

A Bayesian Deep Learning Framework for End-To-End Prediction of Emotion from Heartbeat

Ross Harper¹ and Joshua Southern¹

Abstract—Automatic prediction of emotion promises to revolutionise human-computer interaction. Recent trends involve fusion of multiple modalities – audio, visual, and physiological – to classify emotional state. However, practical considerations ‘in the wild’ limit collection of this physiological data to commoditised heartbeat sensors. Furthermore, real-world applications often require some measure of uncertainty over model output. We present here an end-to-end deep learning model for classifying emotional valence from unimodal heartbeat data. We further propose a Bayesian framework for modelling uncertainty over valence predictions, and describe a procedure for tuning output according to varying demands on confidence. We benchmarked our framework against two established datasets within the field and achieved peak classification accuracy of 90%. These results lay the foundation for applications of affective computing in real-world domains such as healthcare, where a high premium is placed on non-invasive collection of data, and predictive certainty.

Index Terms—Bayesian neural networks, Electrocardiography, Emotion recognition, End-to-end learning

I. INTRODUCTION

HUMANS are social creatures that evolved to think and communicate with emotional information. Cognition and emotion are thus intrinsically linked. Indeed, emotion has been shown to impact attention [1], [2], [3], memory [4], [5], [6], [7], perception [8], [9], and decision-making [10], [11], [12]. Automated analysis of human emotion has correspondingly garnered significant interest across academia and industry in recent years.

A wealth of research within the field of affective computing has focussed on the analysis of face, voice and text [13], [14], [15], [16], [17], [18], [19], [20]. Comparatively few studies, however, investigate the prediction of emotion from physiological signals. This is perhaps unsurprising - humans too rely on audiovisual data for emotion recognition. Artificial systems, however, need not be similarly constrained. For ease of reference, we refer to audio-, visual-, and physiology-based methods of emotion detection as ED_A , ED_V , and ED_P respectively.

ED_P has tremendous potential to compliment existing tools of affective computation. ED_A and ED_V rely heavily on expression, which varies across individuals and cultures [21], [22] and leaves room for deception. By comparison, physiological processes are far less volitional. ED_P further presents an opportunity for non-invasive continuous monitoring. Physiological signals may be passively analysed throughout the day,

whereas audiovisual data is rarely so persistent. ED_P therefore has the option to fill critical gaps in domains that may not permit continuous collection of quality audiovisual data (e.g. healthcare, transport, and hospitality).

To date, there exist a range of affordable wearable monitoring devices that possess the capacity for high quality heartbeat monitoring [23], [24]. These devices have already been used to detect cardiac abnormalities such as atrial fibrillation [25]. However, while heartbeat data is abundant, other physiological signals are markedly less common. To be immediately relevant, ED_P systems must be able to generate accurate predictions with only unimodal heartbeat input.

The link between emotion and heartbeat has a neurobiological correlate in the limbic and autonomic nervous systems. The limbic system, which includes structures such as the amygdala and hippocampus, is important for the processing of emotional information [26], [27], [28]. Physiological responses to emotional stimuli are then coordinated by another limbic structure, the hypothalamus, which regulates heartbeat through antagonistic activity in the sympathetic and parasympathetic branches of the autonomic nervous system (ANS) [29], [30]. This relationship between emotion and heartbeat is recapitulated by neuropsychological theories, which state that the mental component of emotion is simply the cognitive perception of physiological changes elicited by emotion-inducing stimuli [31], [32].

The cardiac cycle is a complex dynamical process. Correspondingly, the heartbeat time series is non-stationary [33] and non-linear [34]. In order to adequately describe these characteristics, ED_P systems must model complex temporal structure. Furthermore, for ED_P systems to be applied in real-world applications where confidence is a key ingredient to decision-making (e.g. healthcare [35]), the model must describe uncertainty over the emotional state output.

In this study, we develop an end-to-end deep learning model for classifying emotional valence from unimodal heartbeat data. We implement recurrent and convolutional architectures to model temporal structure in the input signal, and propose a Bayesian framework for modelling uncertainty over the output. We go on to describe a procedure for tuning model output for varying demands on certainty. This will be critical for applications of affective computing in domains such as healthcare, where a high premium is placed on predictive interpretability. We believe this is the first such model of its kind, and accelerates near-term relevance of ED_P in real-world settings.

¹Ross Harper and Joshua Southern are with Limbic Ltd, 67-71 Shoreditch High St, London, E1 6JJ, UK.

Correspondence to Ross Harper (ross@limbic.ai)

II. RELATED WORK

This section provides an overview of relevant work, with a focus on (A) unimodal heartbeat and temporal models for ED_P, and (B) Bayesian neural networks.

A. Unimodal Heartbeat and Temporal Models

Physiological markers of autonomic nervous activity include galvanic skin response (GSR), electroencephalogram (EEG), electromyogram (EMG), respiration, skin temperature (ST), electrocardiogram (ECG) and photoplethysmogram (PPG) [36]. Note that a heartbeat time series can easily be extracted from both ECG and PPG in the form of inter-beat-intervals (IBIs).

Existing approaches for ED_P typically pool a number of biosignals as multimodal input to classifier algorithms [37], [38], [39]. However, this directly contrasts the near-unimodal nature of affordable wearable devices in use today. Comparatively few studies narrow their scope in accordance with these practical limitations.

Those studies that have explored unimodal heartbeat models for emotion detection tend to ignore temporal structure of the signal. Instead, they employ ‘static’ classifiers that process global features from the input time series (or for a small number of segments). Such approaches include Naive Bayes (NB) [40], [41], linear discriminant analysis (LDA) [42], and support vector machine (SVM) [43], [44], [45]. A summary can be found in Table I.

A number of studies have sought to model temporal information within EEG signals, using hidden Markov models [46], Gaussian Process models [47], continuous conditional random fields [48], and long short-term memory (LSTM) neural networks [49]. Such temporal treatment, however, is rare for other physiological data. One notable exception to this involved the use of a temporal neural network to predict valence from ECG input [50]. Here, a combination of convolutional and recurrent layers performed end-to-end learning, improving on computationally expensive manual feature engineering schemes. In this study, we too implement end-to-end learning, while further limiting model input to IBI time series to simulate the type of data generated by consumer wearables.

At this point, we wish to point out a stark absence of consensus within the literature around data subsetting for machine learning. A typical experimental setup can yield multiple input-output pairs for a single study participant. Many studies partition train, validation, and test datasets without reference to the source (the study participant from which the data was generated). However, ECG has been shown to exhibit subject-specificity [51]. It might therefore be less suitable to include data from a given participant in both the train and test/validation subsets. Moreover, real-world applications may not permit subject-specific calibration, making models that can generalise to new individuals necessary. We propose that a sensible evaluation method is leave-k-subjects-out (LkSO) cross-validation, which has been used previously [40], [52], [42] and will be adopted in this study.

B. Bayesian Neural Networks

Despite the widespread success of deep learning, traditional neural networks lack probabilistic considerations. This is an issue for applications where representing uncertainty is of critical importance (e.g. medical diagnosis) [53].

To combat this, neural networks may be re-cast as Bayesian models to capture probability in the output. In this formalism, network weights belong to some prior distribution with parameters θ . Posterior distributions are then conditioned on the data according to Bayes’ rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (1)$$

where D is the data.

While useful from a theoretical perspective, Equation 1 is infeasible to compute. Indeed, the evidence term in the denominator amounts to the integral over all possible values of the network weights:

$$\begin{aligned} p(D) &= p(x, y) \\ &= \int p(y|x, \theta)p(\theta)d\theta \end{aligned} \quad (2)$$

where the data D can be written as x, y input-output pairs for a supervised task.

For obvious reasons, exact posterior inference is rarely achievable. Instead, we seek to approximate these posterior distributions. Early attempts at this include Monte Carlo (MC) [54] or Laplace [55] approximation methods. However, these are slow and computationally expensive when applied to modern deep learning architectures. Research in the field has focussed on identifying faster inference methods such as stochastic gradient Langevin diffusion [56], expectation propagation [57], and variational methods [58].

Interestingly, Bayesian neural networks can also be constructed using Monte-Carlo dropout - a common approach to reduce over-fitting [59]. Dropout is a process by which individual nodes within the network are randomly removed during training according to a specified probability [60]. By implementing dropout at *test*, and performing N stochastic forward passes through the network, we can approximate a posterior distribution over model predictions (approaching the true distribution as $N \rightarrow \infty$). In this paper, we implement Monte-Carlo dropout as an efficient way to describe uncertainty over emotional state predictions.

III. PROPOSED FRAMEWORK

An overview of our model is shown in Figure 2. Data flows through two concurrent streams. One stream comprises four stacked convolutional layers that extract local patterns along the length of the time series. Each convolutional layer is followed by dropout and a ReLU activation function. A global average pooling layer is then applied to reduce the number of parameters in the model and decrease over-fitting. The second stream comprises a bidirectional LSTM followed by dropout. This models both past and future sequence structure in the input. The output of both streams are then concatenated before

TABLE I
SUMMARY OF RELEVANT WORK - CLASSIFICATION OF EMOTION FROM HEARTBEAT

Author	Stimulus	Modality	Subjects	LkSO	Model	Classes	Performance
Katsigiannis & Ramzan 2018	Videos	ECG	23	No	SVM	High/Low Valence	F1. 0.5305 (Chance: 0.500)
Subramanian et al 2018	Videos	ECG	58	No	NB	High/Low Valence	Acc. 60% (Chance: 50%)
Miranda-Correa et al 2017	Videos	ECG	40	Yes	NB	High/Low Valence	F1. 0.545 (Chance: 0.500)
Guo et al 2016	Videos	ECG	25	No	SVM	High/Low Valence	Acc. 71.40 (Chance: 50%)
Ferdinando et al 2016	Videos & IAPS	ECG	27	Yes	KNN	High/Medium/High Valence	Acc. 59.2% (Chance: 33.3%)
Valenza et al 2014	IAPS	ECG	30	No	SVM	High/Low Valence	Acc. 79.15% (Chance: 50%)
Agrafioti et al 2012	IAPS	ECG	32	Yes	LDA	Gore, Erotica	Acc. 46.56% (Chance: 50%)

passing through a dense layer to output a regression estimate for valence.

In order to capture uncertainty in model predictions, dropout is applied at test time. For a single input sample, stochastic forward propagation is run N times to generate a distribution over model output. This empirical distribution approximates the posterior probability over valence, given the input IBI time series.

For regression problems, the reader may stop here. In order to translate from a regression to a classification scheme, we introduce decision boundaries in continuous space. For a binary class problem, this decision boundary is along the central point of the valence scale to delimit two class zones (high and low valence). We next introduce a confidence threshold, α , to tune predictions according to a specified level of confidence. For example, when $\alpha = 0.95$, at least 95% of the output distribution must lie within a given class zone in order for the input sample to be classified as belonging to that class (Fig. 1). If this is not the case, no prediction is made (the model respectfully makes no comment). As our model may not classify all instances, we adopt the term ‘coverage’ to denote the set of cases for which it is confident enough to make a prediction.

Note that for a binary classification problem, there will always be at least 50% of the output distribution within one of the two class zones. Thus, when $\alpha = 0.5$, classification is determined by the median of the output distribution (Fig. 1), and the coverage is 100%. As α increases, model behaviour moves from risky to cautious – lower coverage, but more certain.

IV. DATA

We applied our Bayesian deep learning framework for end-to-end prediction of emotion using heartbeat (IBI) data from two established datasets – AMIGOS [40] and DREAMER [43]. In this section, we provide details on these data, which were chosen for their quality, clarity, and close comparability.

A. AMIGOS

These data include 40 healthy participants (13 female; 27 male) aged between 21 and 40 years old (mean: 28.3). ECG data was recorded using a Shimmer™ ECG wireless monitoring device (256 Hz, 12 bit resolution). Subjects watched 16 short videos (duration <250 seconds) that had been previously scored for emotional content. The videos were presented in a random order with each trial comprising a 5-second baseline recording showing a fixation cross, presentation of the video stimulus, followed by self-assessment of valence on a scale of 1 to 9 using the self-assessment manikin (SAM) [61].

B. DREAMER

These data include 25 healthy participants (11 female; 14 male) aged between 22 and 33 years old (mean: 26.6). ECG data was recorded using a Shimmer™ ECG wireless monitoring device (256 Hz, 12 bit resolution). Subjects watched 18 short film clips (duration: <395 seconds), which had been previously scored for their ability to elicit emotional responses [62]. Each film clip was followed by self-assessment of valence on a scale of 1 to 5 using SAM [61], and preceded by a neutral video presentation to establish baseline emotional state [62].

V. METHODS

A. Pre-processing

To obtain data of the kind generated by consumer wearables, IBIs were extracted from the ECG time-series using a combined adaptive threshold approach [63]. This markedly reduces the information content of the input signal. Nevertheless, inter-beat dynamics have previously been shown suitable for emotional state classification [45]. The IBI time series was z-score normalised and zero padded to the length of the longest training sample.

B. Training and Hyperparameters

Parameter search was used to select model hyperparameters. For this, a LkSO validation set of 4 subjects was used to assess best-performance (lowest mean-squared loss) for a given

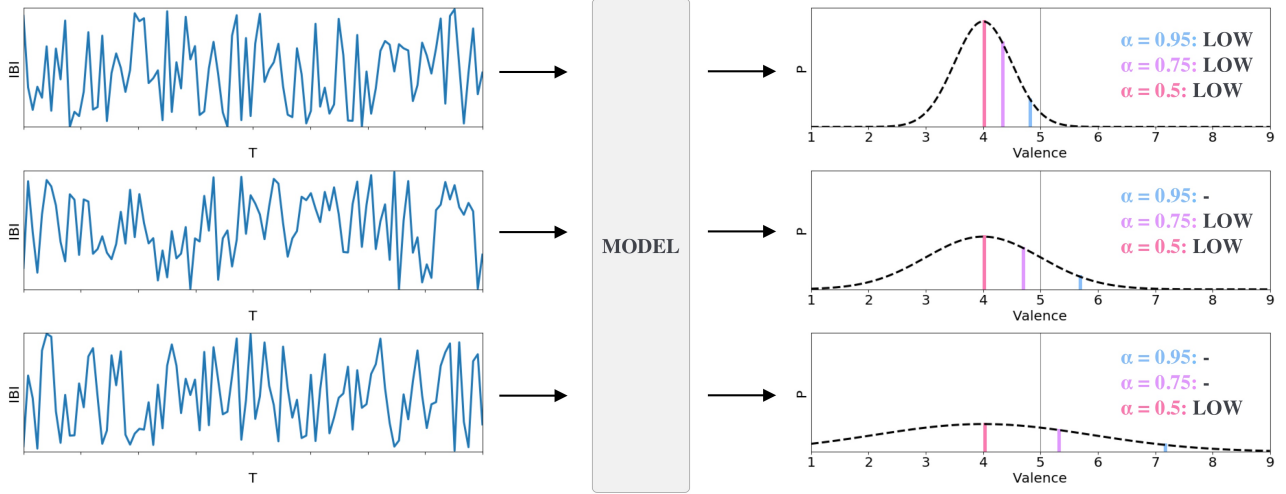


Fig. 1. Probabilistic framework for a binary classification problem. Input IBI time series (left) are passed through the Bayesian model (middle), which outputs a posterior probability over valence (right). Inputs are classified according to confidence threshold, α (illustrated for $\alpha = 0.95, 0.75$, and 0.5 on three example output distributions, which have the same mode but vary in certainty).

combination of hyperparameters. Convolution kernels were initialised as He normal [64] with filter size set to 128, and window size decreasing from 8 to 2 time steps with network depth. 50% dropout was applied after each convolutional block, and 80% dropout followed the bi-directional LSTM comprising 32 hidden units. Training was run for 1500 epochs using Adam optimisation [65]. Learning rate decreased from e^{-3} to e^{-4} , halving with a patience of 100 epochs. Final model parameters were set to those associated with the lowest mean-squared loss on the validation set during training. The model was implemented using Tensorflow [66].

C. Evaluation

As discussed in Section II-A, model performance was assessed using 10-fold leave-one-subject-out cross validation in order to generalise to new participants. Dropout was applied at test time with $N = 1000$ forward propagations made through the network to generate an empirical distribution over model output. In accordance with the original studies from which we obtained our data, labels for valence were divided into high and low classes using the midpoint value of the SAM scale (5 for AMIGOS; 3 for DREAMER). As outlined in Section III, a given test input sample was classified as high/low valence provided a proportion of at least α posterior distribution mass fell within a given class zone. If this was not the case, no prediction was made. Model accuracy was then calculated as the fraction of correct classifications over total predictions covered by the model.

VI. RESULTS

To identify the benefit conferred by our temporal network architecture, we first evaluated our model without dropout at test time. In this non-Bayesian setting, model output was a single point estimate, that fell either in the high or low class zones, and was classified accordingly. Here, we achieved

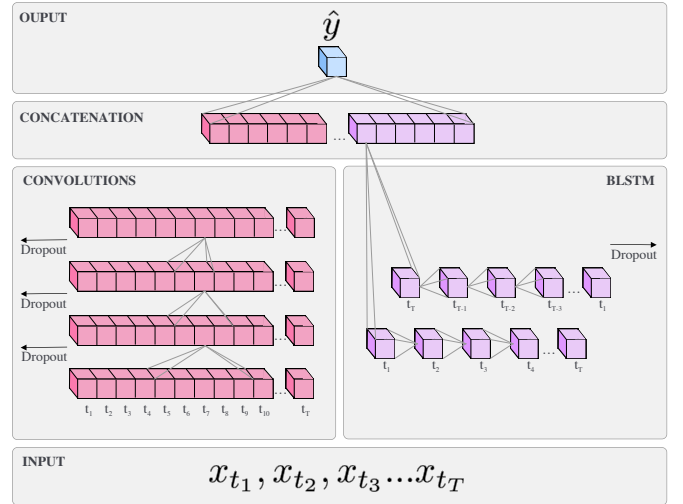


Fig. 2. End-to-end model architecture. Data flows through two temporal processing streams: 1D convolutions (pink) and a bi-directional LSTM (purple). The output from both streams is then concatenated before passing through a dense layer to output a regression estimate for valence, \hat{y} .

higher accuracy across both datasets than previously reported [40], [43] (Table II).

We next implemented our Bayesian framework with confidence threshold set to 50% ($\alpha = 0.5$). As expected, maximal model coverage was observed (Fig. 3 B). Furthermore, classification accuracy outperformed the non-Bayesian setting (Table. II), illustrating the performance increase conferred by our probabilistic framework, which can be considered a special case of ensemble learning.

As the certainty threshold, α increases, so too does classification accuracy, demonstrating a clear relationship between model confidence and propensity to make accurate predictions (Fig. 3 A, and Table II). Naturally, as α increases, model cov-

erage decreases due to the fact that fewer output distributions meet the necessary threshold for a prediction to be made. We see that with a 90% confidence threshold ($\alpha = 0.9$), our model achieved peak accuracy for both datasets (Fig. 3 A, and Table II).

Interestingly, we found that certainty over model output was significantly greater for input time series that belong to the low valence class, for both datasets, as shown by MannWhitneyWilcoxon test (Fig. 3 G,H). This pattern is also reflected in the consistently better performance observed for the low valence class (Fig. 3 C,D,E,F).

TABLE II
COMPARISON OF MEAN ACCURACY AND F1 SCORES.

		[40]	[43]	Non-Bayes	Bayes; $\alpha = 0.5$	Bayes; $\alpha = 0.9$
AMIGOS	Acc.	0.54	-	0.79	0.81	0.90
	F1	-	-	-	0.80	0.88
DREAMER	Acc.	-	0.62	0.70	0.71	0.86
	F1	-	0.53	-	0.66	0.83

VII. DISCUSSION

The growing prevalence of high-fidelity, affordable wearable monitoring devices has introduced an opportunity for continuous emotion detection ‘in the wild’. The vast majority of approaches in the literature rely on fusion of multiple physiological signals for physiology-based emotion detection, or ED_p. Although this multimodal treatment provides significant performance benefits, it is limited in practice. Indeed, the existing landscape for consumer electronics has near-unimodal sensor availability, limiting physiological signals primarily to IBI time series. Timely application of ED_p in real world settings, therefore, requires models that comply with these restrictions.

It has been shown previously that IBI extracted from PPG corresponds closely with IBI extracted from ECG [67], [68]. This allowed us to exploit existing high-quality ECG datasets for this study. We developed an end-to-end neural network capable of modelling temporal structure in the IBI time series, which outperformed previous classifiers on this task [40], [43].

We went on to re-cast our model as a probabilistic neural network to capture uncertainty in the output. Through the use of a confidence threshold parameter, α , we demonstrated a framework for tuning model predictions in order to trade off accuracy against coverage. Indeed, we report peak accuracy of 90%. Further flexibility was achieved by framing our model as a regression problem, which allows the experimenter to specify decision boundaries appropriate for binary- or multi-class tasks.

Incorporating Bayesian considerations could drastically improve the applicability of affective computing in tasks where confidence is critical. For example, emotion detection for mental health monitoring might reasonably require high levels of certainty to predict the onset of major depressive disorder. Additionally, clinical triaging is possible, where uncertain model predictions are sent to a human expert (or a more computationally expensive model) for review. Similar levels

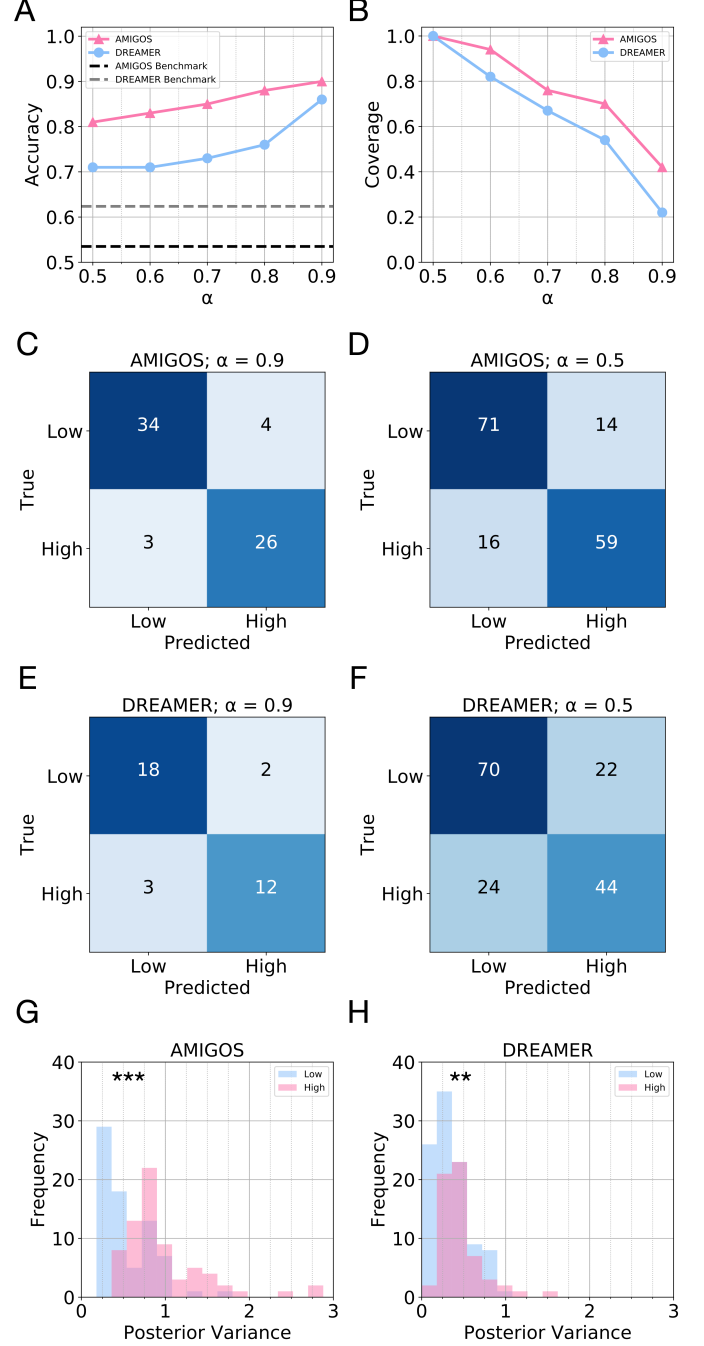


Fig. 3. Results. (A) Model accuracy as a function of α . Results are shown for AMIGOS (pink, triangles) and DREAMER (blue, circles) with benchmarks achieved by these original publications shown (black and grey dashed lines respectively). (B) Model coverage as a function of α . Results are shown for AMIGOS (pink, triangles) and DREAMER (blue, circles). (C,D,E,F) Confusion matrices shown for AMIGOS (C,D) and DREAMER (E,F), with α set to 0.9 (C,E) or 0.5 (D,F). (G,H) Model uncertainty (as measured by variance of posterior output distribution) for low (blue) and high (pink) valence class labels. Significant differences observed for both AMIGOS (G, $p < 1 \times 10^{-11}$) and DREAMER (H, $p < 0.001$)

of certainty may not, however, be absolutely necessary in consumer products. For example, content recommendations based on user mood can afford lower accuracy to ensure a greater number of recommendations.

Our probabilistic framework also provides deeper insight into the underlying properties of the data. Indeed, we found greater levels of model certainty when classifying the low valence class. This could be attributed to the subjective certainty of participants during their own self-reports. Alternatively, signatures of low valence within the heartbeat signal might contain more information than their high valence counterparts. We look forward to future investigation in this area, and further experimentation using affordable wearable monitoring devices in the wild.

VIII. CONCLUSIONS

In this study, we developed an end-to-end deep learning model for classifying emotional valence from unimodal heartbeat data. Our temporal neural network architecture outperformed previous models on the AMIGOS [40] and DREAMER [43] datasets. We further proposed a Bayesian framework for modelling uncertainty over emotional state predictions, providing a means to tune confidence requirements for different tasks. That model accuracy improved with increasing certainty threshold, α , illustrates that probabilistic modelling meaningful impacts performance. Taken together, the component parts of this study represent an important step towards application of affective computing in real-world settings, and provide a probabilistic standard for future work.

ACKNOWLEDGMENTS

The authors would like to thank Nadia Berthouze and Bjrn Schuller for helpful discussions.

REFERENCES

- [1] A. Öhman, D. Lundqvist, and F. Esteves, "The face in the crowd revisited: A threat advantage with schematic stimuli," *Journal of Personality and Social Psychology*, vol. 80, no. 3, pp. 381–396, 2001.
- [2] Y. Bar-Haim, D. Lamy, L. Pergamin, M. J. Bakermans-Kranenburg, and M. H. Van Ijzendoorn, "Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study," *Psychological Bulletin*, vol. 133, no. 1, pp. 1–24, 2007.
- [3] T. Brosch, G. Pourtois, D. Sander, and P. Vuilleumier, "Additive effects of emotional, endogenous, and exogenous attention: Behavioral and electrophysiological evidence," *Neuropsychologia*, vol. 49, no. 7, pp. 1779–1787, 2011.
- [4] F. Dolcos, K. S. LaBar, and R. Cabeza, "Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events," *Neuron*, vol. 42, no. 5, pp. 855–863, 2004.
- [5] E. A. Phelps, "Human emotion and memory: Interactions of the amygdala and hippocampal complex," *Current Opinion in Neurobiology*, vol. 14, no. 2, pp. 198–202, 2004.
- [6] T. Sharot, M. R. Delgado, and E. A. Phelps, "How emotion enhances the feeling of remembering," *Nature Neuroscience*, vol. 7, no. 12, pp. 1376–1380, 2004.
- [7] E. A. Phelps and T. Sharot, "How (and why) emotion enhances the subjective sense of recollection," *Current Directions in Psychological Science*, vol. 17, no. 2, pp. 147–152, 2008.
- [8] E. A. Phelps, S. Ling, and M. Carrasco, "Emotion facilitates perception and potentiates the perceptual benefits of attention," *Psychological Science*, vol. 17, no. 4, pp. 292–299, 2006.
- [9] T. Brosch, G. Pourtois, and D. Sander, "The perception and categorisation of emotional stimuli: A review," *Cognition and Emotion*, vol. 24, no. 3, pp. 377–400, 2010.
- [10] S. Spence, "Descartes' Error: Emotion, Reason and the Human Brain," *BMJ*, vol. 310, pp. 1213–1213, 1995.
- [11] A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio, "Deciding advantageously before knowing the advantageous strategy," *Science*, vol. 275, no. 5304, pp. 1293–1295, 1997.
- [12] A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio, "The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers," *Trends in Cognitive Sciences*, vol. 9, no. 4, pp. 159–162, 2005.
- [13] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.
- [14] A. Hanjalic and L. Q. Xu, "Affective video content representation and modeling," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 143–154, 2005.
- [15] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 154–154, 2004.
- [16] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.
- [17] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.
- [18] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendenmuth, and G. Rigoll, "Cross-Corpus acoustic emotion recognition: Variances and strategies," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 119–131, 2010.
- [19] T. Polzehl, A. Schmitt, F. Metze, and M. Wagner, "Anger recognition in speech using acoustic and linguistic cues," *Speech Communication*, vol. 53, no. 9–10, pp. 1198–1209, 2011.
- [20] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Communication*, vol. 53, no. 9–10, pp. 1062–1087, 2011.
- [21] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, K. Scherer, M. Tomita, and A. Tzavaras, "Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion," *Journal of Personality and Social Psychology*, vol. 53, no. 4, pp. 712–717, 1987.
- [22] K. R. Scherer, R. Banse, and H. G. Wallbott, "Emotion inferences from vocal expression correlate across languages and cultures," *Journal of Cross-Cultural Psychology*, vol. 32, no. 1, pp. 76–92, 2001.
- [23] S. E. Stahl, H.-S. An, D. M. Dinkel, J. M. Noble, and J.-M. Lee, "How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough?," *BMJ Open Sport & Exercise Medicine*, vol. 2, 2016.
- [24] M. P. Wallen, S. R. Gomersall, S. E. Keating, U. Wisløff, and J. S. Coombes, "Accuracy of heart rate watches: Implications for weight management," *PLoS ONE*, vol. 11, pp. 1–11, 2016.
- [25] G. H. Tison, J. M. Sanchez, B. Ballinger, A. Singh, J. E. Olgin, M. J. Fletcher, E. Vittinghoff, E. S. Lee, S. M. Fan, R. A. Gladstone, C. Mikell, N. Sohoni, J. Hsieh, and G. M. Marcus, "Passive detection of atrial fibrillation using a commercially available smartwatch," *JAMA Cardiology*, vol. 3, no. 5, pp. 409–416, 2018.
- [26] M. L. Phillips, W. C. Drevets, S. L. Rauch, and R. Lane, "Neurobiology of emotion perception I: The neural basis of normal emotion perception," *Biological psychiatry*, vol. 54, no. 5, pp. 504–514, 2003.
- [27] T. Dalgleish, "The emotional brain," *Nat Rev Neurosci*, vol. 5, no. July, pp. 252–256, 2004.
- [28] R. L. Redondo, J. Kim, A. L. Arons, S. Ramirez, X. Liu, and S. Tonegawa, "Bidirectional switch of the valence associated with a hippocampal contextual memory engram," *Nature*, vol. 513, 2014.
- [29] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–1210, 1983.
- [30] S. D. Kreibitz, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394–421, 2010.
- [31] W. James, "What is an Emotion?," *Mind*, vol. 9, no. 34, pp. 188–205, 1884.
- [32] W. B. Cannon, "Bodily Changes in Pain, Hunger, Fear and Rage," in *Southern Medical Journal*, p. 311, 1929.

- [33] E. J. Weber, P. C. Molenaar, and M. W. van der Molen, "A Nonstationarity Test for the Spectral Analysis of Physiological Time Series with an Application to Respiratory Sinus Arrhythmia," *Psychophysiology*, vol. 29, no. 1, pp. 55–65, 1992.
- [34] K. Sunagawa, T. Kawada, and T. Nakahara, "Dynamic nonlinear vagosympathetic interaction in regulating heart rate," *Heart and Vessels*, vol. 13, no. 4, pp. 157–174, 1998.
- [35] A. E. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. Clifton, and G. D. Clifford, "Machine Learning and Decision Support in Critical Care," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 444–466, 2016.
- [36] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: A review," in *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, pp. 410–415, 2011.
- [37] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
- [38] O. Alzoubi, S. K. D'Mello, and R. A. Calvo, "Detecting naturalistic expressions of nonbasic affect using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 298–310, 2012.
- [39] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, "An accurate emotion recognition system using ECG and GSR signals and matching pursuit method," *Biomedical Journal*, vol. 40, no. 6, pp. 355–368, 2017.
- [40] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Transactions on Affective Computing*, vol. PP, 2017.
- [41] R. Subramanian, S. Member, J. Wache Student Member, M. Khomami Abadi, S. Member, R. L. Vieri, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and Personality Recognition using Commercial Sensors," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2018.
- [42] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2012.
- [43] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018.
- [44] H. W. Guo, Y. S. Huang, C. H. Lin, J. C. Chien, K. Haraikawa, and J. S. Shieh, "Heart Rate Variability Signal Features for Emotion Recognition by Using Principal Component Analysis and Support Vectors Machine," in *Proceedings - 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering, BIBE 2016*, pp. 274–277, 2016.
- [45] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: A personalized assessment based on heartbeat dynamics," *Scientific Reports*, vol. 4, pp. 1–13, 2014.
- [46] C. A. Torres-Valencia, H. F. Garcia-Arias, M. A. Lopez, and A. A. Orozco-Gutierrez, "Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models," in *2014 19th Symposium on Image, Signal Processing and Artificial Vision, STSIVA*, pp. 1–5, 2014.
- [47] H. F. García, M. A. Álvarez, and Á. A. Orozco, "Gaussian process dynamical models for multimodal affect recognition," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 850–853, IEEE, 2016.
- [48] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.
- [49] S. Alhagry, A. Aly, and R. A., "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.
- [50] G. Keren, T. Kirschstein, E. Marchi, F. Ringeval, and B. Schuller, "End-to-end learning for dimensional emotion recognition from physiological signals," in *Proceedings - IEEE International Conference on Multimedia and Expo*, pp. 985–990, 2017.
- [51] V. Kolodyazhnyi, S. D. Kreibig, J. J. Gross, W. T. Roth, and F. H. Wilhelm, "An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions," *Psychophysiology*, vol. 48, no. 7, pp. 908–922, 2011.
- [52] H. Ferdinando, T. Seppanen, and E. Alasaarela, "Comparing features from ECG pattern and HRV analysis for emotion recognition system," in *CIBCB 2016 - Annual IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1–6, 2016.
- [53] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452–459, 2015.
- [54] R. M. Neal, *Bayesian Learning for Neural Networks*. 1996.
- [55] D. J. MacKay, *Bayesian Methods for Adaptive Models*. PhD thesis, California Institute of Technology, 1992.
- [56] M. Welling and Y.-W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *ICML Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 681–688, 2011.
- [57] L. Hasenclever, S. Webb, T. Lienart, S. Vollmer, B. Lakshminarayanan, C. Blundell, and Y. W. Teh, "Distributed Bayesian Learning with Stochastic Natural-gradient Expectation Propagation and the Posterior Server," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 3744–3780, 2017.
- [58] A. Graves, "Practical Variational Inference for Neural Networks," in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2348–2356, 2011.
- [59] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of the 33rd International Conference on Machine Learning (ICML-16)*, pp. 1050–1059, 2016.
- [60] I. Sutskever, G. Hinton, A. Krizhevsky, and R. R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [61] J. D. Morris, "SAM: The Self-Assessment Manikin - An efficient cross-cultural measurement of emotional response," *Journal of Advertising Research*, vol. 35, no. 6, pp. 63–68, 1995.
- [62] C. A. Gabert-Quillen, E. E. Bartolini, B. T. Abravanel, and C. A. Sanislow, "Ratings for emotion film clips," *Behavior Research Methods*, vol. 47, no. 3, pp. 773–787, 2015.
- [63] I. I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold," *BioMedical Engineering Online*, vol. 3, no. 1, p. 28, 2004.
- [64] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
- [65] P. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations*, 2014.
- [66] GoogleResearch, "TensorFlow: Large-scale machine learning on heterogeneous systems," *Google Research*, 2015.
- [67] J. Pietilä, S. Mehrang, J. Tolonen, E. Helander, H. Jimison, M. Pavel, and I. Korhonen, "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities," in *IFMBE Proceedings*, 2017.
- [68] J. Harju, A. Tarniceriu, J. Parak, A. Vehkaoja, A. Yli-Hankala, and I. Korhonen, "Monitoring of heart rate and inter-beat intervals with wrist plethysmography in patients with atrial fibrillation," *Physiological Measurement*, vol. 39, no. 6, 2018.

Ross Harper completed a PhD in computational neuroscience and an MRes in mathematical modelling at University College London. Prior to this, Ross studied natural sciences at Cambridge University. Ross is the CEO of Limbic Ltd, a London start-up developing emotion recognition technologies. Ross is the recipient of the Lloyds Bank New Entrepreneur of the Year 2018 award. His current research interests include, affective computing, probabilistic modelling of biometric time series data, and wearable computing.

Joshua Southern obtained an Msc in applied mathematics from Imperial College London, and a Bsc in physics from Bath University. He is a machine learning researcher at Limbic Ltd, and previously completed an internship at IBM. His research interests include deep learning, time series analysis, and complex networks.