes, R., Losoya, S., Valiente, C., . . . Shepard, S. (2005). The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity : Concurrent relations and prediction of change. *Developmental Psychology, 41*(1), 193-211.

Ekman P, Friesen WV, Hager JC. *Facial Action Coding System: The Manual on CD Rom.* Salt Lake City, UT: A Human Face, 2002.

Erol, N., & Simsek, Z (2000). Mental health of Turkish children: Behavioral and emotional problems reported by parents, teachers and adolescents. In N. N. Singh, J. P. Leung, & A. N. Singh (Eds.), *International perspectives on child and adolescent mental health* (pp. 223–247). Oxford: Elsevier Science.

Ekman, P. (1994). All emotions are basic. *In: Ekman P and Davidson RJ (eds) The nature of emotion: Fundamental questions.* New York: Oxford University Press.

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y. C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage*, *50*(2), 572-576.

Fazli, S., Popescu, F., Danóczy, M., Blankertz, B., Müller, K. R., & Grozea, C. (2009). Subject-independent mental state classification in single trials. *Neural networks*, *22*(9), 1305-1312.

Feldman, R. (2003). Infant–mother and infant–father synchrony: The coregulation of positive arousal. *Infant Mental Health Journal*, 24(1), 1-23.

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self.* New York: Other Press.

Gratz, K. L., Weiss, N. H., & Tull, M. T. (2015). Examining emotion regulation as an outcome, mechanism, or target of psychological treatments. *Current Opinion in Psychology*, 3, 85-90.

Greenberg, L. S., & Pascual-Leone, A. (2006). Emotion in psychotherapy: A practice-friendly research review. *Journal of Clinical Psychology*, 62(5), 611-630.

Gunes, H., & Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions (IJSE)*, 1(1), 68-99.

Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing,* 31(2), 120-136.

Halfon, S. (2017). Play profile constructions: An empirical assessment of children's play in psychodynamic play therapy. *Journal of Infant, Child, and Adolescent Psychotherapy, 16*(3), 219–233.

Halfon, S. (2020). Psychodynamic technique and therapeutic alliance in prediction of outcome in psychodynamic child psychotherapy. Manuscript submitted for publication.

Halfon, S., Bekar, O., & Gürleyen, B. (2017). An empirical analysis of mental state talk and affect regulation in two single-cases of psychodynamic child therapy. *Psychotherapy, 54*(2), 207–219.

Halfon, S., & Bulut, P. (2019). Mentalization and the growth of symbolic play and affect regulation in psychodynamic therapy for children with behavioral problems. *Psychotherapy Research, 2*(2), 1-13.

Halfon, S., Aydın Oktay, E., & Salah, A. A. (2016). Assessing affective dimensions of play in psychodynamic child psychotherapy via text analysis. In *International workshop on human behavior understanding (*pp. 15–34). Amsterdam: Springer International Publishing.

Halfon, S., Yılmaz, M., & Çavdar, A. (2019). Mentalization, session-to-session negative emotion expression, symbolic play, and affect regulation in psychodynamic child psychotherapy. *Psychotherapy (Chicago, Ill.)*.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics),* 28(1), 100-108.

Hoffman, L., Rice, T., & Prout, T. (2016). *Manual of Regulation-Focused Psychotherapy for Children (RFP-C) with externalizing behaviors: A psychodynamic approach*. New York, NY: Routledge/Taylor & Francis Group.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.

Huang, G., Zhou, H., Ding, X. & Zhang, R (2012). Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513-529.

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology*, 61(1), 146.

Jaiswal, S., & Valstar, M. (2016). Deep learning the dynamic appearance and shape of facial action units. In *Proc. of IEEE Winter Conf. Applications of Computer Vision* (pp. 1-8).

Jones, E. E., & Pulos, S. M. (1993). Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology*, 61(2), 306.

Khan, R.A., A. Crenn, A. Meyer & S. Bouakaz. (2019) A novel database of children's spontaneous facial expressions (LIRIS-CSE). *Image and Vision Computing,* 83, 63-69.

Kaya, H., Gürpınar, F., & Salah, A. A. (2017a). Video-based emotion recognition in the wild using deep transfer learning and score fusion. *Image and Vision Computing,* 65, 66-75.

Kaya, H., Salah, A. A., Karpov, A., Frolova, O., Grigorev, A., & Lyakso, E. (2017b). Emotion, age, and gender classification in children's speech by humans and machines. *Computer Speech & Language*, 46, 268-283.

Kernberg, P. F., & Chazan, S. E. (1991). *Children with conduct disorders: A psychotherapy manual*. New York, NY: Basic Books.

Kernberg, P. F., Chazan, S. E., & Normandin, L. (1998). The children's play therapy instrument (CPTI): description, development, and reliability studies. *The Journal of Psychotherapy Practice and Research, 7*(3), 196–207.

Kleinbub, J. R. (2017). State of the art of interpersonal physiology in psychotherapy: a systematic review. *Frontiers in Psychology*, 8, 2053.

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing,* 4(1), 15-33.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, *38*.

Larochette, A. C., Chambers, C. T., & Craig, K. D. (2006). Genuine, suppressed and faked facial expressions of pain in children. *Pain*, 126(1-3), 64-71.

Lengua, L. J. (2003). Associations among emotionality, self-regulation, adjustment problems, and positive adjustment in middle childhood. *Journal of Applied Developmental Psychology*, 24(5), 595-618.

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing,* in press.

Littlewort, G. C., Bartlett, M. S., Salamanca, L. P., & Reilly, J. (2011, March). Automated measurement of children's facial expressions during problem solving tasks. In *Proceedings of 9th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 30-35).

Liu, C., Moore, G. A., Beekman, C., Pérez-Edgar, K. E., Leve, L. D., Shaw, D. S., ... & Neiderhiser, J. M. (2018). Developmental patterns of anger from infancy to middle childhood predict problem behaviors at age 8. *Developmental Psychology*, 54(11), 2090.

Lous, A. M., De Wit, C. A., De Bruyn, E. E., & Marianne Riksen- Walraven, J. (2002). Depression markers in young children's play: A comparison between depressed and nondepressed 3-to 6-year-olds in various play situations. *Journal of Child Psychology and Psychiatry, 43*(8), 1029–1038.

Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive/ expressive treatment.* New York: Basic Books.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 94-101).

Malan, D. (2001). *Individual psychotherapy and the science of psychotherapy (2nd ed.)*. London, United Kingdom: Butterworths.

Mehrabian, A. (1970). A semantic space for nonverbal behavior. *Journal of Consulting and Clinical Psychology*, *35*(2), 248-257.

Meinhold, R. J., & Singpurwalla, N. D. (1983). Understanding the Kalman filter. *The American Statistician,* 37(2), 123-127.

Midgley, N., Ensink, K., Lindqvist, K., Malberg, N., & Muller, N. (2017). *Mentalization-based treatment for children: A time-limited approach*. American Psychological Association.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion measurement* (pp. 201-237).

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *arXiv preprint* arXiv:1708.03985.

Munezero, M. D., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Transactions on Affective Computing,* 5(2), 101-111.

Oflazer, K., & Saraçlar, M. (Eds.). (2018). *Turkish Natural Language Processing*. Springer.

Oldehinkel, A., Hartman, C., Winter, A., Veenstra, R., & Ormel, J. (2004). Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Development and Psychopathology, 16*(2), 421-440.

Picard, R. W. (1995). *Affective computing.* MIT Press.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

Ramseyer, F., & Tschacher, W. (2014). Nonverbal synchrony of head-and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in Psychology*, 5, 979.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). *A user's guide to MLwiN, v2. 10 Bristol: Centre for Multilevel Modelling*. University of Bristol.

Rice, T. R., & Hoffman, L. (2014). Defense mechanisms and implicit emotion regulation: a comparison of a psychodynamic construct with one from contemporary neuroscience. *Journal of the American Psychoanalytic Association*, 62(4), 693-708.

Russ, S. W., & Niec, L. N. (2011). *Play in clinical practice: Evidence-based approaches.* New York, NY Guilford Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Scott, T. J. L., Short, E. J., Singer, L. T., Russ, S. W., & Minnes, S. (2006). Psychometric properties of the Dominic interactive assessment: a computerized self-report for children. *Assessment*, 13(1), 16-26.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242-264). IGI global.

Verheugt-Pleiter, A. J. E., Zevalkink, J., & Schmeets, M. G. C. (2008). *Mentalizing in child therapy*. London: Karnac.

Von Klitzing, K., Kelsay, K., Emde, R. N., Robinson, J., & Schmitz, S. (2000). Gender-specific characteristics of 5-year-olds' play narratives and associations with behavior ratings. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(8), 1017-1023.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191-1207.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS one*, 12(5), e0177239.

Wu, F., García-PLP, P. D., & Khudanpur, S. (2019). Advances in automatic speech recognition for child speech using factored time delay neural network. In *Proceedings of Interspeech* (pp. 1-5).

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).

Zeinstra, G. G., Koelen, M. A., Colindres, D., Kok, F. J., & De Graaf, C. (2009). Facial expressions in school-aged children are a good indicator of ‘dislikes’, but not of ‘likes’. *Food Quality and Preference*, 20(8), 620-624.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence,* 31(1), 39-58.

Table 1.

Mean and Standard Deviation Comparisons of Automated Predictions and CPTI Affect Scores

	Anger	Anxiety	Pleasure	Sadness
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
CPTI Affect Scores	2.71 (1.60)	2.01 (1.49)	2.71 (1.13)	1.73 (1.29)
Automated Predictions	2.84 (0.91)	1.91 (1.07)	2.71 (0.51)	1.68 (0.69)

Note. CPTI = Children's Play Therapy Instrument

Table 2.

Correlations between Automated Affect Predictions using Different Regression Approaches and CPTI Affect Dimensions

Automated Affect Predictions		CPTI Affect Dimensions											
		Anger			Anxiety			Pleasure			Sadness		
		r	95 % CI		*r*	95 % CI		*r*	95 % CI		*r*	95 % CI	
Face	Linear Regression	0.063	-0.100	0.222	-0.079	-0.237	0.084	0.339**	0.188	0.475	0.061	-0.102	0.220
	Polynomial Regression	-0.004	-0.165	0.158	0.032	-0.130	0.193	0.276**	0.119	0.419	0.134	-0.028	0.289
	SVR (Linear)	0.166*	0.005	0.319	-0.213*	-0.362	-0.053	0.311**	0.158	0.450	0.052	-0.111	0.212
	SVR (RBF)	0.105	-0.058	0.262	-0.028	-0.189	0.134	0.243**	0.085	0.389	0.025	-0.137	0.186
	ELM	0.109	-0.054	-0.265	0.031	-0.131	0.192	0.340**	0.189	0.475	0.118	-0.044	0.275
Text	Linear Regression	0.265**	0.108	0.409	0.188*	0.027	0.339	0.267**	0.111	0.411	0.219**	0.060	0.368
	Polynomial Regression	0.352**	0.202	0.486	0.319**	0.166	0.457	0.374**	0.226	0.505	0.152	-0.010	0.306
	SVR (Linear)	0.216**	0.056	0.364	0.279**	0.123	0.422	0.183*	0.022	0.335	0.262**	0.105	0.406
	SVR (RBF)	0.536**	0.410	0.642	0.309**	0.155	0.448	0.264**	0.107	0.408	0.201*	0.041	0.351
	ELM	0.294**	0.139	0.435	0.386**	0.239	0.515	0.308**	0.154	0.447	0.192*	0.031	0.343
Fusion	Linear Regression	0.236**	0.078	0.383	0.089	-0.074	0.247	0.385**	0.238	0.515	0.193*	0.032	0.343
	Polynomial Regression	0.234**	0.075	0.381	0.235**	0.076	0.382	0.417**	0.274	0.542	0.181*	0.020	0.333
	SVR (Linear)	0.256**	0.099	0.401	0.116	-0.046	0.273	0.349**	0.198	0.483	0.211**	0.051	0.360
	SVR (RBF)	0.520**	0.391	0.629	0.245**	0.087	0.391	0.344**	0.194	0.479	0.185*	0.024	0.336
	ELM	0.310**	0.157	0.449	0.300**	0.146	0.441	0.430**	0.289	0.553	0.227**	0.068	0.375

Notes. CPTI = Children's Play Therapy Instrument, SVR = Support Vector Regressor, RBF = Radial Basis Function, ELM = Extreme Learning Machine

**p < 0.05; **p < 0.01.*

Table 3.
Multilevel Model in Prediction of Automated Affect Scores

	Automated Affect Predictions															
	Sadness				Anxiety				Pleasure				Anger			
Baseline Characteristics	*B*	SE	95 % CI		*B*	SE	95 % CI		*B*	SE	95 % CI		*B*	SE	95 % CI	
Intercept	1.681	0.055	1.573	1.790	1.901	0.103	1.699	2.103	2.706	0.042	2.624	2.788	2.839	0.070	2.702	2.976
Age	0.027	0.030	-0.033	0.086	0.017	0.054	-0.090	0.123	0.013	0.023	-0.032	0.058	-0.090*	0.038	-0.164	-0.016
Sex	-0.159	0.122	-0.397	0.080	0.337	0.224	-0.102	0.777	-0.187*	0.092	-0.367	-0.007	0.542*	0.152	0.244	0.841
Internalizing problems	0.006	0.007	-0.007	0.019	-0.002	0.012	-0.026	0.022	0.002	0.005	-0.008	0.012	-0.018*	0.008	-0.035	-0.002
Externalizing problems	-0.000	0.007	-0.013	0.013	-0.016	0.013	-0.041	0.008	-0.003	0.005	-0.013	0.007	-0.005	0.009	-0.021	0.012

Notes. Sex was dummy coded as 0 = female, 1 = male.

**p < 0.05; **p < 0.01.*

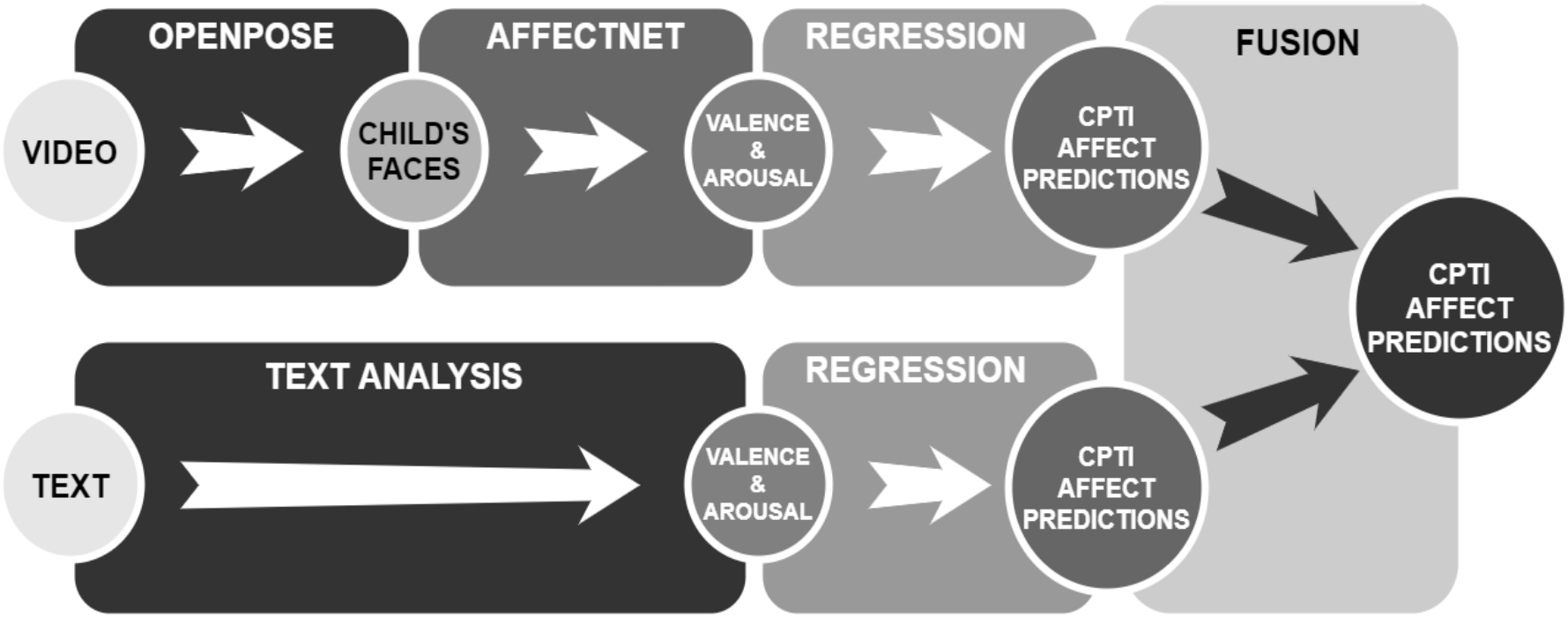
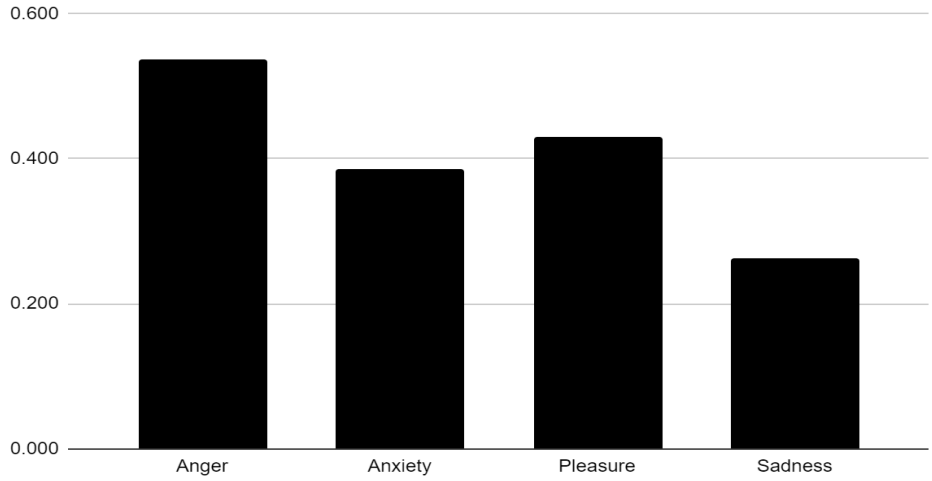
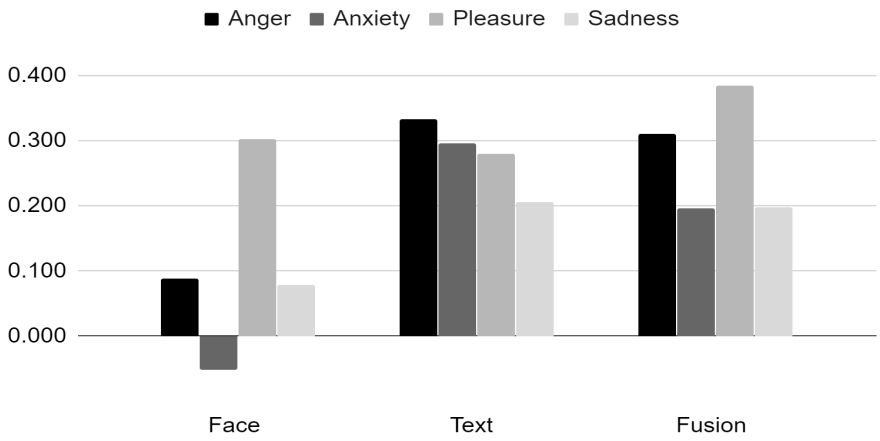


Figure 1.
Pipeline of the system
Notes: CPTI = Children's Play Therapy Instrument

A



B



C

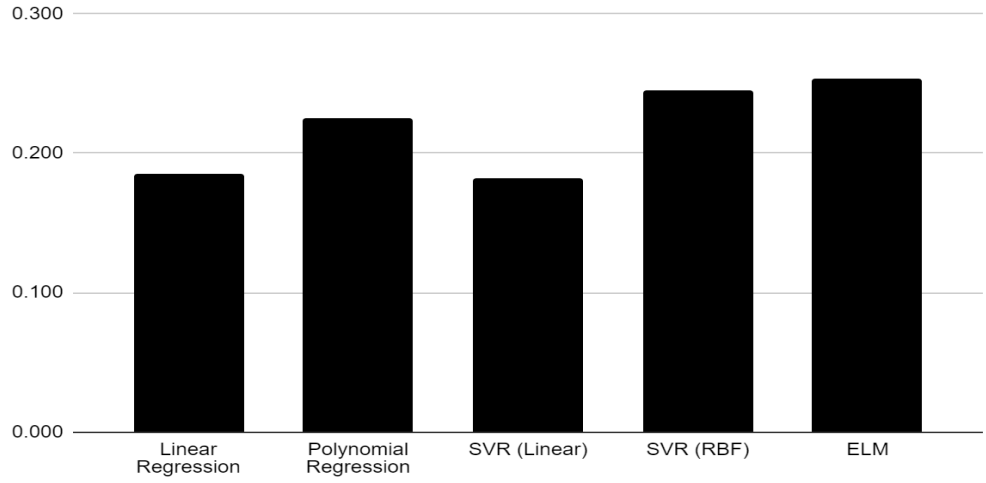


Figure 2.

Performance comparison for A: highest correlations between automated affect predictions and CPTI affect classes, B: average correlations across modalities, C: average correlations of regressors.

Notes. SVR = Support Vector Regressor, RBF = Radial Basis Function, ELM = Extreme Learning Machine