
Bayesian Active Learning for Wearable Stress and Affect Detection

Abhijith Ragav*
Solarillion Foundation
abhijithragav@ieee.org

Gautham Krishna Gudur*
Global AI Accelerator, Ericsson
gautham.krishna.gudur@ericsson.com

Abstract

In the recent past, psychological stress has been increasingly observed in humans, and early detection is crucial to prevent health risks. Stress detection using on-device deep learning algorithms has been on the rise owing to advancements in pervasive computing. However, an important challenge that needs to be addressed is handling unlabeled data in real-time via suitable ground truthing techniques (like Active Learning), which should help establish affective states (labels) while also selecting only the most informative data points to query from an oracle. In this paper, we propose a framework with capabilities to represent model uncertainties through approximations in Bayesian Neural Networks using Monte-Carlo (MC) Dropout. This is combined with suitable acquisition functions for active learning. Empirical results on a popular stress and affect detection dataset experimented on a Raspberry Pi 2 indicate that our proposed framework achieves a considerable efficiency boost during inference, with a substantially low number of acquired pool points during active learning across various acquisition functions. Variation Ratios achieves an accuracy of 90.38% which is comparable to the maximum test accuracy achieved while training on about 40% lesser data.

1 Introduction

Psychological stress is encountered by humans in response to physical, mental or emotional challenges presented to them by the environment. Physiological signals can be easily measured using wearable devices and are good indicators of stress, making them suitable for continuous stress monitoring. Deep learning architectures have recently performed well for such tasks, due to their automatic feature extraction capabilities, in contrast to conventional machine learning models which mandate domain knowledge to craft shallow heuristic features [11], [2]. There has been a special interest in bringing deep learning to wearable devices to provide real-time stress and affect monitoring, whilst preserving sensitive user data [13].

Typically, we assume that the physiological user data is inherently labeled, however it is almost practically never the case. One of the least explored areas involving deep learning for such tasks is *Active Learning* – a technique which gives a model, the ability to learn from real-world unlabeled data by querying an oracle (user). By using *Bayesian Neural Networks (BNNs)*, which integrates *Monte Carlo Dropout* with traditional neural networks [4], it is possible to estimate predictive uncertainties. This is coupled with active learning *acquisition functions* for querying the most uncertain data points from the oracle [5]. However, these works have not been considerably discussed in on-device mobile health scenarios, particularly for stress and affect detection. More related work on stress and affect detection, BNNs and active learning can be found in Appendix A.

Scientific contributions: (1) A study of Bayesian deep neural networks for real-time stress and affect detection. (2) Leveraging the benefits of *Bayesian Active Learning* to model uncertainties, and

*Both authors contributed equally to this work.

exploiting several acquisition functions on-device to instantaneously acquire ground truths (affective states) on-the-fly, thereby substantially reducing the labeling load on oracle.

2 Our Approach

In this section, we discuss in detail our proposed system and approach to perform Active Learning with physiological data.

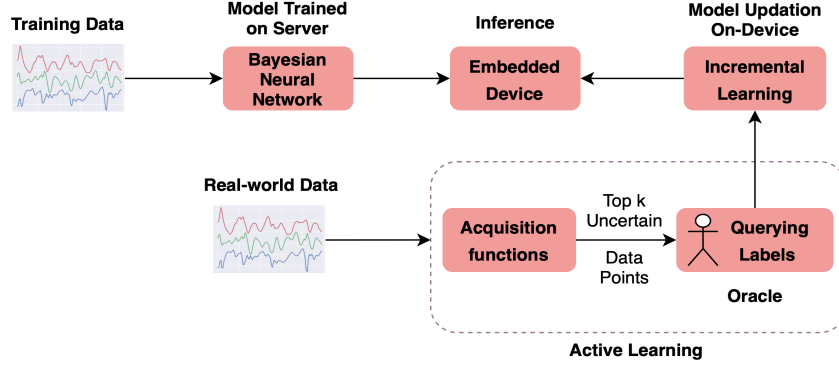


Figure 1: Proposed Architecture

2.1 Background on Modeling Uncertainties

Bayesian Neural Networks (BNNs) offer a probabilistic interpretation to deep learning models by incorporating Gaussian prior (probability distributions) – $p(\omega)$ over our model parameters (weights – ω), thereby modeling output uncertainties. The likelihood model for a classification setting with c classes and x input points is given by,

$$p(y = c|x, \omega) = \text{softmax}(f^\omega(x))$$

where $f^\omega(x)$ is the model output. However, the posterior distribution of BNNs are not easily tractable, hence it becomes computationally intensive for training and inference.

Gal et al. propose that, *Dropout* – a stochastic regularization technique [17], can also perform approximate inference over a deep Gaussian process [4], and thereby learn to model posterior uncertainties without high computational complexities. This is equivalent to performing Variational Inference (VI), where the posterior distribution is approximated by finding another distribution $q_\theta^*(\omega)$, parameterized by θ , within a family of simplified tractable distributions, while minimizing the Kullback-Leibler (KL) divergence between $q_\theta^*(\omega)$ and the true model posterior $p(\omega|D_{train})$.

During inference, we can estimate the mean and variance of the BNN’s output by applying dropout before every fully-connected layer during train and test time for multiple stochastic passes (T). This is equivalent to obtaining predictions and uncertainty estimates respectively from the approximate posterior output of the neural network, thereby making the Bayesian NN *non-deterministic* [4]. The predictive distribution for a new data point input x^* can be obtained by,

$$p(y^*|x^*, D_{train}) = \int p(y^*|x^*, \omega)p(\omega|D_{train})d\omega$$

where $p(\omega|D_{train}) = q_\theta^*(\omega)$, and $q_\theta^*(\omega)$ is the dropout distribution approximated using VI. Dropout, being a light-weight operation in most existing NN architectures, enables easier and faster approximation of posterior uncertainties.

2.2 Model Architecture

In our model, 1D-Convolutional (Conv1D) layers are used for effective temporal extraction of features from physiological data. We use a four-layer Conv1D network with 4, 8, 16 and 32 filters, kernel size 3, and a max-pooling layer of size 2 between each Conv1D layer. This is followed by two

Fully-Connected (FC) layers with 32 and 16 neurons each and ReLU activations. An MC-dropout layer with a probability of 0.3 is applied, followed by a softmax layer to obtain the probability scores. The categorical-cross entropy loss of the model is minimized using Adam optimizer.

In order to make our model a Bayesian NN so as to obtain uncertainty estimates, we introduce a standard Gaussian prior on the set of our model parameters. Also, to perform approximation inference in our model, we perform dropout at train and test-time as discussed in Section 2.1 to sample from the approximate posterior using multiple stochastic forward passes [4]. After experimenting with multiple dropout iterations (forward passes – T), an optimal $T=10$ is utilized in this paper to determine uncertainties. Effectively, our uncertainty-aware model now is a *Bayesian ensembled Convolutional Neural Network (B-CNN)* which can model uncertainties and be used with existing acquisition functions for AL.

2.3 Acquisition functions for Active Learning

As stated in [5], given a classification model M , pool data D_{pool} obtained from real-world, and inputs $x \in D_{pool}$, an acquisition function $a(x, M)$ is a function of x that the active learning system uses to infer the next query point:

$$x^* = \operatorname{argmax}_{x \in D_{pool}} a(x, M).$$

Acquisition functions are used in active learning scenarios for approximations in Bayesian CNNs, thereby arriving at the most efficient set of data points to query from D_{pool} . We examine the following acquisition functions to determine the most suitable function for on-device computation:

Max Entropy: Pool points are chosen that maximize the predictive entropy [15].

$$\mathbb{H}[y|x, D_{train}] := - \sum_c p(y = c|x, D_{train}) \log p(y = c|x, D_{train})$$

Bayesian Active Learning by Disagreement (BALD): In BALD, pool points are chosen that maximize the mutual information between predictions and model posterior [8]. The points that maximize the acquisition function are the points that the model finds uncertain on average, and information about model parameters are maximized under the posterior that disagree the most about the outcome.

$$\mathbb{I}[y, \omega|x, D_{train}] = \mathbb{H}[y|x, D_{train}] - E_{p(\omega|D_{train})} [\mathbb{H}[y|x, \omega]]$$

where $\mathbb{H}[y|x, \omega]$ is the entropy of y , given model weights ω .

Variation Ratios (VR): The LC (Least Confident) method for uncertainty based pool sampling is performed in VR [3].

$$\text{variation} - \text{ratio}[x] := 1 - \max_y p(y|x, D_{train})$$

Random Sampling: This acquisition function is equivalent to selecting a point from a pool of data points uniformly at random.

3 Experiments and Results

SWELL-KW Dataset [10]: The dataset consists of stress data from 25 participants recorded when they were performing knowledge work tasks such as writing reports, making presentations, reading emails, and searching for information. Stress was induced in the participants by stressors such as time-pressure and email interruptions, and their physiological modalities (Heart Rate and Skin Conductance) were recorded. Based on the participants' self-reports, 3 mental conditions were analyzed – Neutral (N), Interruption (I) and Time-Pressure (T). We report results for the N vs I&T classification task.

The train samples are split in random into D_{train} and D_{pool} points, (30-70 ratio) as an approximation of real-world incoming data, while the unseen D_{test} is used for evaluation purposes only in our experiment. To evaluate our Bayesian CNN framework, we initially pre-train our model for a maximum of 10 epochs on a server to establish the baseline efficiencies on the SWELL dataset, and stock the model in the embedded system. To approximate our posterior and obtain predictive uncertainties, we test our BNN model over $T=10$ stochastic iterations, and average our predictions to calculate our final efficiencies. The average initial baseline accuracy without any active learning incorporated is observed to be 79.12%.

3.1 Bayesian Active Learning

In order to handle ground truth labeling and incorporate model weight updation on incoming test user data, we experiment our proposed framework by deploying the system on a Raspberry Pi 2. The number of acquisition windows used for active learning from D_{pool} can be governed by the **acquisition adaptation factor** $\eta \in [0, 1]$. The incremental model updation is simulated for a maximum of 10 epochs for every acquisition iteration, owing to its non-convergence in loss thereafter.

We analyze various AL acquisition functions mentioned in section 2.3, and observe that Variation Ratios (VR) acquisition function performs the best, while Random Sampling has the least classification accuracy as expected. In VR , only 60% ($\eta=0.6$) of the acquisition windows from D_{pool} are required to achieve a test accuracy of 90.38% from a baseline accuracy of 79.12%. This is very close to the maximum test accuracy of 91.92% achieved with $\eta=1.0$ (all D_{pool} windows). Note that, $\eta = 0.0$ gives the efficiency of the pre-trained model without any data points acquired during incremental learning.

Table 1: *Bayesian-CNN* with *SWELL* – Acquisition Windows (η) vs Accuracy (%).

η	Max Entropy	BALD	Variation Ratios	Random Sampling
0.0	79.12	79.12	79.12	79.12
0.2	82.66	81.21	83.11	81.91
0.4	86.43	86.58	88.29	86.76
0.6	89.40	90.22	90.38	88.19
0.8	89.98	90.63	90.82	88.95
1.0	91.92	91.92	91.92	91.92

3.2 On-Device Performance

Raspberry Pi 2 is used for evaluating our proposed active learning framework as it has similar hardware and software specifications to predominant contemporary IoT devices. We observe an average time of about 0.5 sec is utilized for each stochastic forward pass (T) with dropout for acquisition of top η windows. Since we perform $T=10$ dropout iterations in our experiment, we observe that an average of about 5 seconds are needed for querying the most uncertain data points.

Table 2: Computation Time Taken for Execution per window

Process	Time
Inference time	9 ms
Time taken per epoch	0.6 sec

The model size for *SWELL* is approximately 115 kB, which is substantially small compared to conventional deep learning models. For real-time deployment feasibility, it is practical to have a threshold (upper limit) on the number of D_{pool} collected in a single acquisition iteration. This can be quantified by either number of windows (w_a) or time taken (in seconds).

4 Conclusion

This paper presents empirical contributions with emphasis on Bayesian Active Learning for efficient stress and affect detection. First, we benchmark our efficiencies on the *SWELL* dataset using a Bayesian-CNN model with stochastic dropout, which incorporates information on predictive uncertainties. Second, we perform active learning on-device using the uncertainty approximations by systematically analyzing various acquisition functions for extracting the most informative data points to be queried by the oracle. This can further be extended to other behavior monitoring tasks in mobile and pervasive healthcare. We also believe this work would trigger more research in under explored areas like handling unlabeled data, and active learning for mobile health scenarios.

References

- [1] Sourav Bhattacharya, Petteri Nurmi, Nils Hammerla, and Thomas Plötz. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing*, pages 242–262, 2014.
- [2] Flavio Di Martino and Franca Delmastro. High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks. In *2020 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–6. IEEE, 2020.
- [3] Linton C Freeman. *Elementary Applied Statistics: For Students in Behavioral Science*. John Wiley & Sons, 1965.
- [4] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML’16*, pages 1050–1059, 2016.
- [5] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML’17*, pages 1183–1192, 2017.
- [6] Gautham Krishna Gudur, Prahalathan Sundaramoorthy, and Venkatesh Umaashankar. Active-harnet: Towards on-device deep bayesian active learning for human activity recognition. In *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications, EMDL’19*, page 7–12, 2019.
- [7] H. M. S. Hossain, N. Roy, and M. A. A. H. Khan. Active learning enabled activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9, 2016.
- [8] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.
- [9] Fitri Indra Indikawati and Sri Winiarti. Stress detection from multimodal wearable sensor data. In *IOP Conference Series: Materials Science and Engineering*, page 012028. IOP Publishing, 2020.
- [10] Saskia Koldijk, Maya Sappelli, Suzan Verberne, Mark A Neerincx, and Wessel Kraaij. The swell knowledge work dataset for stress and user modeling research. In *Proceedings of the 16th international conference on multimodal interaction*, pages 291–298, 2014.
- [11] Dušan Marković, Dejan Vujičić, Dijana Stojić, Željko Jovanović, Uroš Pešović, and Siniša Randić. Monitoring system based on iot sensor data with complex event processing and artificial neural networks for patients stress detection. In *18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–6. IEEE, 2019.
- [12] Rajdeep Kumar Nath, Himanshu Thapliyal, Allison Caban-Holt, and Saraju P Mohanty. Machine learning based solutions for real-time stress monitoring. *IEEE Consumer Electronics Magazine*, 2020.
- [13] Abhijith Ragav, Nanda Harishankar Krishna, Naveen Narayanan, Kevin Thelly, and Vineeth Vijayaraghavan. Scalable deep learning for stress and affect detection on resource-constrained devices. In *18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1585–1592. IEEE, 2019.
- [14] Burr Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2012.
- [15] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.
- [16] Yanyao Shen, Hyokun Yun, Zachary Lipton, Yakov Kronrod, and Animashree Anandkumar. Deep active learning for named entity recognition. *Proceedings of the 2nd Workshop on Representation Learning for NLP*, 2017.

- [17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

A Appendix: Related Work

Technological advancements in pervasive and ubiquitous healthcare have drastically improved the quality of human life, and hence has been an actively explored research area. Detecting stress in real-time is crucial for continuous monitoring of mental health. Deep learning using sensor data has been an evolving domain for computational behavior analysis and healthcare research. Efficient deep learning for stress and affect detection has become the need of the hour, and such approaches have been widely used with promising empirical results by effectively capturing the most discriminative features [12], [9]. However, these works do not address the important problem of concept drift, which must be handled by techniques like Active Learning (AL) to combat changes in data distribution over time.

Conventional AL literature [14] mostly handle low-dimensional data for uncertainty estimations, but do not generalize to deep neural networks as the data is inherently high-dimensional [5]. Deep active learning using uncertainty representations – which presently are the state-of-the-art AL techniques for high-dimensional data, have had very sparse literature. With the advent of Bayesian approaches to deep learning, Gal et al. [5] proposed Bayesian Active Learning for image classification tasks, and is proven to learn from small amounts of unseen data, while Shen et al. [16] incorporate similar techniques for NLP sequence tagging. However, these techniques are predominantly not discussed for sensor and time-series data.

Incorporating AL for obtaining ground truth in mobile sensing systems has been addressed in a few previous works. Most of them have been addressed with Human Activity Recognition (HAR) tasks [1], [6], [7]. Although these works seem to achieve impressive results, the feasibility of on-device learning in mobile health stress and affect detection scenarios with unlabeled data still have not been explored.