nature portfolio

Peer Review File

A semi-automated framework for global detection of previously undocumented submarine calderas

Corresponding Author: Dr Andrea Verolino

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Verolino,

Your manuscript titled "Global machine-learning detection of submarine calderas" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised substantial concerns that must be addressed. In light of these comments, extensive revisions will be required before we can further consider the manuscript for publication. We would, however, be interested in considering a revised version that fully addresses these serious concerns. Specifically, a revised manuscript must:

1. Provide a comprehensive and structured description of the methodology along with a transparent discussion of its assumptions and limitations.

2. Enhance the robustness of conclusions by compellingly demonstrating that the detection framework can reliably identify the most significant submarine calderas, particularly those central to hazard assessments.

3. Fully enhance the algorithm by addressing its current detection gaps and providing a thorough explanation of how it can be adapted or improved to capture overlooked calderas.

We hope you will find the reviewers' comments useful as you decide how to proceed. If additional work allows you to either incorporate or refute these criticisms, we will be happy to look at a substantially revised manuscript. If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

When resubmitting, please provide a point-by-point response to the reviewers' comments. Please submit your responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Domenico M. Doronzo, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-6866-8870

Alireza Bahadori, PhD
Senior Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

For Manuscripts that fall into the following fields:
• Behavioural and social science
• Ecological, evolutionary & environmental sciences
• Life sciences
An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript
You can download the form here:
https://www.nature.com/documents/nr-reporting-summary.pdf

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide (https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The manuscript by Verolino et al. presents an interesting application of a global machine-learning detection algorithm to identify submarine calderas. Using their adapted workflow, the authors report 78 candidate calderas and highlight eight shallow-water examples as high-priority targets given their perceived hazard potential. The approach builds on the algorithm by Lee and Horgan (2021), originally developed to detect Martian craters, which here is modified and applied to the GEBCO 2023 bathymetric dataset at 500 m resolution. The algorithm initially detects 514 depressions, which are subsequently filtered and classified through a combination of automatic and manual procedures. Manual inspection results in five broad classes (calderas, craters/small calderas, atolls, flank scars/eroded surfaces, and undefined features), with further subdivisions based on bathymetric contours, hillshade visualizations, and profiles. Principal component analysis is then applied to the resulting classes for quality control.

While the manuscript addresses an important observational gap in our knowledge of submarine volcanism, I find that it has major shortcomings that currently prevent publication. Below I detail the primary issues, followed by more minor comments.

Major Issues
1. Failure to identify key calderas
The algorithm fails to detect several of the most significant calderas, including Hunga Tonga, Kikai, Krakatau, Santorini, Kolumbo (note: this is more accurately a crater), and Axial. Only two previously known calderas (Niuatahi and Kemp) are correctly identified. This omission undermines the reliability of the method, particularly since these volcanoes are central to hazard assessments. The authors briefly acknowledge possible causes (e.g., DEM coverage or “technical limitations of the CDA”), but this needs deeper discussion. At a minimum, they should test their method on DEMs that include these well-documented systems, many of which have open-access high-resolution data. If such validation is not feasible, the limitations

must be explicitly and prominently addressed in the discussion, as they directly contradict the claim that the approach "demonstrates high efficacy for submarine caldera identification."

2. Heavy reliance on manual interpretation

Although framed as a machine-learning study, the approach is heavily dependent on manual inspection and classification. The ML algorithm only provides a first-order set of depressions; the actual categorization into volcanic and non-volcanic features is entirely manual. This is only briefly noted in passing (Lines 132–134). The paper should clearly state from the abstract onward (potentially also in the title) that manual interpretation plays a central role. Ideally, the authors should explore a more data-driven workflow, e.g., training an ML model on pre-defined classes to reduce reliance on subjective interpretation and align more closely with the stated goals.

3. Ambiguity between submarine and emergent calderas

The manuscript is titled as a study of "submarine calderas" and the introduction emphasizes submarine volcanism. Yet, emergent calderas (e.g., Santorini, Krakatau, Kikai) are also included in the discussion and hazard framing. The terminology needs to be clarified consistently. I recommend using "oceanic calderas" (covering both submarine and emergent systems) or otherwise restricting the discussion to truly submarine structures.

4. Global context of results

The manuscript emphasizes that existing databases list only a handful of submarine calderas (e.g., 6 in GVP, 2% in CCDB). However, it is unclear how the authors' 78 candidates advance these records. Does this meaningfully fill the global observational gap? This needs a dedicated discussion to situate the results in a broader context.

5. Morphological classes introduced abruptly

The inclusion of atolls, flank scars, and eroded surfaces as classes comes as a surprise. The introduction should prepare the reader by briefly outlining which other morphologies may resemble calderas in bathymetry, thereby justifying the classification scheme.

6. Inconsistent definitions of classes

The manuscript is confusing in how it defines calderas versus craters. For instance, Line 102–103 states that features >3 km are classified as calderas and <3 km are classified as craters, but then Lines 112–113 suggest both are grouped inclusively under a "Submarine Caldera Dataset." Later, the first-order classes again separate calderas from "craters/small calderas," followed by further subdivisions. Figure captions then introduce yet another variant of wording. These inconsistencies make it difficult for readers to follow the classification logic. A careful revision is needed to ensure consistent terminology throughout.

7. Priority sites lack visualization

The identification of eight "priority" calderas is a central outcome, yet the manuscript does not provide bathymetric figures for all of them. I would like to see them in the main text to substantiate their hazard significance.

8. The manuscript lacks an open data availability statement

Minor Comments

Line 14: Reword opening sentence to avoid overstatement. Suggested: "Submarine calderas are far less documented than their subaerial counterparts."

Lines 38–42: Include recent seismic crisis at Santorini linked to dike intrusion (Isken et al., 2025, Nature).

Isken, M. P., Karstens, J., Nomikou, P., Parks, M. M., Drouin, V., Rivalta, E., ... & Berndt, C. (2025). Volcanic crisis reveals coupled magma system at Santorini and Kolumbo. Nature, 645(8082), 939-945.

Lines 52–54: Clarify that "The eruption" refers to the Hunga Tonga event.

Lines 68–70: The sentence on hazardous calderas at convergent boundaries is unclear; rephrase for clarity.

Lines 195–224: This section largely repeats the introduction and should be shortened.

Line 284: The depression inside Kolumbo Volcano has to be classified as a crater, not a caldera:

Karstens, J., Crutchley, G. J., Hansteen, T. H., ... & Berndt, C. (2023). Cascading events during the 1650 tsunamigenic eruption of Kolumbo volcano. Nature Communications, 14(1), 6606.

Line 287: Cite Isken et al. (2025) to illustrate advances from integrated offshore–onshore monitoring.

Figure 2: Increase font size on scale bars; omit sub-kilometer scale bars.

Reviewer #2 (Remarks to the Author):

Review by William Chadwick

General Comments

This manuscript presents the results of a new machine-learning technique, combined with manual filtering and human-confirmation, applied to the task of identifying submarine calderas from a (relatively low-resolution, 500 m) global compilation of seafloor bathymetry (GEBCO). This is an interesting and potentially useful method with societally relevant implications, and (if confirmed) it would be a significant advance and of interest to a broad range of earth and marine scientists. However, I find that the current manuscript jumps to the results and their potential implications without first describing the method in enough detail to convince the reader of its efficacy and clearly exploring its limitations.

Therefore, I recommend major revisions before the manuscript be reconsidered for publication. I think a suitably revised manuscript could possibly be suitable for this journal, but I also wonder if a journal with a longer format (word-count) might be more appropriate to allow a fuller exploration and description of the method and its results.

In any case, I have made many editorial comments and suggestions in an annotated manuscript, with the main comments listed below.

Specific comments keyed to line numbers in the annotated manuscript

• Line 44: I feel that so far the Introduction is almost over emphasizing the hazards at calderas. Would it be worth

acknowledging, here or elsewhere, that not ALL calderas are silicic and explosive? Some are basaltic and have a different set of hazards. I assume your machine learning technique is composition agnostic, right? So some of the submarine calderas that will be detected by this method are not KILLERS, but still maybe be of scientific interest for other reasons. In other words, deadly hazards are not the ONLY reason that this method may be useful.
• Line 47: I think Crisp (1984) is a better (original) reference for this statement:
Crisp, J. A. (1984), Rates of magma emplacement and volcanic output, J. Volcanol. Geotherm. Res., 20, 177–211, doi:10.1016/0377-0273(84)90039-8.
• Line 57: Yeah, but I just looked at the GVP database, and many submarine volcanoes that I would ID as calderas are not identified as such in their database; for example, Havre and Monowai in the Kermadec arc (they are ID-ed as "composite"). Similarly, West Rota is a submarine caldera in the Mariana arc, but is not even in the GVP database, perhaps because the date of its last eruption is not known. Anyway, the GVP database can clearly be improved in this regard...
• Line 59: What does "visually assessed" mean here? Assessed from their geographic location?
• Line 62: I just want to make sure you are aware of this recent paper that includes a new database of documented submarine eruptions:
Maeno, F. (2025), Global database on surface phenomena and hazards of explosive submarine eruptions with application to hazards of the Kolumbo Volcanic Field, Bull. Volcanol., 87(87), doi:10.1007/s00445-025-01871-8.
• Lines 75-77: And even Kilauea volcano, Hawaii, has a history of highly explosive eruptions in its past.
Swanson, D. A., T. R. Rose, R. S. Fiske, and J. P. McGeehin (2012), Keanakāko'i Tephra produced by 300 years of explosive eruptions following collapse of Kīlauea's caldera in about 1500 CE, J. Volcanol. Geotherm. Res., 215-216, 8–25, doi:10.1016/j.jvolgeores.2011.11.009.
Swanson, D. A., T. R. Rose, A. E. Mucek, M. O. Garcia, R. S. Fiske, and L. G. Mastin (2014), Cycles of explosive and effusive eruptions at Kīlauea Volcano, Hawai'i, Geology, 42(7), 631–634, doi:10.1130/G35701.1.
• Line 82: You need to add what you are applying the ML algorithm to.
• Line 103: Personally, I would consider anything wider than 1 km a "caldera". Is the 3 km distinction from reference #22 (Branney & Acocella, 2015)?
• Line 115: I think this paragraph is premature and should either be omitted or moved to later in the paper. Before the authors can highlight 8 "high-priority targets", they need to describe the CDA method in more detail, discuss its limitations, show the results, and convince the reader that the method is effective and the results are valid. This has not been done yet.
To this point, on my first reading of the paper, I was wondering how many of the submarine calderas I know about were detected by this technique (like West Rota caldera in the Mariana arc, or Brothers and Monowai caldera in the Kermadec arc, or Axial Seamount in the NE Pacific, etc).. Turns out the answer was only 1 (Niuatahi), but that fact had not been revealed yet. So there needs to be much more discussion about what can and cannot be detected by this technique and why. And it needs to be openly stated that only 2 of the many already known submarine calderas were ID-ed by the technique. Then the paper must also convince the reader that the other 75 "hits" are valid.
• Line 129: I think there needs to be a Methods section inserted here in the main body of the paper (perhaps abbreviated, and expounded upon later) before the Results can be discussed (instead of putting ALL in the information about the methods in a separate section at the end of the main text). As is, there are too many questions left unanswered in the main text, and too many surprises found in the Method section. If the length limit of Communications Earth & Environment (5000 words) is not long enough to accommodate this, then another journal should be chosen instead. To me, this is essential for an effective paper, and the current organization of having all the information about Methods only described in a "supplemental" section is inadequate.
• Line 134: In Figure 2 and in the text, "horseshoe" does not need a hyphen in the middle.
• Line 136: I think you should include and cite Figure 5 here (after renumbering the figures) However, you'll need to clearly explain the difference between the CDA positive (blue) and negative (red) paths in the flow chart of Figure 5, which is not currently explained in the text until the Methods section.
• Line 144: This is the first mention of the PCA (which should be spelled out the first time it is used, even if it's in the section title). But it's unclear to me at what point in the work flow the PCA is used. Is it what narrowed the 514 candidates in the first sentence of this paragraph (before the CDA)? Or is it used after the CDA? Also, why it is needed or how is it useful? There should also be a sentence or two explaining what the PCA does and how it was set up and used.
• Line 153: What is the resolution of the Multibeam Bathymetry Database compared to GEBCO?
• Line 154: Again, clarification is needed here about the PCA and how it was used. Was it used to create the classifications in Figure 2? Or was is only applied afterward to generate Figure 3?
• Line 171: I think Figure 4 is inadequate. Or rather, another figure should be added with more detailed maps of sub-regions. For example, NE Pacific, equatorial eastern Pacific, equitorial western Pacific, SW Pacific, S Atlantic. Readers need better maps to see where these things are located.
• Line 172: The fact that only 2 previously known submarine calderas were detected is a big eye opener! This fact should be in the Abstract and should be revealed much earlier in the text! One implication of this revelation is that there are a LOT MORE submarine calderas out there, and that the detection level of your technique working with the current resolution of global bathymetry is relatively low.
How does the reader know which of the calderas in Supplementary Database 2 are Niuatahi & Kemp? They should be ID-ed in there somehow, or the database IDs given in the text here.
This makes me wonder what the global bathymetry (500m resolution) looks like over known calderas that are NOT detected (Monowai, Brothers, Havre, West Mata, Axial, etc) and then some analysis of why the ML method didn't identify them. All of those examples have been mapped by much higher resolution ship-based multibeam sonar data, by the way (what effect does that have in the GEBCO database, compared with areas without multibeam data? Also is multibeam data from NOAA/NCEI used with the ML technique at all?). I assume this issue of detection is a matter of bathymetry resolution, but perhaps also caldera size and depth (?). I think the paper should have a paragraph or two about that.
Another thought: shouldn't you add previously known and named submarine calderas to your submarine caldera database (a subset of known calderas NOT detected by your method). If one of the motivations of this study is hazards, then adding

other known calderas would be valuable from that perspective, I would think.
• Line 173: For Niuatahi caldera, you should add these references:
Embley, R. W., and K. H. Rubin (2018), Extensive young silicic volcanism produces large deep submarine lava flows in the NE Lau Basin, Bull. Volcanol., 80(36), doi:10.1007/s00445-018-1211-7.
Baker, E. T., S. L. Walker, G. J. Massoth, and J. A. Resing (2019), The NE Lau Basin: Widespread and abundant hydrothermal venting in the back-arc region behind a superfast subduction zone, Front. Mar. Sci., 6:382, doi:10.3389/fmars.2019.00382.
And for Kemp, you should add this reference:
Kürzinger, V., M. Römer, A. Lichtschlag, and G. Bohrmann (2023), Seafloor investigations of the Kemp Caldera, the southernmost arc caldera volcano from the South Sandwich island arc, Geochemistry, Geophysics, Geosystems, 24, e2023GC011024, doi:10.1029/2023GC011024.
• Line 174: Krakatau and Santorini seem like problematic examples here, since they are partly subaerial. There are many others in the Kermadec and Izu/Bonin/Mariana arcs.
I don't understand what "their location is outside the DEM" means, or why that is the case. Also, the "technical limitations of the CDA" should be discussed in the main text, not in the supplementary Methods section.
• Line 177: Only 9 calderas identified in volcanic arcs is certainly an underestimate!
• Line 196: The Maeno (2025) paper is relevant here.
• Line 199: An example reference about the biogeography of hydrothermal vents.:
Van Dover, C. L., C. R. German, K. G. Speer, L. M. Parson, and R. C. Vrijenhoek (2002), Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates, Science, 295(5558), 1253–1257.
• Lines 202-224: This paragraph is much longer than it needs to be. This is especially true since the ML method didn't even detect Hunga. The Hunga eruption demonstrated a wide range of hazards, we get it. But to me, this paper should be more about clearly demonstrating the effectiveness and limitations of this method, and less about hyping the hazards of calderas.
• Lines 240-263: You should add a figure that shows the detailed bathymetric maps of these 8 calderas.
• Line 261: Summit depth should be added to this table.
• Line 269: There is no reason to say "near future"!
• Line 280: Can you demonstrate that having ship-based bathymetry underlying the GEBCO (coarser) database makes a difference in submarine caldera detection vs. areas with only satellite-based inferred bathymetry? If so, that would be an interesting thing to add to the paper.
• Line 284: I think in-situ monitoring and real-time data is a pretty high and unrealistic bar to aim for in this context. A more realistic effort to advocate for might be remote hydroacoustic monitoring of calderas of interest, for example from the permanent CTBTO-IMS hydrophone arrays at Wake Island in the Pacific.
In any case, a good reference for an overview of the OOI cabled observatory at Axial Seamount is in:
Kelley, D. S., J. R. Delaney, and S. K. Juniper (2014), Establishing a new era of submarine volcanic observatories: Cabling Axial Seamount and the Endeavour Segment of the Juan de Fuca Ridge, Mar. Geol., 352, 426–450, doi:10.1016/j.margeo.2014.03.010.
And a paper on seafloor geodesy observations to complement the Wilcock et al. reference is:
Nooner, S. L., and W. W. Chadwick, Jr. (2016), Inflation-predictable behavior and co-eruption deformation at Axial Seamount, Science, 354(6318), 1399–1403, doi:10.1126/science.aah4666.
• Line 300: I don’t think you can say that this technique “demonstrates high efficacy”, since all but 2 of the known submarine caldera are NOT detected. Instead, I would say this method has pretty significant limitations! And, in my opinion, the paper does not really adequately explain why (is it mainly due to the limited resolution of the global bathymetric data? Or is it the limitations of the CDA?). Tests with satellite-only data vs. GEBCO data with underlying multibeam data vs. multibeam data at higher resolution would help better demonstrate that.
• Line 312: Seems like you should add a concluding comment like this at the end: “ Future improvements in the resolution of the GEBCO global bathymetric data compilation will likely build on these results and will identify even more submarine calderas that are below the current detection limit of this method.”
• Line 320: I think this figure should be in the main body of the paper and should be cited earlier. (see comment on line 136) This would help clarify the questions I had about when the PCA is applied, etc. However, as noted in my previous comment, the difference between the positive and negative results, needs to be explained before the figure is cited.
• Lines 328-329: This is an important piece of information that is missing from the main text, but I think should be included. The main text sort of implies that you surveyed the entire GEBCO dataset for calderas, but in fact you only surveyed a subset of GEBCO where seamounts had previously been identified. Is that correct?
• Line 370: There is no caption or description of Supplementary Figure 3, so it is difficult for the reader to understand its significance and relevance to the methods described here. Also the bathymetric maps for each of the 514 candidate calderas lack a color scale for the bathymetry, but the depths are shown in the radial profile cross-sections, so perhaps that is enough.
Also there is no explanation for the naming convention (A-# vs E-# vs SCD-# vs U-#). The reader can find this information in the "Dataset 1" Excel file (A=Atoll, E=Eroded, SCD= full caldera or eroded caldera or crater or small caldera, U=unidentified), but it shouldn't be this hard for the reader to figure this out! It should be explained in the text and in a caption.
I note that some text is overwritten and illegible on some adjacent images in Supp Figure 3. Specifically, the bottom x-axis label "Longitude" overwrites the top Lat/Lon location of the maps below due to bad vertically spacing between adjacent maps on the page - for example, SCD-0007/0008, SCD-0010/0011, SCD-0011/0012, etc. Perhaps the best way to correct this is to make the repeating legend at the top of every page smaller, so that the 3 individual maps on each page can be spaced further apart.
Maybe the SCD's (the real caldera candidates) should come first in the file, so the reader does not have to search for them?
All that said, I think this Supp. Figure 3 presents critical information that shows how the CDA evaluated the bathymetric data at each caldera candidate site, so it should definitely be kept as part of the manuscript.

Some other notes: SCD-0045 is actually a known seamount in the Mariana arc called Toto Caldera:
Gamo, T., et al. (2004), Discovery of a new hydrothermal venting site in the southernmost Mariana Arc: Al-rich hydrothermal plumes and white smoker activity associated with biogenic methane, Geochemical Journal, 38, 527–534.
Baker, E. T., R. W. Embley, S. L. Walker, J. A. Resing, J. E. Lupton, K. Nakamura, C. E. J. de Ronde, and G. J. Massoth (2008), Hydrothermal activity and volcano distributions along the Mariana Arc, J. Geophys. Res., 113, B08S09, doi:10.1029/2007JB005423.
Stern, R. J., Y. Tamura, H. Masuda, P. Fryer, F. Martinez, O. Ishizuku, and S. H. Bloomer (2013), How the Mariana Volcanic Arc ends in the south, Island Arc, 22, 133–148, doi:10.1111/iar.12008.
Nakagawa, T., K. Takai, Y. Suzuki, H. Hirayama, U. Konno, U. Tsunogai, and K. Horikoshi (2006), Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana Volcanic Arc, Environmental Microbiology, 8(1), 37–49, doi:10.1111/j.1462-2920.2005.00884.x.
Also, candidate U-11 appears pretty bogus, with a clearly bad pixel screwing up the color map, and a supposed caldera diameter of over 200 km! U-171 is similarly obviously bogus with "depths" up to 30 km ABOVE sealevel, and a caldera diameter over 100 km! Why weren't these ID-ed as bogus candidates? Or maybe the better question is: why weren't data gaps ignored (NaN) instead of given real (bogus) depth/elevation values?
I'm surprised there is not an upper limit constraint on possible caldera sites. This could eliminate many bogus candidate sites such as U-34 and U-35, which are clearly ridge-transform-fracture zone intersections, but with supposed (unrealistic) caldera diameters of 100-150 km.
• Lines 372-374: This sounds different than I thought I understood from the description at line 329, above, where it sounded like the Gevorgian database was used to restrict what parts of the GEBCO data to look at. Here, it sounds like the Gevorgian filter came AFTER a global search, in addition to a coral reef database. Which is it? This need clarification.
• Line 375: Also need clarification on how higher-resolution bathymetry was used. What does "inserted in the final dataset" mean? Was it used at the higher resolution (90m)? Or was it downgraded to the GEBCO resolution (500m)? Was the CDA applied to the higher-resolution data? If so, why didn't it identify more of the known submarine calderas?
• Line 376: In the Dataset 1 Excel file, units need to be added to columns L through S. Likewise, in the Dataset 2 Excel file, units need to be added to columns M through T.
• Line 403-415: I think this section on "Limitations" is inadequate. There should be more discussion in the paper (either here or in the main body of the text) of why many known submarine calderas were not detected by this method.

Reviewer #3 (Remarks to the Author):

#Reviewer comments
Manuscript: Global machine-learning detection of submarine calderas

The manuscript entitled “Global machine-learning detection of submarine calderas” presents a novel machine-learning-based approach to identify and characterise submarine calderas. This is a very interesting and valuable application of an existing algorithm, adapted for use in the marine environment. The study makes a meaningful contribution to the field, exposing the needs for improved methodologies to identify different features such as submarine volcanic calderas, particularly under-surveyed regions.
As noted in the manuscript, high-resolution seafloor mapping is time consuming and, in many areas, is logistically challenging. In this context, the proposed method provides an important step towards complementing field-based observations and expanding our capacity to investigate remote marine settings. However, I have some concerns regarding the datasets used, as well as the identification and interpretation of the output features. These issues require substantial clarification. If well revised, the paper might have its place in this journal.
Overall, I believe the paper has the potential to make a valuable contribution to the journal, provided that the authors address the concerns raised. Therefore, I recommend publication following major revision.

Essential Reviews:
- Dataset used for the analysis is GEBCO2023 bathymetry, which has a very low spatial resolution (~500m / 15 arc-second). Although the authors acknowledge this limitation, and refer to the Seabed2030 initiative as a future improvement, they do not discuss or incorporate other higher-resolution global bathymetric datasets that are currently available, such as DEM Global Mosaic and the Multibeam Bathy Mosaic, which could improve spatial resolution and enhance the feature detection. It is unclear why these higher-resolution datasets are not considered (e.g., these datasets are even mentioned in one of the references used in this study). Combining multiple global datasets, or using different datasets for different regions, could significantly improve the quality of the input data, and consequently, the reliability of the analysis.

-The manuscript lacks a sensitivity analysis of the CDA methodology. Including this is essential to demonstrate the robustness and reliability of the approach described here, and will help quantify how uncertainty of the inputs (e.g., low-resolution bathymetry) propagate through the analysis and influence the outputs of the analyses (i.e., submarine caldera identification and characterisation).

- The authors make a good point on the limitations of the existing database of submarine calderas. However, the dataset of documented calderas used in this study appears to be underrepresented (i.e., Fig. 1, Global distribution of documented calderas). For example, in the Tonga-Kermadec Arc region alone, a significantly large number of submarine calderas has been reported in the literature, yet this regions seems underrepresented. This raises concerns regarding the completeness of the reference dataset used for training and validation. It is not clearly explained how submarine calderas were selected from the CCDB; this information cannot be found in the cited reference provided, and the authors only states that features were ‘visually assessed’, without explaining the assessment criteria or validation procedure.

Although the authors emphasise the importance of developing an updated and comprehensive submarine caldera database, the current description of the data sources and selection methodology does not allow for proper validation or reproducibility. Further description of classification criteria is needed.

- It is not sufficiently clear why, out of the 78 detected volcanic depressions, only 8 submarine calderas were selected as high-priority targets. The authors refer to 'factors that collectively suggest an elevated likelihood of future eruptive activity and potential societal impact', yet the parameters described (i.e., water depth, tectonic settings, and caldera morphological type) appear insufficient to support determining the future activity and potential societal impact. Additional parameters, including magma composition, documented eruptive history, and proximity to coastal communities or critical infrastructure are needed. Without incorporating this additional information, the prioritisation of selected calderas appears oversimplified, and almost speculative.

- A major concern is that the CDA fails to identify several well-documented calderas, such as Hunga Tonga, Kikai, Krakatau, Santorini, etc…. The authors attribute this limitation to the fact that these calderas fall outside the DEM used. However, this explanation highlights a fundamental limitation in the study design rather than an unavoidable constrain. This issue could be addressed by adopting a more comprehensive or higher-resolution DEM used, or run the analyses in a region base, instead of global analyses. The omission of such well-characterised and widely studied calderas represents a major gap in the current implementation of the CDA, and raises concerns about the completeness and reliability of the detection of submarine calderas.
Additionally, these well documented calderas surveyed in high-resolution, provide an excellent opportunity for training and validating the algorithm. Incorporating these examples could substantially improve the model calibration, enhancing the detection performance and increasing the confidence of the outcome when identifying unknown submarine calderas worldwide.

- The growth stage classification used in this study to determine the volcanic hazard potential, should be reconsidered or complemented with more comprehensive framework presented by Acocella (2007). This classification represents one of the most extensive and systematic approaches to caldera development and evolution, incorporating not only water depth and volcano height, but also the magma composition, eruptive history, and also the magma reservoir characteristics (depth and volume). In its current form, the growth stage used in this study seems simplified and may not properly capture the complexity of caldera evolution and associated hazard potential. Integrating, or at least discussing the Acocella (2007) classification will provide a more robust basis of hazard interpretation and improve the scientific grounding of the prioritisation scheme.

- I'm not fully convinced with the morphological classification and features description presented in this manuscript. The classification is described as being based on 'a combination of GIS-based analyses' and 'visual assessments'; however, neither of these are described in sufficient detail in the methodology section. This lack of clarity and details limits the reproducibility and raises concerns regarding subjectivity in the interpretation of morphological types and sub-types.
The use of more explicit, semi -automated GIS-based tools will help reduce the subjectivity and strengthen the classification of the identified features, as these tools incorporate description of features, the definition of which are referenced and accepted by the scientific community. In this regard, established tools such as CoMMa (Arosio et al., 2024) and GA-SaMMT (Dove et al., 2020; Nanson et al., 2023), should be incorporated into the workflow, as they provide systematic geomorphometric analyses and would partially replace or complement manual visual interpretations. This would substantially improve methodological transparency, repeatability, and confidence in the proposed feature description.

Overall, limitations acknowledged in the manuscript suggest that the methodology, in its current implementation, is providing a minimum number of potential submarine calderas. This raises an important question regarding the effectiveness and completeness of the proposed detection framework.
While this approach is promising, significant methodological improvements are necessary to enhance the reliability of the presented CDA analysis, including: 1) include higher resolution multibeam datasets as baseline. Expanding the DEM coverage or alternatively conducting regional-scale analyses rather than a purely global assessment, could substantially improve detection accuracy and reduce false negatives; 2) use of well-documented submarine calderas mapped at high resolution, should be systematically incorporated into the training and validation dataset; 3) use of semi-automated GIS-based tools for features classification, should complement or partially replace visual assessments in order to reduce subjectivity and improve reproducibility in feature classification. Moreover, additional parameters should be also incorporated to classify the growth stage of identified calderas, such as the ones included in the framework presented by Acocella (2007).

In summary, while the study presents a promising and innovative approach to identify unknown submarine calderas, substantial methodological refinement is required before the proposed CDA framework can be considered robust for global application. Addressing these points would significantly strengthen the scientific contribution and overall impact of the manuscript.

Minor comments included in the attached pdf.

** Visit Nature Portfolio's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

For more information please visit http://www.springernature.com/orcid

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Verolino,

Your revised manuscript titled "A semi-automated framework for global detection of submarine calderas" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment, provided you reduce overstated claims and limit speculative implications, while clearly and systematically detailing the methodological strengths, limitations, and uncertainties of identifying submarine calderas from the GEBCO bathymetry data, along the lines recommends by our Reviewer #2.

If you can address these requests, we invite you to revise your paper one last time to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Portfolio, please visit https://www.nature.com/commsenv/open-access

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Domenico M. Doronzo, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-6866-8870

Alireza Bahadori, PhD
Senior Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have addressed my previous comments, and I find that the manuscript has improved significantly. I still have some doubts about the proposed method given that it misses some key volcanic calderas, but this limitation is now better discussed and I consider the manuscript a valuable step forward. Therefore I recommend publication after minor revision.
General comments:
• Please ensure that all figures use colour schemes that are accessible to colour-blind readers. The frequent use of red and green may be problematic.
• For all figures, ensure that all abbreviations are clearly defined in the captions.
• The font size jumps in the manuscript. Please check for consistency.
Minor comments:
• Please write out the abbreviation PCA in full in the abstract.
• Line 161: Please check the reference formatting for “2023^(25).”
• Line 300: Please expand the caption and define the abbreviations CDA and PCA.
• Lines 339–344: Please revise the wording to avoid starting multiple consecutive sentences with “This” or “These.”
• Figure 3: I recommend using distinct colours for panels A and B to avoid confusion. Additionally, please increase the font size of the y-axis labels in panels a and B.
• Please ensure consistent capitalization of specific features throughout the manuscript. For example, in lines 648–650, Axial Seamount is capitalized, whereas Kolumbo volcano is not.
• Figure 1 & 5: Please review the axis ticks. They appear to overlap with the figure content and are inconsistent (e.g., 90° is missing “N,” while the southern equivalent is not labeled).
• Line 652: Please define the abbreviation CTBTO-MS.

Reviewer #2 (Remarks to the Author):

Review by William Chadwick

General Comments
This is the second time I have reviewed this manuscript (this time the revised version). While the authors have made very substantial revisions that address some of the issues I raised in my previous review, I feel that it still needs additional revision before it is ready for publication. To me the main problem is that the authors are still overselling the significance and implications of the study, instead of providing more detailed information about the strengths and weaknesses of their method to identify submarine calderas from the GEBCO global bathymetry database. This just weakens their credibility and the actual impact of the work. Too much of the text is exaggerated claims and tangential hyping of volcanic hazards. As before, I have made many editorial comments and suggestions in an annotated manuscript, with the main comments listed below.

Specific comments keyed to line numbers in the annotated manuscript
• Lines 24-25: (Abstract) To me, this whole aspect of ID-ing 8 "hazardous" calderas is a bit arbitrary and unconvincing. I would omit this sentence from the abstract and replace it with one that reveals that many known calderas were not ID-ed by this method and a brief explanation of why not. That would be more relevant and honest information about the strengths and weaknesses of how this new method performs.
• Line 94: I think the following reference might be more appropriate here (and wherever else it is cited), since ref#14 is just a conference abstract:

Mastin, L. G., A. R. V. Eaton, and S. J. Cronin (2024), Did steam boost the height and growth rate of the giant Hunga eruption plume?, Bull. Volcanol., 86(64), doi:10.1007/s00445-024-01749-1.
• Line 180: The PDF format of the Tables in Supplementary files 1 & 2 as submitted for review, are unusable because the (unnumbered) columns are spread out and appear on many different pages, so the (unnumbered) rows are impossible to easily correlate between pages. This is a sloppy and ineffective way to present information to readers. Instead, these should be presented in Excel or tab-delimited or comma-delimited (.csv) files, or otherwise better formatted into more easily readable tables.
• Lines 227-228: The order and/or association is wrong here: Toto (listed 3rd) is in the Mariana arc (listed 2nd), and Kemp (listed 2nd) is in the South Sandwich arc (listed 3rd).
Also, "three" should be "four" in this sentence. In reviewing the maps of the 78 identified calderas in Supplementary File 3, I have identified another previously known caldera among the 78, just by eyeballing the bathymetry that happens to be familiar to me. SCD-0025 is a caldera on one of the Vance Seamounts, which is an off-axis chain on the western flank of the Juan de Fuca Ridge in the NE Pacific. Here are several references about the Vance Seamounts:
Portner, R. A., D. A. Clague, and J. B. Paduan (2014), Caldera formation and varied eruption styles on North Pacific seamounts: the clastic lithofacies record, Bull. Volcanol., 76(845), doi:10.1007/s00445-014-0845-3.
Clague, D. A., J. R. Reynolds, and A. S. Davis (2000), Near-ridge seamount chains in the northeastern Pacific Ocean, J. Geophys. Res., 105, 16541–16562.
Hammond, S. R. (1997), Offset caldera and crater collapse on Juan de Fuca ridge-flank volcanoes, Bull. Volcanol., 58, 617–627.
The fact that I keep finding previously known calderas among the 78 ID-ed in this study (I also found that Toto caldera was among them) shows to me that the authors are not very knowledgeable about existing calderas and have not been very careful about checking whether or not their 78 ID-ed calderas are previously known or not. Are there more known calderas in there?
This is also why I requested in my previous review that a figure be added to show the 78 ID-ed calderas on more detailed regional maps, to make it easier to see them in the context of known volcanic arcs, hot-spot trails, and mid-ocean ridge settings. Regional maps could likely reveal additional examples just by better knowing the geographic context of where their calderas are identified and not identified. This has not been done in this revised manuscript, and I still think that is a serious omission that makes it harder for readers to evaluate the performance of this method and the validity of the newly ID-ed calderas.
• Line 232: A short summary of what the sensitivity analysis showed should be added here.
• Line 245: (Figure 5) These maps of the 8 "high-priority" calderas would be more useful if they included depth information - with either labeled color ramps or labeled depth contours. Geographic information (lat, long axis labels) should also be added to each map. There needs to be some information added to the caption about how to identify these 8 calderas in the Supplementary File 3, which includes all 514 caldera candidates. It would help to re-order the 8 maps in ascending ID order, to make it easier to identify them in the Supplementary File.
• Lines 260-269: This paragraph is repetitive and hyperbolic. Delete.
• Line 287: I would greatly reduce this paragraph. To me, this "hazard analysis", based on supposed growth stage, is so crude and based on so little concrete information, that it does not warrant the amount of text devoted to it. To me, the overselling of potential hazards substantially weakens the paper and seems like exaggerated claims made hoping to make the results "more relevant", but instead just makes the whole endeavor less convincing and more suspect, especially because the hazards have already been hyped in earlier parts of the paper.
• Line 299: More detailed regional maps would help the reader see the setting of these calderas.
• Lines 360-369: To me, this stuff is not relevant to the scope of this study. I would omit.

Reviewer #3 (Remarks to the Author):

Dear Editor,

After careful evaluation, I find that the authors have responded satisfactory to all of the reviewers’ comments and concerns, and have followed the feedback provided.

After this second round of revisions, the revised manuscript has significantly improved in clarifying the methodological explanations and in the addition of a new sensitivity test section, which strengthens the overall quality and rigor of the presented framework.

The authors have also revised the discussion section to include the limitation of the methodology while highlighting the value and contribution of the framework.

Overall, I believe the manuscript now meets the standards required for publication in the journal. I have added a few minor reviews and comments, but I recommend acceptance of the manuscript pending these minor revisions.

benefits**

Response to reviewers

Dear reviewers,

We have substantially revised the manuscript in line with the proposed suggestions and have added clarifying statements where needed. In a few cases, we did not adopt specific suggestions because they were not aligned with the goals of this study; in those instances, we provided clear justifications. Our detailed responses are provided below (in blue), with updated line numbers reported.

To have the correct line number references as reported below, in Review mode in Word (for the tracked version > All Markup), please use “Show Markup” > “Balloons” > ”Show Revisions in Balloons”.

Best regards,
Dr Verolino, on behalf of all authors

Reviewer 1

The manuscript by Verolino et al. presents an interesting application of a global machine-learning detection algorithm to identify submarine calderas. Using their adapted workflow, the authors report 78 candidate calderas and highlight eight shallow-water examples as high-priority targets given their perceived hazard potential. The approach builds on the algorithm by Lee and Horgan (2021), originally developed to detect Martian craters, which here is modified and applied to the GEBCO 2023 bathymetric dataset at 500 m resolution. The algorithm initially detects 514 depressions, which are subsequently filtered and classified through a combination of automatic and manual procedures. Manual inspection results in five broad classes (calderas, craters/small calderas, atolls, flank scars/eroded surfaces, and undefined features), with further subdivisions based on bathymetric contours, hillshade visualizations, and profiles. Principal component analysis is then applied to the resulting classes for quality control.

While the manuscript addresses an important observational gap in our knowledge of submarine volcanism, I find that it has major shortcomings that currently prevent publication. Below I detail the primary issues, followed by more minor comments.

<u>Major Issues</u>
1. Failure to identify key calderas
The algorithm fails to detect several of the most significant calderas, including Hunga Tonga, Kikai, Krakatau, Santorini, Kolumbo (note: this is more accurately a crater), and Axial. Only two previously known calderas (Niuatahi and Kemp) are correctly identified. This omission undermines the reliability of the method, particularly since these volcanoes are central to hazard assessments. The authors briefly acknowledge possible causes (e.g., DEM coverage or “technical limitations of the CDA”), but this needs deeper discussion. At a minimum, they should test their method on DEMs that include these well-documented systems, many of which have open-access high-resolution data. If such validation is not feasible, the limitations must be explicitly and prominently addressed in the discussion, as they directly contradict the claim that the approach “demonstrates high efficacy for submarine caldera identification.”

We thank the reviewer for highlighting the importance of evaluating the CDA against well-documented calderas. In response, we added a dedicated Discussion subsection (“Under-detection of known submarine calderas") and performed a targeted sensitivity analysis using GEBCO images extracted around several key systems (Santorini, Kolumbo, Kikai, Krakatau, Hunga Tonga, Havre, West Rota, Brothers, Axial). This analysis, now presented in Supplementary File 4 and Table 4, shows which calderas can be detected by the CDA, at which scales, and with what confidence. The test results clarify that, besides known calderas being outside the original DEM, non-detections arise primarily from morphological complexity (e.g. highly irregular, eroded, or composite rims) that reduces circularity below the CDA threshold. We now explicitly discuss these limitations and their implications for future users, with new text in the discussion and methods. Importantly, the CDA remains effective for circular to sub-circular depressions and provides a reproducible, scalable first-order screening tool for global reconnaissance.

- New sub-section in the methods:

Edited/Added text (new lines: 904-934): "*Sensitivity test on well-documented submarine calderas*

Through our framework we detected 78 potential calderas, of which only three were previously documented (Niuatahi, Kemp, and Toto). While this aligns with the aim of identifying previously undocumented calderas, it also highlights some limitations of our CDA. Several factors contributed to the non-detection of known calderas. A primary reason is that many well studied systems fall outside the cropped DEM used in the main analysis (Table 4). To evaluate whether these calderas would have been detected by our CDA, and to understand why those located within the cropped DEM were missed, we performed a targeted sensitivity analysis. We extracted from the global GEBCO grid the regions surrounding several known submarine calderas (Table 4) and ran the CDA on these areas to assess detectability and confidence levels. The GEBCO grid has a resolution of 15 arc seconds and was divided into 256×256 pixels tiles (~1°×1°), which formed the basis of the global survey. For this test, each site was additionally oversampled to examine scale sensitivity by generating tiles at 0.25°, 0.50°, 0.75°, 1.00°, 1.26°, and 1.59°. The smaller tiles (0.25°, 0.50°, 0.75°) represent bilinear zoom-ins of the DEM, whereas the larger ones (1.00°, 1.26°, and 1.59°) correspond to the original mapping workflow. For each scale, the DEM was projected orthographically and paired with the corresponding CDA output, which highlights edges interpreted as potential caldera rims. These edge maps were then processed with a circle detection algorithm; rims reconstructed at ≥75% completeness were drawn as solid rings, and those at ≥50% as partial rings (Supplementary File 4).

The test results (Supplementary File 4 and Table 4) show that among the volcanoes excluded from the original analysis because they lay outside the cropped DEM, only Santorini and Kolumbo were detected, with Kolumbo identified at low confidence, probably due to its relatively limited size. Kikai and Krakatau were not detected at any resolution. We interpret this as a consequence of their irregular morphology, and presence of eroded features both underwater and subaerially, reducing their apparent circularity below the 50% threshold required by the CDA. In contrast, most volcanoes located within the original cropped DEM (Havre, Hunga Tonga, West Rota, Brothers) were detected when using the additional image sizes tested here (0.25°, 0.50°, 0.75°). Axial remained undetected due to its pronounced U-shaped geometry, which also falls below the 50% circularity criterion."

- New sub-section in the Discussion:

Edited/Added text (new lines: 603-621): "*Under-detection of well-documented submarine calderas*

Based on the sensitivity test conducted on known submarine calderas (see Methods for details; test results in Table 4 and Supplementary File 4), our CDA performs well in identifying calderas that exhibit certain structural characteristics, but it also presents clear limitations that future users should consider. The method is particularly effective at detecting circular to sub-circular depressions within the spatial scales and resolution for which it was designed, and it successfully identifies most known calderas within the cropped DEM when additional resolutions are applied. However, the CDA struggles with systems that have highly irregular, eroded and composite shapes, as these often fall below the circularity threshold required for detection (50%). Kikai and Krakatau are clear examples of *this behaviour. Kikai consists of two eroded islands with multiple cones, an irregular submerged rim, and a subaqueous lava dome on the caldera floor[66], producing a highly composite morphology. Similarly, Krakatau comprises Anak Krakatau growing on the rim of the older caldera, three eroded islands that represent remnants of the older rim, and an uneven, irregularly shaped submerged caldera floor[67]. These complex geometries reduce their detectability, as well as the detectability of other similarly irregular features worldwide (e.g. Apolaki caldera[68]). Despite these constraints, the CDA provides a reproducible and scalable component of the overall framework for global scale reconnaissance, offering a valuable first order screening tool that can guide future high resolution mapping, targeted surveys, and hazard assessments.*"

- Additionally, we added more motivations (In the methods – CDA section) and an extra sensitivity test (SF5) on high-resolution bathymetry, explaining why we did not incorporate it at this stage.

Edited/Added text (new lines: 722-782): "…*we used GEBCO because it is the only bathymetric dataset that provides full global coverage on a continuous ~500 m/pixel grid. Other globally complete bathymetric products do exist, but at even lower resolution (e.g. ETOPO1, ~1,800 m at the equator). Having a homogeneous grid is essential for running an automated machine-learning detection algorithm consistently across all ocean basins. Higher-resolution datasets such as the MBBDB[42] do not offer global coverage, are spatially heterogeneous with substantial data gaps, and contain numerous artefacts arising from variable data quality and post-processing procedures. These issues currently prevent their integration into a uniform global-scale detection workflow. While we acknowledge the potential utility of multibeam compilations, and anticipate that future targeted studies building on approaches similar to ours are underway and will help address these limitations, integrating them into our current framework would introduce additional uncertainty. To evaluate this briefly, we applied our CDA to the MBBDB, and the results are presented in Supplementary File 5. This test shows that, although some detected features appear real, the majority of the output features are false positives, particularly in areas with incomplete coverage or processing artefacts. For these reasons, we did not incorporate multibeam-based detections into our global detection/classification framework at this stage. Importantly, where multibeam data were available, we used them in a qualitative way for the post-detection manual classification*"

2. Heavy reliance on manual interpretation
Although framed as a machine-learning study, the approach is heavily dependent on manual inspection and classification. The ML algorithm only provides a first-order set of depressions; the actual categorization into volcanic and non-volcanic features is entirely manual. This is only briefly noted in passing (Lines 132–134). The paper should clearly state from the abstract onward (potentially also in the title) that manual interpretation plays a central role. Ideally, the authors should explore a more data-driven workflow, e.g., training an ML model on pre-defined classes to reduce reliance on subjective interpretation and align more closely with the stated goals.

We thank the reviewer for this important observation. We agree that manual interpretation represents an important component of our workflow. Our intention was not to present the study as fully automated, but rather semi-automated, to demonstrate how a machine-learning–based detection algorithm can be used as a global pre-screening tool to identify possible caldera depressions, then refined and validated through expert-based classification and principal component analysis. To address this comment, we have modified the title, expanded the abstract, and clarified the workflow in the introduction:

New title: "*A semi-automated framework for global detection of submarine calderas*"

Edited abstract:

Edited/added text (new lines: 19-21): "*Here we apply a hybrid framework to the GEBCO global bathymetry that combines a depression detection algorithm, expert classification, and PCA-supported validation, to systematically identify previously undocumented submarine calderas.*"

Edited Introduction:

Edited/added text (new lines: 192-207): "*In this study, we applied a* ***hybrid framework (Fig. 2) that includes a*** *machine-learning-based caldera detection algorithm (CDA),* ***applied to the GEBCO[25] global bathymetric database, and an expert-based component****, to systematically identify submarine calderas worldwide* ***(refer to the Methods section for all details on the framework). To note that we use the term "submarine calderas" to refer broadly to calderas formed in oceanic settings, including both fully submerged structures and partially emerged systems (e.g. Santorini, Kikai, Krakatau).*** *Our CDA builds upon the crater detection algorithm originally developed by Lee and Hogan[41] for detecting Martian impact craters, which demonstrated high detection precision (>75%) for sub-circular depressions ≥4 km in diameter with minimal training.* ***To adapt the algorithm from Lee and Hogan[41] for submarine environments, we incorporated*** *a series of morphometric and resolution filters to* ***identify potential*** *caldera* ***candidates*** *within* ***the*** *low-resolution GEBCO 2023 global bathymetry dataset[25], ~500 m resolution,* ***cropped around the ~43,000 seamounts from Gevorgian et al.[23], with a 100 km radius (see Methods for rational and details).*** *These filters* ***substantially*** *reduce****d*** *the number of false positives and enabled* ***subsequent expert-based*** *classification of the retained features, followed by Principal Component Analysis (PCA) to assess and validate the nature of the identified* ***edifices*** *(see Methods for details)*"

The issue of the model training has been largely addressed later in this file.

3. Ambiguity between submarine and emergent calderas
The manuscript is titled as a study of "submarine calderas" and the introduction emphasizes submarine volcanism. Yet, emergent calderas (e.g., Santorini, Krakatau, Kikai) are also included in the discussion and hazard framing. The terminology needs to be clarified consistently. I recommend using "oceanic calderas" (covering both submarine and emergent systems) or otherwise restricting the discussion to truly submarine structures.

We thank the reviewer for highlighting this important point. Our study focuses on calderas formed in oceanic settings, including both fully submarine and partially emergent systems. We agree that this distinction was not sufficiently explicit in the original manuscript, and we prefer to retain the term "submarine" rather than adopt "oceanic". To address this concern, we have now broadened and clarified our definition of "submarine caldera" in the Introduction, explicitly stating that it includes partially emerged systems.

Edited/added text (new lines: 195-197): "*To note that we use the term "submarine calderas" to refer broadly to calderas formed in oceanic settings, including both fully submerged structures and partially emerged systems (e.g. Santorini, Kikai, Krakatau).*"

4. Global context of results
The manuscript emphasizes that existing databases list only a handful of submarine calderas (e.g., 6 in GVP, 2% in CCDB). However, it is unclear how the authors' 78 candidates advance these records. Does this meaningfully fill the global observational gap? This needs a dedicated discussion to situate the results in a broader context.

We thank the reviewer for this helpful comment. We have now added an explicit paragraph in the Discussion section to situate our results within a global context. Specifically, we clarify that only ~30 submarine calderas were previously documented worldwide, and that our identification of 75 new calderas represents an increase of at least 200% in the global inventory. We also emphasize that this figure remains a minimum estimate due to bathymetric resolution and methodological limitations. This addition highlights both the significance of our contribution and the remaining observational gap

Edited/added text (new lines: 673-695): "*Importantly, it successfully identified 78 calderas, 75 of which were detected for the first time, spanning diverse morphological types and growth stages, which in turn reflect their hazard potential[21]. Given that only about 30 to 40 submarine calderas were previously documented worldwide, our results represent an increase of at least 200% in the global inventory. While we acknowledge that these results constitute a minimum estimate due to bathymetric and methodological limitations, this expansion fills a relevant observational gap and generates actionable data to guide hazard prioritization, targeted field campaigns, and eventually future monitoring strategies. As such, this framework not only advances submarine volcanism research but also lays the foundation for integrating machine learning into future planetary-scale hazard assessments.*"

5. Morphological classes introduced abruptly

The inclusion of atolls, flank scars, and eroded surfaces as classes comes as a surprise. The introduction should prepare the reader by briefly outlining which other morphologies may resemble calderas in bathymetry, thereby justifying the classification scheme.

We appreciate the reviewer's comment and have revised the introduction to better prepare the reader for the classification scheme.

Edited/added text (new lines: 207-215): "*In addition to true calderas, the filtered CDA output also included several features that can resemble calderas from a geomorphological point of view. These were manually identified during expert classification and include atolls (i.e. drowned or shallow), and erosion or landslide related depressions (Fig. 3). Accounting for these alike structures is essential for the PCA based validation component of our framework, as their inclusion demonstrates the need for a classification scheme capable of distinguishing calderas from other circular or sub circular bathymetric features. After separating these alike structures from true calderas, our new resulting Submarine Calderas Dataset (SCD)….*"

6. Inconsistent definitions of classes

The manuscript is confusing in how it defines calderas versus craters. For instance, Line 102–103 states that features >3 km are classified as calderas and <3 km are classified as craters, but then Lines 112–113 suggest both are grouped inclusively under a "Submarine Caldera Dataset." Later, the first-order classes again separate calderas from "craters/small calderas," followed by further subdivisions. Figure captions then introduce yet another variant of wording. These inconsistencies make it difficult for readers to follow the classification logic. A careful revision is needed to ensure consistent terminology throughout.

We appreciate the reviewer's comment which highlights the need for clarity in these regards. The reason why we initially classified potential calderas as "calderas" or "craters" was solely for the purpose of the internal classification exercise, in order to test the performance of the CDA (e.g. does it detect small depressions?) and to evaluate how well the PCA could distinguish different morphological groups. However, when it comes to the final calderas dataset (SCD), we adopted an inclusive approach for the reasons largely explained in the manuscript (e.g. there is no universal cutoff size threshold between calderas and crater, and we could not apply one here). Nevertheless, when we "shortlist" the 8 calderas of interest, we simply exclude the small depressions because less relevant from a hazard point of view.

We have added a clarifying paragraph in the introduction explaining the rationale behind the temporary classification scheme and the inclusive nature of the final dataset.

Edited/added text (new lines: 214-297): "***After separating these alike structures from true calderas, our new resulting Submarine Calderas Dataset (SCD)*** *comprises 78 volcanic depressions larger than 1 km in diameter* ***(only 3 of which were previously identified as known calderas)****, which we* ***initially*** *classified as either calderas (diameter >3 km) or volcanic craters (diameter <3 km). Based on global consensus[27], calderas are broad topographic depressions formed by subvertical subsidence into a partially drained magma reservoir, typically associated with large eruptive events, whereas craters are generally smaller, bowl- or funnel-shaped features produced by explosive fragmentation and ejection of host rock near volcanic vents. Importantly, these definitions lack a definitive size threshold for distinguishing calderas from craters, and notable databases such as the CCDB list calderas even smaller than 1 km, creating potential overlap with crater dimensions. Given this ambiguity, compounded by the limited resolution of* ***the GEBCO*** *global bathymetric dataset and the tendency for both feature types to share comparable size ranges in submarine settings,* ***we adopted an inclusive classification approach in this study, grouping both calderas and craters under the SCD. We emphasize that the initial subdivision into calderas and craters was used solely for the internal classification workflow to evaluate the performance of the CDA and to support PCA-based validation. For the final dataset, all volcanic depressions >1 km in diameter are treated collectively. This inclusive approach ensures that potentially significant volcanic structures are not excluded due to arbitrary size cutoffs and is particularly suited to submarine environments where direct observation remains limited****.*"

7. Priority sites lack visualization
The identification of eight "priority" calderas is a central outcome, yet the manuscript does not provide bathymetric figures for all of them. I would like to see them in the main text to substantiate their hazard significance.

We have revised Fig. 4 (now Fig. 5) to include hillshade visualizations of all eight high-priority calderas (GEBCO + MBBDB where available). These visualizations were not included initially because high-resolution MBBDB bathymetry is not available for all sites, however, we now follow the reviewer suggestion, as we realise that it is important to show them. We opted to use hillshades in this figure because they provide a more intuitive representation of caldera morphology, and overlaying bathymetry with transparency was not feasible given that both GEBCO and MBBDB datasets are present in most of these panels, however, the readers can still refer to the corresponding bathymetric maps and profiles in Supplementary File 3, as now reported in the figure caption.

Edited figure caption: "*Fig.* ***5|*** *Global map with the location of the classified sub-types of calderas and craters, and high-priority calderas. Undefined features are also reported (white circles), some of which may be calderas. The basemap is from GEBCO 202323.* ***The lower panels display the eight high priority calderas with underlying hillshades derived from GEBCO and, where available, MBBDB[42], with also the approximate caldera rims shown (dashed lines). Bathymetric maps and elevation profiles for these calderas are provided in Supplementary File 3.***"

8. The manuscript lacks an open data availability statement

The manuscript already includes a combined **Data and Code Availability** section, where we specify the accessibility of both the CDA code and the resulting datasets (filtered and unfiltered). This section appears at the end of the manuscript (immediately before the References), following the journal's standard structure for data availability statements.

To ensure full clarity, we have slightly revised the section, specifying that all datasets generated in this study are openly accessible.

Edited/added text (new lines: 964-968): "*The modified CDA used **in this study is openly available** at https://github.com/eelsirhc/deepcaldera_project. **All datasets generated in this work, including the unfiltered and filtered outputs of the CDA, are openly accessible through** the Nanyang Technological University data repository: https://researchdata.ntu.edu.sg/privateurl.xhtml?token=643346d9-b122-49cc-9db6-ad1072a9958f*"

Minor Comments

Line 14: Reword opening sentence to avoid overstatement. Suggested: "Submarine calderas are far less documented than their subaerial counterparts."

Change applied as requested (new line: 16)

Lines 38–42: Include recent seismic crisis at Santorini linked to dike intrusion (Isken et al., 2025, Nature).
Isken, M. P., Karstens, J., Nomikou, P., Parks, M. M., Drouin, V., Rivalta, E., ... & Berndt, C. (2025). Volcanic crisis reveals coupled magma system at Santorini and Kolumbo. Nature, 645(8082), 939-945.

Reference with associated text added (new lines: 108-110): "…*and more recently, the seismic crisis along the Santorini-Kolumbo structural line (Greece), driven by dike intrusion, underscores the importance of monitoring active caldera systems*[11]"

Lines 52–54: Clarify that "The eruption" refers to the Hunga Tonga event.

Change applied (new line: 124) ("*The **Tonga** eruption triggered*…")

Lines 68–70: The sentence on hazardous calderas at convergent boundaries is unclear; rephrase for clarity.

We have added (new line: 169) "*and more mafic compositions are expected*", to guarantee more clarity to the sentence

Lines 195–224: This section largely repeats the introduction and should be shortened.

This paragraph has been substantially shortened (reduced by approximately half) to avoid repetition with the Introduction. We have retained a concise version here because it provides essential context for the Discussion and frames the relevance of our results.

Edited/added text (new lines: 481-490): "*In the context of risk, submerged calderas can generate high impact hazards, as demonstrated by the 2022 Hunga Tonga eruption, which produced powerful explosions, submarine density currents that severed communication cables, and basin scale tsunamis*[53,54]*. Such events show that shallow (stage 3) and partially emerged systems (stage 4)*[24]*, are especially prone to intensified explosivity due to magma–water interaction*[35–38]*. Historical examples such as the 1883 Krakatau eruption (Krakatau is now stage 4) similarly illustrate how partially submerged activity can trigger catastrophic tsunamis and widespread atmospheric effects*[55,56]*. These impacts highlight the potential vulnerability of submarine infrastructure and the need to better assess hazards associated with shallow and partially emerged submarine calderas*[57,58]*.*"

Line 284: The depression inside Kolumbo Volcano has to be classified as a crater, not a caldera: Karstens, J., Crutchley, G. J., Hansteen, T. H., ... & Berndt, C. (2023). Cascading events during the 1650 tsunamigenic eruption of Kolumbo volcano. Nature Communications, 14(1), 6606.

(new line: 647) We replaced “caldera” with “volcano” to be more generic in this case.

Line 287: Cite Isken et al. (2025) to illustrate advances from integrated offshore–onshore monitoring.

Citation added (new line: 653)

Figure 2: Increase font size on scale bars; omit sub-kilometer scale bars.

Changes applied to Fig. 2 (now Fig. 3) as suggested

Reviewer 2 (William Chadwick)

General Comments

This manuscript presents the results of a new machine-learning technique, combined with manual filtering and human-confirmation, applied to the task of identifying submarine calderas from a (relatively low-resolution, 500 m) global compilation of seafloor bathymetry (GEBCO). This is an interesting and potentially useful method with societally relevant implications, and (if confirmed) it would be a significant advance and of interest to a broad range of earth and marine scientists. However, I find that the current manuscript jumps to the results and their potential implications without first describing the method in enough detail to convince the reader of its efficacy and clearly exploring its limitations.

Therefore, I recommend major revisions before the manuscript be reconsidered for publication. I think a suitably revised manuscript could possibly be suitable for this journal, but I also wonder if a journal with a longer format (word-count) might be more appropriate to allow a fuller exploration and description of the method and its results.

In any case, I have made many editorial comments and suggestions in an annotated manuscript, with the main comments listed below.

Specific comments keyed to line numbers in the annotated manuscript

• Line 44: I feel that so far the Introduction is almost over emphasizing the hazards at calderas. Would it be worth acknowledging, here or elsewhere, that not ALL calderas are silicic and explosive? Some are basaltic and have a different set of hazards. I assume your machine learning technique is composition agnostic, right? So some of the submarine calderas that will be detected by this method are not KILLERS, but still maybe be of scientific interest for other reasons. In other words, deadly hazards are not the ONLY reason that this method may be useful.

We thank the reviewer for this helpful suggestion. We agree that not all calderas are silicic or highly explosive, and that many basaltic systems produce lower-intensity activity and different hazard profiles. Although this distinction was already mentioned later in the Introduction, we have now added an explicit statement earlier in the section to ensure this point is clear.

Edited/Added text (new lines: 110-115): “*It is important to note that not all calderas are silicic or highly explosive; many, particularly those in basaltic ridge settings, produce lower intensity activity and pose different types of hazards*[12]*. Identifying previously undocumented calderas, hence, remains essential not only for hazard assessment but also for reconstructing their eruptive histories and understanding their broader geological and ecological significance.*”

• Line 47: I think Crisp (1984) is a better (original) reference for this statement:
Crisp, J. A. (1984), Rates of magma emplacement and volcanic output, J. Volcanol. Geotherm. Res., 20, 177–211, doi:10.1016/0377-0273(84)90039-8.

(new line: 156) Reference added that replaces the previous one

• Line 57: Yeah, but I just looked at the GVP database, and many submarine volcanoes that I would ID as calderas are not identified as such in their database; for example, Havre and Monowai in the Kermadec arc (they are ID-ed as "composite"). Similarly, West Rota is a submarine caldera in the Mariana arc, but is not even in the GVP database, perhaps because the date of its last eruption is not known. Anyway, the GVP database can clearly be improved in this regard...

This is a valuable observation. We fully agree that the GVP mis-classifies or omits some submarine calderas, but despite of that, the misrepresentation of submarine calderas vs subaerial ones remains obvious. This is now more clearly addressed in the edited text

Edited/added text (new lines: 128-137): "*Existing global caldera databases highlight a **strong** imbalance between subaerial and submarine records. For example, the Global Volcanism Program (GVP) **lists** 140 Pleistocene-Holocene calderas, of which 94% (n= 132) are **above sea level** and only 6% (n= 8) **are fully submarine**[21]**. Several well known systems such as Santorini, Krakatau, and Kikai are predominantly subaqueous, yet their reported GVP elevations place them in the subaerial category; even if reclassified as submarine, the proportion would remain below ~8%. Other confirmed submarine calderas, such as Axial, Havre, and Monowai, are assigned to different volcanic types in the GVP (fissure feature, stratovolcano, and composite volcano, respectively), therefore are not included in the caldera count. Even when accounting for these misclassifications, submarine calderas would still represent <10% of the global inventory.***"

Additionally, the issue of this misrepresentation is not because the GVP only reports calderas with confirmed eruptions, since we also included Pleistocene calderas from the GVP, and this was already reported in the text ("*the Global Volcanism Program (GVP) lists 140 **Pleistocene-Holocene calderas***")

• Line 59: What does "visually assessed" mean here? Assessed from their geographic location?

That is correct, we used the locations of calderas from the CCDB, plotted in GIS, to assess which one was oceanic and below sea level. We added a clarifying part to that sentence ("*visually assessed **through a GIS-based assessment**, as water depth information is missing from the database*")

• Line 62: I just want to make sure you are aware of this recent paper that includes a new database of documented submarine eruptions:
Maeno, F. (2025), Global database on surface phenomena and hazards of explosive submarine eruptions with application to hazards of the Kolumbo Volcanic Field, Bull. Volcanol., 87(87), doi:10.1007/s00445-025-01871-8.

We thank the reviewer for the suggestion, and we were aware of this work, however, we believe that this reference will better fit later in the text, as suggest by the reviewer in a later comment.

• Lines 75-77: And even Kilauea volcano, Hawaii, has a history of highly explosive eruptions in its past.
Swanson, D. A., T. R. Rose, R. S. Fiske, and J. P. McGeehin (2012), Keanakākoʻi Tephra produced by

300 years of explosive eruptions following collapse of Kīlauea's caldera in about 1500 CE, J. Volcanol. Geotherm. Res., 215-216, 8–25, doi:10.1016/j.jvolgeores.2011.11.009.
Swanson, D. A., T. R. Rose, A. E. Mucek, M. O. Garcia, R. S. Fiske, and L. G. Mastin (2014), Cycles of explosive and effusive eruptions at Kīlauea Volcano, Hawai'i, Geology, 42(7), 631–634, doi:10.1130/G35701.1.

We added some text and one of the two references suggested by the reviewer (Swanson et al., 2014).

Added/edited text (new lines: 175-177): "*Intraplate volcanism, similarly, tends to produce basaltic calderas (e.g. Hawaii[29]), **explosive at times[30]**, but can also yield a broad compositional spectrum, from mafic to highly silicic, such as the Canary Islands[31] and Galápagos[32]*"

• Line 82: You need to add what you are applying the ML algorithm to.

Text added: "*In this study, we applied a hybrid framework (Fig. 2) that includes a machine-learning-based caldera detection algorithm (CDA), **applied to the GEBCO[25] global bathymetric database**, and …*"

• Line 103: Personally, I would consider anything wider than 1 km a "caldera". Is the 3 km distinction from reference #22 (Branney & Acocella, 2015)?

This size-based distinction between calderas and crater was only for classification-validation purpose. This has been clarified in the text already, as a response to a comment from reviewer 1. The choice of 3 km as a threshold is arbitrary, but again, this choice does not affect the final Submarine Caldera Dataset, as calderas and craters are both included under the same dataset (SCD). The reference of Branney and Acocella (2015) was about the next sentence (i.e. caldera definition).

• Line 115: I think this paragraph is premature and should either be omitted or moved to later in the paper. Before the authors can highlight 8 "high-priority targets", they need to describe the CDA method in more detail, discuss its limitations, show the results, and convince the reader that the method is effective and the results are valid. This has not been done yet.

We thank the reviewer for this observation. We agree that introducing a specific shortlist of high-priority calderas in the Introduction was premature, but since the indication of calderas of interest is part of this work, we wanted to retain this information in the introduction. To do so, we replace the old specific paragraph with a more general one.

Edited/added text (new lines: 301-306): "*In addition to expanding the global inventory of submarine calderas, our approach also enables the identification of calderas that may warrant closer scientific attention due to their hazard potential, as inferred from their water depths, tectonic settings, and morphological characteristics. Such features are important for first-order volcanic hazard assessments prior to direct exploration, underscoring the broader value of systematic detection frameworks for guiding subsequent investigation and hazard evaluation.*"

To this point, on my first reading of the paper, I was wondering how many of the submarine calderas I know about were detected by this technique (like West Rota caldera in the Mariana arc, or Brothers and Monowai caldera in the Kermadec arc, or Axial Seamount in the NE Pacific, etc). Turns out the answer was only 1 (Niuatahi), but that fact had not been revealed yet. So there needs to be much more discussion about what can and cannot be detected by this technique and why. And it needs to be openly stated that only 2 of the many already known submarine calderas were ID-ed by the technique. Then the paper must also convince the reader that the other 75 "hits" are valid.

This has been largely addressed now in response to a comment by reviewer 1. In short, we added a sub-section within the discussions (Under-detection of known submarine calderas), supported by a sensitivity test on several known submarine calderas (new section in the Methods too), with results reported in the new Table 4 and Supplementary File 4.

• Line 129: I think there needs to be a Methods section inserted here in the main body of the paper (perhaps abbreviated, and expounded upon later) before the Results can be discussed (instead of putting ALL in the information about the methods in a separate section at the end of the main text). As is, there are too many questions left unanswered in the main text, and too many surprises found in the Method section. If the length limit of Communications Earth & Environment (5000 words) is not long enough to accommodate this, then another journal should be chosen instead. To me, this is essential for an effective paper, and the current organization of having all the information about Methods only described in a "supplemental" section is inadequate.

We thank the reviewer for this comment. We fully agree that, in principle, presenting the Methods before the Results would improve the narrative flow of the manuscript. However, we followed the standard format required by *Communications Earth & Environment* (and Nature journals more broadly), in which the Methods are placed after the main text rather than integrated within it. Reordering the sections would therefore conflict with the journal's prescribed structure. To partially address the reviewer's concern, we have: 1) strengthened cross-referencing throughout the main text whenever details are needed to understand the Results. This ensures that no methodological information is overlooked and that the workflow remains transparent, while still adhering to the journal's formatting requirements; 2) added brief descriptions in the main text where necessary, but still referring to the methods for more details; 3) in some cases, we simply followed the reviewer's suggestion where we thought suitable (e.g. moving the figure of the framework, now Fig. 2, from the Methods to the Introduction).

• Line 134: In Figure 2 and in the text, "horseshoe" does not need a hyphen in the middle.

We applied this change to all figures and text.

• Line 136: I think you should include and cite Figure 5 here (after renumbering the figures) However, you'll need to clearly explain the difference between the CDA positive (blue) and negative (red) paths in the flow chart of Figure 5, which is not currently explained in the text until the Methods section.

We cited figure 5 (now fig. 2) as requested, and added a new paragraph introducing the positive and negative datasets. However, for full details we refer to the Methods section, for the reasons explained in response to a previous comment.

Edited/added text (new lines: 329-337): "*Our framework is composed of different components, a Caldera Detection Algorithm (CDA), an expert-based feature classification, and a validation step through Principal Component Analysis (PCA). The CDA was applied in parallel to two subsets to search for depressions (positive pathway) and mounds (negative pathway) (Fig. 2). Unlike meteorite impact craters, which are simple depressions in the ground, calderas are depressions superimposed on a pre-existing topographic mound. Therefore, in addition to the original algorithm proposed by Lee and Hogan (2021), which only looked for depressions (positive dataset), we introduced here a negative pathway (obtained by inverting the original CDA) to identify also mounds (see Methods for more details)*"

• Line 144: This is the first mention of the PCA (which should be spelled out the first time it is used, even if it's in the section title). But it's unclear to me at what point in the work flow the PCA is used. Is it what narrowed the 514 candidates in the first sentence of this paragraph (before the CDA)? Or is it used after the CDA? Also, why it is needed or how is it useful? There should also be a sentence or two explaining what the PCA does and how it was set up and used.

Now we added a paragraph in the results, briefly explaining when the PCA was conducted and why.

Edited/added text (new lines: 378-382): "*The PCA analysis (Fig. 4) was conducted after the manual feature classification and only applied to the CDA-positive filtered dataset (Table 2) (see Methods for rationale and details). Its purpose was to evaluate the consistency and reliability of our expert-based classification, serving as an independent validation step that assesses whether the manually defined groups form coherent and statistically meaningful clusters.*"

• Line 153: What is the resolution of the Multibeam Bathymetry Database compared to GEBCO?

The resolution (90 m) of the Multibeam bathymetry dataset has been added to the caption of Fig. 2 (now Fig. 3).

• Line 154: Again, clarification is needed here about the PCA and how it was used. Was it used to create the classifications in Figure 2? Or was is only applied afterward to generate Figure 3?

We thank the reviewer for pointing out the need for clarification. The PCA was not used to create the classifications shown in Fig. 2 (now Fig. 3). Instead, all features were first manually classified based on expert interpretation. PCA was then applied after this manual classification, and only to the subset of CDA-positive features, to independently assess the coherence and reliability of the expert-defined groups. We have revised the Results section to make this workflow explicit (the edited/added text is the same as in response to comment on line 144).

• Line 171: I think Figure 4 is inadequate. Or rather, another figure should be added with more detailed maps of sub-regions. For example, NE Pacific, equatorial eastern Pacific, equitorial western Pacific, SW Pacific, S Atlantic. Readers need better maps to see where these things are located.

We have modified Fig. 4 (now Fig. 5) to show the hillshades of the eight high-priority calderas, as already addressed in response to a previous comment. Ideally, two figures could be included: one showing the macro-regions highlighted by the reviewer, and another presenting the zoomed-in views of the calderas (as now provided). However, we believe it is more important to show the individual calderas rather than the broader regions, and since we have already moved one figure from the Methods into the main text, the manuscript is already around the upper limit of number of figures allowed by the journal. Nevertheless, if the editors prefer to have both figures and allow the additional insertion, we would be open to consider the implementation.

• Line 172: The fact that only 2 previously known submarine calderas were detected is a big eye opener! This fact should be in the Abstract and should be revealed much earlier in the text! One implication of this revelation is that there are a LOT MORE submarine calderas out there, and that the detection level of your technique working with the current resolution of global bathymetry is relatively low.

We thank the reviewer for this comment. We have now added this information to the Abstract and introduced it earlier in the manuscript to better prepare the reader for this key result. As noted in our responses to previous comments, the goal of this study is not to achieve completeness of the submarine calderas inventory, given the limitations of global bathymetry, but to establish a reproducible framework capable of identifying previously undocumented submarine calderas. Although the detection of 75 new calderas may appear modest at a global scale, it represents an increase of more than 200% relative to the number of submarine calderas previously documented worldwide. This contextualization has now been incorporated into the main text.

How does the reader know which of the calderas in Supplementary Database 2 are Niuatahi & Kemp? They should be ID-ed in there somehow, or the database IDs given in the text here.

We added a new column to Supplementary File 2 called "Known-caldera name". Now the reader can easily identify the previously known calderas from the Submarine Caldera Dataset.

This makes me wonder what the global bathymetry (500m resolution) looks like over known calderas that are NOT detected (Monowai, Brothers, Havre, West Mata, Axial, etc) and then some analysis of why the ML method didn't identify them. All of those examples have been mapped by much higher resolution ship-based multibeam sonar data, by the way (what effect does that have in the GEBCO database, compared with areas without multibeam data? Also is multibeam data from NOAA/NCEI used with the ML technique at all?). I assume this issue of detection is a matter of bathymetry resolution, but perhaps also caldera size and depth (?). I think the paper should have a paragraph or two about that.

All this has been now addressed in response to previous comments. We conducted 2 sensitivity analyses, 1) using GEBCO but just focused on areas where known calderas are located, and 2) using the MBBDB. New text and new supplementary files are reported, as previously indicated.

Another thought: shouldn't you add previously known and named submarine calderas to your submarine caldera database (a subset of known calderas NOT detected by your method). If one of the motivations of this study is hazards, then adding other known calderas would be valuable from that perspective, I would think.

We appreciate the reviewer's suggestion. However, the primary aim of this study is to develop and demonstrate a framework for the detection of previously undocumented submarine calderas using global bathymetry, rather than to compile an exhaustive global catalogue of all known submarine calderas.

That said, we agree that from a hazard-assessment perspective, previously known and well-studied submarine calderas remain highly relevant. We have therefore added a sentence in the Discussion noting that these systems should also be considered when evaluating regional volcanic hazard, together with the newly detected calderas.

Edited/added text (new lines: 582-585): "*This selection of newly detected calderas does not preclude the fact that previously known and well studied submarine calderas, such as Hunga Tonga, Kikai, Krakatau, Santorini, and others, remain critical for hazard assessments and should be considered alongside our prioritized targets when evaluating regional volcanic risk.*"

• Line 173: For Niuatahi caldera, you should add these references:
Embley, R. W., and K. H. Rubin (2018), Extensive young silicic volcanism produces large deep submarine lava flows in the NE Lau Basin, Bull. Volcanol., 80(36), doi:10.1007/s00445-018-1211-7.
Baker, E. T., S. L. Walker, G. J. Massoth, and J. A. Resing (2019), The NE Lau Basin: Widespread and abundant hydrothermal venting in the back-arc region behind a superfast subduction zone, Front. Mar. Sci., 6:382, doi:10.3389/fmars.2019.00382.
And for Kemp, you should add this reference:
Kürzinger, V., M. Römer, A. Lichtschlag, and G. Bohrmann (2023), Seafloor investigations of the Kemp Caldera, the southernmost arc caldera volcano from the South Sandwich island arc, Geochemistry, Geophysics, Geosystems, 24, e2023GC011024, doi:10.1029/2023GC011024.

(new line: 409) We added one reference per volcano (Embley and Rubin, 2018; Kürzinger et al., 2023)

• Line 174: Krakatau and Santorini seem like problematic examples here, since they are partly subaerial. There are many others in the Kermadec and Izu/Bonin/Mariana arcs.

As already rectified in response to a comment from reviewer 1, we added a statement in the introduction specifying that we include also partially emerged calderas in the submarine category.

I don't understand what "their location is outside the DEM" means, or why that is the case. Also, the "technical limitations of the CDA" should be discussed in the main text, not in the supplementary Methods section.

We acknowledge the reviewer's concern, but we would like to specify that the Methods are not considered as part of the supplementary material, they are simply placed outside the main body text at the end, as per Nature journals' format. Indeed, the requested explanations are reported already in the methods, and the reference to the methods was already included here.

• Line 177: Only 9 calderas identified in volcanic arcs is certainly an underestimate!

We agree that the number of calderas detected in volcanic arcs is likely an underestimate. This reflects the limitations of the framework. However, as noted earlier, the aim of this study is not to produce a complete global inventory, but to demonstrate a reproducible framework capable of identifying previously undocumented calderas using currently available low-resolution bathymetry. What we did, however, was to expand the limitations section, and clearly state in the introduction that our aim here was not to reach completeness of the dataset. We also state, however, that future works will address this aspect to improve this framework.

Edited/added text (new lines: 939-962): "***There are*** *several limitations* ***to our approach****. The principal component analysis (PCA) was conducted exclusively on the positive dataset, which comprises 186 of the 514 total features, as opposed to the negative dataset's 328 features* ***(Fig. 2)****. This imbalance reduces the statistical robustness of the analysis. The use of GEBCO global low-resolution bathymetry further constrains detection capabilities, resulting in a high number of undefined features (229 of 514), some of which may actually represent calderas.* ***In addition, as detailed in the Discussion, the CDA shows reduced performance when confronted with highly irregular, composite, or strongly eroded morphologies. The sensitivity test on known systems (Table 4; Supplementary File 4) demonstrates that several well-documented calderas were not detected because their geometries fall below the circularity thresholds required for algorithmic identification (e.g. Kikai, Krakatau, Axial). Other systems were missed because they lay outside the cropped DEM used as input (e.g. Kolumbo, Santorini), or because they were initially detected but subsequently discarded during filtering due to low confidence scores (e.g. Hunga Tonga, Havre). Collectively, these cases illustrate intrinsic constraints of applying a circular-feature-oriented algorithm to morphologically diverse volcanic systems and highlight the conservative nature of our detection approach. A further limitation concerns the non-use of high-resolution multibeam bathymetry (e.g. MBBDB) in the CDA workflow. This choice was necessary to maintain global consistency and reproducibility, given the highly uneven spatial coverage of high-resolution datasets (Supplementary File 5). However, relying solely on GEBCO inevitably limits the level of detail that can be resolved and likely contributes to the conservative nature of our final inventory. We do outline, however, how ongoing improvements in machine-learning architectures, the integration of geological context, and the progressive incorporation of higher-resolution bathymetry will enhance the accuracy and scope of global submarine caldera detection. As multibeam coverage continues to expand and more detailed datasets become publicly accessible, the framework presented here can be***

systematically refined, enabling more complete inventories and more robust hazard assessments."

Edited-added text (new lines: 311-312): "*Our aim is not to produce a complete global inventory of submarine calderas…*"

• Line 196: The Maeno (2025) paper is relevant here.

(new line: 708) Reference added

• Line 199: An example reference about the biogeography of hydrothermal vents.:
Van Dover, C. L., C. R. German, K. G. Speer, L. M. Parson, and R. C. Vrijenhoek (2002), Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates, Science, 295(5558), 1253–1257.

(new line: 710) Reference added

• Lines 202-224: This paragraph is much longer than it needs to be. This is especially true since the ML method didn't even detect Hunga. The Hunga eruption demonstrated a wide range of hazards, we get it. But to me, this paper should be more about clearly demonstrating the effectiveness and limitations of this method, and less about hyping the hazards of calderas.

We thank the reviewer for this comment. In response to Reviewer 1, we had already shortened this paragraph substantially, and we have now streamlined it even further. While we agree that here we demonstrate the capabilities and limitations of the CDA, a brief hazard-context paragraph remains important, as it explains why identifying submarine calderas matters and sets up the discussion that follows.

• Lines 240-263: You should add a figure that shows the detailed bathymetric maps of these 8 calderas.

We modified Fig. 4 (now Fig. 5), also in response to previous comments.

• Line 261: Summit depth should be added to this table.

Summit water depths were already reported in brackets for stage 3 calderas, while for stage 4 calderas we just reported zero. Now we changed the column heading from water depth to summit elevation, updated the stage 4 elevations (from zero to their summit elevation value), and added the negative symbols where elevations were below sea level.

• Line 269: There is no reason to say "near future"!

"Near" removed

• Line 280: Can you demonstrate that having ship-based bathymetry underlying the GEBCO (coarser) database makes a difference in submarine caldera detection vs. areas with only satellite-based inferred bathymetry? If so, that would be an interesting thing to add to the paper.

We appreciate the reviewer's suggestion. We have already addressed the reasons why we did not quantitatively use higher-resolution bathymetry at this stage, and we have added clarifications in the main text along with new supplementary material to support this point. We fully agree that integrating ship-based bathymetry has strong potential to improve submarine caldera detection, and in parallel to the present study we are developing work that specifically explores this potential. However, for the reasons already explained, the present study focuses exclusively on detection based on the GEBCO dataset.

• Line 284: I think in-situ monitoring and real-time data is a pretty high and unrealistic bar to aim for in this context. A more realistic effort to advocate for might be remote hydroacoustic monitoring of calderas of interest, for example from the permanent CTBTO-IMS hydrophone arrays at Wake Island in the Pacific.
In any case, a good reference for an overview of the OOI cabled observatory at Axial Seamount is in: Kelley, D. S., J. R. Delaney, and S. K. Juniper (2014), Establishing a new era of submarine volcanic observatories: Cabling Axial Seamount and the Endeavour Segment of the Juan de Fuca Ridge, Mar. Geol., 352, 426–450, doi:10.1016/j.margeo.2014.03.010.
And a paper on seafloor geodesy observations to complement the Wilcock et al. reference is: Nooner, S. L., and W. W. Chadwick, Jr. (2016), Inflation-predictable behavior and co-eruption deformation at Axial Seamount, Science, 354(6318), 1399–1403, doi:10.1126/science.aah4666.

We added both references as suggested, and related text on the CTBTO-IMS.

Edited/added text (new lines: 648-651): "*Before this step, however, there should be an initial effort to implement hydroacoustic monitoring at calderas of interest through permanent systems, such as the CTBTO IMS hydrophone arrays*[73]*, which provide basin scale and near global monitoring capability*"

• Line 300: I don't think you can say that this technique "demonstrates high efficacy", since all but 2 of the known submarine caldera are NOT detected. Instead, I would say this method has pretty significant limitations! And, in my opinion, the paper does not really adequately explain why (is it mainly due to the limited resolution of the global bathymetric data? Or is it the limitations of the CDA?). Tests with satellite-only data vs. GEBCO data with underlying multibeam data vs. multibeam data at higher resolution would help better demonstrate that.

We modified the sentence, removing "high-efficacy"

Edited/new text (new lines: 665-668): "*This work introduces a global-scale detection framework for submarine calderas.* ***While the approach has acknowledged limitations, it nonetheless enabled the identification of 75 previously undocumented calderas, several of which may be relevant from a hazard perspective.***"

• Line 312: Seems like you should add a concluding comment like this at the end: " Future improvements in the resolution of the GEBCO global bathymetric data compilation will likely build on these results and will identify even more submarine calderas that are below the current detection limit of this method."

Statement added at the end of the discussion section.

Edited/added text (new lines: 695-697): "*Future improvements in the resolution of the GEBCO global bathymetric data compilation will likely build on these results and will identify even more submarine calderas that are below the current detection limit of this method.*"

• Line 320: I think this figure should be in the main body of the paper and should be cited earlier. (see comment on line 136) This would help clarify the questions I had about when the PCA is applied, etc. However, as noted in my previous comment, the difference between the positive and negative results, needs to be explained before the figure is cited.

The figure has been moved to the main body as suggested, and a brief introduction to the negative and positive datasets has been added in the relevant section. However, in order to follow the journal's structure, we refer the reader to the Methods section for the full detailed explanation.

• Lines 328-329: This is an important piece of information that is missing from the main text, but I think should be included. The main text sort of implies that you surveyed the entire GEBCO dataset for calderas, but in fact you only surveyed a subset of GEBCO where seamounts had previously been identified. Is that correct?

We have added to the main text the clarification that the GEBCO dataset used as input was cropped around the 43,000 seamounts identified by Gevorgian et al. However, we still refer the reader to the Methods section for the full details of this preprocessing step.

Edited/added text (new lines: 201-204): "*…we incorporated a series of morphometric and resolution filters to identify potential caldera candidates within the low-resolution GEBCO 2023 global bathymetry dataset[25], ~500 m resolution,* ***cropped around the ~43,000 seamounts from Gevorgian et al.[23], with a 100 km radius (see Methods for rationale and details)***"

• Line 370: There is no caption or description of Supplementary Figure 3, so it is difficult for the reader to understand its significance and relevance to the methods described here. Also the bathymetric maps for each of the 514 candidate calderas lack a color scale for the bathymetry, but the depths are shown in the radial profile cross-sections, so perhaps that is enough.

We thank the reviewer for pointing this out. The purpose of Supplementary File 3 is to provide a graphical overview of all features, both calderas and non-calderas, through their bathymetric maps and elevation profiles. While we intend to keep this file primarily visual, we have now added a brief description in the Methods section to clarify its content and purpose.

Edited/added text (new lines: 835-837): "(*Supplementary File 3 - compilation of all features, calderas and non-calderas, with their respective bathymetric maps and elevation profiles)*"

Also there is no explanation for the naming convention (A-# vs E-# vs SCD-# vs U-#). The reader can find this information in the "Dataset 1" Excel file (A=Atoll, E=Eroded, SCD= full caldera or eroded caldera or crater or small caldera, U=unidentified), but it shouldn't be this hard for the reader to figure this out! It should be explained in the text and in a caption.

We now added a legend on every page of the Supplementary File 3, to help the readers associating the short names with the feature

I note that some text is overwritten and illegible on some adjacent images in Supp Figure 3. Specifically, the bottom x-axis label "Longitude" overwrites the top Lat/Lon location of the maps below due to bad vertically spacing between adjacent maps on the page - for example, SCD-0007/0008, SCD-0010/0011, SCD-0011/0012, etc. Perhaps the best way to correct this is to make the repeating legend at the top of every page smaller, so that the 3 individual maps on each page can be spaced further apart.

We thank the reviewer for highlighting this issue, we initially missed it. This has now been fixed and the new Supplementary file 3 is attached with this submission

Maybe the SCD's (the real caldera candidates) should come first in the file, so the reader does not have to search for them?

All the calderas (SCD) of our dataset are now reported at the top of the file

All that said, I think this Supp. Figure 3 presents critical information that shows how the CDA evaluated the bathymetric data at each caldera candidate site, so it should definitely be kept as part of the manuscript.

Well noted, thank you.

Some other notes: SCD-0045 is actually a known seamount in the Mariana arc called Toto Caldera:
Gamo, T., et al. (2004), Discovery of a new hydrothermal venting site in the southernmost Mariana Arc: Al-rich hydrothermal plumes and white smoker activity associated with biogenic methane, Geochemical Journal, 38, 527–534.
Baker, E. T., R. W. Embley, S. L. Walker, J. A. Resing, J. E. Lupton, K. Nakamura, C. E. J. de Ronde, and G. J. Massoth (2008), Hydrothermal activity and volcano distributions along the Mariana Arc, J. Geophys. Res., 113, B08S09, doi:10.1029/2007JB005423.
Stern, R. J., Y. Tamura, H. Masuda, P. Fryer, F. Martinez, O. Ishizuku, and S. H. Bloomer (2013), How the Mariana Volcanic Arc ends in the south, Island Arc, 22, 133–148, doi:10.1111/iar.12008.
Nakagawa, T., K. Takai, Y. Suzuki, H. Hirayama, U. Konno, U. Tsunogai, and K. Horikoshi (2006), Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana Volcanic Arc, Environmental Microbiology, 8(1), 37–49, doi:10.1111/j.1462-2920.2005.00884.x.

We thank the reviewer for pointing this out. We now included Toto among the previously documented calderas of the dataset and revised the text where necessary (e.g. previously 76 new calderas, and now 75 new calderas) and tables, and added one representative reference among those suggested by the reviewer (Gamo et al., 2004).

Also, candidate U-11 appears pretty bogus, with a clearly bad pixel screwing up the color map, and a supposed caldera diameter of over 200 km! U-171 is similarly obviously bogus with "depths" up to 30 km ABOVE sealevel, and a caldera diameter over 100 km! Why weren't these ID-ed as bogus candidates? Or maybe the better question is: why weren't data gaps ignored (NaN) instead of given real (bogus) depth/elevation values?

We thank the reviewer for the comment. These artefacts arise because the CDA was applied to a cropped version of the GEBCO DEM (restricted to the seamount regions of Gevorgian et al.), rather than the full global grid. At the edges of this cropped DEM, the algorithm can occasionally interpret the artificial boundary as real topography, producing spurious features such as U-11 and U-171. Importantly, these features were manually classified as “Undefined”, and therefore excluded from the caldera dataset and from all subsequent analyses. Consequently, they do not influence any of the results, statistics, or interpretations presented in the manuscript. Since with our framework we already remove such features, we believe no changes to the text are necessary in these regards.

I'm surprised there is not an upper limit constraint on possible caldera sites. This could eliminate many bogus candidate sites such as U-34 and U-35, which are clearly ridge-transform-fracture zone intersections, but with supposed (unrealistic) caldera diameters of 100-150 km.

We thank the reviewer for this observation. We did not impose an upper diameter limit in the CDA because very large calderas do exist on Earth, with documented examples reaching ~150 km in diameter, such as the Apolaki caldera east of the Philippines (Barreto et al., 2020; https://doi.org/10.1016/j.margeo.2019.106052). Applying a strict upper threshold would therefore potentially risk excluding legitimate large calderas. The non-detection of the Apolaki caldera itself is discussed (i.e. the algorithm is not trained to detec highly irregular and eroded features, such as

Kikay, Krakatau and Apolaki). Anyway, the presence of a few large false positives in the initial CDA output does not affect the results, as they are filtered out during manual classification, and not necessarily for the size itself.

• Lines 372-374: This sounds different than I thought I understood from the description at line 329, above, where it sounded like the Gevorgian database was used to restrict what parts of the GEBCO data to look at. Here, it sounds like the Gevorgian filter came AFTER a global search, in addition to a coral reef database. Which is it? This need clarification.

We thank the reviewer for pointing out this ambiguity. The short answer is "both". The Gevorgian seamount database served two distinct purposes in our workflow, and we now clarify this explicitly in the manuscript. First, as described around previous line 329, we used the GEBCO bathymetry cropped around volcanic seamounts (100 km radius) from Gevorgian et al., as the input DEM for the CDA, in order to avoid searching for depressions in regions with no volcanic context. Second, during the manual classification stage, we used the presence or absence of a nearby volcanic seamount (i.e. within the depression, or in proximity of the depression, hence much closer than thee 100 km radius) as contextual information. Depressions lacking any volcanic context were assigned lower confidence as potential calderas than those located near or over a volcanic edifice identified in the Gevorgian dataset. In parallel, we used the coral-reef database, exclusively in this manual classification step, to confirm which depressions corresponded to young atolls, with 100% confidence in this case. We have revised the text in previous lines 372-374 to make aspect more explicit.

Edited/new text (new lines: 840-844): "*We also used the global seamount dataset of Gevorgian et al.*[23] *to determine whether each depression was located on, or immediately adjacent to, a mapped seamount, providing an additional line of evidence for confirming the volcanic nature of the feature. For young atolls, classification confidence was 100% because these features were additionally cross checked against the global coral reef database*[74]*, providing independent confirmation of reef rimmed structures.*"

• Line 375: Also need clarification on how higher-resolution bathymetry was used. What does "inserted in the final dataset" mean? Was it used at the higher resolution (90m)? Or was it downgraded to the GEBCO resolution (500m)? Was the CDA applied to the higher-resolution data? If so, why didn't it identify more of the known submarine calderas?

We thank the reviewer for raising this point. We would like to clarify that this section of the Methods refers exclusively to the feature classification stage, which is the manual component of the workflow. At this point, the CDA had already been applied, and we did not run the CDA on higher-resolution multibeam bathymetry (as now widely justified), only to the cropped GEBCO dataset (except for the sensitivity test with results reported in the new SF5). The higher-resolution MBBDB data (90 m) were used qualitatively during manual classification to better interpret individual features when available. Whether a feature had multibeam coverage or not was recorded as an information in the feature dataset (SF1 and SF2). We have revised the text to make this clearer.

Edited/added text (new lines: 830-835): "*When available, higher resolution multibeam bathymetry (MBBDB*[42]*) (90 m resolution) was consulted qualitatively to refine the interpretation and verify whether apparent rims or depressions in GEBCO were artefacts of low resolution (whether a particular feature had multibeam coverage or not, was inserted as an information in the feature dataset in Supplementary Files 1, and Submarine Caldera Dataset in Supplementary Files 2)*"

Also, such qualitative tools used to support our classification approach are summarised in Table 3, and we now emphasised that in the following short paragraph:

Edited/added text (new lines: 845-848): "*Classifications were assigned only when at least one supporting parameter was met (Table 3, "Support" column), and features lacking sufficient morphological or contextual evidence were labelled as undefined. This systematic approach ensured consistency across the dataset and provided a reproducible basis for the subsequent PCA based validation.*"

• Line 376: In the Dataset 1 Excel file, units need to be added to columns L through S. Likewise, in the Dataset 2 Excel file, units need to be added to columns M through T.

Changes applied to Suppl. Files 1 and 2

• Line 403-415: I think this section on "Limitations" is inadequate. There should be more discussion in the paper (either here or in the main body of the text) of why many known submarine calderas were not detected by this method.

We have expanded the Limitations section and added explicit references to the relevant parts of the main text where these issues are now addressed in detail, supported by new text, tables (Table 4) and new supplementary material (SF4 and SF5)

Edited/new text (new lines: 939-962): "***There are** several limitations **to our approach**. The principal component analysis (PCA) was conducted exclusively on the positive dataset, which comprises 186 of the 514 total features, as opposed to the negative dataset's 328 features **(Fig. 2)**. This imbalance reduces the statistical robustness of the analysis. The use of GEBCO global low-resolution bathymetry further constrains detection capabilities, resulting in a high number of undefined features (229 of 514), some of which may actually represent calderas. **In addition, as detailed in the Discussion, the CDA shows reduced performance when confronted with highly irregular, composite, or strongly eroded morphologies. The sensitivity test on known systems (Table 4; Supplementary File 4) demonstrates that several well-documented calderas were not detected because their geometries fall below the circularity thresholds required for algorithmic identification (e.g. Kikai, Krakatau, Axial). Other systems were missed because they lay outside the cropped DEM used as input (e.g. Kolumbo, Santorini), or because they were initially detected but subsequently discarded during filtering due to low confidence scores (e.g. Hunga Tonga, Havre). Collectively, these cases illustrate intrinsic constraints of applying a circular-feature-oriented algorithm to morphologically diverse volcanic systems and highlight the conservative nature of our detection approach. A further limitation concerns the non-use of high-resolution multibeam bathymetry (e.g. MBBDB) in the CDA workflow. This choice was necessary to maintain global consistency and reproducibility, given the highly uneven spatial coverage of high-resolution datasets (Supplementary File 5). However, relying solely on GEBCO inevitably limits the level of detail that can be resolved and likely contributes to the conservative nature of our final inventory. We do outline, however, how ongoing improvements in machine-learning architectures, the integration of geological context, and the progressive incorporation of higher-resolution bathymetry will enhance the accuracy and scope of global submarine caldera detection. As multibeam coverage continues to expand and more detailed datasets become publicly accessible, the framework presented here can be systematically refined, enabling more complete inventories and more robust hazard assessments**.*"

Reviewer 3

The manuscript entitled "Global machine-learning detection of submarine calderas" presents a novel machine-learning-based approach to identify and characterise submarine calderas. This is a very interesting and valuable application of an existing algorithm, adapted for use in the marine environment. The study makes a meaningful contribution to the field, exposing the needs for improved methodologies to identify different features such as submarine volcanic calderas, particularly under-surveyed regions.

As noted in the manuscript, high-resolution seafloor mapping is time consuming and, in many areas, is logistically challenging. In this context, the proposed method provides an important step towards complementing field-based observations and expanding our capacity to investigate remote marine settings. However, I have some concerns regarding the datasets used, as well as the identification and interpretation of the output features. These issues require substantial clarification. If well revised, the paper might have its place in this journal.

Overall, I believe the paper has the potential to make a valuable contribution to the journal, provided that the authors address the concerns raised. Therefore, I recommend publication following major revision.

Essential Reviews:

- Dataset used for the analysis is GEBCO2023 bathymetry, which has a very low spatial resolution (~500m / 15 arc-second). Although the authors acknowledge this limitation, and refer to the Seabed2030 initiative as a future improvement, they do not discuss or incorporate other higher-resolution global bathymetric datasets that are currently available, such as DEM Global Mosaic and the Multibeam Bathy Mosaic, which could improve spatial resolution and enhance the feature detection. It is unclear why these higher-resolution datasets are not considered (e.g., these datasets are even mentioned in one of the references used in this study). Combining multiple global datasets, or using different datasets for different regions, could significantly improve the quality of the input data, and consequently, the reliability of the analysis.

Individual comments on these aspects mentioned here are addressed individually below

-The manuscript lacks a sensitivity analysis of the CDA methodology. Including this is essential to demonstrate the robustness and reliability of the approach described here, and will help quantify how uncertainty of the inputs (e.g., low-resolution bathymetry) propagate through the analysis and influence the outputs of the analyses (i.e., submarine caldera identification and characterisation).

The CDA represents only the first step of the framework, where depressions, purely geometrical features, without any assessment of their geological nature, are automatically detected. This component of the workflow has already been tested on Martian impact craters and shown to perform with relatively small uncertainty (over 75% accuracy – Lee and Hogan, 2021). If the reviewer is referring to a sensitivity analysis of the manual classification, this is already provided through the Principal Component Analysis (PCA). Although the PCA was conducted on a subset of candidates (186 out of 514 features), a limitation we acknowledge already, the results show strong agreement with our manual classification. Most importantly, calderas and craters form a distinct cluster, clearly separated from other feature types such as atolls and eroded surfaces.

- The authors make a good point on the limitations of the existing database of submarine calderas. However, the dataset of documented calderas used in this study appears to be underrepresented (i.e., Fig. 1, Global distribution of documented calderas). For example, in the Tonga-Kermadec Arc region alone, a significantly large number of submarine calderas has been reported in the literature, yet this regions seems underrepresented. This raises concerns regarding the completeness of the reference dataset used for training and validation. It is not clearly explained how submarine calderas were selected from the CCDB; this information cannot be found in the cited reference provided, and the authors only states that features were ‘visually assessed’, without explaining the assessment criteria or validation procedure.

We thank the reviewer for this comment and welcome the opportunity to clarify. Importantly, we did not use the CCDB for model training, validation, or for generating any of the data in this study. The CCDB (and GVP) locations were used solely to illustrate the broad underrepresentation of submarine calderas in existing global databases, a point we already emphasize in the Introduction. Our intention in Fig. 1 was not to present a complete or curated reference dataset, but to highlight the mismatch between the abundance of submarine and subaerial calderas from existing and well established global databases.

We fully acknowledge that additional submarine or partially submarine calderas have been described in individual studies (including in the Tonga–Kermadec region), but these are not systematically incorporated in the GVP or CCDB, or at least not classified as calderas. This inconsistency is precisely the motivation for our work: the global underreporting of submarine calderas. To make this clearer, we have added a paragraph in the introduction, also as a response to a previous comment.

Edited/added text (new lines: 128-137): “*Existing global caldera databases highlight a* ***strong*** *imbalance between subaerial and submarine records. For example, the Global Volcanism Program (GVP)* ***lists*** *140 Pleistocene-Holocene calderas, of which 94% (n= 132) are* ***above sea level*** *and only 6% (n= 8***) are fully submarine[21]. Several well known systems such as Santorini, Krakatau, and Kikai are predominantly subaqueous, yet their reported GVP elevations place them in the subaerial category; even if reclassified as submarine, the proportion would remain below ~8%. Other confirmed submarine calderas, such as Axial, Havre, and Monowai, are assigned to different volcanic types in the GVP (fissure feature, stratovolcano, and composite volcano, respectively), therefore are not included in the caldera count. Even when accounting for these misclassifications, submarine calderas would still represent <10% of the global inventory***”

Although the authors emphasise the importance of developing an updated and comprehensive submarine caldera database, the current description of the data sources and selection methodology does not allow for proper validation or reproducibility. Further description of classification criteria is needed.

For clarity, we would like to point out that we did not mention or aim at producing an updated and comprehensive list of submarine calderas. We did mention, however, that we wanted to provide an upgradable framework (“*Our dataset fills a major observational gap and offers a reproducible and upgradeable framework for submarine volcano characterization and geohazard estimation, underscoring the need to integrate submarine calderas into future global risk assessments.”).*

To clarify this even further, we added a sentence in the introduction

Edited/added text (new lines: 311-314): “*Our aim is not to produce a complete global inventory of submarine calderas, but to establish a transparent and reproducible framework that yields a conservative baseline dataset, which can be expanded as higher resolution bathymetry and future studies become available*”.

Regarding the reproducibility of our work, the framework we proposed is quite straightforward. The only component that may introduce uncertainty is the manual (expert-based) classification. To overcome this, we added more clarity of the methods in the respective methods section, as specifically reported earlier in this file.

- It is not sufficiently clear why, out of the 78 detected volcanic depressions, only 8 submarine calderas were selected as high-priority targets. The authors refer to 'factors that collectively suggest an elevated likelihood of future eruptive activity and potential societal impact', yet the parameters described (i.e., water depth, tectonic settings, and caldera morphological type) appear insufficient to support determining the future activity and potential societal impact. Additional parameters, including magma composition, documented eruptive history, and proximity to coastal communities or critical infrastructure are needed. Without incorporating this additional information, the prioritisation of selected calderas appears oversimplified, and almost speculative.

We thank the reviewer for this comment. We would like to clarify that our designation of "high-priority calderas" does not imply that these systems are expected to erupt soon. Rather, it reflects their relative importance as targets for future study, especially given that the majority of them are newly identified and occur in settings where hazardous processes are more likely to be consequential. Parameters such as magma composition and eruptive history are indeed critical for assessing future activity and hazards, however, these data are not available for the newly detected calderas, which have no prior documentation. In some cases, magma composition can be inferred from tectonic setting, but this is not always reliable without direct sampling. Similarly, proximity to coastal populations is only one dimension of societal impact: submarine eruptions can generate tsunamis that affect regions thousands of kilometres away, and they can also threaten seafloor infrastructure such as communication cables and pipelines, even when far from shore. Our prioritisation therefore relies on the parameters that can be robustly assessed from global datasets, such as water depth, tectonic setting, and caldera morphology, which together provide a conservative, first-order indication of systems where future detailed investigation is most urgently needed.

- A major concern is that the CDA fails to identify several well-documented calderas, such as Hunga Tonga, Kikai, Krakatau, Santorini, etc…. The authors attribute this limitation to the fact that these calderas fall outside the DEM used. However, this explanation highlights a fundamental limitation in the study design rather than an unavoidable constrain. This issue could be addressed by adopting a more comprehensive or higher-resolution DEM used, or run the analyses in a region base, instead of global analyses. The omission of such well-characterised and widely studied calderas represents a major gap in the current implementation of the CDA, and raises concerns about the completeness and reliability of the detection of submarine calderas.

We thank the reviewer for this comment. The concern regarding the documented calderas not initially detected, has been now widely addressed in this revision, in response to several comments. We now included additional text, tables and supplementary material, that address this specific limitation. However, we would like to point out, once again, that the goal of this work was not to reach "completeness" of the dataset, or perfection in the detection framework. This has also been addressed widely throughout this file and clearly reported in the manuscript.

Additionally, these well documented calderas surveyed in high-resolution, provide an excellent opportunity for training and validating the algorithm. Incorporating these examples could substantially improve the model calibration, enhancing the detection performance and increasing the confidence of the outcome when identifying unknown submarine calderas worldwide.

We thank the reviewer for this suggestion. In principle, well-documented calderas mapped at high resolution would be valuable for validating the CDA. However, using these features to train the CDA is not appropriate for this study for several reasons.

1. The number of known submarine calderas is extremely limited compared to the typical datasets required for effective machine-learning training (on the order of hundreds, versus hundreds of thousands or millions used in planetary crater detection). Training on such a small sample would risk biasing the results towards the few well-studied examples.

2. The CDA must be trained on data with the same resolution as the dataset it is applied to. High-resolution multibeam grids (metre-scale) cannot be used to train a model intended to operate on GEBCO-scale bathymetry (~500 m resolution), as the learned features would not be transferable. Training a model on high-resolution calderas and then applying it to low-resolution global data would therefore not improve detection performance.
3. Finally, the CDA is designed as a geometric detector that identifies circular depressions, and all features emerging from it are, by definition, “undefined” until the subsequent filtering and expert-based classification steps. Training could reduce some false positives, but it would not eliminate the need for manual or rule-based post-processing, nor would it resolve the fundamental trade-off between false positives and false negatives.

For these reasons, we opted to use the CDA as a geometric detection tool, following the same training principles applied to the depression-detection model used for Mars, and applied the necessary morphological constraints outside the machine-learning step, to add an extra element of differentiation between impact craters (simple circular depressions) from volcanic calderas (depressions, more or less regular, on a mound). Besides noting in the manuscript that future work may explore fine-tuning or additional model training, we also added new text in the methods to explain why we did not train the model for caldera detection:

Edited/new text (new lines: 708-714): “*In our case,* ***we did not conduct further model training because the number of calderas worldwide represents only a small fraction of the number of impact craters on Mars (10s versus 100,000s), and using such a relatively small dataset for training could bias the results toward the few known examples. However, we did modify the original algorithm to reflect key differences between impact craters (depressions in the ground) and calderas (depressions over a mound), specifically through the introduction of a series of water depth parameters (Table 2). We refer to this modified algorithm as the*** *caldera detection algorithm (CDA).*”

- The growth stage classification used in this study to determine the volcanic hazard potential, should be reconsidered or complemented with more comprehensive framework presented by Acocella (2007). This classification represents one of the most extensive and systematic approaches to caldera development and evolution, incorporating not only water depth and volcano height, but also the magma composition, eruptive history, and also the magma reservoir characteristics (depth and volume). In its current form, the growth stage used in this study seems simplified and may not properly capture the complexity of caldera evolution and associated hazard potential. Integrating, or at least discussing the Acocella (2007) classification will provide a more robust basis of hazard interpretation and improve the scientific grounding of the prioritisation scheme.

We thank the reviewer for this suggestion. We fully agree that the Acocella (2007) framework provides a comprehensive and systematic approach to caldera evolution and could be useful for our hazard-based caldera prioritization. However, the parameters discussed by Acocella (2007) are not available for the newly detected submarine calderas in our dataset. Because these edifices (or at least the majority of them) have not been previously documented, information on magma composition, reservoir depth/volume, or eruptive history does not exist, and cannot be inferred reliably without data. For this reason, we rely on the growth-stage classification, which in turn relies on edifice size and water depth, providing a conservative, first-order proxy for the likelihood of e.g. hazardous magma–water interaction and the potential for tsunami generation. While simplified, this parameter is consistently measurable from global bathymetry and therefore applicable to all 78 detected calderas. Our prioritisation scheme is not intended to predict future eruptive activity, but rather to identify which newly discovered calderas warrant further investigation based on the limited information currently available.

We have added text to clarify this rationale and to acknowledge the value of more comprehensive frameworks such as from Acocella (2007), which could be applied once additional geological and geophysical data become available. However, in absence of information we have to rely on other parameters globally available.

Edited/added text (new lines: 507-513): "*In the absence of geological and volcanological information that will allow a more quantitative prioritization*[62]*, we rely on parameters that are consistently available at the global scale. One of these is the growth stage classification*[63]*, which reflects volcano size and water depth and provides a first-order, relative indication of hazard potential (e.g. shallow and large volcanoes, Stage 3, are potentially more hazardous than small and deep volcanoes, stage 1, both in terms of explosive magma-water interactions and tsunami generation potential)*[24]*. Therefore…*"

- I'm not fully convinced with the morphological classification and features description presented in this manuscript. The classification is described as being based on 'a combination of GIS-based analyses' and 'visual assessments'; however, neither of these are described in sufficient detail in the methodology section. This lack of clarity and details limits the reproducibility and raises concerns regarding subjectivity in the interpretation of morphological types and sub-types.

We thank the reviewer for the comment. We now expanded the description in the methods, giving more clear guidelines indicating more precisely how the manual classification was carried out:

Edited/added text (new lines: 827-848): "*Here we report guidelines on how the classification was conducted, to ensure consistency across users. Manual classification of the 514 filtered features was carried out using multiple independent lines of evidence (Table 3). Each feature was first visually inspected in GEBCO to assess its overall morphology (e.g. rim continuity, sub circularity, presence of a surrounding edifice). When available, higher resolution multibeam bathymetry (MBBDB*[42]*) (90 m resolution) was consulted qualitatively to refine the interpretation and verify whether apparent rims or depressions in GEBCO were artefacts of low resolution (whether a particular feature had multibeam coverage or not, was inserted as an information in the feature dataset in Supplementary Files 1, and Submarine Caldera Dataset in Supplementary File 2). We also examined the topographic profiles extracted for each feature (Supplementary File 3 - compilation of all features, calderas and non-calderas, with their respective bathymetric maps and elevation profiles), which provided quantitative information on rim elevation, depression depth, and inner slope geometry. These profiles were particularly useful for confirming the presence of steep inner walls characteristic of calderas, as opposed to the gentle, reef modified inner slopes typical of sunken atolls and guyots. We also used the global seamount dataset of Gevorgian et al.*[23] *to determine whether each depression was located on, or immediately adjacent to, a mapped seamount, providing an additional line of evidence for confirming the volcanic nature of the feature. For young atolls, classification confidence was 100% because these features were additionally cross checked against the global coral reef database*[74]*, providing independent confirmation of reef rimmed structures. Classifications were assigned only when at least one supporting parameter was met (Table 3, "Support" column), and features lacking sufficient morphological or contextual evidence were labelled as undefined. This systematic approach ensured consistency across the dataset and provided a reproducible basis for the subsequent PCA-based validation.*"

The use of more explicit, semi -automated GIS-based tools will help reduce the subjectivity and strengthen the classification of the identified features, as these tools incorporate description of features, the definition of which are referenced and accepted by the scientific community. In this regard, established tools such as CoMMa (Arosio et al., 2024) and GA-SaMMT (Dove et al., 2020; Nanson et al., 2023), should be incorporated into the workflow, as they provide systematic geomorphometric analyses and would partially replace or complement manual visual interpretations.

This would substantially improve methodological transparency, repeatability, and confidence in the proposed feature description.

We appreciate the reviewer’s suggestion and believe that semi-automated geomorphometric tools such as CoMMa and GA-SaMMT are valuable for high-resolution, regional-scale analyses. However, these tools are not directly applicable to the objectives and data constraints of our global-scale study. Our workflow is necessarily designed around the ~500 m resolution GEBCO 2023 dataset, for the reasons now widely explained in the manuscript. Tools like CoMMa and GA-SaMMT rely on fine-scale morphometric descriptors (e.g. curvature, slope breaks, rim sharpness, terrace morphology) that cannot be reliably extracted at GEBCO resolution. For example, CoMMa was used on DEMs where features can be resolved, far below the resolution of our global dataset. As noted by Arosio et al. (2024), even under optimal conditions its performance “*does not reach the accuracy of expert interpretation*,” reinforcing that expert-guided classification remains necessary. On the other hand, GA-SaMMT seems conceptually closer to our approach but is designed to classify predefined bathymetric highs and lows using morphometric parameters that cannot be robustly computed at the GEBCO scale. Moreover, GA-SaMMT has been applied only to limited regional domains with high-resolution grids, not to low-resolution global datasets spanning >43,000 seamounts.

In summary, while these tools are valuable for studies using high-resolution bathymetry and targeting standardised geomorphic classes, they are not suitable for the scale, geological context, or feature types addressed in our work.

Overall, limitations acknowledged in the manuscript suggest that the methodology, in its current implementation, is providing a minimum number of potential submarine calderas. This raises an important question regarding the effectiveness and completeness of the proposed detection framework.

While this approach is promising, significant methodological improvements are necessary to enhance the reliability of the presented CDA analysis, including: 1) include higher resolution multibeam datasets as baseline. Expanding the DEM coverage or alternatively conducting regional-scale analyses rather than a purely global assessment, could substantially improve detection accuracy and reduce false negatives; 2) use of well-documented submarine calderas mapped at high resolution, should be systematically incorporated into the training and validation dataset; 3) use of semi-automated GIS-based tools for features classification, should complement or partially replace visual assessments in order to reduce subjectivity and improve reproducibility in feature classification. Moreover, additional parameters should be also incorporated to classify the growth stage of identified calderas, such as the ones included in the framework presented by Acocella (2007).

All this has been addressed above in previous comments.

In summary, while the study presents a promising and innovative approach to identify unknown submarine calderas, substantial methodological refinement is required before the proposed CDA framework can be considered robust for global application. Addressing these points would significantly strengthen the scientific contribution and overall impact of the manuscript.

We thank the reviewer for the constructive summary. We fully agree that continued methodological refinement will further strengthen the CDA framework, and we are actively working in this direction as part of the current revision and future developments. The primary goal of this study, however, is to establish the first global-scale baseline dataset of submarine calderas using a workflow that is compatible with the resolution and coverage of existing global bathymetry. This initial framework is intentionally designed to be upgradeable, so that higher-resolution DEMs, more advanced geomorphometric tools, and fully automated approaches can be incorporated as they become available or when applied to regional studies. By first providing a reproducible global detection tool and dataset, we create the foundation upon which these more sophisticated methods can be systematically tested and integrated.

Minor comments included in the attached pdf.

Line 58: Dataset nowhere to be found. From the links in the reference cited here, the CCDB is no longer available in the link provided in the reference. I was redirected to a site called 'VICTOR' from Columbia University - still, no way to find the CCDB. Would the authors be able to provide a link to this dataset?

The CCDB official webpage must be under renovation, however, there are several other sources online for the CCDB, which, for reviewing purpose only, we provide the link here: https://zenodo.org/records/10636011. However, we emphasize here as well that we did not use the GVP or CCDB data to produce new data for this work, they were only used to show the underrepresentation of submarine calderas in global datasets.

Lines 87-88: Why only use this low-resolution global multibeam bathymetric dataset? This can be combined with other global datasets that have higher-resolution, such as the DEM Global Mosaic, and/or Multibeam Bathy Mosaic (these are even listed as used in the reference cited here, Ref. 20)

This has been widely addressed now. We agree that higher-resolution bathymetric datasets can provide valuable detail, however, they do not offer global coverage and are spatially heterogeneous, which prevents their integration into a uniform global machine-learning workflow. Because our goal was to perform detection consistently across all ocean basins, we relied on GEBCO, the only dataset with full global coverage. Following the reviewer's suggestion, we clarified this rationale in the Methods and explicitly stated that multibeam datasets were used where available for post-detection validation and visualization.

Edited/added text (new lines: 722-782): "*… we used GEBCO because it is the only bathymetric dataset that provides full global coverage on a continuous ~500 m/pixel grid. Other globally complete bathymetric products do exist, but at even lower resolution (e.g. ETOPO1, ~1,800 m at the equator). Having a homogeneous grid is essential for running an automated machine-learning detection algorithm consistently across all ocean basins. Higher-resolution datasets such as the MBBDB[42] do not offer global coverage, are spatially heterogeneous with substantial data gaps, and contain numerous artefacts arising from variable data quality and post-processing procedures. These issues currently prevent their integration into a uniform global-scale detection workflow. While we acknowledge the potential utility of multibeam compilations, and anticipate that future targeted studies building on approaches similar to ours are underway and will help address these limitations, integrating them into our current framework would introduce additional uncertainty. To evaluate this briefly, we applied our CDA to the MBBDB, and the results are presented in Supplementary File 5. This test shows that, although some detected features appear real, the majority of the output features are false positives, particularly in areas with incomplete coverage or processing artefacts. For these reasons, we did not incorporate multibeam-based detections into our global detection/classification framework at this stage. Importantly, where multibeam data were available, we used them in a qualitative way for the post-detection manual classification*"

The reviewer suggested combining low- and high-resolution DEMs. While this is conceptually appealing, it is not compatible with the CDA workflow, at least at this stage. As described in the Methods ("*The CDA splits the input GEBCO bathymetry dataset into small orthographically projected regions that overlap in location and step logarithmically in size from about 1 degree to about 30 degrees*"), and as widely described in the original Crater Detection Algorithm of Lee and Hogan

(2021), the algorithm operates by dividing the input DEM into multiple sub-images of different sizes and resolutions. Using a merged DEM with heterogeneous resolutions would introduce inconsistencies across sub-images, making it difficult to track which depressions were detected at what resolution and potentially biasing the detection process. In addition, such a merged DEM would be computationally heavy to process at global scale. Using a DEM with homogeneous resolution avoids these issues and ensures methodological consistency.

Nevertheless, to address the reviewer's concern, we also performed a sensitivity test using the multibeam (MBBDB) dataset alone to evaluate CDA performance at higher resolution. This analysis is now included in the Supplementary Material (SF5) and clearly shows what expected, with the majority of the features being false positives because the CDA wasn't trained to deal with data gaps. This is also addressed in the text now, as previously reported in response to other comments.

Fig.1: Calderas are widely underrepresented in this map. It makes a good point on flagging the lack of calderas in the global datasets, however, looks like in this map there are still calderas missing from the datasets cited here (especially in the area Kermadec-Tonga Arc)

We thank the reviewer for this observation. The issue highlighted mainly concerns the GVP dataset. The GVP does not classify several known submarine calderas as calderas (e.g. Havre), and partially emerged systems (e.g. Santorini, Krakatau) are reported as subaerial and therefore were not initially counted. We have now clarified this issue in the Introduction and added a qualitative estimate of these missing or misclassified calderas in the revised text. Importantly, even after accounting for these omissions, the overall conclusion remains unchanged: submarine calderas are still markedly underrepresented compared to subaerial ones in existing global inventories. The purpose of Fig. 1 is precisely to illustrate this imbalance, rather than to provide a complete global catalogue.

Edited/Added text (new lines: 129-137): "*For example, the Global Volcanism Program (GVP) lists 140 Pleistocene-Holocene calderas, of which 94% (n= 132) are above sea level and only 6% (n= 8) are fully submarine[21]. Several well known systems such as Santorini, Krakatau, and Kikai are predominantly subaqueous, yet their reported GVP elevations place them in the subaerial category; even if reclassified as submarine, the proportion would remain below ~8%. Other confirmed submarine calderas, such as Axial, Havre, and Monowai, are assigned to different volcanic types in the GVP (fissure feature, stratovolcano, and composite volcano, respectively), therefore are not included in the caldera count. Even when accounting for these misclassifications, submarine calderas would still represent <10% of the global inventory*"

Lines 115-118: Not clear what the selection is based on. How are the 'potential societal impacts' defined? is this different for different regions? is distance to coastal communities a factor being considered here? Further explanation, and references, is needed here.

This part on which the reviewer refers to was revised in response to a different reviewer's request. Nevertheless, as clarified in the revised text, the identification of the eight highlighted calderas was qualitative rather than quantitative. Our aim was not to compute absolute values (e.g. societal impacts or region-specific risk metrics), but to identify calderas that, based on relative measures, such as their depth, tectonic setting, and morphological type (e.g. shallow water depth = relatively more hazardous, because of more efficient magma-water explosive interaction; convergent setting =higher hazard than intraplate or ridge setting, because magma composition is likely more viscous), are more likely to exhibit hazardous eruptive behaviour, but relative to the other calderas identified.

Edited/new text (new lines: 301-306): "*In addition to expanding the global inventory of submarine calderas, our approach also enables the identification of calderas that may warrant closer scientific attention due to their hazard potential, as inferred from their water depths, tectonic settings, and* *morphological characteristics. Such features are important for first-order volcanic hazard assessments prior to direct exploration, underscoring the broader value of systematic detection frameworks for guiding subsequent investigation and hazard evaluation.*"

This aspect is now also explained in the discussion:

Edited/new text (new lines: 507-513): "*In the absence of geological and volcanological information that will allow a more quantitative prioritization[62], we rely on parameters that are consistently available at the global scale. One of these is the growth stage classification[63], which reflects volcano size and water depth and provides a first-order, relative indication of hazard potential (e.g. shallow and large volcanoes, Stage 3, are potentially more hazardous than small and deep volcanoes, stage 1, both in terms of explosive magma-water interactions and tsunami generation potential)[24]*"

Line 124: which highlighted the need for surveying areas such as the Kermadec-Tonga Arc in higher resolution, to have baseline information before big events, such as the 2022 Hunga event. This event showed the importance of having pre- and post-eruption high-resolution seafloor mapping, which prove to be crucial information to better understand volcanic processes and thus the hazard potential of eruptions on these submarine calderas.

We fully agree with the reviewer that high-resolution pre- and post-eruption seafloor mapping is essential for understanding submarine volcanic processes and assessing hazard potential. The 2022 Hunga Tonga eruption is a fantastic example of that. Our contribution is complementary to this need. Well-studied volcanoes that have erupted in recent decades have already benefited from targeted expeditions, high-resolution mapping, and detailed investigations of their magmatic systems and eruptive behaviour. In contrast, many submarine calderas remain entirely undocumented simply because they have not erupted in historical time and therefore have never been surveyed. The aim of our study is to identify and prioritize these previously unexplored systems, providing a global-scale baseline that highlights where future high-resolution mapping efforts would be most impactful. Once these priority calderas are recognized, targeted surveys can then acquire the detailed bathymetry, samples, and geophysical data needed to characterise their magmatic systems and eruptive histories before the next eruption occurs.

Lines 132-134: There is a need for a standarised definition of all the features that are here 'manually categorised'. At the moment it is not clear how the morphologies are distinguished, and what are the criteria followed to define them and classify them. The use of other GIS-based semi-automatic tools to determine the morphological characteristics of the output features would also reduce the need for manual categorisation and classification. GIS-based tools such as CoMMa and GA-SaMMT, provide these standarised criteria and should be considered as a step on the presented methodology

We thank the reviewer for these suggestions. However, the manual classification component is a fundamental part of our framework and cannot be replaced by fully automated tools. We believe that expert-based assessment is essential in any machine-learning or semi-automated workflow, particularly for volcanic and erosion-related morphologies. For example, distinguishing an eroded flank surface from a horseshoe caldera requires geological context and process understanding that cannot be reliably captured by automated rules, even after intense model training. Our approach intentionally combines an automated component (CDA + filtering) to reduce the number of candidates, with a subsequent expert-guided classification step, which we further validate through Principal Component Analysis. This hybrid structure is central to the robustness of the method.

Regarding the tools suggested by the reviewer, we responded already in a previous comment.

Line 138: the 'undefined' features could be reduced if there is some sort of algorithm training. If well documented calderas, mapped in high-resolution, are used to train the algorithm, this uncertainty will be reduced.

The model training aspect has been largely addressed in response to a previous comment

Line 145: how is this defined? is this morphologic degradation? eruptive degradation? This needs clarification supported with references on what is this referring to and how it is defined

In this context, the term *degradation* is used in a broad morphological sense to describe observable modifications to a feature's shape. This generic usage is intentional: nearly all calderas in our dataset are newly identified and lack any documented geological history, so we cannot determine whether their modification reflects erosional, eruptive, or other processes. For this reason, we use *degradation* as a neutral descriptor of morphological alteration (consistent with the Oxford definition), without implying a specific mechanism.

Lines 173-177: See comment above, if well documented calderas, mapped in high-resolution, are used to train the algorithm, this uncertainty will be reduced. Additionally, the use of other higher-resolution global multibeam datasets are used, or the selection of regional DEM are defined, this will also used the uncertainties of the CDA methodology

The model training aspect has been largely addressed in response to a previous comment

Lines 181-182: This classification should be completed with the one defined by Acocella 2007

We have already largely responded to this concern by the reviewer in response to previous comments.

Lines 202-224: This section seems disconnected to the overall discussion. Link to the importance of this study is absent and needs to clarified. This is especially because the two examples presented here are not even included in the CDA analysis (which in turn, could have been included improving the outputs of the analysis - see comments above)

As previously requested, we have substantially shortened this paragraph, and we now ensure that it is directly connected to the purpose of our study. The revised paragraph now clarifies that shallow and partially emerged systems (stages 3 and 4) are the relatively most prone to high-impact explosive behaviour, and therefore constitute the key targets that our study aims to prioritise. We also now explicitly state in the new subsection in the discussions which well-known calderas were not detected by the CDA and why. The examples of Hunga Tonga and Krakatau are retained here to contextualise the importance of our work, rather than to comment on the CDA performance.

Edited/added text (new lines: 481-490): "*In the context of risk, submerged calderas can generate high impact hazards, as demonstrated by the 2022 Hunga Tonga eruption, which produced powerful explosions, submarine density currents that severed communication cables, and basin scale tsunamis*[53,54]*. Such events show that shallow (stage 3) and partially emerged systems (stage 4)*[24]*, are especially prone to intensified explosivity due to magma–water interaction*[35–38]*. Historical examples such as the 1883 Krakatau eruption (Krakatau is now stage 4) similarly illustrate how partially submerged activity can trigger catastrophic tsunamis and widespread atmospheric effects*[55,56]*. These impacts highlight the potential vulnerability of submarine infrastructure and the need to better assess hazards associated with shallow and partially emerged submarine calderas*[57,58]"

Lines 206-207: this highlights the need to include into this classification the parameters described in Acocella 2007

We have already addressed this concern in response to previous comments

Lines 209-211: references of different eruptive events where these hazards where triggered are needed here

That sentence was deleted in the process of shortening that paragraph

Line 2012: can you clarify what stage would be Krakatau?

It's been added in brackets: "*Historical examples such as the 1883 Krakatau eruption (Krakatau is now stage 4)…*"

Line 214: There are more recent studies on this volcano, such as Madden-Nadeau et al., 2021

We replaced Self and Rampino (1981) with Madden-Nadeau et al. (2021), and retained Rampino and Self (1982), because it focuses on the climatic effects of Krakatau 1883.

Lines 223-224: I don't see how this is relevant for this study and the discussion here, how this fits into the work presented here.

This sentence has been deleted

Line 234: are these fractures or ring faults? Please reword

Here we mentioned structural weaknesses, which could be either ring faults or fractures. We now specify that ("*These morphologies can direct future eruptions along structural weaknesses (e.g. ring fault or fractures), with magma migrating preferentially along them*")

Line 235: which one? Not clear. Please rewrite this section as it is not very clear

We now specify that the morphological category is horseshoe calderas ("Alternatively, **horseshoe calderas** may reflect")

Lines 248-249: not clear. Please rewrite to clarify

This statement was deleted from the discussion

Lines 261-262: reasoning/justification of why this is assumed here is needed. Are there any examples that can be referenced to support this assumption?

We have clarified the reasoning behind this assumption. What we mean is that small calderas or craters (those that we initially classified <3 km in diameter) generally reflect past lower eruptive volume, which is consistent with smaller collapse volumes and small collapse features (if it's a horseshoe morphology), also reflecting past reduced tsunami generation potential (this has been referenced now). This is an approximation and we acknowledge that small edifices can sometimes produce large caldera-forming eruptions. However, because most of the newly identified calderas lack geological information, morphology currently provides the most reliable first-order proxy for future hazard potential. We have revised the text to make this justification more explicit.

Edited/added text (new lines: 573-580): "*Craters and small calderas (classified here as volcanic depressions <3 km in diameter) are not prioritized at this stage due to their limited size, which is assumed to correlate with* ***past*** *lower-****intensity*** *explosive* ***activity*** *and reduced tsunami hazard from smaller collapse volumes*[64,65]*,* ***which we interpret at this stage as lower future hazard potential***[24]***. We do acknowledge that small calderas or craters can still produce large caldera-forming eruptions, however, in absence of geological information, we must rely on current morphology, which is the closest approximation of the behaviour of that volcano.***"

Line 275: integration (Including other high-resolution multibeam datasets (instead of limiting it to using only GEBCO) will improve the analysis, as well as not limiting the DEM used, or run the analyses in a regional base, instead of globally.)

This concern was already addressed in response to previous comments. However, for the purpose of this section, which is on future directions, we have revised the sentence to explicitly reflect this point, as suggested by the reviewer. The updated text now emphasises not only the importance of improving access to high-resolution bathymetric mapping technologies, but also the value of integrating these datasets into global bathymetric products, which would enhance the detection and characterization of submarine calderas.

Edited/added text (new lines: 631-639): "*Additionally, improving accessibility to high resolution bathymetric mapping technologies, and integrating existing and new multibeam datasets into global bathymetric products, would substantially accelerate the discovery and characterization of submarine calderas.*"

Lines 300-301: I am not really convinced about this, as it has not been fully demonstrated, and there are serious limitations on the outputs of the analysis - which in turn are acknowledged here. Please change the wording

The sentence has been modified to reflect the reviewer's suggestion ("*This work introduces a global-scale detection framework for submarine calderas.* ***While the approach has acknowledged limitations, it nonetheless enabled the identification of 75 previously undocumented calderas, several of which may be relevant from a hazard perspective.***")

Line 304: semi-automatic GIS-based tools need to be integrated in the analyses to improve its outcomes

This point has already been addressed in response to a previous comment

Line 306: what is the sensitivity of the method and the accuracy of this output?

The sensitivity and accuracy of the automated component of the framework (the CDA) can be evaluated through the number of duplicate detections it produces (definition in Table 2 > now Table 3; and outputs in the SF1 and SF2 files), which reflects how consistently the algorithm identifies depressions with a given confidence level. However, the goal of this study was not only to detect depressions but to determine whether these correspond to calderas, which requires expert interpretation and validation through PCA. Because this interpretive step depends on geological context and expert judgement, its accuracy cannot be quantified in a strict statistical sense. Nevertheless, we provide all elements necessary to reproduce the workflow, ensuring transparency and repeatability. This is analogous to geological mapping: in the absence of direct outcrop constraints (particularly on other planets or underwater), the accuracy of interpretations cannot be expressed numerically, but the methodology can be fully documented and reproducible. We mentioned in the manuscript that our framework is upgradable, so is the Submarine Caldera Dataset (SF2). As new information and data become available, our SCD can be upgraded too.

Line 310: this is a bit of a strecht here. This could be considered in the future, if/when there is a better output of the analysis and limitations are reduced... could be if the analyses would be run in a regional basis, but not at the stage presented here, in a global basis.

The sentence has been modified (“*this expansion fills a relevant observational gap and generates actionable data to guide hazard prioritization, targeted field campaigns, and eventually future monitoring strategies*”)

Lines 317-318: see all the comments about the existing GIS-based semi-automatic tools that should be considered to be included as a step for this analysis, such as CoMMA and GA-SaMMT, providing further standardised feature identification and classification (with definitions that are referenced and already accepted by the scientific community)

This point has already been addressed in response to a previous comment

Line 325: I do not agree with this statement. An algorithm can always be trained further by including more examples, espcially if there area available high-resolution surveyed calderas that could be useful to improve the outputs of the analyses.

The sentence the reviewer refers to was actually referring to the original Crater Detection Algorithm from Lee and Hogan (2021). We now clarify in the following paragraph that we did not conduct training of the model because the data sample is too small for proper training and unbiased results.

Edited/added text (new lines. 708-714): “*In our case,* ***we did not conduct further model training because the number of calderas worldwide represents only a small fraction of the number of impact craters on Mars (10s versus 100,000s), and using such a relatively small dataset for training could bias the results toward the few known examples. However, we did modify the original algorithm to reflect key differences between impact craters (depressions in the ground) and calderas (depressions over a mound), specifically through the introduction of a series of water depth parameters (Table 2). We refer to this modified algorithm as the*** *caldera detection algorithm (CDA)*”

Lines 358-359: see comments on training the algorithm with real examples

The training aspect of our Caldera Detection Algorithm has been addressed now separately in the manuscript, as previously addressed

Lines 371-372: see comments on using GIS-based semi-automatic tools for feature identification and classification, using standarised definitions

This point has already been addressed in response to a previous comment

Line 404: not clear where this number is coming from, from the paper there was: 56 calderas, 22 crater or small calderas, 62 atolls, and 145 flank/erosion. The numbers do not add up, in any (logical) combination...Please clarify

These numbers are reported in Fig. 5 (now Fig. 2), where we clearly show how many features were detected and filtered from both the positive and negative CDAs. To add clarity on where these numbers come from, we now cite Fig. 2 ("*which comprises 186 of the 514 total features, as opposed to the negative dataset's 328 features **(Fig. 2)***").

Line 405: sensitivity analysis should be provided

This has been addressed already

Line 410: see comments on using existing GIS-based semi-automatic tools. These should help reducing these limitations

Comments related to the GIS tools have been already addressed

Lines 413-415: This is concerning, and it does not provide high confidence this is the best methodology to use, at it's current state.

As reported in response to previous comments, we want to clarify again that demonstrating completeness in detecting submarine calderas worldwide was not the goal of this study. We now emphasise that. Although our framework identified 75 new submarine calderas, and this represents a minimum and conservative number, it still corresponds to an increase of 200% or more, relative to current knowledge. At present global bathymetric resolutions, achieving detection completeness is likely impossible, even with further automation (e.g. GIS-based tools) or with the integration of high-resolution local bathymetries. We acknowledge that the number 75 may increase, but it will still not reach completeness. Increased automation may also introduce additional uncertainty. Similarly, training the model would not guarantee full detection of all features and could introduce new biases depending on the training data used. For these reasons, and given the purpose of this study, we continue to value the expert-based component of our framework.

Nevertheless, we expanded the description of the manual classification, turning it into guidelines, to ensure reproducibility by different users. This is now reported in the methods, at the start of the feature classification section:

Edited/added text (new lines: 827-848): "*Here we report guidelines on how the classification was conducted, to ensure consistency across users. Manual classification of the 514 filtered features was carried out using multiple independent lines of evidence (Table 3). Each feature was first visually inspected in GEBCO to assess its overall morphology (e.g. rim continuity, sub circularity, presence of a surrounding edifice). When available, higher resolution multibeam bathymetry (MBBDB[42]) (90 m resolution) was consulted qualitatively to refine the interpretation and verify whether apparent rims or depressions in GEBCO were artefacts of low resolution (whether a particular feature had multibeam coverage or not, was inserted as an information in the feature dataset in Supplementary Files 1, and Submarine Caldera Dataset in Supplementary Files 2). We also examined the topographic profiles extracted for each feature (Supplementary File 3 - compilation of all features, calderas and non-calderas, with their respective bathymetric maps and elevation profiles), which provided quantitative* *information on rim elevation, depression depth, and inner slope geometry. These profiles were particularly useful for confirming the presence of steep inner walls characteristic of calderas, as opposed to the gentle, reef modified inner slopes typical of sunken atolls and guyots. We also used the global seamount dataset of Gevorgian et al.*[23] *to determine whether each depression was located on, or immediately adjacent to, a mapped seamount, providing an additional line of evidence for confirming the volcanic nature of the feature. For young atolls, classification confidence was 100% because these features were additionally cross checked against the global coral reef database*[74]*, providing independent confirmation of reef rimmed structures. Classifications were assigned only when at least one supporting parameter was met (Table 3, "Support" column), and features lacking sufficient morphological or contextual evidence were labelled as undefined. This systematic approach ensured consistency across the dataset and provided a reproducible basis for the subsequent PCA based validation.*"

Response to reviewers

Reviewer 1

The authors have addressed my previous comments, and I find that the manuscript has improved significantly. I still have some doubts about the proposed method given that it misses some key volcanic calderas, but this limitation is now better discussed and I consider the manuscript a valuable step forward. Therefore I recommend publication after minor revision.

We thank the Reviewer for their constructive assessment and for recognising the improvements made during revision. Their comments helped us clarify the limitations of the method and strengthen the manuscript.

General comments:

• Please ensure that all figures use colour schemes that are accessible to colour-blind readers. The frequent use of red and green may be problematic.

We changed symbol colours and/or symbol types of figures 1, 3 and 5, to make them colour blind friendly, and also adopted a colour blind friendly style for the new figure 6. It was not necessary for figure 4 as the key symbols were already different.

• For all figures, ensure that all abbreviations are clearly defined in the captions.

All abbreviations have now been defined in the figure captions.

• The font size jumps in the manuscript. Please check for consistency.

The font-size differences between the main text and the Methods section are intentional and follow the standard formatting used in Nature journals. We have checked for internal consistency within each section. Captions and tables also use a smaller font size, as per the journal's style. If a different formatting is required, we will be happy to adjust this at the proof stage.

Minor comments:

• Please write out the abbreviation PCA in full in the abstract.

Done.

• Line 161: Please check the reference formatting for "2023^(25)."

Reference format fixed.

• Line 300: Please expand the caption and define the abbreviations CDA and PCA.

Done.

• Lines 339–344: Please revise the wording to avoid starting multiple consecutive sentences with "This" or "These."

Text fixed.

• Figure 3: I recommend using distinct colours for panels A and B to avoid confusion. Additionally, please increase the font size of the y-axis labels in panels a and B.

Figure 3 has been replaced with an improved version, following the reviewer's suggestions.

• Please ensure consistent capitalization of specific features throughout the manuscript. For example, in lines 648–650, Axial Seamount is capitalized, whereas Kolumbo volcano is not.

This part was removed from the text following reviewer 2's suggestion.

• Figure 1 & 5: Please review the axis ticks. They appear to overlap with the figure content and are inconsistent (e.g., 90° is missing "N," while the southern equivalent is not labeled).

Figure 1 and 5 have been improved as requested.

• Line 652: Please define the abbreviation CTBTO-MS.
This part was removed from the text following reviewer 2's suggestion.

Reviewer 2 (William Chadwick)
General Comments
This is the second time I have reviewed this manuscript (this time the revised version). While the authors have made very substantial revisions that address some of the issues I raised in my previous review, I feel that it still needs additional revision before it is ready for publication. To me the main problem is that the authors are still overselling the significance and implications of the study, instead of providing more detailed information about the strengths and weaknesses of their method to identify submarine calderas from the GEBCO global bathymetry database. This just weakens their credibility and the actual impact of the work. Too much of the text is exaggerated claims and tangential hyping of volcanic hazards. As before, I have made many editorial comments and suggestions in an annotated manuscript, with the main comments listed below.
Many thanks for the time and effort spent re-reviewing, we appreciate the comments provided and have revised the manuscript to address all of them. In particular, we have modified the abstract, modified Figure 5 as requested, and produced an additional figure (Fig. 6) showing regional maps and labelled calderas, and greatly reduced the paragraph that included an initial expanded reference to hazard, treating the analysis instead as a first step towards prioritising future sea missions to calderas.

Specific comments keyed to line numbers in the annotated manuscript
• Lines 24-25: (Abstract) To me, this whole aspect of ID-ing 8 "hazardous" calderas is a bit arbitrary and unconvincing. I would omit this sentence from the abstract and replace it with one that reveals that many known calderas were not ID-ed by this method and a brief explanation of why not. That would be more relevant and honest information about the strengths and weaknesses of how this new method performs.
Abstract modified accordingly (new lines: 24-26): "*A sensitivity test shows that under detection of known calderas is largely associated with low circularity indexes (<50%). Despite this limitation, our…*"

• Line 94: I think the following reference might be more appropriate here (and wherever else it is cited), since ref#14 is just a conference abstract:
Mastin, L. G., A. R. V. Eaton, and S. J. Cronin (2024), Did steam boost the height and growth rate of the giant Hunga eruption plume?, Bull. Volcanol., 86(64), doi:10.1007/s00445-024-01749-1.
Reference replaced with Mastin et al. (2024) throughout the text.

• Line 180: The PDF format of the Tables in Supplementary files 1 & 2 as submitted for review, are unusable because the (unnumbered) columns are spread out and appear on many different pages, so the (unnumbered) rows are impossible to easily correlate between pages. This is a sloppy and ineffective way to present information to readers. Instead, these should be presented in Excel or tab-delimited or comma-delimited (.csv) files, or otherwise better formatted into more easily readable tables.
This must have been an automatic conversion from the journal submission system, as we indeed uploaded supplementary files 1 and 2 as excel files. Now, following the editorial standards for this journal, we renamed the Supplementary files 1 and 2 with Supplementary Data 1 and 2. These files should be available as .xlsx.

• Lines 227-228: The order and/or association is wrong here: Toto (listed 3rd) is in the Mariana arc (listed 2nd), and Kemp (listed 2nd) is in the South Sandwich arc (listed 3rd).
The order of caldera locations has been fixed

Also, "three" should be "four" in this sentence. In reviewing the maps of the 78 identified calderas in Supplementary File 3, I have identified another previously known caldera among the 78, just by eyeballing the bathymetry that happens to be familiar to me. SCD-0025 is a caldera on one of the Vance Seamounts, which is an off-axis chain on the western flank of the Juan de Fuca Ridge in the NE Pacific. Here are several references about the Vance Seamounts:

Portner, R. A., D. A. Clague, and J. B. Paduan (2014), Caldera formation and varied eruption styles on North Pacific seamounts: the clastic lithofacies record, Bull. Volcanol., 76(845), doi:10.1007/s00445-014-0845-3.

Clague, D. A., J. R. Reynolds, and A. S. Davis (2000), Near-ridge seamount chains in the northeastern Pacific Ocean, J. Geophys. Res., 105, 16541–16562.

Hammond, S. R. (1997), Offset caldera and crater collapse on Juan de Fuca ridge-flank volcanoes, Bull. Volcanol., 58, 617–627.

The fact that I keep finding previously known calderas among the 78 ID-ed in this study (I also found that Toto caldera was among them) shows to me that the authors are not very knowledgeable about existing calderas and have not been very careful about checking whether or not their 78 ID-ed calderas are previously known or not. Are there more known calderas in there?

We thank the reviewer for this careful assessment and for identifying an additional previously documented caldera among the 78 features. Our identification of known calderas relied primarily on the two existing global reference datasets, the GVP and the CCDB, which, to our knowledge, represent the only comprehensive global compilations of calderas. Because the broader literature on submarine calderas is heterogeneous and often regionally focused, some documented systems outside these databases were initially omitted. Following the reviewer's suggestions, and after an additional verification on our side, we have now incorporated the caldera on the Vance Seamounts chain (SCD-0025), citing Portner et al. (2014) as a representative reference, and we have added one further documented example (Niuafo'ou caldera – Kempe & Kazmierczak, 2012). This brings the total number of previously known calderas among the 78 features to five. Niuafo'ou (SCD-0007) was in fact listed in GVP but had been mistakenly categorized as undocumented in our initial screening because the GVP reports it primarily as a shield volcano. Additionally, due to its partially emergent nature (Stage 4), it had also been included in the shortlist of calderas of interest for sea exploration. We have now corrected this oversight and removed Niuafo'ou from the shortlist, which is intended to focus exclusively on calderas without prior documentation. The manuscript text and figures (Figs. 5–6) have been updated accordingly. We appreciate the reviewer's input, which has helped us strengthen the completeness and clarity of the documented-vs-undocumented listing.

This is also why I requested in my previous review that a figure be added to show the 78 ID-ed calderas on more detailed regional maps, to make it easier to see them in the context of known volcanic arcs, hot-spot trails, and mid-ocean ridge settings. Regional maps could likely reveal additional examples just by better knowing the geographic context of where their calderas are identified and not identified. This has not been done in this revised manuscript, and I still think that is a serious omission that makes it harder for readers to evaluate the performance of this method and the validity of the newly ID-ed calderas.

We have produced an additional figure (Fig. 6) that shows regional maps of key areas with labelled calderas from our dataset, and also highlighting key tectonic landmarks.

• Line 232: A short summary of what the sensitivity analysis showed should be added here.

Short summary added (new lines: 304-307): "*In this test, we extracted areas surrounding several known calderas and re-ran the CDA using additional oversampled resolutions, generating more tiles than in the original workflow. The results show that a circularity index above ~50% is critical for successful caldera identification…*"

• Line 245: (Figure 5) These maps of the 8 "high-priority" calderas would be more useful if they

included depth information - with either labeled color ramps or labeled depth contours. Geographic information (lat, long axis labels) should also be added to each map. There needs to be some information added to the caption about how to identify these 8 calderas in the Supplementary File 3, which includes all 514 caldera candidates. It would help to re-order the 8 maps in ascending ID order, to make it easier to identify them in the Supplementary File.
The figure and caption have been improved as requested. Bathymetric contours have been preferred over coloured maps because we overlaid multibeam graphic where available, hence a double colour scheme would create confusion.

• Lines 260-269: This paragraph is repetitive and hyperbolic. Delete.
Paragraph deleted

• Line 287: I would greatly reduce this paragraph. To me, this "hazard analysis", based on supposed growth stage, is so crude and based on so little concrete information, that it does not warrant the amount of text devoted to it. To me, the overselling of potential hazards substantially weakens the paper and seems like exaggerated claims made hoping to make the results "more relevant", but instead just makes the whole endeavor less convincing and more suspect, especially because the hazards have already been hyped in earlier parts of the paper.
This paragraph has been reduced from 417 to 226 words, and now it has less emphasis on hazard, and treats this analysis as a preliminary prioritization for future sea missions, rather than a hazard assessment. However, we do retain the growth stage information to justify the selection.
New paragraph (new lines: 401-434): "*Based on this analysis, we identify seven submarine calderas as suitable candidates for future deep sea exploration (Table 1 and Fig. 6). Because higher resolution geological and volcanological information, which will allow a more hazard-based quantitative assessment58, is not available for most regions, we rely only on parameters that are consistently mapped at the global scale. As a first order filter, we consider the growth stage classification, which reflects volcano size and water depth and provides a broad indication of morphological maturity48, and associated hazard potential19,20,24,59. We focus on stage 3 (<700 m water depth) and stage 4 (partially emergent) calderas, as these are more accessible for seafloor investigations and have experienced more substantial eruptive or constructional histories. We also consider tectonic setting, prioritizing calderas along volcanic arcs, where magmatic systems tend to be longer lived. Four of the selected calderas lie in volcanic arcs (SCD 0027, SCD 0048, SCD 0054 in the Tonga arc, and SCD 0006 in the Sangihe arc; Fig. 6A,E), two are associated with spreading centres (SCD 0034 in the Manus spreading centre and SCD 0049 in the Cocos–Nazca spreading centre; Fig. 6E and Fig. 6C), and one is a shallow intraplate caldera (SCD 0030) in the Bismarck Sea (Fig. 6E). Smaller volcanic depressions (<3 km diameter) are not considered at this stage because their morphology provides limited constraints on past activity, and global datasets do not yet allow a more detailed assessment.*"

• Line 299: More detailed regional maps would help the reader see the setting of these calderas.
An additional figure (Fig. 6) has been added to address this concern.

• Lines 360-369: To me, this stuff is not relevant to the scope of this study. I would omit.
Discussions about the future monitoring strategies have been removed.

Reviewer 3

Dear Editor,
After careful evaluation, I find that the authors have responded satisfactory to all of the reviewers' comments and concerns, and have followed the feedback provided.
After this second round of revisions, the revised manuscript has significantly improved in clarifying the methodological explanations and in the addition of a new sensitivity test section, which strengthens the overall quality and rigor of the presented framework.

The authors have also revised the discussion section to include the limitation of the methodology while highlighting the value and contribution of the framework.
Overall, I believe the manuscript now meets the standards required for publication in the journal. I have added a few minor reviews and comments, but I recommend acceptance of the manuscript pending these minor revisions.
We thank the Reviewer for their thorough evaluation and for recognising the improvements made during the revision process. Their comments were instrumental in strengthening the methodological explanations, the sensitivity analysis, and the overall clarity of the manuscript.

Comments extracted from PDF:

Lines 18-19: not providing much information. Either delete or expand (e.g., what logistic constraints)
Constraints now specified (new line: 18): "…*the scarcity of high resolution bathymetry data*"

Line 20: Principal Component Analysis
Now written in full

Line 20: 'depression detection' is a bit odd...suggest: combines the use of an algorithm for detecting depression areas in seafloor, expert classification, and Principal Component Analysis supported validation.....
Sentence edited (new lines: 19-21): "…*combining an automated algorithm for detecting depressions on the seafloor, expert (user-based) classification, and Principal Component Analysis validation….*"

Line 39: (VEI)
Text fixed

Lines 55-56: maybe clarify why it is so difficult... time consuming and challenging to map (multibeam bathymetry surveys) and to sample, challenging to collect sediment samples
Text edited (new lines: 58-59): "…*mostly due to lack of high-resolution bathymetry in over 70% of Earth's ocean floors. Hence…*"

Line 123: assessed / evaluated
Text fixed

Line 137: peer-reviewed global datasets
Text fixed

Line 137: Ref 22
Reference added

Line 243: add space
Typo fixed

Line 294: why favoring? the use of this term is confusing
(new line: 410) We replaced it with "*prioritising*", which is more appropriate.

Line 356: Autonomous Underwater Vehicles (AUV). Also a tool to consider, as these are complementary to USV. USV being best for shallow-water surveying and AUV being better for deep-water systems and close to seafloor high-resolution bathymetry acquisition
Text edited (new line: 546-547): "…*and Autonomous Underwater Vehicles (AUVs) offer cost-effective alternatives and complementary capabilities to conventional ship-based surveys…*"

Global machine-learning detection of submarine calderas

Andrea Verolino[1*], Christopher Lee[2], Susanna F. Jenkins[1,3], Martin Jutzeler[4], Adam D. Switzer[1,3]

[1] Earth Observatory of Singapore, Nanyang Technological University, Singapore
[2] Department of Physics, University of Toronto, Canada
[3] Asian School of the Environment, Nanyang Technological University, Singapore
[4] Centre for Ore Deposit and Earth Sciences (CODES) and School of Natural Sciences, University of Tasmania, Australia

* Corresponding author (email address: andrea_verolino@hotmail.it)

Abstract

Submarine calderas are among the least documented volcanic structures on Earth, yet recent impactful events highlight their potential for severe geohazard consequences globally, including tsunamis, seafloor damage, and atmospheric impacts. Logistical limits have historically hindered global detection. Here we apply a machine-learning-based caldera detection algorithm to global bathymetry, enabling systematic identification of previously undetected submarine calderas. We found 78 calderas spanning different water depths (up to 5,600 m), diameters (up to 20 km) and tectonic settings (divergent, convergent, intraplate). A subset of eight shallow-water calderas, primarily within volcanic arcs, are identified as high-priority targets due to their high hazard potential. Our dataset fills a critical observational gap and offers a reproducible and upgradeable framework for submarine volcano characterization, geohazard assessment, and deep-sea exploration. These findings underscore the need to integrate submarine calderas into future global hazard modelling and monitoring strategies.

Introduction

Calderas are kilometer-scale volcanic depressions that result from the voluminous eruption of magma from the Earth's crust. Caldera-forming eruptions produce very large amounts of material and are capable of re-shaping landscapes and influencing global climate[1]; they can form from passive effusions of magma[2] or high-intensity explosive eruptions[3]. Explosive caldera-forming events have triggered periods of abrupt surface cooling due to massive ash and gas emissions into the atmosphere[4]. For instance, the Toba eruption (~74,000 years ago) is considered one of the largest volcanic events in the last two million years, releasing approximately 2,800 km³ of magma and potentially inducing a global volcanic winter[1]. Similarly, the Tambora eruption of 1815, one of the most powerful events recorded (Volcanic Explosivity Index= 7, following Newhall and Self[5]), injected 60-80 megatons of SO_2 into the stratosphere, leading to the infamous "Year Without a Summer", which caused widespread crop failures and famine across Europe and North America[6,7]. Beyond their catastrophic past eruptions, many calderas remain active, with evidence of ongoing seismicity, ground deformation, and magmatic movement. For instance Yellowstone Caldera (USA), continues to exhibit signs of deep reservoir activity[8], while Campi Flegrei (Italy) has shown increasing seismic unrest since 2005, suggesting evolving magmatic processes[9,10]. To identify and better understand the formation of calderas is therefore crucial for reconstructing past eruptive histories and assessing potential future hazards.

The vast majority of research into calderas focuses on subaerial structures, and a considerable knowledge gap remains regarding their submarine counterparts. Although the majority of global volcanism occurs beneath the ocean[11], submarine calderas remain largely unexplored and poorly documented. Submarine volcanoes are critical to study because they play a key role in understanding geological hazards, marine ecosystems, global cycles, and resource potentials. Major eruptions from submarine calderas, such as the 2022 Hunga Tonga event, in the Kingdom of Tonga, have recently demonstrated their potential for extreme explosivity and global impact[12–14]. The eruption triggered far-reaching tsunamis (>10,000 km), with fatalities as far away as Peru[15], and resulted in local economic losses equivalent to approximately 18% of Tonga's GDP[16].

Existing global caldera databases highlight the imbalance between subaerial and submarine records. For example, the Global Volcanism Program (GVP) identifies 140 Pleistocene-Holocene calderas, of which 94% (n= 132) are subaerial and only 6% (n= 8) submarine[17]. Whereas, the Collapse Caldera Database (CCDB) lists 640 mapped calderas[18], yet reports fewer than 2% as submarine (visually assessed as water depth information is missing from the database) (Fig. 1). Meanwhile, recent discoveries of over 43,000 seamounts[19] suggest that many unidentified submarine calderas likely exist, and some are almost certainly volcanically active[20].

Tectonic setting exerts a primary control on submarine volcanic activity, with most of Earth's active volcanism concentrated along plate boundaries, either at divergent margins (spreading ridges) or convergent zones (subduction arcs)[21]. Calderas in continental convergent settings typically exhibit more silicic compositions, owing to thicker crust and longer magma residence times, which favor volatile enrichment and high viscosity. These conditions, in turn, enhance explosive potential compared to basaltic calderas[22]. However, such characteristics are less commonly reached in convergent oceanic regions, where the crust is relatively thin. Nevertheless, silicic oceanic arc calderas do exist and are among the most powerful[12]. Calderas also occur at spreading ridges, though they are typically basaltic and associated with fast magma replenishment and limited volatile content, conditions that generally suppress explosive behavior. However, even ridge-associated systems can display sudden energy release, particularly when volatile-rich magma ascends from locally heterogenous mantle domains[23]. Intraplate volcanism, similarly, tend to produce basaltic calderas (e.g. Hawaii[24]), but can also yield a broad compositional spectrum, from mafic to highly silicic examples such as the Canary Islands[25] and Galápagos[26]. In general, intraplate submarine volcanism is dominated by seamounts of mafic compositions, the majority of which are ancient and likely inactive, posing minimal risk. Regardless of composition, explosive submarine calderas can develop across all tectonic settings[27], and intense magma-water interaction, easier to reach for low-viscosity magmas (i.e. basalts)[28], can substantially amplify explosive behavior[12,29–34].

In this study, we applied a machine-learning-based caldera detection algorithm (CDA) to systematically identify submarine calderas worldwide. Our CDA builds upon the algorithm originally developed by Lee and Hogan[35] for detecting Martian impact craters, which demonstrated high detection precision (>80%) for sub-circular depressions ⩾4 km in diameter with minimal training. We implemented the CDA using a series of morphometric and resolution filters to effectively detect submarine calderas within a low-resolution bathymetric digital elevation model (GEBCO 2023 global bathymetry dataset[36], ~500 m resolution). These filters helped reduce the number of false positives and enabled manual classification of the retained features, followed by Principal Component Analysis (PCA) to assess and validate the nature of the identified structures (see Methods for details).

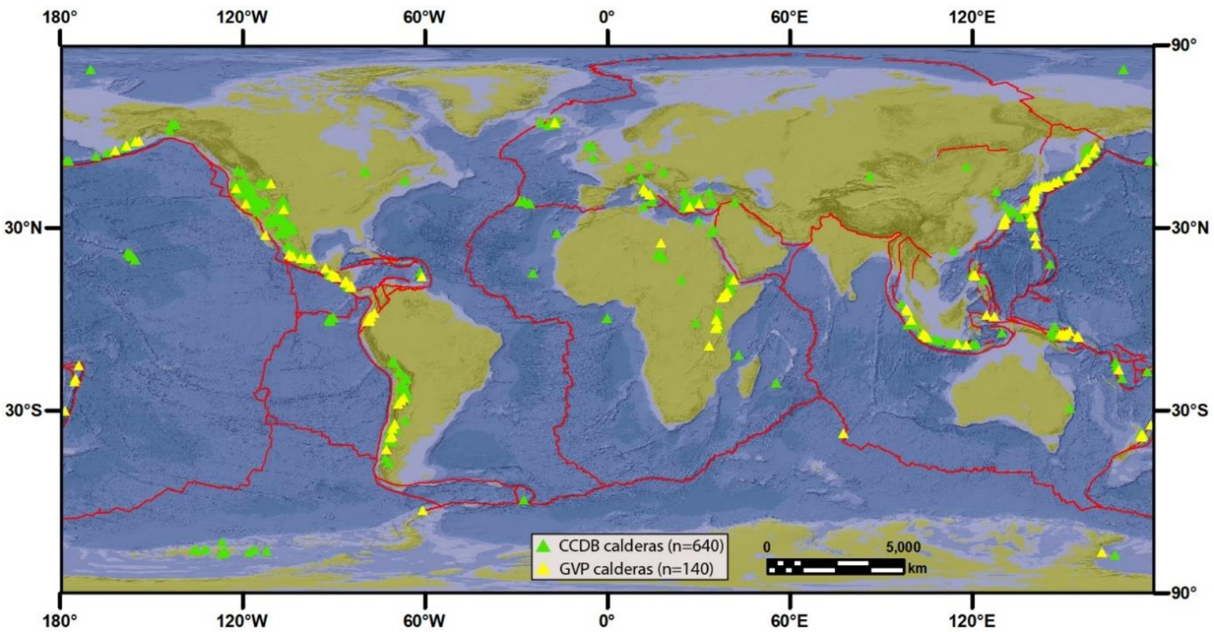


Fig. 1| Global distribution of documented calderas. Caldera locations from the Global Volcanism Program (GVP) (yellow triangles) and from the Collapse Caldera Database (CCDB) (green triangles). Note the vast lack of calderas within ocean basins. The red lines are tectonic plate boundaries. The basemap is from GEBCO 2023[36].

In addition to evaluating the effectiveness of the CDA in the semi-automated identification of previously undocumented submarine calderas, this study primarily aimed to establish a global and upgradable dataset to inform future targeted research, including for example the deployment of monitoring instruments and targeted ~~remote sensing~~ observations at key calderas. Our dataset comprises 78 volcanic depressions larger than 1 km in diameter, which we classified as either calderas (diameter >3 km) or volcanic craters (diameter <3 km). Based on global consensus[22], calderas are broad topographic depressions formed by subvertical subsidence into a partially drained magma reservoir, typically associated with large eruptive events, whereas craters are generally smaller, bowl- or funnel-shaped features produced by explosive fragmentation and ejection of host rock near volcanic vents. Importantly, these definitions lack a definitive size threshold for distinguishing calderas from craters, and notable databases such as the CCDB list calderas even smaller than 1 km, creating potential overlap with crater dimensions. Given this ambiguity, compounded by the limited resolution of global bathymetric datasets and the tendency for both feature types to share comparable size ranges in submarine settings, we adopt an inclusive classification approach in this study, with both calderas and craters grouped under the ‘Submarine Calderas Dataset’. This designation is particularly suited to submarine environments where direct observation remains limited.

Out of the 78 detected volcanic depressions, we propose an initial set of 8 submarine calderas (Table 1) as high-priority targets for scientific investigation, selected based on their water depths, tectonic settings, and caldera morphological type, factors that collectively suggest an elevated likelihood of future eruptive activity and potential societal impact.

This study lays a critical foundation for evaluating submarine volcanic risks and improving global preparedness. By applying the CDA, we establish a new global dataset of submarine calderas, advancing our understanding of explosive subaqueous volcanism and guiding the prioritization of regions where volcanic hazards may be most consequential. The urgency of this endeavor is underscored by recent high-impact events, such as the 2022 Hunga Tonga eruption, and the increasing vulnerability of seafloor and coastal infrastructure. The proposed list of high-priority calderas represents a pivotal starting point for targeted deep-sea exploration and integrated geohazard assessment.

Results

Feature classification and validation through Principal Component Analysis (PCA)

The initial CDA output of 87,435 potential features was narrowed to 514 candidates (Supplementary File 1) based on morphometric criteria and reliability indicators generated by the algorithm; this was accomplished to reduce the initial high number of false positives. These filtered features were then manually categorized through a first-order and second-order classification (Fig. 2), further distinguishing calderas from non-calderas. Details of the workflow, filtering methodology, and classification procedure are provided in the Methods section. Specifically, from the first-order identification (Fig. 2A) we found 56 *calderas*, 22 *craters or small calderas*, 62 *atolls*, 145 *flank scars or erosion surfaces*, and 229 remained *undefined* (i.e. not classifiable due to poor bathymetry and/or unclear structure). Excluding *undefined* features, *calderas* and *craters/small calderas* account for 27% of the filtered dataset, *flank scars and erosion surfaces* 51%, and *atolls* 22%. The results of the second-order classification indicate that, of the caldera subset (Fig. 2B), *horse-shoe or eroded calderas* dominate (~53%), followed by *full calderas* (~19%) and *craters/small calderas*, which are evenly split between full craters/calderas and horse-shoe types (~14% each). The second-order classification exercise was performed to: (1) assess the efficacy of PCA as a supporting classification tool; and (2) better constrain eruptive and volcano degradation history, providing a preliminary hazard assessment based on edifice collapse (e.g. a horse-shoe caldera may indicate that it has already experienced large lateral collapses and/or underwent major erosional processes; hence, potentially less hazardous than a full caldera).

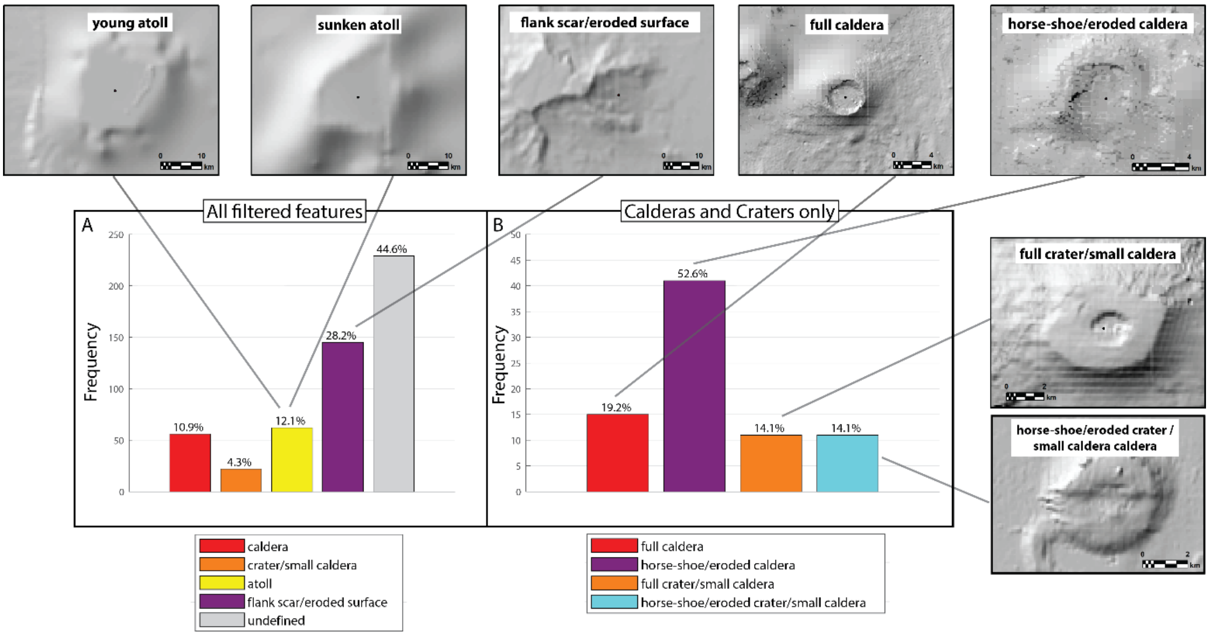


Fig. 2| Feature classification results. (A) First order classification for the filtered features. (B) Second-order classification for the calderas and craters only. Also reported are close-up images of each sub-type. Basemap for the calderas and craters examples is from the Multibeam bathymetry Database (MBBDB)[37].

The PCA analysis was conducted both for the feature types (first-order classification) and sub-types (second-order classification). The first-order classification PCA (Fig. 3A) is very efficient at distinguishing clusters that correspond to atolls, calderas, and flank scars/eroded surfaces. For the second-order classification, the PCA successfully differentiates clusters for the two atoll sub-types (young vs sunken; see Methods). Interestingly, full calderas and horse-shoe/eroded calderas largely overlap within the same field (Fig. 3B). Flank scars and eroded surfaces remain well defined, though a few outliers fall into other clusters. Such errors may partly be linked to locally low bathymetry resolution.

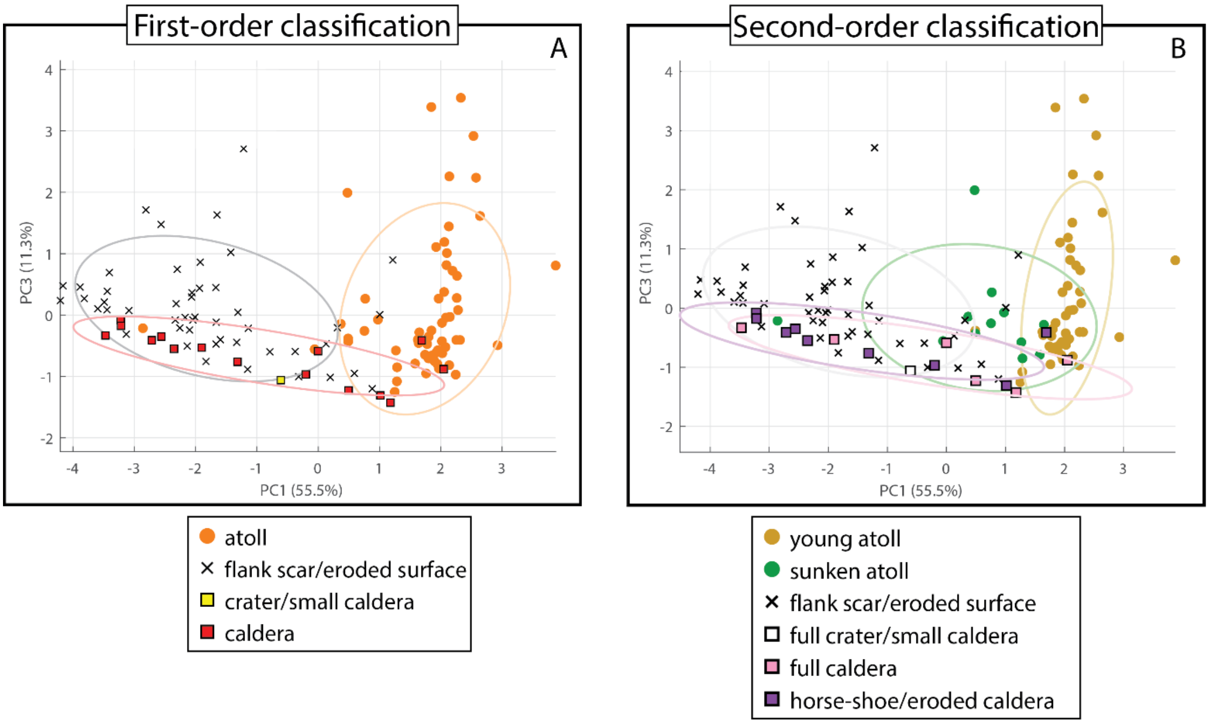


Fig. 3| Principal Component Analysis results. (A) PCA for first-order classification by type. (B) PCA for second-order classification by sub-type. The ellipses were obtained using a 75% confidence level, assuming a multivariate normal distribution, and colored based on the respective types and sub-types.

Submarine calderas dataset

Our CDA and feature classification identified 56 potential submarine calderas and 22 potential small calderas or craters, all grouped into a 'submarine calderas' dataset, totaling 78 sub-circular landforms with >1 km depression diameter (Fig. 4 and Supplementary File 2). Of the detected candidates, only two calderas were previously known, Niuatahi[38] and Kemp[39] calderas in Tonga and along the South Sandwich arc, respectively. Several other well documented calderas, including Hunga Tonga, Kikai, Krakatau, Santorini, Kolumbo and Axial were not identified by the CDA due to two primary factors, either because their location lay outside the DEM used, or because of technical limitations of the CDA itself (more details are reported in the Methods). Among the 78 potential calderas, 8 occur in divergent settings, 9 in volcanic arcs, and 61 in intraplate environments (Fig. 4 and Extended Data Fig. 1). In terms of water depths, the calderas we detected span between 5,600 m below sea level and 90 m above sea level (partially emerged). Water depth, together with volcano height, have been previously associated with volcanic hazard potential[20], through the volcano growth stage[40], with stages 3 and 4 being the more hazardous, and stages 1 and 2 the least hazardous[20]. Following this categorization, most calderas (n= 54) are classified as stage 1 (>700 m water depth, <1000 m height), followed by stage 2 (>700 m depth, >1000 m height) with 16 calderas, stage 3 (<700 m water depth) with 6 calderas, and stage 4 (partially emerged) with 2 calderas

(Extended Data Fig. 2).These calderas have a diameter between ~ 1 and 20 km, with the majority being in the range 1-6 km (Supplementary File 2).

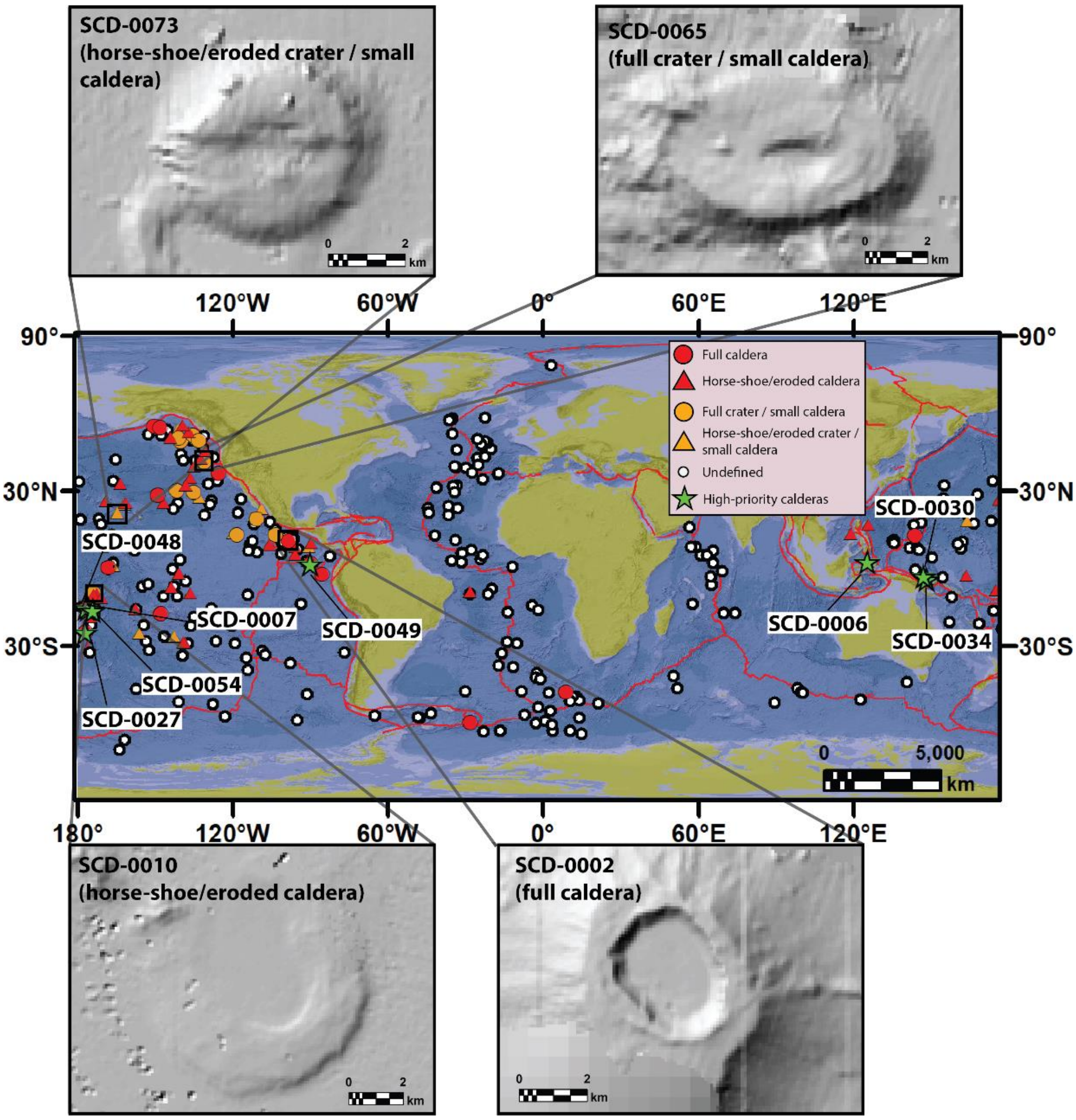


Fig. 4| Global map with the location of the classified sub-types of calderas and craters, and high-priority calderas. Undefined features are also reported (white circles), some of which may be calderas. The basemap is from GEBCO 2023[36]. The smaller panels show examples of sub-types of caldera and crater identified in this study. Basemap for the calderas and craters examples is the Multibeam bathymetry Database (MBBDB)[37].

Discussion

Submarine explosive eruptions remain poorly documented compared to their subaerial counterparts, yet recent events underscore their devastating potential[12,13,15]. They also host hydrothermal vent systems that support unique, chemosynthetic life, offering insights into early Earth conditions and ecosystem resilience in extreme environments[41,42]. Therefore, identifying submarine calderas is essential not only for hazard-focused research, but also for advancing understanding in fields such as deep-sea ecology[43].

In the context of risk, the 2022 Hunga Tonga caldera, submerged 200 m below sea level prior to eruption, demonstrated that submerged silicic calderas can produce high-energy explosive activity, initiate submarine density currents able to destroy critical underwater infrastructure, and generate tsunamis and atmospheric disturbances on a global scale[44–46]. This highlights the need for improved hazard assessments, particularly for shallow and partially emerged volcanic systems, where magma-water interactions intensify explosivity[29–32]. This is particularly the case for stage 3 (<700 m b.s.l.) and stage 4 (partially emerged) calderas[20]. Eruptions at these volcanoes can trigger pyroclastic density currents, inject into the atmosphere considerable amounts of ash and gas, and be accompanied by large-scale tsunamis, with direct consequences for nearby to global coastal and regional populations. A notable historical example is the 1883 eruption of Krakatau, Indonesia, where partially submerged explosive activity triggered catastrophic tsunami waves, with pyroclastic density currents entering the surrounding sea and playing a central role in the tsunamis generation[47]. The eruption also injected large volumes of ash and sulfur-rich aerosols into the atmosphere, causing measurable global cooling and disrupting atmospheric circulation patterns[48]. Submarine infrastructure, including communication cables and deep-sea pipelines, is particularly vulnerable to caldera-related activity[49]. The 2022 Hunga Tonga eruption severed Tonga's primary communication cable, isolating the nation for weeks[50]. Additionally, post-eruption earthquakes and submarine landslides continued to reshape the seafloor in the months following the eruption. These secondary events cut Tonga's domestic cable several months later, compounding the disruption to communication and emergency response[50]. Similarly, hydrothermal vent systems near active submarine calderas can induce cable degradation, affecting global data transmission networks[49].

In our classification, we distinguish full calderas from horse-shoe/eroded calderas. Their morphological features offer insights into their past and future eruptive behavior. Full calderas, with their enclosed circular structure, generally result from high-magnitude explosive eruptions, where rapid magma evacuation causes structural collapse[51]. In contrast, the breached structure of horse-shoe calderas may reflect past sector collapse, which may have generated large debris avalanches that could have triggered tsunamis[52]. Caldera collapse and subsequent constructional volcanism may occur cyclically, and the formation of one caldera does not preclude the possibility of recurrent caldera-forming or collapse events within the same volcanic system[53]. These morphologies can direct future eruptions along structural weaknesses, with magma migrating preferentially to open flanks. Alternatively, this morphological category may reflect extensive post-eruptive erosion, suggesting prolonged inactivity, hence less potentially hazardous. Understanding these morphological distinctions improves hazard assessments, particularly in submarine regions where direct observations remain limited.

Based on this analysis, we prioritize eight submarine calderas for future deep-sea exploration using multiple criteria (Table 1). First, we considered stage 3 (depth <700 m) and stage 4 (partially emerged) calderas, where explosive activity is more likely to occur and produce high-impact scenarios[20]. Second, we assessed tectonic settings, mostly favoring calderas situated along volcanic arcs, where tectonic activity can facilitate the sustained accumulation and recharge of volatile-rich magmas. Some shallow calderas in extensional and intraplate settings are also considered as high-priority, given their increased potential for explosive magma-water interactions involving basaltic compositions[28]. Within each of these categories, we also consider the morphological sub-type, treating full calderas as more hazardous than horse-shoe and eroded calderas, based on the reasoning outlined above. Five of the eight selected calderas are located in volcanic arc settings, four of which in the Tonga arc (SW Pacific) and one in the Sangihe arc (east Indonesia). Two calderas are associated with spreading ridges: SCD-0034 (Manus spreading center, NE of Papua New Guinea) and SCD-0049 (Cocos-Nazca spreading center, ~150 km N of Galapagos). Additionally, we included the only shallow (<700 m) intraplate caldera of the entire dataset (SCD-0030), which is located in the Bismarck Sea (NE of Papua New Guinea) on the South Bismark plate, at <300 km from populated coastlines. Although many of these calderas are situated relatively far from major coastlines (Fig. 4), they nonetheless pose significant hazards, including tsunami generation, tephra dispersal, and ~~potential~~ damage to submarine infrastructure, as exemplified by the 2022 Hunga Tonga eruption[12,13,15,50]. Craters and small calderas (classified here as volcanic depression <3 km in diameter) are not prioritized at this stage due to their limited size, which is assumed to correlate with lower explosive potential and reduced tsunami hazard from smaller collapse volumes. This prioritization should be interpreted as a preliminary recommendation rather than a comprehensive or definitive list.

Table. 1| List of high-priority calderas. They were selected based on their growth stage (i.e. water depth), tectonic setting and morphological sub-type.

Caldera name	Long (°)	Lat (°)	Caldera sub-type	Growth stage (water depth)	Tectonic setting
SCD-0007	-175.64	-15.60	full caldera	4 (~0 m)	Volcanic arc (Tonga arc)
SCD-0006	125.44	2.70	horse-shoe/eroded caldera	4 (~0 m)	Volcanic arc (Sangihe arc)
SCD-0048	-176.60	-15.10	full caldera	3 (~100 m)	Volcanic arc (Tonga arc)
SCD-0027	-176.91	-24.59	full caldera	3 (~500 m)	Volcanic arc (Tonga arc)

SCD-0054	-173.90	-16.22	horse-shoe/eroded caldera	3 (~400 m)	Volcanic arc (Tonga arc)
SCD-0034	147.16	-2.97	horse-shoe/eroded caldera	3 (~600 m)	Divergent (Manus spreading ridge)
SCD-0049	-90.01	1.78	full caldera	3 (~700 m)	Divergent (Cocos-Nazca spreading center)
SCD-0030	148.23	-4.10	horse-shoe/eroded caldera	3 (~700 m)	Intraplate (South Bismarck plate)

Despite the limitations (see Methods section), this work provides the basis for (1) enhanced submarine caldera discovery; and (2) targeted sea exploration to assess potential risk from poorly known submarine calderas that may experience hazardous volcanic activity in the near future. In terms of future directions, several key advancements can enhance the study and monitoring of submarine calderas in the coming years. Refinements to machine-learning caldera detection algorithms will improve the accuracy of caldera identification. Algorithmic improvements should focus on distinguishing complex morphologies, integrating geological context, and reducing false positives to enhance detection reliability at global scales. Additionally, improving accessibility to high-resolution bathymetric mapping technologies could accelerate the discovery and characterization of submarine calderas. Emerging tools such as wave gliders and Uncrewed Surface Vehicles (USVs) offer cost-effective alternatives to conventional ship-based surveys, but are still relatively expensive for research institutions. In this direction, the *Seabed 2030* initiative aims to achieve complete global coverage of the ocean floor by 2030, contributing increasingly higher-resolution data to the GEBCO bathymetric dataset. Expanding seafloor monitoring networks at high-risk calderas, particularly through pressure sensors and hydroacoustic arrays, will be essential for early detection of submarine eruptions. Existing networks, such as those deployed at Axial Caldera (Juan de Fuca Ridge)[54] and Kolumbo caldera (NE of Santorini, Greece)[55], provide a model for establishing real-time monitoring systems in other volcanic regions. Increasing observational capacity will be crucial for surveying volcanic activity, such as tracking magma movement, gas emissions, and seismic precursors to submarine eruptions. Translating the information gathered from advances in the monitoring and identification of submarine calderas to an understanding of the hazards and risks they pose to societies and critical infrastructure, such as shipping and telecommunications, requires additional steps. A greater understanding of the physical processes underpinning explosive submarine eruptions requires increased empirical and experimental data; associating the process with the hazard footprint typically requires tailored numerical modelling; and understanding how societal assets respond to the hazard requires analysis and expansion of the limited empirical and experimental dataset. Addressing these barriers through collaborative research initiatives and technological innovation will be key to advancing submarine volcanic hazard and risk assessments, and will represent a critical step toward geohazard preparedness for coastal populations and global infrastructure networks.

This work introduces a novel global-scale detection framework that demonstrates high efficacy for submarine caldera identification. The CDA enables automated recognition of sub-circular volcanic depressions across tectonic domains with minimal ~~geological~~ input. Unlike traditional mapping efforts constrained by spatial coverage or regional bias, our framework integrates morphometric filtering, GIS-based manual classification, and PCA-supported refinement, allowing reproducible and scalable feature extraction from low-resolution global bathymetry. Importantly, it successfully identified 78 calderas, 76 of which detected for the first time, spanning diverse morphological types and growth stages, which in turn reflect their hazard potential[20]. This expansion of the submarine caldera inventory, by an order of magnitude, fills a longstanding observational gap and generates actionable data to guide hazard prioritization, monitoring strategies, and targeted field campaigns. As such, this framework not only advances submarine volcanism research but also lays the foundation for integrating machine learning into future planetary-scale hazard assessments.

Methods

The framework adopted to conduct this work included four main processes (Fig. 5): (1) Use of the machine learning-based caldera detection algorithm (CDA), modified from the original crater detection algorithm created by Lee and Hogan[35]; (2) Post-processing filtering to reduce the number of false positives from the initial output list; (3) Feature manual classification, through GIS-based analysis and expert judgment; and (4) Final inspection of the outputs, through principal component analysis (PCA). Details are ~~reported~~ in the following sections.

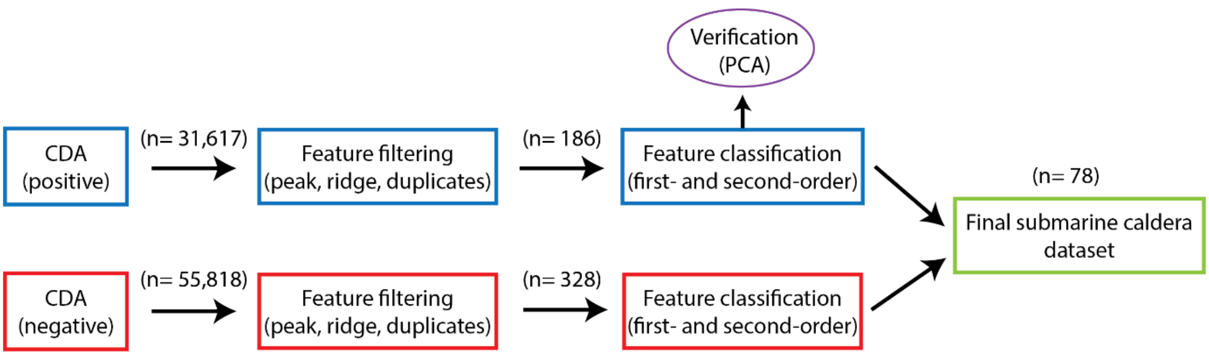


Fig. 5| Workflow for the creation of the submarine caldera dataset.

Caldera Detection Algorithm (CDA), inputs, outputs, filtering and post-processing

The original algorithm developed by Lee and Hogan[35] was trained to detect sub-circular depressions with the algorithm reported to not require further training for similar application. The only input is a digital elevation model; in our case, we used a modified version of the algorithm and refer to it as caldera detection algorithm (CDA). As input DEM we used the GEBCO 2023 bathymetry dataset[36], which has a resolution of 15 arc-seconds, corresponding to approximately 500 m at the equator, and was cropped around the over 43,000 seamounts from the Gevorgian et al.[19] dataset (100 km radius). This DEM had a coverage of about 80% of the Earth's seafloor. This approach not only supported verification on the volcanic origin of some of the potential calderas detected through the CDA, but also reduced computational effort within the GIS environment (ESRI® ArcMap™). Additionally, using a DEM cropped around previously unstudied seamounts, aligns with the scope of this work, which is to detect previously undocumented calderas. Despite its relatively low resolution, GEBCO provides global coverage, enabling the automated detection of depressions across all ocean floors. The cropped GEBCO bathymetry was then reprojected in WGS 1984 World Mercator to run the CDA.

The CDA splits the input GEBCO bathymetry map into small orthographically projected regions that overlap in location and step logarithmically in size from about 2 degrees to about 30 degrees. Each image is processed to find sub-circular features, and the location and size are stored in an output Excel table; this means that the same feature can be found in more than one image, and in that case, we will have *Duplicates*>1. The CDA main outputs include potential calderas of at least ~4 km in diameter (at the equator). This means that features <4 km could still be detected but with a lower level of confidence, particularly if they were away from the equator, due to distortion effects of the projected DEM. Each potential caldera included several output parameters, including *lat* and *long*, number of *duplicates*, the *diameter*, several water depth parameters of the feature itself (*center mean*, *inner mean*, *ring mean* and *outer mean*), and the flags *peak* and *ridge*. They are all listed and detailed in Table 2.

The original algorithm used by Lee and Hogan[35] only detects depressions, regardless of whether they are at the top of a mound or on a flat surface. Therefore, in order to "guide" the algorithm towards possible depressions on a mound, as we would expect from a caldera, we inverted the CDA and applied two water depth flags, *peak* and *ridge* (explained in detail below). This approach provided us with two datasets of potential calderas: *positive* (depressions), from the original algorithm, and *negative* (mounds), from the inverted algorithm.

Across both datasets, we had a total of over 87,000 possible calderas, most of which were false positive. Therefore, to reduce this number to only potentially realistic calderas, we used the output flags *peak* and *ridge*. *Peak*=TRUE when the potential caldera has a positive morphology in its center, and *Ridge*=TRUE when the potential caldera is elevated respect to the surrounding seafloor. Therefore, to only isolate features that resemble a classic caldera (i.e. a mound with a depression in the middle), we set as filters *peak*=FALSE and *ridge*=TRUE . This process brought the number down to nearly 1,200 potential calderas. Then, to remove the still many false positives, both datasets were further processed by eliminating features that have the lowest number of *Duplicates* (n=1), which means they are found only once in the overlapping images of different resolution mentioned earlier, hence with the lowest confidence. This brought the number of potential calderas down to 514, which were then inspected and classified individually, followed by final verification through principal component analysis.

Table. 2| Output parameters of the CDA. Note that to remove as many false positive as possible, we set Duplicates>1, Peak=false and Ridge=true as filters.

PARAMETER	DESCRIPTION
Diameter	Mean diameter of the feature
Ring mean	Mean water depth of the feature ring
Centre mean	Mean water depth measured within a narrow ring concentric to the feature center, positioned at 10% of the caldera radius, to represent near-central bathymetric conditions
Inner mean	Mean water depth measured within a concentric ring located at 75% of the feature radius, to represent intermediate bathymetric conditions within the structure.

Outer mean	Mean water depth measured within a concentric ring located at 125% of the caldera radius, characterizing the bathymetric conditions just beyond the caldera rim.
Duplicates	Number of times a feature is independently detected at consistent location and diameter across multiple reprojected sub-images from the original DEM.
Peak	If P=true, center mean > Inner mean; if P=false, center mean < Inner mean
Ridge	If P=true, Inner mean > Outer mean; if P=false, Inner mean < Outer mean

Feature classification and final inspection

To conduct the classification, we used a combination of GIS-based analyses and visual assessments. As a first step we plotted the filtered 514 potential calderas in a GIS environment and overlaid them on the hillshade derived from the GEBCO bathymetry for visual morphological inspection, bathymetric contour analysis and profiling (Supplementary File 3). As additional elements for the classification we used the seamounts catalogue from Gevorgian et al. (2023), to confirm whether the features were volcanic or not, and the global coral reef database[56], to confirm if the feature was an atoll. Some areas also had higher-resolution multibeam bathymetry (MBBDB)[37] (90 m resolution), and that information was inserted in the final dataset (Supplementary Files 1 and 2) and supported our classification.

We conducted a first-order classification, considering a *Type*, and a second-order classification, considering a *Sub-type* (where applicable), for each of the filtered features. As types we identified *calderas*, *atolls*, *craters*, and *flank scar/eroded surface*. For the sub-types we divided the types into *full calderas*, *horse-shoe/eroded calderas*, *full craters / small calderas*, *horse-shoe/eroded craters / small calderas*, *flank scar/eroded surface*, *young atoll*, and *sunken atoll*. The flank scar/eroded surface category was retained without further subdivision at both the type and sub-type levels, as it was deemed irrelevant to the objective of this study, namely, distinguishing calderas from non-calderas. In contrast, atolls were subdivided into young and sunken sub-types, as their morphology could potentially be misinterpreted as calderas. This distinction was therefore considered necessary. The features that could not be confidently classified were categorized as *undefined* and not considered further. Detailed descriptions and reasoning of classification for types and sub-types are reported in Table 3.

For the PCA, we only used the *positive* dataset, because the parameters obtained from the CDA were referring to the feature itself rather than the depression, hence in the case of the *negative* dataset, the parameters referred to the mound, therefore not useable for the classification of the depression. The PCA was conducted separately for *types* (first-order classification) and *sub-types* (second-order classification), with 75% confidence ellipses, to verify its efficacy and to identify clusters that otherwise would not be distinct. The parameters used for the PCA are *diameter*, *ring mean*, *center mean*, *inner mean*, *outer mean*, and different ratios between the ring mean and the other water depth parameters, to understand the slope relationships between the inside and outside of the depression respect to the ring (*center mean/ring mean*; *inner mean/ring mean*; *outer ring/ring mean*).

Table 3| Type and sub-type descriptions and classification methods.

Type	Sub-type	Description	Support
Atoll	*Young atoll*	Emerged or very shallow atoll (i.e. <100 m below sea level)	Satellite imagery and/or coral reef database[56]
	Sunken atoll	Deep sea old atoll (> 100 m below sea level) with gentle gradients	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling (SF3)

		between center and edges	
Flank scar/eroded surface	*Flank scar/eroded surface*	Valley between volcanic ridges, or eroded volcanic surface or edifices	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling (SF3)
Crater/small caldera	*Full crater/small caldera*	Enclosed volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
	Horse-shoe/eroded crater / small caldera	Open/eroded volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
Caldera	*Full caldera*	Enclosed volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
	Horse-shoe/eroded caldera	Open/eroded volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
Undefined	*Undefined*	Not classifiable (poor bathymetry or unclear structure)	N/A

Abbreviation: SF= Supplementary File

Limitations

~~Our approach presents~~ several limitations. The principal component analysis (PCA) was conducted exclusively on the *positive* dataset, which comprises 186 of the 514 total features, as opposed to the *negative* dataset's 328 features. This imbalance reduces the statistical robustness of the analysis. The use of GEBCO global low-resolution bathymetry further constrains detection capabilities, resulting in a high number of undefined features (229 of 514), some of which may actually represent calderas. Finally, several previously documented calderas were not detected by the CDA for one of the following reasons: (1) their locations lay outside the cropped DEM used as input (e.g. Kikai, Kolumbo, Santorini, Krakatau); (2) they lacked sub-circular morphologies suitable for algorithmic detection (e.g. Axial); or (3) they were initially identified but discarded during filtering due to low confidence scores (e.g. Hunga Tonga and Havre), reflecting certain limitations of the method. Notwithstanding these exclusions, the primary goal of this study was to identify previously undocumented submarine calderas. Collectively, these limitations highlight the fact that those we provided represent a minimum number of potential submarine calderas, rather than an exhaustive inventory.

Data and Code Availability

The modified CDA used for this work can be found at https://github.com/eelsirhc/deepcaldera_project. The unfiltered and filtered datasets can be found in the repository of the Nanyang Technological University: https://researchdata.ntu.edu.sg/privateurl.xhtml?token=643346d9-b122-49cc-9db6-ad1072a9958f.

References

1. Petraglia, M. D., Korisettar, R. & Pal, J. N. The Toba volcanic super-eruption of 74,000 years ago: Climate change, environments, and evolving humans. *Quaternary International* **258**, 1–4 (2012).
2. Michon, L., Massin, F., Famin, V., Ferrazzini, V. & Roult, G. Basaltic calderas: Collapse dynamics, edifice deformation, and variations of magma withdrawal. *J. Geophys. Res.* **116**, (2011).
3. Bond, A. & Sparks, R. S. J. The Minoan eruption of Santorini, Greece. *JGS* **132**, 1–16 (1976).
4. Oppenheimer, C. Climatic, environmental and human consequences of the largest known historic eruption: Tambora volcano (Indonesia) 1815. *Progress in Physical Geography: Earth and Environment* **27**, 230–259 (2003).

5. Newhall, C. G. & Self, S. The volcanic explosivity index (VEI) an estimate of explosive magnitude for historical volcanism. *J. Geophys. Res.* **87**, 1231–1238 (1982).
6. Wirakusumah, A. D. & Rachmat, H. Impact of the 1815 Tambora Eruption to global climate change. *IOP Conf. Ser.: Earth Environ. Sci.* **71**, 012007 (2017).
7. Raible, C. C. *et al.* Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. *WIREs Climate Change* **7**, 569–589 (2016).
8. Bennington, N. *et al.* The progression of basaltic–rhyolitic melt storage at Yellowstone Caldera. *Nature* **637**, 97–102 (2025).
9. Costa, A. *et al.* Tephra fallout hazard assessment at the Campi Flegrei caldera (Italy). *Bull Volcanol* **71**, 259–273 (2009).
10. Bevilacqua, A. *et al.* Accelerating upper crustal deformation and seismicity of Campi Flegrei caldera (Italy), during the 2000–2023 unrest. *Commun Earth Environ* **5**, 742 (2024).
11. White, J. D. L., Schipper, C. I. & Kano, K. Submarine Explosive Eruptions. in *The Encyclopedia of Volcanoes* 553–569 (Elsevier, 2015). doi:10.1016/B978-0-12-385938-9.00031-6.
12. Cronin, S. *et al. Extreme Explosivity of the 15 January 2022 Hunga Eruption, Tonga, Driven by Magma-Mixing, Caldera Collapse and Magma-Water Interaction*. https://hdl.handle.net/10523/43599 (2023).
13. Chacón-Barrantes, S., Rivera-Cerdas, F. & Murillo-Gutiérrez, A. Impact of the tsunami caused by the Hunga Tonga–Hunga Ha'apai eruption in Costa Rica on 15 January 2022. *Bull Volcanol* **85**, 36 (2023).
14. Dogan, G. G., Yalciner, A. C., Annunziato, A., Yalciner, B. & Necmioglu, O. Global propagation of air pressure waves and consequent ocean waves due to the January 2022 Hunga Tonga-Hunga Ha'apai eruption. *Ocean Engineering* **267**, 113174 (2023).
15. Lynett, P. *et al.* Diverse tsunamigenesis triggered by the Hunga Tonga-Hunga Ha'apai eruption. *Nature* **609**, 728–733 (2022).
16. The World Bank. *The January 15, 2022 Hunga Tonga-Hunga Ha'apai Eruption and Tsunami, Tonga Global Rapid Post Disaster Damage Estimation (Grade) Report. Https://Thedocs.Worldbank.Org/En/Doc/B69af83e486aa652d4232276ad698c7b-0070062022/Original/GRADE-Report-Tonga-Volcanic-Eruption.Pdf*. (2022).
17. Global Volcanism Program & Venzke, E. Volcanoes of the World, v.5.2.0. Global Volcanism Program (2024).
18. Geyer, A. & Martí, J. The new worldwide collapse caldera database (CCDB): A tool for studying and understanding caldera processes. *Journal of Volcanology and Geothermal Research* **175**, 334–354 (2008).
19. Gevorgian, J., Sandwell, D. T., Yu, Y., Kim, S. & Wessel, P. Global Distribution and Morphology of Small Seamounts. *Earth and Space Science* **10**, e2022EA002331 (2023).
20. Verolino, A., Wee, S. F., Jenkins, S. F., Costa, F. & Switzer, A. D. SEATANI: hazards from seamounts in Southeast Asia, Taiwan, and Andaman and Nicobar Islands (eastern India). *Nat. Hazards Earth Syst. Sci.* **24**, 1203–1222 (2024).
21. Acocella, V. Calderas. in *Volcano-Tectonic Processes* 163–203 (Springer International Publishing, Cham, 2021). doi:10.1007/978-3-030-65968-4_5.
22. Branney, M. & Acocella, V. Calderas. In The encyclopedia of volcanoes (pp. 299-315). Academic Press. https://doi.org/10.1016/B978-0-12-385938-9.00016-X. in (2015).
23. Wang, W. *et al.* Volatile Element Evidence of Local MORB Mantle Heterogeneity Beneath the Southwest Indian Ridge, 48°–51°E. *Geochem Geophys Geosyst* **22**, e2021GC009647 (2021).
24. Poland, M. P., Takahashi, T. J. & Landowski, C. M. *Characteristics of Hawaiian Volcanoes*. https://pubs.usgs.gov/publication/pp1801 (2014) doi:10.3133/pp1801.
25. Martí, J., Mitjavila, J. & Araña, V. Stratigraphy, structure and geochronology of the Las Cañadas caldera (Tenerife, Canary Islands). *Geol. Mag.* **131**, 715–727 (1994).
26. Geist, D. J. *et al.* Wolf Volcano, Galápagos Archipelago: Melting and Magmatic Evolution at the Margins of a Mantle Plume. *Journal of Petrology* **46**, 2197–2224 (2005).
27. Cocchi, L., Caratori Tontini, F., Muccini, F. & De Ronde, C. E. J. Magnetic Expression of Hydrothermal Systems Hosted by Submarine Calderas in Subduction Settings: Examples from the Palinuro and Brothers Volcanoes. *Geosciences* **11**, 504 (2021).
28. Zimanowski, B., Büttner, R., Lorenz, V. & Häfele, H.-G. Fragmentation of basaltic melt in the course of explosive volcanism. *Journal of Geophysical Research: Solid Earth* **102**, 803–814 (1997).
29. Verolino, A., White, J. D. L. & Brenna, M. Eruption dynamics at Pahvant Butte volcano, Utah, western USA: insights from ash-sheet dispersal, grain size, and geochemical data. *Bull Volcanol* **80**, 1–18 (2018).

30. Verolino, A., White, J. D. L., Dürig, T. & Cappuccio, F. Black Point – Pyroclasts of a Surtseyan eruption show no change during edifice growth to the surface from 100 m water depth. *Journal of Volcanology and Geothermal Research* **384**, 85–102 (2019).
31. Dürig, T. *et al.* Deep-sea eruptions boosted by induced fuel–coolant explosions. *Nat. Geosci.* **13**, 498–503 (2020).
32. Verolino, A., White, J. D. L., Baxter, R. J. M., Schipper, C. I. & Thordarson, T. Characteristics of Sub-Aerially Emplaced Pyroclasts in the Surtsey Eruption Deposits: Implications for Diverse Surtseyan Eruptive Styles. *Geosciences* **12**, 79 (2022).
33. Verolino, A. *et al.* Assessing volcanic hazard and exposure to lava flows at remote volcanic fields: a case study from the Bolaven Volcanic Field, Laos. *J Appl. Volcanol.* **11**, 6 (2022).
34. Murch, A. P., Portner, R. A., Rubin, K. H. & Clague, D. A. Deep-subaqueous implosive volcanism at West Mata seamount, Tonga. *Earth and Planetary Science Letters* **578**, 117328 (2022).
35. Lee, C. & Hogan, J. Automated crater detection with human level performance. *Computers & Geosciences* **147**, 104645 (2021).
36. GEBCO Compilation Group. GEBCO 2023 Grid. https://doi.org/doi:10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b. (2023).
37. MBBDB. National Centers for Environmental Information. 2004: Multibeam Bathymetry Database (MBBDB). NOAA National Centers for Environmental Information. Accessed [April 2025]. https://doi.org/doi:10.7289/V56T0JNC.
38. Falkenberg, J. J. *et al.* Spatial Variations in Magmatic Volatile Influx and Fluid Boiling in the Submarine Hydrothermal Systems of Niuatahi Caldera, Tonga Rear-Arc. *Geochem Geophys Geosyst* **23**, e2021GC010259 (2022).
39. Leat, P. T. *et al.* Bathymetry and geological setting of the South Sandwich Islands volcanic arc. *Antarctic Science* **28**, 293–303 (2016).
40. Staudigel, H. & Clague, D. The Geological History of Deep-Sea Volcanoes: Biosphere, Hydrosphere, and Lithosphere Interactions. *Oceanog.* **23**, 58–71 (2010).
41. Danovaro, R. *et al.* A submarine volcanic eruption leads to a novel microbial habitat. *Nat Ecol Evol* **1**, (2017).
42. Carey, R. J., Butterfield, D. A. & Clark, M. R. Submarine Volcanic Eruptions and Their Impacts on Hydrothermal Systems and Biological Communities. *Annual Review of Earth and Planetary Sciences* **53**, 397–423 (2025).
43. Fouquet, Y. *et al.* Volcanic and hydrothermal processes in submarine calderas: The Kulo Lasi example (SW Pacific). *Ore Geology Reviews* **99**, 314–343 (2018).
44. Omira, R. *et al.* Global Tonga tsunami explained by a fast-moving atmospheric source. *Nature* **609**, 734–740 (2022).
45. Ren, Z., Higuera, P. & Li-Fan Liu, P. On Tsunami Waves Induced by Atmospheric Pressure Shock Waves After the 2022 Hunga Tonga-Hunga Ha'apai Volcano Eruption. *JGR Oceans* **128**, (2023).
46. Verolino, A. *et al.* Forced and free waves of simulated volcanic meteo-tsunamis in the South China Sea. *npj Nat. Hazards* **2**, 25 (2025).
47. Self, S. & Rampino, M. R. The 1883 eruption of Krakatau. *Nature* **294**, 699–704 (1981).
48. Rampino, M. R. & Self, S. Historic Eruptions of Tambora (1815), Krakatau (1883), and Agung (1963), their Stratospheric Aerosols, and Climatic Impact. *Quat. res.* **18**, 127–143 (1982).
49. Clare, M. A. *et al.* Volcanic eruptions and the global subsea telecommunications network. *Bull Volcanol* **87**, 51 (2025).
50. Clare, M. A. *et al.* Fast and destructive density currents created by ocean-entering volcanic eruptions. *Science* **381**, 1085–1092 (2023).
51. Sparks, R. S. J. *et al.* Large silicic magma bodies and very large magnitude explosive eruptions. *Bull Volcanol* **84**, (2022).
52. Watt, S. F. L. *et al.* Combinations of volcanic-flank and seafloor-sediment failure offshore Montserrat, and their implications for tsunami generation. *Earth and Planetary Science Letters* **319–320**, 228–240 (2012).
53. Preine, J. *et al.* Hazardous explosive eruptions of a recharging multi-cyclic island arc caldera. *Nat. Geosci.* **17**, 323–331 (2024).
54. Wilcock, W. S. D. *et al.* Seismic constraints on caldera dynamics from the 2015 Axial Seamount eruption. *Science* **354**, 1395–1399 (2016).
55. Nomikou, P. *et al.* SANTORY: SANTORini's Seafloor Volcanic ObservatorY. *Front. Mar. Sci.* **9**, 796376 (2022).
56. UNEP-WCMC. Global Coral Reef Database: WCMC008_CoralReef2018_Py_v4_1. UN Environment Programme World Conservation Monitoring Centre, Cambridge, UK (2018). (2018).

Acknowledgements

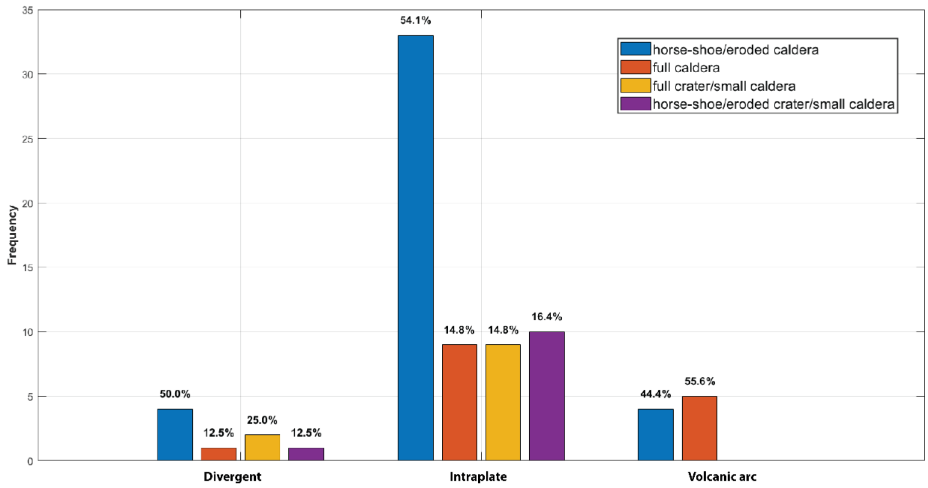
This project was supported by the Ministry of Education, Singapore, under its MOE Academic Research Fund Tier 3 InVEST project (Award MOE-MOET32021-0002). This work comprises EOS contribution number 661.

Author Contributions

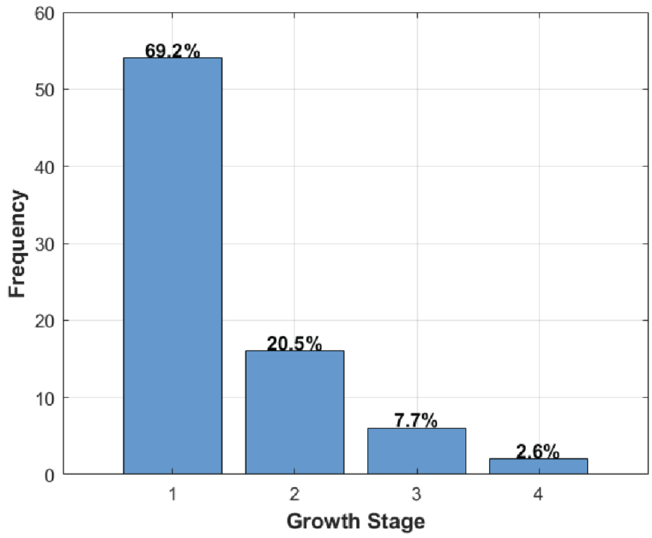
AV: work idealization, paper drafting, editing; CL: assistance with the Caldera Detection Algorithm, editing; SFJ: work idealization, editing; MJ: editing; ADS: editing

Competing Interests

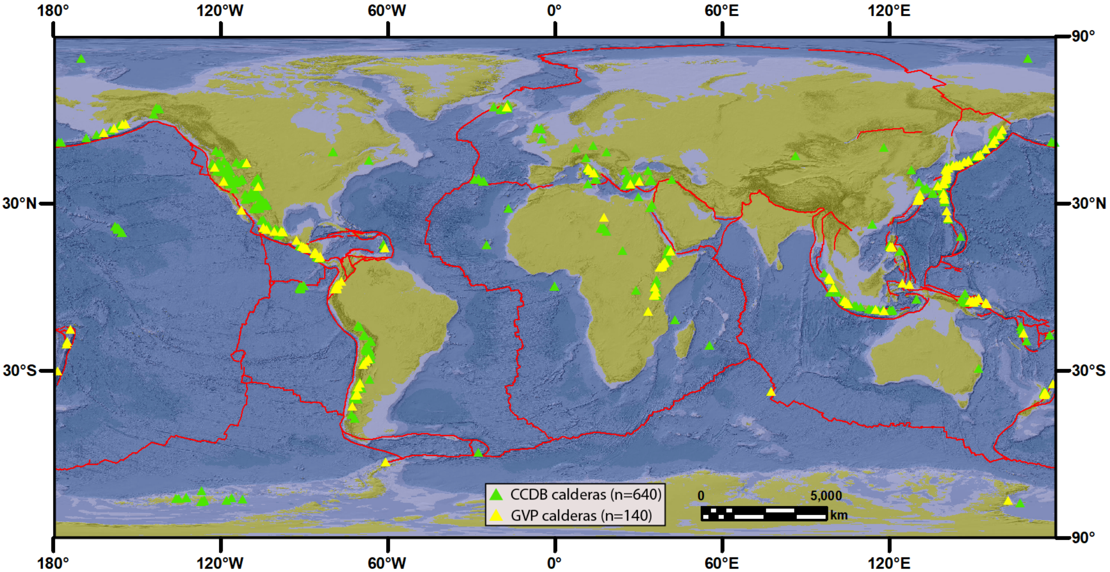
We declare that co-author A.D. Switzer is a member of the editorial board for Communications Earth & Environment (Ocean and Cryosphere).

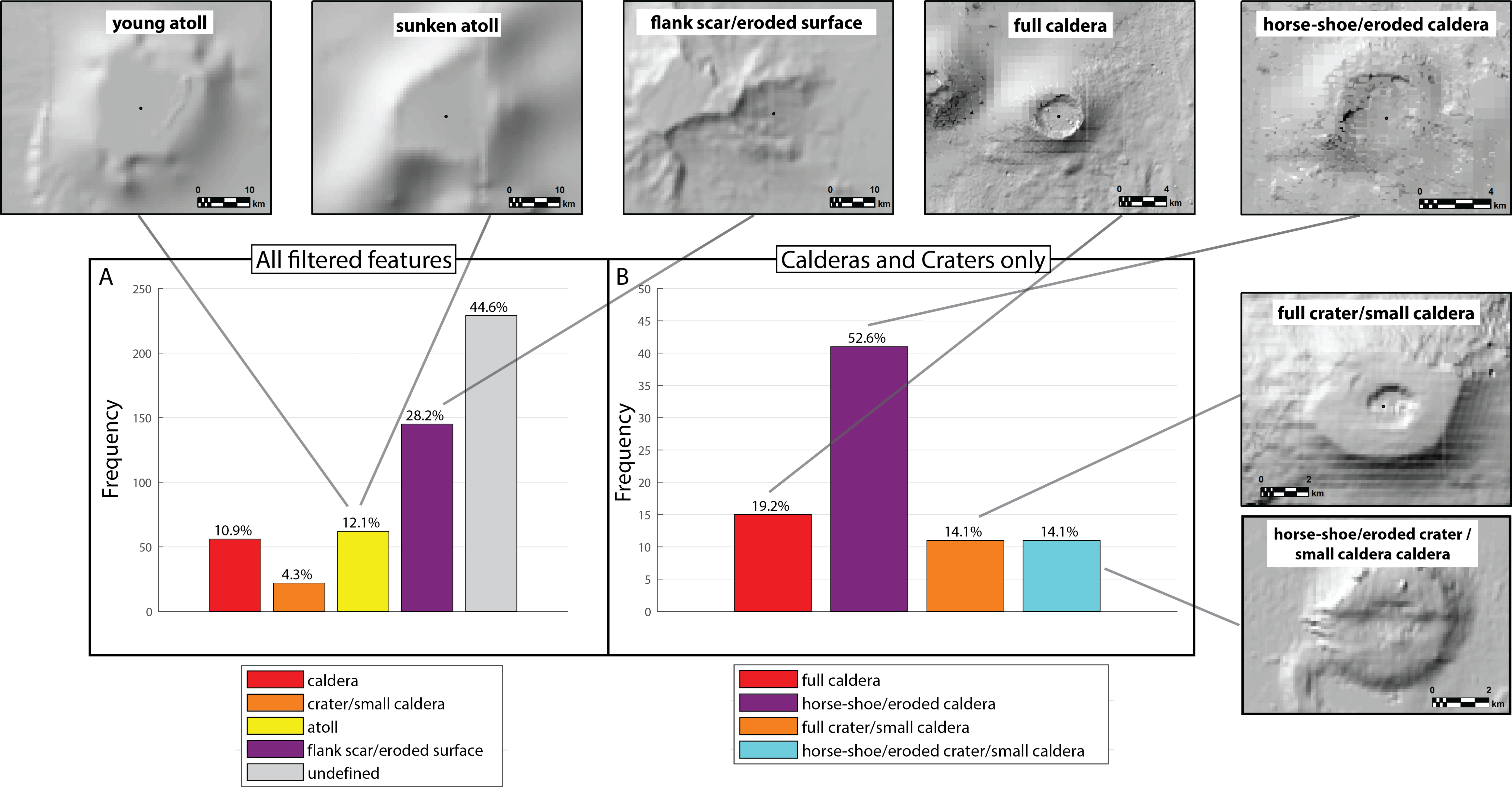


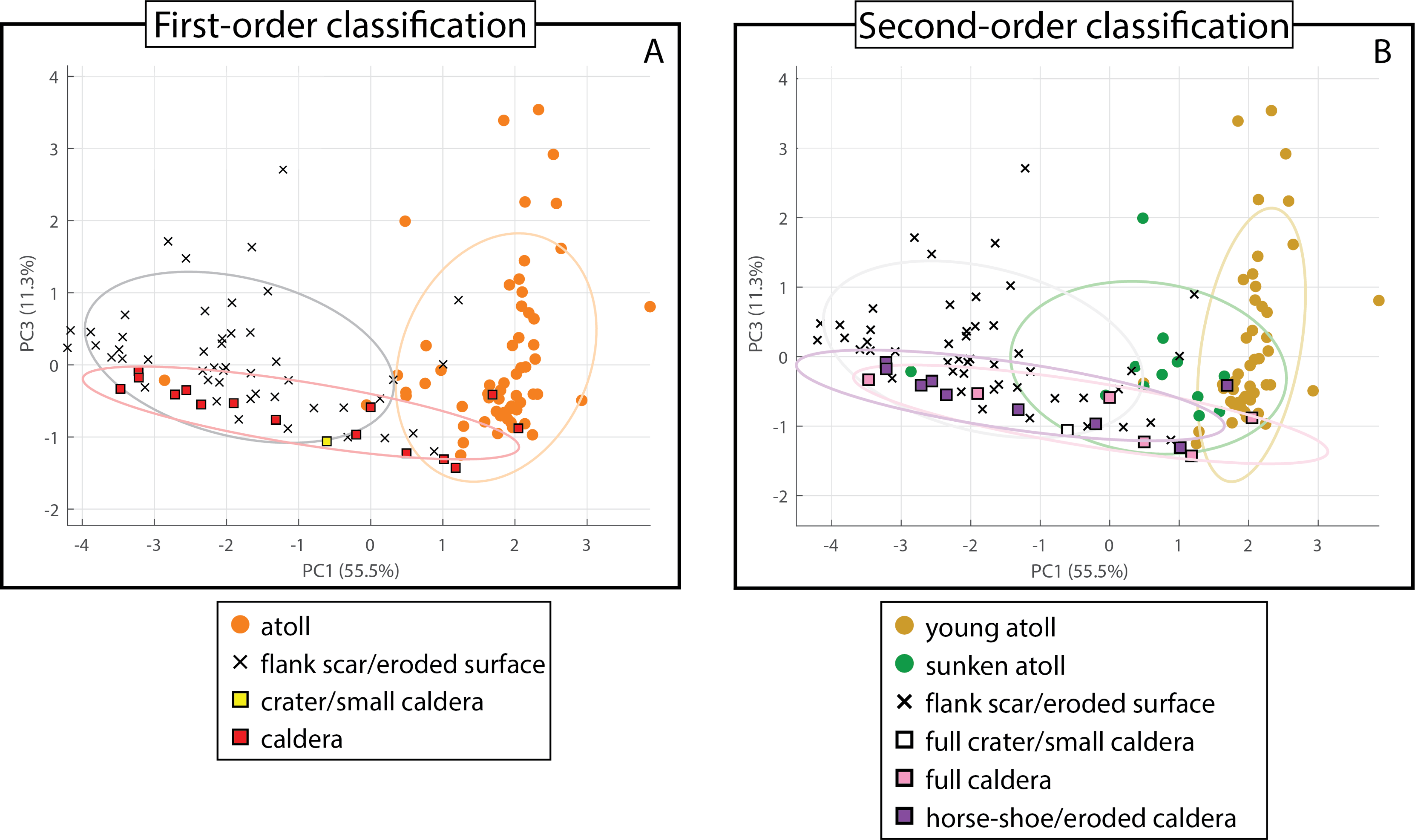
Extended Data Fig. 1| Distribution of caldera sub-types across tectonic settings.



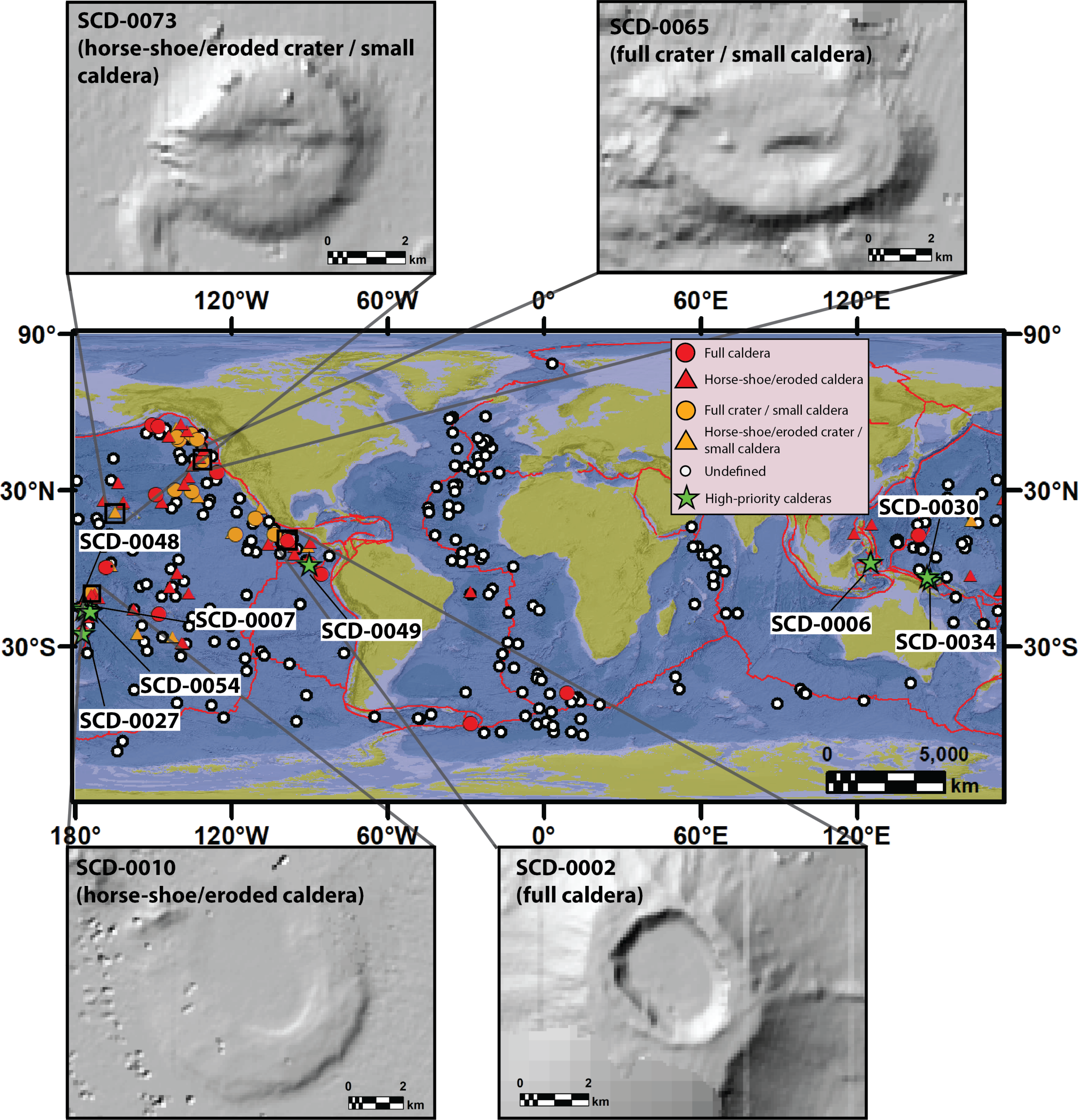
Extended Data Fig. 2| Distribution of caldera growth stages.


180°
120°W
60°W
0°
60°E
120°E
90°
30°N
30°S
90°
CCDB calderas (n=640)
GVP calderas (n=140)
0
5,000
km

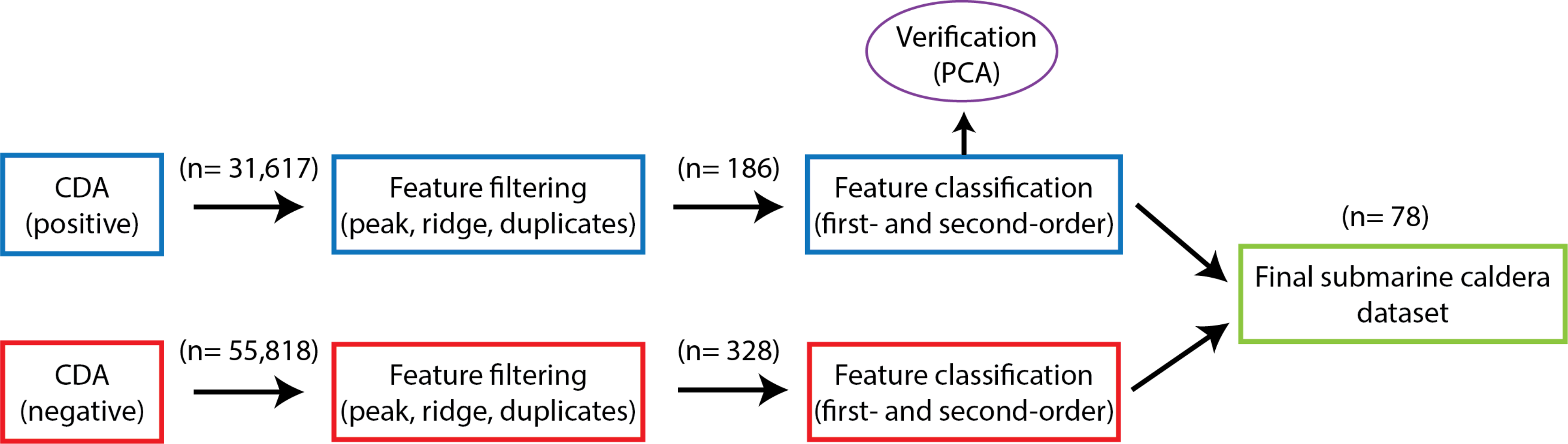

young atoll
sunken atoll
flank scar/eroded surface
full caldera
horse-shoe/eroded caldera
full crater/small caldera
horse-shoe/eroded crater / small caldera caldera
km
All filtered features
A
Frequency
10.9%
4.3%
12.1%
28.2%
44.6%
Calderas and Craters only
B
Frequency
19.2%
52.6%
14.1%
14.1%
caldera
crater/small caldera
atoll
flank scar/eroded surface
undefined
full caldera
horse-shoe/eroded caldera
full crater/small caldera
horse-shoe/eroded crater/small caldera



First-order classification
A
PC3 (11.3%)
PC1 (55.5%)
atoll
flank scar/eroded surface
crater/small caldera
caldera
Second-order classification
B
PC3 (11.3%)
PC1 (55.5%)
young atoll
sunken atoll
flank scar/eroded surface
full crater/small caldera
full caldera
horse-shoe/eroded caldera



SCD-0073
(horse-shoe/eroded crater / small caldera)
0
2
km
SCD-0065
(full crater / small caldera)
0
2
km
120°W
60°W
0°
60°E
120°E
90°
90°
30°N
30°N
30°S
30°S
Full caldera
Horse-shoe/eroded caldera
Full crater / small caldera
Horse-shoe/eroded crater / small caldera
Undefined
High-priority calderas
SCD-0048
SCD-0007
SCD-0049
SCD-0054
SCD-0027
SCD-0030
SCD-0006
SCD-0034
0
5,000
km
180°
120°W
60°W
0°
60°E
120°E
SCD-0010
(horse-shoe/eroded caldera)
0
2
km
SCD-0002
(full caldera)
0
2
km



Verification
(PCA)
CDA
(positive)
(n= 31,617)
Feature filtering
(peak, ridge, duplicates)
(n= 186)
Feature classification
(first- and second-order)
(n= 78)
Final submarine caldera
dataset
CDA
(negative)
(n= 55,818)
Feature filtering
(peak, ridge, duplicates)
(n= 328)
Feature classification
(first- and second-order)

Table. 1| List of high-priority calderas. They were selected based on their growth stage (i.e. water depth), tectonic setting and morphological sub-type.

Caldera name	Long	Lat	Caldera sub-type
SCD-0007	-175.64	-15.60	full caldera
SCD-0006	125.44	2.70	horse-shoe/eroded caldera
SCD-0048	-176.60	-15.10	full caldera
SCD-0027	-176.91	-24.59	full caldera
SCD-0054	-173.90	-16.22	horse-shoe/eroded caldera
SCD-0034	147.16	-2.97	horse-shoe/eroded caldera
SCD-0049	-90.01	1.78	full caldera
SCD-0030	148.23	-4.10	horse-shoe/eroded caldera

wth stage (water de	Tectonic setting
4 (~0 m)	Volcanic arc (Tonga arc)
4 (~0 m)	Volcanic arc (Sangihe arc)
3 (~100 m)	Volcanic arc (Tonga arc)
3 (~500 m)	Volcanic arc (Tonga arc)
3 (~400 m)	Volcanic arc (Tonga arc)
3 (~600 m)	Divergent (Manus spreading ridge)
3 (~700 m)	Divergent (Cocos-Nazca spreading centre)
3 (~700 m)	Intraplate (South Bismarck plate)

Table. 2| Output parameters of the CDA. Note that to remove as many false positive as possible, we set Duplicates>1, Peak=false and Ridge=true as filters.

PARAMETER	DESCRIPTION
Diameter	Mean diameter of the feature
Ring mean	Mean water depth of the feature ring
Centre mean	Mean water depth measured within a narrow ring concentric to the feature center, positioned at 10% of the caldera radius, to represent near-central bathymetric conditions
Inner mean	Mean water depth measured within a concentric ring located at 75% of the feature radius, to represent intermediate bathymetric conditions
Outer mean	Mean water depth measured within a concentric ring located at 125% of the caldera radius, characterizing the bathymetric conditions just beyond the caldera rim.
Duplicates	Number of times a feature is detected within the 20 images (1-20)
Peak	If P=true, centre mean > Inner mean; if P=false, centre mean < Inner mean
Ridge	If P=true, Inner mean > Outer mean; if P=false, Inner mean < Outer mean

Table 3| Type and sub-type descriptions and classification methods.

Type	Sub-type
Atoll	*Young atoll*
	Sunk atoll
Flank scar/eroded surface	*Flank scar/eroded surface*
Crater/small caldera	*Full crater/small caldera*
	horse-shoe/eroded crater / small caldera
Caldera	*Full caldera*
	Horse-shoe/eroded caldera
Undefined	*Undefined*

Description	Support
Emerged or shallow atoll	Satellite imagery and/or coral reef database[38]
Deep sea old atoll with gentle gradients between centre and edges	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling
Valley between volcanic ridges, or eroded volcanic surface	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling
Enclosed volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Open/eroded volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Enclosed volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Open/eroded volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Not classifiable (poor bathymetry or unclear structure)	N/A

Global machine-learning detection of submarine calderas

Andrea Verolino[1*], Christopher Lee[2], Susanna F. Jenkins[1,3], Martin Jutzeler[4], Adam D. Switzer[1,3]

[1] Earth Observatory of Singapore, Nanyang Technological University, Singapore
[2] Department of Physics, University of Toronto, Canada
[3] Asian School of the Environment, Nanyang Technological University, Singapore
[4] Centre for Ore Deposit and Earth Sciences (CODES) and School of Natural Sciences, University of Tasmania, Australia

* Corresponding author (email address: andrea_verolino@hotmail.it)

Abstract

Submarine calderas are among the least documented volcanic structures on Earth, yet recent impactful events highlight their potential for severe geohazard consequences globally, including tsunamis, seafloor damage, and atmospheric impacts. Logistical limits have historically hindered global detection. Here we apply a machine-learning-based caldera detection algorithm to global bathymetry, enabling systematic identification of previously undetected submarine calderas. We found 78 calderas spanning different water depths (up to 5,600 m), diameters (up to 20 km) and tectonic settings (divergent, convergent, intraplate). A subset of eight shallow-water calderas, primarily within volcanic arcs, are identified as high-priority targets due to their high hazard potential. Our dataset fills a critical observational gap and offers a reproducible and upgradeable framework for submarine volcano characterization, geohazard assessment, and deep-sea exploration. These findings underscore the need to integrate submarine calderas into future global hazard modelling and monitoring strategies.

Introduction

Calderas are kilometer-scale volcanic depressions that result from the voluminous eruption of magma from the Earth's crust. Caldera-forming eruptions produce very large amounts of material and are capable of re-shaping landscapes and influencing global climate[1]; they can form from passive effusions of magma[2] or high-intensity explosive eruptions[3]. Explosive caldera-forming events have triggered periods of abrupt surface cooling due to massive ash and gas emissions into the atmosphere[4]. For instance, the Toba eruption (~74,000 years ago) is considered one of the largest volcanic events in the last two million years, releasing approximately 2,800 km³ of magma and potentially inducing a global volcanic winter[1]. Similarly, the Tambora eruption of 1815, one of the most powerful events recorded (Volcanic Explosivity Index= 7, following Newhall and Self[5]), injected 60-80 megatons of SO_2 into the stratosphere, leading to the infamous "Year Without a Summer", which caused widespread crop failures and famine across Europe and North America[6,7]. Beyond their catastrophic past eruptions, many calderas remain active, with evidence of ongoing seismicity, ground deformation, and magmatic movement. For instance Yellowstone Caldera (USA), continues to exhibit signs of deep reservoir activity[8], while Campi Flegrei (Italy) has shown increasing seismic unrest since 2005, suggesting evolving magmatic processes[9,10]. To identify and better understand the formation of calderas is therefore crucial for reconstructing past eruptive histories and assessing potential future hazards.

The vast majority of research into calderas focuses on subaerial structures, and a considerable knowledge gap remains regarding their submarine counterparts. Although the majority of global volcanism occurs beneath the ocean[11], submarine calderas remain largely unexplored and poorly documented. Submarine volcanoes are critical to study because they play a key role in understanding geological hazards, marine ecosystems, global cycles, and resource potentials. Major eruptions from submarine calderas, such as the 2022 Hunga Tonga event, in the Kingdom of Tonga, have recently demonstrated their potential for extreme explosivity and global impact[12–14]. The eruption triggered far-reaching tsunamis (>10,000 km), with fatalities as far away as Peru[15], and resulted in local economic losses equivalent to approximately 18% of Tonga's GDP[16].

Existing global caldera databases highlight the imbalance between subaerial and submarine records. For example, the Global Volcanism Program (GVP) identifies 140 Pleistocene-Holocene calderas, of which 94% (n= 132) are subaerial and only 6% (n= 8) submarine[17]. Whereas, the Collapse Caldera Database (CCDB) lists 640 mapped calderas[18], yet reports fewer than 2% as submarine (visually assessed as water depth information is missing from the database) (Fig. 1). Meanwhile, recent discoveries of over 43,000 seamounts[19] suggest that many unidentified submarine calderas likely exist, and some are almost certainly volcanically active[20].

Tectonic setting exerts a primary control on submarine volcanic activity, with most of Earth's active volcanism concentrated along plate boundaries, either at divergent margins (spreading ridges) or convergent zones (subduction arcs)[21]. Calderas in continental convergent settings typically exhibit more silicic compositions, owing to thicker crust and longer magma residence times, which favor volatile enrichment and high viscosity. These conditions, in turn, enhance explosive potential compared to basaltic calderas[22]. However, such characteristics are less commonly reached in convergent oceanic regions, where the crust is relatively thin. Nevertheless, silicic oceanic arc calderas do exist and are among the most powerful[12]. Calderas also occur at spreading ridges, though they are typically basaltic and associated with fast magma replenishment and limited volatile content, conditions that generally suppress explosive behavior. However, even ridge-associated systems can display sudden energy release, particularly when volatile-rich magma ascends from locally heterogenous mantle domains[23]. Intraplate volcanism, similarly, tend to produce basaltic calderas (e.g. Hawaii[24]), but can also yield a broad compositional spectrum, from mafic to highly silicic examples such as the Canary Islands[25] and Galápagos[26]. In general, intraplate submarine volcanism is dominated by seamounts of mafic compositions, the majority of which are ancient and likely inactive, posing minimal risk. Regardless of composition, explosive submarine calderas can develop across all tectonic settings[27], and intense magma-water interaction, easier to reach for low-viscosity magmas (i.e. basalts)[28], can substantially amplify explosive behavior[12,29–34].

In this study, we applied a machine-learning-based caldera detection algorithm (CDA) to systematically identify submarine calderas worldwide. Our CDA builds upon the algorithm originally developed by Lee and Hogan[35] for detecting Martian impact craters, which demonstrated high detection precision (>80%) for sub-circular depressions ⩾4 km in diameter with minimal training. We implemented the CDA using a series of morphometric and resolution filters to effectively detect submarine calderas within a low-resolution bathymetric digital elevation model (GEBCO 2023 global bathymetry dataset[36], ~500 m resolution). These filters helped reduce the number of false positives and enabled manual classification of the retained features, followed by Principal Component Analysis (PCA) to assess and validate the nature of the identified structures (see Methods for details).

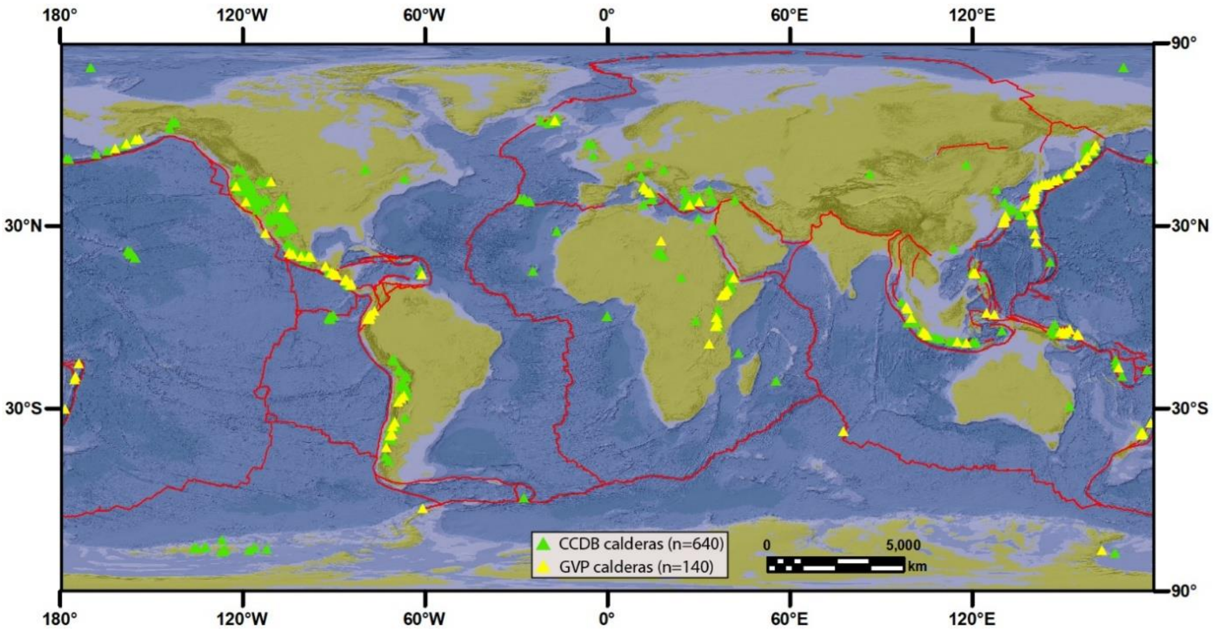


Fig. 1| Global distribution of documented calderas. Caldera locations from the Global Volcanism Program (GVP) (yellow triangles) and from the Collapse Caldera Database (CCDB) (green triangles). Note the vast lack of calderas within ocean basins. The red lines are tectonic plate boundaries. The basemap is from GEBCO 2023[36].

In addition to evaluating the effectiveness of the CDA in the semi-automated identification of previously undocumented submarine calderas, this study primarily aimed to establish a global and upgradable dataset to inform future targeted research, including for example the deployment of monitoring instruments and targeted remote sensing observations at key calderas. Our dataset comprises 78 volcanic depressions larger than 1 km in diameter, which we classified as either calderas (diameter >3 km) or volcanic craters (diameter <3 km). Based on global consensus[22], calderas are broad topographic depressions formed by subvertical subsidence into a partially drained magma reservoir, typically associated with large eruptive events, whereas craters are generally smaller, bowl- or funnel-shaped features produced by explosive fragmentation and ejection of host rock near volcanic vents. Importantly, these definitions lack a definitive size threshold for distinguishing calderas from craters, and notable databases such as the CCDB list calderas even smaller than 1 km, creating potential overlap with crater dimensions. Given this ambiguity, compounded by the limited resolution of global bathymetric datasets and the tendency for both feature types to share comparable size ranges in submarine settings, we adopt an inclusive classification approach in this study, with both calderas and craters grouped under the 'Submarine Calderas Dataset'. This designation is particularly suited to submarine environments where direct observation remains limited.

Out of the 78 detected volcanic depressions, we propose an initial set of 8 submarine calderas (Table 1) as high-priority targets for scientific investigation, selected based on their water depths, tectonic settings, and caldera morphological type, factors that collectively suggest an elevated likelihood of future eruptive activity and potential societal impact.

This study lays a critical foundation for evaluating submarine volcanic risks and improving global preparedness. By applying the CDA, we establish a new global dataset of submarine calderas, advancing our understanding of explosive subaqueous volcanism and guiding the prioritization of regions where volcanic hazards may be most consequential. The urgency of this endeavor is underscored by recent high-impact events, such as the 2022 Hunga Tonga eruption, and the increasing vulnerability of seafloor and coastal infrastructure. The proposed list of high-priority calderas represents a pivotal starting point for targeted deep-sea exploration and integrated geohazard assessment.

Results

Feature classification and validation through Principal Component Analysis (PCA)

The initial CDA output of 87,435 potential features was narrowed to 514 candidates (Supplementary File 1) based on morphometric criteria and reliability indicators generated by the algorithm; this was accomplished to reduce the initial high number of false positives. These filtered features were then manually categorized through a first-order and second-order classification (Fig. 2), further distinguishing calderas from non-calderas. Details of the workflow, filtering methodology, and classification procedure are provided in the Methods section. Specifically, from the first-order identification (Fig. 2A) we found 56 *calderas*, 22 *craters or small calderas*, 62 *atolls*, 145 *flank scars or erosion surfaces*, and 229 remained *undefined* (i.e. not classifiable due to ~~poor~~ bathymetry and/or unclear structure). Excluding *undefined* features, *calderas* and *craters/small calderas* account for 27% of the filtered dataset, *flank scars and erosion surfaces* 51%, and *atolls* 22%. The results of the second-order classification indicate that, of the caldera subset (Fig. 2B), *horse-shoe or eroded calderas* dominate (~53%), followed by *full calderas* (~19%) and *craters/small calderas*, which are evenly split between full craters/calderas and horse-shoe types (~14% each). The second-order classification exercise was performed to: (1) assess the efficacy of PCA as a supporting classification tool; and (2) better constrain eruptive and volcano degradation history, providing a preliminary hazard assessment based on edifice collapse (e.g. a horse-shoe caldera may indicate that it has already experienced large lateral collapses and/or underwent major erosional processes; hence, potentially less hazardous than a full caldera).

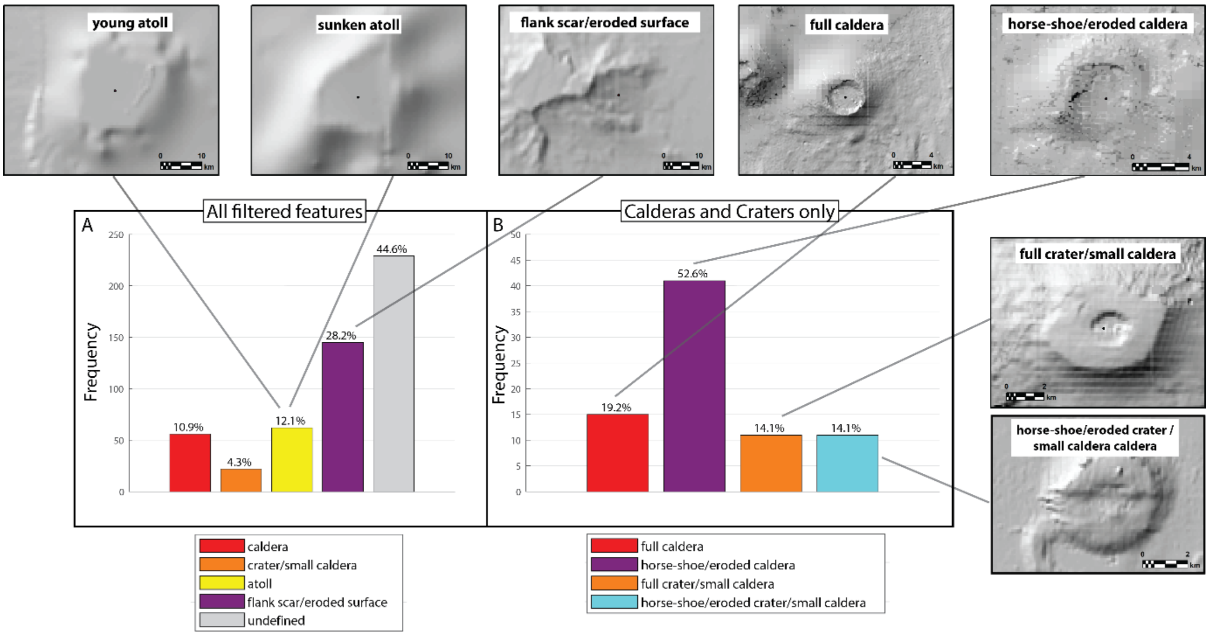


Fig. 2| Feature classification results. (A) First order classification for the filtered features. (B) Second-order classification for the calderas and craters only. Also reported are close-up images of each sub-type. Basemap for the calderas and craters examples is from the Multibeam bathymetry Database (MBBDB)[37].

The PCA analysis was conducted both for the feature types (first-order classification) and sub-types (second-order classification). The first-order classification PCA (Fig. 3A) is very efficient at distinguishing clusters that correspond to atolls, calderas, and flank scars/eroded surfaces. For the second-order classification, the PCA successfully differentiates clusters for the two atoll sub-types (young vs sunken; see Methods). Interestingly, full calderas and horse-shoe/eroded calderas largely overlap within the same field (Fig. 3B). Flank scars and eroded surfaces remain well defined, though a few outliers fall into other clusters. Such errors may partly be linked to locally low bathymetry resolution.

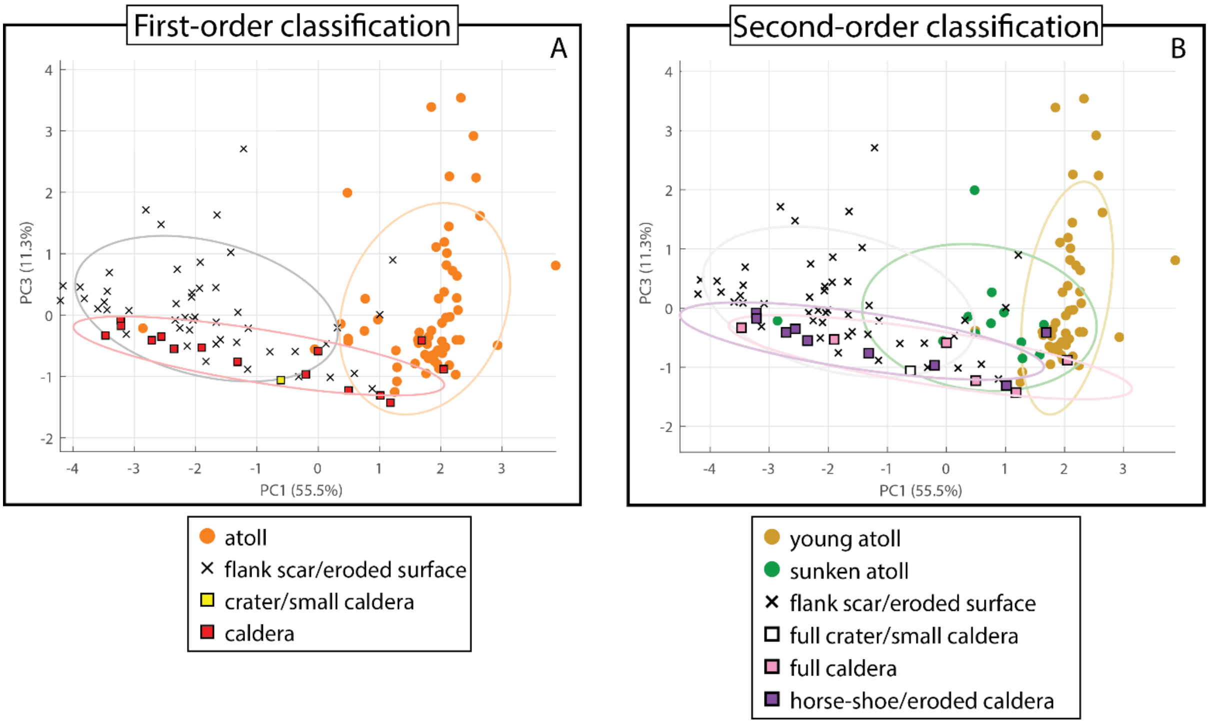


Fig. 3| Principal Component Analysis results. (A) PCA for first-order classification by type. (B) PCA for second-order classification by sub-type. The ellipses were obtained using a 75% confidence level, assuming a multivariate normal distribution, and colored based on the respective types and sub-types.

Submarine calderas dataset

Our CDA and feature classification identified 56 potential submarine calderas and 22 potential small calderas or craters, all grouped into a ‘submarine calderas’ dataset, totaling 78 sub-circular landforms with >1 km depression diameter (Fig. 4 and Supplementary File 2). Of the detected candidates, only two calderas were previously known, Niuatahi[38] and Kemp[39] calderas in Tonga and along the South Sandwich arc, respectively. Several other well documented calderas, including Hunga Tonga, Kikai, Krakatau, Santorini, Kolumbo and Axial were not identified by the CDA due to two primary factors, either because their location lay outside the DEM used, or because of technical limitations of the CDA itself (more details are reported in the Methods). Among the 78 potential calderas, 8 occur in divergent settings, 9 in volcanic arcs, and 61 in intraplate environments (Fig. 4 and Extended Data Fig. 1). In terms of water depths, the calderas we detected span between 5,600 m below sea level and 90 m above sea level (partially emerged). Water depth, together with volcano height, have been previously associated with volcanic hazard potential[20], through the volcano growth stage[40], with stages 3 and 4 being the more hazardous, and stages 1 and 2 the least hazardous[20]. Following this categorization, most calderas (n= 54) are classified as stage 1 (>700 m water depth, <1000 m height), followed by stage 2 (>700 m depth, >1000 m height) with 16 calderas, stage 3 (<700 m water depth) with 6 calderas, and stage 4 (partially emerged) with 2 calderas

(Extended Data Fig. 2).These calderas have a diameter between ~ 1 and 20 km, with the majority being in the range 1-6 km (Supplementary File 2).

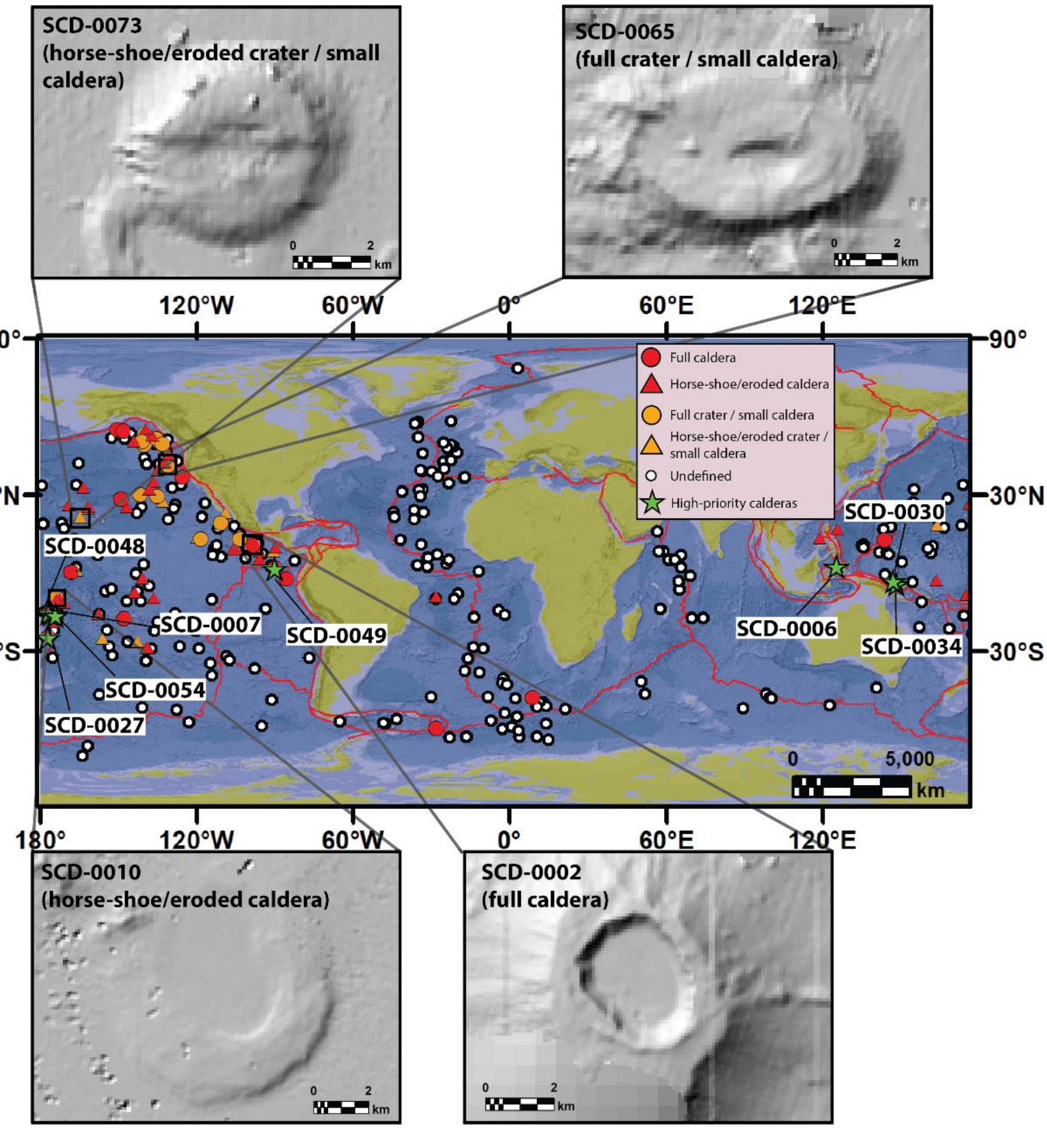


Fig. 4| Global map with the location of the classified sub-types of calderas and craters, and high-priority calderas. Undefined features are also reported (white circles), some of which may be calderas. The basemap is from GEBCO 2023[36]. The smaller panels show examples of sub-types of caldera and crater identified in this study. Basemap for the calderas and craters examples is the Multibeam bathymetry Database (MBBDB)[37].

Discussion

Submarine explosive eruptions remain poorly documented compared to their subaerial counterparts, yet recent events underscore their devastating potential[12,13,15]. They also host hydrothermal vent systems that support unique, chemosynthetic life, offering insights into early Earth conditions and ecosystem resilience in extreme environments[41,42]. Therefore, identifying submarine calderas is essential not only for hazard-focused research, but also for advancing understanding in fields such as deep-sea ecology[43].

In the context of risk, the 2022 Hunga Tonga caldera, submerged 200 m below sea level prior to eruption, demonstrated that submerged silicic calderas can produce high-energy explosive activity, initiate submarine density currents able to destroy critical underwater infrastructure, and generate tsunamis and atmospheric disturbances on a global scale[44–46]. This highlights the need for improved hazard assessments, particularly for shallow and partially emerged volcanic systems, where magma-water interactions intensify explosivity[29–32]. This is particularly the case for stage 3 (<700 m b.s.l.) and stage 4 (partially emerged) calderas[20]. Eruptions at these volcanoes can trigger pyroclastic density currents, inject into the atmosphere considerable amounts of ash and gas, and be accompanied by large-scale tsunamis, with direct consequences for nearby to global coastal and regional populations. A notable historical example is the 1883 eruption of Krakatau, Indonesia, where partially submerged explosive activity triggered catastrophic tsunami waves, with pyroclastic density currents entering the surrounding sea and playing a central role in the tsunamis generation[47]. The eruption also injected large volumes of ash and sulfur-rich aerosols into the atmosphere, causing measurable global cooling and disrupting atmospheric circulation patterns[48]. Submarine infrastructure, including communication cables and deep-sea pipelines, is particularly vulnerable to caldera-related activity[49]. The 2022 Hunga Tonga eruption severed Tonga's primary communication cable, isolating the nation for weeks[50]. Additionally, post-eruption earthquakes and submarine landslides continued to reshape the seafloor in the months following the eruption. These secondary events cut Tonga's domestic cable several months later, compounding the disruption to communication and emergency response[50]. Similarly, hydrothermal vent systems near active submarine calderas can induce cable degradation, affecting global data transmission networks[49].

In our classification, we distinguish full calderas from horse-shoe/eroded calderas. Their morphological features offer insights into their past and future eruptive behavior. Full calderas, with their enclosed circular structure, generally result from high-magnitude explosive eruptions, where rapid magma evacuation causes structural collapse[51]. In contrast, the breached structure of horse-shoe calderas may reflect past sector collapse, which may have generated large debris avalanches that could have triggered tsunamis[52]. Caldera collapse and subsequent constructional volcanism may occur cyclically, and the formation of one caldera does not preclude the possibility of recurrent caldera-forming or collapse events within the same volcanic system[53]. These morphologies can direct future eruptions along structural weaknesses, with magma migrating preferentially to open flanks. Alternatively, this morphological category may reflect extensive post-eruptive erosion, suggesting prolonged inactivity, hence less potentially hazardous. Understanding these morphological distinctions improves hazard assessments, particularly in submarine regions where direct observations remain limited.

Based on this analysis, we prioritize eight submarine calderas for future deep-sea exploration using multiple criteria (Table 1). First, we considered stage 3 (depth <700 m) and stage 4 (partially emerged) calderas, where explosive activity is more likely to occur and produce high-impact scenarios[20]. Second, we assessed tectonic settings, mostly favoring calderas situated along volcanic arcs, where tectonic activity can facilitate the sustained accumulation and recharge of volatile-rich magmas. Some shallow calderas in extensional and intraplate settings are also considered as high-priority, given their increased potential for explosive magma-water interactions involving basaltic compositions[28]. Within each of these categories, we also consider the morphological sub-type, treating full calderas as more hazardous than horse-shoe and eroded calderas, based on the reasoning outlined above. Five of the eight selected calderas are located in volcanic arc settings, four of which in the Tonga arc (SW Pacific) and one in the Sangihe arc (east Indonesia). Two calderas are associated with spreading ridges: SCD-0034 (Manus spreading center, NE of Papua New Guinea) and SCD-0049 (Cocos-Nazca spreading center, ~150 km N of Galapagos). Additionally, we included the only shallow (<700 m) intraplate caldera of the entire dataset (SCD-0030), which is located in the Bismarck Sea (NE of Papua New Guinea) on the South Bismark plate, at <300 km from populated coastlines. Although many of these calderas are situated relatively far from major coastlines (Fig. 4), they nonetheless pose significant hazards, including tsunami generation, tephra dispersal, and potential damage to submarine infrastructure, as exemplified by the 2022 Hunga Tonga eruption[12,13,15,50]. Craters and small calderas (classified here as volcanic depression <3 km in diameter) are not prioritized at this stage due to their limited size, which is assumed to correlate with lower explosive potential and reduced tsunami hazard from smaller collapse volumes. This prioritization should be interpreted as a preliminary recommendation rather than a comprehensive or definitive list.

Table. 1| List of high-priority calderas. They were selected based on their growth stage (i.e. water depth), tectonic setting and morphological sub-type.

Caldera name	Long (°)	Lat (°)	Caldera sub-type	Growth stage (water depth)	Tectonic setting
SCD-0007	-175.64	-15.60	full caldera	4 (~0 m)	Volcanic arc (Tonga arc)
SCD-0006	125.44	2.70	horse-shoe/eroded caldera	4 (~0 m)	Volcanic arc (Sangihe arc)
SCD-0048	-176.60	-15.10	full caldera	3 (~100 m)	Volcanic arc (Tonga arc)
SCD-0027	-176.91	-24.59	full caldera	3 (~500 m)	Volcanic arc (Tonga arc)

SCD-0054	-173.90	-16.22	horse-shoe/eroded caldera	3 (~400 m)	Volcanic arc (Tonga arc)
SCD-0034	147.16	-2.97	horse-shoe/eroded caldera	3 (~600 m)	Divergent (Manus spreading ridge)
SCD-0049	-90.01	1.78	full caldera	3 (~700 m)	Divergent (Cocos-Nazca spreading center)
SCD-0030	148.23	-4.10	horse-shoe/eroded caldera	3 (~700 m)	Intraplate (South Bismarck plate)

Despite the limitations (see Methods section), this work provides the basis for (1) enhanced submarine caldera discovery; and (2) targeted sea exploration to assess potential risk from poorly known submarine calderas that may experience hazardous volcanic activity in the near future. In terms of future directions, several key advancements can enhance the study and monitoring of submarine calderas in the coming years. Refinements to machine-learning caldera detection algorithms will improve the accuracy of caldera identification. Algorithmic improvements should focus on distinguishing complex morphologies, integrating geological context, and reducing false positives to enhance detection reliability at global scales. Additionally, improving accessibility to high-resolution bathymetric mapping technologies could accelerate the discovery and characterization of submarine calderas. Emerging tools such as wave gliders and Uncrewed Surface Vehicles (USVs) offer cost-effective alternatives to conventional ship-based surveys, but are still relatively expensive for research institutions. In this direction, the *Seabed 2030* initiative aims to achieve complete global coverage of the ocean floor by 2030, contributing increasingly higher-resolution data to the GEBCO bathymetric dataset. Expanding seafloor monitoring networks at high-risk calderas, particularly through pressure sensors and hydroacoustic arrays, will be essential for early detection of submarine eruptions. Existing networks, such as those deployed at Axial Caldera (Juan de Fuca Ridge)[54] and Kolumbo caldera (NE of Santorini, Greece)[55], provide a model for establishing real-time monitoring systems in other volcanic regions. Increasing observational capacity will be crucial for surveying volcanic activity, such as tracking magma movement, gas emissions, and seismic precursors to submarine eruptions. Translating the information gathered from advances in the monitoring and identification of submarine calderas to an understanding of the hazards and risks they pose to societies and critical infrastructure, such as shipping and telecommunications, requires additional steps. A greater understanding of the physical processes underpinning explosive submarine eruptions requires increased empirical and experimental data; associating the process with the hazard footprint typically requires tailored numerical modelling; and understanding how societal assets respond to the hazard requires analysis and expansion of the limited empirical and experimental dataset. Addressing these barriers through collaborative research initiatives and technological innovation will be key to advancing submarine volcanic hazard and risk assessments, and will represent a critical step toward geohazard preparedness for coastal populations and global infrastructure networks.

This work introduces a novel global-scale detection framework that demonstrates high efficacy for submarine caldera identification. The CDA enables automated recognition of sub-circular volcanic depressions across tectonic domains with minimal geological input. Unlike traditional mapping efforts constrained by spatial coverage or regional bias, our framework integrates morphometric filtering, GIS-based manual classification, and PCA-supported refinement, allowing reproducible and scalable feature extraction from low-resolution global bathymetry. Importantly, it successfully identified 78 calderas, 76 of which detected for the first time, spanning diverse morphological types and growth stages, which in turn reflect their hazard potential[20]. This expansion of the submarine caldera inventory, by an order of magnitude, fills a longstanding observational gap and generates actionable data to guide hazard prioritization, ~~monitoring strategies,~~ and targeted field campaigns. As such, this framework not only advances submarine volcanism research but also lays the foundation for integrating machine learning into future planetary-scale hazard assessments.

Methods

The framework adopted to conduct this work included four main processes (Fig. 5): (1) Use of the machine learning-based caldera detection algorithm (CDA), modified from the original crater detection algorithm created by Lee and Hogan[35]; (2) Post-processing filtering to reduce the number of false positives from the initial output list; (3) Feature manual classification, through GIS-based analysis and expert judgment; and (4) Final inspection of the outputs, through principal component analysis (PCA). Details are reported in the following sections.

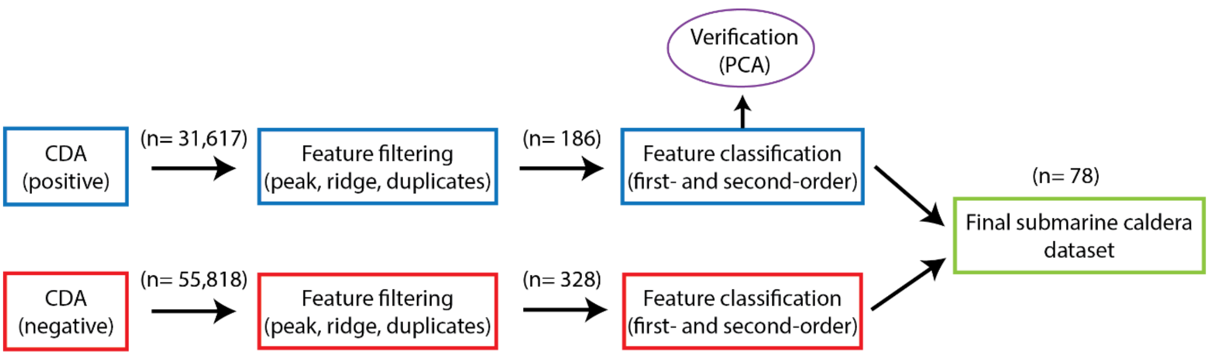


Fig. 5| Workflow for the creation of the submarine caldera dataset.

Caldera Detection Algorithm (CDA), inputs, outputs, filtering and post-processing

The original algorithm developed by Lee and Hogan[35] was trained to detect sub-circular depressions with the algorithm reported to ~~not require further training for similar application~~. The only input is a digital elevation model; in our case, we used a modified version of the algorithm and refer to it as caldera detection algorithm (CDA). As input DEM we used the GEBCO 2023 bathymetry dataset[36], which has a resolution of 15 arc-seconds, corresponding to approximately 500 m at the equator, and was cropped around the over 43,000 seamounts from the Gevorgian et al.[19] dataset (100 km radius). This DEM had a coverage of about 80% of the Earth's seafloor. This approach not only supported verification on the volcanic origin of some of the potential calderas detected through the CDA, but also reduced computational effort within the GIS environment (ESRI® ArcMap™). Additionally, using a DEM cropped around previously unstudied seamounts, aligns with the scope of this work, which is to detect previously undocumented calderas. Despite its relatively low resolution, GEBCO provides global coverage, enabling the automated detection of depressions across all ocean floors. The cropped GEBCO bathymetry was then reprojected in WGS 1984 World Mercator to run the CDA.

The CDA splits the input GEBCO bathymetry map into small orthographically projected regions that overlap in location and step logarithmically in size from about 2 degrees to about 30 degrees. Each image is processed to find sub-circular features, and the location and size are stored in an output Excel table; this means that the same feature can be found in more than one image, and in that case, we will have *Duplicates*>1. The CDA main outputs include potential calderas of at least ~4 km in diameter (at the equator). This means that features <4 km could still be detected but with a lower level of confidence, particularly if they were away from the equator, due to distortion effects of the projected DEM. Each potential caldera included several output parameters, including *lat* and *long*, number of *duplicates*, the *diameter*, several water depth parameters of the feature itself (*center mean*, *inner mean*, *ring mean* and *outer mean*), and the flags *peak* and *ridge*. They are all listed and detailed in Table 2.

The original algorithm used by Lee and Hogan[35] only detects depressions, regardless of whether they are at the top of a mound or on a flat surface. Therefore, in order to "guide" the algorithm towards possible depressions on a mound, as we would expect from a caldera, we inverted the CDA and applied two water depth flags, *peak* and *ridge (e*xplained in detail below). This approach provided us with two datasets of potential calderas: *positive* (depressions), from the original algorithm, and *negative* (mounds), from the inverted algorithm.

Across both datasets, we had a total of over 87,000 possible calderas, most of which were false positive. Therefore, to reduce this number to only potentially realistic calderas, we used the output flags *peak* and *ridge*. *Peak*=TRUE when the potential caldera has a positive morphology in its center, and *Ridge*=TRUE when the potential caldera is elevated respect to the surrounding seafloor. Therefore, to only isolate features that resemble a classic caldera (i.e. a mound with a depression in the middle), we set as filters *peak*=FALSE and *ridge*=TRUE . This process brought the number down to nearly 1,200 potential calderas. Then, to remove the still many false positives, both datasets were further processed by eliminating features that have the lowest number of *Duplicates* (n=1), which means they are found only once in the overlapping images of different resolution mentioned earlier, hence with the lowest confidence. This brought the number of potential calderas down to 514, which were then inspected and classified individually, followed by final verification through principal component analysis.

Table. 2| Output parameters of the CDA. Note that to remove as many false positive as possible, we set Duplicates>1, Peak=false and Ridge=true as filters.

PARAMETER	DESCRIPTION
Diameter	Mean diameter of the feature
Ring mean	Mean water depth of the feature ring
Centre mean	Mean water depth measured within a narrow ring concentric to the feature center, positioned at 10% of the caldera radius, to represent near-central bathymetric conditions
Inner mean	Mean water depth measured within a concentric ring located at 75% of the feature radius, to represent intermediate bathymetric conditions within the structure.

Outer mean	Mean water depth measured within a concentric ring located at 125% of the caldera radius, characterizing the bathymetric conditions just beyond the caldera rim.
Duplicates	Number of times a feature is independently detected at consistent location and diameter across multiple reprojected sub-images from the original DEM.
Peak	If P=true, center mean > Inner mean; if P=false, center mean < Inner mean
Ridge	If P=true, Inner mean > Outer mean; if P=false, Inner mean < Outer mean

Feature classification and final inspection

To conduct the classification, we used a combination of GIS-based analyses and visual assessments. As a first step we plotted the filtered 514 potential calderas in a GIS environment and overlaid them on the hillshade derived from the GEBCO bathymetry for visual morphological inspection, bathymetric contour analysis and profiling (Supplementary File 3). As additional elements for the classification we used the seamounts catalogue from Gevorgian et al. (2023), to confirm whether the features were volcanic or not, and the global coral reef database[56], to confirm if the feature was an atoll. Some areas also had higher-resolution multibeam bathymetry (MBBDB)[37] (90 m resolution), and that information was inserted in the final dataset (Supplementary Files 1 and 2) and supported our classification.

We conducted a first-order classification, considering a *Type*, and a second-order classification, considering a *Sub-type* (where applicable), for each of the filtered features. As types we identified *calderas*, *atolls*, *craters*, and *flank scar/eroded surface*. For the sub-types we divided the types into *full calderas*, *horse-shoe/eroded calderas*, *full craters / small calderas*, *horse-shoe/eroded craters / small calderas*, *flank scar/eroded surface*, *young atoll*, and *sunken atoll*. The flank scar/eroded surface category was retained without further subdivision at both the type and sub-type levels, as it was deemed irrelevant to the objective of this study, namely, distinguishing calderas from non-calderas. In contrast, atolls were subdivided into young and sunken sub-types, as their morphology could potentially be misinterpreted as calderas. This distinction was therefore considered necessary. The features that could not be confidently classified were categorized as *undefined* and not considered further. Detailed descriptions and reasoning of classification for types and sub-types are reported in Table 3.

For the PCA, we only used the *positive* dataset, because the parameters obtained from the CDA were referring to the feature itself rather than the depression, hence in the case of the *negative* dataset, the parameters referred to the mound, therefore not useable for the classification of the depression. The PCA was conducted separately for *types* (first-order classification) and *sub-types* (second-order classification), with 75% confidence ellipses, to verify its efficacy and to identify clusters that otherwise would not be distinct. The parameters used for the PCA are *diameter*, *ring mean*, *center mean*, *inner mean*, *outer mean*, and different ratios between the ring mean and the other water depth parameters, to understand the slope relationships between the inside and outside of the depression respect to the ring (*center mean/ring mean*; *inner mean/ring mean*; *outer ring/ring mean*).

Table 3| Type and sub-type descriptions and classification methods.

Type	Sub-type	Description	Support
Atoll	*Young atoll*	Emerged or very shallow atoll (i.e. <100 m below sea level)	Satellite imagery and/or coral reef database[56]
	Sunken atoll	Deep sea old atoll (> 100 m below sea level) with gentle gradients	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling (SF3)

		between center and edges	
Flank scar/eroded surface	*Flank scar/eroded surface*	Valley between volcanic ridges, or eroded volcanic surface or edifices	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling (SF3)
Crater/small caldera	*Full crater/small caldera*	Enclosed volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
	Horse-shoe/eroded crater / small caldera	Open/eroded volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
Caldera	*Full caldera*	Enclosed volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
	Horse-shoe/eroded caldera	Open/eroded volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling (SF3), presence of seamounts[19]
Undefined	*Undefined*	Not classifiable (poor bathymetry or unclear structure)	N/A

Abbreviation: SF= Supplementary File

Limitations

Our approach presents several limitations. The principal component analysis (PCA) was conducted exclusively on the *positive* dataset, which comprises 186 of the 514 total features, as opposed to the *negative* dataset's 328 features. This imbalance reduces the statistical robustness of the analysis. The use of GEBCO global low-resolution bathymetry further constrains detection capabilities, resulting in a high number of undefined features (229 of 514), some of which may actually represent calderas. Finally, several previously documented calderas were not detected by the CDA for one of the following reasons: (1) their locations lay outside the cropped DEM used as input (e.g. Kikai, Kolumbo, Santorini, Krakatau); (2) they lacked sub-circular morphologies suitable for algorithmic detection (e.g. Axial); or (3) they were initially identified but discarded during filtering due to low confidence scores (e.g. Hunga Tonga and Havre), reflecting certain limitations of the method. Notwithstanding these exclusions, the primary goal of this study was to identify previously undocumented submarine calderas. Collectively, these limitations highlight the fact that those we provided represent a minimum number of potential submarine calderas, rather than an exhaustive inventory.

Data and Code Availability

The modified CDA used for this work can be found at https://github.com/eelsirhc/deepcaldera_project. The unfiltered and filtered datasets can be found in the repository of the Nanyang Technological University: https://researchdata.ntu.edu.sg/privateurl.xhtml?token=643346d9-b122-49cc-9db6-ad1072a9958f.

References

1. Petraglia, M. D., Korisettar, R. & Pal, J. N. The Toba volcanic super-eruption of 74,000 years ago: Climate change, environments, and evolving humans. *Quaternary International* **258**, 1–4 (2012).
2. Michon, L., Massin, F., Famin, V., Ferrazzini, V. & Roult, G. Basaltic calderas: Collapse dynamics, edifice deformation, and variations of magma withdrawal. *J. Geophys. Res.* **116**, (2011).
3. Bond, A. & Sparks, R. S. J. The Minoan eruption of Santorini, Greece. *JGS* **132**, 1–16 (1976).
4. Oppenheimer, C. Climatic, environmental and human consequences of the largest known historic eruption: Tambora volcano (Indonesia) 1815. *Progress in Physical Geography: Earth and Environment* **27**, 230–259 (2003).

5. Newhall, C. G. & Self, S. The volcanic explosivity index (VEI) an estimate of explosive magnitude for historical volcanism. *J. Geophys. Res.* **87**, 1231–1238 (1982).
6. Wirakusumah, A. D. & Rachmat, H. Impact of the 1815 Tambora Eruption to global climate change. *IOP Conf. Ser.: Earth Environ. Sci.* **71**, 012007 (2017).
7. Raible, C. C. *et al.* Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. *WIREs Climate Change* **7**, 569–589 (2016).
8. Bennington, N. *et al.* The progression of basaltic–rhyolitic melt storage at Yellowstone Caldera. *Nature* **637**, 97–102 (2025).
9. Costa, A. *et al.* Tephra fallout hazard assessment at the Campi Flegrei caldera (Italy). *Bull Volcanol* **71**, 259–273 (2009).
10. Bevilacqua, A. *et al.* Accelerating upper crustal deformation and seismicity of Campi Flegrei caldera (Italy), during the 2000–2023 unrest. *Commun Earth Environ* **5**, 742 (2024).
11. White, J. D. L., Schipper, C. I. & Kano, K. Submarine Explosive Eruptions. in *The Encyclopedia of Volcanoes* 553–569 (Elsevier, 2015). doi:10.1016/B978-0-12-385938-9.00031-6.
12. Cronin, S. *et al. Extreme Explosivity of the 15 January 2022 Hunga Eruption, Tonga, Driven by Magma-Mixing, Caldera Collapse and Magma-Water Interaction*. https://hdl.handle.net/10523/43599 (2023).
13. Chacón-Barrantes, S., Rivera-Cerdas, F. & Murillo-Gutiérrez, A. Impact of the tsunami caused by the Hunga Tonga–Hunga Ha'apai eruption in Costa Rica on 15 January 2022. *Bull Volcanol* **85**, 36 (2023).
14. Dogan, G. G., Yalciner, A. C., Annunziato, A., Yalciner, B. & Necmioglu, O. Global propagation of air pressure waves and consequent ocean waves due to the January 2022 Hunga Tonga-Hunga Ha'apai eruption. *Ocean Engineering* **267**, 113174 (2023).
15. Lynett, P. *et al.* Diverse tsunamigenesis triggered by the Hunga Tonga-Hunga Ha'apai eruption. *Nature* **609**, 728–733 (2022).
16. The World Bank. *The January 15, 2022 Hunga Tonga-Hunga Ha'apai Eruption and Tsunami, Tonga Global Rapid Post Disaster Damage Estimation (Grade) Report. Https://Thedocs.Worldbank.Org/En/Doc/B69af83e486aa652d4232276ad698c7b-0070062022/Original/GRADE-Report-Tonga-Volcanic-Eruption.Pdf*. (2022).
17. Global Volcanism Program & Venzke, E. Volcanoes of the World, v.5.2.0. Global Volcanism Program (2024).
18. Geyer, A. & Martí, J. The new worldwide collapse caldera database (CCDB): A tool for studying and understanding caldera processes. *Journal of Volcanology and Geothermal Research* **175**, 334–354 (2008).
19. Gevorgian, J., Sandwell, D. T., Yu, Y., Kim, S. & Wessel, P. Global Distribution and Morphology of Small Seamounts. *Earth and Space Science* **10**, e2022EA002331 (2023).
20. Verolino, A., Wee, S. F., Jenkins, S. F., Costa, F. & Switzer, A. D. SEATANI: hazards from seamounts in Southeast Asia, Taiwan, and Andaman and Nicobar Islands (eastern India). *Nat. Hazards Earth Syst. Sci.* **24**, 1203–1222 (2024).
21. Acocella, V. Calderas. in *Volcano-Tectonic Processes* 163–203 (Springer International Publishing, Cham, 2021). doi:10.1007/978-3-030-65968-4_5.
22. Branney, M. & Acocella, V. Calderas. In The encyclopedia of volcanoes (pp. 299-315). Academic Press. https://doi.org/10.1016/B978-0-12-385938-9.00016-X. in (2015).
23. Wang, W. *et al.* Volatile Element Evidence of Local MORB Mantle Heterogeneity Beneath the Southwest Indian Ridge, 48°–51°E. *Geochem Geophys Geosyst* **22**, e2021GC009647 (2021).
24. Poland, M. P., Takahashi, T. J. & Landowski, C. M. *Characteristics of Hawaiian Volcanoes*. https://pubs.usgs.gov/publication/pp1801 (2014) doi:10.3133/pp1801.
25. Martí, J., Mitjavila, J. & Araña, V. Stratigraphy, structure and geochronology of the Las Cañadas caldera (Tenerife, Canary Islands). *Geol. Mag.* **131**, 715–727 (1994).
26. Geist, D. J. *et al.* Wolf Volcano, Galápagos Archipelago: Melting and Magmatic Evolution at the Margins of a Mantle Plume. *Journal of Petrology* **46**, 2197–2224 (2005).
27. Cocchi, L., Caratori Tontini, F., Muccini, F. & De Ronde, C. E. J. Magnetic Expression of Hydrothermal Systems Hosted by Submarine Calderas in Subduction Settings: Examples from the Palinuro and Brothers Volcanoes. *Geosciences* **11**, 504 (2021).
28. Zimanowski, B., Büttner, R., Lorenz, V. & Häfele, H.-G. Fragmentation of basaltic melt in the course of explosive volcanism. *Journal of Geophysical Research: Solid Earth* **102**, 803–814 (1997).
29. Verolino, A., White, J. D. L. & Brenna, M. Eruption dynamics at Pahvant Butte volcano, Utah, western USA: insights from ash-sheet dispersal, grain size, and geochemical data. *Bull Volcanol* **80**, 1–18 (2018).

30. Verolino, A., White, J. D. L., Dürig, T. & Cappuccio, F. Black Point – Pyroclasts of a Surtseyan eruption show no change during edifice growth to the surface from 100 m water depth. *Journal of Volcanology and Geothermal Research* **384**, 85–102 (2019).
31. Dürig, T. *et al.* Deep-sea eruptions boosted by induced fuel–coolant explosions. *Nat. Geosci.* **13**, 498–503 (2020).
32. Verolino, A., White, J. D. L., Baxter, R. J. M., Schipper, C. I. & Thordarson, T. Characteristics of Sub-Aerially Emplaced Pyroclasts in the Surtsey Eruption Deposits: Implications for Diverse Surtseyan Eruptive Styles. *Geosciences* **12**, 79 (2022).
33. Verolino, A. *et al.* Assessing volcanic hazard and exposure to lava flows at remote volcanic fields: a case study from the Bolaven Volcanic Field, Laos. *J Appl. Volcanol.* **11**, 6 (2022).
34. Murch, A. P., Portner, R. A., Rubin, K. H. & Clague, D. A. Deep-subaqueous implosive volcanism at West Mata seamount, Tonga. *Earth and Planetary Science Letters* **578**, 117328 (2022).
35. Lee, C. & Hogan, J. Automated crater detection with human level performance. *Computers & Geosciences* **147**, 104645 (2021).
36. GEBCO Compilation Group. GEBCO 2023 Grid. https://doi.org/doi:10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b. (2023).
37. MBBDB. National Centers for Environmental Information. 2004: Multibeam Bathymetry Database (MBBDB). NOAA National Centers for Environmental Information. Accessed [April 2025]. https://doi.org/doi:10.7289/V56T0JNC.
38. Falkenberg, J. J. *et al.* Spatial Variations in Magmatic Volatile Influx and Fluid Boiling in the Submarine Hydrothermal Systems of Niuatahi Caldera, Tonga Rear-Arc. *Geochem Geophys Geosyst* **23**, e2021GC010259 (2022).
39. Leat, P. T. *et al.* Bathymetry and geological setting of the South Sandwich Islands volcanic arc. *Antarctic Science* **28**, 293–303 (2016).
40. Staudigel, H. & Clague, D. The Geological History of Deep-Sea Volcanoes: Biosphere, Hydrosphere, and Lithosphere Interactions. *Oceanog.* **23**, 58–71 (2010).
41. Danovaro, R. *et al.* A submarine volcanic eruption leads to a novel microbial habitat. *Nat Ecol Evol* **1**, (2017).
42. Carey, R. J., Butterfield, D. A. & Clark, M. R. Submarine Volcanic Eruptions and Their Impacts on Hydrothermal Systems and Biological Communities. *Annual Review of Earth and Planetary Sciences* **53**, 397–423 (2025).
43. Fouquet, Y. *et al.* Volcanic and hydrothermal processes in submarine calderas: The Kulo Lasi example (SW Pacific). *Ore Geology Reviews* **99**, 314–343 (2018).
44. Omira, R. *et al.* Global Tonga tsunami explained by a fast-moving atmospheric source. *Nature* **609**, 734–740 (2022).
45. Ren, Z., Higuera, P. & Li-Fan Liu, P. On Tsunami Waves Induced by Atmospheric Pressure Shock Waves After the 2022 Hunga Tonga-Hunga Ha'apai Volcano Eruption. *JGR Oceans* **128**, (2023).
46. Verolino, A. *et al.* Forced and free waves of simulated volcanic meteo-tsunamis in the South China Sea. *npj Nat. Hazards* **2**, 25 (2025).
47. Self, S. & Rampino, M. R. The 1883 eruption of Krakatau. *Nature* **294**, 699–704 (1981).
48. Rampino, M. R. & Self, S. Historic Eruptions of Tambora (1815), Krakatau (1883), and Agung (1963), their Stratospheric Aerosols, and Climatic Impact. *Quat. res.* **18**, 127–143 (1982).
49. Clare, M. A. *et al.* Volcanic eruptions and the global subsea telecommunications network. *Bull Volcanol* **87**, 51 (2025).
50. Clare, M. A. *et al.* Fast and destructive density currents created by ocean-entering volcanic eruptions. *Science* **381**, 1085–1092 (2023).
51. Sparks, R. S. J. *et al.* Large silicic magma bodies and very large magnitude explosive eruptions. *Bull Volcanol* **84**, (2022).
52. Watt, S. F. L. *et al.* Combinations of volcanic-flank and seafloor-sediment failure offshore Montserrat, and their implications for tsunami generation. *Earth and Planetary Science Letters* **319–320**, 228–240 (2012).
53. Preine, J. *et al.* Hazardous explosive eruptions of a recharging multi-cyclic island arc caldera. *Nat. Geosci.* **17**, 323–331 (2024).
54. Wilcock, W. S. D. *et al.* Seismic constraints on caldera dynamics from the 2015 Axial Seamount eruption. *Science* **354**, 1395–1399 (2016).
55. Nomikou, P. *et al.* SANTORY: SANTORini's Seafloor Volcanic ObservatorY. *Front. Mar. Sci.* **9**, 796376 (2022).
56. UNEP-WCMC. Global Coral Reef Database: WCMC008_CoralReef2018_Py_v4_1. UN Environment Programme World Conservation Monitoring Centre, Cambridge, UK (2018). (2018).

Acknowledgements

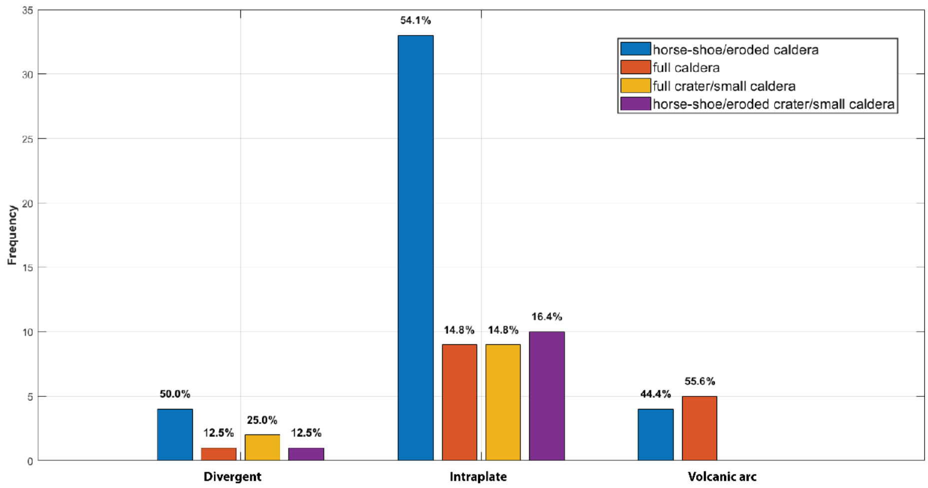
This project was supported by the Ministry of Education, Singapore, under its MOE Academic Research Fund Tier 3 InVEST project (Award MOE-MOET32021-0002). This work comprises EOS contribution number 661.

Author Contributions

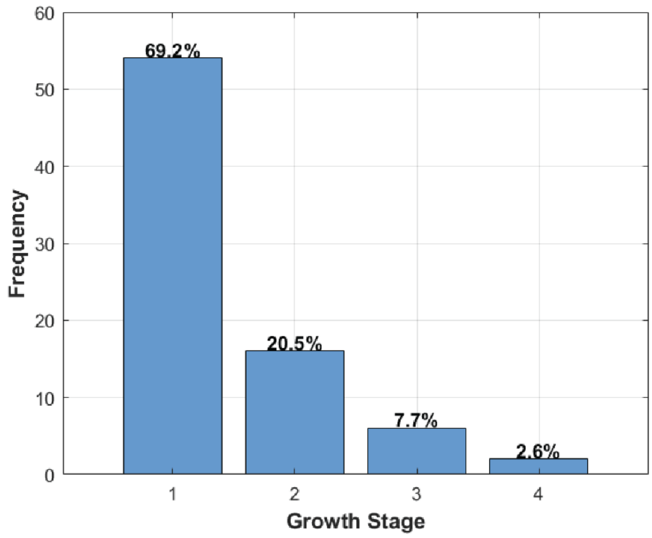
AV: work idealization, paper drafting, editing; CL: assistance with the Caldera Detection Algorithm, editing; SFJ: work idealization, editing; MJ: editing; ADS: editing

Competing Interests

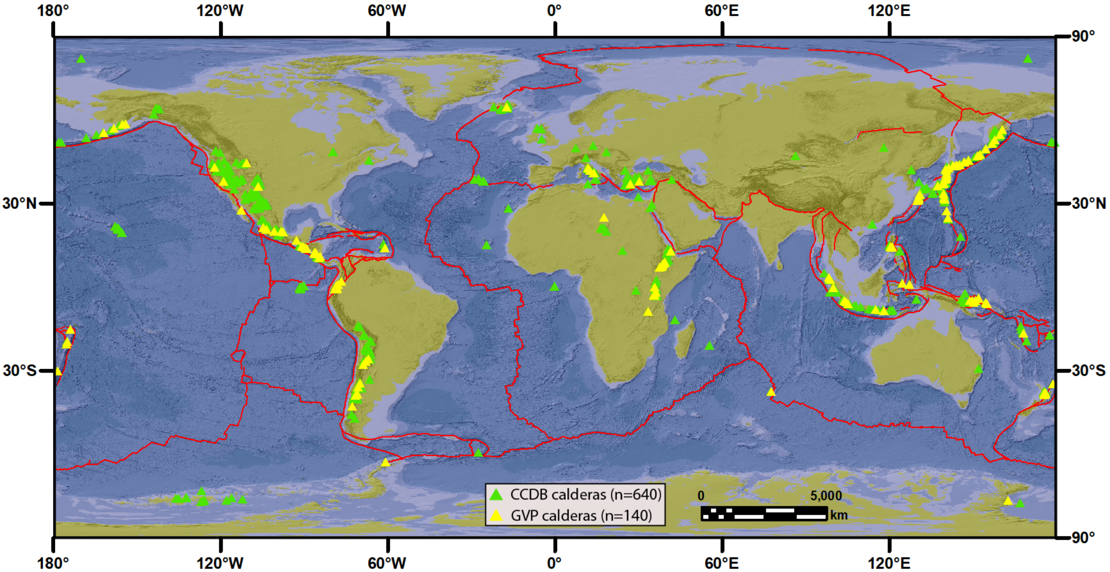
We declare that co-author A.D. Switzer is a member of the editorial board for Communications Earth & Environment (Ocean and Cryosphere).



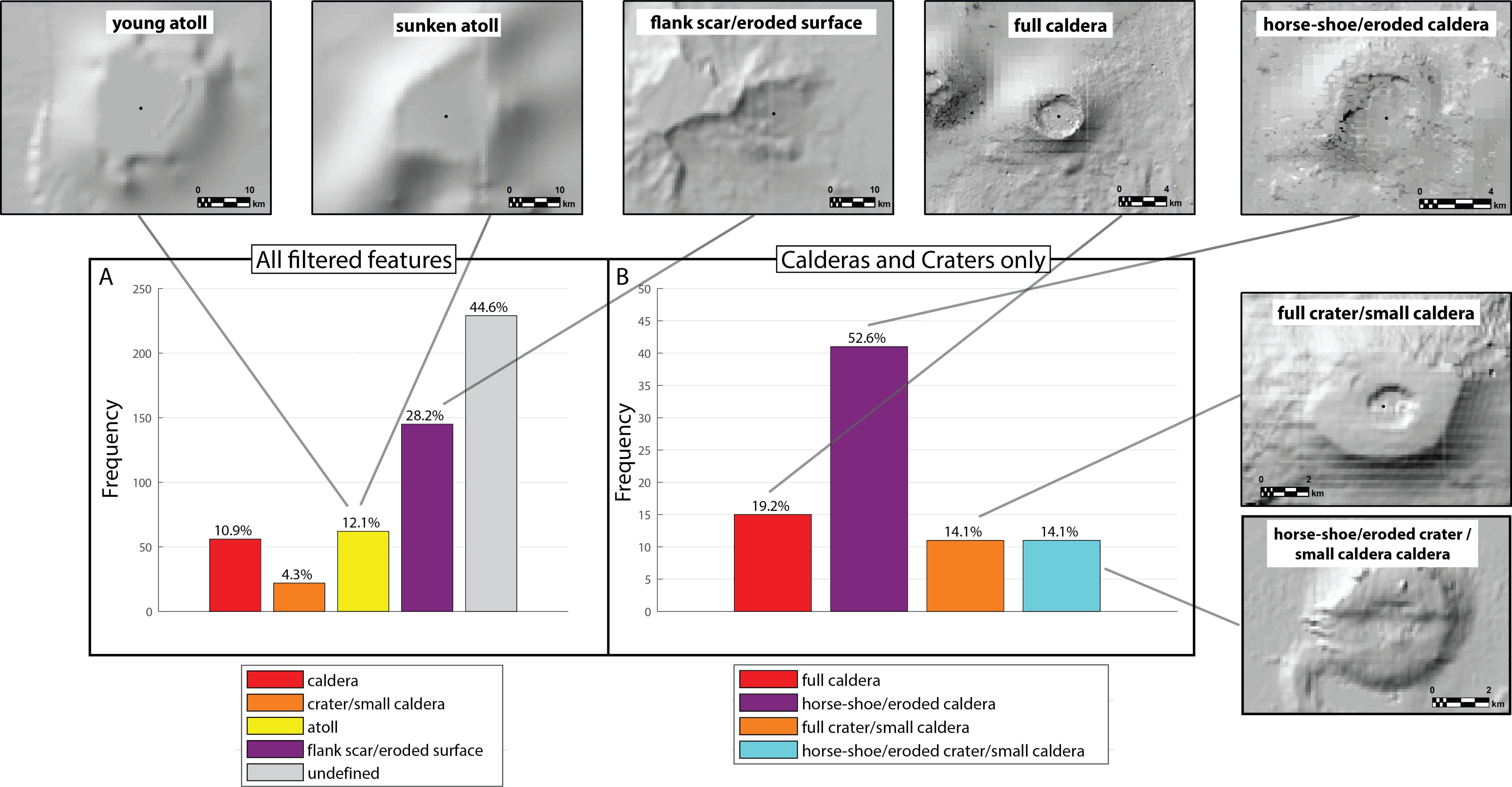
Extended Data Fig. 1| Distribution of caldera sub-types across tectonic settings.

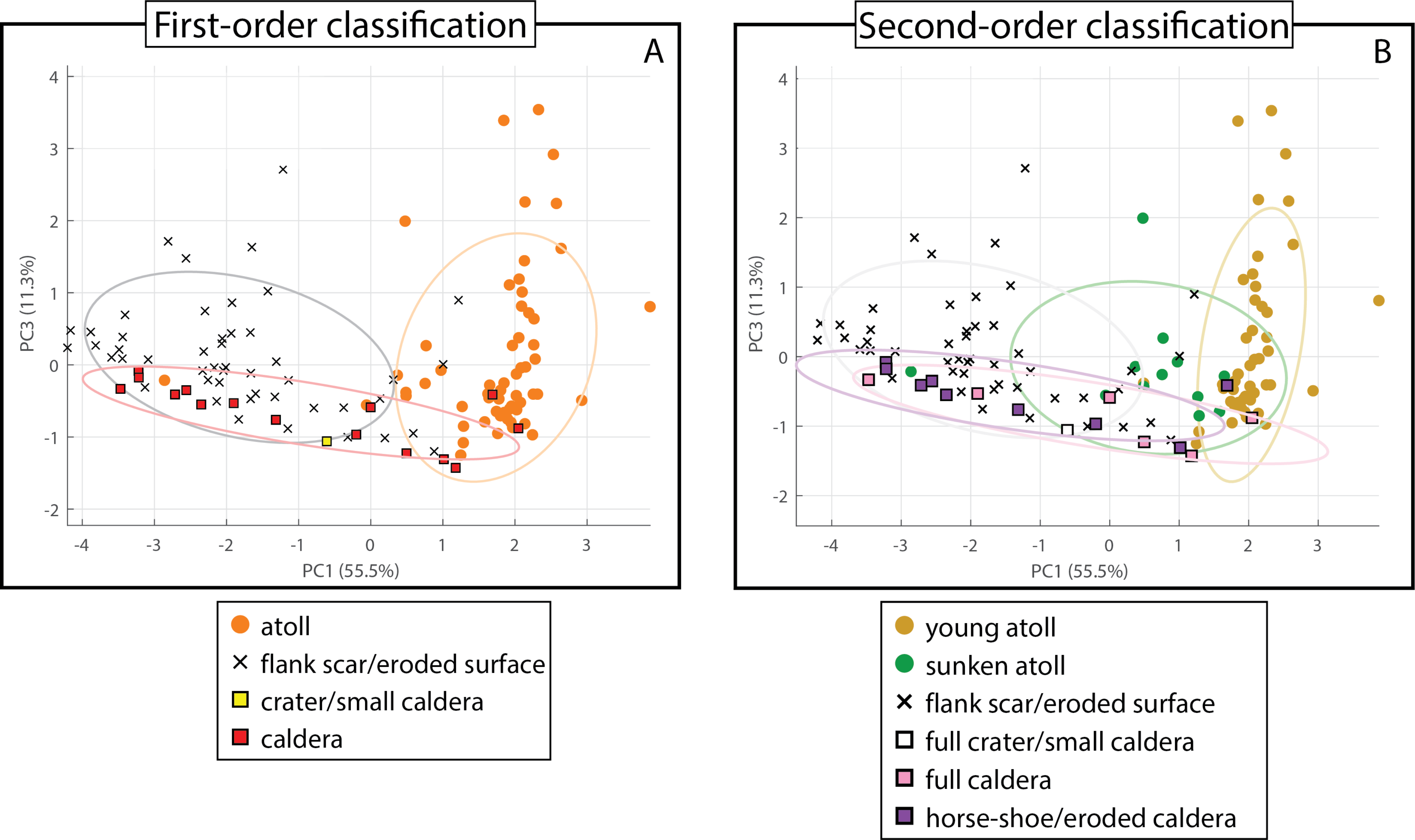


Extended Data Fig. 2| Distribution of caldera growth stages.

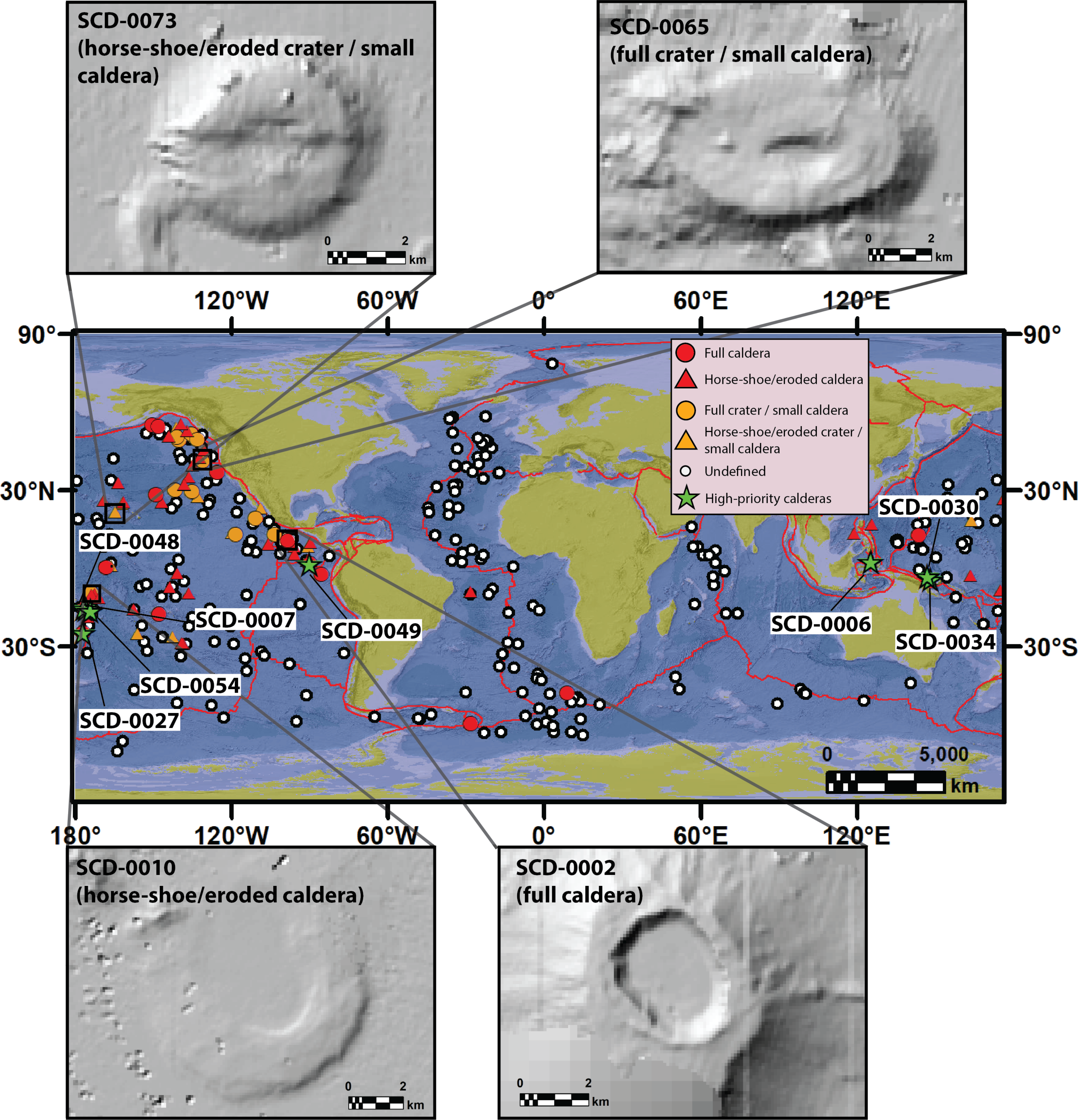


180°
120°W
60°W
0°
60°E
120°E
90°
30°N
30°S
CCDB calderas (n=640)
GVP calderas (n=140)
0
5,000
km


young atoll
sunken atoll
flank scar/eroded surface
full caldera
horse-shoe/eroded caldera
full crater/small caldera
horse-shoe/eroded crater /
small caldera caldera
0 10 km
0 4 km
0 2 km
All filtered features
A
Frequency
250
200
150
100
50
0
10.9%
4.3%
12.1%
28.2%
44.6%
Calderas and Craters only
B
Frequency
50
45
40
35
30
25
20
15
10
5
0
19.2%
52.6%
14.1%
14.1%
caldera
crater/small caldera
atoll
flank scar/eroded surface
undefined
full caldera
horse-shoe/eroded caldera
full crater/small caldera
horse-shoe/eroded crater/small caldera



First-order classification
A
PC3 (11.3%)
PC1 (55.5%)
atoll
flank scar/eroded surface
crater/small caldera
caldera
Second-order classification
B
PC3 (11.3%)
PC1 (55.5%)
young atoll
sunken atoll
flank scar/eroded surface
full crater/small caldera
full caldera
horse-shoe/eroded caldera


SCD-0073
(horse-shoe/eroded crater / small caldera)
0 2 km
SCD-0065
(full crater / small caldera)
0 2 km
120°W
60°W
0°
60°E
120°E
90°
30°N
30°S
180°
Full caldera
Horse-shoe/eroded caldera
Full crater / small caldera
Horse-shoe/eroded crater / small caldera
Undefined
High-priority calderas
SCD-0048
SCD-0007
SCD-0049
SCD-0054
SCD-0027
SCD-0030
SCD-0006
SCD-0034
0 5,000 km
SCD-0010
(horse-shoe/eroded caldera)
0 2 km
SCD-0002
(full caldera)
0 2 km

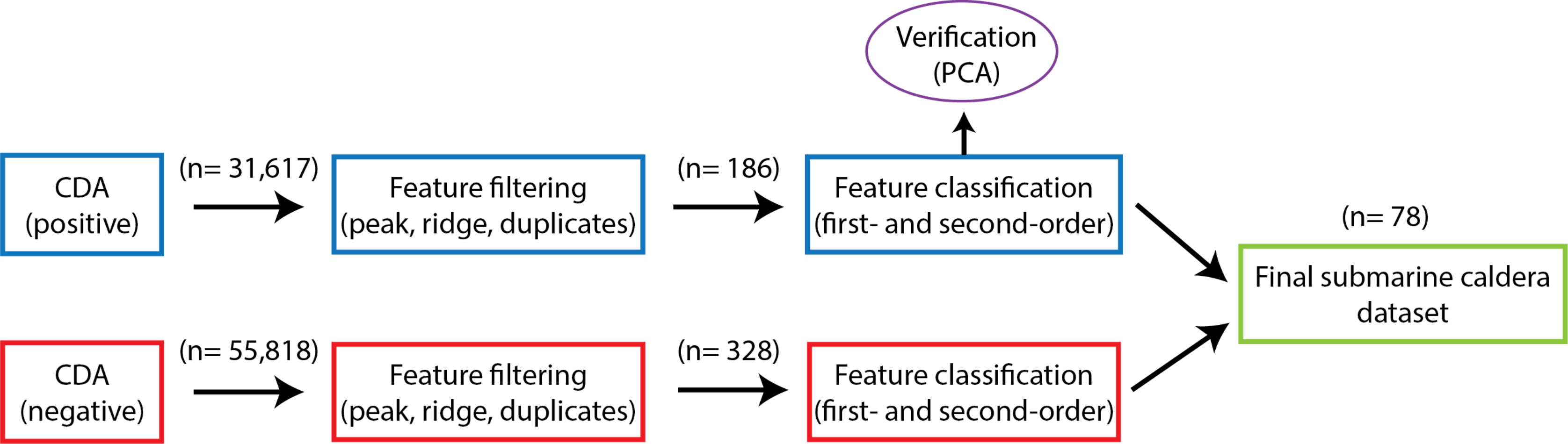

Verification
(PCA)
CDA
(positive)
(n= 31,617)
Feature filtering
(peak, ridge, duplicates)
(n= 186)
Feature classification
(first- and second-order)
(n= 78)
Final submarine caldera
dataset
CDA
(negative)
(n= 55,818)
Feature filtering
(peak, ridge, duplicates)
(n= 328)
Feature classification
(first- and second-order)

Table. 1| List of high-priority calderas. They were selected based on their growth stage (i.e. water depth), tectonic setting and morphological sub-type.

Caldera name	Long	Lat	Caldera sub-type
SCD-0007	-175.64	-15.60	full caldera
SCD-0006	125.44	2.70	horse-shoe/eroded caldera
SCD-0048	-176.60	-15.10	full caldera
SCD-0027	-176.91	-24.59	full caldera
SCD-0054	-173.90	-16.22	horse-shoe/eroded caldera
SCD-0034	147.16	-2.97	horse-shoe/eroded caldera
SCD-0049	-90.01	1.78	full caldera
SCD-0030	148.23	-4.10	horse-shoe/eroded caldera

wth stage (water de	Tectonic setting
4 (~0 m)	Volcanic arc (Tonga arc)
4 (~0 m)	Volcanic arc (Sangihe arc)
3 (~100 m)	Volcanic arc (Tonga arc)
3 (~500 m)	Volcanic arc (Tonga arc)
3 (~400 m)	Volcanic arc (Tonga arc)
3 (~600 m)	Divergent (Manus spreading ridge)
3 (~700 m)	Divergent (Cocos-Nazca spreading centre)
3 (~700 m)	Intraplate (South Bismarck plate)

Table. 2| Output parameters of the CDA. Note that to remove as many false positive as possible, we set Duplicates>1, Peak=false and Ridge=true as filters.

PARAMETER	DESCRIPTION
Diameter	Mean diameter of the feature
Ring mean	Mean water depth of the feature ring
Centre mean	Mean water depth measured within a narrow ring concentric to the feature center, positioned at 10% of the caldera radius, to represent near-central bathymetric conditions
Inner mean	Mean water depth measured within a concentric ring located at 75% of the feature radius, to represent intermediate bathymetric conditions
Outer mean	Mean water depth measured within a concentric ring located at 125% of the caldera radius, characterizing the bathymetric conditions just beyond the caldera rim.
Duplicates	Number of times a feature is detected within the 20 images (1-20)
Peak	If P=true, centre mean > Inner mean; if P=false, centre mean < Inner mean
Ridge	If P=true, Inner mean > Outer mean; if P=false, Inner mean < Outer mean

Table 3| Type and sub-type descriptions and classification methods.

Type	Sub-type
Atoll	*Young atoll*
	Sunk atoll
Flank scar/eroded surface	*Flank scar/eroded surface*
Crater/small caldera	*Full crater/small caldera*
	horse-shoe/eroded crater / small caldera
Caldera	*Full caldera*
	Horse-shoe/eroded caldera
Undefined	*Undefined*

Description	Support
Emerged or shallow atoll	Satellite imagery and/or coral reef database[38]
Deep sea old atoll with gentle gradients between centre and edges	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling
Valley between volcanic ridges, or eroded volcanic surface	Hillshade visual analysis, bathymetric contour analysis, bathymetric profiling
Enclosed volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Open/eroded volcanic depression <3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Enclosed volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Open/eroded volcanic depression >3km in diameter	Hillshade visual analysis, bathymetric contours, bathymetric profiling, presence of seamounts[18]
Not classifiable (poor bathymetry or unclear structure)	N/A