Apathy Classification by Exploiting Task Relatedness

Anonymous FG2020 submission
Paper ID 198

***Abstract*— Apathy is characterized by symptoms such as reduced emotional response, lack of motivation, and limited social interaction. Current methods for *apathy diagnosis* require the patient's presence in a clinic and time consuming clinical interviews, which are costly and inconvenient for both patients and clinical staff, hindering among others large-scale diagnostics. In this work we propose a multi-task learning (MTL) framework for apathy classification based on facial analysis, entailing both *emotion* and *facial movements*. In addition, it leverages information from other auxiliary tasks (i.e., clinical scores), which might be closely or distantly related to the main task of apathy classification. Our proposed MTL approach (termed MTL+) improves apathy classification by jointly learning model weights and the relatedness of the auxiliary tasks to the main task in an iterative manner. Our results on 90 video sequences acquired from 45 subjects obtained an apathy classification accuracy of up to 80%, using the concatenated emotion and motion features. Our results further demonstrate the improved performance of MTL+ over MTL.**

I. INTRODUCTION

Apathy is defined as the quantitative reduction of goal-directed activity either in behavioral, cognitive, emotional or social dimensions [1]. It is mainly characterized by *limited emotional responses to positive or negative events, diminished empathy*, and *reduced verbal or physical reactions*. While experts suggest that the early indication of apathy could improve the intervention effects and decrease the global burden of the disease [2], apathy has been highly underdiagnosed. Its diagnosis, presently, is based on interviews between patients and their caregivers through a series of questionnaires. Towards *assisting such subjective assessment*, an objective and automated analysis carries the promise to *enable early apathy diagnostics*, leading to improved intervention effects, potentially increasing the performance of apathy detection in a non-invasive and efficient manner.

Motivated by the above, we introduce an automated system for apathy classification based on facial behavior analysis. Specifically, we investigate the effect of apathy on *facial movements* and *expressions*. To validate the reduced emotional responses of apathetic subjects, spontaneous expressions were elicited by asking the subjects to briefly narrate past positive and negative experiences. The *clinical diagnosis (clinical scores)* of the subjects was carried out by the clinicians along with the recording of facial videos during *positive* and *negative narration*. We explore the video data for apathy classification, while leveraging the information in clinical scores during model training.

When a model is trained for one task, there is a possibility of overfitting the task, as it does not generalize the noise pattern. Multi-task learning (MTL) [3] helps in learning the relevant and irrelevant features for different tasks, thereby learning a suitable general representation that ignores the data-dependent noise. However, the central issue in MTL is to learn and explore the relatedness among tasks. We propose an iterative algorithm, which we refer as MTL+, to jointly learn the MTL deep network parameters, as well as the task relationship alternatively. Different to other MTL works, MTL+ focuses on improving the performance of a main task (and ignoring the performance of other tasks) by automatically assigning lower weights to unrelated tasks and higher weights to similar tasks. We use apathy classification as the main task, whereas prediction of other clinical scores are considered as the auxiliary tasks.

In a nutshell, our proposed approach analyzes the facial expressions and face movement patterns to infer the apathy state. Video-level representation is utilized to regress the apathy related clinical scores through MTL. Further, we propose to jointly learn the model weights and the value of regularization parameters in an iterative manner (MTL+), thereby exploring the relationships of auxiliary tasks to the main task. The contributions of the paper are following.

- This paper is among the first to investigate automatic facial behavior analysis, i.e., facial movement and expression, for classifying apathy.
- We show that using MTL to estimate the clinical scores as auxiliary tasks improves the performance of apathy classification.
- We propose to learn the value of regularization parameters used in the loss function, thereby exploring the relationships of auxiliary tasks to the main task.

II. LITERATURE REVIEW

Apathy classification is a relatively new area of research in computer vision. Currently, clinical interviews or questionnaires are the only methods of a reliable apathy diagnosis [4], which are known to be time consuming and require the presence of patients at a hospital. Recently, literature reported the use of the neuroimaging modalities for apathy diagnosis [5], [6]. Structural and functional alteration of frontal-subcortical networks was observed in apathy patients through single-photon emission computed tomography, positron emission tomography, and diffusion tensor imaging [6]. However, diagnosis of apathy using facial analysis is a challenging problem and of utmost interest due to its non-invasive nature in field deployment.

Recently, a few works have been reported on facial analysis based apathy classification. Video-based apathy diagnosis by visual scanning behavior analysis is proposed in [7].

Specifically, here the authors investigated the sequences of *eye fixations and saccades* for emotional and non-emotional visual stimuli. The group difference and individual difference of scanning patterns were explored using recurrent neural networks (RNN). In [8], the authors reported a framework for apathy diagnosis based on facial motion and expression analysis from videos. Deviating from the above works, this paper proposes a novel multi-task leaning approach to jointly learn the apathy classification task along with predicting the other clinical scores to improve the model performance.

Multi-task Learning (MTL): We are interested in a scenario, where we have (1) insufficient annotated data for multiple tasks, and further we have that (2) the related tasks may or may not be related to the main task of 'apathy classification'. MTL exploits this scenario by learning the common information that is shared between multiple related tasks. MTL improves the performance of a specific task by learning the common information that is shared between multiple related tasks. An overview of MTL methods adopted in deep neural networks (DNN) is discussed in [3]. However, sharing parameters with unrelated and dissimilar tasks usually degrades the performance, which is known as *negative transfer*. Many solutions are proposed to overcome this problem. Kang et al. [9] and Liu et al. [10] grouped multiple tasks into several groups, so that the shared features can be learned jointly. Lee et al. [11] proposed an asymmetric MTL that jointly learns a regularization graph along with the task predictors to avoid performance degradation due to negative transfer. Multilinear relationship networks have been proposed by Long et al. [12], which networks model the task relatedness through the covariance structures over tasks based on novel tensor normal priors. Pan et al. [13] explored the common feature, task auxiliary feature, as well as task specific feature to indicate the shared features for each tasks. Hai et al. [14] carried out spam detection by jointly learning the correlation of each pair of tasks and the model parameters through stochastic alternating method. Similar learning of class relationship was carried out by Wu et al. [15] to train a unified framework for jointly learning feature relationship and class relatedness by imposing regularizations on the weights of the final output layer of DNNs.

In contrast to the aforementioned works, our method explores the relatedness of the predefined auxiliary tasks to the main task via *alternating optimization* (iteratively optimized). Our goal is to dynamically learn the relatedness of the auxiliary tasks to the main task so that the neural network model improves the performance of the main task, while avoiding the negative transfer. Thus, our method allows weights of task-related loss to dynamically evolve in an adaptive manner for improving the performance of the main task.

III. Proposed Method

A. Feature Extraction

While performing video level classification, researchers usually leverage the temporal dynamics [16] using long short-term memory (LSTM) or recurrent neural network (RNN). However, literature, which leverages the temporal patterns, predominantly focuses on categories like human activities (e.g., *walking*, *bowling*, *typing*, etc.) or visual content (e.g., *parade*, *wedding dance*, *bird*, *birthday*, etc.). These categories are quite distinguishable from human perspective. However, the problem at hand (apathy classification from facial videos) is more challenging, as opposed to the categories discussed above. We here note that even psychology experts are challenged in predicting the apathy state, by analyzing merely the face of a subject. Since there is no particular temporal pattern associated with the collected video data, we approached the problem with a model based on visual words, in which features extracted from each frame are further pooled for a codebook-based representation of the whole video. Though this model lacks temporal relation, it is advantageous for us, as the present data possesses no particular temporal pattern.

1) Emotion Features: As facial expressions are related to internal emotions, facial expression recognition has widely focused on emotion analysis [17]. While expression recognition is predominantly based on the six-expression model [17], as agreed with involved clinicians, we here use three categories of expressions, namely: *positive, negative*, and *neutral*. This choice stems from the highly limited expressions exhibited by the participants in the relative short video sequences under clinical conditions. Thus, we trained a convolutional neural network (CNN) model for *expression classification* with these three categories.

The use of pretrained VGG-Face [18], [19], [20] is a prominent architecture choice among recent works on face analysis. Since VGG-Face is trained with 2.6 million faces, it has been reported as a robust facial feature extractor, achieving promising results in facial expression recognition. In such works, the last few layers of VGG-Face are typically fine-tuned for respective applications. Publicly available expression datasets generally contain the universal classes (namely: anger, disgust, fear, happiness, sadness, and surprise). We here directly use 'happy' and 'neutral' samples from the dataset as positive and neutral classes respectively during training, while grouping the 'sad', 'anger' and 'disgust' samples into the negative class. In our experiments, the first five convolutional blocks use the pretrained weights of VGG-Face, while the latter layer weights are trained with the publicly available datasets.

Symptoms of apathy include reduced emotional responses. Hence, we hypothesize that (a) *expression intensity distribution*, as well as (b) *duration of each expression* throughout the video can be utilized to infer the apathy state. Therefore, we concatenated these two types of features for both, positive and negative narration videos, and call them "emotional features".

Expression intensity representation: We assume that the log probabilities of the softmax layer represent the emotion intensities corresponding to each category and pooled the frame-wise expression intensities into a histogram vector. Thus, we obtained a histogram vector (b bins in each histogram) for each expression, which are further combined

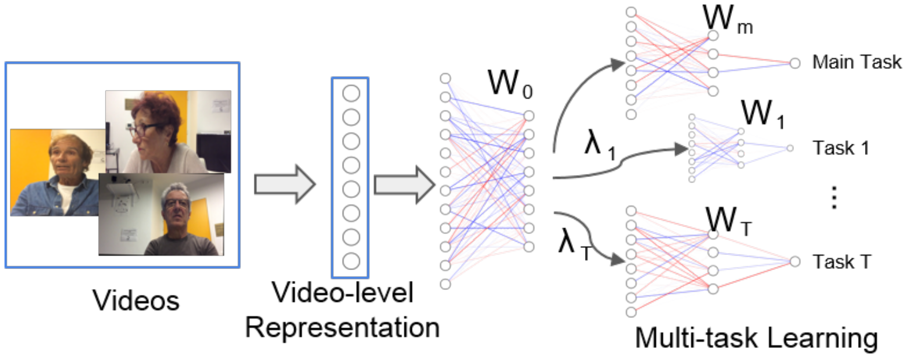


Fig. 1: The proposed MTL+ framework for jointly learning the task relatedness and model parameters.

together ($3 \times b$ dimensional feature vector for 3 classes) as a representation of expression intensities for the whole video. Similar to the *bag of words* analysis, here the histogram features represent the probable occurrence of an expression with certain intensity. As per our hypothesis, apathetic subjects will show less expressions with subtle intensities, thereby having higher bin counts in the first few bins.

Expression duration of dominant expressions in a video is an important cue for accessing the overall emotional display. If e-th expression is dominant for n_e number of frames out of total N number of video frames, then we used $t_e = \frac{n_e}{N}$ as the expression duration of e-th expression. The expression durations $(t_{pos}, t_{neg}, t_{neut})$ are appended to the expression representation, resulting in a $3 \times (b+1)$ dimensional feature vector.

2) Motion Features: From the psychological point of view, facial motion can be an additional prominent indicator of apathy. Since apathy is characterized by limited verbal or nonverbal interaction along with lack of interest in surrounding environment, we investigate head and facial movement in apathy detection. Inspired by Hammal et al. [21], we extract the dynamics of head and facial landmarks. We estimate the non-rigid head movements by tracking the facial points. The non-rigid facial landmark movements are associated with lips, eyes, eyebrows, and chin, while having a conversation or showing an expression. Specifically, the average movement of facial landmarks around these regions in successive frames is computed as motion feature.

Video sequences in our dataset are not time-limited and hence entail different video length. We obtain the statistical features (such as, minimum, maximum, mean, median, standard deviation, skewness, and kurtosis) of the motion information as the motion representation of a video. In addition, we include the b-bin histograms of motion values to preserve the motion intensity distribution information in motion representation, thereby making a vector of $b + 7$ dimensions.

B. Proposed MTL+ for learning task relatedness

We here propose an MTL method, namely MTL+, to leverage information that can be learned from other related tasks and proceed to learn a general representation. In this work, *apathy classification* is considered as the *main task*, while the prediction of *clinical scores* are used as *auxiliary tasks*. We streamline MTL+ to improve the performance of the main task (apathy classification) by leveraging the information from the auxiliary tasks (prediction of clinical scores). Generally, MTL approaches consider the contribution of different tasks to be equal or based on a prior. However, in our case, the relatedness of the tasks is not evident from facial analysis point of view. This motivates us to learn the relatedness scores, in order to avoid negative transfer during training of the MTL model.

MTL+: Suppose that there are T number of auxiliary supervised tasks, for which we seek to find the relatedness to the main task ($\{\lambda^a\}_{a=1}^T$, as shown in Figure 1). Thus, for each data sample $\mathbf{x}_i$, we have $T+1$ number of label information (y_i^m : main task label and $\{y_i^a\}_{a=1}^T$: auxiliary task labels). Let $\mathbf{W}_0$ be the weight of the network which learns the shared parameters, whereas $\mathbf{W}_m$ and $\{\mathbf{W}_a\}_{a=1}^T$ be the network weights related to the main task and the auxiliary tasks respectively. Thus, all the tasks are instrumental in learning the same feature space $f \in \mathcal{F}$ through $\mathbf{W_0}$ followed by learning the weights for individual tasks. For a sample $\mathbf{x}_i$, let $\mathcal{L}(y_i^m, f(\mathbf{x}_i; \mathbf{W}_0, \mathbf{W}_m))$ and $\mathcal{L}(y_i^a, f(\mathbf{x}_i; \mathbf{W}_0, \mathbf{W}_a))$ be the loss functions associated with the main and auxiliary tasks respectively. MTL optimizes

$$\begin{aligned}\underset{\mathbf{W}_0, \mathbf{W}_m, \{\mathbf{W}_t\}_{a=1}^T}{\arg\min} \quad & \sum_{i=1}^{N} \mathcal{L}(y_i^m, f(\mathbf{x}_i; \mathbf{W}_0, \mathbf{W}_m)) + \\ & \sum_{a=1}^{T}\sum_{i=1}^{N} \lambda^a \mathcal{L}(y_i^a, f(\mathbf{x}_i; \mathbf{W}_0, \mathbf{W}_a)). \quad (1)\end{aligned}$$

For simplicity, we will use $\mathbf{W} : \{\mathbf{W}_0, \mathbf{W}_m, \{\mathbf{W}_a\}_{a=1}^T\}$, $\mathcal{L}^a : \sum_{i=1}^N \mathcal{L}(y_i^a, f(\mathbf{x}_i; \mathbf{W}_0, \mathbf{W}_a))$. In MTL, λ^a is set based on prior knowledge. However, we jointly learn $\mathbf{W}$ and $\{\lambda^a\}_{a=1}^T$ in an alternating manner. Note that we aim to *improve the performance of main task, irrespective of its performance in auxiliary tasks*. Thus, the optimization performed in our framework is

$$\underset{\mathbf{W},\{\lambda^a\}_{a=1}^T}{\arg\min} \; \mathcal{L}^m + \sum_{a=1}^{T} \lambda^a \mathcal{L}^a. \tag{2}$$

Since there are two parameters to be learned, it is challenging to optimize the loss against both at the same time. Doing so will result in a trivial solution of $\lambda^a = 0, \forall a$, thus nullifying the loss incurred due to the auxiliary tasks. Therefore, we optimize $\mathbf{W}$, while updating λ^a after some epochs iteratively. The weight update is given by

$$\mathbf{W} \longleftarrow \mathbf{W} - \eta_1 \frac{\partial \mathcal{L}}{\partial \mathbf{W}}. \tag{3}$$

We propose to update λ^a in a similar manner using the gradient of loss function. However, $\frac{\partial \mathcal{L}}{\partial \lambda^a} = \mathcal{L}^a$. Therefore, λ^a can be updated as

$$\lambda^a \longleftarrow \lambda^a - \eta_2 \mathcal{L}^a. \tag{4}$$

λ^a can be interpreted as the importance coefficient of the loss, contributed by the a-th task to the total loss, thus signifies the relatedness of tasks. This learning process might get stuck in a local minima during the initial stages of training. To avoid that, we initialize $\lambda^a = 1, \forall a$ and traine $\mathbf{W}$ to have fair performance. Then, λ^a is updated intermittently, while further optimizing $\mathbf{W}$, in order to improve the performance of the main task. This method is termed as MTL+ in this work. Note that η_1 and η_2 are learning rates.

Early stopping criterion: Since λ_a is penalized intermittently according to the loss value incurred by task a, it is possible that this value will reduce close to zero after sufficient number of updates. To avoid that, we propose an effective early stopping criterion, which stops the further update of λ_a before they begin to over-fit the main task. For a given constant $\epsilon > 0$, we stop updating λ_a for the task that satisfies

$$\frac{k.\text{med}_{j=t-k}^{t} E_{val}^{a}(j)}{\sum_{j=t-k}^{t} E_{val}^{a}(j) - k.\text{med}_{j=t-k}^{t} E_{val}^{a}(j)} > \epsilon, \tag{5}$$

where $E_{val}^a(j)$ represent the loss values of the ath task at jth iteration for validation data; and med stands for median.

IV. Experiments and Results

A. Dataset Description

We acquired the dataset at the Nice Memory Research Center, located at the Institute Claude Pompidou in the Nice University Hospital. Patients, who suffer from subjective memory complaint to severe cognitive impairment were included in the study. The dataset comprises 45 subjects, out of which 18 constitute patients with apathy condition. Among the apathy and control subjects, the number of female patients were 38% and 62%, respectively.

The dataset includes the scenario, where the clinician is interviewing a patient involving (i) the collection of demographic details, (ii) standardized neuropsychological assessment tests, and (iii) a short positive and negative experience narration. The neuropsychological assessment was carried out by one-on-one interview with a battery of cognitive tests to access the anxiety, affect, interest, etc. In our experiment, we use nine clinical scores: mini mental state examination (MMSE) [22], and neuropsychiatric apathy inventory (NPI-apathy, NPI-anxiety, NPI-depression, NPI-total) [23], clinical dementia score (CDR), and apathy inventory (IA-affect, IA-initiative, IA-interest) [24]. To elicit spontaneous facial expressions, in (iii) the participants were asked to narrate some positive and negative events or experiences from their past ("tell me a positive/negative event of your life in one minute"). The video data was recorded with a tablet controlled by the psychologist. Though most of the videos have near-frontal face, there exists a set of pose variations, as well as facial occlusions in the dataset. Moreover, the expressions were highly subtle. We note that the average video length was about one minute.

B. Implementation details

Prior to the main framework, we detect faces from the dataset-videos using MTCNN [25], followed by face alignment by positioning both eyes at a fixed distance parallel to the horizontal axis. The aligned faces are re-sized to 224×224 resolution, constituting the input for the CNN model. The CNN model is trained to classify the face into three expression classes, namely: positive, negative and neutral. We use AffectNet [26] dataset to train the CNN model. The Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of 0.0001 is used for training the deep model. The facial landmarks are detected using DLIB [27] in order to compute the motion features. In all our experiments, we consider the histograms with 10 bins ($b = 10$) for both motion and emotion feature extraction. The extracted features are further normalized to zero mean and unit variance before feeding into the MTL framework, which consists of two dense layers with 128 and 32 units respectively with dropout ratio of 0.5 followed by the output layer with 10 units for classification/regression tasks. In our case, apathy classification is the main task, while predicting the values of the clinical scores through regression is considered as the auxiliary tasks.

All the results reported here are obtained by performing 10-fold cross validation. Note that, the validation set used in our experiments is different from the test set. 10% of the samples from the train data are selected randomly to constitute the validation set.

C. Experimental Results

Combining features of positive and negative narration: The performance of apathy classification for various features is provided in Table I. Note that our dataset contains two videos (positive and negative narration) per subject. Here '*after fusion*' refers to the experiments where the features from these two videos are concatenated and used for model training. We first show the results without using MTL in order to observe the performance improvements with MTL.

TABLE I: Performance improvement by fusion of features from positive and negative narration. **(without using MTL)**

Features used	**Without fusion** of positive and negative narration Accuracy	F1-score	**After fusion** of positive and negative narration Accuracy	F1-score
Motion Features	58.88	0.505	57.77	0.555
Emotion features	52.68	0.532	64.44	0.622
Emotion features + Motion Features	53.86	0.526	**77.77**	**0.757**

TABLE II: Performance comparison when features from positive and negative narrations are concatenated. (MTL: mutlti-task learning considering equal contribution of each task; MTL+: proposed method that exploits task relatedness.) The proposed MTL+ improves the classification accuracy.

Features used	Accuracy	F1-score
MTL with Motion Features	62.22	0.582
MTL with Emotion features	66.66	0.638
MTL with Emotion + Motion Features	71.11	0.716
MTL+ with Motion Features	71.11	0.679
MTL+ with Emotion features	77.77	0.776
MTL+ with Emotion + Motion Features	**80.00**	**0.786**

The results shown in Table I is obtained without using MTL on auxiliary tasks.

As per Table I, the performance achieved '*without fusion*' is very poor given the data imbalance and the binary classification problem. After fusion, the accuracy of the framework improves from 52% to 64.44% using the emotion features. The best performance (accuracy = 77.77% and F1-score = 0.757) is achieved, when emotion features are combined with motion features. This proves the complementary information present in motion and emotion features.

From psychological point of view, apathetic persons are indifferent toward any emotions and hence expressions. However, the healthy subjects exhibit limited expressions in a clinical environment (such as ours) as well, which challenges the classification task. Hence, the presence of both, positive and negative narrations per subject is pertinent, and combining the features that are extracted from both videos-sequences provides a broader spectrum of facial expressions, instrumental for apathy classification.

Performance of proposed MTL: Table II reports the performance of MTL with and without learning task relatedness. Note that all results reported in Table II are obtained by combining the features from both, positive and negative narrations. Comparing the accuracy in Table I and II, we observe performance improvement with MTL when individual features are used (57% to 62% for *Motion*; and 64% to 66% for *Emotion*). However, the performance of Motion + Emotion features is degraded from 77% to 71% by using MTL, which is further boosted to 80% by MTL+. We believe that the reduced performance was due to the negative transfer, which was avoided in MTL+ to obtain better performance.

As can be seen from Table II, the performance of MTL improves significantly by learning the task relatedness (MTL+). For instance, the accuracy of motion features improved from 62% to 71%, and the accuracy of emotion features improved from 66% to 77%. Similarly, we observe significant improvement of the F1-score for MTL+ as well. The best performance (classification accuracy = 80% and F1-score = 0.78) is achieved by MTL+ using the concatenated motion and emotion features. The confusion matrices of different feature extraction methods are reported in Figure 2.

Task relatedness: The task relatedness, learned by the proposed method is shown in Figure 3. The variation of the task relatedness is visualized here by using median and lower-upper quartile values of $\{\lambda_t\}_{t=1}^{T}$ obtained during training over the 10-folds. As can be seen, IA-affect and NPI-depression are found to be highly related to the apathy classification task. More importantly, this illustrates the importance of emotion and motion features for estimating other clinical scores. The tasks that are assigned low relatedness-score contribute less toward the loss function, thus avoiding the negative transfer during model training.

V. Conclusions

We present an automatic apathy detection method, which analyzes facial emotion and motion while sharing the knowledge contained in the training signals of other tasks. Due to the uncertainty in the relatedness of auxiliary tasks, we propose a framework (MTL+) to jointly learn the model weights along with the task relatedness, thereby avoiding negative transfer by the distantly related tasks. Our framework benefits from the concatenation of features from both positive and negative narration videos. Experimental results show that the performance may improve or degrade by MTL, while the proposed MTL+ consistently achieves outperforms other methods. For example, *emotion* based apathy detection achieved an accuracy of 64%, 66%, and 77% in case of without MTL, MTL, and MTL+, respectively. We obtain the best results (accuracy = 80%, F1-score = 0.78) by using MTL+ on the combination of *emotion* and *motion* features.

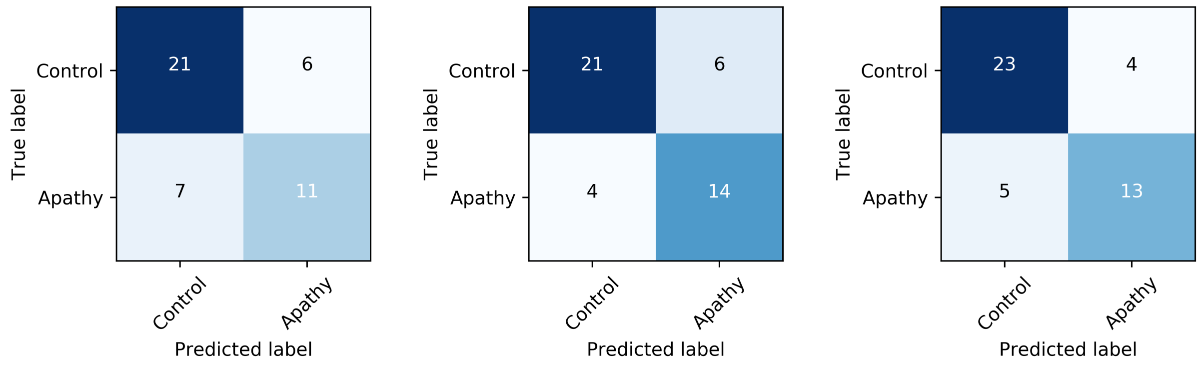
References

[1] P. Robert, K. Lanctôt, L. Agüera-Ortiz, P. Aalten, F. Bremond, M. Defrancesco, C. Hanon, R. David, B. Dubois, K. Dujardin *et al.*, "Is it time to revise the diagnostic criteria for apathy in brain disorders? the 2018 international consensus group," *European Psychiatry*, vol. 54, pp. 71–76, 2018.

[2] H. Hampel, R. Frank, K. Broich, S. J. Teipel, R. G. Katz, J. Hardy, K. Herholz, A. L. Bokde, F. Jessen, Y. C. Hoessler *et al.*, "Biomarkers for alzheimer's disease: academic, industry and regulatory perspectives," *Nature reviews Drug discovery*, vol. 9, no. 7, p. 560, 2010.

[3] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[4] J. Pagonabarraga, J. Kulisevsky, A. P. Strafella, and P. Krack, "Apathy in parkinson's disease: clinical features, neural substrates, diagnosis, and treatment," *The Lancet Neurology*, vol. 14, no. 5, pp. 518–531, 2015.



(a) (b) (c)

Fig. 2: The confusion matrices obtained with proposed MTL for (a) motion features, (b) emotion features, and (c) combination of motion and emotion features.

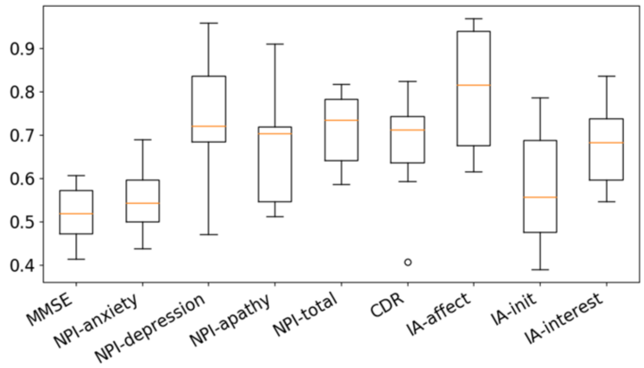


Fig. 3: The relatedness of different tasks to apathy classification for the concatenated motion and emotion features.

[5] L. Agüera-Ortiz, J. A. Hernandez-Tamames, P. Martinez-Martin, I. Cruz-Orduña, G. Pajares, J. López-Alvarez, R. S. Osorio, M. Sanz, and J. Olazarán, "Structural correlates of apathy in alzheimer's disease: a multimodal mri study," *International journal of geriatric psychiatry*, vol. 32, no. 8, pp. 922–930, 2017.

[6] C. Theleritis, A. Politis, K. Siarkos, and C. G. Lyketsos, "A review of neuroimaging findings of apathy in alzheimer's disease," *International psychogeriatrics*, vol. 26, no. 2, pp. 195–207, 2014.

[7] J. Chung, S. A. Chau, N. Herrmann, K. L. Lanctôt, and M. Eizenman, "Detection of apathy in alzheimer patients by analysing visual scanning behaviour with rnns," in *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 149–157.

[8] S. Happy, A. Dantcheva, A. Das, R. Zeghari, P. Robert, and F. Bremond, "Characterizing the state of apathy with facial expression and motion analysis," in *14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–8.

[9] Z. Kang, K. Grauman, and F. Sha, "Learning with whom to share in multi-task feature learning." in *ICML*, 2011, pp. 521–528.

[10] S. Liu and S. J. Pan, "Adaptive group sparse multi-task learning via trace lasso," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 2358–2364.

[11] G. Lee, E. Yang, and S. Hwang, "Asymmetric multi-task learning based on task relatedness and loss," in *International Conference on Machine Learning*, 2016, pp. 230–238.

[12] M. Long, Z. Cao, J. Wang, and S. Y. Philip, "Learning multiple tasks with multilinear relationship networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 1594–1603.

[13] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang, "Task sensitive feature exploration and learning for multitask graph classification," *IEEE transactions on cybernetics*, vol. 47, no. 3, pp. 744–758, 2017.

[14] Z. Hai, P. Zhao, P. Cheng, P. Yang, X.-L. Li, and G. Li, "Deceptive review spam detection via exploiting task relatedness and unlabeled data," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1817–1826.

[15] Z. Wu, Y.-G. Jiang, J. Wang, J. Pu, and X. Xue, "Exploring inter-feature and inter-class relationships with deep neural networks for video classification," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 167–176.

[16] Y.-G. Jiang, Z. Wu, J. Tang, Z. Li, X. Xue, and S.-F. Chang, "Modeling multimodal clues in a hybrid deep learning framework for video classification," *IEEE Transactions on Multimedia*, 2018.

[17] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.

[18] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[19] H. Ding, S. K. Zhou, and R. Chellappa, "Facenet2expnet: Regularizing a deep face recognition net for expression recognition," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE, 2017, pp. 118–126.

[20] Y. Liu, X. Yuan, X. Gong, Z. Xie, F. Fang, and Z. Luo, "Conditional convolution neural network enhanced random forest for facial expression recognition," *Pattern Recognition*, vol. 84, pp. 251–261, 2018.

[21] Z. Hammal, J. F. Cohn, C. Heike, and M. L. Speltz, "Automatic measurement of head and facial movement for analysis and detection of infants' positive and negative affect," *Frontiers in ICT*, vol. 2, p. 21, 2015.

[22] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""mini-mental state": a practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.

[23] J. L. Cummings, M. Mega, K. Gray, S. Rosenberg-Thompson, D. A. Carusi, and J. Gornbein, "The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia," *Neurology*, vol. 44, no. 12, pp. 2308–2308, 1994.

[24] M. Benoit, S. Clairet, P. Koulibaly, J. Darcourt, and P. Robert, "Brain perfusion correlates of the apathy inventory dimensions of alzheimer's disease," *International journal of geriatric psychiatry*, vol. 19, no. 9, pp. 864–869, 2004.

[25] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[26] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, 2017.

[27] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.