An Always-On tinyML Acoustic Classifier for Ecological Applications

H.R. Sabbella*, A.R. Nair*, V. Gumme*, S.S. Yadav*, S. Chakrabartty†, C.S. Thakur*
{hemanthrs, abhisheknair, vishnugumme, satyapreets, csthakur}@iisc.ac.in, shantanu@wustl.edu
* Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore, India, 560012
† Department of Electrical and Systems Engineering, Washington University in St. Louis, USA, 63130

***Abstract*—Long-term monitoring and tracking of wildlife and endangered species in their natural environment is challenging due to human factors and logistical limitations. We present a light-weight, always-on acoustic classification system that can identify the density of specific wildlife species in an ecological environment where human presence may be undesirable. The system uses a template-based support-vector-machine (SVM) classifier that combines acoustic filtering and classification into an in-filter computing and a hardware-friendly platform. We demonstrate the system's capabilities for identifying the density of different bird species using ARM Cortex-M4 based AudioMoth hardware. The embedded software, designed specifically for the AudioMoth hardware, can generate the programmable parameters, given limited training samples corresponding to different wildlife species. We show that the system can identify four different bird species with an accuracy of more than 95% and consumes a memory footprint of 14 KB SRAM and 149 KB Flash memory that can run for 48 days on battery without any human intervention.**

***Index Terms*—tinyML, template-SVM, Ecology, acoustic**

I. Introduction

The field of ecology and nature conservation is being revolutionized by Machine Learning (ML) [1]. ML algorithms can prove to be an essential tool in identifying and locating the density of different bird species in environments where temporal and spatial limitations due to human observation pose a challenge [2]. ML has been used in predicting extinction risk of different species using wildlife sensor data [3] [4]. Many initiatives like Wildlife Insights for wildlife conservation have been started, which relies heavily on data gathered from sensors [5]. One such system can be visualized, as shown in Fig. 1, where an ecological conservation protocol can be triggered based on acoustic classification inputs for Bird species conservation or Human-wildlife conflict. In this regard, there have been significant advancements in the detection and classification of birds using sounds, i.e., bird vocalization [6]. Improvements in signal processing and subsequent ML techniques have enabled efficient edge devices to perform bird detection using vocalization sensing [7].

Most of these light-weight ML or tinyML systems use Deep Learning (DL) as the ML technique for bird classification [8] [9]. However, in the case of rare or near-extinct species detection, the available training data is sparse [10]. DL systems do not perform well with sparsely available data. Moreover, using a DL system with limited resources requires quantizations like Binary Neural Networks (BNNs) [11]. Retraining a BNN requires full-precision parameters, which becomes a challenge for IoT-based systems. K-Nearest Neighbour (KNN) [12] [13] and Support Vector Machines (SVMs) [14] [15] have known to perform well with sparse training data. However, SVM is more robust to outliers since only support vectors determine the classification boundary, and noisy outliers don't affect the classification of the system [16] [17]. Also, the convexity of SVM ensures retraining is stable and interpretable [18].

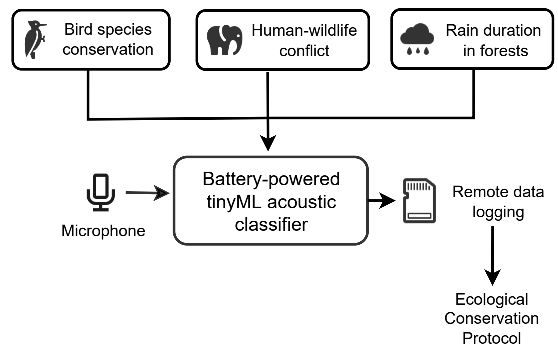


Fig. 1: Ecological Conservation Protocol can be triggered based on inputs from always-On acoustic classifiers

Due to power and resource constraints, more often than not, preprocessing of the sounds is a challenge to implement on tinyML systems. In this paper, we present a bird density detection system using the in-filter classification system described in [19]. Here, the acoustic front-end is the neuromorphic cochlea-based system that acts as the preprocessing unit and also the non-linear SVM kernel. The resulting template-based SVM architecture [20] can be used to identify critically endangered bird species by tuning the Cascade of Asymmetric Resonators with Inner Hair Cells (CAR-IHC) filters [21]. We show the successful deployment of this system on a smart acoustic device, i.e., AudioMoth [22]. This IoT-based system identifies the density of bird species in an area using the classification data from the device and is capable of running on battery power.

II. Design and Methodology

Our system is proven computationally efficient, with user-defined resource constraints leveraging in-filter computation by using CAR-IHC filters, requiring less training data. The

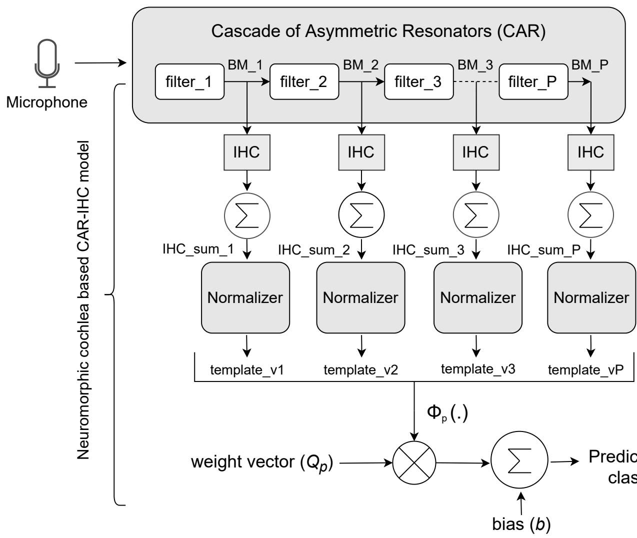


Fig. 2: Functional diagram of template-SVM

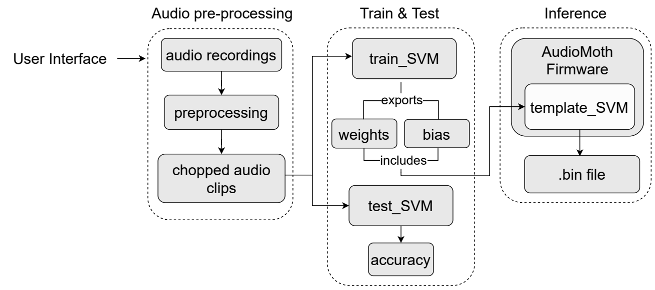


Fig. 3: Software architecture of the proposed framework

CAR-IHC filters acting as feature extractors are computationally less expensive than Mel Frequency Cepstral Coefficients (MFCCs) that typically include log-domain operations [19]. These attributes make it ideal for deployment as a tinyML acoustic classifier for rare species detection. We use AudioMoth hardware and a unique software framework for quick and easy deployment.

A. In-Filter Compute Classification

We have developed a software implementation of in-house designed template-SVM formulation [19]. As shown in Fig. 2, the audio samples are provided to the CAR filters that are tuned to the required application, i.e., bird species identification. Further, the processed samples are passed through IHC filters. The IHC filters are mathematically modeled as half-wave rectifiers. The output of these filters is summed over the audio samples and normalized to get the templates. The template-SVM formulation can be expressed as,

$$f(X_n) = \sum_{p=1}^{P} Q_p \Phi_p(X_n) + b. \tag{1}$$

Here, $X_n \in \mathbb{R}^W$ is constructed using a window $X_n = \{x_n, x_{n-1}, ..., x_{n-W+1}\}$ at a time-instant n for previous W samples of input signal x_n. $Q_p \in \mathbb{R}$ and $b \in \mathbb{R}$ are parameters that are determined by a template-SVM training which is described in [19]. The key advantage of template-SVM training is that the memory and computational footprint of the ML architecture can be constrained by the number and the nature of the template functions $\Phi_p : \mathbb{R}^W \to \mathbb{R}$ with $p = 1, .., P$, while retaining the large-margin properties of conventional SVM training [18]. In this work, the template functions represent P CAR-IHC filters used as feature extractors as shown in Fig. 2

B. AudioMoth as the acoustic classifier device

AudioMoth hardware is an ARM Cortex-M4 microcontroller based system [22] [23]. It has been widely used as a low-cost acoustic data logger for ecological purposes. It has an EFM32 Gecko 32-bit processor with an operating speed of 48 MHz. There is 256 KB of Flash RAM along with 32 KB internal and 256 KB external SRAM. It can store audio recordings of uncompressed WAV files in an SD card which supports up to 128 GB. The software used for programming the hardware is open-sourced and employs the low-level programming language, C. It has internal operational amplifiers to strengthen the analog microphone signal without additional external components and has sufficiently low energy consumption to allow powering by three lithium AA-cell batteries.

C. Software Design and Framework

A bare-metal implementation of template-SVM inference is designed and written in C and is incorporated into AudioMoth firmware.

We have designed an easy to deploy framework as a one-stop solution to incorporate different bird species classification for AudioMoth. As shown in the Fig. 3, the entire framework is run using a single user-friendly interface. The interface reads the audio dataset and chops the audio samples into a user-defined time frame and amplitude threshold. It also has a feature to discard the samples by listening to the chopped audio samples and curates a clean dataset. The chopped audio dataset is segregated into train and test samples. These train audio samples are used to train the template-SVM model using one vs. all training and exports the weights (Q_p), bias (b), and also parameters defined during the model training, i.e, Sampling rate (Fs), Number of filters ($Nfilter$) and Number of classes ($Nclasses$), that are described in header files. Further, the model is tested on the test audio samples, and the accuracy of each bird species is reported. Finally, the trained parameters are included in the AudioMoth firmware which has inference engine programmed on it. The entire firmware is complied, and a binary machine file (.bin file) is generated.

D. System Design

The AudioMoth device can be configured to sample audio up to 384 kHz. The gain of the audio samples is set to be

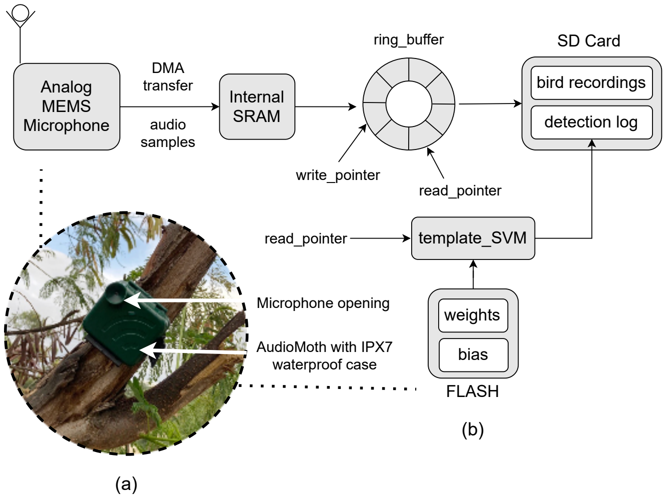


Fig. 4: (a) AudioMoth setup with IPX7 waterproof casing strapped to a tree at IISc for deployment; (b) Data flow diagram in the AudioMoth

32dB, and it is time-synchronized with the local time of the programming computer. Fig. 4(a) shows the deployment setup of AudioMoth in an ecological environment. It is programmed to work for 24-hours with zero-sleep cycles, making it an always-on system.

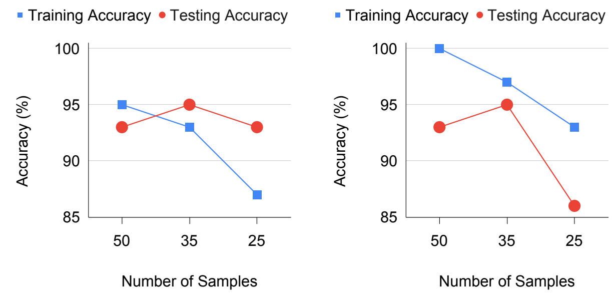
Fig. 4(b) describes the data flow of audio samples and the implementation of template-SVM. Initially, the Analog MEMS Microphone is enabled, and the 16-bit audio samples are transferred through the Direct Memory Access (DMA) to the internal SRAM. Using DMA as a transfer protocol without involving the Cortex-M4 processor during the data flow can optimize the device for low-power consumption. The internal SRAM works as a ping-pong buffer that simultaneously acquires and transfers the audio samples from the microphone and to a $ring\_buffer$ on the external SRAM. The $ring\_buffer$ is an 8-element buffer with a buffer size of 32 KB. Hence it can accommodate 16-kilo samples per second (KS/s). The 16-kilo samples as an audio chunk, i.e., one-second audio, are written into the $ring\_buffer$ using the $write\_pointer$. The $read\_pointer$ points to 16-kilo audio samples and is passed as an input to the template-SVM inference and, the weights and bias are fetched from the FLASH memory. The one-second audio data is processed, and the classification results are dumped into a detection log stored along with time-stamps in the SD card. The processed audio chunk is transferred into the SD card using the $read\_pointer$ based on the classification result, and the $read\_pointer$ is incremented. This data gets processed in real-time as long as $write\_pointer > read\_pointer$ and the classification results are dumped continuously on SD Card. The demonstration of this system is available at [24].

III. Results and Discussion

The template-SVM model is trained on 4-different bird audio samples (Common Cuckoo, White Throated Kingfisher, Asian Koel, and Red Wattled lapwing). All the recordings are sampled at 8 kHz and trained on 100 samples. TABLE I shows the accuracy values of both training and testing. The template-SVM model is able to classify the different species with accuracies greater than 95%.

TABLE I: Accuracy results in % of 4 different bird species using 30 filters ($Nfilter$ = 30)

Bird Species	Training accuracy	Testing accuracy
Common Cuckoo	100	97
White Throated Kingfisher	98	95
Asian Koel	99	97
Red Wattled lapwing	97	95



(a) Forest Owlet (b) Spoon Billed Sandpiper

Fig. 5: Accuracy vs Number of Samples for endangered species

With the minimal data samples, template-SVM can perform very well. This could be used in the case of critically endangered bird species where the recordings of those birds are limited. We have selected Forest Owlet and Spoon Billed Sandpiper, which are tagged as globally threatened birds, and their audio samples are very limited. We have collected only available audio recordings of Spoon Billed Sandpiper [25] [26] and Forest Owlet [27] [28] from xeno-canto and e-bird community, the global repositories for bird sounds. Out of these recordings, we have extracted 50 clean audio samples of both the birds using the proposed framework. Fig. 5 shows the accuracy of the model against the number of samples available for the training. The template-SVM performs significantly well with a limited number of audio samples in the training dataset. We have incorporated an optional noise class for segregating the environmental-specific noise for better robustness on the hardware to eliminate misclassification.

Often in real-world scenarios, the SNR of the bird chirps depends on the distance the AudioMoth is acquiring the sound signatures. To understand the system's coverage area, we have analyzed the maximum distance at which the system is able to detect the bird sounds. It is practically observed that the system could classify sounds up to 13 meters radius. Multiple systems are used across the hotspot region of endangered species to cover larger areas.

The entire system, including template-SVM inference, consumes a small memory footprint of 14 KB SRAM and FLASH memory of 149 KB on AudioMoth. The weights and bias consume 2.7 KB for four different bird species. With 256 KB of FLASH memory available on the board, we can approximately accommodate 35 more bird species. Also, with the increase of bird species, the number of training samples should proportionally increase to retain accuracy of the system. The overall current consumption of the AudioMoth is measured by sourcing 4.5 V, and it is reported to consume 9.1 mA of current. So, the daily battery capacity consumed by the AudioMoth is 218.4 mAh. With three AA batteries of 3500 mAh capacity, it can last for 48 days with the bird recordings on SD card and higher battery capacities like 6 V, 26000 mAh can be used for long-term deployments up to 6 months.

This system can generate valuable information regarding bird species. The following are the insights we derived:

- It can determine the number of occurrences of bird species, i.e., the density of birds in a particular area.
- For critically endangered species, it can be used to detect bird habitats by deploying this system across different places.
- Analysing the timestamps in the detection log can help us understand the seasonal migrations patterns to a location when deployed for long durations.

IV. Conclusion

In this paper, we presented a unique always-on tinyML acoustic classifier that can monitor and track ecological wildlife. To demonstrate the capabilities of this system, we used AudioMoth hardware to identify four different bird species with minimal resource footprint and can last up to 48 days on three AA batteries. The potential of identifying rare species or even the presence of previously assumed extinct species is possible as our system can train on limited or sparse data. Such hardware can be deployed in remote ecological locations, and the density data can be reviewed after a few days to provide necessary species protection. We also developed a software framework for AudioMoth hardware, resulting in the flexibility of training different species and enabling easy system deployment. We intend to extend this system to create a species map of entire forests using multiple deployments and wireless transmissions to control the systems. We can also improve and extend the battery life in such scenarios by leveraging energy harvesting and making a self-sustained system. This would result in easy tracking and monitoring of wildlife ecosystems with the least human intervention.

References

[1] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics*, vol. 4, no. 4, pp. 206–214, 2009.

[2] P. F. Donald and R. J. Fuller, "Ornithological atlas data: a review of uses and limitations," *Bird study*, vol. 45, no. 2, pp. 129–145, 1998.

[3] R. G. Pearson, J. C. Stanton, K. T. Shoemaker, M. E. Aiello-Lammens, P. J. Ersts, N. Horning, D. A. Fordham, C. J. Raxworthy, H. Y. Ryu, J. McNees *et al.*, "Life history and spatial traits predict extinction risk due to climate change," *Nature Climate Change*, vol. 4, no. 3, pp. 217–221, 2014.

[4] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. VerCauteren, N. P. Snow, J. M. Halseth, P. A. Di Salvo, J. S. Lewis, M. D. White *et al.*, "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 585–590, 2019.

[5] Google, 2021, https://https://www.wildlifeinsights.org/.

[6] N. Priyadarshani, S. Marsland, and I. Castro, "Automated birdsong recognition in complex acoustic environments: a review," *Journal of Avian Biology*, vol. 49, no. 5, pp. jav–01 447, 2018.

[7] C. Kwan, G. Mei, X. Zhao, Z. Ren, R. Xu, V. Stanford, C. Rochet, J. Aube, and K. Ho, "Bird classification algorithms: Theory and experimental results," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, 2004, pp. V–289.

[8] J. Xie, K. Hu, M. Zhu, J. Yu, and Q. Zhu, "Investigation of different cnn-based models for improved bird sound classification," *IEEE Access*, vol. 7, pp. 175 353–175 361, 2019.

[9] S. Disabato, G. Canonaco, P. G. Flikkema, M. Roveri, and C. Alippi, "Birdsong detection at the edge with deep learning," in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2021, pp. 9–16.

[10] Y. Shan, D. Paull, and R. McKay, "Machine learning of poorly predictable ecological data," *Ecological Modelling*, vol. 195, no. 1-2, pp. 129–138, 2006.

[11] X. Xiang, Y. Qian, and K. Yu, "Binary deep neural networks for speech recognition," *Proc. Interspeech 2017*, pp. 533–537, 2017.

[12] Y. T. Quek, W. L. Woo, and L. Thillainathan, "Iot load classification and anomaly warning in elv dc picogrids using hierarchical extended k-nearest neighbors," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 863–873, 2019.

[13] H. Yang, S. Liang, J. Ni, H. Li, and X. S. Shen, "Secure and efficient knn classification for industrial internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10 945–10 954, 2020.

[14] A. Palaniappan, R. Bhargavi, and V. Vaidehi, "Abnormal human activity recognition using svm based approach," in *2012 International Conference on Recent Trends in Information Technology*. IEEE, 2012, pp. 97–102.

[15] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "Svms modeling for highly imbalanced classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 281–288, 2008.

[16] S. S. Heydari and G. Mountrakis, "Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 landsat sites," *Remote Sensing of Environment*, vol. 204, pp. 648–658, 2018.

[17] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *European conference on machine learning*. Springer, 2004, pp. 39–50.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] A. R. Nair, S. Chakrabartty, and C. S. Thakur, "In-filter computing for designing ultra-light acoustic pattern recognizers," *IEEE Internet of Things Journal*, 2021.

[20] P. Kumar, A. R. Nair, O. Chatterjee, T. Paul, A. Ghosh, S. Chakrabartty, and C. S. Thakur, "Neuromorphic in-memory computing framework using memtransistor cross-bar based support vector machines," in *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2019, pp. 311–314.

[21] C. S. Thakur, T. J. Hamilton, J. Tapson, A. van Schaik, and R. F. Lyon, "Fpga implementation of the car model of the cochlea," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2014, pp. 1853–1856.

[22] A. P. Hill, P. Prince, E. Piña Covarrubias, C. P. Doncaster, J. L. Snaddon, and A. Rogers, "Audiomoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment," *Methods in Ecology and Evolution*, vol. 9, no. 5, pp. 1199–1211, 2018.

[23] A. Hill and A. Rogers, "Audiomoth hardware," 2017.

[24] H. R. Sabbella, A. R. Nair, and V. Gumme, 2021, https://labs.dese.iisc.ac.in/neuronics/demos-and-products-2/in-filter-computation-for-acoustic-classification/.

[25] X.-C. Team, 2021, https://https://www.xeno-canto.org/species/Calidris-pygmaea.

[26] ebird, 2021, https://ebird.org/species/spbsan1.

[27] X.-C. Team, 2021, https://www.xeno-canto.org/species/Athene-blewitti.

[28] ebird, 2021, https://ebird.org/species/forowl1.