

# Demo: On-device Puff Detection System for Smoking Cessation

Shalini Mukhopadhyay Swarnava Dey Avik Ghose  
 {shalini.mukhopadhyay,swarnava.dey,avik.ghose}@tcs.com  
 TCS Research, Kolkata, India

## ABSTRACT

Customized, on-device applications that provide timely interventions about smoking episodes are very helpful for smoking cessation. For this, real-time detection of smoking puffs are necessary through unobtrusive wearable devices. This work demonstrates auto-generated tiny puff detection models for on-device inference on low-power wearable devices.

## CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile devices**.

## KEYWORDS

Wearable Sensing, Smoking cessation, Edge Computing, NAS

### ACM Reference Format:

Shalini Mukhopadhyay Swarnava Dey Avik Ghose. 2023. Demo: On-device Puff Detection System for Smoking Cessation. In *The 21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23)*, June 18–22, 2023, Helsinki, Finland. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3581791.3597284>

## 1 INTRODUCTION

Cigarette and Tobacco Smoking is a very prominent cause of death and illness around the world. It affects both active as well as passive smokers. Thus, smoking cessation is a necessary wellness requirement due to the range of related illnesses from cardiopulmonary diseases to cancer. Customized applications can help users who wish to quit smoking by timely intervention. This requires detection of smoking puffs on-device and in real-time, through round-the-clock monitoring on unobtrusive devices like wearables. The puff counts can be stored for episode estimation and cigarette pack computation and used later for analysis and design of personalized aid systems for users over a more extended period. The idea overview is presented in Fig. 1. There are various techniques for smoking activity detection from inertial sensors e.g., Accelerometer and Gyroscope, and physiological sensors e.g., Photoplethysmography (PPG) sensor from wearable devices. Some solutions such as [4] use Respiratory Inductance Plethysmography (RIP) and data from additional devices such as smart lighters for smoking episode detection. Some solutions use time series classification techniques such as Long Short Time Memory (LSTM) [7] or Squeeze and Excitation (SNE) [6] for better results. These models are complex and manually designed

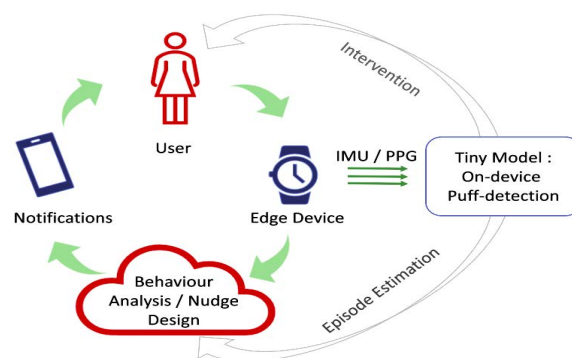


Figure 1: Overview of the Puff Inference and Intervention [6]

by domain experts and it is not trivial, more so when the target application runs continuously on a battery-powered device. The aim is to optimize the operations, parameters, and training hyperparameters in an automated way to look for globally optimal models and minimize the time in a manual process. There are well-known techniques on neural network architecture search (NAS) such as Evolutionary Algorithms [9], or Differential Search (DARTS) [10] techniques. We have adopted the RL-based technique [3] for edge devices and modified it for the puff classification task. In our work, we reduce the models to smaller, sequential, and efficient architectures with basic operations of a Convolutional Neural Network (CNN). We also present two versions of the model, one dependent on IMU and RIP / PPG signals, and the other with only IMU signals, generally available on most wearable devices. This eliminates the need for PPG signal since its variation and corruption across users and activities result in a less robust outcome for smoking event detection. Since microcontroller-based wearables can afford only small IMU models for 24 x 7 use, this makes the system suitable for users throughout the day.

In this work, we present two instances of auto-generated, CNN-

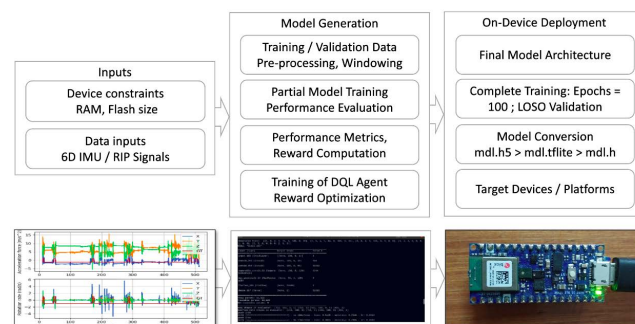


Figure 2: Visualization of the complete System

based classifiers implemented on a tiny device, for detecting smoking activity in real time. The specifications of our selected device

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
 MobiSys '23, June 18–22, 2023, Helsinki, Finland  
 © 2023 Copyright held by the owner/author(s).  
 ACM ISBN 979-8-4007-0110-8/23/06.  
<https://doi.org/10.1145/3581791.3597284>

represent a limited-resource and affordable wearable device. The models are compatible with TensorFlow Lite and TensorFlow Lite Micro (TFLM) and can be deployed on a range of tiny machine-learning microcontroller platforms. Moreover, due to the model generation process being automated, it can be re-used with a different set of constraints to fit any target device and platform. The overview of our demonstration system is depicted in Fig. 2.

## 2 PUFF DETECTION SYSTEM ON EDGE DEVICE: SYSTEM & DEMO

We demonstrate the automated model generation technique for the puff classification task, along with two of our Tiny CNN-based puff classification models [6] deployed on-device, to visualize the detected output in real-time. This will help the user by sending immediate notifications for timely intervention. Latency and power

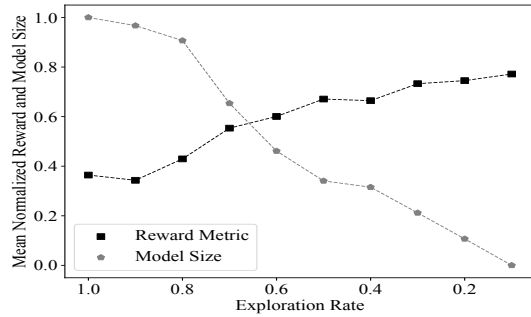


Figure 3: Reward and Model size optimization for Puff detection from IMU+RIP signals [6]

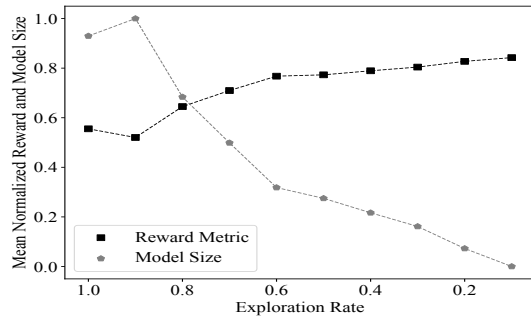


Figure 4: Reward and Model size optimization for Puff detection from IMU signals [11]

are estimated using the substitute measurement of Multiply Accumulate operations (MACs) which depicts a high correlation with the two aforementioned metrics ( $r=0.9946$ ), as discussed in detail in previous work [2]. Since latency is also related to the number of layers in the network architecture, we have limited our models to a maximum of 5 layers of individual operations. Our auto-generated models produce comparable results with domain expert-designed, state-of-the-art models, although tinier by a large factor. This is shown in Table 1. The model generation and optimization process is based on an RL-based NAS Technique, with the epsilon-greedy policy. The improvement of reward and reduction in size with the progress of episodes and decrease in epsilon (exploration rate)

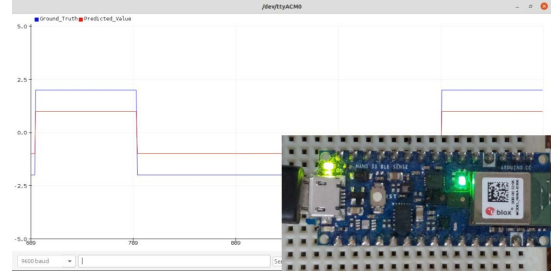


Figure 5: On-device Puff Inference and Actuation System

for puff activity detection with IMU + RIP data [5] and IMU-only data [8], are depicted in Fig. 3 and Fig. 4 respectively.

Table 1: Comparison with State-of-the-Art model Performance

Model	Technique	Input	F1 Score	Size(kB)
Manual [6]	CNN+SNE	RIP+IMU	0.81	472.8
NAS [6]	Tiny CNN	RIP+IMU	0.75	111.77
Manual [7]	CNN+LSTM	IMU	0.86	669.9
NAS [11]	Tiny CNN	IMU	0.81	114.07

Our on-device execution is on an Arduino Nano 33 BLE [1] board with a limited on-board RAM of 256 kB and Flash size of 1 MB. The deployed models performed a maximum of 4M Multiply-Accumulate operations. The model, along with the TensorFlow Lite Micro runtime overhead, utilized 186.5 kB of available flash memory and a peak SRAM usage of 231.3 kB, with a latency of 2.1 s for a 10 s window. We used the onboard *Green LED* (Fig. 5) to indicate the detection of the Puff activity.

## REFERENCES

- [1] Arduino.cc. 2023 [Online]. Nano 33 BLE Sense. (2023 [Online]). <https://docs.arduino.cc/hardware/nano-33-ble-sense>
- [2] Colby R. Banbury et al. 2020. MicroNets: Neural Network Architectures for Deploying TinyML Applications on Commodity Microcontrollers. *CoRR* abs/2010.11267 (2020).
- [3] S. Mukhopadhyay et al. 2023a. Generating Tiny Deep Neural Networks for ECG Classification on Micro-controllers. In *IEEE International Conference on Pervasive Computing and Communications: PerCom Industry Track*.
- [4] Volkan Senyurek et al. 2019a. Cigarette smoking detection with an inertial sensor and a smart lighter. *Sensors* 19, 3 (2019), 570.
- [5] Volkan Senyurek et al. 2019b. In-smoking respiration and hand gesture IMU signals. (2019). DOI: <http://dx.doi.org/10.21227/7d8x-yr60>
- [6] V. Sharma et al. 2023b. PuffConv: A System for Online and On-device Puff Detection for Smoking Cessation. In *Ninth Workshop on sensing systems and applications using wrist worn smart devices (WristSense)*.
- [7] Athanasios Kirmizis, Konstantinos Kyritsis, and Anastasios Delopoulos. 2021a. A Bottom-up method Towards the Automatic and Objective Monitoring of Smoking Behavior In-the-wild using Wrist-mounted Inertial Sensors. In *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 6867–6870.
- [8] A. Kirmizis, K. Kyritsis, and A. Delopoulos. 2021b. Wrist-mounted IMU data towards the investigation of free-living smoking behavior - the Smoking Event Detection (SED) and Free-living Smoking Event Detection (SED-FL) datasets [Data set]. Zenodo. (2021). <https://doi.org/10.5281/zenodo.4507451>
- [9] Edgar Liberis, Łukasz Dudziak, and Nicholas D. Lane. 2021.  $\mu$ NAS: Constrained Neural Architecture Search for Microcontrollers. *Proceedings of the 1st Workshop on Machine Learning and Systems, EuroMLSys* (2021).
- [10] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2019. DARTS: Differentiable architecture search. *7th International Conference on Learning Representations, ICLR* (2019).
- [11] S. Mukhopadhyay, S. Dey, and A. Ghose. 2023. TinyPuff: Automated design of Tiny Smoking Puff Classifiers for Body Worn Devices (In press). (2023). <https://doi.org/10.1145/3597061.3597259>