

NRC Publications Archive Archives des publications du CNRC

Sustainable live sound monitoring and classification system enabled by a triboelectric nanogenerator and machine learning techniques

Haji Bagheri, Majid; Rajabi-Abhari, Araz; Gibbs, Owen; Xi, Pengcheng; Khan, Asif Abdullah; Huang, Fangzheng; Hassan, Md Soyaeb; Yan, Ning; Ban, Dayan

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1002/eem2.70044>

Energy & Environmental Materials, 2025-06-19

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=69eda814-9bab-4847-9ef9-dc7a73932698>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=69eda814-9bab-4847-9ef9-dc7a73932698>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>



LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Sustainable Live Sound Monitoring and Classification System Enabled by a Triboelectric Nanogenerator and Machine Learning Techniques

Majid Haji Bagheri , Araz Rajabi-Abhari, Owen Gibbs, Pengcheng Xi, Asif Abdullah Khan, Fangzheng Huang, Md Soyaeb Hassan, Ning Yan , and Dayan Ban*


The growing demand for sustainable, real-time audio processing drives innovations in sound classification and energy harvesting. Traditional sound monitoring systems often struggle with scalability, energy efficiency, and adaptability, particularly in remote or resource-limited environments. The expansion of IoT applications intensifies power demands in widely distributed wireless sensor networks, highlighting the need for sustainable solutions. Moreover, the volume of data generated by these sensors frequently exceeds the capacity for efficient human analysis, necessitating the integration of machine learning and deep learning techniques. These methods must be optimized for fine-tuning with minimal data from new sensors, enabling efficient and accurate sound classification without extensive retraining. This paper presents a Triboelectric Nanogenerator (TENG)-based microphone that addresses energy consumption and data processing challenges by integrating advanced materials with sound classification systems. The proposed device uses polyimine/graphite polypropylene (PI/GP) coated paper to capture sound and harvest energy from ambient noise. It delivers an output power of 25.67 μW at 94 dB, powering a wireless transmission circuit while achieving high acoustic sensitivity and a frequency response of up to 20 kHz. Performance evaluations show 92.7% classification accuracy in simulated live environments and a processing time of 0.342 s for 5-s audio clips using the MobileNet V1 model. Pre-trained models fine-tuned with minimal data from the TENG microphone enable efficient sound classification without extensive retraining. This innovation offers a sustainable alternative to conventional microphones, supporting self-powered, real-time monitoring systems with wireless data transmission and energy storage capabilities.

1. Introduction

The global pursuit of sustainable energy solutions has intensified in recent years, driven by escalating energy demands and the urgent need for environmentally friendly alternatives. In parallel, the rapid advancement of Artificial Intelligence (AI) and digital media has increased the demand for real-time audio processing systems, which play a vital role in various applications such as surveillance, environmental monitoring, human-computer interaction, and accessibility for the hearing-impaired.^[1–3] However, the growing reliance on AI and Machine Learning (ML) technologies contributes to a significant rise in energy consumption. This highlights the pressing need for more energy-efficient systems operating independently of non-renewable energy sources. The use of non-renewable energy and disposable batteries further exacerbates environmental degradation, raising sustainability concerns as the demand for audio processing systems continues to grow.^[4–6] Consequently, there is a critical need to design AI and ML-integrated systems that are not only capable of harvesting energy from natural sources but also able to function sustainably in diverse conditions. This approach would reduce reliance on fossil fuels and other environmentally harmful energy sources, paving the way for clean, renewable alternatives in developing next-generation audio processing technologies.^[5]

M. Haji Bagheri, O. Gibbs, Dr. A. A. Khan, F. Huang, M. S. Hassan, Prof. D. Ban
Department of Electrical and Computer Engineering, Waterloo Institute for Nanotechnology, University of Waterloo, 200 University Ave. West, Waterloo, Ontario N2L 3G1, Canada
E-mail: dban@uwaterloo.ca
Dr. A. Rajabi-Abhari, Prof. N. Yan
Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, Ontario M5S 3E5, Canada
Dr. P. Xi
Digital Technologies Research Centre, National Research Council of Canada, 1200 Montreal Rd., Ottawa, Ontario K1A 0R6, Canada

Prof. D. Ban
Waterloo Institute for Nanotechnology, University of Waterloo, Mike & Ophelia Lazaridis Quantum Nano Centre, 200 University Ave W, Waterloo N2L 0A4, Ontario, Canada

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/eem2.70044>.

DOI: 10.1002/eem2.70044

Traditional sound classification and audio processing systems typically depend on centralized frameworks and commercial microphones that require continuous power from external sources, restricting their deployment in remote or resource-constrained environments. Even though the power consumption of a single sensor is minimal, the combined energy usage of lots of sensing units can become incredibly high, significantly increasing the overall power demand of IoT networks. These systems face substantial scalability, energy efficiency, and adaptability challenges, especially in dynamic or resource-constrained environments.^[7,8] Moreover, as the scale of monitoring increases, traditional systems' energy consumption and data processing requirements escalate, posing challenges in managing large-scale deployments without substantial infrastructure support.

Hardware constraints also play a role in limiting the versatility of traditional systems.^[9] The bulky or fragile components often used in conventional microphones and audio sensors are not well-suited for integration into portable, wearable, or embedded devices, narrowing the range of potential applications.^[10] Additionally, these systems often struggle to adapt to unpredictable changes in ambient conditions, leading to reduced accuracy and performance in varying environments.^[8,11–13]

To overcome these challenges, researchers have explored innovative approaches to real-time audio processing, including incorporating energy harvesting technologies. Among these, Triboelectric Nanogenerators (TENGs) have garnered significant attention for their ability to convert mechanical energy from ambient sources into electrical energy.^[14–17] One such application is the development of TENG-based acoustic energy-harvesting devices and microphones.^[18–20] TENGs produce electricity through the repetitive physical interaction and separation of two materials with different tribo-polarities under external forces, such as contact separation, lateral sliding, single-electrode, or freestanding modes.^[21,22]

TENGs offer notable advantages, including high power density, efficient energy conversion, cost-effectiveness, and a compact, lightweight design that is easy to fabricate. They are also easily integrated into wearable devices, IoT platforms, and other embedded systems. TENGs excel at capturing low-frequency and irregular mechanical energy, effectively converting it into electricity.^[23–25] In contrast, conventional microphones, such as condenser, dynamic, and electret microphones, rely on electrical power for their transducer elements and signal amplification, which restricts their mobility and versatility, especially in outdoor or resource-constrained environments.^[26] They also limit the choice of vibrating film materials, and their complex fabrication process leads to higher production costs. By harnessing energy from ambient mechanical vibrations or movements, TENG-based systems can operate independently and function not only as a conventional microphone, capturing sound waves even on a broader frequency range, but also as a generator that harvests energy from ambient noise, eliminating the necessity for external power sources or frequent battery changes.^[20,27–29] This enhances the sustainability of audio processing applications and allows for deployment in remote or challenging environments with limited access to power infrastructure.^[30]

In the development of TENG-based microphones, attention must extend beyond device structure. Factors like optimizing new composite materials, applying chemical modifications, modulating surface roughness, innovative device designs, and system integration are crucial.^[31,32] The efficiency and effectiveness of TENGs are primarily determined by the materials used in their construction.^[33] These materials play a crucial role in defining the triboelectric properties, energy

conversion efficiency, and durability of the devices. Various materials, including polymers, metals, ceramics, and composites, have been explored to optimize TENG performance.^[34] By incorporating flexible electrode materials, a TENG becomes highly responsive to external disturbances, positioning it as a promising option for harvesting acoustic wave energy. Various attempts have been made to convert acoustic vibrations into electrical energy using TENGs. Yang et al. proposed an ultra-wide spectrum triboelectric acoustic sensor, demonstrating the ability to perform high-quality music recording.^[19] Similarly, Guo et al. developed a triboelectric auditory sensor using Au and Fluorinated ethylene propylene (FEP) materials, designed for constructing an electronic auditory system and an architecture for external hearing aids in intelligent robotic applications.^[20]

Additionally, lightweight materials are essential for TENG-based microphones to effectively capture sound waves, making them susceptible to mechanical damage.^[35] Common tribo-materials for TENG construction, including metals, ceramics, semiconductors, and biobased or synthetic polymers, are prone to damage, leading to material fractures and performance deterioration. Therefore, designing sustainable TENGs requires the development of repairable friction layers for long-term applications.^[36] In this context, a study group developed an acoustic nanocomposite fibrous membrane-based triboelectric nanogenerator using polyvinylidene fluoride (PVDF) and multi-walled carbon nanotubes (MWCNTs).^[37] Moreover, another study utilized electrospun nylon nanofibers to create a TENG-based sound sensor for voice recognition applications.^[38]

Furthermore, incorporating high-dielectric constant fillers or conductive fillers can further enhance the electrical output of TENGs by increasing relative permittivity and creating electron-trapping sites and charge transport pathways.^[39,40] By applying strategies such as filler incorporation and surface modifications, the performance of TENGs can be significantly improved, offering both enhanced mechanical strength and electrical output.^[36]

Despite the advancements in tribo-materials used in TENG-based microphones, many studies have not integrated their devices with sustainable systems.^[41] Some focused solely on integrating ML without utilizing the device for active sound harvesting and recording purposes for ML-based sound classification. The reported studies lacked dual functionality, such as real-time sound classification and energy harvesting. They also fell short in providing high accuracy and fast response times, such as the 0.34 s achieved in our system.

TENG sensors capture data in both the time and frequency domains, encompassing features with local and long-term dependencies. Employing a combination of algorithms to extract these features significantly improves prediction accuracy.^[42] We can achieve high accuracy and enhanced performance by incorporating favorable materials and integrating suitable ML models, ensuring the device operates sustainably and efficiently. A key advantage of integrating TENG technology into live-streaming pipelines for sound classification and live monitoring is eliminating the need for external power sources.^[35] The vast amounts of data generated by sound sensors have paved the way for leveraging advanced ML techniques, such as Convolutional Neural Networks (CNNs), for tasks like environmental sound monitoring and classification.^[11,43,44] A popular method for sound classification involves using image-based classification CNNs and spectrograms. Spectrograms are image representations of audio signals throughout time. Mel spectrograms use a mel frequency scale, which simulates how humans perceive pitch, allowing mel spectrograms to represent how humans perceive sound accurately.^[45] Some examples of

image-based classification models that can be repurposed for audio classification are MobileNet V1,^[46,47] Inception V3,^[48] and Xception.^[49] These models' unique architectures allow them to perform differently across different datasets and use cases. Xuefeng Liu et al. used both MobileNet V1 and Inception V3 for real-time marine animal classification, achieving accuracies of 91.1% and 89.9%, respectively.^[50] These models have also been used by Li Gang et al. for honeycomb lung recognition, where MobileNet V1's accuracy outperformed Inception V3's by 12%, showing that due to the different architectures of the models, the performances on different datasets will vastly vary between the models.^[51] MobileNet V1's architecture focuses on its building blocks being depthwise separable convolutions.^[46,47] Inception V3 is built on inception modules.^[46] Finally, Xception combines an extreme application of depthwise separable convolutions, inception modules, and residual connections.^[49] These models were pre-trained on image datasets such as ImageNet, meaning that in order to be repurposed for sound classification, they need to be fine-tuned on large audio datasets. Using datasets such as Environmental Sound Classification (ESC-50), which is a benchmark for audio classification and consists of recordings captured using conventional microphones, the models mentioned above can be re-used for sound classification by implementing simple transfer learning techniques.^[52] This work proposes MobileNet V1 for its use in a live pipeline for audio classification due to its lightweight architecture, high performance, and fast inference, which is ideal for real-time applications.^[53,54] However, integrating data from these sensors with existing ML models, which were trained on large datasets collected by conventional microphones, presents significant challenges. This integration has not yet been systematically tested or fully explored, highlighting an area of ongoing research and development. For TENG sensors, the challenge lies in developing a device that not only harvests sound energy efficiently but also integrates seamlessly with ML models capable of high accuracy using limited training data.

This paper explores a TENG-based sound sensor's design, fabrication, and performance utilizing polyimide/graphite polypropylene (PI/GP) material for live-streaming sound classification. The PI/GP material is highlighted for its high output, which is attributed to factors such as the presence of amine groups, surface roughness, and potentially its dielectric constant.^[36] The study aims to advance the development of energy-efficient audio processing systems by covering the entire workflow from data acquisition to final output, integrating signal processing, ML algorithms, and classification modules. A vital aspect of this approach is integrating TENG-based energy harvesting units, constructed from an innovative material with live-streaming audio classification and monitoring, providing the necessary power for sustained operation in real-world scenarios. This study also investigates the energy harvesting capabilities, assesses the quality of captured audio, and explores their integration with ML models for real-time audio classification. Furthermore, the performance of the complete pipeline is evaluated with emulated real-life scenarios. The device's versatility is demonstrated through comparisons with various existing models, underscoring its adaptability to different configurations. Furthermore, the device was employed not only for energy harvesting and charging capacitors but also to power wireless transmission circuits, providing proof of concept that data can be both recorded by the microphone and transmitted wirelessly in a sustainable manner and effectively showcasing its potential for sustainable recording and wireless data transmission. The PI/GP material's high charge density enhances energy harvesting efficiency and durability. This allows the device to operate sustainably in harsh environments, providing the necessary power for

continuous real-time operation without external sources. Integrating TENG technology into live-streaming audio monitoring and sound classification represents a significant advancement in real-time audio processing. We demonstrate that the device can operate in harvesting and recording modes, making it a versatile tool for various applications. This innovation opens up exciting possibilities for applications in environmental monitoring, healthcare, smart cities, Internet of Things (IoT), agricultural fields, remote device malfunction monitoring, Human-Robot Interaction, and assistive technologies. Ultimately, it paves the way for a more sustainable and interconnected future.^[20,36]

2. Results and Discussions

2.1. Structure and Operating Mechanism of the Device

The PI/GP-coated paper was employed as a tribo-positive material and coupled with a tribo-negative polytetrafluoroethylene (PTFE) film to create TENG devices. The PI, with its backbone rich in electron-donor amino and imine groups, enhanced the TENG's charge density (**Figure 1a**).^[55,56] Furthermore, PI significantly enhanced the tensile strength of paper, which improves the long-term durability of the coated paper and the final device (**Figure S1a**, Supporting Information). The conductive GP particles also facilitated charge transfer and reduced surface charge loss.^[57,58] The GP particles further improved the TENG's output by increasing the composite's dielectric constant. The distribution of GP particles on the coated paper also increased surface roughness, effectively enlarging the contact area between the PI/GP composite and the opposing material. Based on a previous study, the 30% optimized percentage of PI/GP was used to fabricate the TENG device.^[36] This increased contact area led to generating more triboelectric charges, thereby enhancing the TENG's output power.^[39,59,60] The morphology of the GP particles is shown in **Figure 1b**. The surface and cross-sectional images of the base paper and PI/GP-coated papers are illustrated in **Figure S1b,c**, Supporting Information and **Figure 1c,d**.

The operation of the acoustic microphone fundamentally relies on the principles of contact electrification and electrostatic induction, as illustrated in **Figure 1e**.^[37,61] In this setup, the PTFE membrane functions as the tribo-negative layer due to its higher electron affinity compared to the PI/GP layer, which acts as the tribo-positive layer.^[62] When the PI/GP-coated paper comes into contact with the Al-coated PTFE membrane in response to acoustic pressure, charge transfer occurs from the PI/GP surface to the PTFE surface (**Figure 1e-i**). This process leaves behind positive charges on the PI/GP surface and an equal amount of negative charges on the PTFE membrane (**Figure 1e-ii**).^[63] As the acoustic pressure causes the two layers to move apart, an electric potential difference is generated between the triboelectric materials. This potential difference drives a flow of electrons from the bottom electrode to the top layer through an external circuit (**Figure 1e-iii**). When the acoustic wave compresses the layers together again, this contact separation cycle reverses the electron flow. This repetitive motion creates an alternating electrical signal corresponding to the acoustic wave's variations, as depicted in **Figure 1e-v**. Additionally, the arrow indicator in **Figure 1e**, connected to the device, illustrates the direction of current flow and indirectly represents the electron transfer process during this cycle.

Figure 1f presents the design of the TENG device used as a microphone. The device consists of a 300 nm thick layer of aluminum (Al) deposited on a PTFE membrane and a paper coated with a PI/GP

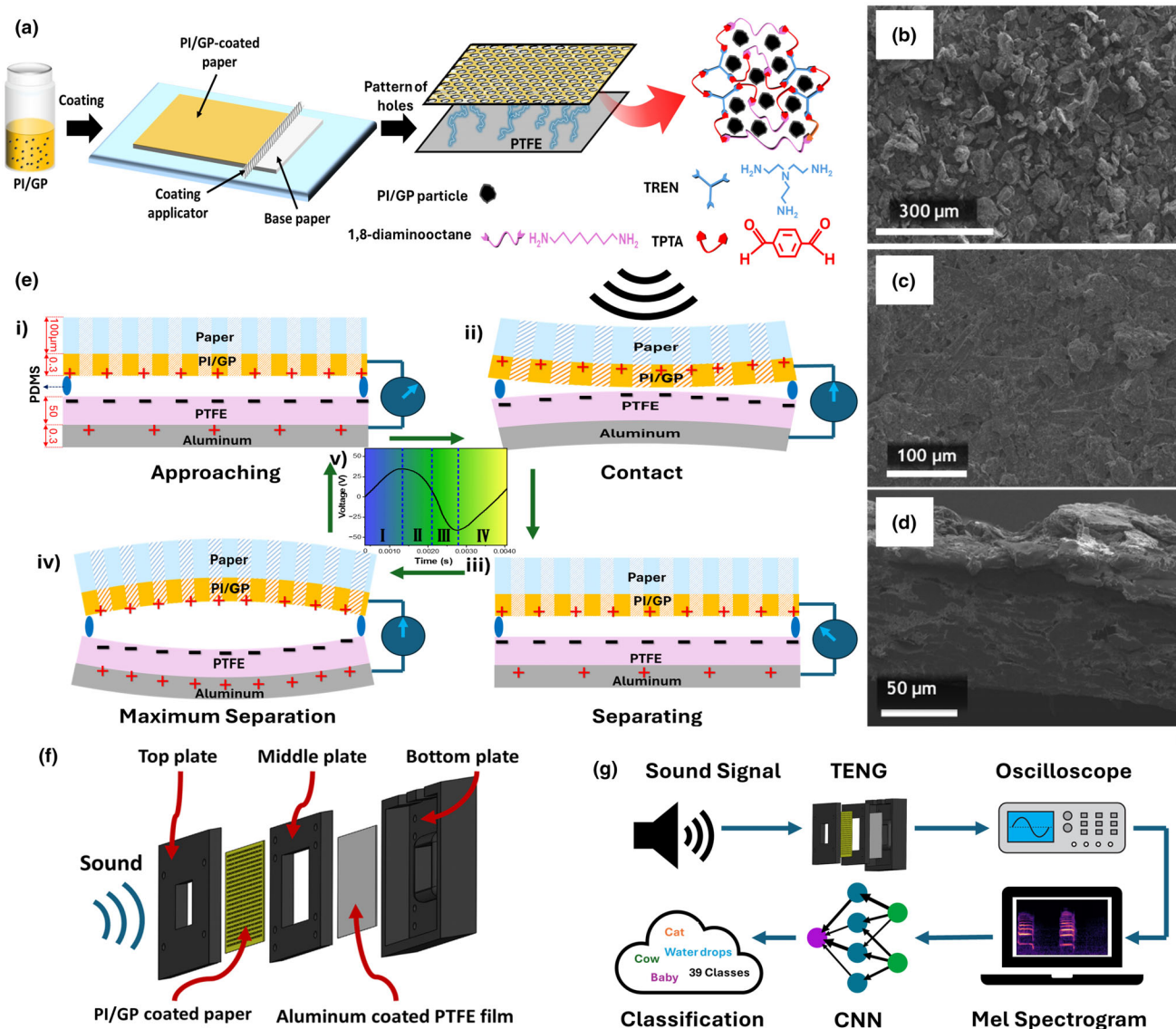


Figure 1. a) Diagram of the PI/GP film material preparation; SEM images showing. b) The morphology of the GP particles and c) the surface and d) the cross-sectional images of the PI/GP-coated paper. e) Schematic view of the working principle of the PI/GP-TENG device. f) Layer-by-layer diagram of the PI/GP-TENG device. g) Illustration of the developed live pipeline for audio classification.

composite featuring hole patterns (Figure 1e-i). The paper serves primarily as a support for the tribo-positive PI/GP layer. The aluminum coating on the tribo-negative PTFE enhances charge collection efficiency. This innovative design surpasses previous models in performance due to its advanced structural features. The compact, sandwich-like configuration of the device, produced using 3D printing technology, supports the creation of thin, susceptible triboelectric acoustic sensors. The device was designed with three primary goals. These goals were to simplify the fabrication process, tension the PTFE layer effectively to eliminate wrinkles, and allow for adjustable spacing.^[64] This design ensures the PTFE layer vibrates effectively when exposed to sound, enhancing its performance, reproducibility, and durability. Most existing devices have utilized tape to affix layers to the substrate, which can lead to issues with reproducibility and extended assembly time.^[30]

Our proposed sandwich-like structure addresses these problems, improving the ease of preparation and the device's overall efficiency. Maintaining the PTFE layer tight and flat achieves maximal contact with the PI/GP layer, significantly increasing the device's output and efficiency. The spacing between the two triboelectric layers critically influences the device's output (Figure 3a); thus, adjusting this spacing dynamically allows the device to perform optimally and can be tuned based on the desired frequency range and environment. The sensor developed in this study features a thickness of 15 mm and a diameter of 70 mm and includes a 20 mm × 20 mm paper PI/GP section, as illustrated in Figure 1f.

The front of the device features the PI/GP-TENG and incorporates patterned holes on the paper created with a UV laser cutter. These holes are designed to interact with acoustic waves, allowing air pressure to

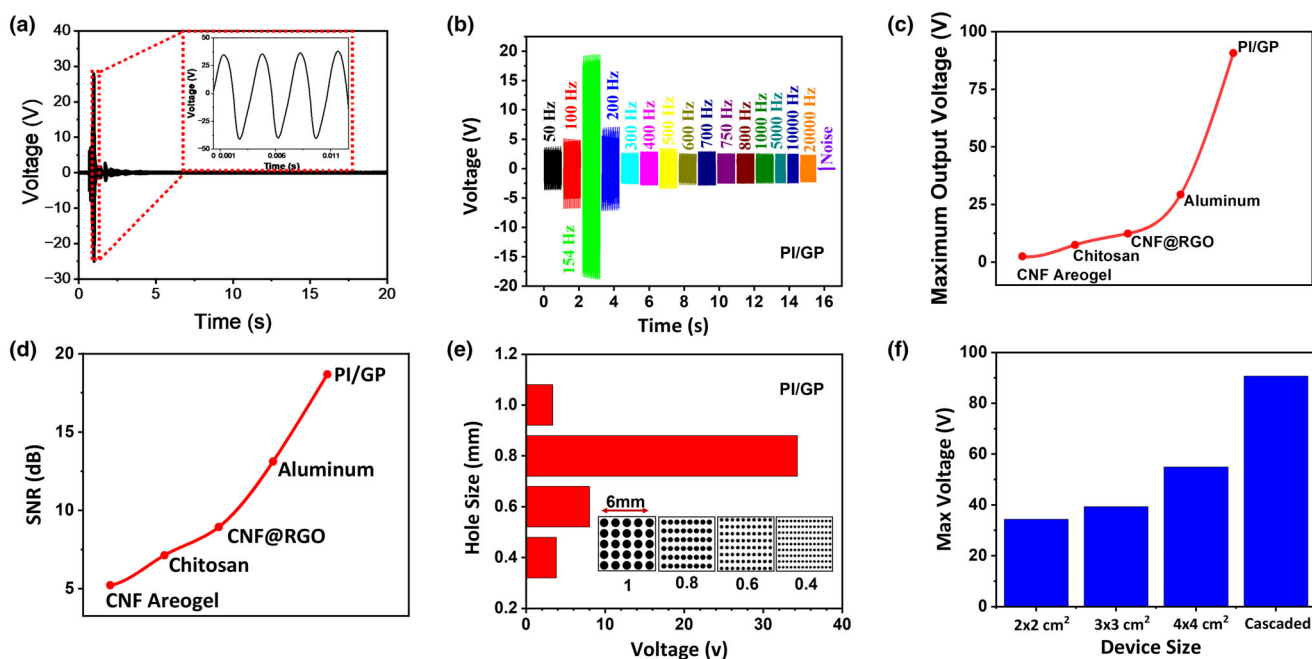


Figure 2. a) Spectrogram and voltage output from the $2 \times 2 \text{ cm}^2$ device when a linear frequency sweep is played. b) Voltage output at different frequencies shows a resonance frequency of 154 Hz and a high SNR. c) Maximum voltage outputs using different top film materials. d) Varying SNRs of different top film materials. e) Varying maximum voltage output from different PI/GP film hole sizes. f) Maximum voltage outputs from the different PI/GP-TENG device sizes.

pass through and thereby enhancing the vibration of the triboelectric membrane. By minimizing the effects of air resistance that can dampen vibrations, these holes create small resonant cavities that amplify the vibrations as sound waves travel through them.^[64,65] All components are assembled, as shown in Figure 1f, and the final device is secured with screws to adjust the thickness of the spacer.

The developed triboelectric acoustic sensor features a simple yet effective design to leverage the high efficiency of the PI/GP-TENG in converting acoustic signals to electrical ones. This sensor's adjustable spacer thickness allows for a highly sensitive, tunable, and self-powered acoustic detection system with potential applications in human-machine interfaces, the IoT, industrial manufacturing, transportation, and other advanced technology sectors, as depicted in Figure 1g.

2.2. Performance of Acoustic TENG

A linear frequency sweep test was performed to evaluate the performance of the PI/GP-TENG device. This test covered frequencies ranging from 20 Hz to 20 kHz over a 20-s duration. The results in Figure 2a illustrate the TENG's frequency response. Figure 2a inset demonstrates the device has a swift response time of 0.004 s, highlighting its rapid and efficient performance. Figure S2a, Supporting Information, compares the linear frequency sweep audio recorded by the TENG-based microphone with the original playback sound. Figure S2b, c, Supporting Information, present examples of comparing the TENG-recorded audio and the original ESC-50 dataset recordings, which were captured using condenser microphones.^[13] Together, these comparisons provide direct evidence of the TENG's accuracy in capturing a

wide frequency range and the complex harmonic structures of various environmental sounds.

Various frequencies were used to activate the sensor and assess the performance of the proposed acoustic sensor across different sound frequencies. As demonstrated in Figure 2b, the TENG acoustic sensor exhibited a robust output across a wide frequency range of 50–20 000 Hz, showcasing its ability to detect high-frequency sounds with minimal noise interference effectively. This capability was further substantiated in Figure S2d,e, Supporting Information, which present the power spectral density (PSD) analysis and the output voltage response of the TENG acoustic sensor as a function of frequency, respectively. The PSD graph illustrates power distribution across the frequency spectrum and confirms the device's ability to record up to 20 kHz sound signals accurately.^[66,67] This highlights the sensor's effectiveness in capturing high-frequency acoustic data while maintaining low noise levels. As shown in Figure S2e, Supporting Information inset, the output voltage gradually declines as frequency increases from 1 to 20 kHz. This behavior is attributed to the reduced displacement amplitude of the vibrating membrane at higher frequencies, which limits the effective contact area between the PTFE and PI/GP layers. Additionally, the device demonstrates sensitivity comparable to commercial electret microphones, which are expensive and consume more energy, offering a sensitivity of around -40 dB for frequencies under 5 kHz.^[68] The peak voltage was recorded at 154 Hz under a consistent sound pressure of 90 dB. This happens due to the natural limitations in the deformability of the vibrating membrane, causing the contact between the PI/GP layer and the PTFE layer to change with sound frequency under constant sound pressure.^[69,70] These findings demonstrate the TENG's potential for integration with sound monitoring

systems, as it can detect a wide range of frequencies while delivering high output performance.

To investigate and compare using different materials as tribo-positive layers output performance, different eco-friendly materials, including aluminum (Al), chitosan film, cellulose nanofibrils (CNF) aerogel, and an aerogel containing CNF and reduced graphene oxide (CNF@rGO) films were made with the same thickness and structure and tested with the same linear frequency sweep. The maximum output voltage results are shown in Figure 2c among different materials, and the PI/GP shows a maximum output of ~90 volts. In our comparative analysis of various materials used in the proposed device, the PI/GP composite demonstrated the highest signal-to-noise ratio (SNR) among all tested materials, as illustrated in Figure 2d. This superior SNR indicates that the PI/GP composite is highly effective in minimizing background noise while accurately capturing sound signals, making it an optimal choice for real-time sound classification applications. The enhanced performance of the PI/GP-based TENG microphone underscores the importance of material selection in achieving high-precision acoustic sensing.^[71]

Additionally, the impact of hole size was also investigated. The maximum voltage amplitude was attained with a hole diameter of 0.8 mm, as shown in Figure 2e. All samples maintained a consistent hole area percentage of 20%. This observation can be attributed to changes in hole pressure resulting from tiny holes, leading to fluctuations in output voltage.^[64]

The cascade device consists of three interconnected units with dimensions of 2 cm × 2 cm, 3 cm × 3 cm, and 4 cm × 4 cm, respectively, as illustrated in Figure S3, Supporting Information. Each unit is constructed using a 3D-printed design featuring different square holes. This arrangement ensures a thorough coverage of the sound frequency range from 20 Hz to 20 kHz and the maximum output efficiency (Figure 2f).^[72-74] Inspired by the function of the basilar membrane in acoustic sensors, this design enhances coverage across the frequency spectrum, thereby boosting the sensor's sensitivity and accuracy in detecting a wide range of sound signals. Figure 2f illustrates that the cascaded device can obtain maximum output. As shown in Figure S4a-d, Supporting Information, among the waveforms of 2 × 2 cm², 3 × 3 cm², 4 × 4 cm², and the cascaded devices, cascading the devices significantly boosts both voltage and current outputs, thereby increasing sensitivity and power, making the device suitable for energy harvesting and driving low-energy consumption circuits. The study further compares series and parallel connections of the generator units, revealing that while series connections yield higher voltage output, parallel connections provide more significant current, aligning with existing literature and highlighting the importance of connection configuration in maximizing desired output (Figure S4e,f, Supporting Information).^[75,76]

Simulations were performed using COMSOL Multiphysics to theoretically verify the device's behavior with varying film spacings. The electrostatics interface in COMSOL was utilized to model the electric potential between the two layers. In the simulation, two 2D layers were represented, separated by a gap, with the top PI/GP layer oscillating up and down to contact and then separate from the bottom layer. The films were modeled with 2 cm × 2 cm dimensions to simulate the 2 × 2 device. Based on the datasheet information, the relative permittivity values were set to 2.1 for the PTFE film and 3.4 for the PI/GP film.^[76,77]

The simulation results indicate that increased spacing between the two film layers results in the device's higher maximum voltage output.

According to Gauss's laws and the relationship $V = E \cdot d$, the potential difference between the electrodes can be expressed as^[78-80]:

$$V = \frac{-Q}{S\epsilon_0}(d + x(t)) + \frac{\sigma x(t)}{\epsilon_0} \quad (1)$$

Here, V is the output voltage, S is the contact surface area, ϵ_0 is the vacuum permittivity, Q is the charge transferred between the electrodes, d is the effective thickness constant, σ is the surface charge density, and $x(t)$ is the distance between two contact surfaces.

In open-circuit conditions, where no charge transfers ($Q = 0$), the open-circuit voltage (V_{OC}) is:

$$V_{OC} = \frac{\sigma x(t)}{\epsilon_0} \quad (2)$$

In the simulation, the electric field (E) is constant due to fixed surface charge densities, implying that an increase in separation distance ($x(t)$) results in a corresponding increase in voltage (V), as illustrated in Figure 3a. This trend is consistent with the experimental data shown in Figure 3b, which illustrates that as the spacing increases from 0 to 1 mm, the device's output voltage and current also increase. This observation supports the notion that current and voltage are directly proportional to the tribo-charge surface density (σ) and the distance between the tribo-layers.

However, the data indicates that the device's output begins to decline beyond a separation distance of 1 mm. This decline is attributed to a reduction in contact between the two films and constraints on the motion of the top film.^[81] When the separation becomes excessively large, the diminished contact leads to less charge transfer, reducing surface charge density and overall output. Consequently, the optimal spacing for the device is achieved by balancing separation distance and surface contact, which, as illustrated in Figure 3b, is optimally maintained at 1 mm.

Figure 3c illustrates the maximum output power (dB) versus spacer distance (mm) across the first to fourth resonant frequencies. The data show that at a spacer distance of 1 mm, the maximum output power is achieved for the first resonant frequency. The highest output power occurs at a 0.5 mm spacer distance for the second and third resonant frequencies. Additionally, for the fourth resonant frequency, the output power remains relatively constant across all spacer distances. All that data indicate that the device is tunable. By adjusting the spacer distance, it is possible to optimize the output at different frequencies. This tunability is further demonstrated in Figure S5, Supporting Information, which presents frequency versus power and spacer distance.

The acoustic sensitivity of a microphone is a crucial performance metric.^[82] As the sound pressure level (SPL) increases from 60 to 100 dB, the maximum output voltage achieved at the resonant frequency of 154 Hz rises from ~1 to ~35 V (see Figure 3d and Figure S6a, Supporting Information). The formula for calculating acoustic sensitivity is:

$$SPL = 20 \log \left(\frac{V}{V_{ref}} \right) \quad (3)$$

SPL represents the sound pressure level, V is the output voltage, and V_{ref} is the reference voltage. The exponential fitting results exhibit excellent linearity ($R^2 = 99.85\%$), as shown in Figure 3d. To test our device further, we measured the output voltage and current at varying

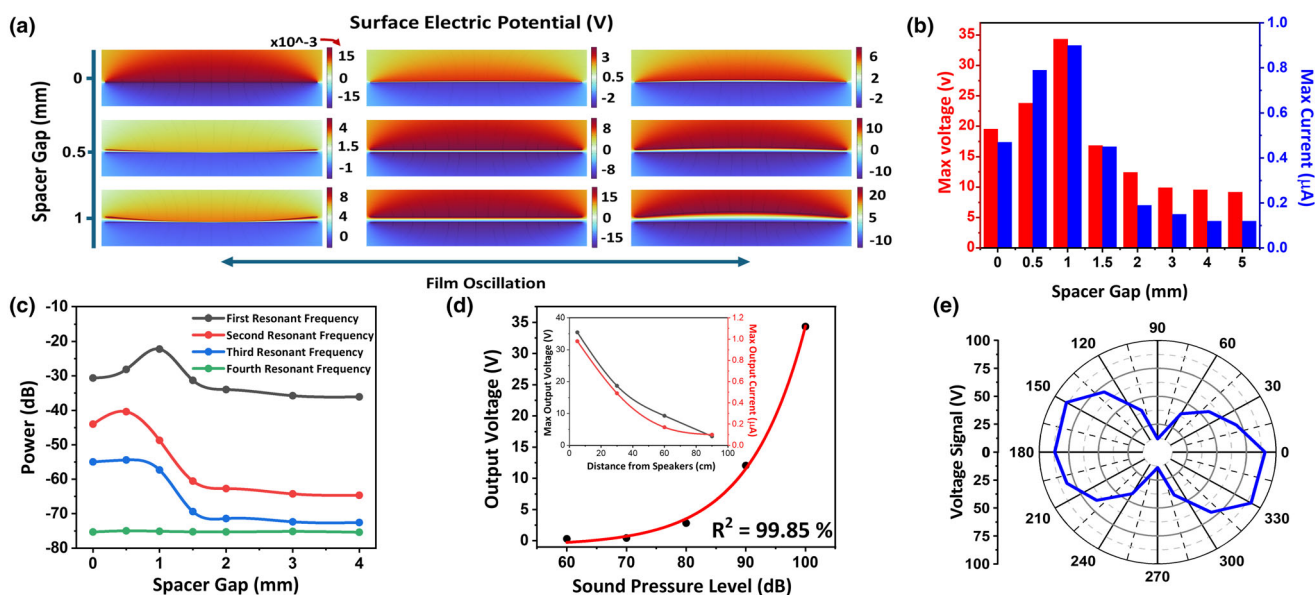


Figure 3. a) COMSOL simulations of the surface electric potential with varying spacer gaps and oscillation of the top film due to sound pressure. b) Voltage and current outputs of the PI/GP-TENG device with different spacer gaps. c) Varying power of different resonance frequencies at different spacer gaps. d) Voltage output from varying sound pressure levels and voltage and current output from different distances from the speakers. e) Directionality of the PI/GP-TENG device showing a bi-directional pattern.

distances of the speakers from the device (Figure S6b, Supporting Information). The results show that the microphone produces an acceptable output of 3 V at a distance of ~ 90 cm, indicating the device's capability to detect small signals (Figure 3d).^[83]

Figure 3e demonstrates the directional response of the PI/GP-TENG device. Directionality tests, commonly employed for conventional microphones, classify microphones into types such as cardioid, bi-directional, and omnidirectional.^[84,85] The symmetric pattern observed in our device identifies it as a bi-directional microphone, capable of capturing sound from both the front and back. This bi-directional pickup pattern significantly enhances its utility in sensing applications by effectively capturing sounds from multiple directions, which is particularly advantageous given the typically omnipresent nature of sound sources.

All of the outlined results demonstrate that the TENG-based microphone output remained consistent and accurate, even for high-amplitude sound sources, with negligible deviation that would impact the recorded audio. This ensures that the device is suitable for use in machine learning applications.

2.3. ML and Data Processing Results

After confirming the device's outstanding output and optimizing its parameters to achieve consistent performance, we conducted sound classification tests to evaluate its audio quality and compatibility with advanced ML models. A TENG-recorded ESC-50 dataset was created by playing the ESC-50 sounds through a speaker and capturing them with the TENG microphone (Figure S7a and Video S1, Supporting Information). Video S1, Supporting Information, demonstrates the device's high-quality sound recording capabilities. This recorded data was then used to train and test 4-ML models: MobileNet V1, Inception V3, ResNet 50, and Xception.

As previously mentioned, these models were chosen for their varying architectures and proven effectiveness in audio and image classification tasks. MobileNet V1 is known for its efficiency and suitability for real-time applications; Inception V3 offers a balanced trade-off between accuracy and computational cost; ResNet 50 is renowned for its deep architecture and ability to handle complex patterns, and Xception leverages depthwise separable convolutions for improved performance.^[46–49]

To assess audio quality, we performed both qualitative and quantitative analyses through human listening analyses by tuning the mentioned models on the ESC-50 and TENG ESC-50 training datasets. Random samples were taken, and their mel spectrograms were compared with the original recordings to ensure high-quality sound capture, as previously demonstrated (Figure S2, Supporting Information). The models were then tested on their respective test datasets. If the models performed comparably on the TENG-recorded ESC-50 dataset and the original ESC-50 dataset, it would indicate that the TENG device produces high-quality recordings despite its distinct audio characteristics.

The results of the classification tests are summarized in Table 1. The table demonstrates the models' performances under four scenarios: training and testing on the ESC-50 dataset, training on the ESC-50 dataset and testing on the TENG dataset, training and testing on the TENG dataset, and training on the TENG dataset and testing on the ESC-50 dataset. The difference in audio characteristics between conventional microphones and the TENG device was evident in the models' performances. Models tuned on the ESC-50 dataset performed poorly when tested on the TENG dataset, whereas models tuned on the TENG dataset performed very well. This discrepancy highlights the inherent differences in sensing characteristics between the devices and underscores the necessity of a TENG-recorded dataset for accurate classification.

When models were tested on datasets matching their tuning datasets, their performances were comparable for ESC-50 and TENG ESC-50

Table 1. Data was collected from MobileNet V1, Inception V3, ResNet 50, and Xception models, including accuracies from different tuning and testing datasets, model inference times, and model parameters.

Model	Training: ESC50 Test: ESC50	Training: ESC50 Test: TENG	Training: TENG Test: TENG	Training: TENG Test: ESC50	Inference time (ms)	Parameters (Millions)
MobileNet V1	79.80%	38.10%	81.70%	42.00%	69.2	3
Inception V3	73.40%	34.90%	70.50%	32.70%	91.8	22
ResNet 50	56.10%	22.80%	60.60%	21.50%	117.2	24
Xception	76.90%	30.80%	76.90%	38.80%	112.3	21

datasets. Notably, Xception achieved the same accuracy with both datasets, demonstrating the high quality of the audio produced by the TENG device. This compatibility with state-of-the-art ML models indicates that the TENG device's high-quality audio output can effectively be used for sound classification tasks.

For real-time applications, MobileNet V1 emerged as the best-performing model, achieving an accuracy of 81.7%, the highest among the tested models. Besides, it had the lowest inference time of only 69.2 ms. Leveraging MobileNet V1's low parameter count, it can be a highly efficient and lightweight model for live applications using the TENG device.

The TENG device shows great potential for high-quality audio recording and compatibility with state-of-the-art ML models. The comparable performances of the models on the TENG-recorded dataset and the benchmark ESC-50 dataset demonstrate its ability to produce high-quality audio output. MobileNet V1 is particularly well-suited for real-time applications due to its high accuracy and low inference time.

Figure 4 presents the confusion matrix for the individual performance of each class using MobileNet V1, offering a comprehensive view of the classification performance across various sound classes. The matrix reveals a strong diagonal dominance, signifying that most sounds were accurately classified, as indicated by the high number of true positives along the diagonal.

The analysis reveals that the classifier performs exceptionally well with familiar sounds such as “dog,” “cat,” and “vacuum cleaner,” underscoring its reliability in detecting these frequently encountered noises. Most classes achieve correct classification rates of at least 5 out of 8 test audio clips; however, some misclassifications occur, particularly among sounds with similar acoustic profiles, such as “crying baby” versus “drinking sipping” or “wind” versus “waterdrops.” These misclassifications can be attributed to the distinctive acoustic features of these sounds, as illustrated by the spectrogram analysis. This suggests that refining feature extraction techniques or enhancing model training could improve differentiation between these similar sound types.^[86]

Furthermore, the limited size of the ESC-50 dataset, with only 8 test audio clips per class and a relatively small number of training samples, constrains the model's ability to generalize effectively. This amplifies the impact of each misclassification during testing and highlights the challenge of achieving high accuracy. These results provide a benchmark for evaluating classifier performance while emphasizing the dataset's limitations, particularly regarding the training set size. However, the high classification accuracy achieved despite the limited training on the ESC-50 dataset reflects the exceptional audio quality of the TENG device. The device's capture of detailed and high-fidelity sound features significantly enhances model performance, enabling accurate classification even with minimal training data. This highlights the device's

potential for robust performance in real-world live sound detection tasks. Furthermore, considering that the ESC-50 dataset includes environmental sounds recorded with conventional microphones—some of which contain natural background noise or are of lower sound quality—the model's success with this dataset showcases both the robustness of the TENG device and the model's ability to handle varying sound qualities in real-time sound monitoring tasks.

2.4. Multifunctional Applications as a Sustainable Sensor

Due to TENG devices' inherent energy harvesting capabilities, they can be repurposed as energy harvesters within a system. By capturing background noise energy, this energy can be stored in a power unit, such as a rechargeable battery or capacitor, potentially transforming the live sound monitoring system into a sustainable, self-powered setup.^[23,24]

Leveraging the high output current and voltage of the PI/GP-TENG and its advantage in converting acoustic signals to electrical ones, the developed triboelectric acoustic sensor is used for self-power and the wireless communication circuit. To illustrate this potential and the practical electrical applications of PI/GP-TENG, the device was employed to charge different capacitors. The energy generated by the composite PI/GP-TENG device at its resonant frequency (154 Hz) with an SPL of 94 dB was stored in various commercial capacitors by connecting it to a full-wave bridge rectifier (**Figure 5a**). The rectifier converts the alternating current (AC) produced by PI/GP-TENG into direct current (DC). **Figure 5a** presents the results obtained using capacitors with varying capacitances (1, 2.2, 4.7, 10, and 47 μF) in the rectifier circuit to evaluate the charge storage capacity of the PI/GP-TENG. Along with that, the PI/GP-TENG successfully charged the commercial capacitors within 40 s, demonstrating specific charge rates. Notably, 3 V of 1 μF capacitor charge was achieved after 5 s.

Voltage and current performances were evaluated with various external resistances (ranging from 0 to 100 $\text{M}\Omega$) connected to the circuit to explore the electrical characteristics further. According to the maximum power transfer theorem,^[87] the peak output power was achieved when the external resistance exactly matched the internal resistance of the fabricated PI/GP-TENG. As depicted in **Figure 5b**, the current decreased with increasing load resistance, while the voltage exhibited an opposing trend. The peak output power reached 25.67 μW , corresponding to a power density of 1.6 $\mu\text{W}/\text{cm}^2$, as illustrated in **Figure S7b**, Supporting Information, at a resistance of 25.5 $\text{M}\Omega$ with an effective contact area of $4 \times 4 \text{ cm}^2$. This level of triboelectric power generation can support small electronic devices, such as recording or transmission circuits.^[22,88,89] To evaluate the device's power output at various sound pressure levels (SPLs), measurements were conducted across a range of

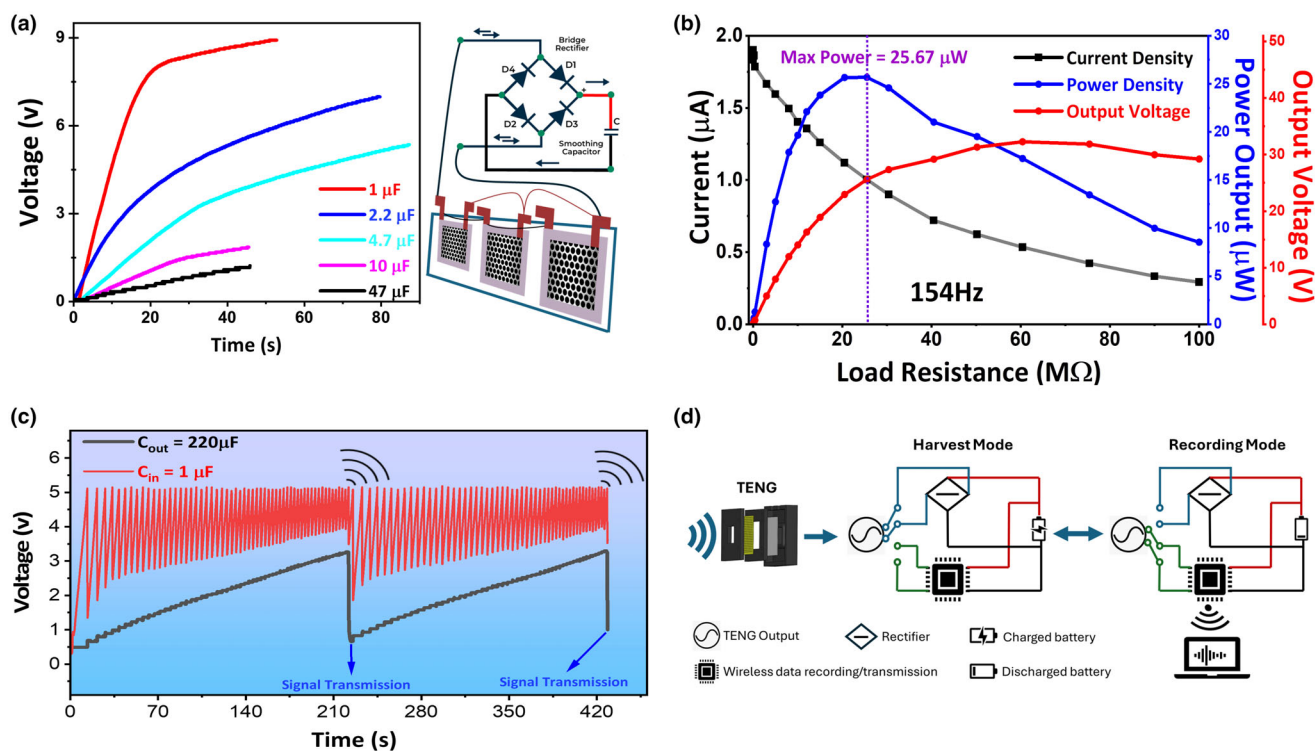


Figure 5. a) Charging curves of various capacitors, charged by the PI/GP-TENG device output, and an illustration of the charging circuit. b) Output current, voltage, and power with varying load resistances from 0 to 100 MΩ. c) Capacitor charging curves from the wireless communication device showing the signal output. d) Concept illustration of the PI/GP-TENG device used in a self-powered sensor circuit.

The device facilitated wireless communication after approximately 3.5 min in the initial charging cycle, while subsequent transmissions required only about 3 min. The longer initial charging time is attributed to the capacitor starting from an empty state.^[22] After the initial cycle, the capacitor did not fully discharge, maintaining a voltage of around 0.9 V after powering the RF module. Consequently, the time needed for subsequent wireless data transmissions was significantly reduced. This demonstration of integrating the PI/GP-TENG device with the self-powered communication system highlights the remarkable potential of the PI/GP material.

Known for its high output performance, PI/GP benefits from the presence of amine groups, increased surface roughness, and a favorable dielectric constant.^[36] The conductive GP particles enhance charge transfer, reduce surface charge loss, and increase the dielectric constant, resulting in greater output power. The composite's morphology also boosts surface roughness, improving the contact area and overall energy output. As a high-performance tribo-negative material, PI/GP enables the development of self-powered wireless electronics, making it a viable solution for sustainable and energy-efficient applications.

2.5. Live Audio Monitoring Evaluation

The capabilities of the PI/GP-TENG device for live sound monitoring and classification were evaluated by constructing a live pipeline, which was then tested in a simulated environment to assess its accuracy (Video S3, Supporting Information). This pipeline was centered around the fine-tuned MobileNet V1 model, chosen for its high performance

on the TENG ESC-50 dataset and rapid inference time, making it ideal for real-time classification.

The first step in the pipeline involves capturing the device's output using an oscilloscope to record the voltage output (Figure 1g and Figure S7a, Supporting Information). This output is recorded in 5-s windows and then sent to the processing device. The pipeline also includes sound event detection (SED), which monitors sounds using Root Mean Square (RMS) thresholding. In this method, the audio signal's energy is compared to a threshold determined by a noise profile, with the threshold set at 160% of the noise profile's RMS energy. The corresponding window is discarded and not classified if no 0.1-s input audio frame exceeds this threshold. The 5-s windows were chosen to match the length of the audio clips used to tune the model.

Once captured, the windows are processed into mel spectrograms and fed to the ML model, which outputs the class with the highest score. The system can be adapted to output the top-ranked classes; however, only the highest-scoring class was considered for quantifiable results, corresponding to Top-1 accuracy. This metric reflects the model's ability to correctly classify the most likely class from the predictions.

Some videos featuring continuous sounds from different classes, including background noises, were selected to simulate a live environment. A total of 500 s of audio from these videos was used, resulting in 100 classifications per class. The correct inferences were tallied to calculate accuracy out of 100. The system's processing time was also averaged across 100 classifications, including the audio-to-mel-spectrogram processing and the model's inference time.

Table 2. Accuracies and processing times from the live pipeline across various sound groups and classes.

Group	Class	Accuracy (%)	Processing time (s)
Animals	Cow	87	0.35
	Cat	96	0.33
Human, non-speech sounds	Crying baby	88	0.34
	Snoring	95	0.34
Interior/domestic sounds	Mouse click	96	0.33
	Clock tick	98	0.33
Natural soundscapes and water sounds	Water drops	100	0.37
	Chirping birds	81	0.35
Exterior/urban noises	Church bells	86	0.35
	Hand saw	100	0.33
Average across classes		92.7	0.342

Ten classes, representing two classes from five different sound groupings, were chosen for testing. **Table 2** presents the accuracy results for each class under each group and the processing time for each class. This table provides a detailed breakdown, showing how the system performs across different sound categories and highlighting the efficiency of the processing times.

The live system with the PI/GP-TENG device achieved an average accuracy of 92.7% across the 10 classes and an average processing time of 0.342 s. This processing latency aligns with typical embedded audio classification pipelines, as reported in recent works. Recent studies have reported that end-to-end latency in smart microphone systems can vary depending on buffering and processing architecture.^[92,93] The high accuracy and low processing time demonstrate the TENG device's capability for live applications. In latency-sensitive applications such as urban monitoring or real-time voice interfaces, even a sub-second delay can impact system responsiveness and user experience.^[94] The device's self-powered, compact, and low-cost nature, combined with lightweight machine learning models, further enables seamless integration into diverse environments, making it an ideal solution for scalable, on-site deployment.

By implementing these approaches, this work demonstrates the proof of concept for a self-powered or sustainable recording, wireless transmission, and live monitoring system using a PI/GP-TENG. Leveraging its high output power and current, this system paves the way for more sustainable environmental and remote sound monitoring applications where energy is crucial and self-powered sensors are needed (Figure 5d). This device can operate in two modes: an energy harvester and a recording circuit that transmits data wirelessly. The device can harvest background noise without specific sound events and store it, as shown in Figure 5d. The device can switch to recording and sending signals when a sound event occurs, ensuring efficient energy usage and reliable monitoring. A summary of relevant works is presented in Table S1, Supporting Information, to compare the device's performance with that of recent triboelectric acoustic energy harvesters. Although the obtained power density is moderate, it was measured at a realistic SPL of 90 dB, reflecting typical ambient conditions. The device also offers broad frequency response, compact design, and real-time ML integration, highlighting its potential for self-powered, real-time sound classification.

3. Conclusion and Future Perspective

This study presents a significant advancement in real-time audio processing through the development of a polyimide/graphite polypropylene (PI/GP) based triboelectric nanogenerator (TENG) sound sensor. The PI/GP material was highly effective in converting acoustic signals into electrical energy, thanks to high output power and enhanced charge density. Additionally, the conductive GP particles increase charge transfer, reduce surface charge loss, and boost the dielectric constant, resulting in higher energy output. With its increased surface roughness, the composite's morphology enhances the contact area, optimizing energy generation across a range of sound frequencies up to 20 KHz. The fabricated PI/GP-TENG device excels in energy conversion and provides a highly tunable structure, enabling output optimization across various sound frequencies.

The ability of the PI/GP-TENG device to operate both as an energy harvester and as a sound recording tool marks a substantial advancement in creating sustainable, self-powered audio monitoring systems. The high output power allows it to charge storage units and power wireless communication devices, showcasing its potential in self-sufficient sound monitoring and data transmission applications. By leveraging its tunable structure, the device can dynamically adjust its resonance frequency to match ambient noise frequencies, thus optimizing energy harvesting from background noise and ensuring efficient operation in real-world environments.

In practical applications, the PI/GP-TENG device was integrated into a live-streaming audio classification pipeline, achieving an impressive accuracy of 92.7% and a processing time of 0.342 s, optimized through fine-tuning on a dataset of 2000 TENG-recorded samples. This performance underscores the device's capability for real-time sound monitoring and classification, along with its integrability and adaptability to replace conventional microphones, making it a sustainable and versatile tool for various fields such as environmental monitoring, smart cities, IoT, and assistive technologies.

Future work will ensure the model's performance under less-than-ideal or extreme conditions, such as high noise levels and background interference. Although the current tests have demonstrated the device's resilience to such challenges, strategies such as adversarial training, data augmentation, and regularization could be employed in future updates to the model, further enhancing its performance. Latency reduction may be achieved through approaches such as model compression and hardware-efficient architectures to support faster edge-device deployment with minimal impact on accuracy.

Overall, this work demonstrates a proof of concept for integrating low-cost TENG technology with Machine Learning-based sound classification, enabling real-time, sustainable sound monitoring, recording, and wireless transmission. The findings from this study lay a strong foundation for future developments in self-powered audio monitoring systems, offering promising solutions for efficient, environmentally friendly real-world sound processing technology applications. This work also paves the way for sustainable, interconnected technologies in IoT, environmental monitoring, and assistive systems.

4. Experimental Section

TENG materials and preparation of the films: To fabricate the basic structure of TENG, it is essential to use at least two materials with different triboelectric properties. The effectiveness of a TENG depends mainly on the difference in

triboelectric properties between the selected materials. However, selecting appropriate materials from the triboelectric series can be challenging.^[95–98]

The two films used in the PI/GP-TENG device were aluminum-coated PTFE film as a tribo-negative material, while a PI/GP-coated paper was used as a tribo-positive material. The PTFE film is prepared using AJA sputtering, where 300 nm of aluminum is coated onto the back side of the film. Commercial office paper was first selected as a base material to prepare the tribo-positive layer. The coating dispersion was prepared using the method provided in our previous study.^[36] Briefly, 1.00 g of terephthaldehyde was first dissolved thoroughly in 10.00 g of dichloromethane (DCM). Then, 5.00 g of ethanol was added to the mixture and stirred well for 1 min. Next, commercially available GP (graphite-to-polypropylene ratio of 75:25, Hydrogenic Corporation, Canada) was added to the solution and stirred for 1 min. Finally, 0.485 g of 1,8-diaminooctane and 0.398 g of tris(2-aminoethyl)amine were added to the mixture and stirred for 5 min. 9 g m⁻² of the resulting mixture was then coated onto the base paper using a coating applicator and dried in the oven at 80 °C overnight. After preparing PI/GP-coated paper, a CO₂ laser was used to make uniform circular patterns on the paper.

TENG design and device fabrication—preparation: The device structures were printed using Prusa i3 MK3S+ 3D printers. PLA was used as the filament for printing the devices. The print settings were 30% infill, supports everywhere, and no brims.

The single and cascaded devices are prepared similarly due to their matching structures with only slight differences. The single structure assembly begins with attaching the copper electrodes. For the PTFE film, a rectangular piece of copper tape is attached to the aluminum side. For the PI/GP film, a square larger than the device's opening is made with copper tape, and finally, a rectangular piece of copper tape is attached to that square. The next step is aligning the square PTFE film with the aluminum deposited side, touching the tensioning mechanism and the copper electrode extending out of the front notch. The middle plate is pressed firmly onto the PTFE film to hold it in place. Using four 10 mm M3 screws, the middle plate is secured in place. The PI/GP is then placed with the square of copper tape touching the middle plate and surrounding the device's opening. The copper electrode is aligned with the notch not used for the PTFE film. Once the PI/GP layer is aligned, the top layer is firmly pressed against the film. Four 20 mm M3 screws are then placed and tightened using nuts. For the cascaded device, the same steps are followed but done 3 times, once for each of the 3 sizes of devices. The difference in the preparations of the single and cascaded device is that once the cascaded device is built, the electrodes need to be attached in either a series or parallel configuration using copper tape. In the series configuration, the positive electrode of each device is connected to the negative electrode of the following device. In the parallel configuration, the positive electrodes are connected, and the negative electrodes are similarly connected.

Characterization techniques and measurements: The GP powder and the surface and cross-section area of the paper and PI/GP-coated papers were observed using a Hitachi S5200 (Japan). Tensile tests of base paper and PI/GP-coated papers were performed using DMA Q800 (TA Instruments, USA) at room temperature.

Machine learning and data processing: The ESC-50 dataset includes 2000 environmental sound recordings, categorized into five main groups with 10 distinct classes, each containing 40 audio clips. The TENG recorded ESC-50 dataset was recorded by concatenating all audio clips of two classes into 6-min and 20-s audio clips, which were then played by speakers and recorded on the TENG device. The audio clips were then split into the original 5-s clips and named according to the original ESC-50 dataset. The audio was converted to mel spectrograms using Librosa, then normalized, converted to RGB, and resized to the model's input of [224, 224] pixels. The data pre-processing was done on the fly. TensorFlow was used to initialize the dataset and to load the ML models. All models were tuned and tested in an identical format.

Each model was pre-trained on ImageNet and loaded without the top layers as a custom classification block was built for our dataset. The classification block included a global average pooling layer to reduce dimensionality, a dropout layer with a 60% dropout rate to lessen overfitting, and a 39-neuron fully connected layer for classification. The fully connected layer also included L2 regularization to help with overfitting. The models were compiled with an Adam optimizer, categorical cross-entropy, and a learning rate of 0.0001. Early stopping and reduced learning rate on the plateau were also implemented. The models were fine-tuned by unfreezing layer by layer, starting from the top until the model reached its

peak accuracy on the test dataset. Once the models reached their peak accuracy, they were tested on the test dataset to produce confusion matrices and a value for their average inference time. Average inference time was calculated by tracking the time it took for the model to output a classification for each audio clip and then averaging those times from all the audio clips in the test dataset. Pre-processing was not included in the inference time calculation as the comparison was between the speed and efficiency of solely the models, not the data processing. The models were tuned and run on an Asus laptop with an AMD Ryzen 9590HS CPU, a Nvidia 3050ti GPU, and 16GB of RAM.¹

Acknowledgements

M.H.B. and A.R.-A. contributed equally to this work. This research is financially supported by The Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant No: CRDPJ 514858-17), Ontario Centers of Excellence (OCE), Canada (Grant No: VIP II – 28314), University of Waterloo, Canada (Grant No: 10001-10643). SX and CW acknowledge the National Science Fund for Excellent Young Scholars (Grant No: 61822503), the Natural Science Foundation of China (Grant Nos: 22075043, 21875034, 61704093), and the Foundation of Jiangsu Province for Outstanding Young Teachers in University (Grant No: BK20180064). The authors thank the Giga-to-Nano Electronics (G2N) lab for the experiment facilities.

Conflict of Interest

The authors declare no conflict of interest.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Keywords

deep learning, energy-harvesting, live sound monitoring, self-powered microphone, triboelectric nanogenerator

Received: February 26, 2025

Revised: April 29, 2025

Published online: May 2, 2025

- [1] Y. Yang, Y. Kartyrnik, Y. Li, J. Tang, X. Li, G. Sung, M. Grundmann, *STREAMVC: Real-Time Low-Latency Voice Conversion*, 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, **2024**.
- [2] F. Ahsan, N. H. Dana, S. K. Sarker, L. Li, S. M. Muyeen, M. F. Ali, Z. Tasneem, M. M. Hasan, S. H. Abhi, M. R. Islam, M. H. Ahamed, M. M. Islam, S. K. Das, M. F. R. Badal, P. Das, *Prot. Control Mod. Power Syst.* **2023**, *8*, 43.
- [3] F. K. Donkor, K. Mearns, in *Affordable and Clean Energy* (Eds: W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, T. Wall), Springer International Publishing, Cham **2020**, pp. 1–9.
- [4] R. Raman, D. Pattnaik, H. H. Lathabai, C. Kumar, K. Govindan, P. Nedungadi, *J. Big Data* **2024**, *11*, 55.
- [5] A. van Wynsberghe, *AI Ethics* **2021**, *1*, 213.
- [6] K. R. Kaja, S. Hajra, S. Panda, M. A. Belal, P. Pakawanit, N. Vittayakorn, C. Bowen, H. Khanbareh, H. J. Kim, *Adv. Sustain. Syst.* **2024**, DOI: [10.1002/adus.202400678](https://doi.org/10.1002/adus.202400678).
- [7] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, S. Misra, *Electronics* **2022**, *11*, 3795.

- [8] J. Yang, K. Hong, Y. Hao, X. Zhu, Y. Qin, W. Su, H. Zhang, C. Zhang, Z. L. Wang, X. Li, *Adv. Mater. Technol.* **2025**, *10*, 2400554.
- [9] R. K. Rajaboina, U. K. Khanapuram, A. Kulandaivel, *Adv. Sensor Res.* **2024**, *3*, 2400045.
- [10] A.-L. Georgescu, A. Pappalardo, H. Cucu, M. Blott, *Eurasip J. Audio Speech Music Process.* **2021**, *2021*, 28.
- [11] A. M. Tripathi, A. Mishra, *Appl. Acoust.* **2021**, *182*, 108183.
- [12] Z. Lin, S. Duan, M. Liu, C. Dang, S. Qian, L. Zhang, H. Wang, W. Yan, M. Zhu, *Adv. Mater.* **2024**, *36*, 2306880.
- [13] M. H. Bagheri, J. Li, E. Gu, K. Habashy, M. M. Rana, A. A. Khan, Y. Zhang, G. Xiao, P. Xi, D. Ban, *Adv. Sensor Res.* **2024**, *4*, 2400156.
- [14] M. Chi, S. Chen, J. Jiao, N. Yu, *J. Appl. Phys.* **2023**, *133*, 245105.
- [15] M. A. Pillai, E. Deenadayalan, *Int. J. Precis. Eng. Manuf.* **2014**, *15*, 949.
- [16] G. Zhu, Y. Zhou, Z. Si, Y. Cheng, F. Wu, H. Wang, Y. Pan, J. Xie, C. Li, A. Chen, R. Wang, J. Sun, *Nano Energy* **2023**, *108*, 108237.
- [17] R. Tabassian, A. Rajabi-Abhari, M. Mahato, H. Yoo, H. Y. Yoon, J. Y. Park, I.-K. Oh, *SmartMat* **2024**, *5*, e1270.
- [18] K. Dong, X. Peng, J. An, A. C. Wang, J. Luo, B. Sun, J. Wang, Z. L. Wang, *Nat. Commun.* **2020**, *11*, 2868.
- [19] H. Yang, J. A. Lai, Q. Li, X. Zhang, X. Li, Q. Yang, Y. Hu, Y. Xi, Z. L. Wang, *Nano Energy* **2022**, *104*, 107932.
- [20] H. Guo, X. Pu, J. Chen, Y. Meng, M.-H. Yeh, G. Liu, Q. Tang, B. Chen, D. Liu, S. Qi, C. W. Wu, C. Hu, J. Wang, Z. L. Wang, *Sci. Robot.* **2018**, *3*, eaat2516.
- [21] T. H. Le, T. Van Le, V.-T. Bui, C. C. Nguyen, M. T. N. Dinh, N. M. Chau, V.-T. Bui, *Nano Energy* **2024**, *129*, 109963.
- [22] N. R. Tanguy, M. Rana, A. A. Khan, X. Zhang, N. Tratnik, H. Chen, D. Ban, N. Yan, *Nano Energy* **2022**, *98*, 107337.
- [23] X. Pu, C. Zhang, Z. L. Wang, *Natl. Sci. Rev.* **2022**, *10*, nwac170.
- [24] J. Wang, Y. Zi, S. Li, X. Chen, *MRS Energy Sustain.* **2020**, *6*, 17.
- [25] A. Chang, C. Uy, X. Xiao, X. Xiao, J. Chen, *Nano Energy* **2022**, *98*, 107282.
- [26] S. A. Zawawi, A. A. Hamzah, B. Y. Majlis, F. Mohd-Yasin, *Micromachines* **2020**, *11*, 484.
- [27] D. Jiang, M. Lian, M. Xu, Q. Sun, B. B. Xu, H. K. Thabet, S. M. El-Bahy, M. M. Ibrahim, M. Huang, Z. Guo, *Adv. Compos. Hybrid Mater.* **2023**, *6*, 57.
- [28] Y. Li, J. Yu, Y. Wei, Y. Wang, Z. Feng, L. Cheng, Z. Huo, Y. Lei, Q. Sun, *Sensors* **2023**, *23*, 1329.
- [29] J. Jo, S. Panda, N. Kim, S. Hajra, S. Hwang, H. Song, J. Shukla, B. K. Panigrahi, V. Vivekananthan, J. Kim, P. G. Raju Achary, H. Keum, H. Kim, *J. Sci. Adv. Mater. Devices* **2024**, *9*, 100693.
- [30] S. Zhou, C. Jia, G. Shu, Z. Guan, H. Wu, J. Li, W. Ou-Yang, *Nano Energy* **2024**, *129*, 109951.
- [31] J. Hu, M. Iwamoto, X. Chen, *Nano-Micro Lett.* **2023**, *16*, 7.
- [32] A. Babu, S. Gupta, R. Katru, N. Madathil, A. Kulandaivel, P. Kodali, H. Divi, H. Borkar, U. K. Khanapuram, R. K. Rajaboina, *Energ. Technol.* **2024**, *12*, 2400796.
- [33] S. Paria, S. K. Si, S. K. Karan, A. K. Das, A. Maitra, R. Bera, L. Halder, A. Bera, A. De, B. B. Khatua, *J. Mater. Chem. A* **2019**, *7*, 3979.
- [34] H. Zhang, X. Gong, X. Li, *J. Mater. Chem. A* **2023**, *11*, 24454.
- [35] A. Panda, K. K. Das, K. R. Kaja, M. Belal, B. K. Panigrahi, *J. Met. Mater. Miner.* **2024**, *34*, 2170.
- [36] A. Rajabi-Abhari, P. Li, M. H. Bagheri, A. A. Khan, C. Hao, N. R. Tanguy, D. Ban, L. Yu, N. Yan, *Nano Energy* **2024**, *131*, 110306.
- [37] W. Sun, J. Chen, T. Yuan, D. Sui, J. Zhou, *Nano Energy* **2024**, *128*, 109913.
- [38] A. Babu, P. Malik, N. Das, D. Mandal, *Small* **2022**, *18*, 2201331.
- [39] X. Tao, X. Chen, Z. L. Wang, *Energy Environ. Sci.* **2023**, *16*, 3654.
- [40] Y. S. Choi, S.-W. Kim, S. Kar-Narayan, *Adv. Energy Mater.* **2021**, *11*, 2003802.
- [41] R. S. Puppala, K. Prakash, R. R. Kumar, M. F. Hashmi, K. U. Kumar, 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS). Nagpur, India, **2023**, pp. 1–5.
- [42] R. Zhang, *Cell Rep. Phys. Sci.* **2024**, *5*, 101888.
- [43] A. O. Albaji, R. B. A. Rashid, S. Z. Abdul Hamid, *J. Electr. Comput. Eng.* **2023**, *2023*, 3615137.
- [44] S. Bhattacharya, N. Das, S. Sahu, A. Mondal, S. Borah, in *Lecture Notes in Networks and Systems* (Eds: V.H. Patil, N. Dey, P.N. Mahalle, M. Shafi Pathan, V.V. Kimbahune), Springer, Singapore **2021**.
- [45] S. Chu, S. Narayanan, C. C. J. Kuo, *IEEE Trans. Audio Speech Lang. Process.* **2009**, *17*, 1142.
- [46] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, *arXiv. preprint arXiv:1704.04861* **2017**, DOI: [10.48550/arXiv.1704.04861](https://doi.org/10.48550/arXiv.1704.04861).
- [47] K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition*, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, **2016**.
- [48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, **2016**.
- [49] F. Chollet, *Xception: Deep Learning with Depthwise Separable Convolutions*, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, **2017**.
- [50] X. Liu, Z. Jia, X. Hou, M. Fu, L. Ma, Q. Sun, *Real-time Marine Animal Images Classification by Embedded System Based on Mobilenet and Transfer Learning*, OCEANS 2019 – Marseille, Marseille, France, **2019**.
- [51] L. Gang, Z. Haixuan, E. Lanning, Z. Ling, L. Yu, Z. Juming, *Med. Phys.* **2021**, *48*, 4304.
- [52] K. J. Piczak, *ESC: Dataset for Environmental Sound Classification*, Proceedings of the 23rd ACM international conference on Multimedia (MM '15), New York, NY, USA, **2015**.
- [53] D. Ranmal, P. Ranasinghe, T. Paranayapa, *Sensors* **2024**, *24*, 3749.
- [54] G. Jekaterýnczuk, Z. Piotrowski, *Sensors* **2023**, *24*, 68.
- [55] L. Chen, Q. Shi, Y. Sun, T. Nguyen, C. Lee, S. Soh, *Adv. Mater.* **2018**, *30*, 1802405.
- [56] S. Wang, Y. Zi, Y. S. Zhou, S. Li, F. Fan, L. Lin, Z. L. Wang, *J. Mater. Chem. A* **2016**, *4*, 3728.
- [57] Y. Tang, B. Xu, Y. Gao, Z. Li, D. Tan, M. Li, Y. Liu, J. Huang, *Nano Energy* **2022**, *103*, 107833.
- [58] G. M. Rani, C.-M. Wu, K. G. Motora, R. Umapathi, C. R. M. Jose, *Nano Energy* **2023**, *108*, 108211.
- [59] S. Kuntharin, V. Harnchana, A. Klamchuen, K. Sinthiptharakoon, P. Thongbai, V. Amornkitbamrung, P. Chindaprasirt, *ACS Sustain. Chem. Eng.* **2022**, *10*, 4588.
- [60] A. Rajabi-Abhari, J.-N. Kim, J. Lee, R. Tabassian, M. Mahato, H. J. Youn, H. Lee, I.-K. Oh, *ACS Appl. Mater. Interfaces* **2021**, *13*, 219.
- [61] K. R. Kaja, S. Hajra, S. Panda, M. A. Belal, U. Pharino, H. Khanbareh, N. Vittayakorn, V. Vivekananthan, C. Bowen, H. J. Kim, *Nano Energy* **2024**, *131*, 110319.
- [62] A. Chen, C. Zhang, G. Zhu, Z. L. Wang, *Adv. Sci.* **2020**, *7*, 2000186.
- [63] W. Sun, Z. Jiang, X. Xu, Q. Han, F. Chu, *Int. J. Non Linear Mech.* **2021**, *136*, 103773.
- [64] N. Arora, T. Starner, G. D. Abowd, *Commun. ACM* **2020**, *63*, 12.
- [65] H. Shao, H. Wang, Y. Cao, X. Ding, J. Fang, H. Niu, W. Wang, C. Lang, T. Lin, *Nano Energy* **2020**, *75*, 104956.
- [66] S. Chai, X. Liu, X. Wu, Y. Xiong, *Sensors* **2021**, *21*, 279.
- [67] J. Wang, H. Liu, S. Han, G. Sun, X. Hu, *Appl. Acoust.* **2025**, *227*, 110258.
- [68] J. Hillenbrand, S. Haberzettl, G. M. Sessler, *J. Acoust. Soc. Am.* **2013**, *134*, EL499.
- [69] M. Qu, X. Chen, D. Yang, D. Li, K. Zhu, X. Guo, J. Xie, *J. Micromech. Microeng.* **2022**, *32*, 014001.
- [70] Y. Li, C. Liu, S. Hu, P. Sun, L. Fang, S. Lazarouk, V. Labunov, W. Yang, D. Li, K. Fan, G. Wang, L. Dong, L. Che, *Acoust. Aust.* **2022**, *50*, 383.
- [71] X. Hui, L. Tang, D. Zhang, S. Yan, D. Li, J. Chen, F. Wu, Z. L. Wang, H. Guo, *Adv. Mater.* **2024**, *36*, 2401508.
- [72] J. H. Han, J.-H. Kwak, D. J. Joe, S. K. Hong, H. S. Wang, J. H. Park, S. Hur, K. J. Lee, *Nano Energy* **2018**, *53*, 198.

- [73] G. Von Békésy, *Nature* **1970**, 225, 1207.
- [74] H. S. Lee, J. Chung, G.-T. Hwang, C. K. Jeong, Y. Jung, J.-H. Kwak, H. Kang, M. Byun, W. D. Kim, S. Hur, S.-H. Oh, K. J. Lee, *Adv. Funct. Mater.* **2014**, 24, 6914.
- [75] J. Li, R. Li, X. Ruan, *IET Renew. Power Generation* **2023**, 17, 1106.
- [76] X. He, H. Zhang, J. Jiang, X. Liu, *Nanotechnology* **2023**, 34, 155403.
- [77] Y. Zhang, T. Hu, R. Hu, S. Jiang, C. Zhang, H. Hou, *Molecules* **2022**, 27, 8896.
- [78] J. Yun, J. Park, M. Ryoo, N. Kitchamsetti, T. S. Goh, D. Kim, *Nano Energy* **2023**, 105, 108018.
- [79] K. Lee, H. Han, J. H. Ryu, S. Kang, K. Jung, Y.-K. Kim, T. Song, S. Mhin, K. M. Kim, *Carbon* **2023**, 212, 118120.
- [80] S. Niu, S. Wang, L. Lin, Y. Liu, Y. S. Zhou, Y. Hu, Z. L. Wang, *Energy Environ. Sci.* **2013**, 6, 3576.
- [81] H. Sun, X. Gao, L.-Y. Guo, L.-Q. Tao, Z. H. Guo, Y. Shao, T. Cui, Y. Yang, X. Pu, T.-L. Ren, *InfoMat* **2023**, 5, e12385.
- [82] X. Yu, Y. Shang, L. Zheng, K. Wang, *ACS Appl. Electron. Mater.* **2023**, 5, 5240.
- [83] F. Jean, M. U. Khan, A. Alazzam, B. Mohammad, *J. Sci. Adv. Mater. Devices* **2024**, 9, 100805.
- [84] Y. Kurihara, T. Kaburagi, K. Watanabe, *IEEE Sensors J.* **2016**, 16, 1772.
- [85] Y. Kurihara, T. Kaburagi, K. Watanabe, *IEEE Sensors J.* **2015**, DOI: [10.1109/JSEN.2019.2894744](https://doi.org/10.1109/JSEN.2019.2894744).
- [86] D. Theng, K. K. Bhojar, *Knowl. Inf. Syst.* **2024**, 66, 1575.
- [87] W.-G. Kim, D.-W. Kim, I.-W. Tcho, J.-K. Kim, M.-S. Kim, Y.-K. Choi, *ACS Nano* **2021**, 15, 258.
- [88] M. H. Bagheri, A. A. Khan, S. Shahzadi, M. M. Rana, M. S. Hasan, D. Ban, *Nano Energy* **2024**, 120, 109101.
- [89] A. A. Khan, R. Saritas, M. M. Rana, N. Tanguy, W. Zhu, N. Mei, S. Kokilathasan, S. Rassel, Z. Leonenko, N. Yan, E. Abdel-Rahman, D. Ban, *ACS Appl. Mater. Interfaces* **2022**, 14, 4119.
- [90] M. M. Rana, A. A. Khan, W. Zhu, M. F. A. Fattah, S. Kokilathasan, S. Rassel, R. Bernard, S. Ababou-Girard, P. Turban, S. Xu, C. Wang, D. Ban, *Nano Energy* **2022**, 101, 107631.
- [91] M. M. Rana, A. A. Khan, G. Huang, N. Mei, R. Saritas, B. Wen, S. Zhang, P. Voss, E. Abdel-Rahman, Z. Leonenko, S. Islam, D. Ban, *ACS Appl. Mater. Interfaces* **2020**, 12, 47503.
- [92] M. Lamrini, M. Y. Chkouri, A. Touhafi, *Sensors* **2023**, 23, 6227.
- [93] B. da Silva, A. W. Happi, A. Braeken, A. Touhafi, *Appl. Sci.* **2019**, 9, 3885.
- [94] R. M. Alsina-Pagès, J. Navarro, F. Alías, M. Hervás, *Sensors* **2017**, 17, 854.
- [95] S. Niu, Z. L. Wang, *Nano Energy* **2015**, 14, 161.
- [96] R. D. I. G. Dharmasena, J. H. B. Deane, S. R. P. Silva, *Adv. Energy Mater.* **2018**, 8, 1802190.
- [97] H. J. Hwang, D. Choi, *Funct. Compos. Struct.* **2021**, 3, 025004.
- [98] S. Chao, H. Ouyang, D. Jiang, Y. Fan, Z. Li, *EcoMat* **2021**, 3, e12072.