

Identification and denoising of radio signals from cosmic-ray air showers using convolutional neural networks

R. Abbasi¹⁶, M. Ackermann⁶³, J. Adams¹⁷, S. K. Agarwalla^{39,*}, J. A. Aguilar¹⁰, M. Ahlers²¹, J. M. Alameddine²², S. Ali³⁵, N. M. Amin⁴³, K. Andeen⁴¹, C. Argüelles¹³, Y. Ashida⁵², S. Athanasiadou⁶³, S. N. Axani⁴³, R. Babu²³, X. Bai⁴⁹, J. Baines-Holmes³⁹, A. Balagopal V.^{39,43}, S. W. Barwick²⁹, S. Bash²⁶, V. Basu⁵², R. Bay⁶, J. J. Beatty^{19,20}, J. Becker Tjus^{9,†}, P. Behrens¹, J. Beise⁶¹, C. Bellenghi²⁶, B. Benkel⁶³, S. BenZvi⁵¹, D. Berley¹⁸, E. Bernardini^{47,‡}, D. Z. Besson³⁵, E. Blaufuss¹⁸, L. Bloom⁵⁸, S. Blot⁶³, I. Bodo³⁹, F. Bontempo³⁰, J. Y. Book Motzkin¹³, C. Boscolo Meneguolo^{47,‡}, S. Böser⁴⁰, O. Botner⁶¹, J. Böttcher¹, J. Braun³⁹, B. Brinson⁴, Z. Brisson-Tsavoussis³², R. T. Burley², D. Butterfield³⁹, M. A. Campana⁴⁸, K. Carloni¹³, J. Carpio^{33,34}, S. Chattopadhyay^{39,*}, N. Chau¹⁰, Z. Chen⁵⁵, D. Chirkin³⁹, S. Choi⁵², B. A. Clark¹⁸, A. Coleman⁶¹, P. Coleman¹, G. H. Collin¹⁴, D. A. Coloma Borja⁴⁷, A. Connolly^{19,20}, J. M. Conrad¹⁴, R. Corley⁵², D. F. Cowen^{59,60}, C. De Clercq¹¹, J. J. DeLaunay⁵⁹, D. Delgado¹³, T. Delmeulle¹⁰, S. Deng¹, P. Desiati³⁹, K. D. de Vries¹¹, G. de Wasseige³⁶, T. DeYoung²³, J. C. Díaz-Vélez³⁹, S. DiKerby²³, M. Dittmer⁴², A. Domi²⁵, L. Draper⁵², L. Dueser¹, D. Durnford²⁴, K. Dutta⁴⁰, M. A. DuVernois³⁹, T. Ehrhardt⁴⁰, L. Eidenschink²⁶, A. Eimer²⁵, P. Eller²⁶, E. Ellinger⁶², D. Elsässer²², R. Engel^{30,31}, H. Erpenbeck³⁹, W. Esmail⁴², S. Eulig¹³, J. Evans¹⁸, P. A. Evenson⁴³, K. L. Fan¹⁸, K. Fang³⁹, K. Farrag¹⁵, A. R. Fazely⁵, A. Fedynitch⁵⁷, N. Feigl⁸, C. Finley⁵⁴, L. Fischer⁶³, D. Fox⁵⁹, A. Franckowiak⁹, S. Fukami⁶³, P. Fürst¹, J. Gallagher³⁸, P. Gálvez Molina⁴³, E. Ganster¹, A. Garcia¹³, M. Garcia⁴³, G. Garg^{39,*}, E. Genton^{13,36}, L. Gerhardt⁷, A. Ghadimi⁵⁸, C. Glaser⁶¹, T. Glüsenkamp⁶¹, J. G. Gonzalez⁴³, S. Goswami^{33,34}, A. Granados²³, D. Grant¹², S. J. Gray¹⁸, S. Griffin³⁹, S. Griswold⁵¹, K. M. Groth²¹, D. Guevel³⁹, C. Günther¹, P. Gutjahr²², C. Ha⁵³, C. Haack²⁵, A. Hallgren⁶¹, L. Halve¹, F. Halzen³⁹, L. Hamacher¹, M. Ha Minh²⁶, M. Handt¹, K. Hanson³⁹, J. Hardin¹⁴, A. A. Harnisch²³, P. Hatch³², A. Haungs³⁰, J. Häußler¹, K. Helbing⁶², J. Hellrung⁹, B. Henke²³, L. Hennig²⁵, F. Henningsen¹², L. Heuermann¹, R. Hewett¹⁷, N. Heyer⁶¹, S. Hickford⁶², A. Hidvegi⁵⁴, C. Hill¹⁵, G. C. Hill², R. Hmaid¹⁵, K. D. Hoffman¹⁸, D. Hooper³⁹, S. Hori³⁹, K. Hoshina^{39,§}, M. Hostert¹³, W. Hou³⁰, M. Hrywniak⁵⁴, T. Huber³⁰, K. Hultqvist⁵⁴, K. Hymon^{22,57}, A. Ishihara¹⁵, W. Iwakiri¹⁵, M. Jacquart²¹, S. Jain³⁹, O. Janik²⁵, M. Jansson³⁶, M. Jeong⁵², M. Jin¹³, N. Kamp¹³, D. Kang³⁰, W. Kang⁴⁸, X. Kang⁴⁸, A. Kappes⁴², L. Kardum²², T. Karg⁶³, M. Karl²⁶, A. Karle³⁹, A. Katil²⁴, M. Kauer³⁹, J. L. Kelley³⁹, M. Khanal⁵², A. Khatzee Zathul³⁹, A. Kheirandish^{33,34}, H. Kimku⁵³, J. Kiryluk⁵⁵, C. Klein²⁵, S. R. Klein^{6,7}, Y. Kobayashi¹⁵, A. Kochocki²³, R. Koirala⁴³, H. Kolanoski⁸, T. Kontrimas²⁶, L. Köpke⁴⁰, C. Kopper²⁵, D. J. Koskinen²¹, P. Koundal⁴³, M. Kowalski^{8,63}, T. Kozynets²¹, N. Krieger⁹, J. Krishnamoorthi^{39,*}, T. Krishnan¹³, K. Kruiswijk³⁶, E. Krupczak²³, D. Kullgren^{43,60}, A. Kumar⁶³, E. Kun⁹, N. Kurahashi⁴⁸, N. Lad⁶³, C. Lagunas Gualda²⁶, L. Lallement Arnaud¹⁰, M. Lamoureux³⁶, M. J. Larson¹⁸, F. Lauber⁶², J. P. Lazar³⁶, K. Leonard DeHolton⁶⁰, A. Leszczyńska⁴³, J. Liao⁴, C. Lin⁴³, Y. T. Liu⁶⁰, M. Liubarska²⁴, C. Love⁴⁸, L. Lu³⁹, F. Lucarelli²⁷, W. Luszczak^{19,20}, Y. Lyu^{6,7}, J. Madsen³⁹, E. Magnus¹¹, Y. Makino³⁹, E. Manao²⁶, S. Mancina^{47,||}, A. Mand³⁹, I. C. Mariş¹⁰, S. Marka⁴⁵, Z. Marka⁴⁵, L. Marten¹, I. Martinez-Soler¹³, R. Maruyama⁴⁴, J. Mauro³⁶, F. Mayhew²³, F. McNally³⁷, J. V. Mead²¹, K. Meagher³⁹, S. Mechbal⁶³, A. Medina²⁰, M. Meier¹⁵, Y. Merckx¹¹, L. Merten⁹, J. Mitchell⁵, L. Molchany⁴⁹, T. Montaruli²⁷, R. W. Moore²⁴, Y. Morii¹⁵, A. Mosbrugger²⁵, M. Moulai³⁹, D. Mousadi⁶³, E. Moyaux³⁶, T. Mukherjee³⁰, R. Naab⁶³, M. Nakos³⁹, U. Naumann⁶², J. Necker⁶³, L. Neste⁵⁴, M. Neumann⁴², H. Niederhausen²³, M. U. Nisa²³, K. Noda¹⁵, A. Noell¹, A. Novikov⁴³, A. Obertacke⁵⁴, V. O'Dell³⁹, A. Olivas¹⁸, R. Orsoe²⁶, J. Osborn³⁹, E. O'Sullivan⁶¹, V. Palusova⁴⁰, H. Pandya⁴³, A. Parenti¹⁰, N. Park³², V. Parrish²³, E. N. Paudel⁵⁸, L. Paul⁴⁹, C. Pérez de los Heros⁶¹, T. Pernice⁶³, J. Peterson³⁹, M. Plum⁴⁹, A. Pontén⁶¹, V. Poojyam⁵⁸, Y. Popovych⁴⁰, M. Prado Rodriguez³⁹, B. Pries²³, R. Procter-Murphy¹⁸, G. T. Przybylski⁷, L. Pyras⁵², C. Raab³⁶, J. Rack-Helleis⁴⁰, N. Rad⁶³, M. Ravn⁶¹, K. Rawlins³, Z. Rechav³⁹, A. Rehman⁴³, I. Reistoffer⁴⁹, E. Resconi²⁶, S. Reusch⁶³, C. D. Rho⁵⁶, W. Rhode²², L. Ricca³⁶, B. Riedel³⁹, A. Rifaie⁶², E. J. Roberts², M. Rongen²⁵, A. Rosted¹⁵, C. Rott⁵², T. Ruhe²², L. Ruohan²⁶, D. Ryckbosch²⁸, J. Saffer³¹, D. Salazar-Gallegos²³, P. Sampathkumar³⁰, A. Sandrock⁶², G. Sanger-Johnson²³, M. Santander⁵⁸, S. Sarkar⁴⁶, J. Savelberg¹, M. Scarnera³⁶, P. Schaile²⁶, M. Schaufel¹, H. Schieler³⁰, S. Schindler²⁵, L. Schlickmann⁴⁰, B. Schlüter⁴², F. Schlüter¹⁰, N. Schmeisser⁶², T. Schmidt¹⁸, F. G. Schröder^{30,43}, L. Schumacher²⁵, S. Schwirn¹, S. Sclafani¹⁸, D. Seckel⁴³, L. Seen³⁹, M. Seikh³⁵, S. Seunarine⁵⁰, P. A. Sevle Myhr³⁶, R. Shah⁴⁸, S. Shefali³¹, N. Shimizu¹⁵, B. Skrzypek⁶, R. Snihur³⁹, J. Soedingrekso²², A. Sjøgaard²¹, D. Soldin⁵², P. Soldin¹, G. Sommani⁹, C. Spannfellner²⁶, G. M. Spiczak⁵⁰, C. Spiering⁶³, J. Stachurska²⁸, M. Stamatikos²⁰, T. Stanev⁴³, T. Stezelberger⁷, T. Stürwald⁶², T. Stuttard²¹, G. W. Sullivan¹⁸, I. Taboada⁴, S. Ter-Antonyan⁵, A. Terliuk²⁶, A. Thakuri⁴⁹, M. Thiesmeyer³⁹, W. G. Thompson¹³, J. Thwaites³⁹, S. Tilav⁴³, K. Tollefson²³, S. Toscano¹⁰, D. Tosi³⁹, A. Trettin⁶³

A. K. Upadhyay^{1b,39,*} K. Upshaw,⁵ A. Vaidyanathan,⁴¹ N. Valtonen-Mattila^{1b,9,61} J. Valverde^{1b,41} J. Vandenbroucke^{1b,39}
 T. Van Eeden,⁶³ N. van Eijndhoven^{1b,11} L. Van Rootselaar,²² J. van Santen^{1b,63} J. Vara,⁴² F. Varsi,³¹ M. Venugopal,³⁰
 M. Vereecken,³⁶ S. Vergara Carrasco,¹⁷ S. Verpoest^{1b,43} D. Veske,⁴⁵ A. Vijai,¹⁸ J. Villarreal^{1b,14} C. Walck,⁵⁴ A. Wang^{1b,4}
 E. H. S. Warrick^{1b,58} C. Weaver^{1b,23} P. Weigel,¹⁴ A. Weindl,³⁰ J. Weldert,⁴⁰ A. Y. Wen^{1b,13} C. Wendt^{1b,39} J. Werthebach,²²
 M. Weyrauch,³⁰ N. Whitehorn^{1b,23} C. H. Wiebusch^{1b,1} D. R. Williams,⁵⁸ L. Witthaus^{1b,22} M. Wolf^{1b,26} G. Wrede,²⁵
 X. W. Xu,⁵ J. P. Yanez^{1b,24} Y. Yao^{1b,39} E. Yildizci,³⁹ S. Yoshida^{1b,15} R. Young,³⁵ F. Yu^{1b,13} S. Yu^{1b,52} T. Yuan^{1b,39}
 A. Zegarelli^{1b,9} S. Zhang^{1b,23} Z. Zhang,⁵⁵ P. Zhelnin^{1b,13} and P. Zilberman³⁹

(IceCube Collaboration)

¹*III. Physikalisches Institut, RWTH Aachen University, D-52056 Aachen, Germany*

²*Department of Physics, University of Adelaide, Adelaide, 5005, Australia*

³*Department of Physics and Astronomy, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, Alaska 99508, USA*

⁴*School of Physics and Center for Relativistic Astrophysics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

⁵*Department of Physics, Southern University, Baton Rouge, Louisiana 70813, USA*

⁶*Department of Physics, University of California, Berkeley, California 94720, USA*

⁷*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

⁸*Institut für Physik, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*

⁹*Fakultät für Physik and Astronomie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*

¹⁰*Université Libre de Bruxelles, Science Faculty CP230, B-1050 Brussels, Belgium*

¹¹*Vrije Universiteit Brussel (VUB), Dienst ELEM, B-1050 Brussels, Belgium*

¹²*Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada*

¹³*Department of Physics and Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge, Massachusetts 02138, USA*

¹⁴*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

¹⁵*Department of Physics and The International Center for Hadron Astrophysics, Chiba University, Chiba 263-8522, Japan*

¹⁶*Department of Physics, Loyola University Chicago, Chicago, Illinois 60660, USA*

¹⁷*Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*

¹⁸*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*

¹⁹*Department of Astronomy, Ohio State University, Columbus, Ohio 43210, USA*

²⁰*Department of Physics and Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, Ohio 43210, USA*

²¹*Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark*

²²*Department of Physics, TU Dortmund University, D-44221 Dortmund, Germany*

²³*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*

²⁴*Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada*

²⁵*Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, D-91058 Erlangen, Germany*

²⁶*Physik-department, Technische Universität München, D-85748 Garching, Germany*

²⁷*Département de physique nucléaire et corpusculaire, Université de Genève, CH-1211 Genève, Switzerland*

²⁸*Department of Physics and Astronomy, University of Gent, B-9000 Gent, Belgium*

²⁹*Department of Physics and Astronomy, University of California, Irvine, California 92697, USA*

³⁰*Karlsruhe Institute of Technology, Institute for Astroparticle Physics, D-76021 Karlsruhe, Germany*

³¹*Karlsruhe Institute of Technology, Institute of Experimental Particle Physics, D-76021 Karlsruhe, Germany*

³²*Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, Ontario K7L 3N6, Canada*

³³*Department of Physics and Astronomy, University of Nevada, Las Vegas, Nevada 89154, USA*

³⁴*Nevada Center for Astrophysics, University of Nevada, Las Vegas, Nevada 89154, USA*

³⁵*Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA*

³⁶*Centre for Cosmology, Particle Physics and Phenomenology—CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

³⁷*Department of Physics, Mercer University, Macon, Georgia 31207-0001, USA*

³⁸*Department of Astronomy, University of Wisconsin—Madison, Madison, Wisconsin 53706, USA*

- ³⁹*Department of Physics and Wisconsin IceCube Particle Astrophysics Center, University of Wisconsin—Madison, Madison, Wisconsin 53706, USA*
- ⁴⁰*Institute of Physics, University of Mainz, Staudinger Weg 7, D-55099 Mainz, Germany*
- ⁴¹*Department of Physics, Marquette University, Milwaukee, Wisconsin 53201, USA*
- ⁴²*Institut für Kernphysik, Universität Münster, D-48149 Münster, Germany*
- ⁴³*Bartol Research Institute and Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA*
- ⁴⁴*Department of Physics, Yale University, New Haven, Connecticut 06520, USA*
- ⁴⁵*Columbia Astrophysics and Nevis Laboratories, Columbia University, New York, New York 10027, USA*
- ⁴⁶*Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*
- ⁴⁷*Dipartimento di Fisica e Astronomia Galileo Galilei, Università Degli Studi di Padova, I-35122 Padova PD, Italy*
- ⁴⁸*Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, USA*
- ⁴⁹*Physics Department, South Dakota School of Mines and Technology, Rapid City, South Dakota 57701, USA*
- ⁵⁰*Department of Physics, University of Wisconsin, River Falls, Wisconsin 54022, USA*
- ⁵¹*Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA*
- ⁵²*Department of Physics and Astronomy, University of Utah, Salt Lake City, Utah 84112, USA*
- ⁵³*Department of Physics, Chung-Ang University, Seoul 06974, Republic of Korea*
- ⁵⁴*Oskar Klein Centre and Department of Physics, Stockholm University, SE-10691 Stockholm, Sweden*
- ⁵⁵*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA*
- ⁵⁶*Department of Physics, Sungkyunkwan University, Suwon 16419, Republic of Korea*
- ⁵⁷*Institute of Physics, Academia Sinica, Taipei, 11529, Taiwan*
- ⁵⁸*Department of Physics and Astronomy, University of Alabama, Tuscaloosa, Alabama 35487, USA*
- ⁵⁹*Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*
- ⁶⁰*Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*
- ⁶¹*Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120 Uppsala, Sweden*
- ⁶²*Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany*
- ⁶³*Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, D-15738 Zeuthen, Germany*



(Received 21 August 2025; accepted 5 February 2026; published 10 June 2026)

Radio pulses generated by cosmic-ray air showers can be used to reconstruct key properties like the energy and depth of the electromagnetic component of cosmic-ray air showers. Radio detection threshold, influenced by natural and anthropogenic radio background, can be reduced through various techniques. In this work, we demonstrate that convolutional neural networks (CNNs) are an effective way to lower the threshold. We developed two CNNs: a classifier to distinguish radio signal waveforms from background noise and a denoiser to clean contaminated radio signals. Following the training and testing phases, we applied the networks to air-shower data triggered by scintillation detectors of the prototype station for the enhancement of IceTop, IceCube’s surface array at the South Pole. Over a four-month period, we identified 554 cosmic-ray events in coincidence with IceTop, approximately five times more compared to a reference method based on a cut on the signal-to-noise ratio. Comparisons with IceTop measurements of the same air showers confirmed that the CNNs reliably identified cosmic-ray radio pulses and outperformed the reference method. Additionally, we find that CNNs reduce the false-positive rate of air-shower candidates and effectively denoise radio waveforms, thereby improving the accuracy of the power and arrival time reconstruction of radio pulses.

DOI: [10.1103/xw99-zq7j](https://doi.org/10.1103/xw99-zq7j)

*Also at: Institute of Physics, Sachivalaya Marg, Sainik School Post, Bhubaneswar 751005, India.

†Also at: Department of Space, Earth and Environment, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

‡Also at: INFN Padova, I-35131 Padova, Italy.

§Also at: Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo 113-0032, Japan.

||Present address: INFN Padova, I-35131 Padova, Italy.

I. INTRODUCTION

High-energy cosmic rays interacting with Earth’s atmosphere generate extensive air showers. These showers comprise numerous charged particles that emit a coherent radio pulse, primarily due to the deflection of electrons and positrons by the Earth’s magnetic field [1,2]. Detecting this impulsive radio emission has proven valuable for reconstructing air-shower properties, such as electromagnetic energy and depth of the shower maximum, X_{\max} [3–7].

Complementary measurements of air-shower events using radio and other detection techniques, particularly particle detectors sensitive to the muonic component, further enhance sensitivity to the mass of the primary particle [8–11]. The radio technique is thus expected to play an important role in next-generation experiments targeting the particle and astrophysics of ultra-high-energy cosmic rays [12].

A primary challenge of using radio signals for air-shower measurements is continuous background noise from natural (Galactic and thermal) and anthropogenic sources. The system noise is often small compared to the external background, e.g., for the SKA log-periodic antennas (SKALA) v2 and their low-noise amplifiers used here, the internal noise is approximately 40K [13,14], only. This noise contaminates radio measurements, obscuring air-shower pulses and limiting the energy threshold for cosmic-ray studies.

Several techniques have been employed by radio experiments to mitigate the impact of background noise on air-shower measurements. One approach is digital radio interferometry [3,15,16], which combines measurements from multiple antennas to improve the signal-to-noise ratio (SNR). Another method involves selecting an optimal frequency band [17]: while air-shower radio emissions are broadband (from MHz to GHz), Galactic radio background dominates below a few 100 MHz. Thus, selecting an optimal frequency range can significantly enhance the SNR. Moreover, narrow-band radio frequency interferences, which are a major component of anthropogenic background, can be effectively filtered using notch filters and similar techniques, because the air-shower radio signals are broad-band. More recently, machine learning tools (see [18] for a living review) have been introduced to further reduce background noise. For example, convolutional neural networks (CNNs) have shown promise in denoising contaminated radio traces when trained on simulated [19] and measured background [20–22]. Building on these advancements, this work presents the development of two independent CNNs: a *classifier* for identifying traces containing radio pulses and a *denoiser* for recovering underlying air-shower pulses from noisy traces [23].

The performance of these CNNs is demonstrated by application to radio measurements of air showers from IceCube’s surface array at the South Pole. The surface component of IceCube, known as IceTop, consists of 81 stations, each with pairs of ice-Cherenkov tanks, spread over an area of approximately 1 km². IceTop is designed to detect

cosmic-ray air showers in the energy range from below a PeV to a few EeV [24]. However, snow accumulation on IceTop tanks has increased the detection threshold and introduced systematic uncertainties in air-shower measurements over time [25]. To improve IceTop’s detection threshold and expand its scientific capabilities, such as precise energy calibration and improved composition studies, a surface array enhancement of IceTop has been planned [26–28]. This enhancement involves deploying elevated scintillation panels and radio antennas across the existing IceTop array.

While a low false-positive rate is critical for self-triggered radio experiments, externally triggered radio setups, such as the IceTop enhancement, focus primarily on improving the measurement quality of air showers. Every triggered event is an air shower, but in many the radio signal is too weak and buried in the noise. The goal is thus to measure that radio signal in as many air-shower events as possible and to reduce the measurement uncertainty on important parameters of the radio pulse in each antenna, in particular, its arrival time and strength.

A prototype station of the IceTop enhancement consisting of eight panels and three SKALA v2 antennas was deployed in 2020 and has regularly been measuring air showers with both the scintillation detectors and the radio antennas [29] (cf. Fig. 1). Since its deployment, the prototype station’s firmware and hardware have undergone several upgrades, including an update to the local data acquisition in January 2022 [30,31]. The radio measurements are broadband (70–350 MHz) with two polarization channels recorded per antenna. We collect two types of data: air-showerdata, triggered by scintillation panels, and

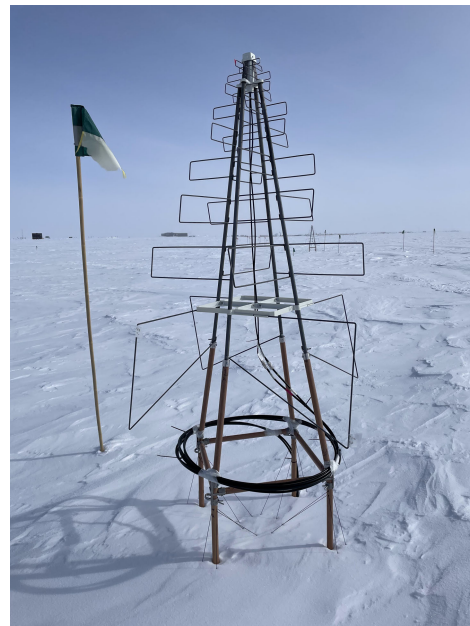


FIG. 1. One of the three SKALA antennas of the prototype surface station at the South Pole (photo from December 2022, courtesy of Roxanne Turcotte [30]).

fixed-rate-triggered (FRT) data, collected by a software trigger at regular intervals and providing a measurement of the radio background. For this study, FRT radio data are used to train and test the CNNs, which are then applied to air-shower data to experimentally demonstrate their performance.

This paper details the architecture, training, testing, and implementation of the networks, and is organized as follows: Sec. II provides the basic architecture of the CNNs. Section III covers data preparation for network training and validation using simulated radio pulses and measured background, followed by Sec. IV on the performance of the networks evaluated on an independent test dataset. Section V describes applying the networks to measured radio data from the prototype station to identify air-shower events. Section VI, compares CNN results with a standard SNR method for identifying air-shower events. Finally, Sec. VII provides a summary and outlook.

II. NETWORK ARCHITECTURE

We developed two networks: a *classifier*, to distinguish signal traces from noise-only traces, and a *denoiser*, to remove noise from contaminated traces. Both networks take the two-polarization-channel waveforms from an antenna as input and are primarily built from 1D convolutional layers. Their basic architecture is depicted in Fig. 2.

Each network's input is a data array of shape 1000×2 , representing 1000 time samples per trace across 2 polarization channels. For both networks, the *input* layer is followed by an encoder composed of several encoding

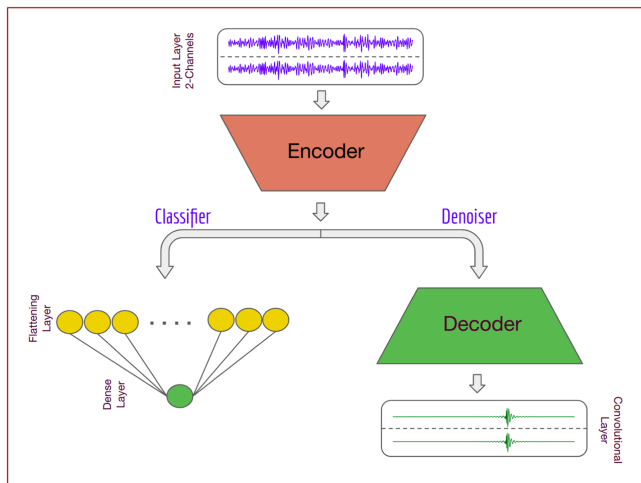


FIG. 2. Schematic of the neural network architectures for the *classifier* and *denoiser*. Both networks take two-channel input waveforms and begin with an encoder composed of 1D convolutional layers paired with max-pooling layers. In the *classifier*, the encoder is followed by a flattening layer and a dense output layer. In contrast, the *denoiser* includes a decoder composed of up-sampling and convolutional layers, followed by a final convolutional output layer.

blocks, each consisting of a 1D convolutional layer and a max-pooling layer with a fixed pooling size of 2 [32].

For the *classifier*, extensive hyperparameter testing led to a final architecture with four encoding layers. Each convolutional layer uses 8 filters with a kernel size of 33 bins (33 ns), followed by a flattening layer and a dense output layer with a single neuron. ReLU activation functions are applied in all layers except the output layer, which uses a sigmoid activation to map the output to a value between 0 (backgroundlike) and 1 (signallike). This configuration results in a total of 7,393 trainable parameters.

The *denoiser* adopts an autoencoder architecture, where the input is first compressed by the encoder (of the same architecture as for the classifier, but trained separately) and then reconstructed to its original dimensions by the decoder. The decoder mirrors the encoder in structure, using the same number of decoding blocks to ensure output dimensionality matches the input. Each decoding block includes a 1D convolutional layer followed by an up-sampling layer with a factor of 2, doubling the output size at each step. Each convolutional layer contains 64 filters with a kernel size of 33. A final convolutional layer with one filter serves as the output of the denoiser. ReLU activation is applied to all layers except the final one, which uses a linear activation. The output retains the same shape as the input radio traces, with 1000 samples per polarization channel, with noise ideally removed from both channels. This architecture has a total of 684,674 trainable parameters.

The networks were implemented and trained using Keras [33] and TensorFlow [34]. Binary cross-entropy (BCE) served as the loss function for the *classifier*, while mean squared error (MSE) was used for the *denoiser*. Training was monitored using an early stopping criterion based on the validation loss, with a patience of 10 epochs. If no improvement in validation loss occurred for 10 consecutive epochs, training was halted, and the model weights corresponding to the lowest validation loss were retained (The network architecture and associated processing scripts are available at [35]).

III. DATA PREPARATION

The neural networks were trained using simulated radio signals and measured background data. This choice was made because the CoREAS simulations have been shown to accurately describe measured radio pulses from air showers, meaning that no significant discrepancies with experimental data are known [36–38]. However, modeled background often oversimplifies the reality of radio experiments, which is why we used radio background recorded with the exact setup at the South Pole.

CORSIKA [39] v7.7401, with the CoREAS extension [40], was used to generate air-shower simulations and their radio signals. The simulations utilized the average April South Pole atmosphere, which closely approximates the

yearly average, with the `Fluka` model [41] for low-energy hadronic interactions and `Sibyll 2.3d` [42] for high-energy interactions. Protons and iron nuclei were used as primary particles, with energies ranging from $10^{16.5}$ to $10^{18.0}$ eV, zenith angles (θ) from 0 to 0.8 in $\sin^2 \theta$ (0° to 63.4°), and uniform azimuth angles (ϕ). More than 40,000 distinct CoREAS simulations were generated, wherein radio emissions were computed on a spatial grid configured as an eight-legged star. The center of this star was positioned at the intersection of the shower axis with the ground [43].

For the radio background, we used FRT event waveforms recorded during the first half of 2022, a period of stable configuration of the updated local data acquisition. Each FRT event contains waveforms of 1024 data points per polarization channel of each of the three antennas of the prototype station, of which the inner 1000 data points (corresponding to $1 \mu\text{s}$ of radio background) were used in this study. The total data collection period analyzed in this work amounts to approximately 106 days of effective runtime, during which over 300,000 background events were recorded and used for data preparation.

To prepare the training data, CoREAS radio signals are combined with FRT waveforms and processed using custom software developed within the IceCube framework for radio air-shower analysis [43]. The CoREAS radio pulses are transformed into mock data which include all known properties of the experimental setup. As a first step, the CoREAS pulses are zero-padded in the time domain to make all waveforms equal in length. These zero-padded signals are then convolved with the SKALA v2 antenna response. Next, the convolved signals are resampled from the native 0.2 ns to the 1 ns sampling rate used for data taking in the prototype station. The resampled signals are further convolved with the full electronics response, based on calibration measurements, including components such as a low-noise amplifier and an analog radio board. They are then digitized at 14 bits with a 1 V peak-to-peak dynamic range (discrete amplitude bins of 2^{-14} V). Measured background waveforms in the same units are added to the digitized signals, and the combined traces are converted back to voltages. Next, pedestal offsets, corresponding to unphysical DC components, are removed from the waveforms, and a box frequency filter is applied to the nominal measurement band of the prototype station of 70–350 MHz [14]. Finally, the electronics response is deconvolved, and the resulting waveforms are saved as the noisy dataset.

The pure simulated signals (before noise addition) are also filtered to 70–350 MHz and used as target labels for the *denoiser*. Background waveforms, processed in the same way, are saved for training the *classifier*. Each waveform is individually normalized to the range $[-1, 1]$ prior to being passed to the networks to ensure consistent scaling of input features across network layers.” To prevent any positional bias, each simulated radio pulse is randomly

time-shifted within the $1 \mu\text{s}$ (1000-bin) waveform before being added to the background.

Following this procedure, approximately 165,000 samples of each type (noisy, pure signal, and background noise) were generated per antenna. Each sample consists of two time series, corresponding to the two antenna polarization channels. Thus, for training and validation of the *denoiser*, approximately 165,000 noisy samples paired with their corresponding pure signal were used. For the *classifier*, about 165,000 noisy samples and an equal number of background samples were utilized. In total, for each antenna, the complete dataset consists of approximately 495,000 samples.

This processed dataset was split into two subsets: 80% for training and 20% for validation. After finalizing the networks, an independent dataset was generated following the same procedure to test the network’s performance. This included approximately 3,000 new CORSIKA simulations, combined with measured background from the same period. The resulting test set contains approximately 24,000 samples of each type.

A key quantity to characterize a noisy radio trace is the SNR. Since definitions vary across experiments, we adopt a version previously used in other experiments, but made it more robust against occasional outliers in the traces by using the median noise level over several time intervals instead of a single time interval [44,45].

The SNR of a voltage trace with amplitudes Y_i in each bin i is calculated as follows:

$$\text{SNR} = \left(\frac{\max[|Y_i|]}{\text{median}[\text{RMS}(n_j)]} \right)^2, \quad (1)$$

where the numerator is the maximum magnitude of the waveform. The denominator uses the root-mean-square (RMS) as an estimator for the noise power in consecutive, nonoverlapping time intervals n_j of equal length. The median is chosen to avoid the impact of occasional outliers, such as human-made or natural radio pulses, as well as the air-shower pulse itself, on the estimated noise power. Although counterintuitive, it is common in air-shower radio detection that pure background traces yield average SNR values well above 1. In our case, only SNR values $\gg 10$ correspond to radio pulses that are clearly distinguishable from noise by eye. Figure 3 shows the SNR distribution for both polarizations of antenna 1.

IV. NETWORK PERFORMANCE

This section describes the performance evaluation of the networks using the test dataset. For the *classifier* evaluation is straightforward, based on correct and false classifications. For the *denoiser* physics motivated metrics are used, specifically focusing on the accurate reconstruction of the radio pulse’s peak time and power, as these are critical parameters for air-shower measurements.

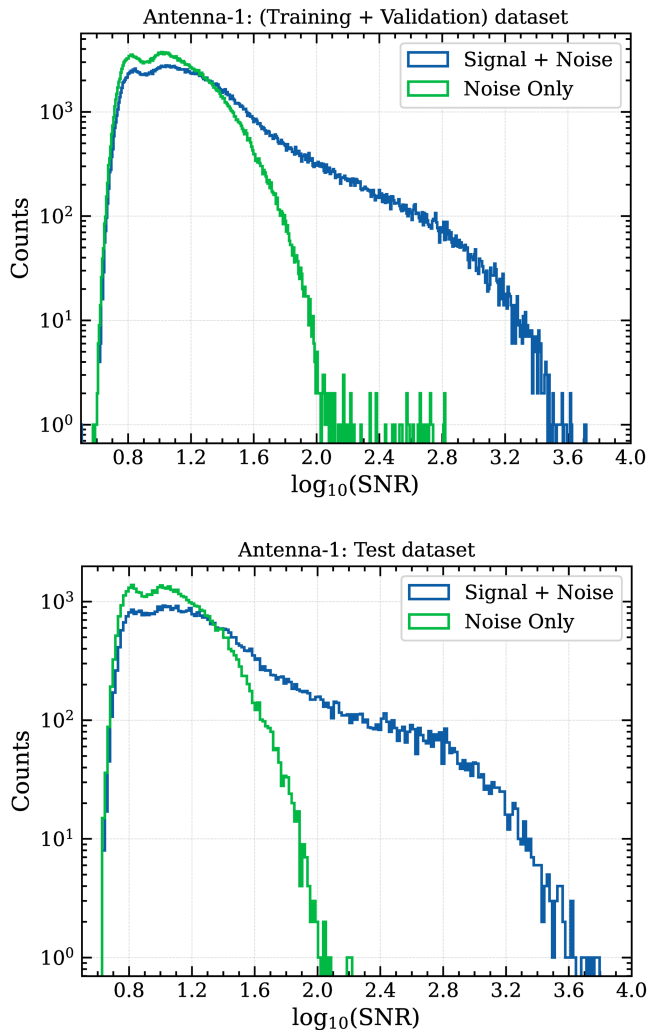


FIG. 3. Distribution of signal-to-noise ratio of the data set used for training and validation (top) and test (bottom). Shown here is only the distribution for the dataset produced for antenna 1, distributions for other antennas exhibit similar structures, differing only in their mean and slight variations in the tails. The mean background SNR is approximately 14 for both datasets.

A. Classifier

The *classifier* is trained on a mix of noisy (signal + noise) and noise-only data, with the integer values 1 and 0 used as labels, respectively. Each input sample, consisting of two polarization waveforms, produces a single output value between 0 and 1. The left panel of Fig. 4 shows the distribution of classifier outputs for the test set. While the network effectively pushes signal and background events toward outputs of 1 and 0, respectively, overlap remains, especially for scores below 0.5, where classification becomes more challenging.

To illustrate the performance of the *classifier*, a threshold of 0.5 is applied, classifying samples with outputs ≥ 0.5 as signallike and others as backgroundlike. This results in an overall false positive rate (FPR) of 2.2%, i.e., background traces misclassified as air-shower pulses.

In contrast to many other applications of classifiers, the concept of a true positive rate (TPR) is not meaningful for externally triggered radio enhancements. All air-shower events are known to contain some radio signal (as by definition also all events in our noisy dataset), but often this radio signal is negligible and too weak to be detected. This depends on many parameters, such as the energy of the air shower and its geometry relative to the antenna. In a naive approach counting all events, the TPR would thus drastically depend on how many air showers with undetectable weak signal would be included in the sample, and any definition of a detectability threshold would be arbitrary.

We have thus divided the test dataset into SNR bins, and determined the TPR and FPR for each bin. Rates are reported as percentages, with the TPR defined as the number of correctly identified signals divided by the total number of noisy waveforms. The right panel of Fig. 4 shows TPR and FPR as a function of SNR. A TPR above 80% was achieved for $\text{SNR} > 40$, while the FPR remained below 5% across all SNR values.

In practice, the particular situation of the experiment can guide how the *classifier* is used. For example, in our case, the IceTop array at the same location allows an additional cross-check, enabling us to tolerate a few percent FPR at the *classifier* stage. Moreover, to identify radio events from air showers, both the *classifier* score and the amplitude of the denoised pulse are considered in a combined selection criterion (see Sec. V for more details).

B. Denoiser

Following training, the *denoiser* was also evaluated using the test dataset. Noisy traces identified as signallike (i.e., with classifier scores ≥ 0.5) were passed through the *denoiser* for cleaning. An example of a successful denoising result is shown in Fig. 5, where the radio pulse shape and timing are recovered in both channels and the background is strongly suppressed inside of the design band of 70–350 MHz (differences outside of the design band do not matter for air-shower analyses, as these usually filter to a subband fully contained in the design band).

To quantify the *denoiser*'s performance with quantities relevant for physics analyses, two accuracy metrics were devised: power ratio and peak time difference. These metrics compare raw (nondenised) and denoised waveforms.

The power ratio is calculated by measuring the power within a 50 ns signal window (P_S) centered around the simulated pulse position, and a 400 ns nonoverlapping noise window (P_N). The power ratio is then given by:

$$\text{Power Ratio} = \frac{[P_S - P_N]_{\text{measured}}}{[P_S - P_N]_{\text{true}}}. \quad (2)$$

In the ideal case, the power ratio is 1, indicating that the calculated power matches the true radio pulse power. Negative values can occur for small signals when the

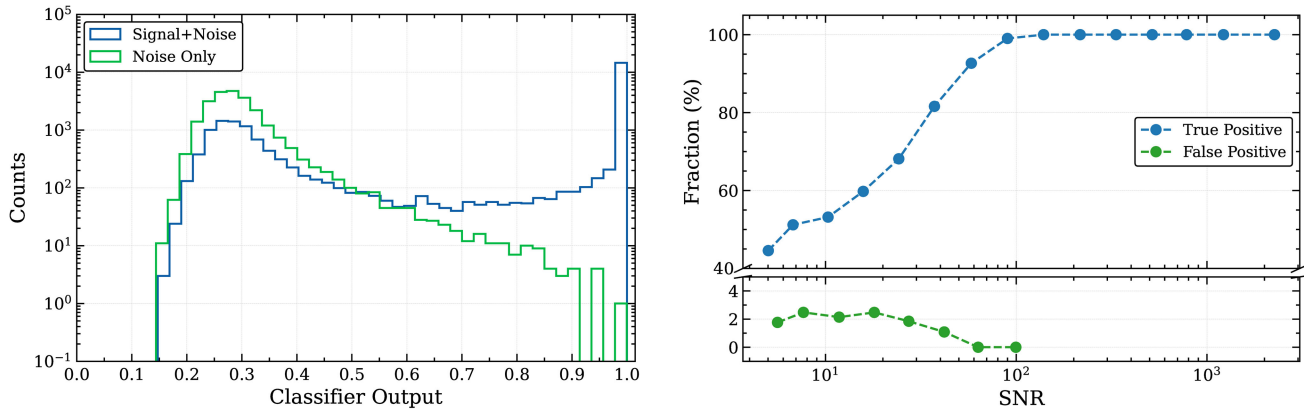


FIG. 4. Left: classifier output obtained for both the noisy waveform data (blue) and for the noise-only waveform data (green). Right: true positive and false positive rate as a function of SNR, obtained after applying a threshold cut of 0.5 on the classifier output. Note that this threshold is used only to illustrate performance on the test set. The y-axis is broken to emphasize low false-positive fractions. The false-positive curve stops at SNR = 100, as almost none of the noise-only waveforms in the data sample exceeds that value (cf. Fig. 3).

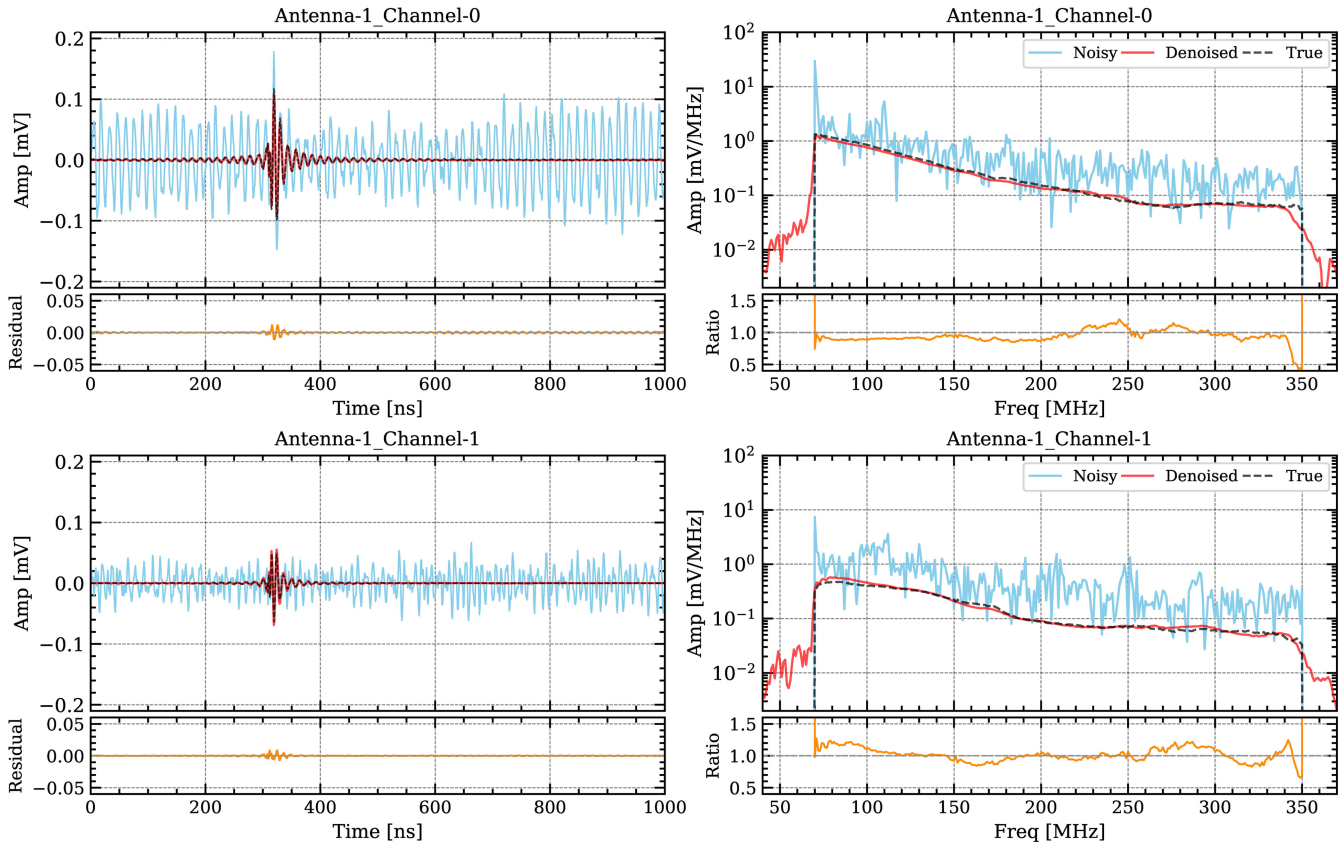


FIG. 5. Example of a denoised waveform, together with the corresponding noisy input and true waveform (CoREAS simulation including the antenna response), shown for the two polarization channels of the antenna. The frequency spectra of the waveforms are displayed in the right panels. The lower subpanels present the difference between the denoised and true waveforms in the time domain, and the ratio between the denoised and true spectra in the frequency domain.

power measured in the background window exceeds the power in the signal window.

The peak time difference, Δt is computed by locating the peak of the waveform’s Hilbert envelope, i.e., the

magnitude of its analytic signal that represents the instantaneous amplitude, and subtracting the true pulse time

$$\Delta t = T_{\text{measured}} - T_{\text{true}}, \quad (3)$$

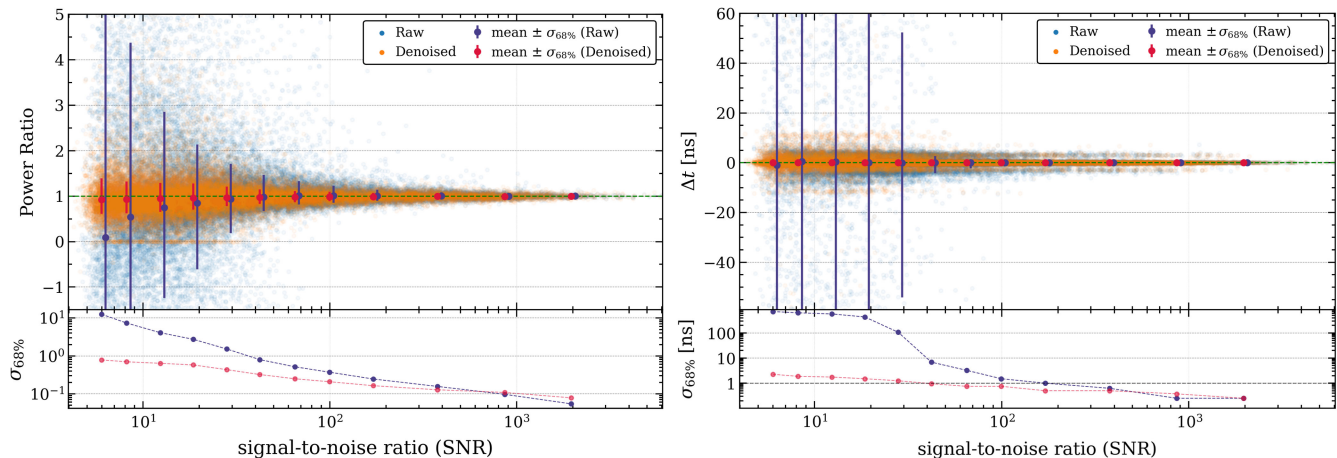


FIG. 6. Left: power ratio as a function of SNR for both waveforms without applying denoiser (blue distribution) and after applying the denoiser (orange distribution). Right: peak time difference over SNR. The navy and red dots are the mean values for (not-)denoised waveforms with the error bars representing 68% containment, which is also shown in the bottom plot for better visibility.

where T_{measured} is the peak time (from either the raw or denoised traces) and T_{true} is the simulated pulse time. Ideally, $\Delta t = 0$, meaning the estimated time matches the true radio pulse time within the trace.

The power ratio as a function of SNR is presented in the left panel of Fig. 6. At low to intermediate SNR values (≤ 100), the *denoiser* reduces both bias (deviation from an ideal power ratio of 1) and spread, improving the estimation of the pulse power. Near-zero power ratios in the denoised waveforms arise when the *denoiser* fails to recognize a pulse and mistakenly removes both pulse and background. At high SNR, both raw and denoised traces yield similarly accurate results.

The right panel of Fig. 6 shows the peak time difference as a function of SNR. Similar to the power ratio, the *denoiser* notably improves time accuracy at low and intermediate SNR. At high SNR, there is only a slight difference between denoised and nondenoised cases, with typical time offsets ≤ 1 ns. The large scatter in raw values is due to background fluctuations being misidentified as the pulse, since the highest peak across the entire trace is taken as the pulse peak time. The horizontal banded pattern seen in the distribution is caused by the algorithm sometimes detecting different oscillations of the waveform for peak time, effectively selecting one oscillation peak over another [44].

In summary, the *denoiser* significantly enhances the accuracy of pulse power and peak time at low to intermediate SNR, while at high SNR the impact of background is anyway low, and the accuracy is fully sufficient in either case.

V. APPLICATION OF NETWORKS TO MEASURED AIR SHOWERS

After the testing phase, the networks were applied to air-shower events simultaneously measured by the three surface antennas of the prototype station, triggered by the scintillation panels. Figure 7 presents an example event: six

waveforms corresponding to the two polarization channels of the three antennas. The blue curves show the measured, preprocessed waveforms (see Sec. III), which serve as inputs to the networks. The red curves depict the output from the *denoiser* and *classifier* scores (CI_score) for each antenna are shown in the text box. In this example, all three antennas received the maximum CI_score, and the *denoiser* successfully suppressed background noise, even in channels with significant interferences. Each antenna recorded a strong signal in at least one polarization channel, while signals in the other channels were weak or comparable to the background. Additionally, in channel 0 of antenna 3, an artifact taller than the true radio signal appears around 250 ns; the network correctly identified and removed this feature, which standard methods looking for the highest peak in the trace would have misidentified as a signal. The resulting directional reconstruction for this event differs by only 2.6° from the IceTop reconstruction of the same air shower, as expected when the network correctly identified the air-shower radio signal (see next section).

However, such optimal results are not always guaranteed. Most scintillator-triggered events have radio signals below the detection threshold. To filter out poorly denoised events without a clear radio signal, a set of quality cuts was developed. The first cut uses the output of both the *classifier* and *denoiser* to retain waveforms likely containing an air-shower radio pulse.

The cut values were determined by applying the networks to background waveforms from FRT events as shown in the top panel of Fig. 8. The plot shows the distribution of classifier scores (y-axis) versus the absolute peak amplitude of the denoised waveform (x-axis). The threshold is defined by a line $y = mx + c$, shown as a red dashed line, with the slope ($m = -20$) and intercept ($c = 0.6$) tuned to achieve 95% rejection of pure background waveforms per antenna, comparable to reference methods based on a threshold in

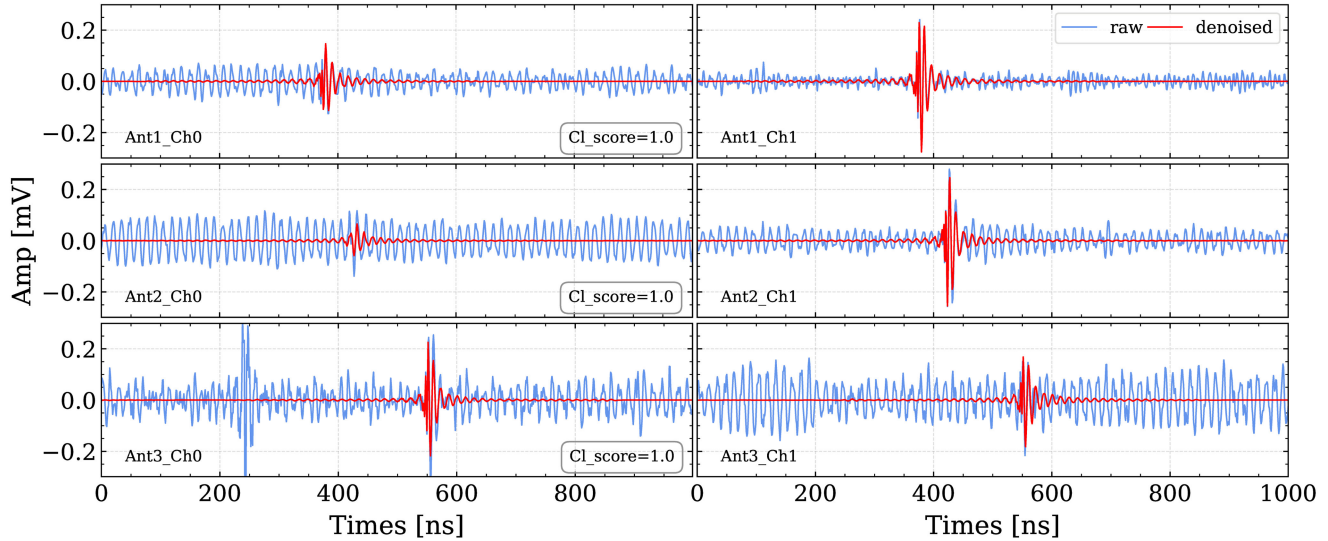


FIG. 7. Measured waveforms for all six channels from the three antennas for an air-shower event triggered by the scintillation panels. The raw waveforms, utilized as input for the networks, are depicted in blue. The cleaned waveforms, obtained through the denoising process, are presented in red. The output of the classifier, represented as the Cl_score , is displayed as text for each of the three antennas. In this particular example, all three antennas yield a Cl_score of 1.0, representing an ideal classification that is not always achieved.

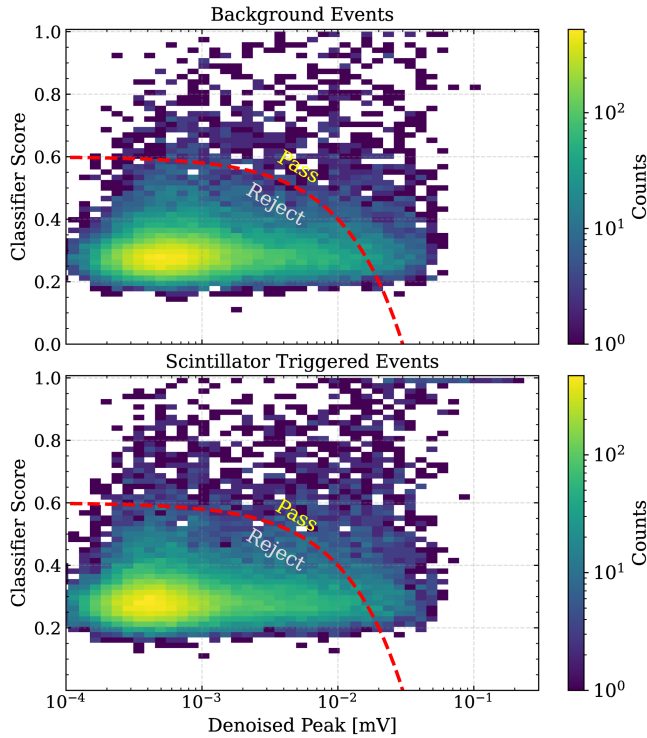


FIG. 8. Top: Distribution of the *classifier* output (y-axis) versus the maximum amplitude of the denoised waveform (x-axis) for pure background events recorded with the fixed-rate trigger (FRT). The red dashed line indicates the quality cut used to reject 95% of background waveforms; only waveforms above this line are considered as passing. Bottom: Same distribution for scintillator-triggered air-shower events. A distinct population in the upper right region indicates clear air-shower signals.

SNR [46]. By this choice of a cut combining the classifier output and the pulse amplitude, we identify air-shower radio pulses of low amplitude without the risk of losing air-shower events with high-amplitude radio pulses (using a flat cut of 0.6 in *classifier* score would result in an approximately 3-fold increase in detected air-shower events over the reference SNR method, while the combined cut results in a 5-fold increase, as shown later). For events surviving this cut across all three antennas, a subsequent cut removes those whose reconstructed radio arrival direction is inconsistent with IceTop.

The bottom panel of Fig. 8 displays the same network outputs, but for scintillator-triggered events. As expected, most outputs cluster in the lower-left corner of the plot, indicating that the networks categorize them as backgroundlike. This is consistent with expectations, as the scintillator panels have lower detection thresholds and trigger on lower-energy air showers with little or no detectable radio signal. Nevertheless, a small cluster of events in the top-right of the plot indicates confident detections of air-shower radio signals.

The distribution of the *classifier* scores and denoised peak amplitudes illustrates the flexibility of the selection strategy: the decision boundary can be adjusted depending on whether a particular analysis prioritizes purity (requiring higher scores/amplitudes) or efficiency (accepting lower values). There generally is no need for perfect classification, as the IceTop measurement of the same air showers is always available for cross-check. The choice used here is motivated by achieving 95% background rejection per antenna, to compare the CNNs with the reference SNR method.

A. Reconstructed air-shower events and their angular distributions

For the successful reconstruction of the air-shower arrival direction, at least one channel from each of the three antennas must pass the quality cuts described in the previous section. When this condition is met, the denoised waveforms are used for direction reconstruction. From each antenna, the channel with the largest peak amplitude is selected to extract the pulse arrival time. Because the prototype station only features three antennas, the arrival direction is reconstructed assuming a simple plane wavefront.

During a 106-day search period, a total of 227,613 scintillator-triggered events were recorded and analyzed using the CNN-based pipeline. Among these, 608 events passed the CNN quality cuts and were considered for arrival direction reconstruction. The reconstructed directions from radio were then compared to those obtained from IceTop. Fig. 9 shows the distribution of opening angles ($\Delta\omega$) between the two reconstructions. As seen in the zoomed-in view, most events cluster within a few degrees of the IceTop direction, indicating accurate identification of the radio pulse in all three antennas.

A small subset of events, however, shows large opening angles, forming a tail in the distribution. These events typically result from incorrect identification of the radio pulse in at least one antenna, for example, when an artifact or background fluctuation is mistaken for a true signal.

To exclude such cases, an additional data-driven quality cut is applied. Events with $\Delta\omega \geq 7^\circ$ are removed, where the distribution shows a significant drop. As the main purpose of the radio enhancement is to increase the measurement accuracy of IceTop for cosmic-ray analyses, this IceTop direction reconstruction is available for all air-shower

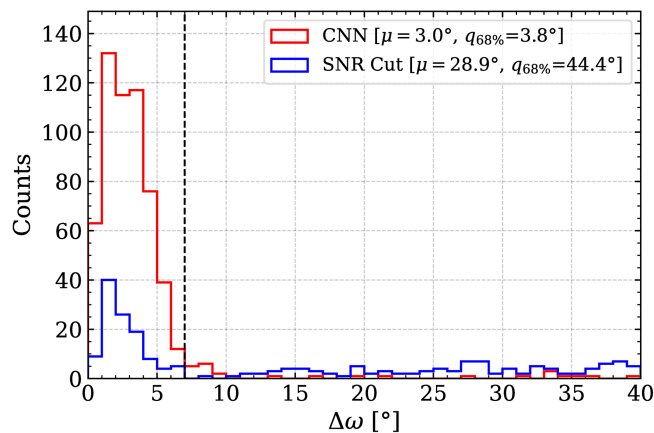


FIG. 9. Opening angle between the radio-reconstructed and IceTop-reconstructed directions for scintillator-triggered air-shower events identified using the CNN method and an SNR method (see Sec. VI). The dashed line at 7° indicates the quality cut applied to select well-reconstructed air-shower events. The mean values of the open angle for events within this quality cut are 3.5° for the CNN and 3.0° for the SNR method.

events of interest. The choice of $\Delta\omega$ threshold can be adjusted based on specific science analysis requirements for purity and efficiency. This cut at 7° removes about 9% of the events found with the CNNs, leaving 554 candidate radio events.

Since the strength of the radio emission depends on the geomagnetic Lorentz force, we expect a suppression of events aligned with the geomagnetic field and an enhancement at larger geomagnetic angles (the angle between the shower axis and the local geomagnetic field). This expectation is reflected in the zenith distribution (Fig. 10, upper panel), which shows a concentration of events at intermediate zenith angles. The decline above 60° is due to the reduced efficiency of the scintillator trigger at large angles.

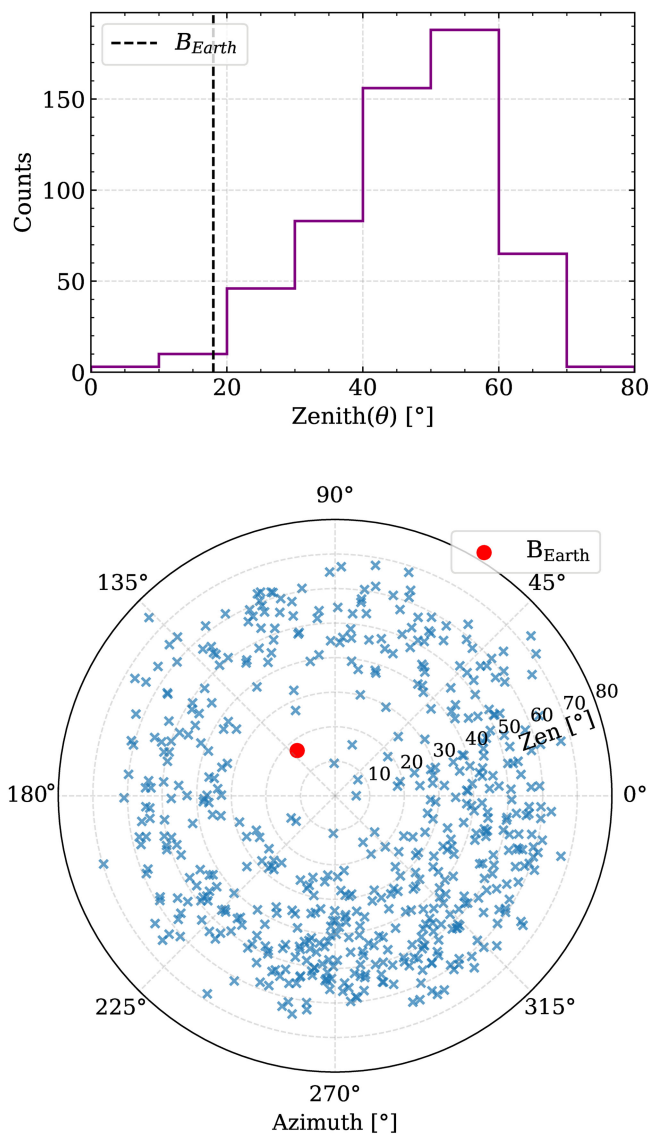


FIG. 10. Zenith angles (top) and Arrival directions (bottom) of the candidate radio events. The black dashed line in the upper plot and the red-filled circle in the lower plot represent the local geomagnetic field at the IceTop location.

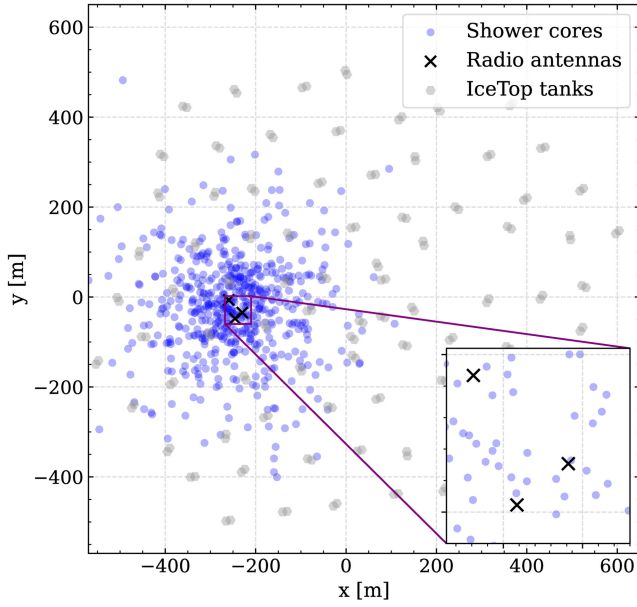


FIG. 11. IceTop core distribution of the air-shower events detected with the radio antennas. Shower core positions are marked with blue circles, antenna locations are indicated by black X's, and IceTop tanks are represented by gray circles.

Additionally, we require coincidence with IceTop, and the IceTop array also shows decreasing sensitivity at high zenith angles [47]. Again, consistent with the expectation of geomagnetic radio emission, the azimuth and polar distributions also show an excess of events in the direction opposite to the geomagnetic field (Fig. 10, lower panel).

B. Core distribution

The shower core positions, defined as the impact point of the shower axis on the ground, were reconstructed using IceTop data and are shown in Fig. 11, along with the locations of the detectors. The figure demonstrates that most candidate events have core positions close to the prototype station. However, several events were detected with cores located more than 200m away from the antennas. These correspond to inclined showers, which produce elongated radio footprints due to their geometry, allowing the radio signal to reach the array from a greater distance.

Figure 12 shows the distance between the prototype station center and the reconstructed shower axis, plotted as a function of the zenith angle (top panel) and the geomagnetic angle (bottom panel). The top panel illustrates that events farther from the array have larger zenith angles, some even exceeding 60° , consistent with expectations for inclined showers that produce larger radio footprints.

The bottom panel further shows that these events also have larger geomagnetic angles, which indicates stronger geomagnetic radio emissions. Overall, these trends provide strong evidence that the CNNs correctly identified radio signals from cosmic-ray air showers.

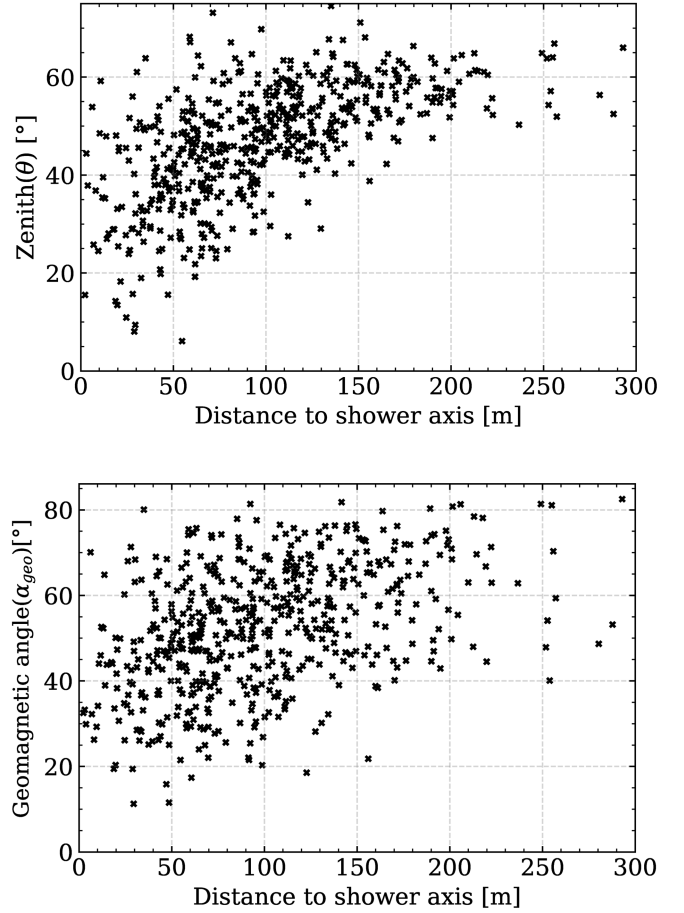


FIG. 12. The distance of the shower core to the center of the prototype station is calculated in the shower plane and plotted over the zenith angle (top) and geomagnetic angle (bottom). The events that are far from the station are also the ones with larger zenith and geomagnetic angles, hence a large radio emission from these events is expected.

VI. COMPARISON WITH AN SNR METHOD

To realistically assess the CNN performance on experimental data, we compare its results with those obtained from a standard SNR method that combines narrow-band radio frequency interference (RFI) suppression via digital filtering and a SNR threshold cut. Variants of this standard method have been used by several radio arrays for air-shower detection (see, e.g., [5,6]), and have also been applied to data from the prototype station at IceTop [29].

In this SNR method, a bandpass filter is applied to limit the signal to the 100–230 MHz range, which has been optimized using CoREAS simulations to enhance the SNR for typical air-shower signals and is consistent with earlier simulation studies [17]. To further suppress narrow-band RFI within this band, frequency-dependent weights are applied [29]. Note that, for the CNN method, the full band of 70–350 MHz is used and does not require explicit RFI suppression. After filtering, a SNR threshold is applied,

chosen to reject 95% of the background waveforms (FRT events) in each channel.

All scintillator-triggered events within the search period are considered for which at least one channel from each of the three antennas exceeds the SNR threshold. These selected events are then reconstructed using a plane-wave front model, as described earlier for the CNN method.

Using this SNR method, 341 events were successfully reconstructed. The opening angle between their radio-reconstructed directions and the corresponding IceTop directions is shown in Fig. 9 (blue). Similar to the CNN-based results, many events are reconstructed within a few degrees of the IceTop direction, however, the distribution features a more pronounced tail at large opening angles. Applying the same quality cut of $\Delta\omega \leq 7^\circ$, yields 111 candidate events, approximately one-fifth the number identified using the CNN method. Despite the much larger number of CNN events passing the quality cut, which are mostly at lower energy and therefore at lower SNR than the events identified by the SNR method, the mean opening angle for those events deteriorates only slightly from 3.0° to 3.5° .

Figure 13 (top-left) compares the $\log_{10}(S_{125}/\text{VEM})$ distributions for candidate events identified by both methods, along with all scintillator-triggered IceTop events. The $\log_{10}(S_{125}/\text{VEM})$ parameter (in units of vertical equivalent muons (VEM)) is used as an energy proxy for IceTop, however, the energy reconstruction is only optimized and calibrated for small zenith angles ($\theta < 30^\circ$) [24,25]. Most high-energy events (on the right-hand side of the $\log_{10}(S_{125}/\text{VEM})$ distribution) were identified by both methods; however, the CNN method recovers significantly more lower-energy events, which are missed by the SNR method. Only 8 of the 111 events found by the SNR method were missed by the CNN method. These 8 events are all at low energy $\log_{10}(S_{125}/\text{VEM}) < 1$, and may therefore be false-positive detections by the SNR method. This would be consistent with their nonidentification by the CNN method. As discussed in the next section, the false-positive rate is indeed expected much larger for the SNR method than for the CNN method.

The zenith angle (top right), azimuth angle (bottom left), and geomagnetic angle (bottom right) distributions in Fig. 13 further demonstrate that the increase in the number

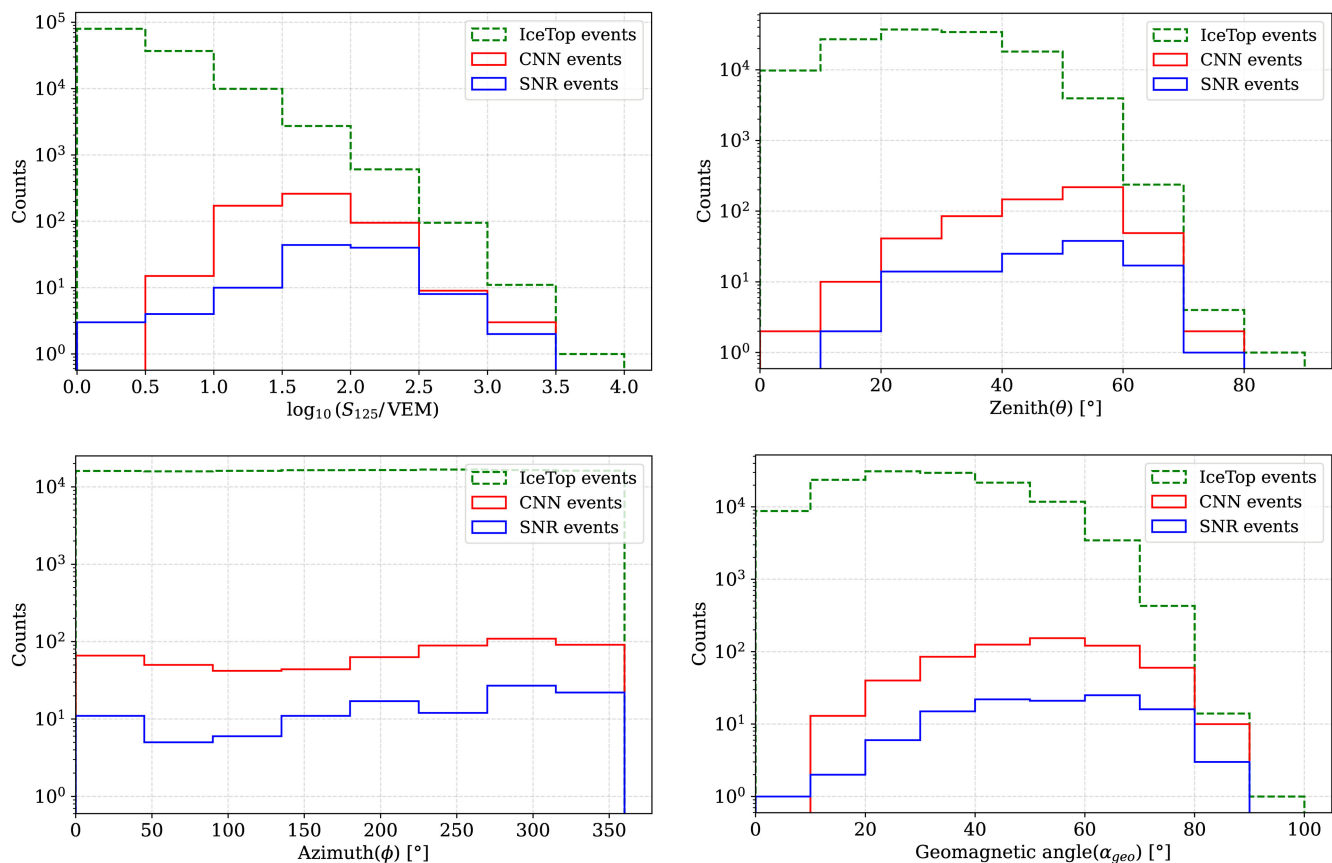


FIG. 13. Distributions of air-shower events identified with the CNN (red) and SNR (blue) methods. The top-left panel shows the distribution of $\log_{10}(S_{125}/\text{VEM})$, i.e., the IceTop signal strength at 125m from the shower axis in units of vertical equivalent muons (VEM). The other panels show the distributions of zenith angle (top right), azimuth angle (bottom left), and geomagnetic angle (bottom right). For reference, all IceTop events with a scintillator trigger are shown as a green dashed line.

TABLE I. Summary of false positive estimates for the SNR and CNN methods. Of all FRT events, 321 passed the SNR selection and were reconstructed, compared to only 8 for the CNN method. After the $\Delta\omega$ cut, the expected number of false positives is 3.7 for SNR and 0.08 for CNN. These values are obtained by multiplying the final false-positive fractions (last row) with the total number of scintillator-triggered events during the search period (227,613 events).

	Total events	Passing Fraction SNR cuts	Passing Fraction CNN method
Fixed-Rate Triggered (FRT)	336070	9.6×10^{-4} (321 events)	2.4×10^{-5} (8 events)
Mean fraction of IceTop events within 7°		0.017	0.014
Resulting false-positive fraction		1.6×10^{-5}	3.3×10^{-7}

of events identified by the CNN method is consistent across all angular ranges.

A. Estimation of the rate of false positive events

Even in an externally triggered radio setup, like the IceTop enhancement, false positive detections of radio pulses may negatively affect physics analyses. Because of the IceTop information available for each air shower, the tolerable false positive rate (FPR) for the additional radio measurement of the same air showers will be higher than in self-triggered radio instruments, typically of at least per mill level, to be determined for each cosmic-ray analysis individually.

The vice-versa concept of a true positive rate (TPR) is less useful for our application, as the radio readout is triggered by the scintillation panels. The absolute value of the TPR would thus depend critically on the trigger threshold, which is purposefully chosen such that the radio signal is well below the noise level for most triggered air showers. As shown in the previous section, we correctly detect the air-shower radio signal in about five times as many events than with the SNR approach, which corresponds to a five-time relative increase of the TPR. Here, we focus on determining how the CNNs impact the FPR compared to the SNR method, i.e., for how many of the air-shower events background would be mistaken for the radio signal.

The selection cuts were designed to reject 95% of background waveforms per antenna. Consequently, some background events will pass this cut in all three antennas simultaneously, and a few of those events may coincidentally yield an arrival direction consistent with IceTop. Because the background across multiple antennas is correlated, it is insufficient to estimate the rate of false positive events by simply multiplying individual probabilities. Therefore, to obtain a realistic estimate, we evaluate the FPR for both methods by applying the full selection criteria to FRT events and subsequently reconstructing the arrival direction using the same procedure used for actual air-shower candidates.

During the 106-day search period, a total of 336,070 FRT events were recorded. Of these, 321 events passed the SNR selection in all three antennas and were successfully reconstructed. In comparison, only 8 FRT events passed the

CNN selection and reconstruction pipeline. For an air-shower search, however, only a small fraction of such events would accidentally align with the IceTop direction.

The expected number of false positives in the air-shower search was estimated using the formula

$$FP_1 = \frac{N_{\text{Pass}}}{N_B} \times N_S, \quad (4)$$

where N_B is the total number of FRT events, N_{Pass} is the number of FRT events passing the SNR or CNN quality cuts and successfully reconstructed, and N_S represents the total number of scintillator-triggered events. This yields a first-order estimate of false positives before imposing a directional consistency requirement with IceTop.

To obtain the final false positive rate after applying the $\Delta\omega$ cut, we calculate for each of the N_{Pass} FRT events the fraction f_i^{IT} of IceTop events within 7° of the given FRT event. The total number of expected false positives is then

$$FP_{\text{tot}} = \text{mean}(f_i^{\text{IT}}) \times FP_1. \quad (5)$$

Using this approach, the SNR method yields an expected 3.7 false positives, while the CNN method results in only 0.08 expected false positives. A summary of the FPR calculation is presented in Table I.

This demonstrates that the use of CNNs not only lowers the detection threshold but at the same time significantly decreases the identification of false positive events. Given the total livetime of the analysis, future analyses could adopt less conservative selection thresholds while still maintaining acceptable background rates, further highlighting the advantages of CNN-based selection over traditional methods.

VII. CONCLUSION

In this paper, we presented the development and successful implementation of deep-learning-based convolutional neural networks (CNNs), namely a *classifier* and a *denoiser*, applied to radio measurements of air showers at the IceCube Neutrino Observatory. Both networks were trained on simulated CoREAS radio pulses added to measured background waveforms from a prototype station of IceCube's surface array enhancement.

To experimentally test the performance of the CNNs, the networks were applied to air-shower events recorded over 106 days, triggered by the scintillation panels of the prototype station. The results were compared to a reference SNR method.

For both methods, the reconstructed arrival directions and comparisons with IceTop energy estimators confirm that we detect the radio pulses originating from air showers. This enables a direct comparison between the CNN and SNR methods, not only with simulated data, but also with these real air-shower measurements.

The CNN-based approach offers multiple advantages, particularly at low and intermediate SNR

- (i) The CNNs improve the precision and bias of the reconstructed radio pulse power and arrival time, as confirmed on an independent validation dataset. At high SNR values, where the uncertainties are already low, the CNNs perform comparably to the reference SNR method, which makes them most useful for waveforms with low and intermediate SNRs.
- (ii) The CNNs increased the number of detected air-shower events by nearly a factor of five compared to the SNR method. Comparison with IceTop shows that most of the additionally identified events are of lower energy than those identified with the SNR method.
- (iii) Despite applying the same 95% background rejection per antenna, the CNNs yielded substantially fewer false positives than the SNR method. This indicates that false positives in the SNR method are more likely due to correlated background across antennas, whereas the CNNs are more robust against such correlations.

Several insights emerged during network development that are relevant for future applications in other radio arrays. For training the networks, it was important to use the background recorded by the same antennas and data acquisition system to which the networks would subsequently be applied. Training on the simulated background proved insufficient, likely due to its oversimplified representation of realistic, complex background noise. Moreover, the real background also varied between antennas, e.g., because anthropogenic RFI depends on the orientation and position of the antennas and because the electronics response of the data-acquisition varies slightly from channel to channel. Additionally, leveraging both polarization channels simultaneously improved network performance, as this captures the real correlations in the background of both channels. Multiantenna correlations of both the background and the signal of an event can be considered in future work. This would need a drastic increase in computational resources, as noise and signal show different correlations, adding another dimension to the phase space. Further improvement of the *denoiser* might be possible, e.g., by upsampling of the input

traces [48], which can be achieved with a modest increase of computational resources.

Finally, the selection criteria based on network outputs, can be adjusted for different needs. For example, if a particular physics analysis demands a purer data sample, a higher classifier score can be set to further reduce the false positive rate. This flexibility could be particularly useful for self-triggering radio experiments like GRAND [49] or PUEO [50], which require low FPR. Radio setups that are triggered by other air-shower detectors, like the IceTop enhancement, can typically tolerate a higher FPR and will primarily benefit from the lower detection threshold achieved with the CNNs. As demonstrated with the three prototype antennas of IceCube's surface enhancement, the CNNs enable detection of air showers down to a few tens of PeV. We thus expect that they will play an important role for radio detection with the IceCube-Gen2 surface array [51], as well as for other externally triggered radio experiments, such as at the Pierre Auger Observatory [52–54] or the POEMMA-Balloon with Radio [55].

In conclusion, the neural networks developed in this work are a valuable tool for improving radio-based cosmic-ray detection. These networks can enhance the science reach of existing radio arrays, e.g., by lowering their detection threshold, and will contribute to achieving the science goals envisioned for future experiments.

ACKNOWLEDGMENTS

The IceCube collaboration acknowledges the significant contributions to this manuscript from Abdul Rehman, Alan Coleman, and Frank Schroeder. We thank the Tunka-Rex Collaboration, in particular, Dmitriy Kostunin and Dmitry Shipilov, for making their code available, which became the initial basis for this study. The authors gratefully acknowledge the support from the following agencies and institutions: USA—U.S. National Science Foundation-Office of Polar Programs, U.S. National Science Foundation-Physics Division, U.S. National Science Foundation-EPSCoR, U.S. National Science Foundation-Office of Advanced Cyberinfrastructure, National Aeronautics and Space Administration-EPSCoR, Wisconsin Alumni Research Foundation, University of Delaware high-performance computing resources, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science Grid (OSG), Partnership to Advance Throughput Computing (PATH), Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), Frontera and Ranch computing project at the Texas Advanced Computing Center, U.S. Department of Energy-National Energy Research Scientific Computing Center, Particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, Astroparticle

physics computational facility at Marquette University, NVIDIA Corporation, and Google Cloud Platform; Belgium—Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (Belspo); Germany—Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), European Research Council (ERC), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen; Sweden—Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation; European Union—EGI Advanced Computing

for research; Australia—Australian Research Council; Canada—Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Digital Research Alliance of Canada; Denmark—Villum Fonden, Carlsberg Foundation, and European Commission; New Zealand—Marsden Fund; Japan—Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University; Korea—National Research Foundation of Korea (NRF); Switzerland—Swiss National Science Foundation (SNSF).

DATA AVAILABILITY

The data that support the findings of this article are openly available [35].

-
- [1] T. Huege, Radio detection of cosmic ray air showers in the digital era, *Phys. Rep.* **620**, 1 (2016).
- [2] F. G. Schröder, Radio detection of cosmic-ray air showers and high-energy neutrinos, *Prog. Part. Nucl. Phys.* **93**, 1 (2017).
- [3] W. D. Apel *et al.* (LOPES Collaboration), Final results of the LOPES radio interferometer for cosmic-ray air showers, *Eur. Phys. J. C* **81**, 176 (2021).
- [4] S. Buitink *et al.* (LOFAR Collaboration), Method for high precision reconstruction of air shower X_{\max} using two-dimensional radio intensity profiles, *Phys. Rev. D* **90**, 082003 (2014).
- [5] A. Aab *et al.* (Pierre Auger Collaboration), Measurement of the radiation energy in the radio signal of extensive air showers as a universal estimator of cosmic-ray energy, *Phys. Rev. Lett.* **116**, 241101 (2016).
- [6] P. A. Bezyazeev *et al.* (Tunka-Rex Collaboration), Reconstruction of cosmic ray air showers with Tunka-Rex data using template fitting of radio pulses, *Phys. Rev. D* **97**, 122004 (2018).
- [7] A. Abdul Halim *et al.* (Pierre Auger Collaboration), Demonstrating agreement between radio and fluorescence measurements of the depth of maximum of extensive air showers at the Pierre Auger Observatory, *Phys. Rev. Lett.* **132**, 021001 (2024).
- [8] E. M. Holt, F. G. Schröder, and A. Haungs, Enhancing the cosmic-ray mass sensitivity of air-shower arrays by combining radio and muon detectors, *Eur. Phys. J. C* **79**, 371 (2019).
- [9] B. Flagg, A. Coleman, and F. G. Schröder, Studying the mass sensitivity of air-shower observables using simulated cosmic rays, *Phys. Rev. D* **109**, 042002 (2024).
- [10] S. Andringa, R. Conceicao, and M. Pimenta, Mass composition and cross-section from the shape of cosmic ray shower longitudinal profiles, *Astropart. Phys.* **34**, 360 (2011).
- [11] T. Antoni *et al.* (KASCADE Collaboration), KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems, *Astropart. Phys.* **24**, 1 (2005).
- [12] A. Coleman *et al.*, Ultra high energy cosmic rays the intersection of the cosmic and energy frontiers, *Astropart. Phys.* **149**, 102819 (2023).
- [13] E. de Lera Acedo *et al.*, Evolution of SKALA (SKALA-2), the log-periodic array antenna for the ska-low instrument, in *2015 International Conference on Electromagnetics in Advanced Applications (ICEAA)* (IEEE, Torino, Italy, 2015) pp. 839–843.
- [14] P. Benthem, R. Wayth, E. de Lera Acedo, K. Z. Adami, M. Alderighi, C. Belli, P. Bolli, T. Boller, J. Borg, J. Broderick *et al.*, The aperture array verification system 1: System overview and early commissioning results, *Astron. Astrophys.* **655**, A5 (2021).
- [15] S. Hoover *et al.* (ANITA Collaboration), Observation of ultra-high-energy cosmic rays with the ANITA balloon-borne radio Interferometer, *Phys. Rev. Lett.* **105**, 151101 (2010).
- [16] H. Schoorlemmer and W. R. Carvalho, Radio interferometry applied to the observation of cosmic-ray induced extensive air showers, *Eur. Phys. J. C* **81**, 1120 (2021).
- [17] A. Balagopal V., A. Haungs, T. Huege, and F. G. Schröder, Search for PeVatrons at the Galactic Center using a radio air-shower array at the South Pole, *Eur. Phys. J. C* **78**, 111 (2018); **78**, 1017(E) (2018); **81**, 483(E) (2021).
- [18] HEP ML Community, A living review of machine learning for particle physics, <https://iml-wg.github.io/HEPML-LivingReview/>, (accessed on 28 Nov 2025).
- [19] M. Erdmann, F. Schlüter, and R. Smida, Classification and recovery of radio signals from cosmic ray induced air showers with deep learning, *J. Instrum.* **14**, P04005 (2019).

- [20] P. A. Bezyazeev *et al.* (Tunka-Rex Collaboration), Measurement of cosmic-ray air showers with the Tunka Radio Extension (Tunka-Rex), *Nucl. Instrum. Methods Phys. Res., Sect. A* **802**, 89 (2015).
- [21] D. Shipilov *et al.* (Tunka-Rex Collaboration), Signal recognition and background suppression by matched filters and neural networks for Tunka-Rex, *EPJ Web Conf.* **216**, 02003 (2019).
- [22] P. Bezyazeev *et al.* (Tunka-Rex Collaboration), Reconstruction of sub-threshold events of cosmic-ray radio detectors using an autoencoder, *Proc. Sci. ICRC2021* (2021) 223 [arXiv:2108.04627].
- [23] A. Rehman, Convolutional Neural Networks for Radio Signals from Cosmic-Ray Air Showers, Ph.D. thesis, University of Delaware, ProQuest Dissertations, 2024, <https://www.proquest.com/dissertations-theses/convolutional-neural-networks-radio-signals/docview/3163000979/se-2>.
- [24] R. Abbasi *et al.* (IceCube Collaboration), IceTop: The surface component of IceCube, *Nucl. Instrum. Methods Phys. Res., Sect. A* **700**, 188 (2013).
- [25] R. Abbasi *et al.* (IceCube Collaboration), Accounting for changing snow over 10 years of IceTop, and its impact on the all-particle cosmic ray spectrum, *Proc. Sci. ICRC2023* (2023) 377.
- [26] A. Leszczyńska and M. Plum (IceCube Collaboration), Simulation and reconstruction study of a future surface scintillator array at the IceCube neutrino observatory, *Proc. Sci. ICRC2019* (2020) 332 [arXiv:1909.02258].
- [27] A. Haungs (IceCube Collaboration), A scintillator and radio enhancement of the IceCube surface detector array, *EPJ Web Conf.* **210**, 06009 (2019).
- [28] F. G. Schröder (IceCube-Gen2 Collaboration), Physics potential of a radio surface array at the south pole, *EPJ Web Conf.* **216**, 01007 (2019).
- [29] R. Abbasi *et al.* (IceCube Collaboration), First air-shower measurements with the prototype station of the IceCube Surface Enhancement *Proc. Sci., ICRC2021* (2021) 314 [arXiv:2107.08750].
- [30] R. Turcotte-Tardif, Radio measurements of cosmic rays at the South Pole, Ph.D. thesis, KIT, Karlsruhe, 2022, [10.5445/IR/1000160782](https://doi.org/10.5445/IR/1000160782).
- [31] F. G. Schröder *et al.* (IceCube Collaboration), Status and plans for the instrumentation of the IceCube Surface Array Enhancement, *Proc. Sci. ICRC2023* (2023) 342 [arXiv:2402.17854].
- [32] B. Graham, Fractional Max-Pooling, [arXiv:1412.6071](https://arxiv.org/abs/1412.6071).
- [33] F. Chollet *et al.*, Keras, <https://keras.io> (2015).
- [34] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [35] IceCube Collaboration, RadioPulseDenoising: Network architecture and processing scripts, <https://github.com/icecube/RadioPulseDenoising> (2025).
- [36] W. D. Apel *et al.* (LOPES Collaboration), Improved absolute calibration of LOPES measurements and its impact on the comparison with REAS 3.11 and CoREAS simulations, *Astropart. Phys.* **75**, 72 (2016).
- [37] K. Belov *et al.* (T-510 Collaboration), Accelerator measurements of magnetically-induced radio emission from particle cascades with applications to cosmic-ray air showers, *Phys. Rev. Lett.* **116**, 141103 (2016).
- [38] W. D. Apel *et al.* (Tunka-Rex Collaboration and LOPES Collaboration), A comparison of the cosmic-ray energy scales of Tunka-133 and KASCADE-Grande via their radio extensions Tunka-Rex and LOPES, *Phys. Lett. B* **763**, 179 (2016).
- [39] D. Heck *et al.*, CORSIKA: A Monte Carlo code to simulate extensive air showers, Forschungszentrum Karlsruhe Report No. FZKA 6019, 1998.
- [40] T. Huege, M. Ludwig, and C. W. James, Simulating radio emission from air showers with CoREAS, *AIP Conf. Proc.* **1535**, 128 (2013).
- [41] T. Böhlen, F. Cerutti, M. Chin, A. Fassò, A. Ferrari, P. Ortega, A. Mairani, P. Sala, G. Smirnov, and V. Vlachoudis, The FLUKA code: Developments and challenges for high energy and medical applications, *Nucl. Data Sheets* **120**, 211 (2014).
- [42] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, Hadronic interaction model Sibyll 2.3d and extensive air showers, *Phys. Rev. D* **102**, 063002 (2020).
- [43] R. Abbasi *et al.* (IceCube Collaboration), Framework and tools for the simulation and analysis of the radio emission from air showers at IceCube, *J. Instrum.* **17**, P06026 (2022).
- [44] A. Rehman, A. Coleman, and F. G. Schröder (IceCube Collaboration), Deep learning for the classification and recovery of cosmic-ray signals against background measured at south pole, *Proc. Sci. ARENA2022* (2023) 012.
- [45] F. G. Schröder, A. L. Connolly, T. Huege, and A. Rehman, Towards a standard definition of the signal-to-noise ratio for radio signals of ultra-high-energy particles, *Proc. Sci. ARENA2022* (2023) 027 [arXiv:2306.05901].
- [46] R. Abbasi *et al.* (IceCube Collaboration), Enhancing air-shower observations: Results from an IceCube Surface Array Prototype Station, *Proc. Sci. ARENA2024* (2024) 038.
- [47] E. N. Paudel (IceCube Collaboration), New IceTop trigger in the context of the planned IceCube surface detector enhancement at the south pole, *J. Instrum.* **19**, C01031 (2024).
- [48] P. Gálvez Molina *et al.* (IceCube Collaboration), Detection of Radio Signals from Cosmic Rays Using Convolutional Neural Networks with Data from SKALA antennas at IceTop, in *Workshop on Machine Learning for Analysis of High-Energy Cosmic Particles, January 2025*, University of Delaware (2025), pp. 27–31 [10.5281/zenodo.14919724](https://doi.org/10.5281/zenodo.14919724).
- [49] J. Álvarez-Muñiz *et al.* (GRAND Collaboration), The Giant Radio Array for Neutrino Detection (GRAND): Science and design, *Sci. China Phys. Mech. Astron.* **63**, 219501 (2020).
- [50] Q. Abarr *et al.* (PUEO Collaboration), The Payload for Ultrahigh Energy Observations (PUEO): A white paper, *J. Instrum.* **16**, P08035 (2020).
- [51] F. G. Schröder (IceCube-Gen2 Collaboration), Design and expected performance of the IceCube-Gen2 surface array and its radio component, *Proc. Sci. ARENA2022* (2023) 058 [arXiv:2306.05900].

- [52] A. Aab *et al.* (Pierre Auger Collaboration), Observation of inclined EeV air showers with the radio detector of the Pierre Auger Observatory, *J. Cosmol. Astropart. Phys.* **10** (2018) 026.
- [53] A. Abdul Halim *et al.* (Pierre Auger Collaboration), Status and expected performance of the AugerPrime Radio Detector, *Proc. Sci. ICRC2023* (2023) 344.
- [54] S. Verpoest *et al.* (IceCube-Gen2 Collaboration and Pierre Auger Collaboration), Observation of air showers with an IceCube-Gen2 prototype station at the Pierre Auger Observatory, *Proc. Sci. ARENA2024* (2024) 037 [arXiv:2409.00713].
- [55] M. Battisti, J. Eser, A. Olinto, and G. Osteria (JEM-EUSO Collaboration), POEMMA-Balloon with radio: A balloon-born multi-messenger multi-detector observatory, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1069**, 169819 (2024).