Reviewer #2:

The adjective "robust," used by the authors in the title, encapsulates the quality of the study and clearly and unambiguously reflects the content of the article. The detection of infrasound associated with geophysical events such as earthquakes, volcanic eruptions, or gravitational movements is well-established and applied both in scientific contexts—such as the USGS's monitoring of volcanic activity or the CNR in Italy, inspired by Giovanni Gregori—and in professional settings for monitoring gravitational phenomena, particularly during major construction projects and infrastructure development. The manuscript addresses scientific and technical issues of significance both within the field of NHESS and within Earth Sciences. The creation of a network to distinguish the type of infrasound produced by natural events from that of anthropogenic origin represents a novel element in this area of research, as does the use of new technological devices compliant with international standards. The methodology adopted for conducting the research is effectively described, and the scientific approach is sound, as evidenced by the results that support the robustness of the technique used and by the study's conclusions. The description of the experimentation is sufficient to stimulate the interest of other research centers to pursue similar work. Both the title and the abstract are aimed at a specialized audience, given their technical nature. The size, quality, and legibility of each figure are not always appropriate for the type and amount of data presented. The authors give proper credit to previous or related work while specifying their own contribution. In this context, the number and quality of the references span a period of about thirty years and are appropriate and accessible for use by other scientists. The presentation of the study is well-structured but intended for scientists in the field, while the length of the paper, though substantial, is appropriate for an international scientific context, thanks in part to precise technical language and high-quality English. The document does not include supplementary material.

The authors could specify whether monitoring is continuous and active 24/7, and add a few words about the acronyms that, while clear to researchers in the field, could help non-specialist researchers—such as the acronym CTBTO. In lines 68, 70, and 72, there are three capital letters after the colons, and Figure 1 could use a bit more explanation. Furthermore, Figure 5, although illustrative, is not entirely clear. In line 314, on page 12, the meaning of "neurons" could be clarified.

Dear Reviewer,

Thank you very much for your careful reading and constructive comments on our manuscript. We highly appreciate the positive and encouraging feedback, and we have carefully addressed all the points raised. The revisions have been highlighted in the revised manuscript. Below is our point-by-point response.

1. Comment on continuous monitoring
The authors could specify whether monitoring is continuous and active 24/7.

Response: We thank the reviewer for this suggestion. We have clarified in the manuscript **(lines 108–111)** that the monitoring is indeed continuous and operates 24/7. The revised sentence reads: "The monitoring system runs continuously and actively 24 hours a day, 7 days a week."
The specific monitoring process is described as follows: The process by which the CTBTO's International Monitoring System (IMS) achieves global infrasound monitoring and identifies various types of events can be summarized as a rigorous workflow from "listening" to "identification." First, at the global monitoring level, the IMS plans to build 60 infrasound stations distributed across different countries (most of which have already been established). Each station consists of an array of 4 to 9 sensors, with inter-sensor distances of 1 to 3 kilometers. These stations operate continuously, 24 hours a day, 7 days a week, constantly collecting infrasound data from the atmosphere and transmitting them in real-time via dedicated communication links to the International Data Centre (IDC) in Vienna. The raw data then enter an automatic processing stage: algorithms such as PMCC (Progressive Multi-Channel Correlation) are used to detect valid infrasound signals from environmental noise, estimate the azimuth and speed of the sound source, and automatically cluster spatiotemporally correlated signals to form preliminary event lists (e.g., Standard Event Lists SEL1/SEL2). This is followed by a crucial manual review step—analysts use specialized interactive workstations, together with auxiliary information such as global atmospheric wind models, to carefully examine each automatically detected event, remove false events caused by anthropogenic interference or natural noise, correct location errors, and produce an analyst-reviewed standard event list. To further ensure reliability, the IMS also performs multi-technology cross-validation: infrasound monitoring results are spatiotemporally correlated with seismic, hydroacoustic, and even radionuclide monitoring data from the same region. Only events confirmed by multiple technologies are accepted as genuine events. Finally, the validated infrasound events are included in the Reviewed Event Bulletin (REB), classified by source type into known events (e.g., earthquakes, volcanic eruptions), unknown events, anthropogenic events (e.g., mining blasts, industrial explosions), and natural events, and serve as an official verification product issued by the CTBTO to its member states.

2. Comment on acronyms (e.g., CTBTO)
Add a few words about the acronyms that, while clear to researchers in the field, could help non-specialist researchers – such as the acronym CTBTO.

Response: We thank the reviewer for the valuable suggestion. In the revised manuscript, at its first occurrence, we have spelled out CTBTO as "Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)". We have also carefully checked other acronyms throughout the text and provided full names where appropriate. Specifically:

SVM: Support Vector Machine
LSTM: Long Short-Term Memory
CNN: Convolutional Neural Network
CTBT: Comprehensive Nuclear-Test-Ban Treaty
LOTIS: Library of Typical Infrasonic Signals
BiLSTM: Bidirectional Long Short-Term Memory
BiGRU: Bidirectional Gated Recurrent Unit
ReLU: Rectified Linear Unit

3. Comment on capital letters after colons (lines 68, 70, 72)
In lines 68, 70, and 72, there are three capital letters after the colons.

Response: We apologize for this oversight. We have corrected these three instances by changing the initial capital letter after the colon to lowercase, in accordance with standard English grammar.

4. Comment on Figure 1
Figure 1 could use a bit more explanation.

Response: We have expanded the caption of Fig. 1 and added a few sentences in the main text **(lines 97–100)** to better explain what is shown in the figure and how it supports the analysis. The revised caption now reads: [Figure 1: Schematic diagram of the infrasound sensor array located at Windless Bight, Antarctica. The array consists of multiple infrasound sensors (channels CHF1–CHF4 are shown as examples, with a bandwidth of 0.01–0.2 Hz, and their relative Cartesian coordinates in kilometers: CHF1 at (0, 0), CHF2 at (−2.4055, 5.6579), CHF3 at (5.4587, 3.0939), and CHF4 at (3.5853, −1.0567)).]

5. Comment on Figure 5
Figure 5, although illustrative, is not entirely clear.

Response: We thank the reviewer for the valuable comment regarding Fig. 5. To address the concern that the figure was "not entirely clear," we have made the following improvements in the revised manuscript:

Revised caption of Fig. 5: the original caption has been expanded to provide a more detailed description. The new caption now clearly indicates the key discriminative features across the three layers, including the time windows of energy differences and the frequency-dependent separations. The revised caption now reads: [Figure 5: Three-layer wavelet scattering features comparing a chemical explosion (left) with an earthquake (right). Key discriminative time windows (60–100 s in

Layer 1, after 120 s in Layer 3) and frequency-dependent energy separations (Layer 2) are indicated. This multi-layer visualization clarifies how scattering features capture source-specific signatures, supporting robust event classification.]

Enhanced readability: we have enlarged the font sizes of all labels, axes, and annotations in the figure to improve legibility.

Additional annotations: important differences between the chemical explosion and the earthquake signals have been directly marked on the figure, making the comparison more intuitive.

We thank the reviewer again for helping us strengthen the presentation of our work.

6. Comment on the meaning of "neurons" (line 314, page 12)
In line 314, on page 12, the meaning of "neurons" could be clarified.

Response: We have added a brief clarification in the manuscript. The revised sentence at **line 309** now reads: "···neurons (i.e., the fundamental computational units that receive input signals, compute a weighted sum, and apply a nonlinear activation function)···"

Additional revisions: We have also proofread the entire manuscript to correct any remaining typographical or formatting issues.

We sincerely thank the reviewer for the thorough and constructive review. We believe the revised manuscript has been significantly improved and hope it is now acceptable for publication in NHESS.

Sincerely,

Xihai Li