frontiers | Frontiers in Earth Science

TYPE Methods
PUBLISHED 26 April 2024
DOI 10.3389/feart.2024.1345104

OPEN ACCESS

EDITED BY
Carmen Benítez,
Universidad de Granada, Spain

REVIEWED BY
Francesco Massimetti,
University of Turin, Italy
Manuel Titos,
Icelandic Meteorological Office, Iceland
Michael Ramsey,
University of Pittsburgh, United States

*CORRESPONDENCE
Pablo Saunders-Shultz,
csaundersshultz@gmail.com

RECEIVED 27 November 2023
ACCEPTED 05 April 2024
PUBLISHED 26 April 2024

CITATION
Saunders-Shultz P, Lopez T, Dietterich H and Girona T (2024), Automatic identification and quantification of volcanic hotspots in Alaska using HotLINK: the hotspot learning and identification network.
Front. Earth Sci. 12:1345104.
doi: 10.3389/feart.2024.1345104

Automatic identification and quantification of volcanic hotspots in Alaska using HotLINK: the hotspot learning and identification network

Pablo Saunders-Shultz[1]*, Taryn Lopez[1], Hannah Dietterich[2] and Társilo Girona[1]

[1]Geophysical Institute, Alaska Volcano Observatory, University of Alaska Fairbanks, Fairbanks, AK, United States, [2]U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK, United States

An increase in volcanic thermal emissions can indicate subsurface and surface processes that precede, or coincide with, volcanic eruptions. Space-borne infrared sensors can detect hotspots—defined here as localized volcanic thermal emissions—in near-real-time. However, automatic hotspot detection systems are needed to efficiently analyze the large quantities of data produced. While hotspots have been automatically detected for over 20 years with simple thresholding algorithms, new computer vision technologies, such as convolutional neural networks (CNNs), can enable improved detection capabilities. Here we introduce HotLINK: the Hotspot Learning and Identification Network, a CNN trained to detect hotspots with a dataset of ~3,800 satellite-based, Visible Infrared Imaging Radiometer Suite (VIIRS) images from Mount Veniaminof and Mount Cleveland volcanoes, Alaska. We find that our model achieves an accuracy of 96% (F1-score 0.92) when evaluated on ~1,700 unseen images from the same volcanoes, and 95% (F1-score 0.67) when evaluated on ~3,000 images from six additional Alaska volcanoes (Augustine Volcano, Bogoslof Island, Okmok Caldera, Pavlof Volcano, Redoubt Volcano, Shishaldin Volcano). In comparison with an existing threshold-based hotspot detection algorithm, MIROVA (Coppola et al., Geological Society, London, Special Publications, 2016, 426, 181–205), our model detects 22% more hotspots and produces 12% fewer false positives. Additional testing on ~700 labeled Moderate Resolution Imaging Spectroradiometer (MODIS) images from Mount Veniaminof demonstrates that our model is applicable to this sensor's data as well, achieving an accuracy of 98% (F1-score 0.95). We apply HotLINK to 10 years of VIIRS data and 22 years of MODIS data for the eight aforementioned Alaska volcanoes and calculate the radiative power of detected hotspots. From these time series we find that HotLINK accurately characterizes background and eruptive periods, similar to MIROVA, but also detects more subtle warming signals, potentially related to volcanic unrest. We identify three advantages to our model over its predecessors: 1) the ability to detect more subtle volcanic hotspots and produce fewer false positives, especially in daytime images; 2) probabilistic predictions provide a measure of detection confidence; and 3) its transferability, i.e., the successful application to multiple sensors and multiple volcanoes

without the need for threshold tuning, suggesting the potential for global application.

KEYWORDS

thermal remote sensing, global volcano monitoring, machine learning, neural network, eruption forecasting, VIIRS, MODIS, MIROVA

1 Introduction

Volcanic eruptions pose hazards to human life and society (Loughlin et al., 2015). To mitigate these hazards, volcano monitoring agencies aim to detect signs of unrest and eruption as early as possible. Local monitoring stations and remote satellite observations are commonly used to monitor volcanic unrest (e.g., Dehn et al., 2000; Cameron et al., 2018; Girona et al., 2021). Here we will focus on one satellite-based approach to monitor thermal unrest: detecting localized volcanic heat emissions, also referred to as volcanic hotspots. In a single satellite image, hotspots may be identified as a few pixels of elevated infrared radiance caused by relatively high temperature volcanic features. Hotspots may be produced by various types of volcanic activity, including lava flows (Dehn et al., 2000; Hirn et al., 2009; Blackett, 2013; Harris, 2013; Wright, 2016), explosive and strombolian activity (Harris and Stevenson, 1997; Coppola et al., 2012; Coppola et al., 2014), dome growth (Carter et al., 2007; Ramsey et al., 2012; Coppola et al., 2022), degassing of a hot vent or fumarole field (Oppenhemier et al., 1993; Harris and Stevenson, 1997; Blackett, 2013; Laiolo et al., 2017), or increased surface meltwater in the case of glaciated volcanoes (Pieri and Abrams, 2005; Blackett, 2013; Bleick et al., 2013; Reath et al., 2016). Therefore, monitoring changes in hotspot activity can provide key insights into a volcano's behavior by indicating the presence of thermal volcanic features and characterizing them over time. Due to the utility of these observations, thermal satellite data are used by volcano observatories as part of their daily monitoring operations (Dehn et al., 2000; Dehn et al., 2002; Harris et al., 2016; Harris et al., 2017; Cameron et al., 2018; Coombs et al., 2018; Coppola et al., 2020; Pritchard et al., 2022; Chevrel et al., 2023). Automating the detection and quantification of volcanic hotspots can provide near-real time information to volcano observatory scientists to inform decision-making and provide a mechanism to generate long time series of thermal activity for volcanoes around the world. Time series observations are useful for determining baseline activity, identifying periods of volcanic unrest, characterizing the thermal evolution of ongoing eruptions, and retrospectively studying eruptive histories and processes (Dehn et al., 2002; Wright, 2016; Girona et al., 2021; Chevrel et al., 2023; Coppola et al., 2023).

Surface hotspots will result in increased spectral radiance (Wm^{-2} sr^{-1} μm^{-1}) in both Mid-Infrared (MIR, 3–5 μm) and Thermal-Infrared (TIR, 5–20 μm) wavelengths (Harris, 2013). This behavior is characterized by Planck's Law, which states that as the temperature of a blackbody increases, the spectrum of energy it emits will increase in radiance, and the peak radiance will shift to shorter wavelengths (Planck, 1914). Therefore, a volcanic hotspot can be identified by an elevated TIR radiance above background and an even greater signal above background in MIR radiance (e.g., Blackett, 2013; Blackett, 2017). For especially hot surfaces (>950 K), the peak radiance emission is in the shortwave infrared (SWIR, 1.4–3 μm) part of the spectrum. The distinct features produced by hotspots in MIR and TIR bands have been exploited to automate their detection by different algorithms (Higgins and Harris, 1997; Pergola et al., 2004; Wright et al., 2004; Ganci et al., 2011; Coppola et al., 2016; Gouhier et al., 2016; Lombardo, 2016; Valade et al., 2019; Castaño et al., 2020; Genzano et al., 2020; Layana et al., 2020; Massimetti et al., 2020; Corradino et al., 2023; Ramsey et al., 2023).

One of the first algorithms to automate volcanic hotspot detection, MODVOLC (Wright et al., 2004), applies a threshold to the Normalized Thermal Index (NTI), constructed from radiance values of MIR and TIR bands:

$$NTI = \frac{MIR - TIR}{MIR + TIR} \quad (1)$$

MODVOLC flags nighttime pixels with NTI greater than −0.8, and daytime pixels with NTI greater than −0.55 as hotspots, because of the large impact of solar reflections and heating on daytime images (Wright et al., 2004; Wright, 2016). These thresholds were found by manual analysis of histograms of NTI at 100 locations to minimize false positive detections (Wright et al., 2004). Another popular approach, the MIROVA algorithm, incorporates a new spectral index in addition to NTI, and spatially filters both spectral indices to improve hotspot detections (Coppola et al., 2016, further details on the MIROVA algorithm and its application in this study can be found in Section 2.4). While these and other algorithms define their own band indices, ratios, spatial filters, and corrections in order to accentuate the differences between hotspot and background pixels, each of these approaches use thresholding to automate the flagging of hotspot pixels. The ability of each algorithm to distinguish hotspots from background pixels depends on how successful their index is in separating the two classes and the accuracy and precision of the threshold set for that index. MODVOLC and MIROVA have successfully generated decades long time series of hotspots at volcanoes across the globe, which has allowed for detection and monitoring of eruptions in near-real time and the study of thermal output from different eruptions and volcanic systems (Wright, 2016; Coppola et al., 2023). Still, both datasets contain false detections and missed hotspots, due to the fact that there will inevitably be non-volcanic thermal signals exceeding the set thresholds, and real volcanic signals lower than the detection thresholds.

In this paper, we aim to enhance the automatic detection of volcanic hotspots in infrared satellite data by applying a convolutional neural network (CNN). CNNs are a machine learning technique commonly employed for image analysis (LeCun et al., 2010). They have been applied to numerous problems in the field of computer vision, including to identify cancer cells in MRIs (El Adoui et al., 2019), facial unlock in cellphones (Apple, 2023), and reverse image search algorithms (Wan et al., 2014). In our approach the use of CNNs can be conceptualized as identifying

hotspots based on what they look like, rather than by thresholding a particular thermal index. While previous methods employ human created indices to highlight hotspot pixels, this approach is data-driven—deriving the spectral and spatial characteristics that define hotspots from a large labeled dataset of the hotspots themselves. In this way, the CNN mimics the pattern recognition of a human analyst.

The type of CNN used here is a U-net (Ronneberger et al., 2015). U-nets are a popular architecture for image segmentation, or tasks in which a prediction is made for each pixel in order to both detect and locate features of interest. A U-net was successfully applied to volcanic hotspot detection in data from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), achieving a high accuracy (Corradino et al., 2023). In this study, we apply a similar method to data from the Visible Infrared Imaging Radiometer Suite (VIIRS) and Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensors. Although ASTER has a finer spatial resolution (90 m in TIR bands, used in Corradino et al., 2023) than VIIRS (375 m) and MODIS (1,000 m), we chose to apply this methodology to VIIRS and MODIS data due to their high acquisition rates and MIR and TIR bands. High acquisition rates result in more frequent opportunities to detect and track changes in volcanic unrest. At the time of this writing, VIIRS sensors provide coverage of each Alaska volcano 8–15 times per day, while MODIS sensors provide coverage 1–6 times per day. Volcanoes at higher latitudes are imaged more frequently than those at lower latitudes by the polar-orbiting satellites used here. Detection frequency will increase in the future with the planned launch of additional VIIRS instruments. Although MODIS has a coarser spatial resolution than VIIRS, it has a longer operational history (satellites Terra and Aqua launched in 1999 and 2002, respectively), so it is useful for studying eruptions prior to the launch of VIIRS (Suomi-National Polar-Orbiting Partnership, SNPP, launched in 2011, and National Oceanic and Atmospheric Administration 20, NOAA-20, launched in 2017).

We incorporate data from eight Alaska volcanoes with a wide range of volcanic thermal signals to develop and test our model for broad applicability to many volcanic settings (Table 1). Alaska volcanoes have frequent eruptions, but are very remote, necessitating remote sensing as a primary method for eruption monitoring, forecasting, and response. We use images of Mount Veniaminof (Alaska) acquired between 2018–2019 covering an effusive-explosive eruption, and images of Mount Cleveland (Alaska) between 2017–2018 with coverage of lava dome growth in order to train our model. The Mount Veniaminof eruption captures high temperature basaltic lava flows into a large, ice-filled caldera (Loewen et al., 2021). Mount Cleveland activity consists of explosions, dome growth, and degassing within the summit crater of a stratovolcano (Werner et al., 2017). These volcanoes are quite different in terms of morphology, eruption style, and governing subsurface processes. They also differ in the source of hotspot detections, namely lava surrounded by ice at Mount Veniaminof, *versus* hot rock surrounded by cold rock at Mount Cleveland. These source differences result in hotspots that may differ slightly in intensity and appearance, leading to a more robust model than it would be if trained on just one of these volcanoes alone.

The other six volcanoes in this study are used for model testing, and were chosen to comprise a wide range of edifice morphologies, magma compositions, eruption frequencies, and eruption styles. These include the frequently erupting and typically mafic volcanoes Okmok Caldera, Shishaldin Volcano and Pavlof Volcano, and the less frequently erupting and typically more silicic volcanoes Augustine Volcano, Bogoslof Island, and Redoubt Volcano. Importantly, all have erupted since the launch of the MODIS sensors. Although our development is focused in Alaska, the volcanoes compiled here range widely in terms of the thermal signatures we expect to identify and the meaning of those signatures in terms of eruptive potential. This dataset can help to evaluate the effectiveness of the model across volcanic systems, and inform future application of the model.

We call the final version of our trained U-net model HotLINK: the Hotspot Learning and Identification Network. After testing and training, HotLINK is applied to VIIRS data from 2012–2022 and MODIS data from 2000–2022 for the eight target volcanoes. The result of these analyses are 22 years of hotspot detections for these volcanoes, 10 years of which have both VIIRS and MODIS observations. We also implement an optimized version of the MIROVA algorithm for our target volcanoes to compare the performance of the machine learning and thresholding approaches. We choose to compare our results with MIROVA because it is one of the most widely used algorithms for global volcanic hotspot monitoring, and was already familiar to the authors. Through this work we hope to improve the accuracy of hotspot detections in infrared satellite data and share our methodology so that it can be applied elsewhere. We aim to address the questions: 1) is a CNN approach able to detect volcanic hotspots in infrared data better than a thresholding approach? 2) Can a computer vision model trained on VIIRS data be reasonably applied to MODIS data with a different resolution? 3) What are the limitations of HotLINK in terms of generalizability to other volcanoes, and detection limits for VIIRS and MODIS, night and daytime images? For each detection we calculate radiative power to quantify the heat emissions over the 22-year study period for the target volcanoes. We then discuss the capabilities and limitations of this approach for volcano monitoring.

2 Methodology

Our model takes as input a VIIRS or MODIS image with MIR and TIR bands, and outputs the probability that each pixel in a central region of the scene contains a volcanic hotspot. Once a hotspot is detected we calculate the total volcanic radiative power (RP in Watts) and area (m^2) of the hotspot. The methodology applied here involves the use of four separate VIIRS datasets to: 1) *train* the network, 2) *validate* hyperparameter selection (i.e., tuning parameters that configure the model and training, as opposed to parameters that are used within the model to make predictions), 3) *test* the model's accuracy when applied to new volcanoes, and 4) *analyze* detections and calculate RP for each volcano over an extended time period. Each of these four datasets (with names italicized above) is assembled for the VIIRS sensor, and additional test and analysis datasets are assembled for the MODIS sensor to produce six datasets in total (Table 2).

HotLINK is trained to detect hotspots in VIIRS infrared images on a manually labeled dataset (VIIRS training) of 3,783 images

TABLE 1 Volcanoes used in this study, in order from west to east. Eruption dates and eruption styles are composited from information available on the Alaska Volcano Observatory website (www.avo.alaska.edu/).

Volcano	Eruptive styles	Eruptions within study period (2000–2022)
Mount Cleveland	Explosive, dome-building	2001, 2005, 2006, 2007, 2009, 2010, 2011, 2013, 2014, 2016, 2017, 2019, 2020
Okmok Caldera	Explosive, phreato-magmatic	2008
Bogoslof Island	Phreato-magmatic, explosive, dome-building	2016–2017
Shishaldin Volcano	Effusive, explosive	2004, 2014–2015, 2019–2020
Pavlof Volcano	Explosive, effusive	2007, 2013, 2014, 2016, 2021
Mount Veniaminof	Effusive, explosive	2002, 2004, 2005, 2006, 2008, 2009, 2013, 2018, 2021
Augustine Volcano	Explosive, dome-building	2006
Redoubt Volcano	Explosive, dome-building	2009

TABLE 2 Datasets used in this study.

Dataset	Labeled	Volcanoes (dates)	Number of images
VIIRS Training	By pixel	Mount Veniaminof (2018), Mount Cleveland (2018–2019)	3,783
VIIRS Validation	By pixel	Mount Veniaminof (2018), Mount Cleveland (2018–2019)	1,275
VIIRS Test	By image	Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (Mar, Jun, Sep, and December 2017)	3,280 (includes 66 ambiguous images moved from the VIIRS validation dataset)
VIIRS Analysis	None	Mount Veniaminof, Mount Cleveland, Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (2012–2022)	160,497
MODIS Test (Aqua)	By image	Mount Veniaminof (2018)	634
MODIS Analysis (Aqua and Terra)	None	Mount Veniaminof, Mount Cleveland, Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (2000–2022)	385,426

of Mount Veniaminof and Mount Cleveland volcanoes. We opt for a manual labeling approach because our goal is to create an automated system that simulates the manual hotspot identification which is done on a daily basis by duty satellite scientists at the Alaska Volcano Observatory (AVO). The same training dataset is used to optimize the thresholds of the MIROVA algorithm (Coppola et al., 2016), and results from both the optimized implementation of the MIROVA algorithm and HotLINK are compared using the same validation dataset, which consists of 1,275 images from the same volcanoes. After training and validation, the accuracy of the model is estimated by applying it to the VIIRS test dataset, which is also manually labeled and consists of images from the six other Alaska volcanoes (Figure 1): Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, and Bogoslof Island.

Although HotLINK is only trained on VIIRS data, we test its applicability to MODIS data simply by inputting the MODIS test dataset into the VIIRS-trained HotLINK model. Data pre-processing for MODIS follows all of the same steps as for VIIRS data (see Section 2.1). Finally, HotLINK is used to detect volcanic hotspots in 10 years of VIIRS data (VIIRS analysis dataset) and 22 years of MODIS data (MODIS analysis dataset) from all eight of the previously mentioned Alaska volcanoes. A subset of the MODIS analysis dataset (MODIS test data, manually labeled for Mount Veniaminof) is reviewed and used to estimate the accuracy of the model when applied to MODIS.

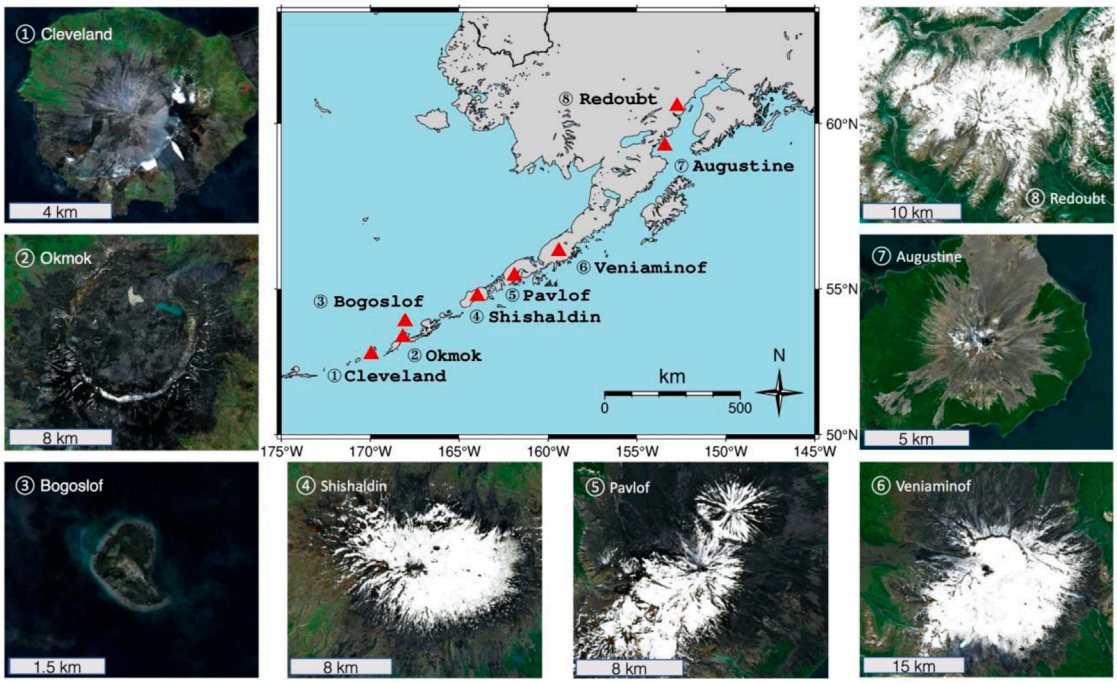

FIGURE 1
Volcanoes used in this study. The map in the center shows all volcano locations in Alaska. Numbered images shows high-resolution satellite data of the volcanoes at various zoom levels, from west to east 1) Mount Cleveland, 2) Okmok Caldera. 3) Bogoslof Island, 4) Shishaldin Volcano, 5) Pavlof Volcano, 6) Mount Veniaminof, 7) Augustine Volcano, and 8) Redoubt Volcano. Satellite data are from Sentinel-2 and composited by CalTopo to provide cloud-free viewing.

2.1 Dataset pre-processing

The pre-processing for all VIIRS and MODIS datasets is the same. First, files containing any of the 8 target volcanoes are downloaded using the Atmosphere Science Investigator-led Processing System API (sips.ssec.wisc.edu) or NASA Earthdata portal (search.earthdata.nasa.gov). Next, terrain and atmospherically corrected radiance data (level 1b) are resampled onto a uniform grid of 64 x 64 pixels centered on the volcano using the nearest neighbor resampling method and the nadir pixel resolution. For VIIRS this corresponds to an area of roughly 24 × 24 km^2 and for MODIS this is an area of 64 × 64 km^2. We use VIIRS image bands I4 (3.55–3.93 μm, MIR) and I5 (10.5–12.4 μm, TIR), and MODIS bands 21 (3.929–3.989 μm, MIR) and 32 (11.77–12.27 μm, TIR). Spectral radiance values have the pixel area (m^2), spectral bandwidth (m), and angular aperture (steradians) factored out of the raw radiative power measurement (W), which allows for direct comparison between data from the two sensors, and normalization using the same factors.

Spectral radiance values (L) are normalized to the minimum (L_{min}) and maximum (L_{max}) possible radiance values for the VIIRS sensor, as determined by scale and offset factors (available in the VIIRS level-1b product user guide; NASA Goddard Space FlightCenter, 2018). Physically, L_{min} and L_{max} represent the limits of the sensor, and possible retrieval values are always within this range. Although the true radiance may be outside this range, the sensor will always return at least L_{min} and will saturate at values greater than L_{max} (NASA Goddard Space Flight Centere, 2018). The equation used to normalize the spectral radiance data is as follows:

$$L_{norm} = \frac{L - L_{min}}{L_{max} - L_{min}} \quad (2)$$

Normalization is important to prevent issues with vanishing or exploding gradients which would make it difficult for the CNN model to converge on a solution (Sola and Sevilla, 1997). We use the same L_{min} and L_{max} for both VIIRS and MODIS data despite the sensors having different minimum and maximum possible spectral radiance values. This is because once the model has been trained on spectral radiance data normalized to a certain range, it must be applied to data normalized in the same way. Lastly, since VIIRS data saturates at a lower spectral radiance than MODIS data, some exceedingly rare MODIS pixels have values higher than one after normalization (<0.002% of pixels in the MODIS test dataset). To remedy this, values are capped at a maximum value of one.

The VIIRS training and validation datasets are assembled by collecting all (day and night) VIIRS data from the SNPP and NOAA-20 satellites with coverage of Mount Veniaminof for the year of 2018 and NOAA-20 VIIRS data (only) with coverage of Mount Cleveland for both 2017 and 2018. These volcanoes and time frames were selected to encompass background non-eruptive behavior, increasing unrest, and eruption. From this dataset, 75% of images are grouped into the VIIRS training dataset, and the remaining 25% are put into the VIIRS validation dataset. The validation dataset is smaller because it is only used to ensure the model is not overfitting, and a representative population is sufficient.

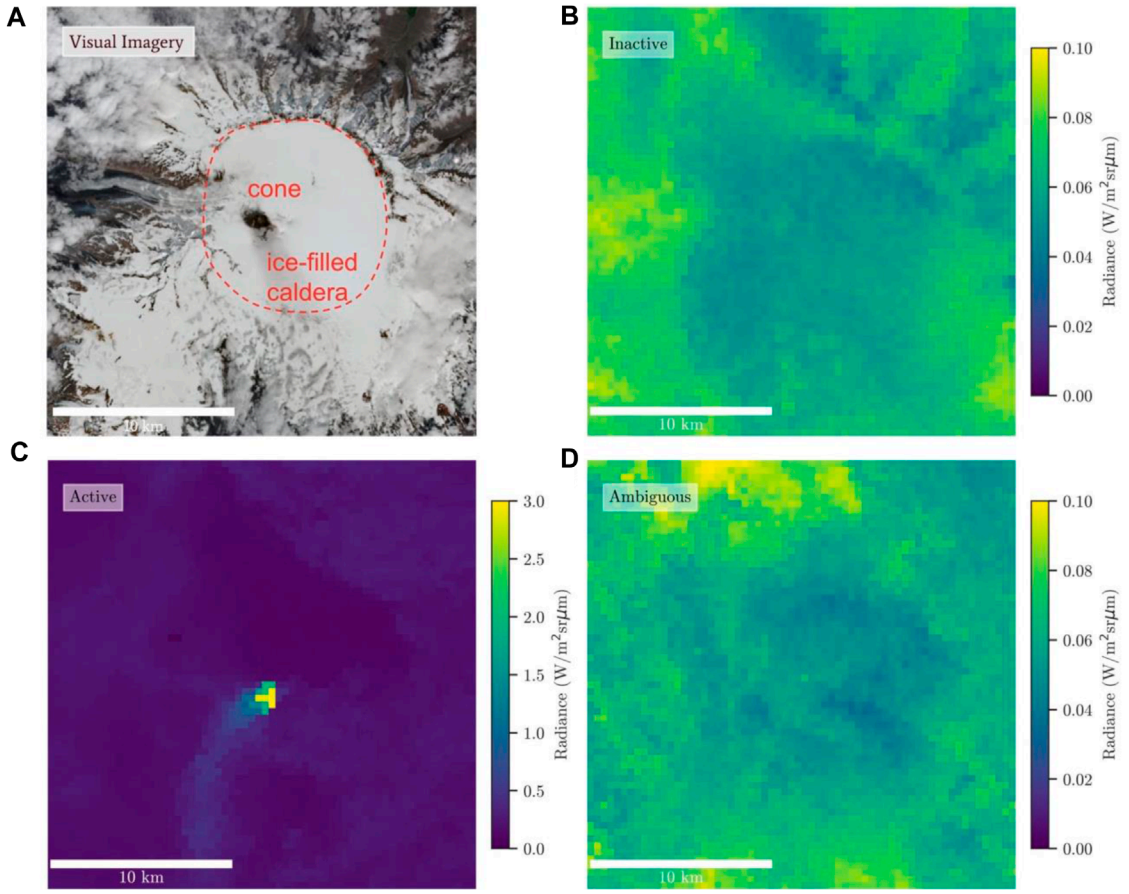

FIGURE 2
Classified example images. **(A)** Sentinel-2 visible RGB image (enhanced natural color visualization) of the ice-filled summit caldera and central cone. The classes of MIR VIIRS images used while training our model shown are **(B)** inactive–not containing a volcanic hotspot as identified by a human analyst, **(C)** active–containing a volcanic hotspot, and **(D)** ambiguous, which could not be confidently categorized into either class. All examples are nighttime images, showing the same cropped region of Mount Veniaminof (24 by 24 km). Note that images C and D have the same color mapping, but image B is scaled differently. Color bars show the range of radiance values in each image.

Whereas the training dataset is larger because data in this group is used to actually train the model, and more data results in better model performance. The grouping between these two datasets is done randomly, with the exception that each image is grouped together with its closest temporal neighbor, since overpasses of SNPP and NOAA-20 satellites can be within −45 min of each other. This prevents having one image in the training dataset and a nearly identical image in the validation dataset.

Images are manually classified into three groups: "active" defined as images containing a volcanic hotspot, "inactive" or images with no volcanic hotspot, and "ambiguous," where we cannot conclusively identify whether or not the image contains a volcanic hotspot (Figure 2). Next, all hotspot pixels within the active-labeled images are identified to construct pixel-wise masks. The ambiguous images are not used for training or validation since we only want images we can characterize with confidence in those datasets. All ambiguous images from the VIIRS validation and training datasets are moved into the VIIRS test dataset, which can have images of any class (66 ambiguous images in total are moved). The final training dataset contains 3,783 images and the final validation dataset contains 1,275 images. In both the VIIRS training and validation datasets, 45% of images are of Mount Veniaminof, 55% are of Mount Cleveland, and 32% of the total are classified as active.

To evaluate how well the model generalizes to other volcanoes not used in training, a test dataset is assembled consisting of 4 months (March, June, September, and December 2017) of VIIRS data for the six additional Alaska volcanoes (Augustine Volcano, Bogoslof Island, Okmok Caldera, Pavlof Volcano, Redoubt Volcano, and Shishaldin Volcano). These months are chosen from throughout the year to capture the full extent of Alaska's seasonal variations. Of our target volcanoes, only Bogoslof Island had an eruption during 2017, so few volcanic hotspots are expected in the VIIRS test dataset. Although choosing data from different volcanoes or times could have yielded more hotspot detections, the volcanoes were chosen with the aim of facilitating future interdisciplinary analysis, and the time period was chosen to ensure standardization across all volcanoes. The resulting dataset is a good indicator of the model's performance when applied to new volcanoes during typical conditions. Images in the test dataset are also manually classified as active, inactive, or ambiguous, but not further classified on a pixel-wise basis. Therefore, the VIIRS test dataset is only used to test the ability of the model to detect images containing hotspots,

not whether it accurately retrieves all of the pixels associated with the hotspot.

The VIIRS analysis dataset consists of the remaining (unlabeled) data, which are analyzed by the trained model and used to generate a hotspot detection time series from 2012–2022 for each of the eight volcanoes in this study. It is the largest VIIRS dataset of our study, consisting of 160,497 individual images of the volcanoes. Note that the VIIRS analysis dataset encompasses data that is already a part of the VIIRS training, validation, and test datasets.

We generate additional MODIS test and analysis datasets in order to test the applicability of our model to MODIS data, compare time series results for VIIRS and MODIS, and extend the time series of detections back to the year 2000. The MODIS test dataset consists of all 2018 MODIS data from the Aqua satellite of Mount Veniaminof classified by image. This volcano and time period were chosen for the MODIS test dataset to encompass a known eruption at Mount Veniaminof that was included in the VIIRS training data. The MODIS analysis dataset consists of all MODIS data from both Aqua and Terra satellites from 2000 to 2022 with coverage of the eight target volcanoes.

2.2 U-net architecture and training

CNNs utilize 3 × 3 (or other sized) matrices, known as convolution kernels, to search for specific patterns within an image (LeCun et al., 2010). The kernel is moved across the image and multiplied with each 3 × 3 subsection to create a new filtered image that shows the degree of correlation between the features of the kernel and the image. This allows the network to identify and locate specific spatial patterns within the image. By stacking multiple layers of convolutions, the network is able to detect increasingly larger and more complex features. At first the network's kernels are populated randomly, but through an iterative training process the kernels are adapted to identify spatial patterns optimized for the task at hand.

Training a CNN involves inputting batches of labeled images into the model. As each image is passed into the model the probabilistic prediction (initially computed by the randomly initialized kernels) is compared to the truth value (the class of each pixel), which is known by prior manual analysis. Then a value, the "loss," is calculated to quantify how well the model prediction compares to the truth value. This is calculated by the "loss function," which, in simple terms, is a quantitative measure of how poorly the model performs—so, a lower loss score indicates better performance. Importantly, the loss function is differentiable with respect to the model—meaning that the gradient of the loss function can be calculated for the entire model. The gradient is very high dimensional, with a value for each trainable parameter of the entire model. By taking a small step in the direction of the gradient, each parameter of the model is adjusted slightly in the optimal direction to decrease the loss, which thereby increases the performance. With each pass over the training dataset, or epoch, each parameter is adjusted slightly, the loss decreases, and the performance of the model improves. This iterative training process is called gradient descent, since the model is descending step-by-step down the gradient of the loss function with the goal of reaching a local minimum. For a more comprehensive explanation of the training, underlying mathematics, and applications of CNNs, see LeCun et al. (2010).

We chose a U-net CNN architecture, because it allows for predictions to be made in the same resolution as the input (Figure 3; Ronneberger et al., 2015). This allows individual pixels to be flagged as hotspots or not. The input for our model is normalized radiance data from the MIR and TIR bands of the VIIRS or MODIS sensor, resampled to uniform resolution and cropped to 64 × 64 pixels centered on the main vent of the volcano of interest (64 × 64 pixels and 2 channels). The output is the probability that each pixel in a central area of the input belongs to one of three classes: background, hotspot, or hotspot-adjacent (24 × 24 pixels and 3 classes). The third class of pixels, hotspot-adjacent, helps the model to train faster; these pixels are considered background pixels during validation and testing. The output region is smaller than the input, due to the fact that convolutions of border pixels are undefined, resulting in a smaller image after each convolution. We consider that a 24 × 24 area of pixels is sufficient for detecting most hotspots (9 × 9 km^2 for VIIRS, and 24 × 24 km^2 for MODIS), but acknowledge that it may miss distal regions of large lava flows, or eruptions which occur far from the main vent.

Many additional parameters can be adjusted in order to alter the architecture, training, or functionality of the model—these are referred to as hyperparameters. We experimented with many of these, selecting the hyperparameters which result in the best performance (as measured by the validation dataset). Parameters that we tested include the random seed and distribution used to initialize the kernels (Glorot and Bengio, 2010), the number of convolutional filters used in each layer (i.e. the width of each rectangle in Figure 3), the gradient descent algorithm (Kingma and Ba, 2014), and the number of training epochs. We also tried many techniques to address the class imbalance in our training dataset. In the VIIRS training dataset approximately 25% of images contain a hotspot, while the remaining 75% do not. We explored several methods to mitigate the effects of the class imbalance, including: oversampling images with hotspots, undersampling the background images, using class weights, and using simple image augmentations to generate more training samples (details in the appendix). Out of all these methods explored, only the image augmentation resulted in an increase in model performance. The rest of this paper only describes the final model, referred to as HotLINK, which uses the best hyperparameters found through dozens of training iterations.

HotLINK is trained on the VIIRS training dataset for 250 epochs, which is the point when the loss ceases to decrease for the validation dataset. During training, input images are augmented using 90° rotations and flips applied randomly after each epoch using the Albumentations library (Buslaev et al., 2020). This produces eight unique orientations for each original input image, which helps the model to learn only the most relevant features for prediction. The model is trained using the Adam optimizer (Kingma and Ba, 2014) with a sparse categorical cross entropy loss function, both of which are a part of the TensorFlow Python library (Abadi et al., 2015). Our U-net took ~2 h to train on a 6-core Intel i7 processor, and after training makes predictions at an average rate of ~5 images per second. Further details on the specific hyperparameters used in the training of the HotLINK model can be found in the code itself, available in the appendix and on GitHub (Saunders-Shultz,

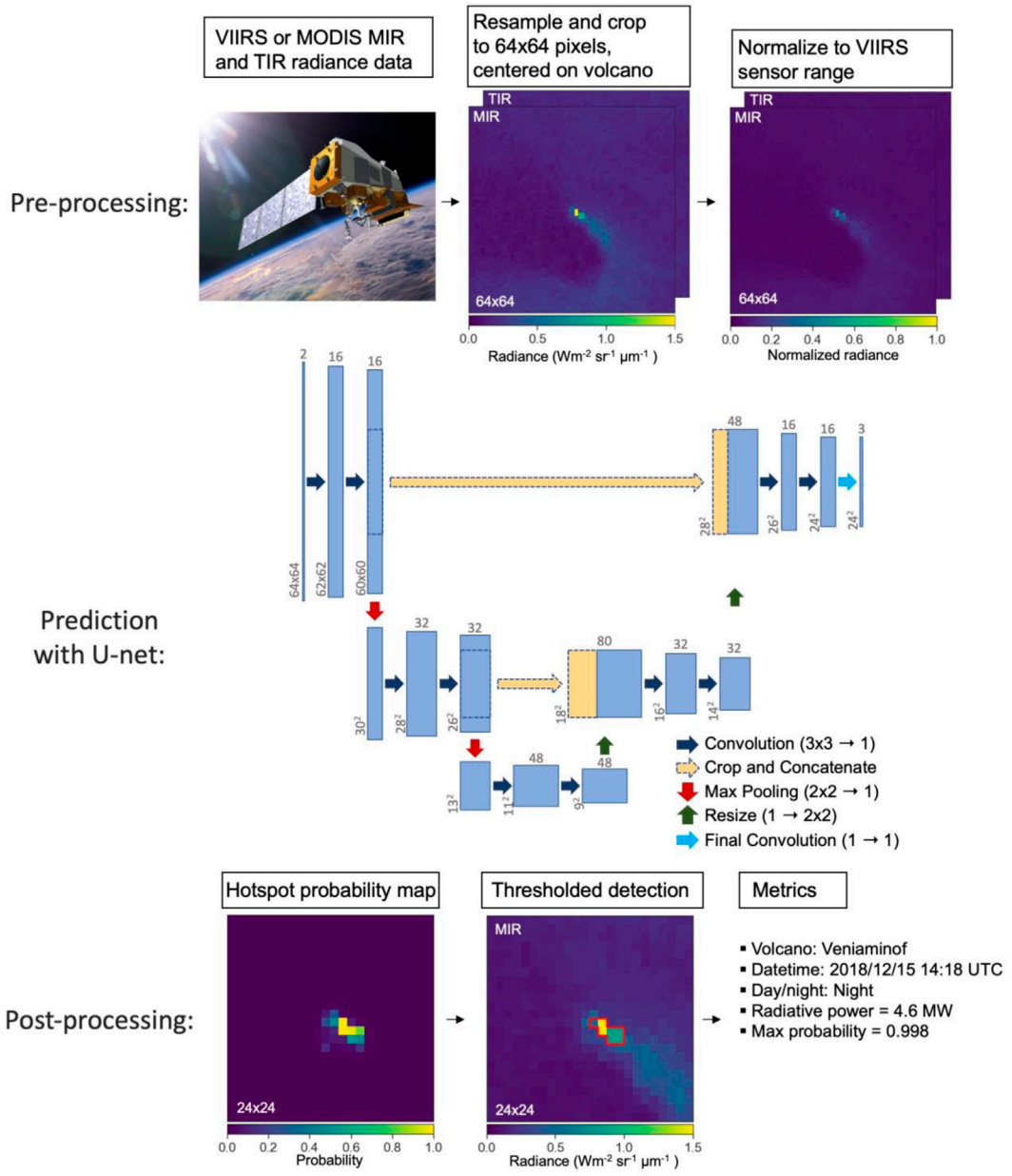

FIGURE 3
Steps of HotLINK processing: pre-processing, prediction with the U-net, and post-processing of a hotspot detection. Blue and tan rectangles of the U-net diagram represent data, the dimensions of which are labeled and denoted by the shape of the rectangles. For example, the input is [64 x 64 x 2] pixels and the output is [24 x 24 x 3] pixels. Note that at each convolution step the height and width of the data are decreased by two, since convolutions on the perimeter pixels are undefined. This progressive loss of perimeter pixels results in a prediction area significantly smaller than the input area. For further description of the motivation and function of the U-net architecture, see Ronneberger et al. (2015).

2023). Although we found these hyperparameters to work best for our problem, they may require modification for other hotspot detection applications.

2.3 Validation and testing

During the training process, we use the validation dataset to try out many different versions of the model in order to test which architectures, model hyperparameters, etc., result in the best hotspot predictions. This process also helps to ensure that the model is learning patterns that are applicable to unseen data and not overfitting. Validation data are also used to tune threshold parameters applied to the output probability maps, and to compare HotLINK and our optimized application of the existing threshold-based algorithm, MIROVA (Coppola et al., 2016). To assess how the trained and validated model performs on new data, we use the test dataset, which is composed

entirely of images from volcanoes the model has not seen during training.

We use two main metrics during validation and testing to evaluate HotLINK and MIROVA's performance: accuracy and F1-score. Accuracy is simply the percentage of images correctly identified by the model. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP, TN, FP, and FN refer to the number of true positives (true hotspot detections), true negatives (true background detections), false positives (erroneous hotspot detections), and false negatives (missed volcanic hotspot detections), respectively, generated by the model. However, accuracy may not be the most appropriate metric for imbalanced datasets, which have higher proportions of some classes than others. For example, in this study a high percentage of images do not contain a volcanic hotspot. Therefore, a high accuracy could be achieved simply by predicting no hotspots in any image. A better metric for evaluating model performance in cases with imbalanced datasets is the F1-score (Ferri et al., 2009), defined as:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \tag{4}$$

The F1-score rewards true positive results and equally punishes false positives and false negatives, while true negatives have no impact on the score. Although our model predicts whether or not each pixel comprises a hotspot, accuracy and F1-scores are calculated on an image-wise basis. Image-wise metrics are used to evaluate the model's ability to detect a hotspot, because image-wise labelling is faster allowing us to create largervtest datasets (Table 2). The training dataset is labeled for each pixel, since the U-net requires every pixel to be labeled in order to train.

Another way to compare HotLINK and our optimized MIROVA algorithm is by using receiver operating characteristic (ROC) curves, which provide a graphical means to characterize the effectiveness of binary classification models (Figure 4). For a given index or predicted probability, an ROC curve plots the true positive rate against the false positive rate achieved by thresholding at different values. In this way it shows the tradeoff between false positives and true positives. For example, setting a low threshold will achieve a high true positive rate at the expense of more false positives, and setting a high threshold will achieve a low true positive rate while providing fewer false positives. ROC curves plot a model's performance at all possible thresholds, thereby showing a particular model's ability to identify hotspots with low FP and FN rates. The ROC curve comparison of HotLINK and MIROVA is further discussed in Section 3.3.

2.4 MIROVA optimization on the VIIRS training dataset

In order to test the performance of HotLINK, we compare our results to the MIROVA algorithm, which was originally developed for use with MODIS data (Coppola et al., 2016). The MIROVA algorithm has already been applied to VIIRS data (Campus et al., 2022, using moderate resolution bands; Aveni et al., 2023, using the same image bands used here). However, these studies use the

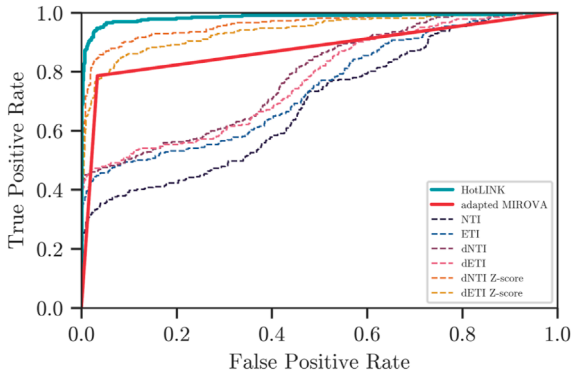

FIGURE 4
Receiver Operating Characteristic (ROC) curve applied to HotLINK and the adapted MIROVA algorithm. HotLINK probabilities are shown in blue, MIROVA prediction is red, and the different indices used in MIROVA are the thinner dashed lines. Preferred classifiers have a high true positive rate (TPR) and low false positive rate (FPR). Note that MIROVA consists of two straight lines because it produces just a binary output.

original thresholds of the MIROVA algorithm that were designed for use with MODIS data. Since VIIRS and MODIS have different spatial resolutions and slightly different spectral bands, it is possible that the original thresholds could be improved for use with VIIRS data. To make a fair comparison between MIROVA's threshold methodology and our model, we optimize the thresholds of the MIROVA algorithm using a grid search over the same VIIRS training dataset that is used to train HotLINK. By using the same training dataset to tune each model and the same validation dataset to evaluate them, we ensure a fair and consistent benchmark between the two approaches. This allows for an unbiased comparison, ensuring that any observed performance differences can be attributed to the inherent capabilities of each model rather than variations in the data they are applied to.

MIROVA employs three thresholds (C1, C2, and K) on multiple indices calculated from the MIR and TIR spectral bands. These indices are the Normalized Thermal Index (NTI), Enhanced Thermal Index (ETI), spatially filtered versions of the first two indices called dNTI and dETI, and the Z-scores of dNTI and dETI. These indices are designed to increase the contrast between hotspot and background pixels, by combining spectral information at each pixel (indices NTI and ETI) with spatial information from surrounding pixels (indices dNTI and dETI) and the scene as a whole (Z_{dNTI} and Z_{dETI}). A full description of the algorithm and definitions of indices are presented in Coppola et al. (2016). In brief, pixels are flagged as active if the index NTI is greater than the threshold K, or if the indices dNTI, dETI, and the Z-scores of both, surpass the C1 and C2 thresholds, respectively:

$$\begin{gathered}(NTI > K)\ or \\ ((\mathrm{dNTI} > C_1)\ \text{or}\ (Z_{dNTI} > C_2))\ \text{and}\ ((\mathrm{dETI} > C_1)\ \text{or}\ (Z_{dETI} > C_2))\end{gathered} \tag{5}$$

In order to optimize MIROVA for use with VIIRS data, we conduct separate grid-searches for nighttime and daytime data to define new threshold values for C1 and C2, which minimize the error

rate on images within the VIIRS training dataset. The daytime grid search is conducted between C1 values of 0.0–0.29 with a stepsize of 0.01, and C2 values of 2.0–11.75 with a stepsize of 0.25. The nighttime samples are more sensitive to the C1 threshold so we use a finer stepsize of 0.005 and smaller range of 0.0–0.095. The C2 range and stepsize remain the same for the nighttime grid search. At each step the accuracy of MIROVA using specific thresholds is calculated. The K threshold was not optimized because it was found to have little effect on the pixel selections made by the algorithm, so it was left as the default value of −0.8 for nighttime images and −0.6 for daytime images. Default MIROVA values for daytime data are C1=0.02 and C2=15, and for nighttime data are C1=0.003 and C2=5. With our grid search we found the highest accuracy using values of C1=0.11 and C2=6.25 for daytime data, and C1=0.075 and C2=5.25 for nighttime data (see Supplementary Figures A.3 and A.4 in the appendix for visualization of both grid searches). The grid searches demonstrate that slight changes to threshold values can result in slight increases in the performance of MIROVA, at least when applied to our particular dataset.

2.5 Hysteresis thresholding and radiative power calculation

Some final considerations for implementing the model are choosing how to threshold pixels in the output probability map (Figure 3), and then calculating useful metrics for each detection to better track changes in volcanic thermal emissions over time. Although each pixel is predicted with an individual probability, we recognize that a pixel is more likely to be a hotspot if it is adjacent to a hotspot pixel. For that reason, we implement hysteresis thresholding, in which a high threshold is used to initialize hotspot detections and a lower threshold is used to continue them. Here, all pixels with a probability greater than 0.5 are classified as hotspots, and pixels with a probability greater than 0.4 are classified as hotspot pixels if they are adjacent to other hotspot pixels. The high threshold is set by optimizing the validation dataset for image F1-score, and then the low threshold is set by optimizing for pixel-wise F1-score. To clarify, these metrics are chosen because only the high threshold determines which images are active, while the low threshold determines which pixels within the image are active.

Once active images are detected and all hotspot pixels within those images are identified, radiative power (RP) is calculated following the method of Wooster, (2003), using the following formula:

$$RP = C \times A_{pix} \times \sum^{n} L_{pix} - L_{BG} \quad (6)$$

where RP is the radiative power measured in Watts, C is a constant of proportionality that is specific to the sensor ($sr^{-1}\mu m^{-1}$, 18.9 for MODIS and 17.34 for VIIRS), A_{pix} is the area of the pixel in kilometers squared (1 km^2 for MODIS, 0.14 km^2 for VIIRS), n is the number of pixels in the hotspot, L_{pix} is the radiance of each hotspot pixel ($Wm^{-2}sr^{-1}\mu m^{-1}$), and L_{BG} is the mean radiance of pixels directly surrounding the hotspot detection ($Wm^{-2}sr^{-1}\mu m^{-1}$, following the established methods of Wooster, 2003). RP is a measure of how much energy is released over the entire hotspot, and includes corrections for pixel size, central wavelength, and background radiance. Since pixel size and central wavelengths are different for VIIRS and MODIS, using RP allows us to make direct comparisons between the two sensors.

3 Results

3.1 Validation and test results

Results on the VIIRS validation dataset (Table 3) show that the final model works well when applied to data that has not been seen during training but comes from the same volcanoes. Specifically, both Mount Veniaminof and Mount Cleveland validation data yield model accuracies >95% and F1-scores >0.9.

On the VIIRS test dataset, which includes data from the six volcanoes that the model has not seen previously, HotLINK achieves a relatively low F1-score of 0.667 (Table 3). This seemingly poor performance is best explained by the lack of true hotspots in the dataset used; out of the six volcanoes, only Bogoslof Island erupted during the sampling period of the test dataset (Table 2). Since F1-score is mainly a function of true positive detections we achieve a poor score on most of the volcanoes since there were not many true hotspots to detect. False negative and false positive rates on all datasets do not exceed 4%, except for the Augustine Volcano false negative rate, which is 7.9%.

3.2 HotLINK results on MODIS test data

The MODIS test dataset consists of all Mount Veniaminof data from the Aqua satellite in 2018, including 634 images in total. HotLINK achieves an accuracy of 98% on the MODIS test dataset, and an F1-score of 0.95 (Table 3). Unexpectedly, this performance is better than the model performs on VIIRS data. In section 4.3 we discuss a possible explanation for this.

3.3 HotLINK and adapted MIROVA algorithm results on the VIIRS validation dataset

The VIIRS validation dataset is used to compare the results of HotLINK and the optimized MIROVA algorithm after both models are trained/optimized with the VIIRS training dataset. On the validation dataset, we find that HotLINK outperforms our implementation of the MIROVA algorithm in all metrics (Table 4). Specifically, HotLINK produces more true positives (fewer missed detections), and more true negatives (fewer false detections) than the MIROVA approach. Both methods score higher on nighttime data than daytime data. The conditions under which each model performs best is further discussed in Section 4.4.

The ROC curve (Figure 4) further demonstrates that HotLINK (blue line) outperforms the MIROVA algorithm implementation (red line) with respect to true and false positives. In this plot, preferred classifiers have a high true positive rate (TPR) and low false positive rate (FPR). So better classifiers are those which plot further into the top left corner. These results show that HotLINK performs better than the overall optimized MIROVA algorithm,

TABLE 3 HotLINK results on training, validation, and test datasets. Metrics shown are: accuracy, F1-scores, ratio of True Negatives, True Positives, False Negatives, and False Positive detections, and the total count of images used to calculate the metrics. Each row shows the average of all volcanoes first, and then specific values for each volcano in the dataset. Note that ambiguous images (195 total) are removed prior to this analysis.

Dataset	Accuracy	F1-score	TN	TP	FN	FP	Count
VIIRS Training	0.952	0.914	0.698	0.254	0.031	0.017	3,781
Mount Cleveland	0.962	0.898	0.795	0.167	0.017	0.021	1,551
Mount Veniaminof	0.945	0.920	0.631	0.314	0.041	0.014	2,230
VIIRS Validation	0.962	0.923	0.731	0.231	0.022	0.016	1,275
Mount Cleveland	0.977	0.933	0.820	0.157	0.011	0.011	527
Mount Veniaminof	0.951	0.919	0.668	0.282	0.029	0.020	748
VIIRS Test	0.947	0.667	0.908	0.049	0.024	0.019	2,956
Augustine Volcano	0.914	0.172	0.901	0.009	0.079	0.011	547
Bogoslof Island	0.955	0.892	0.765	0.189	0.024	0.022	460
Okmok Caldera	0.956	0.512	0.927	0.024	0.024	0.026	468
Pavlof Volcano	0.974	0.723	0.936	0.037	0.008	0.019	483
Redoubt Volcano	0.919	0.608	0.940	0.040	0.002	0.019	530
Shishaldin Volcano	0.979	0.444	0.970	0.009	0.002	0.019	468
MODIS Test (Mount Veniaminof)	0.981	0.954	0.786	0.195	0.019	0.0	646

TABLE 4 Comparison of HotLINK and the adapted MIROVA algorithm on the VIIRS validation dataset. Metrics shown are: accuracy, day/night/combined F1-scores, and ratio of True Negatives, True Positives, False Negatives, and False Positive detections.

Model	Accuracy	F1-score	Night F1-score	Day F1-score	TN	TP	FN	FP
HotLINK	0.962	0.923	0.929	0.916	0.731	0.231	0.022	0.016
Adapted MIROVA algorithm	0.921	0.834	0.894	0.765	0.722	0.198	0.054	0.025

as well as all of the individual indices used by the MIROVA algorithm (thin dashed lines) with respect to TPR and FPR. This indicates that HotLINK is able to better differentiate hotspot and background pixels in comparison with individual indices, regardless of threshold selection. This is due to the CNN's ability to extract additional spatial information compared to manually tuned spatial filters.

3.4 Time series results

After applying HotLINK to the validation and test datasets, we apply HotLINK to the VIIRS and MODIS analysis datasets. This provides 10 years of VIIRS and 22 years of MODIS hotspot detections for the eight target alaska volcanoes. These results can be found in Figure 5. Despite being unlabeled, these results can help provide a qualitative check on the effectiveness of the model when applied to different volcanoes experiencing background, unrest, or eruptive behavior. All detections found in this dataset are plotted as time series in Figure 5, with the Alaska Volcano Observatory (AVO) Aviation Color Code as the background color. In this analysis we use the AVO Aviation Color Code as a proxy for the state of activity of the volcano. A color code of "green" is used to indicate that a volcano is at a background non-eruptive state, "yellow" indicates increasing unrest with the possibility of an eruption in the future, "orange" indicates that effusive or low-level explosive eruptions are occurring or are expected in the immediate future, "red" indicates a significant explosive eruption is occurring or imminent, and

"unassigned" (colored as gray in Figures 5, 7) indicates that there is insufficient ground-based monitoring data to assign a color code (Guffanti and Miller, 2013). While accuracy metrics are useful, the time series plots demonstrate the utility of HotLINK in practical applications. Figure 5 illustrates that HotLINK succeeds at detecting eruptions, which are accompanied by significant increases in the frequency and RP of detected hotspots. This figure also shows patterns of potential false positive detections during non-eruptive periods at all volcanoes, which are discussed in the following paragraphs.

Mount Cleveland erupts frequently, as indicated by many periods of orange color code in the timeline (Figure 5), which represent lava dome eruptions and other elevated activity (e.g., Werner et al., 2017). The Mount Cleveland time series shows numerous hotspot detections, which are much more frequent during periods of orange color code compared to when the color code is unassigned.

Okmok Caldera had only one eruption during our analysis period, in 2008. Only MODIS data is available for this eruption, from which there was one nighttime and three daytime detections during the eruptive period all with RP values >5 MW. Steady detections occur in VIIRS night and daytime data at Okmok Caldera, which we infer may be due to the presence of lakes within the caldera.

At Bogoslof Island we see a strong seasonal trend, in which VIIRS daytime detections and associated RP increase in the summer and decrease during winter. These seasonal trends are observable both before and after the 2017 eruption, but are stronger post-eruption. The 2016–2017 Bogoslof Island eruption is captured well, with VIIRS nighttime detections producing higher RP values than at any other time.

At Shishaldin Volcano, extended eruption periods from 2014–2016 and 2019–2020 are tracked well by HotLINK detections. The onset of these eruptions are accompanied by significant increases in the rate and RP of detections, and the end of eruptions are accompanied by a return to background values.

Pavlof Volcano eruptions are detected well by the HotLINK system, with RP values during eruptive episodes significantly higher than during non-eruptive periods. The 2007 eruption is captured well in MODIS data, and subsequent eruptions are captured well in both VIIRS and MODIS data.

At Mount Veniaminof there have been multiple eruptions that are detected by HotLINK, but there is also a high rate of background detections, which could either be indicative of background heat output or potentially the emissivity and thermal inertia differences between the active cone and surrounding glacier. In Section 4.4 we further discuss the nature of these signals.

Augustine Volcano had one observed eruption in 2006. Augustine Volcano has infrequent VIIRS nighttime detections, but does show a seasonal signal with increased VIIRS daytime detections during winter and increased MODIS daytime detections during summer.

Redoubt Volcano also had only one eruption during our analysis period, in 2009, which was detected well in MODIS data. Since then, no anomalous thermal activity has been detected but there have been frequent hotspot detections in VIIRS nighttime and daytime data, which may be attributed to localized persistent degassing and snow melt on the 2009 lava dome.

4 Discussion

In this section we discuss the time series results at all volcanoes to investigate the strengths and weaknesses of our model. We also discuss the probabilistic output of HotLINK, and our finding that probabilities are well calibrated. Next, we compare VIIRS and MODIS applications of HotLINK, and estimate detection limits for each sensor. Finally, we advance our comparison of HotLINK and the threshold-based MIROVA algorithm by looking at a case study of the Mount Veniaminof time series.

4.1 Analysis of time series results from all volcanoes

Based on the time series of detections at all volcanoes (Figure 5), we find that 1) the HotLINK model, as currently trained, works well for many, but not all volcano morphologies/settings, 2) the VIIRS sensor has a lower detection limit than MODIS due to a finer spatial resolution, which also results in a slightly higher false positive rate for VIIRS, and 3) the RP and relative frequency of daytime and nighttime detections reveals distinct periods in the eruptive chronologies at many volcanoes, which can be used to further discern true and false detections. We discuss how we can discern true and false hotspot detections during non-eruptive periods at volcanoes, why false positive detections appear more often in some volcanoes during certain times of the day and year than others, and how results can be further filtered to remove many of the false detections.

Although HotLINK has a lower false positive rate than MIROVA in the validation dataset (Table 4), in the analysis dataset we still see nearly continuous hotspot detections at all volcanoes even between eruptive periods (Figure 5). Even though HotLINK makes many detections when volcanoes are at "green," or a background state (e.g. Okmok Caldera 2012–2022), that does not mean that all of those detections are false positives as it is common for many volcanoes to be persistently degassing and producing heat at the surface even in absence of an eruption. In this case, increases in the rate and RP of detections, rather than the detection of a single hotspot, may indicate volcanic unrest or eruption. However, as testing shows (Table 3), we expect HotLINK to have a false positive rate ~2%, such that some of the detections during background periods are likely not true volcanic hotspots.

In our analysis of Figure 5, we expect true volcanic hotspot detections to be those which are spaced closely together in time and at higher RP than other detections observed during periods with no eruptive activity. At all volcanoes, likely false positives seem to occur in VIIRS daytime images with RP in the range of ~1–10 MW, and in VIIRS nighttime images with RP ~0–0.5 MW. We determine that most detections with RP above these thresholds are true positives, but that does not preclude the possibility of true (but weak) volcanic hotspot detections within those ranges.

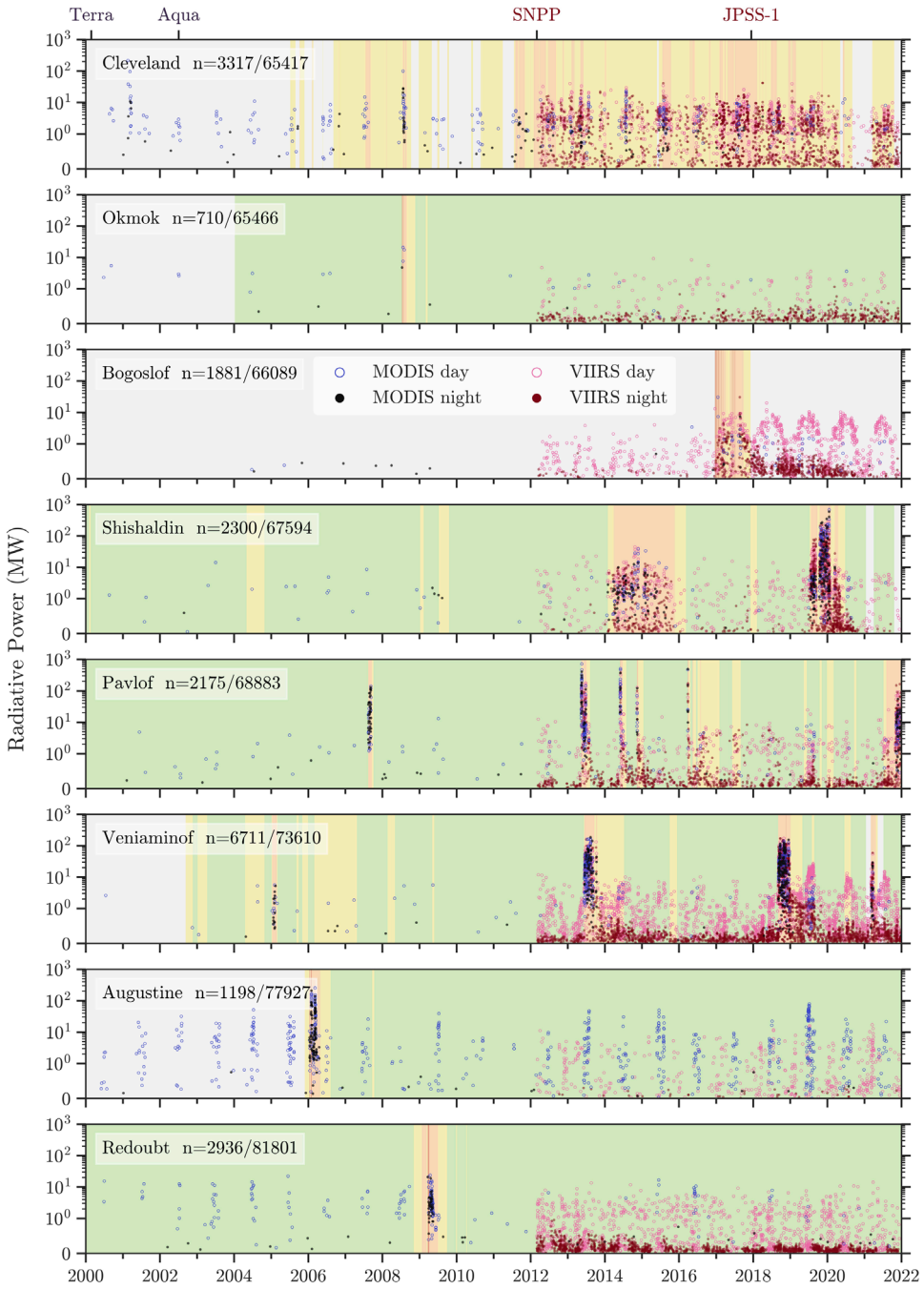

FIGURE 5
Time series results of HotLINK detections and calculated radiative powers for all eight target volcanoes: Mount Cleveland, Okmok Caldera, Bogoslof Island, Shishaldin Volcano, Pavlof Volcano, Mount Veniaminof, Augustine Volcano, Redoubt Volcano. The AVO color code at each volcano is shown as the background color of each figure for general context on the state of activity at the volcano (see Section 3.3 for description of color codes), with gray indicating a period with insufficient monitoring data for AVO to assign a color code ("Unassigned; " Guffanti and Miller, 2013). The RP of individual hotspot detections are shown as points (MODIS = black, VIIRS = red), with solid and open points representing night and daytime acquisitions, respectively. Next to the name of each volcano is the number of total detections at each volcano, and the number of total images for each volcano in both the VIIRS and MODIS analysis datasets. Note that for all plots the *y*-axis scale is linear between 0 and 1 MW, and logarithmic >1 MW. The top axis shows the start of data acquisition from satellites used.

At some volcanoes (Bogoslof Island and Augustine Volcano) there are notable seasonal variations in the number of detections and the RP of those detections. At these volcanoes we believe the source of these detections is primarily from diurnal effects on land/water boundaries. For example, both Bogoslof Island and Augustine Volcano are island volcanoes, which means that during the day the land surface regularly heats up more than the surrounding ocean, creating a temperature difference that is visible in infrared images, centered on the volcano, and thus to our model looks like a volcanic hotspot. Since Bogoslof Island is

~1.5 km in diameter while Augustine Island is ~12 km in diameter, Bogoslof Island tends to appear more like a hotspot in daytime VIIRS data while Augustine Volcano Island regularly is identified as a hotspot in daytime, summer, MODIS data (Figure 5). Similarly, clouds frequently develop during the daytime on land, creating localized solar reflections.

A similar effect occurs at volcanoes that have crater lakes/lagoons (e.g., Okmok Caldera and Bogoslof Island). Since water has a higher thermal inertia than land, it preserves solar heat longer into the night than land and is commonly warmer than land at night, particularly when the land is snow-covered. Volcanic lakes are commonly connected to hydrothermal systems and increasing lake temperature can be linked to volcanic activity (Hurst et al., 1991; Rouwet et al., 2014). However, increasing lake temperatures due to volcanic thermal input are difficult to distinguish from increasing temperatures due to diurnal patterns. With that in mind, a hotspot detection of a lake is not necessarily indicative of increased volcanic or hydrothermal activity. By looking at trends in detections and RP over time, however, HotLINK may have the capability to characterize background lake temperatures and thus detect deviations above background. In our data we did find clear examples of diurnal and seasonal cycles in hotspot detections at Okmok Caldera and Bogoslof Island. However, in neither case did we observe clear deviations in the background radiative power that might have been caused by increased volcanic activity. Example images of false detections at Okmok Caldera and Bogoslof Island and comparison with high resolution true color imagery are available in the Supplementary Material. Other common effects producing non-volcanic hotspot detections are snow melting off rocky areas that then become solar-heated (Mount Veniaminof), and clouds or volcanic plumes reflecting solar radiation.

While these non-volcanic sources of apparent hotspots are considered in our study to be false-positives, they highlight the capability of HotLINK to detect subtle warming signals that could be successfully applied to other research problems. Fundamentally there will always be a tradeoff between the sensitivity of the method to detect real volcanic hotspots, and the number of false positives produced. With this in mind, there are simple ways to minimize the occurrence of the false positives in the dataset through filtering. One easy approach is to only use the nighttime data, which is much less susceptible to false positives, especially those occurring on exposed rocks surrounded by snow and ice fields and solar reflection off clouds or plumes. Another way is to set a specific probabilistic threshold. In Figure 5, we calculated radiative power for all images containing any pixels whose probability exceeds 0.5. However, this probability could be adjusted for different contexts. For example, if conducting a long-term historical analysis, it may be better to set a high confidence threshold and remove as many false positives as possible. Conversely, for near-real-time monitoring it may be important to incorporate as many detections as possible, even if a greater percentage of them might be false.

To illustrate the effects of further filtering the data, we look at time series from Bogoslof Island, Okmok Caldera, Redoubt Volcano, and Augustine Volcanoes, each of which only had one eruption during the time period of study. At all four of these volcanoes combined there are 6,725 total detections made out of 291,283 total images analyzed (Figure 5). These statistics yield a combined detection rate of 2.3% (>97% of images are non-detections). However, if we use only night time data and set a probabilistic threshold of 0.75 at the same volcanoes, HotLINK detects 2,661 hotspots out of 168,400 total images, which is a detection rate of 1.6%. So, with a higher threshold and only using nighttime images HotLINK removes >98% of images as non-detections. These statistics also help us estimate an upper bound on the false positive rate of HotLINK at around 2%, which is similar to what we calculated earlier with the VIIRS test dataset. For comparison to detection rates during eruptions see Section 4.3 in which detection rates of VIIRS and MODIS sensors at Mount Veniaminof during eruptive periods are discussed.

4.2 Analysis of HotLINK probability estimates

In order to use probabilistic predictions from HotLINK for filtering hotspot detections, or for future incorporation into forecasting methods, we must verify that the probabilistic predictions of the model are meaningful. This is especially relevant since modern neural networks have shown a tendency to be overconfident (Guo et al., 2017). Although the model outputs a probability prediction for each pixel in the image, we are most interested in whether the image contains a hotspot at all. Therefore, for the purposes of this analysis we refer to "image probability" as the highest probability of all pixels in the image, since it only takes one hotspot pixel for an image to be classified as active. We evaluate our probability outputs using a reliability diagram, adapted from Hamill, 1997; Figure 6A.

For image probabilities to be well calibrated, we want the accuracy of a thresholded prediction to scale with its probability (Hamill, 1997). For example, if a well-calibrated model predicts five images to contain hotspots at a probability of 80%, four of the images would contain hotspots while one would not. While this may seem counterintuitive, we want some images with high probabilities to be wrong in order to confirm that probabilistic predictions are reliable. We find a strong correlation between the probabilities of HotLINK predictions and whether images contain a hotspot, since they align with the ideal distribution (black line) shown in the reliability diagram below (Figure 6A). This demonstrates that the probabilistic output of HotLINK can be considered a well-calibrated estimate.

While the reliability diagram (Figure 6A) demonstrates that probabilities are well calibrated, we can expand our probabilistic analysis by including the ambiguous images identified by human visual inspection. The ambiguous images contained in the VIIRS validation and test datasets present a great opportunity to compare HotLINK's probability predictions to images we could not confidently classify as volcanic or not. Figure 6B shows that ambiguous images are skewed toward low probabilities, with ~50% of ambiguous images predicted at a probability <0.1. However, ambiguous images are proportionally more represented than each other class in all bins from 0.1–0.8. In other words, ambiguous images are much more likely to be predicted at intermediate probabilities than images labeled as inactive or active. This finding supports the idea that CNNs mimic the visual learning of human experts. It also provides more confidence in the quality

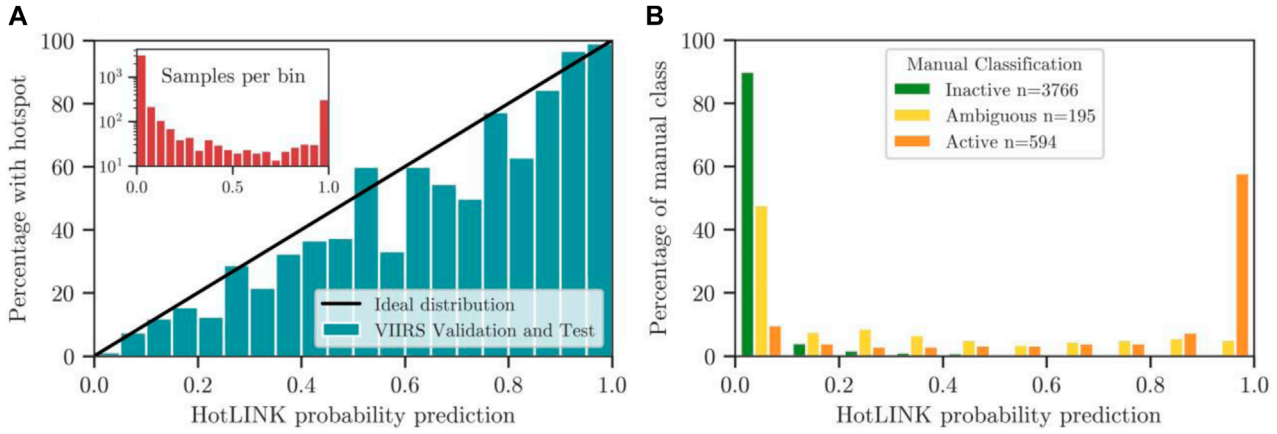

FIGURE 6
Reliability diagram and histogram of VIIRS validation and test datasets. **(A)** Reliability diagram of the HotLINK model applied to the VIIRS training and validation dataset (unambiguous images only). Blue bars represent the proportion of images manually identified as active in 5 percentile bins. The black line represents the ideal probability distribution, indicating that probability predictions are accurate to the true classification. Bars below the black line are overconfident (probability prediction of hotspots is higher than the true probability), and bars above are under-confident (probability prediction of hotspots is lower than true probability). The inset figure shows the number of samples per bin on a logarithmic scale. **(B)** Histogram of the VIIRS validation and VIIRS test datasets, showing the percentage of each class - inactive (green), active (orange), and ambiguous (yellow) - in 10 percentile bins. Ambiguous images are the most represented class at intermediate probabilities (0.1–0.8).

of probabilistic predictions, since images that appear ambiguous to analysts are likely to be predicted at intermediate probabilities by the network.

4.3 Comparison and detection limits of MODIS and VIIRS data

We speculate that the higher accuracy of HotLINK on the MODIS test dataset relative to the VIIRS test and validation datasets is due to the larger pixel size of MODIS preventing small hotspots from being identified by either HotLINK or manual analysis, resulting in an increased number of true negatives for MODIS compared to VIIRS. Similarly, the larger pixel size blurs out smaller scale background variance that is visible in VIIRS data, such that MODIS has a lower false positive rate than VIIRS and a higher F1-score. The larger pixel size of MODIS data results in fewer detections overall than VIIRS.

HotLINK shows a slightly better accuracy on MODIS data than on VIIRS because the MODIS data contains a greater proportion of true negatives and a smaller proportion of false positives. Despite this, the VIIRS data has a higher true positive rate and is able to see smaller and weaker hotspots. To further support this conclusion we compare VIIRS and MODIS detections during three eruptive events at Mount Veniaminof from the analyzed data. From these eruptions we also attempt to quantify a night and daytime detection limit for HotLINK when applied to VIIRS and MODIS data.

Mount Veniaminof had three eruptions between 2012–2022, the time period when both VIIRS and MODIS data are available. These eruptions were effusive-explosive in nature, characterized by lava effusion into and within the intra-caldera glacier, and sporadic ash emissions (Loewen et al., 2021; Waythomas, 2021; Waythomas et al., 2023). Start and end dates for these eruptions are taken from Loewen et al. (2021). During these eruptive periods, both VIIRS and MODIS agree well on RP estimates in our analysis. For the 2013 eruption (June 13—October 17), both MODIS and VIIRS retrieved an average RP of 27.8 MW. During the 2018 eruption (September 4—December 27) MODIS retrieved an average of 27.6 MW and VIIRS 30.2 MW, and for the 2021 eruption (February 28—April 21) MODIS retrieved an average of 6.0 MW and VIIRS 5.0 MW (Figure 7).

Although the average RP retrieved by both sensors is comparable, the VIIRS sensor had a much higher rate of detections during the same eruptive periods. Across all three eruptions, VIIRS had 1,553 detections out of 2,874 total images, for an active percentage of 54%. Meanwhile MODIS had 536 detections out of 1,902 total images, for an active percentage of 28%. We hypothesize VIIRS had a greater active percentage because it was able to capture significantly weaker signals, due to its finer spatial resolution (0.137 km^2 compared to 1 km^2 pixel area at nadir). In future work, this hypothesis could be tested through a more robust analysis of the relative detection rate of VIIRS and MODIS images that are captured at nearly the same time.

To approximate detection limits for both sensors using HotLINK, we use the 5^{th} percentile radiative power of all hotspots detected during the 2013, 2018, and 2021 eruptions at Mount Veniaminof. It is important to acknowledge the possibility of false positives in these data, constituting approximately 2% of samples according to the labeled VIIRS validation and test datasets (Table 3). To mitigate the impact of false positives on the detection limit estimate, we opt for a conservative approach by using the 5th percentile, which is more than twice the estimate for the percentage of false positives in the dataset. This ensures that potential low RP false positives do not artificially lower the detection limit estimate. Still, our estimate for detection limit is not the threshold at which signals are missed, but approximates this by indicating the weakest signals retrieved by HotLINK. This estimate allows us to compare the relative detection limits between sensors. For VIIRS data, we find the 5th percentile of daytime detections to be 0.69 MW, and nighttime detections to be 0.26 MW. For MODIS data, we

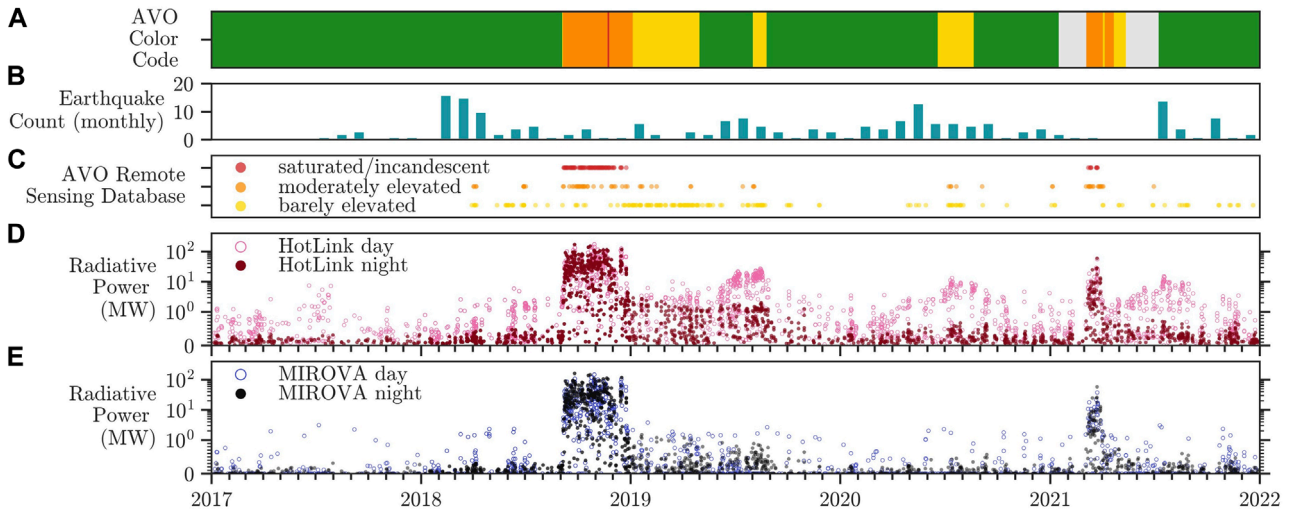

FIGURE 7
Multidisciplinary observations at Mount Veniaminof. Subplots show: **(A)** AVO Color Code timeseries, with color code levels indicated by their respective colors, and gray indicating periods with insufficient monitoring data for AVO to designate a color code (Guffanti and Miller, 2013). **(B)** Monthly earthquake counts within 20 km of Mount Veniaminof, assembled from the USGS ComCat earthquake database (https://earthquake.usgs.gov/earthquakes/search/). **(C)** Analyst flags from the AVO remote sensing database, showing analyst identified thermal signals in VIIRS images, characterized as being "saturated," "moderately elevated," and "barely elevated." **(D)** Mount Veniaminof hotspot detections in VIIRS images from 2017 to 2022 using HotLINK and **(E)** hotspot detections from the adapted MIROVA algorithm.

find the 5th percentile of daytime detections to be 1.4 MW, and nighttime detections to be 0.79 MW. These results demonstrate that HotLINK is 1.8–3 times more sensitive to nighttime observations than daytime observations, and that HotLINK is 2–3x more sensitive when applied to VIIRS data compared to MODIS. To compare with literature values, the MIROVA algorithm applied to MODIS data cites a detection limit of −1 MW irrespective of the time of day (Coppola et al., 2020). This is the first time the authors are aware of a comparison of the detection limits between MODIS and VIIRS I-bands, although the radiative power between MODIS and VIIRS M-bands (750 m at nadir) have been previously compared, finding that the VIIRS M-bands are more sensitive than MODIS bands to thermal signals (Li et al., 2018; Campus et al., 2022). We caution that these detection limits are only approximations, since we are only using one volcano for this analysis and are not looking at the radiative power of missed detections. Detection limits could be more rigorously ascertained by comparing the radiative power of true positive and false negative detections across many volcanoes. Here we only calculated the radiative power for images that were detected as hotspots by HotLINK and statistical analysis of the RP of false negative detections was not done.

4.4 Analysis of HotLINK and adapted MIROVA on the Mount veniaminof time series

Table 4 shows a higher true positive rate of HotLINK relative to our implementation of the MIROVA algorithm, indicating a greater sensitivity to smaller and lower temperature hotspots. Similarly, the high true negative rate of HotLINK relative to this adapted MIROVA indicates that HotLINK is less susceptible to false positive detections. We can expand on this analysis by examining the Mount Veniaminof time series from 2017–2021 to further compare results during eruptive and inter-eruptive periods (Figure 7). During this time period there were two eruptions, one in 2018 and one in 2021. The main difference between HotLINK and the optimized MIROVA detections during this period is that HotLINK detects more hotspots. From an eruption tracking perspective, the MIROVA algorithm does well as it has a similar detection rate as HotLINK during eruptions. In contrast, during non-eruptive periods HotLINK makes a greater number of detections than MIROVA, which may represent volcanic thermal output associated with volcanic unrest, as well as false positives. Therefore, while both models perform well for eruption detection and tracking, HotLINK is able to detect weaker signals that may be relevant for monitoring unrest at Mount Veniaminof.

Figure 7 shows an increase in HotLINK detected RP prior to the 2018 eruption, and more peaks in 2019 and 2020 that are not seen in MIROVA data. These HotLINK detections are consistent with AVO analyst checks of VIIRS MIR images, where analysts observed weakly to moderately elevated surface temperatures qualitatively prior to eruption at Mount Veniaminof, and again during discrete time periods in the summers of 2019 and 2020 (Figures 7C,D; Cameron et al., 2018; Orr et al., 2023). We therefore find that the HotLINK detections are real, capturing weaker, but notable above-background thermal signals as seen in both the rate and radiative power of detections. These HotLINK results also have the advantage of providing quantitative information in comparison to the qualitative AVO remote sensing database classifications of "barely elevated," "moderately elevated," and "saturated/incandescent."

Inspection of these signals through complementary high resolution optical satellite imagery (e.g. Sentinel-2, Maxar) suggests that they comprise a combination of subtle surface heating, potentially due to increased vent degassing behavior at the volcano, as well as a seasonal signature due to the still-warm 2018 lavas

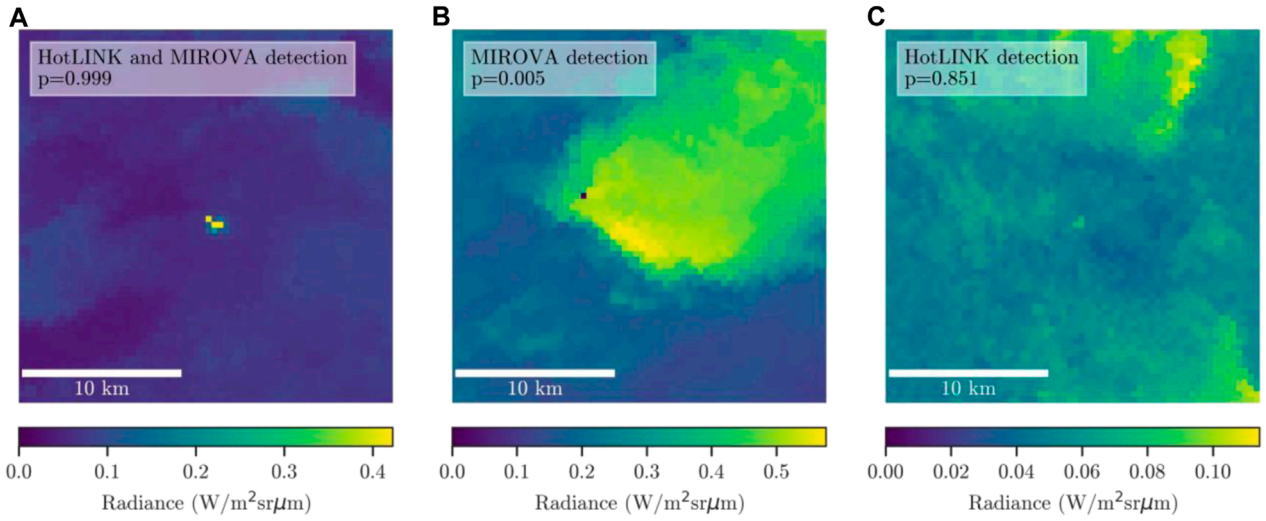

FIGURE 8
Example images from the VIIRS validation dataset. All images show MIR spectral radiance (Wm^{-2} sr^{-1} μm^{-1}) at Mount Veniaminof. **(A)** a true hotspot detection made by both HotLINK and the adapted MIROVA algorithm (nighttime image), **(B)** a false positive detection of a bright cloud made by the adapted MIROVA algorithm (daytime image), and **(C)** a true positive detection of a more subtle hotspot made by HotLINK, which is missed by the adapted MIROVA algorithm (night image). All images are 64 x 64 pixels, or roughly 24 x 24 km. Note that each image has its own colorbar scale in order to show the maximum contrast within each image.

readily melting the overlying snow cover in spring. The 2018 pre-eruptive hotspot signals suggest increased thermal output, perhaps via increased degassing or ground surface temperatures of the active cone (Orr et al., 2023). The 2019 and 2020 peaks in RP coincide with seasonal snow melting that exposed the large and relatively-warm lava flow field, but these signals also coincide with seismic unrest noted by AVO that prompted AVO to raise the color code from green to yellow on 1 August 2019 for 24 days and on 18 June 2020 for 64 days (Cameron et al., 2018; Orr et al., 2023). Further analysis of the detected radiative power in comparison with complementary multiparameter datasets and higher resolution infrared images (e.g., Figure 1) could help tease out the origins and processes associated with these detections.

Our analysis shows that while both HotLINK and MIROVA are able to detect large and high temperature hotspots (e.g. Figure 8A), more subtle hotspots (Figure 8C) are only detected by HotLINK. The MIROVA system struggles to disregard bright and dispersed signals, such as solar reflections off of clouds, which exceed thresholds defined in the algorithm, but are visibly not hotspots in context (Figure 8B). HotLINK is able to detect more subtle hotspots that may be weak but still match the spatial patterns of a discrete thermal signal. The detection capabilities of HotLINK are similar to what an analyst can detect by eye.

5 Conclusion

This study confirms the capability of machine learning, specifically convolutional neural networks, to automate remote sensing tasks usually designated to human experts (Corradino et al., 2023). This technology provides three main improvements relative to threshold-based algorithms: 1) the model is more sensitive to subtle signals and can detect a larger number of hotspots while also detecting fewer false positive hotspots, 2) the probabilistic nature of the detections makes the model useful for different monitoring contexts, and 3) the same model performs well on data from different sensors (MODIS and VIIRS) and different Alaska volcanoes (with some caveats for volcanoes that are islands or have crater lakes).

The ability to detect more and weaker hotspots opens up the possibility of detecting precursory as well as eruptive hotspot signals. Specifically, our network detects subtle increases in volcanic surface temperature from Mount Veniaminof that correspond with both increased number of analyst detections of thermal signals and elevated seismicity. The capability to detect subtle signals associated with volcanic unrest, as well as eruptions, may aid in eruption forecasting efforts. Another advantage of our network is the probabilistic output. This expands the amount of information available to human analysts and will facilitate incorporation into statistical eruption forecasting models.

We found that HotLINK was able to detect hotspots in MODIS data with an even higher accuracy than for VIIRS data. Our model is therefore directly applicable to both VIIRS and MODIS data and is shown to work well on multiple volcanoes, only producing large errors in cases with crater lakes or small island volcanoes, which are especially susceptible to seasonal false detections. These errors could be minimized in the future using a detection threshold that exceeds the seasonal background signals at relevant volcanoes and/or by filtering out daytime images.

In conclusion, with a labeled training dataset of less than 4,000 VIIRS images from two volcanoes we were able to train a model to detect hotspots in both VIIRS and MODIS data that is applicable to many volcanoes. The time series for the eight volcanoes analyzed here captures volcanic unrest and eruption and thus can provide critical input into data-driven volcano monitoring and forecasting

studies, as well as valuable insight into the magmatic and eruptive processes occurring in active volcanic systems across Alaska. The model itself is also readily applicable for near-real-time or historical hotspot detection efforts by volcano observatories.

6 Plain language summary

Volcanoes release heat on their surface, and by monitoring this heat, we can determine if a volcano is erupting or might erupt soon. Heated areas, called hotspots, can be detected by satellite sensors, which generate images from space in infrared wavelengths. Traditionally, volcanologists or simple computer programs would identify the hotspots in infrared images. Now, advanced computer algorithms based on artificial intelligence can accurately identify complex features in images. We used these algorithms to improve the way we detect volcanic hotspots. Our approach detects more subtle heat signals than other algorithms, which is useful for detecting different types of volcanic activity, and may contribute to better forecasting of volcanic eruptions.

Data availability statement

The trained HotLINK model and datasets generated for this study can be found on Github: https://github.com/csaundersshultz/HotLINK.

Author contributions

PS-S: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. TL: Funding acquisition, Supervision, Writing–review and editing, Conceptualization, Investigation, Project administration, Resources. HD: Conceptualization, Software, Writing–review and editing. TG: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this project came from the National Science Foundation through a Prediction of and Resilience against Extreme Events (PREEVENTS) award (number 1855126), and also through U.S. Geological Survey Cooperative Agreement No. G21AC10384.

Acknowledgments

Thanks to Israel Brewster among others at the University of Alaska Fairbanks who aided in the efforts of the authors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the United States Government.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2024.1345104/full#supplementary-material

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: large-scale machine learning on heterogeneous systems. Available at: https://www.tensorflow.org/.

Apple (2023). About Face ID advance technology. Available at: https://support.apple.com/en-us/102381 (Accessed November 22, 2023).

Aveni, S., Laiolo, M., Campus, A., Massimetti, F., and Coppola, D. (2023). The capabilities of FY-3D/MERSI-II sensor to detect and quantify thermal volcanic activity: the 2020–2023 Mount Etna case study. *Remote Sens.* 15 (10), 2528. doi:10.3390/rs15102528

Blackett, M. (2013). Review of the utility of infrared remote sensing for detecting and monitoring volcanic activity with the case study of shortwave infrared data for Lascar Volcano from 2001–2005. *Geol. Soc. Lond. Spec. Publ.* 380, 107–135. doi:10.1144/SP380.10

Blackett, M. (2017). An overview of infrared remote sensing of volcanic activity. *J. Imaging* 3 (2), 13. doi:10.3390/jimaging3020013

Bleick, H. A., Coombs, M. L., Cervelli, P. F., Bull, K. F., and Wessels, R. L. (2013). Volcano–ice interactions precursory to the 2009 eruption of Redoubt Volcano, Alaska. *J. Volcanol. Geotherm. Res.* 259, 373–388. doi:10.1016/j.jvolgeores.2012.10.008

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information* 11 (2), 125. doi:10.3390/info11020125

Cameron, C. E., Prejean, S. G., Coombs, M. L., Wallace, K. L., Power, J. A., and Roman, D. C. (2018). Alaska Volcano Observatory alert and forecasting timeliness: 1989–2017. *Front. Earth Sci.* 6, 1–16. doi:10.3389/feart.2018.00086

Campus, A., Laiolo, M., Massimetti, F., and Coppola, D. (2022). The transition from MODIS to VIIRS for global volcano thermal monitoring. *Sensors* 22, 1713. doi:10.3390/s22051713

Carter, A. J., Ramsey, M. S., and Belousov, A. B. (2007). Detection of a new summit crater on Bezymianny Volcano lava dome: satellite and field-based thermal data. *Bull. Volcanol.* 69 (7), 811–815. doi:10.1007/s00445-007-0113-x

Castaño, L. M., Ospina, C. A., Cadena, O. E., Galvis-Arenas, B., Londono, J. M., Laverde, C. A., et al. (2020). Continuous monitoring of the 2015–2018 Nevado del Ruiz activity, Colombia, using satellite infrared images and local infrasound records. *Earth, Planets Space* 72, 81. doi:10.1186/s40623-020-01197-z

Chevrel, M. O., Villeneuve, N., Grandin, R., Froger, J. L., Coppola, D., Massimetti, F., et al. (2023). Lava flow daily monitoring: the case of the 19 September–5 October 2022 eruption at Piton de la Fournaise. *Volcanica* 6 (2), 391–404. doi:10.30909/vol.06.02.391404

Coombs, M. L., Wech, A. G., Haney, M. M., Lyons, J. J., Schneider, D. J., Schwaiger, H. F., et al. (2018). Short-term forecasting and detection of explosions during the 2016–2017 eruption of Bogoslof volcano, Alaska. *Front. Earth Sci.* 6, 1–17. doi:10.3389/feart.2018.00122

Coppola, D., Cardone, D., Laiolo, M., Aveni, S., Campus, A., and Massimetti, F. (2023). Global radiant flux from active volcanoes: the 2000–2019 MIROVA database. *Front. Earth Sci.* 11. doi:10.3389/feart.2023.1240107

Coppola, D., Laiolo, M., Cigolini, C., Donne, D. D., and Ripepe, M. (2016). Enhanced volcanic hot-spot detection using MODIS IR data: results from the MIROVA system. *Geol. Soc. Lond. Spec. Publ.* 426, 181–205. doi:10.1144/SP426.5

Coppola, D., Laiolo, M., Cigolini, C., Massimetti, F., Delle Donne, D., Ripepe, M., et al. (2020). Thermal remote sensing for global volcano monitoring: experiences from the MIROVA system. *Front. Earth Sci.* 7, 362. doi:10.3389/feart.2019.00362

Coppola, D., Laiolo, M., Delle Donne, D., Ripepe, M., and Cigolini, C. (2014). Hot-spot detection and characterization of strombolian activity from MODIS infrared data. *Int. J. Remote Sens.* 35 (9), 3403–3426. doi:10.1080/01431161.2014.903354

Coppola, D., Piscopo, D., Laiolo, M., Cigolini, C., Delle Donne, D., and Ripepe, M. (2012). Radiative heat power at Stromboli volcano during 2000–2011: twelve years of MODIS observations. *J. Volcanol. Geotherm. Res.* 215, 48–60. doi:10.1016/j.jvolgeores.2011.12.001

Coppola, D., Valade, S., Masias, P., Laiolo, M., Massimetti, F., Campus, A., et al. (2022). Shallow magma convection evidenced by excess degassing and thermal radiation during the dome-forming Sabancaya eruption (2012–2020). *Bull. Volcanol.* 84 (2), 16. doi:10.1007/s00445-022-01523-1

Corradino, C., Ramsey, M. S., Pailot-Bonnetat, S., Harris, A. J. L., and Negro, C. D. (2023). Detection of subtle thermal anomalies: deep learning applied to the ASTER global volcano dataset. *IEEE Trans. Geoscience Remote Sens.* 61, 1–15. doi:10.1109/TGRS.2023.3241085

Dehn, J., Dean, K., and Engle, K. (2000). Thermal monitoring of North Pacific volcanoes from space. *Geology* 28 (8), 755–758. doi:10.1130/0091-7613(2000)028<0755:tmonpv>2.3.co;2

Dehn, J., Dean, K., Engle, K., and Izbekov, P. (2002). Thermal precursors in satellite images of the 1999 eruption of Shishaldin Volcano. *Bull. Volcanol.* 64, 525–534. doi:10.1007/s00445-002-0227-0

El Adoui, M., Mahmoudi, S. A., Larhmam, M. A., and Benjelloun, M. (2019). MRI breast tumor segmentation using different encoder and decoder CNN architectures. *Computers* 8 (3), 52. doi:10.3390/computers8030052

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* 30 (1), 27–38. doi:10.1016/j.patrec.2008.08.010

Ganci, G., Vicari, A., Fortuna, L., and Del Negro, C. (2011). The HOTSAT volcano monitoring system based on combined use of SEVIRI and MODIS multispectral data. *Ann. Geophys.* 54, 544–550. doi:10.4401/ag-5338

Genzano, N., Pergola, N., and Marchese, F. (2020). A google Earth engine tool to investigate, map and monitor volcanic thermal anomalies at global scale by means of mid-high spatial resolution satellite data. *Remote Sens.* 12, 3232. doi:10.3390/rs12193232

Girona, T., Realmuto, V., and Lundgren, P. (2021). Large-scale thermal unrest of volcanoes for years prior to eruption. *Nat. Geosci.* 14 (4), 238–241. doi:10.1038/s41561-021-00705-4

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research, 249–256. Available at: https://proceedings.mlr.press/v9/glorot10a.html.

Gouhier, M., Guéhenneux, Y., Labazuy, P., Cacault, P., Decriem, J., and Rivet, S. (2016). HOTVOLC: a web-based monitoring system for volcanic hot spots. *Geol. Soc. Lond. Spec. Publ.* 426, 223–241. doi:10.1144/SP426.31

Guffanti, M., and Miller, T. P. (2013). A volcanic activity alert-level system for aviation: review of its development and application in Alaska. *Nat. hazards* 69, 1519–1533. doi:10.1007/s11069-013-0761-4

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). "On calibration of modern neural networks," in Proceedings of the 34th International Conference on Machine Learning, USA, August 6 - 11, 2017 (PMLR), 1321–1330.

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather Forecast.* 12 (4), 736–741. doi:10.1175/1520-0434(1997)012<0736:rdfmpf>2.0.co;2

Harris, A. (2013). *Thermal remote sensing of active volcanoes: a user's manual*. Cambridge: Cambridge University Press.

Harris, A. J., De Groeve, T., Garel, F., and Carn, S. A. (2016). "Detecting, modelling and responding to effusive eruptions,"*Geol. Soc. Lond*. doi:10.1144/SP426.0

Harris, A. J., and Stevenson, D. S. (1997). Thermal observations of degassing open conduits and fumaroles at Stromboli and Vulcano using remotely sensed data. *J. Volcanol. Geotherm. Res.* 76 (3-4), 175–198. doi:10.1016/S0377-0273(96)00097-2

Harris, A. J., Villeneuve, N., Di Muro, A., Ferrazzini, V., Peltier, A., Coppola, D., et al. (2017). Effusive crises at Piton de la Fournaise 2014–2015: a review of a multi-national response model. *J. Appl. Volcanol.* 6 (1), 11–29. doi:10.1186/s13617-017-0062-9

Higgins, J., and Harris, A. (1997). Vast: a program to locate and analyse volcanic thermal anomalies automatically from remotely sensed data. *Comput. Geosciences* 23, 627–645. doi:10.1016/S0098-3004(97)00039-3

Hirn, B., Di Bartola, C., and Ferrucci, F. (2009). Combined use of SEVIRI and MODIS for detecting, measuring, and monitoring active lava flows at erupting volcanoes. *IEEE Trans. Geoscience Remote Sens.* 47 (8), 2923–2930. doi:10.1109/TGRS.2009.2014224

Hurst, A. W., Bibby, H. M., Scott, B. J., and McGuinness, M. J. (1991). The heat source of Ruapehu crater lake; deductions from the energy and mass balances. *J. Volcanol. Geotherm. Res.* 46, 1–20. doi:10.1016/0377-0273(91)90072-8

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. doi:10.48550/arXiv.1412.6980

Laiolo, M., Coppola, D., Barahona, F., Benítez, J. E., Cigolini, C., Escobar, D., et al. (2017). Evidences of volcanic unrest on high-temperature fumaroles by satellite thermal monitoring: the case of Santa Ana volcano, El Salvador. *J. Volcanol. Geotherm. Res.* 340, 170–179. doi:10.1016/j.jvolgeores.2017.04.013

Layana, S., Aguilera, F., Rojo, G., Vergara, Á., Salazar, P., Quispe, J., et al. (2020). Volcanic Anomalies monitoring System (VOLCANOMS), a low-cost volcanic monitoring system based on Landsat images. *Remote Sens.* 12 (10), 1589. doi:10.3390/rs12101589

LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). "Convolutional networks and applications in vision," in Proceedings of 2010 IEEE international symposium on circuits and systems, China, 3-6 May 2010 (IEEE), 253–256. doi:10.1109/ISCAS.2010.5537907

Li, F., Zhang, X., Kondragunta, S., and Csiszar, I. (2018). Comparison of fire radiative power estimates from VIIRS and MODIS observations. *J. Geophys. Res. Atmos.* 123, 4545–4563. doi:10.1029/2017JD027823

Loewen, M. W., Dietterich, H. R., Graham, N., and Izbekov, P. (2021). Evolution in eruptive style of the 2018 eruption of Veniaminof volcano, Alaska, reflected in groundmass textures and remote sensing. *Bull. Volcanol.* 83 (11), 72. doi:10.1007/s00445-021-01489-6

Lombardo, V. (2016). AVHotRR: near-real time routine for volcano monitoring using IR satellite data. *Geol. Soc. Lond. Spec. Publ.* 426, 73–92. doi:10.1144/SP426.18

Loughlin, S. C., Vye-Brown, C., Sparks, R. S. J., Brown, S. K., Barclay, J., Calder, E., et al. (2015). An introduction to global volcanic hazard and risk. *Glob. Volcan. Hazards Risk* 2015, 1–80. doi:10.1017/CBO9781316276273.003

Massimetti, F., Coppola, D., Laiolo, M., Valade, S., Cigolini, C., and Ripepe, M. (2020). Volcanic hot-spot detection using SENTINEL-2: a comparison with MODIS–MIROVA thermal data series. *Remote Sens.* 12 (5), 820. doi:10.3390/rs12050820

NASA Goddard Space Flight Center (2018). NASA visible infrared imaging radiometer suite level-1B product user guide. Available at: https://landweb.modaps.eosdis.nasa.gov/NPP/forPage/NPPguide/NASAVIIRSL1BUGMay2018.pdf.

Oppenheimer, C., Rothery, D. A., and Francis, P. W. (1993). Thermal distributions at fumarole fields: implications for infrared remote sensing of active volcanoes. *J. Volcanol. Geotherm. Res.* 55 (1-2), 97–115. doi:10.1016/0377-0273(93)90092-6

Orr, T. R., Cameron, C. E., Dietterich, H. R., Dixon, J. P., Enders, M. L., Grapenthin, R., et al. (2023). 2019 volcanic activity in Alaska—summary of events and response of the Alaska Volcano Observatory. *U. S. Geol. Surv*. doi:10.3133/sir20235039

Pergola, N., Marchese, F., and Tramutoli, V. (2004). Automated detection of thermal features of active volcanoes by means of infrared AVHRR records. *Remote Sens. Environ.* 93, 311–327. doi:10.1016/j.rse.2004.07.010

Pieri, D., and Abrams, M. (2005). ASTER observations of thermal anomalies preceding the April 2003 eruption of Chikurachki volcano, Kurile Islands, Russia. *Remote Sens. Environ.* 99, 84–94. doi:10.1016/j.rse.2005.06.012

Planck, M. (1914). *The theory of heat radiation*. USA: Blakiston.

Pritchard, M. E., Poland, M., Reath, K., Andrews, B., Bagnardi, M., Biggs, J., et al. (2022). Optimizing satellite resources for the global assessment and mitigation of volcanic hazards—suggestions from the USGS powell center volcano remote sensing working group. *U. S. Geol. Surv*. doi:10.3133/sir20225116

Ramsey, M. S., Corradino, C., Thompson, J. O., and Leggett, T. N. (2023). Statistical retrieval of volcanic activity in long time series orbital data: implications for forecasting future activity. *Remote Sens. Environ.* 295, 113704. doi:10.1016/j.rse.2023.113704

Ramsey, M. S., Wessels, R. L., and Anderson, S. W. (2012). Surface textures and dynamics of the 2005 lava dome at Shiveluch Volcano, Kamchatka. *Bulletin* 124 (5-6), 678–689. doi:10.1130/B30580.1

Reath, K. A., Ramsey, M. S., Dehn, J., and Webley, P. W. (2016). Predicting eruptions from precursory activity using remote sensing data hybridization. *J. Volcanol. Geotherm. Res.* 321, 18–30. doi:10.1016/j.jvolgeores.2016.04.027

Ronneberger, O., Fischer, P., and Brox, T. (2015). INet: convolutional networks for biomedical image segmentation. *IEEE Access* 9, 16591–16603. doi:10.1109/ACCESS.2021.3053408

Rouwet, D., Tassi, F., Mora-Amador, R., Sandri, L., and Chiarini, V. (2014). Past, present and future of volcanic lake monitoring. *J. Volcanol. Geotherm. Res.* 272, 78–97. doi:10.1016/j.jvolgeores.2013.12.009

Saunders-Shultz, P. (2023). Hotspot learning and identification network (HotLINK). Available at: https://github.com/csaundersshultz/HotLINK.

Sola, J., and Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* 44 (3), 1464–1468. doi:10.1109/23.589532

Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., et al. (2019). Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: the MOUNTS monitoring system. *Remote Sens.* 11, 1528–1531. doi:10.3390/rs11131528

Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., et al. (2014). "Deep learning for content-based image retrieval: a comprehensive study," in Proceedings of the 22nd ACM international conference on Multimedia, 2014, November (IEEE), 157–166. doi:10.1145/2647868.2654948

Waythomas, C. (2021). Simultaneous effusive and explosive cinder cone eruptions at Veniaminof Volcano, Alaska. *Volcanica* 4 (2), 295–307. doi:10.30909/vol.04.02.295307

Waythomas, C. F., Edwards, B. R., Miller, T. P., and McGimsey, R. G. (2023). Lava-ice interactions during historical eruptions of Veniaminof Volcano, Alaska and the potential for meltwater floods and lahars. *Nat. Hazards* 115 (1), 73–106. doi:10.1007/s11069-022-05523-4

Werner, C., Kern, C., Coppola, D., Lyons, J. J., Kelly, P. J., Wallace, K. L., et al. (2017). Magmatic degassing, lava dome extrusion, and explosions from Mount Cleveland volcano, Alaska, 2011–2015: insight into the continuous nature of volcanic activity over multi-year timescales. *J. Volcanol. Geotherm. Res.* 337, 98–110. doi:10.1016/j.jvolgeores.2017.03.001

Wooster, M. (2003). Fire radiative energy for quantitative study of biomass burning: derivation from the BIRD experimental satellite and comparison to MODIS fire products. *Remote Sens. Environ.* 86 (1), 83–107. doi:10.1016/S0034-4257(03)00070-1

Wright, R. (2016). MODVOLC: 14 years of autonomous observations of effusive volcanism from space. *Geol. Soc. Spec. Publ.* 426, 23–53. doi:10.1144/SP426.12

Wright, R., Flynn, L. P., Garbeil, H., Harris, A. J. L., and Pilger, E. (2004). MODVOLC: near-real-time thermal monitoring of global volcanism. *J. Volcanol. Geotherm. Res.* 135, 29–49. doi:10.1016/j.jvolgeores.2003.12.008