



Benchmarking computer vision architectures for cloud detection from lidar ceilometer backscatter data

Alessio Barbaro Chisari¹ · Luca Guarnera¹ · Alessandro Ortis¹ · Wladimiro Carlo Patatu² · Sebastiano Battiato¹ · Mario Valerio Giuffrida³

Accepted: 24 April 2025 / Published online: 21 May 2025
© The Author(s) 2025

Abstract

Cloud detection is fundamental for accurate weather monitoring, often achieved through remote sensing technology, such as satellite imagery or radar. This study explores the use of lidar ceilometer backscatter data, a rich but noisy source of atmospheric information, to enhance cloud detection. Leveraging data acquired from a Lufft CHM 15k ceilometer over three months near Mount Etna, Italy, we gathered a novel dataset comprising time-height plots derived from backscatter profiles. The Weather Research and Forecasting (WRF) model was used for ground-truth data labeling, ensuring reliable model validation. We benchmarked state-of-the-art deep learning architectures, including CNN-based models (e.g., ResNet50, VGG16, InceptionV3, EfficientNet) and the Vision Transformer (ViT), on our collected dataset. Among these, ResNet50 achieved the highest accuracy (89.57%), closely followed by ViT (89.36%), showcasing the efficacy of residual learning and transformer-based approaches in extracting complex patterns from atmospheric data. Our results highlight the potential of lidar-based systems for accurate cloud detection, complementing other remote sensing technologies. Our work contributes to the field by introducing a publicly available dataset and providing comprehensive benchmarking results that establish a baseline for future research. This study also opens avenues for broader applications of ceilometer data, such as the detection of pollutants and other atmospheric phenomena. Our dataset is publicly available at <https://zenodo.org/records/10616434>.

Keywords Ceilometer · Weather · Environment · Cloud Detection · Deep Learning.

1 Introduction

Monitoring and understanding cloud formations and their dynamics are critical tasks in meteorology, as clouds play a pivotal role in weather prediction, climate modeling, and atmospheric studies. Clouds influence the Earth's energy balance, modulate temperatures, and impact precipitation patterns. Traditional methods for cloud observation often rely on satellite imagery and ground-based radar systems. This work utilizes data from a lidar-based ceilometer, an advanced remote sensing instrument capable of measuring cloud base heights and detecting atmospheric aerosols through backscatter analysis of a modulated light beam emitted into the sky.

Ceilometers offer distinct advantages over other remote sensing devices, providing continuous, high-resolution vertical measurements of cloud and aerosol distribution. This capability makes them particularly valuable for studying rapid changes in atmospheric conditions and for detecting phenomena that are challenging to observe through satellites, such as low-altitude clouds or localized aerosol

✉ Alessio Barbaro Chisari
alessio.chisari@phd.unict.it

✉ Luca Guarnera
luca.guarnera@unict.it

Alessandro Ortis
alessandro.ortis@unict.it

Wladimiro Carlo Patatu
vladimiro.patatu@eht.eu

Sebastiano Battiato
sebastiano.battiato@unict.it

Mario Valerio Giuffrida
valerio.giuffrida@nottingham.ac.uk

¹ Department of Mathematics and Computer Science,
University of Catania, Viale Andrea Doria 6, Catania 95131,
Italy

² EHT S.C.p.A., Viale Africa 31, Catania 95129, Italy

³ School of Computer Science, University of Nottingham,
Wollaton Road, Nottingham NG8 1BB, UK

concentrations. Our study focuses on leveraging these unique capabilities to enhance cloud detection using state-of-the-art deep neural networks.

To better contextualize the role of ceilometers in cloud detection, we briefly discuss alternative lidar-based methods and their characteristics. Our approach employs lidar-based ceilometers for continuous measurement, real-time monitoring of the lower atmosphere, specifically focusing on earth-to-satellite observations. Unlike high-sensitivity lidar systems [1], which integrate SNSPD technology to detect faint backscatter signals from high-altitude clouds with minimal noise, ceilometers are optimized for tracking cloud base heights and aerosol layers within the Planetary Boundary Layer (PBL). While hybrid radar-lidar techniques [2] combine millimeter-wave radar and multi-wavelength lidar for improved penetration and resolution, they require colocated instrumentation, limiting their deployment flexibility. Similarly, Airborne Laser Scanning (ALS) [3] provides high-resolution 3D cloud morphology but is constrained by high operational costs and limited temporal coverage. In contrast, ceilometers operate at a single wavelength, offering a cost-effective, high-frequency sampling of atmospheric processes, making them particularly suitable for studying cloud dynamics in urban and industrial environments, where local emissions and surface heating play a crucial role. These comparisons highlight the unique advantages of ceilometers in providing continuous, cost-effective, and high-frequency atmospheric observations, making them particularly relevant for cloud monitoring in urban and industrial environments.

While existing approaches have utilized lidar-based imagery in combination with machine learning models [4], the availability of publicly accessible datasets remains limited, creating a barrier for broader research and development in this field. To address this gap, our study introduces a newly curated dataset comprising backscatter profiles. Our new dataset differs significantly from existing cloud detection datasets in the literature. Traditional lidar remote sensing systems predominantly follow a satellite-to-earth (top-down) perspective, whereas our approach adopts a bottom-up (Earth-to-satellite) acquisition method. This inversion in the data collection paradigm captures atmospheric dynamics from a rarely explored viewpoint, introducing novel challenges and opportunities for cloud detection.

This dataset represents a challenging benchmark for cloud detection due to its inclusion of diverse atmospheric conditions and varying cloud types observed around Mount Etna, an area known for its complex meteorological phenomena.

Ground-truth labeling of the dataset was performed using a high-resolution Weather Research and Forecasting (WRF) model, providing reliable reference data for model training and evaluation. Our study aims to provide a comprehensive performance benchmark for cloud detection on this dataset using several state-of-the-art deep learning architec-

tures, including both convolutional neural networks (CNNs) and transformer-based models. Specifically, we evaluated VGG16 [5], ResNet50 [6], InceptionV3 [7], EfficientNet [8], and the Vision Transformer (ViT) [9]. Results indicated that ResNet50 achieved the highest accuracy among CNNs at 89.57%, while the transformer-based ViT reached a comparable performance of 89.36%.

The main contributions of this work are as follows:

- *Introduction of a novel dataset* We present a new dataset of lidar ceilometer backscatter profiles, collected over a three-month period near Mount Etna, Italy. This dataset, characterized by high temporal resolution and diverse atmospheric conditions, serves as a valuable benchmark for cloud detection and atmospheric studies.
- *Comprehensive benchmarking of state-of-the-art models* We evaluate the performance of cutting-edge deep learning architectures, including CNN-based models (ResNet50, VGG16, InceptionV3, EfficientNet) and the Vision Transformer (ViT). This benchmarking provides a robust baseline for cloud detection task using lidar backscatter data.
- *High accuracy results* Among the tested models, ResNet50 achieved the highest accuracy (89.57%), closely followed by ViT (89.36%). These results highlight the efficacy of residual learning and transformer-based approaches in analyzing complex atmospheric patterns.
- *Support for future research* By making the dataset publicly accessible and offering detailed performance benchmarks, this work lays the foundation for future advancements in cloud detection and lidar-based atmospheric research.
- *Broader application potential* The dataset and methodology introduced in this study open new opportunities for leveraging lidar ceilometer data to detect other atmospheric phenomena, such as aerosols, pollutants, and volcanic emissions.

These contributions represent a significant step forward in utilizing lidar-based systems and advanced deep learning techniques for accurate and scalable atmospheric monitoring. This work is an extension of the work presented in [10] by the authors. It includes a more in-depth analysis of the state-of-the-art, a larger number of experiments and an in-depth and detailed comparison of results, not present in [10].

The paper is structured as follows: Sect. 2 reviews the main contributions in the existing literature. Section 3 describes the proposed methodology and outlines the processes for dataset acquisition, collection, and preparation. Experimental results are presented in Sect. 4. Section 5 offers an in-depth analysis of the findings, discusses the limitations of the work, and highlights key observations. Finally, Sect. 6 concludes the paper.

2 Related works

For a comprehensive review of cloud detection, including the use of ceilometer data, we invite the readers to refer to [11].

Backscatter profiles acquired by ceilometers have been shown to be highly correlated in the presence of atmospheric particulate matter, as demonstrated in previous studies [12, 13]. The efficacy of ceilometer data has also been instrumental in detecting volcanic emissions during the 2010 eruption of the Icelandic volcano Eyjafjallajökull [14].

Given its potential, the exploration of sophisticated data mining techniques for the analysis of ceilometer-acquired data has been a subject of discussion since the inception of the Data Mining Project [15]. In pursuit of this objective, a noteworthy contribution to the research community emerged from the work of Wiegner et al. [16], wherein an approach to calibrate measurements from a Jenoptik CHM 15kx ceilometer was presented. Later, Arun et al. [17] delved into the synergy between ground-based ceilometer observations and satellite data from remote sensing sources in their study. By combining these distinct datasets, they aimed to enhance the precision of cloud detection, highlighting the evolving landscape of data fusion for atmospheric analysis.

In [18], the authors proposed a technique for detecting specific meteorological phenomena, such as fog and clouds, using a lidar-based ceilometer. The methodology involved the application of classical machine learning methods, including Support Vector Machines (SVM), as well as shallow neural networks. These techniques leveraged raw data obtained from the ceilometer as predictive features, enabling the accurate identification of atmospheric events. Similarly, in [19], the authors undertook cloud classification by taking advantage of both ceilometer data with sky images captured by a camera. Within their study, a random forest approach was employed to perform multi-class classification, effectively discerning various cloud types. This integration of data sources facilitated comprehensive cloud identification. In [20], ceilometer data have been utilized to evaluate a federated learning approach incorporating both labeled and unlabeled samples in a semi-supervised setting. This methodology aims to enhance model performance by leveraging feature extraction from unannotated data, contributing to the broader research on privacy-preserving machine learning for Earth observation applications. Sleeman et al. [21] used lidar-based ceilometer data to detect the Planetary Boundary Layer Height (PBLH) with the use of machine learning techniques. In [22], they introduced an unsupervised methodology for classifying meteorological occurrences, leveraging k -means clustering. An autoencoder was trained to learn a suitable representation of backscatter profiles, subsequently organized into clusters. While demonstrating promise, this technique was presented as a prototype proof-of-concept. Notably, the absence of labeled data and a comprehensive

evaluation hampered its full validation. Conversely, the study in [4] addressed cloud detection through Fully Convolutional Networks. In their approach, backscatter profiles were provided into their model via a *mask algorithm*, and the model was trained in a supervised fashion, as they labeled a dataset of backscatter profiles. This dataset enabled an in-depth quantitative performance analysis of their proposed methodology, setting it apart from prior works in the literature. An et al. [23] developed a cloud detection algorithm based on FY-3E satellite infrared channels for early morning observations. Their method utilizes dynamic thresholds and auxiliary data (such as SST, LST, and snow cover masks) to adjust for varying land surface conditions and improve detection accuracy. In contrast, Li et al. [24] proposed a Residual Dual U-Shape Network (RD-UNet) with improved skip connections, which effectively integrates multi-scale features to better detect thin clouds and refine cloud boundaries. Although both approaches rely on satellite imagery, our work diverges by employing ceilometer lidar backscatter data, offering a bottom-up perspective that captures high temporal resolution and detailed vertical structure information. Our system uses lidar-based ceilometers to gather backscatter data primarily from the atmospheric boundary layer. Conversely, other research efforts have adopted different lidar setups, each with unique functional mechanisms and observational strengths. The method described in [1] features a high-sensitivity atmospheric lidar system equipped with a Superconducting Nanowire Single-Photon Detector (SNSPD), which improves the identification of weak backscatter signals from upper-atmosphere clouds while minimizing noise. Nonetheless, such systems are mainly designed for mid-to-high atmospheric layers and require intricate calibration techniques to address signal degradation. Alternatively, hybrid radar-lidar approaches, such as the one in [2], merge millimeter-wave cloud radar with ground-based lidar operating at multiple wavelengths. In these setups, radar excels at deep penetration, whereas lidar provides finer detail at lower elevations. However, these combined systems struggle to differentiate between drizzle and cloud particles and rely on co-located instruments, restricting their deployment flexibility. Airborne Laser Scanning (ALS) lidar, exemplified in [3], uses aircraft-mounted, pulsed near-infrared lasers to construct high-resolution, three-dimensional representations of cloud structures. Although effective in capturing cloud morphology, ALS faces limitations due to its high operational expenses, restricted temporal coverage, and reliance on flight schedules. Our system stands apart by employing ceilometers for ground-to-satellite measurements. These instruments operate at a single wavelength and offer continuous monitoring of the lower atmosphere. Their key strength lies in their ability to deliver real-time, high-frequency measurements of the Planetary Boundary Layer (PBL)-the atmospheric zone most affected by surface-level weather and human-induced

changes. Unlike powerful research lidar systems, ceilometers are tailored for monitoring cloud base altitudes and aerosol concentrations within a vertical range spanning from a few dozen meters to several kilometers. This makes them especially suitable for examining cloud behavior and evolution in urban and industrial settings, where local emissions and heat exchanges significantly influence atmospheric properties.

The current research landscape demonstrates diverse approaches leveraging ceilometer data, often relying on distinct datasets and traditional machine learning methods (e.g., SVM). However, the complexity of backscatter profiles, as collected in this study, makes them more suitable for deep neural networks. To address gaps in the existing literature, this work introduces a high-resolution dataset tailored for deep learning applications in cloud detection. Collected near an active volcano, this dataset captures unique and challenging atmospheric conditions, enabling rigorous benchmarking of advanced models and fostering future research. In this paper, we benchmark state-of-the-art deep learning models using this novel dataset, offering valuable insights and resources for the scientific community.

3 Proposed method

This section outlines the proposed approach for collecting and labeling the dataset obtained using a lidar-based ceilometer. Data were acquired by deploying a dedicated measuring instrument, commonly utilized in the state-of-the-art for atmospheric profiling (the ceilometer).

Data collection occurred between January 2023 and mid-March 2023. This period was characterized by a series of atmospheric perturbations, leading to a more balanced and diverse dataset. The data acquisition was conducted on a daily basis using the lidar-based ceilometer, strategically located near San Giovanni La Punta (CT), Italy, at the following coordinates: [37° 34' 43.997" N, 15° 6' 11.002" E].

3.1 Ceilometer

Ceilometers are widely employed for various meteorological applications, with two primary configurations being most common. In the rotating transmitter design, a projector sweeps the sky with a modulated light beam, while a detector, placed at a known distance and pointed vertically, detects reflections when the light intersects a cloud base directly overhead. This reflection is captured, and the projection angle at the time of detection is recorded. Conversely, in the scanner-receiver setup, the projector remains stationary and emits a vertical modulated beam. A parabolic detector, positioned at a predetermined distance, moves to scan the beam in both upward and downward directions, capturing light reflected by the cloud base and recording the vertical

intersection angle. Both configurations provide essential data for determining cloud base height using triangulation.

In this study, we utilized a Lufft CHM 15k ceilometer, which leverages LiDAR (Light Detection and Ranging) technology, as depicted in Fig. 1. This device emits short light pulses from a solid-state laser microchip, which are scattered by particles such as aerosols, water droplets, and other molecules in the atmosphere. The portion of light reflected back, termed backscatter, is analyzed by the ceilometer. By measuring the time taken for the laser pulses to return, the device determines the distance to the scattering particles. The height profile of the reflected signals is then used to derive the backscatter intensity β -raw, which, with the aid of a valid calibration constant, is converted into the attenuated backscatter coefficient β -att.

From this information, various atmospheric parameters are determined, including cloud height and aerosol layer distributions. The detection mechanism employs photon counting to ensure precise measurement. The narrow bandwidth of the laser allows for a roughly 1 nm optical filter to be placed in front of the detector, reducing background noise significantly. Signal averaging is employed to improve the signal-to-noise ratio, which is crucial for lidar-based measurements and for obtaining detailed aerosol profiles. Compared to analog measurement methods, this approach offers superior sensitivity and accuracy.

The casing of the ceilometer consists of a dual-layered stainless aluminum structure, with the outer shell designed to shield the inner casing, which houses the measurement unit, from solar radiation, wind, rain, and snow. A protective lid prevents dirt and precipitation from entering, and a window allows the laser beam to exit and re-enter the instrument. Inside the lid, a partition separates the emission zone from the sensitive receiving area, while an air baffle directs airflow from internal fans over the glass panel of the inner casing to keep it clean and functional.

3.2 Data collection & processing

The ceilometer carried out measurements every 15 s, enabling precise quantification of atmospheric particle concentration. By analyzing the reflected signals, cloud layer coverage could be determined. The selected ceilometer generates and processes a substantial volume of raw data, as shown in Table 1. Figure 1 illustrates the data collection procedure.

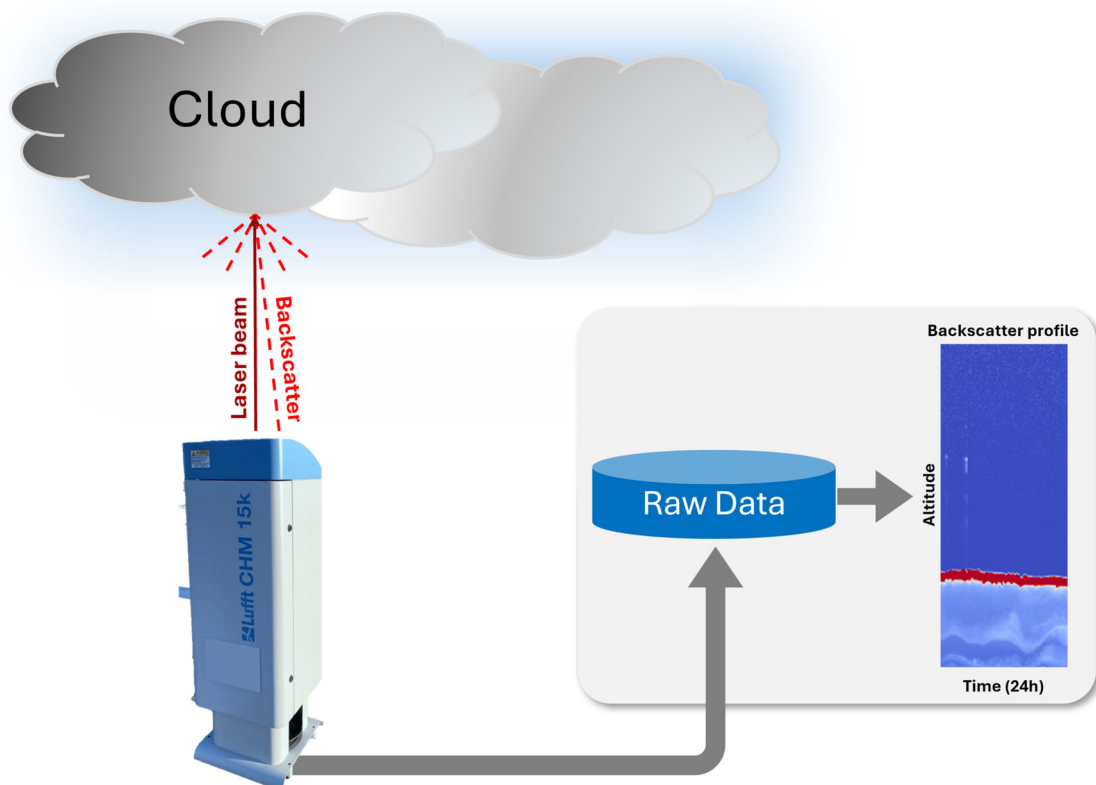
Once the raw data were acquired, the parameters of interest were normalized using a specific calibration factor for the lidar-based ceilometer. Several parameters contributed to constructing backscatter profiles. These profiles were plotted with time represented on the x -axis and particle height (reflected in the backscatter coefficient) on the y -axis. The plot's colors indicate the intensity of the measured particles: deep blue signifies minimal particulate presence, while red

Table 1 Complete list of all parameters determined by the Lufft CHM 15k ceilometer from the backscatter

Parameter	Type	Unit	Description
Time	Double	Secs	End point of the measurement (UTC)
Range	Float	m	Measurement distance of the device (independent of direction and height of installation location)
Range_hr	Float	m	Measurement distance of the device for high resolution
Layer	Fnt	–	Layer index
Latitude	Float	°	Latitude of the installation location
Longitude	Float	°	Longitude of the installation location
Azimuth	Float	°	Azimuth angle of the device (direction of the laser indicator)
Zenith	Float	°	Zenith angle of the device (direction of the laser indicator)
Altitude	Float	m	Height of installation of the device above sea level
Wavelength	Float	nm	Wavelength of the laser in nm
Average_time	Int	ms	Average time per recording
Range_gate	Float	m	Spatial resolution of the measurement
Range_gate_hr	Float	m	Spatial resolution of the high-resolution measurement
Life_time	Int	h	Propagation time of the laser
Error_ext	Int	–	32-bit status code
State_laser	Byte	%	Laser quality index
State_detector	Byte	%	Signal detector quality
State_optics	Byte	%	Optical quality index
Temp_int	Short	K	Internal temperature of the housing
Temp_ext	Short	K	External temperature of the housing
Temp_det	Short	K	Temperature of the detector
Temp_lom	Short	K	Temperature of the measurement unit
Laser_pulses	int	-	Number of laser pulses emitted during a measurement (lp)
p_calc	short	#	Calibration pulse (normalization of the measurement unit over time)
Scaling	float	-	Scaling factor (normalization of measurement units relative to each other) (cs)
Base	float	#	Height of the baseline of the raw signal (primarily influenced by daylight) (b)
stddev	float	#	Standard deviation of the raw signal
Beta_raw	float	-	Normalized backscatter signal, corrected for range $((P_{\text{raw}} / lp) - b) / (cs * o(r) * p_{\text{calc}}) * r * r$, with $P_{\text{raw}} = \sum(P_{\text{raw_hr}} * \text{range_gate_hr} / \text{range_gate})$
Beta_raw_hr	float	-	High-resolution backscatter signal, normalized corrected for range $((P_{\text{raw_hr}} / lp) - b) / (cs * o(r) * p_{\text{calc}}) * r * r$
pbl	short	m	Aerosol layers
pbs	byte	–	Quality index for aerosol layers (1: good, 9: bad)
tcc	byte	–	Degree of coverage (overall)
bcc	byte	–	Degree of coverage of the lower cloud layer
sci	byte	-	Sky Condition Index (0: no precipitation, 1: rain, 2: fog 3: snow, 4: precipitation or particles on the window pane)

Table 1 continued

Parameter	Type	Unit	Description
vor	short	m	Vertical visibility
voe	short	m	Opacity of the detected vertical visibility
mxh	short	m	Maximum detection distance
cbh	short	m	Cloud base height
cbe	short	m	Calculated cloud base blur
cdp	short	m	Cloud penetration depth
cde	short	m	Calculated cloud penetration depth blur
cho	short	m	Height offset (calculated in cbh, mxh, vor, and pbl corresponds to altitude when usealtitude=1, otherwise 0)

**Fig. 1** Visual representation of the data collection process. Backscatter raw data are utilized to generate a time-height plotting of backscatter coefficients (profile)

indicates high concentrations. The scale ranges numerically from 0 to $5 \cdot 10^{-6}$. We generated a backscatter profile for each day of data collection, further dividing it into hourly intervals, resulting in 24 profiles per day. Figure 2 provides an example of the processed data. In total, 1,568 images of dimensions 150×1000 were created, each representing an hour-long measurement period.

The generated backscatter profiles were labeled using the *Weather Research and Forecasting* (WRF) Model, a mesoscale numerical prediction system designed for atmospheric research and operational forecasting. Figure 3 presents

the workflow of the WRF model, which features two dynamic cores, a data assimilation system, and a software architecture optimized for parallel computing. The model serves a broad spectrum of meteorological applications, covering scales from tens of meters to thousands of kilometers. Its spatial resolution of 1×1 km offers greater detail compared to typical global forecast models, which often operate at 27×27 km. The WRF model leverages global weather data from the Global Forecast System (GFS), provided by the *National Center for Atmospheric Research* (NCAR) [25].

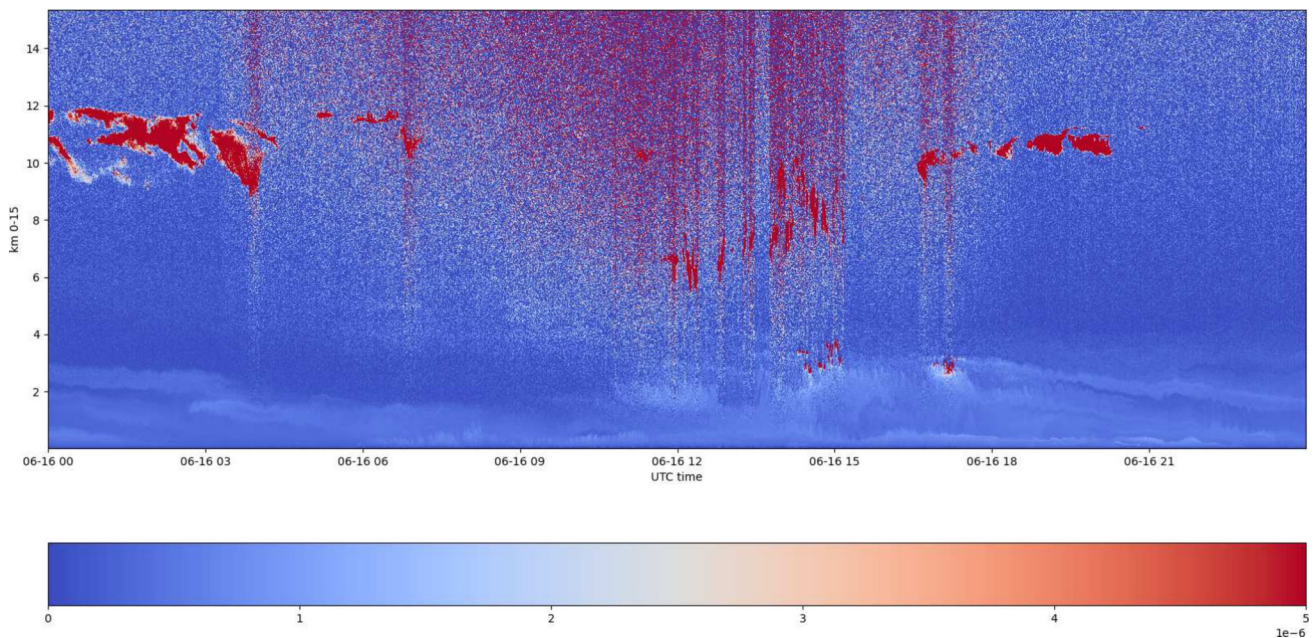


Fig. 2 Backscatter profile of 24-hour measurements taken on the 16th of June 2022. As explained in Sect. 3.2, the color of the plot depends on the intensity of the measured particle: intense blue means absence of particulate; red means intense presence of particulate. **Best viewed in color.**

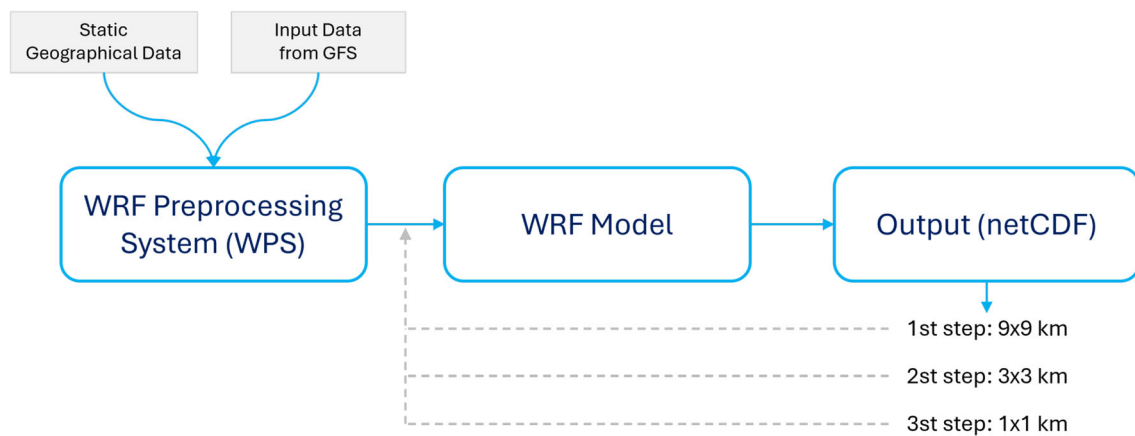


Fig. 3 Adopted workflow for the employed WRF model. Note that the outputs of the first and second steps serve as inputs to the last step. As explained in Sect. 3.2, GFS is the Global Forecast System. Global weather data from the GFS are used as the first input

The WRF model produces netCDF files representing a 3D geographic grid, as depicted in Fig. 4. Latitude and longitude are aligned with the x -axis and y -axis, while 40 pressure levels are represented on the z -axis. The displayed images differ in spatial resolution and the quality of the WRF model outputs. The first image has a resolution of 9×9 km, meaning that each point on the spatial grid is spaced 9 km apart. This results in a relatively coarse depiction of atmospheric conditions, as finer details of meteorological phenomena are not captured at this scale. In contrast, the second image utilizes a higher-resolution grid of 3×3 km, achieved through the nesting technique within the WRF model. Nesting involves embedding one or more high-resolution grids (referred to as

nested domains) within a coarser grid (the parent domain). In this case, the 3×3 km grid is nested inside the parent domain of 9×9 km. During this process, the WRF model incorporates meteorological data from the parent domain to enhance the local representation of atmospheric phenomena, resulting in more precise forecasts for specific areas of interest. Finally, the third image presents data with a resolution of 1×1 km, which offers nine times the precision of the initial grid. At this level, nesting is further refined by adding a third nested domain, enabling the model to capture highly detailed meteorological features, such as localized variations in temperature, wind, and precipitation. This setup enabled us to isolate the central point of the reference domain, correspond-

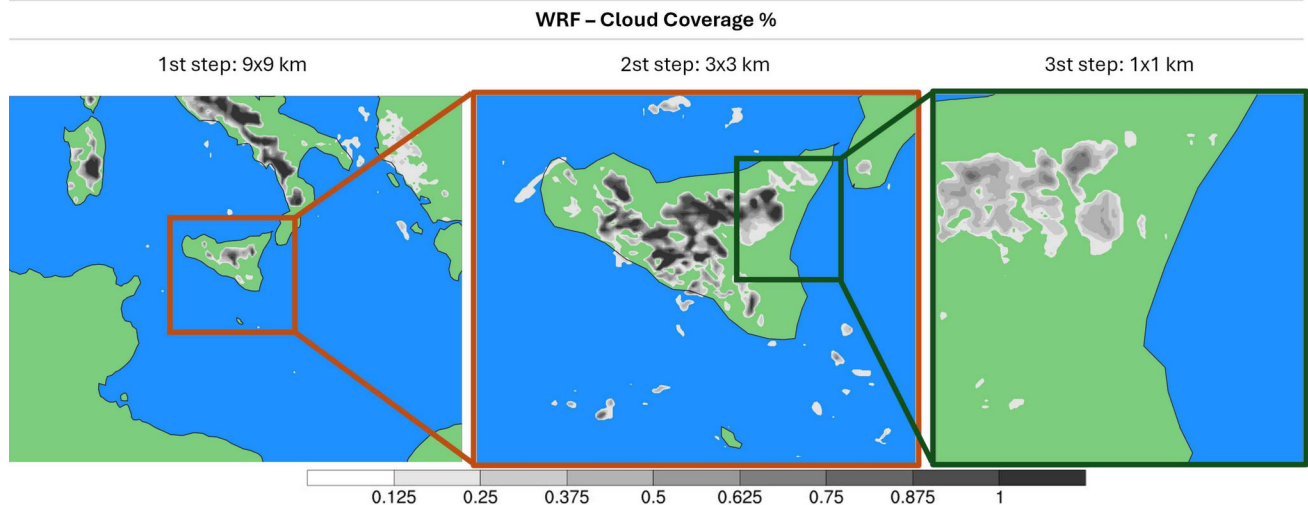


Fig. 4 Visual representation of the output of WRF model showing the percentage cloud cover. **Best viewed in color.**

ing to the ceilometer's geographical location, and determine cloud presence or absence at each pressure level. This process provided hourly cloud cover data for the ceilometer's location, serving as ground-truth labels for each backscatter profile with high reliability.

The labeled backscatter profiles were subsequently used to train several state-of-the-art deep learning models, including VGG16 [5], ResNet50 [6], InceptionV3 [7], EfficientNet [8], and ViT [9], utilizing the PyTorch framework.

VGG16 is a classic Convolutional Neural Network (CNN) known for its simple and uniform architecture, which serves as a strong baseline in image classification tasks. Inception v3 introduces the concept of inception modules, allowing the network to capture multi-scale features efficiently while reducing computational cost. ResNet50 employs residual connections to enable the training of deeper networks and mitigate the vanishing gradient problem, proving highly effective in various vision tasks. EfficientNet scales network width, depth, and resolution in a principled manner, achieving high accuracy with fewer parameters. Lastly, Vision Transformer (ViT) adopts a transformer-based architecture that operates directly on image patches, providing an alternative to convolutional approaches and achieving state-of-the-art results in image recognition.

The dataset is publicly available at <https://zenodo.org/records/10616434>.

4 Experimental results

We trained all models in a standard supervised learning approach using Cross-Entropy loss. Training and inference were conducted on 1-hour-long ceilometer measurements, as detailed in Sect. 3.2. All models were initialized with

ImageNet-pretrained weights to enhance training stability and robustness. To mitigate overfitting during training, we utilized horizontal-flip data augmentation. For each experimental setup, the dataset was divided into training, validation, and test sets, with proportions of 49% (769 samples), 21% (329 samples), and 30% (470 samples), respectively, totaling 1050 samples belonging to the True class.

We explored various hyperparameter configurations and employed two optimizers: Stochastic Gradient Descent (SGD) and Adam. The following subsection outlines the combinations tested to identify the optimal model configuration.

4.1 Performance analysis and comparison of model configurations

Figures 5 and 6 show the performance results of various state-of-the-art deep learning models trained on a dataset of cloud detection using ceilometer backscatter profiles. The models evaluated include VGG16, EfficientNet, InceptionV3, ResNet50, and Vision Transformer (ViT), and were tested using multiple configurations and optimization strategies. The figures primarily differ based on the optimizer employed (SGD for Fig. 5 and ADAM for Fig. 6). Specifically, the diagrams in Fig. 5 are organized according to training parameters like learning rate, momentum, and weight decay, whereas the diagrams in Fig. 6 are grouped based solely on learning rate and weight decay. The models were evaluated using standard metrics, including Accuracy, F1-score, Precision, and Recall, to gauge their efficacy in detecting clouds accurately.

The VGG16 model exhibited notable variability across different configurations. The best performance was observed with configuration 3 using SGD, yielding an accuracy of

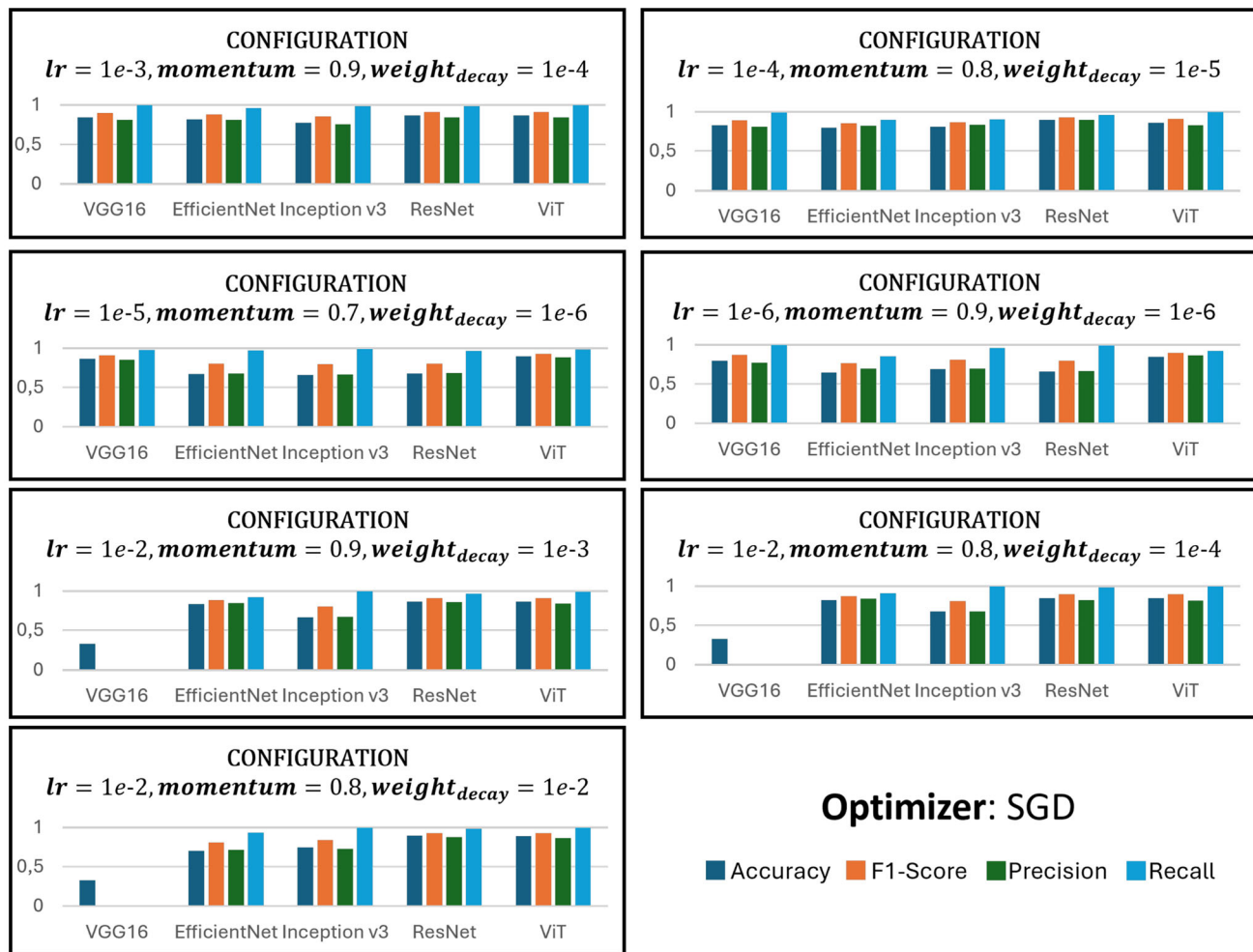


Fig. 5 Performance results of various state-of-the-art deep learning models trained for cloud detection using ceilometer backscatter profiles, evaluated with SGD optimizer and organized by training parameters such as learning rate, momentum, and weight decay. Models include

VGG16, EfficientNet, InceptionV3, ResNet50, and Vision Transformer (ViT), with metrics like Accuracy, F1-score, Precision, and Recall used for assessment. Missing values in some graphs indicate that the value of the metric in question is close to or equal to 0.0

86.38%, an F1-score of 0.9077, a Precision of 0.8495, and a Recall of 0.9744. The high recall value indicates that VGG16 was proficient at capturing relevant positive instances (i.e., cloud presence). However, its slightly lower precision compared to recall highlights the model's tendency to generate false positives. Interestingly, configurations with a higher learning rate (0.01) performed poorly, achieving a mere accuracy of 32.98%. This suggests that higher learning rates caused instability, potentially leading to divergence or overfitting during the training process.

EfficientNet demonstrated solid performance, with its best configuration using SGD achieving an accuracy of 83.19% alongside an F1-score of 0.8824 and a Precision of 0.8450. This result reflects a balance between precision and recall, suggesting that EfficientNet effectively identified cloud instances without a significant number of

false positives or negatives. However, when optimized with Adam, EfficientNet's performance slightly decreased, with the highest accuracy obtained being 81.91%. This difference underscores that certain architectures benefit more from one optimizer over another, with SGD appearing more suitable for EfficientNet in this context.

InceptionV3 showed a moderate performance range, achieving its highest accuracy of 82.77% with Adam optimization. This configuration exhibited a balanced F1-score (0.8751) and Precision (0.8656), indicating robust performance across different metrics. When optimized with SGD, InceptionV3 achieved a comparable accuracy of 81.06% but required the maximum number of epochs (59), reflecting a slower convergence rate compared to Adam. This suggests that InceptionV3 might be more efficient with Adam, espe-

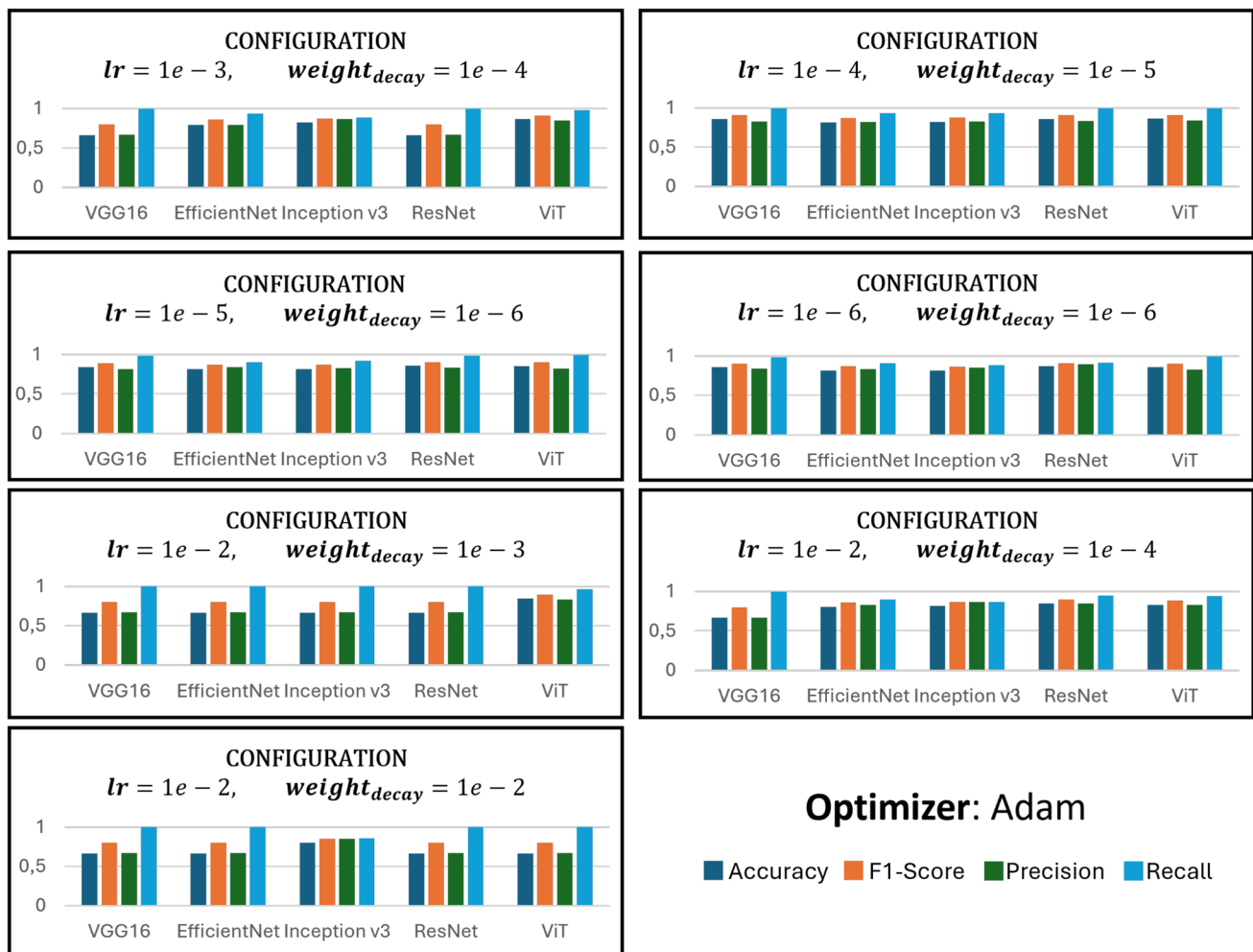


Fig. 6 Performance results of various state-of-the-art deep learning models trained for cloud detection using ceilometer backscatter profiles, evaluated with the ADAM optimizer and organized by training parameters such as learning rate and weight decay. Models include VGG16,

EfficientNet, InceptionV3, ResNet50, and Vision Transformer (ViT), with metrics like Accuracy, F1-score, Precision, and Recall used for assessment

cially for datasets with complex patterns like backscatter profiles.

ResNet50 emerged as the top-performing model, achieving the highest accuracy of 89.57% using SGD with configuration 2 (learning rate = 10^{-4} , momentum = 0.8, weight decay = 10^{-5}). The corresponding F1-score of 0.9273 highlights the model's superior robustness and generalization capabilities. ResNet50 maintained consistent performance even when optimized with Adam, achieving an accuracy of 87.23% under configuration 4. The consistently high recall values indicate the model's strong ability to detect relevant instances, making it well-suited for real-world applications in cloud detection.

The Vision Transformer (ViT) model demonstrated competitive results, closely following ResNet50. The highest accuracy achieved by ViT was 89.36% using SGD with

configuration 3 (learning rate = 10^{-5}), accompanied by strong precision (0.8795) and recall (0.9808). ViT's performance underscores the potential of transformer-based architectures for complex tasks such as cloud detection. Even when optimized with Adam, ViT maintained robust performance, achieving an accuracy of 85.96% with configuration 4. This suggests that transformer-based models, when properly tuned, can rival traditional convolutional networks in such specialized tasks.

Overall, a comparison of optimizers across models revealed that SGD tended to produce higher accuracy scores compared to Adam, though Adam often provided faster convergence, requiring fewer epochs. Lower learning rates generally resulted in more stable and higher accuracies, while higher learning rates (e.g., 0.01) frequently led to poor performance, indicating potential issues with stability and

Table 2 The first half of the table shows the parameters used for experiments with SGD. The second half of the table shows the parameters used for experiments with Adam

#	SGD			ADAM	
	Learning rate	Momentum	Weight decay	Learning rate	Weight decay
1	10^{-3}	0.9	10^{-4}	10^{-3}	10^{-4}
2	10^{-4}	0.8	10^{-5}	10^{-4}	10^{-5}
3	10^{-5}	0.7	10^{-6}	10^{-5}	10^{-6}
4	10^{-6}	0.9	10^{-6}	10^{-6}	10^{-6}
5	10^{-2}	0.9	10^{-3}	10^{-2}	10^{-3}
6	10^{-2}	0.8	10^{-2}	10^{-2}	10^{-2}
7	10^{-2}	0.8	10^{-4}	10^{-2}	10^{-4}

Table 3 Test performance of the considered state-of-the-art models. Bold values highlight the best-performing model for each evaluation metric when using either SGD or Adam as optimizers. Please refer to Table 2 for the hyperparameter configurations

Model	Configuration	Optimizer	Accuracy	F1-score	Precision	Recall	Training time (min.) / Last epoch
VGG16	#3	SGD	86.38	0.91	0.85	0.97	22 m / 39
	#4	Adam	86.17	0.91	0.84	0.98	24 m / 38
EfficientNet	#5	SGD	83.19	0.88	0.85	0.92	16 m / 14
	#4	Adam	81.91	0.87	0.84	0.91	37 m / 59
Inception v3	#2	SGD	81.06	0.87	0.83	0.90	30 m / 59
	#1	Adam	82.77	0.88	0.87	0.88	10 m / 15
ResNet50	#2	SGD	89.57	0.93	0.90	0.96	27 m / 59
	#4	Adam	87.23	0.91	0.90	0.92	23 m / 59
ViT	#3	SGD	89.36	0.93	0.88	0.98	129 m / 59
	#1	Adam	86.81	0.91	0.85	0.98	25 m / 9

Bold values highlights the best results

overfitting during training. In terms of model performance, ResNet50 consistently outperformed other architectures, demonstrating the efficacy of residual connections for feature extraction from backscatter data. ViT, while slightly behind ResNet50 in accuracy, showed promise, especially given its strong recall and balanced performance metrics. The high recall scores across many configurations suggest a strong capability to capture positive instances (cloud presence), but they also highlight the need to balance precision, as seen with models like VGG16 and EfficientNet.

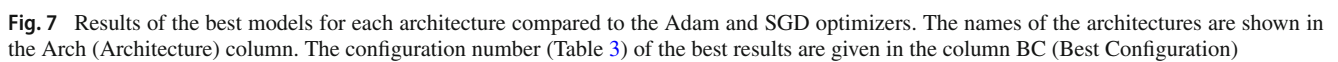
In conclusion, the results highlight that ResNet50 and Vision Transformer are highly effective models for cloud detection using ceilometer backscatter profiles. Their robust performance, particularly in terms of recall, demonstrates their strong suitability for real-world atmospheric monitoring applications. This analysis underscores the potential of these models to accurately identify and classify cloud presence, providing a reliable foundation for further advancements in environmental monitoring and data-driven atmospheric analysis.

4.2 Final model

The final hyperparameters used for training with SGD and Adam optimizers are reported in Table 2. An early stopping criterion was applied, whereby training would terminate if the variation in validation loss remained within a margin of $\delta = 0.05$ for at least three consecutive epochs. In total, we conducted 70 experiments, all of which are available at the following GitHub repository: <https://github.com/alessiochisari/CeilometerDatasetBenchmark>.

These experiments were performed on Google Colab Pro equipped with a *Tesla T4* GPU with 16GB GDDR6 memory. Table 3 shows the best results obtained by training the models with the best set of hyperparameters (*c.f.* Table 2) for each of the two chosen optimizers.

Figure 7 provides a detailed view of the performance trends for several deep learning architectures evaluated across four key metrics: *Accuracy (%)*, *Precision (%)*, *Recall (%)*, and *F1-score (%)*. Each metric is plotted as a function of the number of training epochs, shown on the x-axis, up to a maximum of 60 epochs. However, the training process



incorporates an early stopping mechanism, which terminates the training when the difference in loss between consecutive epochs falls below a predefined threshold for a fixed number of epochs. As a result, the number of epochs varies across models and optimizer configurations, reflecting differences in learning dynamics and convergence. For each architecture, the graphs show the best-performing solutions corresponding to the two optimizers, Adam and SGD, with all the previously mentioned metrics (accuracy, precision, recall and F1-score) obtained according to the best parameter configuration shown in Table 3.

The metrics themselves provide valuable insights into the strengths and weaknesses of the models. For instance, in the case of VGG16, the recall metric consistently increases across epochs, demonstrating the model's ability to effectively identify positive instances (e.g., cloud presence). However, the precision curve shows some fluctuations, suggesting a propensity for occasional false positives. EfficientNet, on the other hand, achieves a balance between precision and recall, with its performance metrics stabilizing effectively as training progresses. This indicates strong generalization capabilities, though the model may take slightly longer to converge compared to others.

InceptionV3 presents an interesting case, with moderate performance across all metrics. It demonstrates reliability in capturing relevant patterns in the data, as evidenced by its recall trends, though its overall accuracy is slightly lower compared to top-performing models. Adam optimization appears to benefit this architecture, as the model converges more quickly and achieves its best results with fewer epochs compared to SGD.

ResNet50 stands out as the best-performing architecture across all metrics. Its accuracy and F1-score remain consistently high, and the recall metric underscores its exceptional ability to detect positive instances with minimal false negatives. This performance is likely due to the advantages of residual connections, which help the model capture hierarchical features more effectively. The Vision Transformer (ViT) also delivers impressive results, rivaling ResNet50 in accuracy and recall. Its performance demonstrates the potential of transformer-based architectures for tasks involving complex patterns in atmospheric data. While Adam leads to faster convergence for ViT, the final performance metrics are marginally better when the model is optimized with SGD.

Regarding training times and last epoch (last column of Table 3), it can be noted that they vary across models, influenced by the early stopping criterion used to prevent overfitting. The VGG16 model requires 22–24 min with 38–39 epochs, demonstrating quick convergence. Similarly, Inception v3 shows a range from 10 min for configuration #1 to 30 min for configuration #2, both stopping after 59 epochs, as early stopping was likely triggered to prevent overfitting. The EfficientNet model, with more complex architecture,

requires longer training times (16–37 min) and up to 59 epochs, reflecting the balance between model complexity and training duration. ResNet50 achieves efficient performance with training times of 23–27 min, stopping after 59 epochs for both configurations. The Vision Transformer (ViT) takes the longest training time (129 min for configuration #3), with 59 epochs, due to its computational intensity, though configuration #1 converges in 25 min and 9 epochs. All models are pretrained, and the use of early stopping ensures that training is halted before overfitting occurs. Thus, while VGG16 and Inception v3 train faster, EfficientNet and ViT offer higher accuracy at the cost of longer training durations. After completing the training of all models, the inference time is generally negligible and takes between 10 to 20 s for a batch size of 12 samples, or approximately 1 to 1.7 s per sample.

The final column of plots (Fig. 7), showing the confusion matrices provided by the best model for each involved architecture, offers additional insights into the classification performance of each model. The diagonal elements of these matrices represent correctly classified instances, while off-diagonal elements indicate misclassifications. For models like ResNet50 and ViT, the confusion matrices reveal a strong ability to correctly classify both positive and negative instances, reinforcing their suitability for the task.

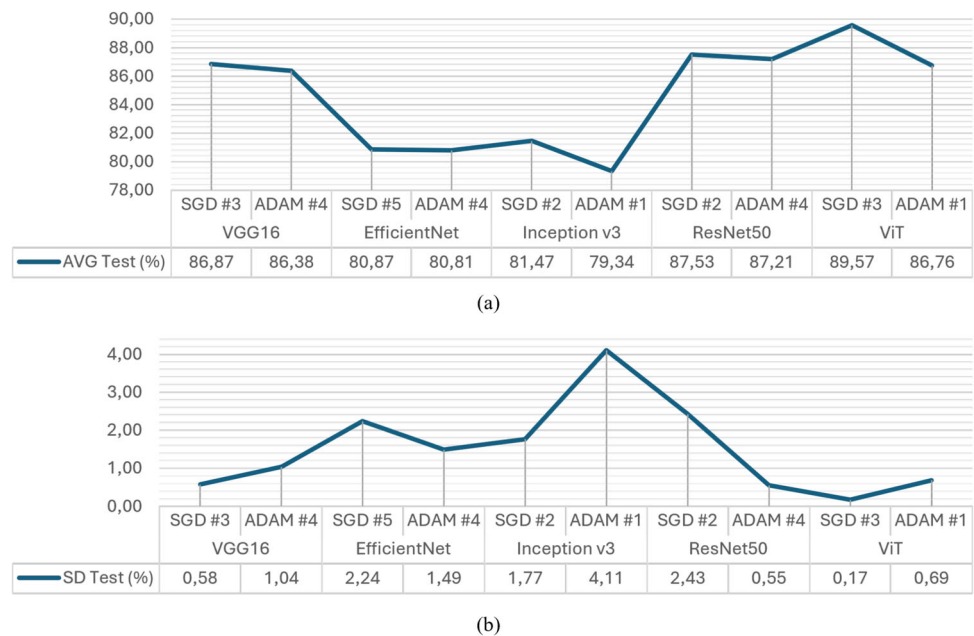
Overall, the analysis highlights the interplay between model architecture, optimizer choice, and the early stopping mechanism. ResNet50 emerges as the most robust and reliable model, followed closely by ViT, while architectures like EfficientNet and InceptionV3 offer competitive alternatives with specific strengths. The early stopping mechanism ensures efficient training, preventing overfitting and reducing computational costs, while still enabling a thorough evaluation of model performance. These results demonstrate the promise of advanced neural network architectures for challenging tasks such as cloud detection from lidar backscatter data.

5 Discussion

The results presented in this study highlight several critical observations regarding the performance of state-of-the-art deep learning architectures for cloud detection using lidar-based ceilometer backscatter data. Below, we discuss the implications of these findings, the strengths and limitations of the proposed methodology.

We aim to emphasize the distinctions between our dataset and those commonly found in the literature. Conventional remote sensing systems employing lidar technology are primarily satellite-based, capturing data from a top-down (satellite-to-Earth) viewpoint. While these methods have been extensively researched and yield valuable atmospheric and surface data, they differ significantly from our approach,

Fig. 8 Average (AVG) **a** and Standard Deviation (SD) **b** of the 10 best experiments of the best model configurations



which takes a bottom-up (Earth-to-satellite) perspective. This shift in the data acquisition model introduces a novel and compelling aspect to cloud detection, as the dataset we propose provides insights into atmospheric dynamics from a rarely explored viewpoint in existing studies. As a result of this unique perspective, our dataset stands apart from others in the field, making it innovative. Therefore, comparing it directly with existing methods is not feasible, as the dataset's nature is substantially different, rendering those traditional approaches unsuitable.

5.1 Performance analysis of models

Our experimental results revealed that ResNet50 achieved the highest accuracy (89.57%) among CNN-based architectures, closely followed by the Vision Transformer (ViT) with 89.36%. This performance gap suggests that residual connections in ResNet50 provide significant advantages in extracting hierarchical features from complex backscatter profiles. Meanwhile, ViT's ability to model global dependencies demonstrates the potential of transformer-based approaches for atmospheric data analysis. These findings are consistent with the growing success of hybrid and transformer models in computer vision.

Other architectures, such as VGG16, EfficientNet, and InceptionV3, showed lower, albeit competitive, performance. Notably, VGG16 exhibited high recall values, indicating its reliability in identifying cloud presence, but at the cost of increased false positives. EfficientNet, while slightly behind in accuracy, offered a balanced trade-off between precision and recall, which could be beneficial for specific real-time

applications where false negatives are particularly detrimental.

The results shown in Fig. 8 present the average test accuracy and corresponding standard deviation for the top 10 experiments using the best configurations of Adam and SGD optimizers across various architectures. These metrics offer complementary insights: average accuracy reflects general performance, while standard deviation reveals how stable the model is across multiple runs.

Starting with VGG16, both optimizers achieve comparable accuracy: SGD reaches 86.87%, slightly outperforming Adam at 86.38%. However, SGD shows more consistent results, with a lower standard deviation (0.58%) compared to Adam (1.04%). This suggests that while performance is similar, SGD yields more stable outcomes. For EfficientNet, results are nearly identical in terms of accuracy—SGD at 80.87% and Adam at 80.81%. Yet, Adam exhibits slightly better stability with a lower standard deviation of 1.49% versus 2.24% for SGD. This makes Adam marginally more reliable in repeated runs, despite similar performance. The Inception v3 architecture reveals a clearer distinction. SGD significantly outperforms Adam in both metrics: it achieves a higher average accuracy (81.47% vs. 79.34%) and a notably lower standard deviation (1.77% vs. 4.11%). These results indicate that SGD is both more accurate and considerably more stable, making it the preferred choice for this architecture. In contrast, ResNet50 shows minimal differences in accuracy—SGD slightly leads with 87.53% over Adam's 87.21%. However, stability tells a different story: Adam has an impressively low standard deviation of 0.55%, compared to SGD's 2.43%. This highlights Adam's robustness for ResNet50, despite the small accuracy gap. Lastly, in the

case of the Vision Transformer (ViT), SGD achieves the highest overall accuracy at 89.57%, surpassing Adam's 86.76%. Additionally, SGD offers superior consistency with a remarkably low standard deviation of 0.17%, while Adam records 0.69%. Here, SGD stands out as both the most accurate and the most stable optimizer.

The comparison across architectures underscores the importance of evaluating both accuracy and stability when selecting an optimizer:

- Adam is particularly effective with ResNet50 and EfficientNet, where it provides higher consistency.
- SGD clearly outperforms Adam in Inception v3 and ViT, excelling in both accuracy and robustness.
- In VGG16, the two optimizers are comparable, but SGD provides slightly better and more stable results.

These findings suggest that optimizer selection should balance both performance and repeatability, especially in real-world applications where consistency across multiple training runs is crucial.

5.2 Optimizer sensitivity and hyperparameter impact

A notable observation was the sensitivity of model performance to optimizer choice and hyperparameter configurations. Models optimized with SGD generally outperformed those trained with Adam, particularly in achieving higher accuracy and stability. However, Adam demonstrated faster convergence, which could be advantageous for computationally constrained scenarios. This optimizer-dependent performance underscores the importance of hyperparameter tuning for specialized tasks like cloud detection.

5.3 Dataset characteristics and challenges

The dataset curated for this study, derived from lidar ceilometer backscatter profiles, presented unique challenges due to its high temporal resolution and the variability of atmospheric conditions. The presence of diverse cloud types, combined with the influence of Mount Etna's complex meteorological phenomena, created a demanding environment for model training and evaluation. Despite these challenges, the models achieved promising results, demonstrating the potential of lidar data for real-world cloud detection tasks.

One limitation of the dataset is its geographical specificity, as data were collected exclusively in the vicinity of San Giovanni La Punta, Catania, Italy. Future studies could benefit from expanding the dataset to include backscatter profiles from multiple regions with varying climatic conditions. This would enhance the generalizability of the models and facilitate cross-regional comparisons.

5.4 Practical implications and applications

The high accuracy and recall achieved by ResNet50 and ViT make these models viable candidates for deployment in operational meteorological systems. Their robust performance in detecting cloud presence from lidar data could complement existing satellite and radar systems, particularly for identifying low-altitude clouds and localized aerosol concentrations.

Moreover, the potential for applying this approach to detect other atmospheric phenomena, such as pollutants or volcanic emissions, is noteworthy. The ability of lidar-based systems to capture fine-grained vertical profiles of the atmosphere could pave the way for monitoring air quality, early warning systems for natural disasters, and climate research.

5.5 Strengths, limitations and future directions

The proposed approach for cloud detection using lidar-based ceilometer backscatter data demonstrates several notable strengths that distinguish it from traditional remote sensing methodologies. One of the main strengths lies in its novel bottom-up data acquisition. Unlike conventional satellite-based techniques, our method captures atmospheric dynamics from an Earth-to-satellite perspective, allowing for high-frequency, real-time monitoring of the lower atmosphere. This unique vantage point enables the detailed observation of phenomena such as low-altitude cloud formations and localized aerosol concentrations that are often underrepresented in traditional approaches.

Another strength is the robust performance of the deep learning models. The benchmarking experiments highlighted that architectures such as ResNet50 and Vision Transformer (ViT) are particularly effective in extracting hierarchical and global features from complex backscatter profiles. Their high accuracy and recall not only validate the efficacy of the method but also underscore the potential for integrating these models into operational meteorological systems for tasks like early warning detection and real-time environmental monitoring.

Despite these significant advantages, several limitations must be acknowledged. The geographical scope and temporal window of the dataset are relatively limited, as the data were collected exclusively near San Giovanni La Punta, Catania, Italy over a three-month period. This geographical and temporal confinement may affect the generalizability of the models when applied to regions with different climatic conditions or extended seasonal variations. Moreover, the performance of the models is notably sensitive to the choice of hyperparameters and optimizer configurations. As observed in our experiments, variations in learning rate, momentum, and weight decay can lead to significant fluctuations in accuracy and stability, indicating a potential

challenge in achieving consistent performance across diverse training scenarios.

Looking ahead, several future directions can be pursued to build upon the current work. First, expanding the dataset both in terms of geographic diversity and duration would enhance model robustness and facilitate cross-regional comparisons, making the approach more broadly applicable. Additionally, investigating alternative and hybrid architectures, such as integrating CNNs with transformer-based models, could leverage the complementary strengths of local feature extraction and global context modeling, potentially leading to further performance improvements. Finally, significant efforts should be directed toward real-time deployment. This involves optimizing the computational efficiency of the models and reducing latency to facilitate their incorporation into practical, operational meteorological systems, especially in edge computing scenarios where resources are limited.

6 Conclusion

In this study, we proposed and evaluated a novel approach for cloud detection using lidar-based ceilometer backscatter data, benchmarked across state-of-the-art deep learning models. By leveraging a newly curated dataset characterized by high temporal resolution and diverse atmospheric conditions, we demonstrated the efficacy of advanced neural network architectures in accurately detecting cloud presence.

The experimental results highlight ResNet50 as the top-performing model, achieving an accuracy of 89.57%, with the Vision Transformer (ViT) closely following at 89.36%. These findings underscore the advantages of residual learning and transformer-based global attention mechanisms in capturing complex patterns in backscatter profiles. Additionally, models such as VGG16, EfficientNet, and InceptionV3 provided competitive performance, offering alternative solutions based on specific application requirements, such as reduced false negatives or computational efficiency.

This work contributes to the field in two significant ways:

- **Dataset Availability:** We provide a publicly accessible dataset of labeled backscatter profiles acquired over three months near Mount Etna, Italy. This dataset represents a valuable resource for advancing research in cloud detection and atmospheric studies.
- **Comprehensive Benchmarking:** By evaluating several deep learning architectures under varying hyperparameter configurations and optimization strategies, we establish a robust baseline for future developments in lidar-based cloud detection.

Despite these achievements, our study identifies areas for improvement, including expanding the dataset to cover diverse geographic regions, exploring alternative architectures, and optimizing models for real-time deployment. Future research could also explore the application of this approach to detect other atmospheric phenomena, such as aerosols, pollutants, or volcanic emissions, further broadening the utility of lidar-based systems in environmental monitoring.

In conclusion, our results validate the potential of combining lidar ceilometer data with advanced deep learning techniques to enhance cloud detection capabilities. This integration represents a significant step forward in developing accurate, scalable, and automated solutions for atmospheric monitoring, with implications for meteorology, climate research, and environmental protection.

Author contributions A.C. code and experiments. A.C., L.G. wrote the main manuscript text. A.C., L.G. prepared figures. All authors reviewed the manuscript. S.B., V.G. supervisors.

Funding Open access funding provided by Università degli Studi di Catania within the CRUI-CARE Agreement.

Data Availability The dataset generated during this study is publicly available at <https://zenodo.org/records/10616434>. It is licensed under the Creative Commons Attribution- NonCommercial-ShareAlike 2.0 Generic license.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Salvoni, D., Sannino, A., Parlato, L., Amoroso, S., Pepe, G.P., Boselli, A., Ejrnaes, M., Wang, X., Cristiano, R., Zhang, C., et al.: Lidar measurement of clouds profile with a superconducting nanowire single photon detector. In: 2021 IEEE 14th Workshop on Low Temperature Electronics (WOLTE), pp. 1–4 (2021). <https://doi.org/10.1109/WOLTE49037.2021.9555454>. IEEE
2. Zou, W., Yin, Z., Dai, Y., Chen, Y., Bu, Z., Li, S., He, Y., Hu, X., Müller, D., Lu, T., et al.: Robust lidar-radar composite cloud boundary detection method with rainfall pixels removal. IEEE Trans.

- Geosci. Remote Sensing **62**, 1–16 (2024). <https://doi.org/10.1109/TGRS.2024.3476127>
3. Yan, W.Y.: Airborne lidar data artifacts: what we know thus far. IEEE Geosci. Remote Sensing Magaz. **11**(3), 21–45 (2023). <https://doi.org/10.1109/MGRS.2023.3285261>
4. Cromwell, E., Flynn, D.: Lidar cloud detection with fully convolutional networks. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 619–627. IEEE, Waikoloa Village, HI, USA (2019). <https://doi.org/10.1109/WACV.2019.00071>
5. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015) (2015). Computational and Biological Learning Society
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
8. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR
9. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
10. Chisari, A.B., Ortis, A., Guarnera, L., Patatu, W.C., Giandolfo, R.A., Spampinato, E., Battiatto, S., Giuffrida, M.V.: On the cloud detection from backscattered images generated from a lidar-based ceilometer: current state and opportunities. In: 2024 IEEE International Conference on Image Processing (ICIP), pp. 144–150. IEEE, Abu Dhabi, United Arab Emirates (2024). <https://doi.org/10.1109/ICIP51287.2024.10647352>
11. Mahajan, S., Fataniya, B.: Cloud detection methodologies: variants and development-a review. Complex Intell. Syst. **6**(2), 251–261 (2020). <https://doi.org/10.1007/s40747-019-00128-0>
12. Munkel, C., Emeis, S., Müller, W.J., Schaefer, K.P.: Aerosol concentration measurements with a lidar ceilometer: results of a one year measuring campaign. In: Remote Sensing of Clouds and the Atmosphere VIII, vol. 5235, pp. 486–496 (2004). SPIE
13. Munkel, C., Eresmaa, N., Räsänen, J., Karppinen, A.: Retrieval of mixing height and dust concentration with lidar ceilometer. Bound. Layer Meteorol. **124**(1), 117–128 (2007). <https://doi.org/10.1007/s10546-006-9103-3>
14. Flentje, H., Claude, H., Elste, T., Gilge, S., Köhler, U., Plass-Dülmer, C., Steinbrecht, W., Thomas, W., Werner, A., Fricke, W.: The Eyjafjallajökull eruption in April 2010—detection of volcanic plume using in-situ measurements, ozone sondes and lidar-ceilometer profiles. Atmos. Chem. Phys. **10**(20), 10085–10092 (2010). <https://doi.org/10.5194/acp-10-10085-2010>
15. Bartok, J., Habala, O., Bednar, P., Gazak, M., Hluchý, L.: Data mining and integration for predicting significant meteorological phenomena. Proc. Comput. Sci. **1**(1), 37–46 (2010). <https://doi.org/10.1016/j.procs.2010.04.006>
16. Wiegner, M., Geiß, A.: Aerosol profiling with the Jenoptik ceilometer CHM15kx. Atmos. Measure. Tech. **5**(8), 1953–1964 (2012). <https://doi.org/10.5194/amt-5-1953-2012>
17. Arun, S.H., Sharma, S.K., Chaurasia, S., Vaishnav, R., Kumar, R.: Fog/low clouds detection over the Delhi earth station using the ceilometer and the INSAT-3D/3DR satellite data. Int. J. Remote Sens. **39**(12), 4130–4144 (2018). <https://doi.org/10.1080/01431161.2018.1454624>
18. Kim, Y.-H., Moon, S.-H., Yoon, Y.: Detection of precipitation and fog using machine learning on backscatter data from lidar ceilometer. Appl. Sci. **10**(18), 6452 (2020). <https://doi.org/10.3390/app10186452>
19. Huertas-Tato, J., Rodríguez-Benítez, F.J., Arbizu-Barrena, C., Aler-Mur, R., Galvan-Leon, I., Pozo-Vázquez, D.: Automatic cloud-type classification based on the combined use of a sky camera and a ceilometer: cloud class using ceilometer and camera. J. Geophys. Res. Atmos. **122**(20), 11045–11061 (2017). <https://doi.org/10.1002/2017JD027131>
20. Casella, B., Chisari, A.B., Aldinucci, M., Battiatto, S., Giuffrida, M.V.: Federated learning in a semi-supervised environment for earth observation data. In: ESANN 2024 Proceedings-32th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp. 93–98 (2024). Michel Verlesian
21. Sleeman, J., Halem, M., Yang, Z., Caicedo, V., Demoz, B., Delgado, R.: A deep machine learning approach for lidar based boundary layer height detection. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 3676–3679 (2020). <https://doi.org/10.1109/IGARSS39084.2020.9324191>
22. Dammann, M., Mattis, I., Neitzke, M., Möller, R.: Towards the automatic analysis of ceilometer backscattering profiles using unsupervised learning. In: NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning (2022). <https://www.climatechange.ai/papers/neurips2022/99>
23. An, N., Shang, H., Lesi, W., Ri, X., Shi, C., Tana, G., Bao, Y., Zheng, Z., Xu, N., Chen, L., et al.: A cloud detection algorithm for early morning observations from the fy-3e satellite. IEEE Trans. Geosci. Remote Sens. **61**, 1–15 (2023)
24. Li, A., Li, X., Ma, X.: Residual dual u-shape networks with improved skip connections for cloud detection. IEEE Geosci. Remote Sens. Lett. **21**, 1–5 (2023)
25. National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce: NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Place: Boulder CO (2015). <https://doi.org/10.5065/D65D8PWK>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Alessio Barbaro Chisari is a Ph.D. student in Computer Science at the University of Catania and member of Image Processing Laboratory (IPLab). Alessio received the B.Sc. degree in Computer Engineering in 2019 and the M. Sc. degree in Management Engineering in 2021 from the University of Catania. Alessio was also a visiting Ph.D. student at the University of Nottingham (UK) in 2023. His current research interests include Computer Vision and Deep Learning techniques applied to atmospheric phenomena. Alessio is also a Technical Business Developer and Technical Project Manager of the R&D Business Unit for EHT S.C.p.A., in which he manages and checks the scientific relevance of R&D projects in the IT field.



Luca Guarnera is a research fellow in Computer Science at the University of Catania. He graduated as Ph.D. in Computer Science on October 14, 2021, discussing the thesis entitled “Discovering Fingerprints for Deepfake Detection and Multimedia-Enhanced Forensic Investigations” at the Department of Mathematics and Computer Science, University of Catania. Part of the Ph.D. research was carried out at the University of Hertfordshire, College Lane Campus, Hatfield, UK, under the supervision of Prof. Salvatore Livatino, working on the creation of a software for forensic ballistics analysis and firearms comparison, using the Oculus Rift S headset. His main research interests are Computer Vision, Machine Learning, Multimedia Forensics, and its related fields with a focus on the Deepfake phenomenon.

vision of Prof. Salvatore Livatino, working on the creation of a software for forensic ballistics analysis and firearms comparison, using the Oculus Rift S headset. His main research interests are Computer Vision, Machine Learning, Multimedia Forensics, and its related fields with a focus on the Deepfake phenomenon.



Alessandro Ortis is a Researcher and Lecturer at the University of Catania, specializing in Artificial Intelligence for Computer Vision and security. He earned a Ph.D. in Mathematics and Computer Science, with part of his research conducted at Imperial College London. Alessandro is a Senior IEEE Member, a member of the IEEE Biometric Council, and part of the IPLAB (Image Processing Laboratory) research group, which is involved in international collaborations and funded projects. His work focuses on AI vulnerabilities, such as Deepfakes, Adversarial Machine Learning, and Biometrics. He also teaches Multimedia Security and Biometrics at the University of Catania.

work focuses on AI vulnerabilities, such as Deepfakes, Adversarial Machine Learning, and Biometrics. He also teaches Multimedia Security and Biometrics at the University of Catania.



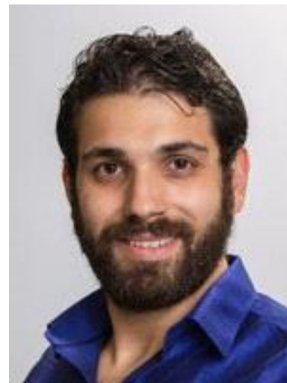
Wladimiro Carlo Patatu is an Electronic Engineer with an MBA and extensive experience in driving complex IT innovation projects. As an EHT manager of the R&D Business Unit leads the team that provides scientific guidance for the innovation projects. With longstanding experience as a scientific coordinator for complex R&D funded projects, his research focuses in leveraging innovative IT technologies, models and methods to drive innovation across multiple sectors, including agritech, life science, smart

communities, cultural heritage, smart industry, energy, and environment.



Sebastiano Battiato received his Ph.D. in computer science and applied mathematics from the University of Naples in 1999, after earning his degree in computer science from the University of Catania in 1995. He has held various positions at the University of Catania, starting as an Assistant Professor in 2004, then becoming an Associate Professor in 2011, and finally a Full Professor in 2016. Throughout his career, he has been involved in leading roles within academic programs and research

initiatives, including chairing the Undergraduate Program in Computer Science and serving as the Rector's delegate for Education. Currently, he serves as the Deputy Rector for Strategic Planning and Information Systems. Prof. Battiato is highly active in research, directing the IPLab research laboratory and participating in numerous national and international projects. He has supervised numerous Ph.D. students and postdocs, edited several books, and authored hundreds of papers in various publications. His research interests lie in computer vision, imaging technology, and multimedia forensics. Prof. Battiato is recognized for his contributions, receiving accolades such as the 2017 PAMI Mark Everingham Prize and the 2011 Best Associate Editor Award from the IEEE Transactions on Circuits and Systems for Video Technology. He also plays key roles in organizing international events and serves as an Associate Editor for the SPIE Journal of Electronic Imaging. Additionally, he is the Director and Co-Founder of both the International Computer Vision Summer School (ICVSS) and the International Forensics Summer School (IFOSS).



Mario Valerio Giuffrida is an Assistant Professor in Computer Vision at the School of Computer Science, University of Nottingham, and a member of the Computer Vision Lab (CVL). His research focuses on deep learning applied to biological problems, including plant and medical imaging analysis. As a Co-Investigator in the PhenomUK scoping project, Dr. Giuffrida contributes to advancing phenotyping technologies in the UK. He has also organized international workshops on computer vision in plant

phenotyping at leading international conferences, fostering collaboration and knowledge exchange in this field.