Multiple knowledge categorising behavioural states and communication attempts in people with profound intellectual and multiple disabilities. [*]

Matej Cigale
Jožef Stefan Institute,
Jožef Stefan IPS
Jamova cesta 39, 1000 Ljubljana
matej.cigale@ijs.si

Mitja Luštrek
Jožef Stefan Institute,
Jožef Stefan IPS
Jamova cesta 39, 1000 Ljubljana
mitja.lustrek@ijs.si

Abstract

People with profound an intellectual disabilities are not often discussed in our society. They have a hard time in their daily lives, to a large extent unable to fulfil their basic needs without assistance and mostly unable to communicate their desires to the outside world. The INSENSION project uses computer vision, to extract gestures and facial expressions from multiple cameras in the room, capturing the movement of people with PIMD. It combines that with other noninvasive sensors to build a coherent picture of the environment in order to provide more independence and ease the work of the caregivers. This paper presents specialised Machine Learning algorithms that attempt to classify the behavioural states and communication attempts of people with PIMD based on annotations of non-verbal signals(NVS) and expert knowledge provided by their caregivers. Two methods are presented. First is based on the idea of unique NVS that classifies the behavioural state e.g. a smile in happy healthy individuals. The second builds on the Arousal-Valence model to generate a value for valiance based on a group of NVS.

1 Introduction

People with profound intellectual and multiple disabilities (PIMD), face extreme difficulties in everyday life. They are a heterogeneous group, suffering from different ailments and conditions. Severe cognitive, motor and sensory disabilities make this population reliant on outside care for most daily tasks. While these individuals are exactly the ones that would benefit most from intelligent systems in their vicinity, they are unable to use them due to relative high complexity. The INSENSION project aims to develop a system that will observe behavioural state and non-symbolic communication attempts of people with PIMD and interpret them to people in the vicinity in order to allow them to render assistance or support if needed; and even automatically control their environment using external services.

[*] This work was supported by the European Union's Horizon 2020 research and innovation program [grant agreement No 780819 (INSENSION)

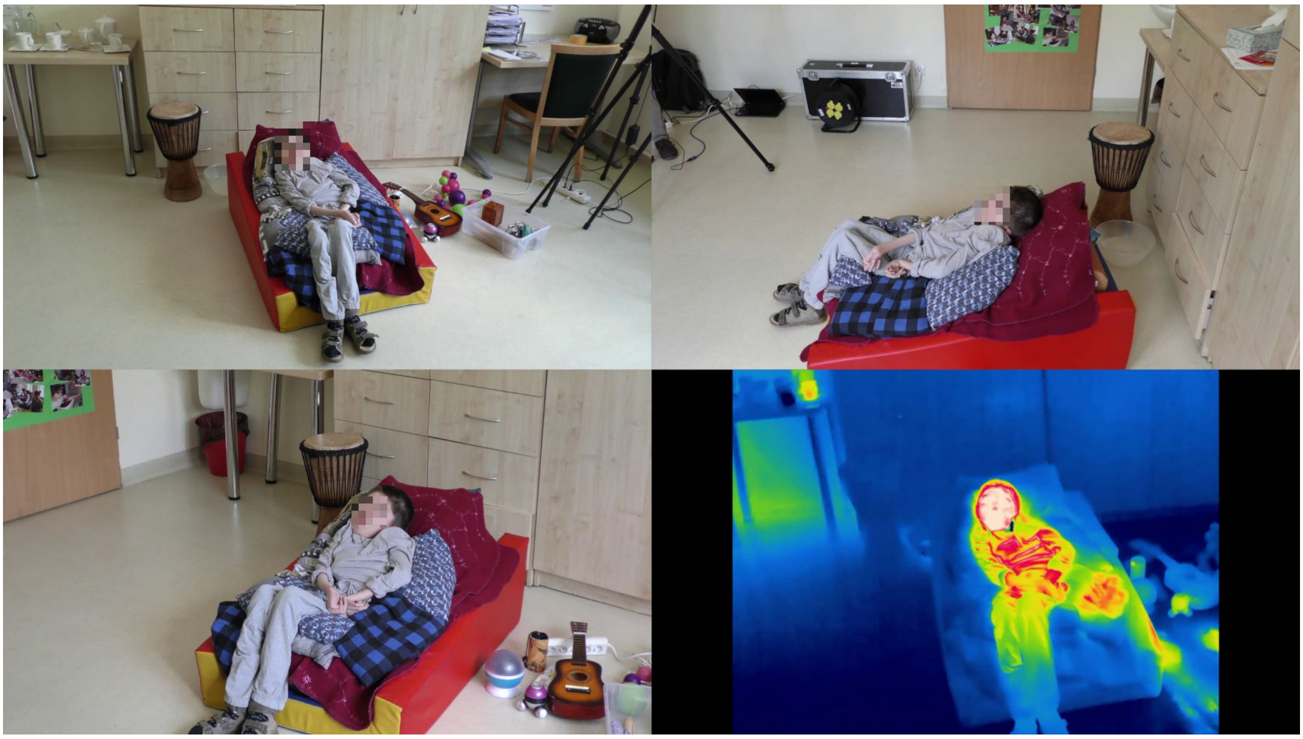


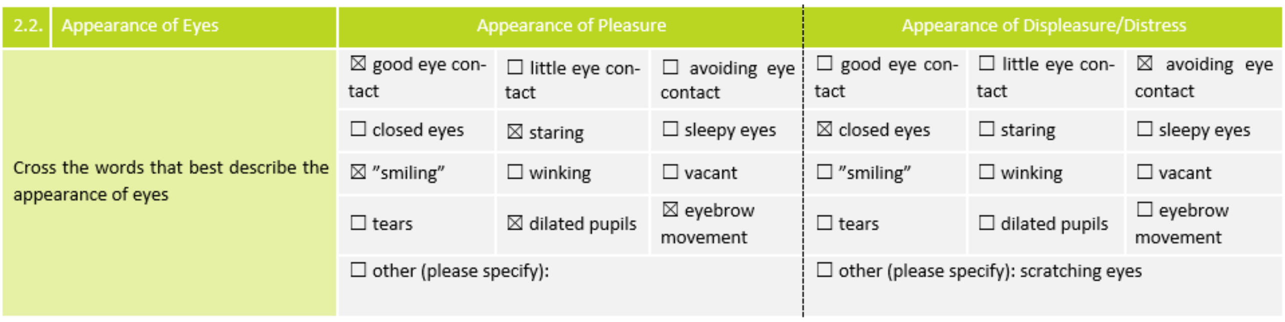
Figure 1: An example of the videos recorded. Computer vision algorithms are ran on all the streams and the results collated based on probability.

The first step is to recognise Non-Verbal Signals (NVS) expressed by people with PIMD (e.g., certain gestures [CSWS17] and facial expressions [SLQR18]) primarily using video to capture these movements (Figure 1). The extracted NVS are then used to determine behavioural states and communication attempts. Important features of their environment (e.g., presence of a caregiver and objects, temperature) are used to provide recommendations to a caregiver or control external services. This paper deals with the interpretation of NVS once they are recognised by machine vision systems.

Each individual is unique with different abilities and signals. This holds doubly so with people with PIMD is as, due to their condition there are no general signals that could be associated with behavioural states and communication attempts. General-purpose system are mot feasible and personalised classification methods must be used. Collecting a large set of annotated data for each individual is unpractical, as it would make the setup of the system prohibitively time consuming. Additionally, mappings between certain NVS and behavioural states are known to those close to the person, and this expert knowledge should be used in the decision making process. Taking this into account, we developed two machine-learning (ML) algorithms that are designed to work with a limited amount of data and can incorporate expert knowledge. They are presented in the paper and evaluated on a preliminary collection of data.

To compound the problem the limited vocabulary of people with PIMD means that their communication is heavily reliant on interpretation and usually cannot be interpreted directly as statements. To this end a database of context will be build that will take the environment into account. This would in turn allow control of light to suit the mood of the user, the music that is being played and the availability of certain toys or tools in the environment. The system takes this into account when it is dealing with the recommendations and can take additional clues form the knowledge base to suit the needs of the users.

This paper is organised as follows. In Section 2 we present the state of the art on the subject. Section 3 discusses the data collected. Section 4 presents the two ML algorithms: the Unique Non Verbal Signals model optimised for extremely small data sets, and the Valence model that works better with limited but somewhat larger data sets. Section 5 draws the conclusions based on this paper.



2.2. Appearance of Eyes	Appearance of Pleasure			Appearance of Displeasure/Distress		
Cross the words that best describe the appearance of eyes	☒ good eye contact	☐ little eye contact	☐ avoiding eye contact	☐ good eye contact	☐ little eye contact	☒ avoiding eye contact
	☐ closed eyes	☒ staring	☐ sleepy eyes	☒ closed eyes	☐ staring	☐ sleepy eyes
	☒ ”smiling”	☐ winking	☐ vacant	☐ ”smiling”	☐ winking	☐ vacant
	☐ tears	☒ dilated pupils	☒ eyebrow movement	☐ tears	☐ dilated pupils	☐ eyebrow movement
	☐ other (please specify):			☐ other (please specify): scratching eyes		

Figure 2: An example of questionnaire on eye movements.

2 State of the art

Since people with PIMD have limited ability to express their desires and intentions, it is these ambiguous feelings that are key to understanding them. Consequently, recognising them is the focus of this section. One of the approaches that seems most objective is the EEG signal [YWL+17]. Even with this signal, we need to rely on advances in machine learning to extract the information about the feelings based on the brain activity measurements. State of the art methods for extracting feelings from EEG are done on DEAP dataset and predict arousal and valence with an accuracy of 74.65% and 78% [CK18, KV18, LLS17]. Arousal and valence are the standard metrics that are used to map human feelings onto a 2D plane. Arousal can be understood as the strength of a feeling, while Valence is the positive or negative connotation of the feeling. There are several ways to map discrete feelings to this 2D space and the actual mapping is not agreed upon leaving to some ambiguity on this subject, but it is at this point one of the standard models [SBS+18]. A step towards understanding feelings closest to what INSENSION will use (from video and audio) was done by Metallinou et al. [MKN13]. They also use a different space, which also includes the dominance dimension. They used USC CreativeIT database consisting of acted-out scenes. Behoora et al. [BT15] tackled a similar problem, focusing on real-life setting with designers in a team. They used an infrared imaging sensor (i.e., Microsoft Kinect) to extract the body positions, velocity and acceleration of all the joints in the upper part of the body. The recorded scenario assumed sitting down. This information was then used to train several different machine-learning models of the actions. (C4.5, IBK, Random forest, Naive Bayes). They used a static table to map the resulting body language poses into feelings. The accuracy of detecting the poses is understandably quite high at around 99%. This is not surprising, as the infrared imaging for pose acquisition is a robust approach.

Another part of the project is identifying and using context of the interaction to infer the meaning. As context we understand data from IoT devices that are present in the vicinity, objects and people that are detected by the cameras. YOLO [Red16] is one such detector that is considered for the detection of objects, mostly because of its reported low overhead. For people detection Zoom-RNN [AAA+18] seems to be a good fit, but other solutions are investigated.

When it comes to extracting the context of the interaction there are several approaches that produce interesting results. Probabilistic Event Calculus [SPA14] is one of the approaches that can be used and extended to the case at hand. In the cited work the cited work the authors propose a method (MLN–EC) that deals with uncertainty in the detected environment to classify the event that is occurring. They mostly deal with the movement of people in the system, but the method is flexible enough that it could be adapted to our needs.

Another field that can be used to deal with broader context of events is Case-Based Reasoning (CBR) [LJPM17]. In this paradigm knowledge is represented as a set of cases - events that happened and the solutions that were used to solve the problem. Events that are detected are conformed into the closest case that is stored in the database and the solution of the problem is used. The solution is then evaluated and stored in the system based on the success of the solution. CBR seems a promising candidate, but there are, at this time no solutions that are capable of using expert knowledge, so it does not meet our desired criteria.

3 The data used

Our project currently works with six people with PIMD. Expert knowledge was collected from their caregivers of the people with PIMD in the form of an extensive questionnaire (see Fig. 2 for an excerpt). This data was then incorporated into the behavioural state recognition to improve decisions.

Visual data was collected in the facilities where the people with PIMD are cared for, and took the form of multiple-angle recordings with normal and heat-vision cameras (Figure 1).

Parts of these videos were then annotated using the ELAN [SW08] software. In the first step, all NVS had to be annotated by hand, and this is the data that we used to train our behavioural state classifiers. NVS recognisers will be developed to recognise NVS directly form video using computer vision, based on the same annotations (Figure 3a). Annotators were asked to annotate pre-defined facial expressions, gestures, vocalisations, presence of caregivers, and to note any special cases including elemental factors, such as music or light, that might play a role in the behavioural state of the subjects.

In addition to NVS, behavioural states of *pleasure* and *displeasure* were annotated, while *neutral or undefined* state was assumed to be any state that was not specifically marked. This simplification was used since recognising more subtle behavioural states of people with PIMD is extremely hard. A second category of communication attempts, *(comment, demand, protest)* was also annotated, but is not discussed here, as the recognition works in a similar manner.

In our experiments we take the annotations of behavioural states as ground truth. This is somewhat debatable, as there is in fact no way to know what the people with PIMD are actually experiencing at any given time, and there is no way for them to actually explain it to us. Nevertheless we feel that people tasked with annotation were familiar enough with their subjects so that they could render as accurate picture of their behavioural state as is possible [VDCP$^+$10].

4 Behaviour state recognition

The core of the system is developed in Prolog. Our method assumes that the person with PIMD has distinct NVS that correlate to his internal behavioural states. Each of the detected signals can have a meaning, but in people with PIMD that is not a guaranty. The NVS can have no meaning or the same NVS is used to convey multiple dissimilar meanings ie. the person with PIMD could clap to signify happiness, but also to signify sadness. These signals do not necessary follow social conventions, for instance lifting the corners of the mouth up can signify pain not pleasure as in normative individuals.

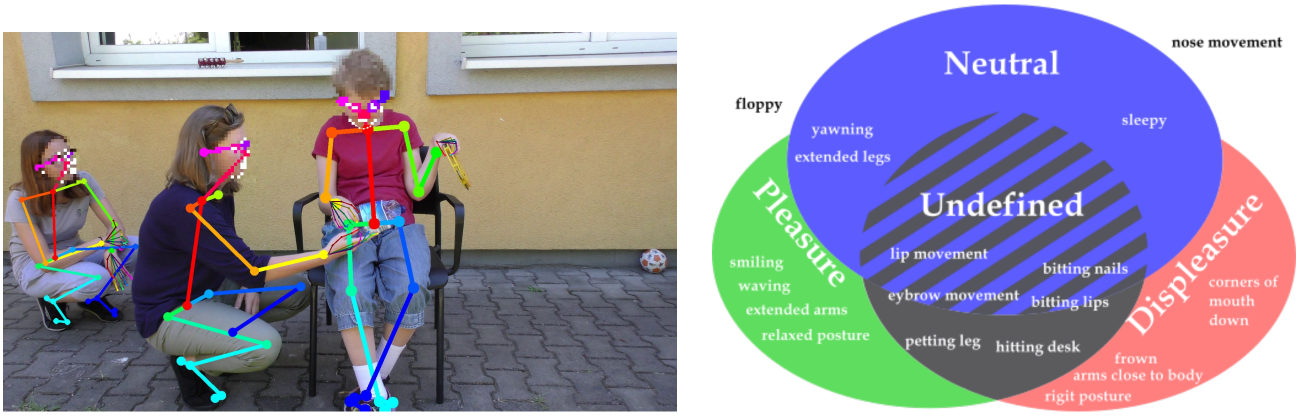
The code in the listings that follow is simplified in order to help understanding and not bog down the user with details.

4.1 The Unique Non Verbal Signals method

The Unique Non Verbal Signals Model makes it's decision based on the idea that there exists a NVS that will signify pleasure, but will never be used to signify any other behavioural state. In Figure 3b, the green represents all the NVS that correspond only to pleasure, red are those that would correspond solely to displeasure and blue are the those that would represent the neutral state. The grey represent NVS that cannot be robustly categorised. They could have no meaning or the meaning simply cannot be extracted at this time. There are also some NVS that are not part of the annotated dataset, as they did not occur in the particular individual. They are shown on the outside of the set.

In order for us to robustly detect pleasure we must remove all NVS, that are associated with displeasure or neutral state. This leaves us with a set of NVS that uniquely represent the behavioural state of pleasure. Additionally if experts annotated that a certain NVS corresponds with a behavioural state we must take that into account.

Deciding on the behavioural state based on the NVS is simple (Listing 1). We check if there are any NVS that are specific to pleasure, either from the expert knowledge or from the annotated examples. The term *assessment('Pleasure', NVS)* will return a NVS that was annotated as denoting pleasure by the experts. The second part of *pleasure_marker* term will check if there is a NVS in the annotations associated with pleasure and not with displeasure or neutral state. The term *window(Interval, NVS, _)* unifies for any three second window that is annotated with a given NVS or other annotation. If either of these rules holds true we identify the behavioural state as pleasure. This system works surprisingly well, with the limited dataset available, as discussed previously. But it is expected to become less viable with more data. The results of this can be seen in Figure 4a. Due to uniqueness of the people with PIMD a model is trained for each individual. The dataset for each individual is small, consisting of less than 10 annotated examples of each behavioural state of various lengths. In order to verify the results of our method we trained the model on all the examples baring one for each state and compared it to the Ground truth - as based on annotations. The example here is a contiguous set of



(a) NVS detection

(b) The visualisation of the NVS set interactions.

Figure 3: NVS detection and interpretation.

```
decide_state(Interval, 'Pleasure') :-
    pleasure_marker(Pleasure),
    window(Interval, NVS, Annotation),
    member(NVS, Pleasure).

pleasure_marker(NVS) :-
    assessment('Pleasure', NVS).
pleasure_marker(NVS) :-
    window(Interval, NVS, _),
    window(Interval, 'Pleasure', _),
    not(displeasure_marker(NVS)),
    not(neutral_marker(NVS)).
```

Listing 1: Querying the behavioural state.

windows that annotate pleasure, so the data is not cross contaminated. The accuracy is the correct classification of the behavioural state that was not used for training for each possible example of a behavioural state.

4.2 The Valence method

The second method treats the significance of the NVS as an indicator of behavioural state on a continuous interval. We assume that each NVS has a certain correlation with valence. In our case valence is a number that is correlated with the three behavioural states *(displeasure, neutral, pleasure)*, a simplified case of mapping feelings to an Arousal-Valence plane. Valence is assumed to be a value in [-1, 1] interval where displeasure is associated with negative and pleasure with positive numbers.

Listing 2 contains the pseudo code for valence calculation. If there is little or no correlation between pleasure and the expression it should gravitate towards negative values. Inverse must be true for displeasure. *correlation_set(NVS, Behavoural_state, Num_correlations)* returns the number of all annotated intervals that contain a NVS at the same time as the behavioural state. The *intervals(Behavioural_state, Num_examples)* returns the number of all annotated intervals of a certain behavioural state.

If we want to classify the behavioural state based on NVS we add the valence off all the NVS that are expressed. We determine the behavioural state based on the value of valence (Listing 3). The *calculate_valence* is a recursive function that sums the valence of a set of NVS, and returns 0 for an empty set.

The P_Cut and D_Cut variables determine the intervals of pleasure, displeasure or neutral behavioural state. We use Constraint Logic Programming to determine the optimal values for these values. At its core it is a minimisation problem where we try to find the thresholds for the intervals that produce the smallest classification error. It uses standard architecture for clpfd labeling from the SWI-Prolog library, adapting it to the problem.

The rationale for the system is as follows. We take all the windows in the annotations we have and attempt to find values where we cut the valence dimension so that our classification error is the smallest possible.

The function *behaviour_state(NVS_Set, Decision, Pleasure, Displeasure)* in Listing 3 can also be used in

```
valence(NVS, Valence) :-
    correlation_set(NVS, Pleasure, NVS_P),
    correlation_set(NVS, Displeasure, NVS_D),
    correlation_set(NVS, Neutral, NVS_N),
    intervals(Pleasure, P_Set),
    intervals(Displeasure, D_Set),
    intervals(Neutral, N_Set),
    Valence_direction is NVS_P/P_Set
                       - NVS_D/D_Set,
    Valence_strength is 1 + P_Set
                      + N_Set+D_Set,
    Valence is Valence_direction/Valence_strength.
```

Listing 2: The function that calculates the valence.

```
behaviour_state(NVS_Set, Decision, P_Cut, D_Cut) :-
    calculate_Valence(NVS_Set, Valence),
        (valence > P_Cut ->
            (Decision = Pleasure);
        (valence < D_Cut ->
            (Decision = Displeasure);
            (Decision = Neutral))).

calculate_valence([], 0).
calculate_valence([NVS | Rest], Valence) :-
    valence(NVS, V1),
    calculate_valence(Rest, V2),
    Valence is V1 + V2.
```

Listing 3: Determining the behavioural State.

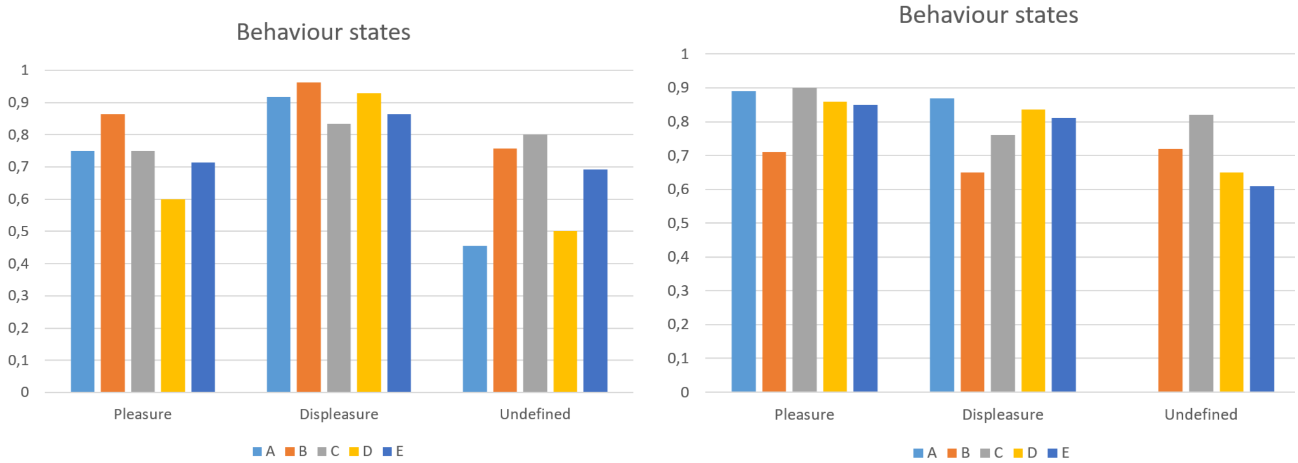
deciding future behavioural states if provided withe the optimised values for P_cut and D_cut. Note that these values are dependant on the function that is used.

Using the same methodology as before, this model performs worse then the somewhat naive Unique non Verbal Signals method, as seen in Figure 4b. Person A has very high miss classification of neutral state, owning to a small example size for this state. The Valence method seems to perform better for subjects with more annotations, perhaps indicating that it does not benefit as much from expert knowledge. Due to the algorithm the methods work on opposite spectrum, as Unique non Verbal Signals method with infinite data converges toward expert knowledge while Valence method diverges from it.

5 Conclusions

In this paper we presented two Machine Learning algorithms, specialised for learning behavioural states of people with PIMD. The advantage over the more common Machine Learning algorithms is the ability to incorporate prior knowledge from the assessments. This is important as detecting feelings is hard even when the subjects are healthy people who exhibit appropriate socially conditioned verbal and von-verbal signals. With the population of people with PIMD it is much harder as they exhibit little to no standardised expressions due to mental or physical disabilities. The cost of personalising behaviour state algorithms must be as small as possible since collecting the data to train them will require outside help. We presented two candidates for this system that give promising results.

The work at this time assumes that the detected NVS are robustly detected. As the work progresses the recognisers will return the probability of the detected NVS. This will add another level of complexity to the system. While the Valence method is easily adapted to the non deterministic nature of the recognisers the Unique Non Verbal Signals method will require additional handling of probability logic to cope with the expected data. Furthermore other machine learning methods will be investigated in order to see if we can achieve better results with them on a larger dataset.



(a) Unique Non Verbal Signals method.

(b) The valence method

Figure 4: Model classification accuracy.

References

[AAA+18] Sina Mokhtarzadeh Azar, Sajjad Azami, Mina Ghadimi Atigh, Mohammad Javadi, and Ahmad Nickabadi. Zoom-RNN: A Novel Method for Person Recognition Using Recurrent Neural Networks. *arXiv preprint arXiv:1809.09189*, 2018.

[BT15] Ishan Behoora and Conrad S Tucker. Machine learning classification of design team members' body language patterns for real time emotional state detection. *Design Studies*, 39:100–127, July 2015.

[CK18] Eun Jeong Choi and Dong Keun Kim. Arousal and Valence Classification Model Based on Long Short-Term Memory and DEAP Data for Mental Healthcare Management. *Healthcare Informatics Research*, 24(4):309, 2018.

[CSWS17] Zhe Cao, Tomas Simon, Shih En Wei, and Yaser Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 1302–1310, 2017.

[KV18] Piyush Kawde and Gyanendra K. Verma. Deep belief network based affect recognition from physiological signals. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017*, volume 2018-Janua, pages 587–592. IEEE, October 2018.

[LJPM17] Eduardo Lupiani, Jose M Juarez, Jose Palma, and Roque Marin. Knowledge-Base d Systems Monitoring elderly people at home with temporal Case-Based Reasoning. *Knowledge-Based Systems*, 134:116–134, 2017.

[LLS17] Wenqian Lin, Chao Li, and Shouqian Sun. Deep Convolutional Neural Network for Emotion Recognition Using EEG and Peripheral Physiological Signal. In Yao Zhao, Xiangwei Kong, and David Taubman, editors, *Image and Graphics*, volume 10667 of *Lecture Notes in Computer Science*, pages 385–394. Springer International Publishing, Cham, 2017.

[MKN13] Angeliki Metallinou, Athanasios Katsamanis, and Shrikanth Narayanan. Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information. *IMAVIS*, 31(2):137–152, 2013.

[Red16] Joseph; Santosh Divvala; Ross Girshick; Ali Farhadi Redmon. (YOLO) You Only Look Once. *Cvpr*, 2016.

[SBS+18] Zangeneh Soroush, Behav Brain, Morteza Zangeneh Soroush, Keivan Maghooli, Seyed Kamaledin Setarehdan, and Ali Motie Nasrabadi. A novel approach to emotion recognition using local subset feature selection and modified Dempster - Shafer theory. *Behavioral and Brain Functions*, 4:1–15, 2018.

[SLQR18] Xiao Sun, Man Lv, Changqin Quan, and Fuji Ren. Improved facial expression recognition method based on ROI deep convolutional neutral network. *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, 2018-Janua:256–261, 2018.

[SPA14] Anastasios Skarlatidis, Georgios Paliouras, and Alexander Artikis. Probabilistic Event Calculus for Event Recognition. (September 2015), 2014.

[SW08] Han Sloetjes and Peter Wittenburg. Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, pages 816–820, 2008.

[VDCP$^+$10] Pieter Vos, Paul De Cock, Katja Petry, Wim Van Den Noortgate, and Bea Maes. What makes them feel like they do? investigating the subjective well-being in people with severe and profound disabilities. *Research in developmental disabilities*, 31(6):1623–1632, 2010.

[YWL$^+$17] Zhong Yin, Yongxiong Wang, Li Liu, Wei Zhang, and Jianhua Zhang. Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination. *Frontiers in Neurorobotics*, 11(April):1–16, April 2017.