Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Afri-COVID: A Comprehensive COVID-19 Dataset for Statistical Modeling and Public Health Research in Algeria

Mohamed Zemmouri [1, *], Fatma Zohra Laallam [1], Okba Kazar [2], Adel Oulefki [3], Yassine Himeur [3], Sos Agaian [4]

[1] *Artificial Intelligence and Information Technologies Laboratory, Kasdi Merbah University, Ouargla, Algeria*

[2] *Computer Science Department, University of Kalba, Sharjah, UAE*

[3] *College of Engineering and Information Technology, University of Dubai, Dubai, UAE*

[4] *Computer Science Department, College of Staten Island, NY, USA*

**Corresponding Author:*

ARTICLEINFO

Received: 10 May 2026

Revised: 03 July 2026

Accepted: 12 July 2026:

ABSTRACT

CT imaging is an invaluable tool for screening suspected COVID-19 patients and monitoring individuals undergoing treatment. This approach can potentially spot clinical condition lesions with false-negative RT-PCR test results. Thus, high-quality CT scan data is essential for high-quality lesion detection algorithms. Yet, there is no dataset gathering positive patients from all variants of coronavirus (e.g., Alpha, Beta, Gamma, Delta, Lambda, and Omicron). Moreover, existing data contain many flaws, such as scarcity due to the low number of patients, CT exams and slices, mixed modality (X-ray, CT scan, and MRI), mixed sources and devices, and short acquisition time; which make them unsuitable for machine learning applications. This paper presents Afri-COVID, an Algerian COVID-19 dataset, which represents a large-scale repository containing 2500 anonymized CT images and a roundup of the latest statistical data and trends about COVID-19 world-wide. CT data were acquired between January 1, 2020, and October 31, 2022, from three Algerian cities. Additionally, comparative results on different public CT scan COVID-based datasets are presented to inform the state-of-the-art.

Keywords: COVID-19 Dataset, Decision Analytics, Statistics, Machine Learning, Segmentation, CT Scan, Medical Imaging, Public Health Data.

1. INTRODUCTION

The novel COVID-19 (coronavirus disease 2019) pandemic has spread all over the world since its appearance at the end of 2019, according to the WHO [1] and other statistical centers [2, 3, 4], causing a major outbreak and a severe impact on the health and lives of many people universally. Globally, over 773 million confirmed cases and over 6.87 million deaths had been reported as of December 24, 2023.

Computed Tomography (CT) imagery is a crucial tool in the fight against COVID-19 [5]. CT has been shown to be more sensitive in the early diagnosis of COVID-19 infection compared to Reverse Transcription-Polymerase Chain Reaction (RT-PCR) tests [6]. On the other hand, compared to X-rays, CT screening is widely preferred due to its merit and three-dimensional view of the lung [7]. Moreover, CT images show more pathological information with higher accuracy [8], such as the lung burden, the percentage of high opacity [9, 7], and the lung severity score, which can be used to detect the infection in the early stage, monitor the disease progression, and help us understand the course of COVID-19 [7, 10, 11].

Machine Learning (ML) approaches have been widely applied in medical image analysis [12, 13, 14] and spatially to combat COVID-19 [15, 16]. For instance, ML can be used for building a contactless imaging workflow to prevent transmission from patients to healthcare providers [17]. Furthermore, most screening and segmentation algorithms for COVID-19 are developed with ML models. The automatic diagnosis and COVID-19 infection quantification systems usually rely on the segmentation results generated by Deep Neural Networks (DNN) [8, 7, 18].

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Convolutional neural networks (CNNs) have shown impressive performance in many computer vision tasks. However, CNNs heavily rely on large datasets to avoid overfitting [19, 20], which is biased training of the model. Unfortunately, many application domains, particularly in medical image analysis, do not have access to big data [21]. Therefore, having a comprehensive dataset is crucial in the fight against this pandemic [22, 23]. On the other hand, detecting and diagnosing COVID-19 relies heavily on both RT-PCR and CT tests, with each playing a crucial role in patient management [24]. While RT-PCR is the standard test for diagnosis, it has been known to yield false-negative results in some cases. This is where chest CT imaging can prove to be very useful, especially in patients who are clinically suspected of having COVID-19 [25, 26]. CT imaging can provide detailed images of the lungs and can help in identifying lesions that may not have been detected through RT-PCR testing, enabling clinicians to make more informed decisions about patient management [27]. However, to achieve high accuracy in lesion detection, it is important to have access to high-quality CT scan data, which can serve as input for developing effective algorithms. Such algorithms can play a vital role in improving the accuracy of diagnosis and the quality of patient care [28].

Currently, there is a lack of a comprehensive database that collects COVID-19 positive cases from all variants of the virus, including Alpha, Beta, Gamma, Delta, Lambda, and Omicron [29, 30]. The existing data suffer from various limitations, including a scarcity of patients, a small number of CT exams and slices, mixed imaging modalities (including X-ray, CT scan, and MRI), data collected from mixed sources and devices, and short acquisition time [31]. These limitations make the existing data unsuitable for use in artificial intelligence and machine learning applications. To that end, this article introduces Afri-COVID, a dataset that focuses on collecting images from various COVID-19 patients in Algeria (Africa). The dataset contains 2500 anonymized CT images and provides an overview of the latest statistical data and trends related to the Coronavirus (COVID-19) globally. To highlight specific areas of interest, some studies were marked by radiologists using binary pixel masks. These areas of interest include consolidation and ground-glass opacities. The CT data was collected from three Algerian cities between January 1, 2020, and October 31, 2022. Furthermore, a comparative study of different COVID-19 datasets used with deep learning models vs. Afri-COVID is presented. Typically, their advantages and limitations in lesion prediction for accurate diagnosis and timely decision-making are identified. Lastly, the paper derives the challenges that exist in this area and highlights inspiring future directions.

2. RELATED WORK

In spite of the prospects of CT imagery in the contribution to the research in the era of COVID-19, available databases are still limited to a few cases, are not accompanied by other types of respiratory diseases to facilitate comparisons, and are not associated with suitably labelled masks [20]. Moreover, patients may be acquired from many sources using various imaging procedures, restricting the scope of a single investigation. Available CT exams in a few selected datasets are limited to only diseased slices rather than the entire volume. Another factor to consider in the accessible datasets is if labels are provided at the patient, study, and slice levels. In this section, we describe the most popular COVID-19 datasets we based on: patients' number, time interval acquisition, and slices' number.

a) **NCCID Dataset:** The National COVID-19 Chest Imaging Database (NCCID) is a centralized database containing chest X-ray (CXR), magnetic resonance imaging (MRI), and computed tomography (CT) images from hospital patients across the UK. The database was created to support a better understanding of the COVID-19 virus and develop technology that will enable the best care for patients hospitalized with a severe infection [32]. Unfortunately, the NCCID data is not open access; therefore, the metrics of the dataset can not be accessed. In May 2020, data access was initiated through online applications by filling and submitting according to the instructions on the NCCID website. A central data access committee will then assess applications [33].

b) **MASHHAD Dataset:** This dataset consists of unenhanced chest CTs from 1000+ patients with confirmed COVID-19 infections. The age distribution of patients who underwent CT imaging was 47.18 ± 16.32 (mean ± standard deviation) years, and the age range was between 6 and 89 years. Gender distribution was 60.9% male and 39.1% female. Images were obtained in the March 2020–January 2021 period. They were acquired at the point of care in an inpatient setting from patients with positive RT-PCR[1] tests for COVID-19, accompanied by supporting clinical symptoms [34]. The CT exams were performed with a NeuViz 16-slices CT scanner machine

[1] RT-PCR: Reverse Transcription Polymerase Chain Reaction

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(Neusoft medical systems). All images are in Digital Imaging and Communication in Medicine (DICOM) format and consist of 16-bit grayscale images composed of 512×512 pixels. Patient privacy was preserved by removing all patientspecific information from image headers. Subsequently, all images corresponding to each patient are compressed and stored in RAR format [34].

c) **MOSMED Dataset:** This dataset was obtained between the 1st of March 2020 and the 25th of April 2020 and provided by municipal hospitals in Moscow, Russia. This dataset contains anonymized human lung CT scans. There are 1110 studies in the dataset. Population: 1110 persons, males: 42%, females: 56%, other/unknown: 2%; age from 18 to 97 years, median: 47 years [35]. Every study has been saved in NIFTI format and archived into a Gzip archive. During this process, only every 10th image (Instance) was preserved in the final study file. A small subset of studies (n=50) has been annotated by professionals. The resulting masks have been saved in NIFTI format and then transformed into Gzip archives [35].

d) **COVID AMP Dataset:** The COVID AMP dataset provides an extensive collection of global COVID-19 response policies, capturing nearly 50,000 policy measures from 150 countries, 124 intermediate areas, and 235 local areas between January 2020 and June 2022. It encodes up to 40 structured and unstructured characteristics per policy, including original source and policy text. This dataset offers a broad perspective on governance strategies for pandemic response and is valuable for research in legal epidemiology and political science [36].

e) **SARI COVID-CTset:** The SARI COVID-CT dataset contains 63,849 CT scan images from 377 patients, including 15,589 images from 95 COVID-19 patients and 48,260 images from 282 non-COVID-19 individuals, gathered from Negin radiology located at Sari in Iran between March 5th to April 23rd, 2020. The scans, originally in 16-bit grayscale DICOM format, were converted to 16-bit grayscale TIFF format to protect patient privacy. The dataset is designed to aid in the training and validation of deep neural networks for COVID-19 detection. Images are available for download and visualization using a provided scripts [37].

Table 1: Comparison of Available COVID-19 Datasets

Dataset		Tracking		Patients		Metrics	
Ref.	**Name**	**Date**	**Locat.**	**COVID+**	**Total**	**Modality**	**Slices**
[38, 39]	RADIOPEDIA	2021-01			100	CT	829
[40]	ZENODO	2020-04		20	20	CT	
[41]	COHEN	2020-10		468	661	Mixed	950
[35]	MOSMED	2020-03 → 2020-04	Russia	860	1,110	CT	
[22]	MEDRXIV	2020-04	China	216		CT	746
[42]	SAO PAULO	2020-04	Brazil	60	120	CT	2,482
[37]	SARI	2020-03 → 2020-04	Iran	95	377	CT	63,849
[34]	MASHHAD	2020-03 → 2021-01	Iran	1,000		CT	
[43]	COVID-CT-MD	2020-03 → 2020-04	Iran	169	305	CT	11,944
[33, 32]	NCCID[2]	2020-05 → 2022-10	UK		9,640[3]	Mixed	7,060
	Afri-COVID	**2020-01 → 2022-12**	**Algeria**	**1,149**	**2,496**	**CT**	**> 10⁶**

Using CT imaging datasets for screening COVID-19 presents several challenges. Typically, CT imaging datasets specifically focused on COVID-19 screening may be limited in size and availability. Collecting a large and diverse

[2] The information of NCCID set was last updated on 31 October 2022
[3] The number of patients -in NCCID set- of all modalities (chest X-ray, CT and MR) combined is 18, 723 with 6, 582 positive cases.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

dataset that accurately represents the various manifestations of COVID-19 can be challenging. They can also vary in terms of image quality, resolution, and scanning protocols. These variabilities can affect the performance and generalizability of algorithms trained on such datasets. Additionally, the process of annotating CT images for COVID-19 can be subjective and time-consuming. Agreement among experts on the presence and extent of COVID-19 features in the images may vary, leading to inconsistencies in annotations and ground truth labels.

Moving forward, CT imaging datasets used for COVID-19 screening may suffer from biases due to variations in patient demographics, imaging protocols, and disease severity. These biases can limit the generalizability of models trained on such datasets to different populations and settings. Besides, these datasets contain sensitive patient information, and ensuring patient privacy and data protection is crucial. Ethical considerations must be addressed when collecting, storing, and sharing these datasets. Lastly, longitudinal CT imaging datasets that track the progression and treatment outcomes of COVID-19 patients are valuable for understanding the disease dynamics. However, obtaining longitudinal data can be challenging due to logistical constraints and follow-up limitations.

To that end, addressing these challenges is important to ensure the reliability, generalizability, and ethical use of CT imaging datasets for screening COVID-19.

3. AFRI-COVID DATASET

3.1 Description

This study introduces the Afri-COVID dataset, which contains the full original **2 700 CT scans** of more than **2,500 patient**. There are **1.4 million slice** images belonging to 1 149 COVID-19 cases and 830 healthy persons - more than 1 000 studies pending radiology reports -respectively. It was collected from three Algerian cities. The format of the exported radiology images was 16-bit grayscale DICOM with 512×512 pixels resolution. Since DICOM files[4] contain personal identifiers, all private patient information was removed during preprocessing.

A key novelty of our dataset is the use of 16-bit DICOM images, rather than converting them to 8-bit. This preserves subtle image details that are of-ten lost during conversion, particularly in early-stage infections where lesions are small and difficult to detect, even for clinicians. This causes ambiguity and reduces the precision of the network. Another advantage of using the DICOM format is the ease of 3D reconstruction after segmentation.

3.2 Metadata

DICOM files contain metadata that provides information about the pa-tient (name, age, and sex) and the image data, such as the size, dimensions, bit depth, modality used to create the data, and equipment settings used to capture the image. To read metadata from a DICOM file, we use the *PyDicom* Library. This library allows us to modify DICOM files to hide sensitive information such as patient names and physician names.

Another essential source of data is the reports, which contain the evalu-ation of doctors; these reports let us know even the patient is infected with COVID-19 or not, and the lesion rate of the infection. The reports are re-trieved in PDF format. To get the data, we should do document scraping using *PdfPlumber* library. We extract two kinds of metadata:

[4] DICOM files will be provided upon request

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 2: Study data description.

	Attribute	Source	Description
Patient	ID	*DICOM*	Patient Identifier
	Sex	*DICOM*	Male (M), Female (F)
	Age	*DICOM*	Age of the patient in years
Study	UID	*DICOM*	A unique identifier of the study
	Description	*DICOM*	Description of the study
	Date	*DICOM*	The date of the CT scan was done
	Modality	*DICOM*	We used only CT scan
Result	COVID-19	*Reports*	True | False
	Lesion Rate	*Reports*	Lesion Rate evaluated by professionals

- **The series data:** which contains all the metadata of the DICOM file - after hiding some elements - as the result file is large was separated by month.
- **The studies data:** which contains information about the patient and in-stead of the result of the study, you can find the details of the attributes in Table 2.

4. PROCESS AND METHODOLOGY

The section describes the process of importing and organizing data. The process -which is shown in Figure 1- of extracting and organizing medical imaging data from the Syngo solution and Electronic Document Management (EDM) system involves several steps. First, the data is extracted using filters based on date and type of study, and then PyDicom and PdfPlumber libraries are used to import and organize the data into a structured format that can be analyzed and processed. The dataset is then de-identified to protect patient privacy by removing or obfuscating personal information while maintaining useful DICOM attributes such as gender and age. Data cleaning is performed to fix or remove incorrect, corrupted, duplicate, or incomplete data within the data set, and patient age is computed based on birth date. Finally, the collected data is reorganized and grouped from extracted DICOM metadata and the scraped studies reports to make it easier to work with.

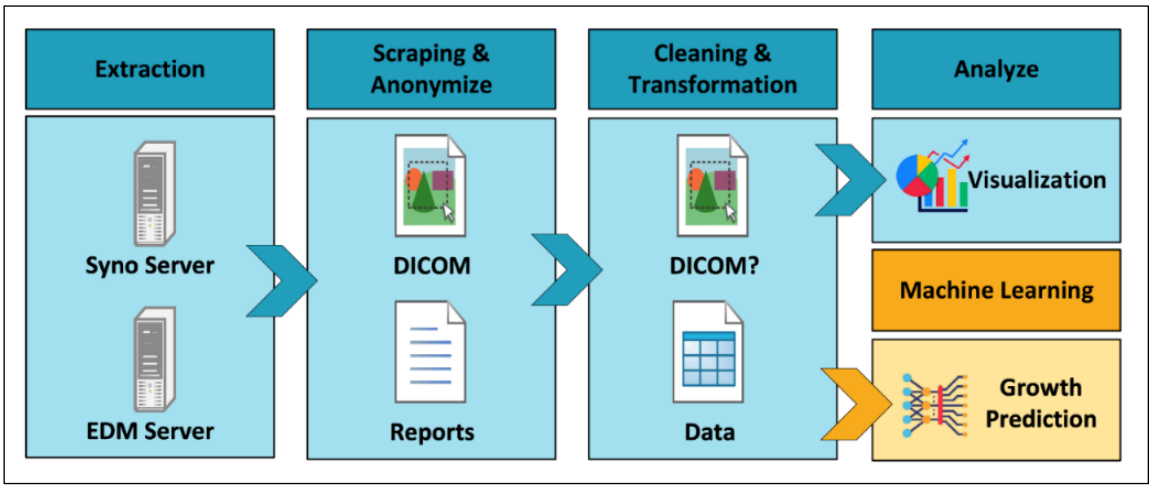


Figure 1: Process and methodology

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.1 Data extraction

This operation involves retrieving DICOM files from the Syngo solution. To do this, a filter is applied based on two criteria: date and type of study. The date filter is used to retrieve DICOM files for a specific time period, while the type of study filter is used to select only those DICOM files that correspond to a specific type of medical study. In this case, the type of study chosen is "Thorax CT", which refers to computed tomography (CT) scans of the chest area. In addition to retrieving the DICOM files, the operation also involves extracting the reports from the Electronic Document Management (EDM) system. The EDM system is used to store and manage medical reports, such as radiology reports or pathology reports, which provide the results of a medical study. In this case, the reports are extracted as PDF files to provide a record of the study's results.

4.2 Data scraping

Once the DICOM files and PDF reports are extracted from the Syngo so-lution and EDM system, respectively, the next step is to import the extracted data into a spreadsheet. This involves organizing the data in a structured format that can be easily analyzed and processed. To do this, the team uses two Python libraries: PyDicom and PdfPlumber.

- **PyDicom** is an open-source Python package that provides functional-ity for reading and writing DICOM files. It allows the team to access and manipulate the DICOM files extracted from the Syngo solution.

  PyDicom provides methods to read the metadata and image data of the DICOM files, as well as to modify the metadata if necessary.

- **PdfPlumber**, on the other hand, is a Python library that provides tools for extracting data from PDF files. In this case, the team uses PdfPlumber to extract the text and tables from the PDF reports ex-tracted from the EDM system. PdfPlumber provides a simple interface to extract text and tables from PDF documents, which can then be converted to a structured format that can be easily imported into a spreadsheet.

By using PyDicom and PdfPlumber together, the team is able to extract and organize data from both DICOM files and PDF reports, and then import this data into a spreadsheet for further analysis and processing. This allows the team to efficiently manage and analyze medical imaging data, and to generate insights that can inform medical decision-making.

4.3 De-identification (Data anonymization)

We de-identified all CT studies by deleting or obfuscating all names, UIDs, dates, times, comments, and center-related information to protect the patients' privacy. To maintain the dataset's statistical features, sev-eral useful DICOM attributes related to the patient's gender, age, scanner type, and picture acquisition parameters have been maintained. Patient ID and UID properties, which are required to maintain CT study uniformity, are replaced with freshly generated values that prevent patients from being iden-tified. This phase is ensured using the Deid package. This Python module is intended for the basic coding of medical images, which means "cleaning" image headers and pixel data and replacing identifiers with other anonymous identifiers.

4.4 Data cleaning

In general, data cleaning is the process of fixing or removing incorrect, cor-rupted, incorrectly formatted, duplicate, or incomplete data within a dataset. Our work consists of removing duplicated data and computing patient age based on the Birth date.

4.5 Data transformation

After cleaning data, we reorganize and group collected data from ex-tracted DICOM metadata and the scraped studies reports; as we have seen in the dataset description section, we have:

- **Patients' data:** contains information about the patient and the study (for more detail, see Table 2).
- **DICOM data:** contains DICOM files and their metadata extracted - to make the interaction easy-. Given their size, and to facilitate the in-formation retrieval, the data are separated monthly, and they are structured - as shown in Figure 2 - as follows: {Month}/{Patient}/{Study}/{Slice}

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

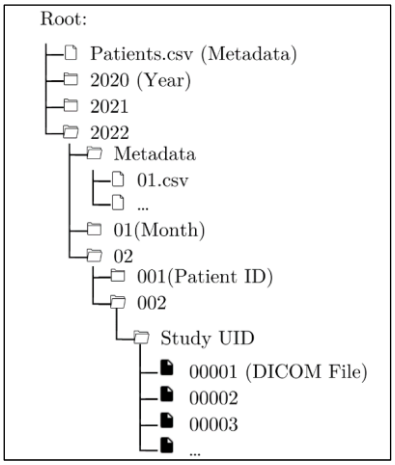


Figure 2: Dataset structure

5. DESCRIPTIVE ANALYSIS OF THE DATASET

This section provides statistical analysis and visualization of Afri-COVID. It begins with section 5.1, which delves into the lesion rate, providing insights into the prevalence and severity of COVID-19-associated lesions. Moving forward, section 5.2 sheds light on the various strains of COVID-19 that have surfaced and their effects on disease spread and severity. Lastly, section 5.3 examines the distribution of COVID-19 patients across age and gender, highlighting its significance for healthcare resource allocation and public health measures.

Table 3: Descriptive Statistics of COVID-19 Patients' Dataset

Age (Year)	Lesion Rate	Male	Female
57 ± 15.88	25% ± 20.93%	647 (56.31%)	502 (43.69%)

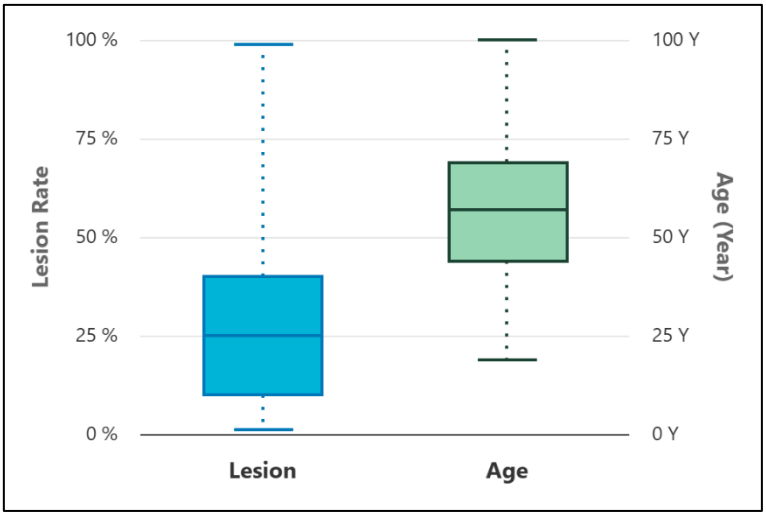


Figure 3: Descriptive statistics of COVID-19 patients' dataset

Table 3 presents descriptive statistics for the dataset, covering age, lesion rate, and gender distribution. Age and lesion rate are continuous variables and are displayed as means ± standard deviations (SD), while gender is a

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

categorical variable represented as a total count and percentage of the overall population. To obtain a better grasp of the age and lesion rate distribution within the dataset, we employ a boxplot visualization in Figure 3. The box-plot showcases the data distribution, including the median, quartiles, and outliers. In the age boxplot, we notice a slightly right-skewed distribution, signifying that positive cases are more prevalent among the elderly. Conversely, the lesion rate boxplot suggests that COVID-19 cases are mainly concentrated in lower percentage lesions.

In summary, the descriptive statistics and boxplot visualization offer valuable insights into the dataset and provide a foundation for further analysis. Additionally, Table 4 displays the lesion distribution by gender, while Figure 4 offers a visual representation of the same. It is noteworthy that males have a higher median lesion rate than females, though the difference lacks statistical significance ($p = 0.116$).

5.1 Distribution of COVID-19 patients by Age and Sexe

As depicted in Figure 4, males are more numerous than females in this dataset. This male dominance is a prevalent trend in many COVID-19-related datasets [44], possibly due to the fact that men are more susceptible to COVID-19 than women [45].

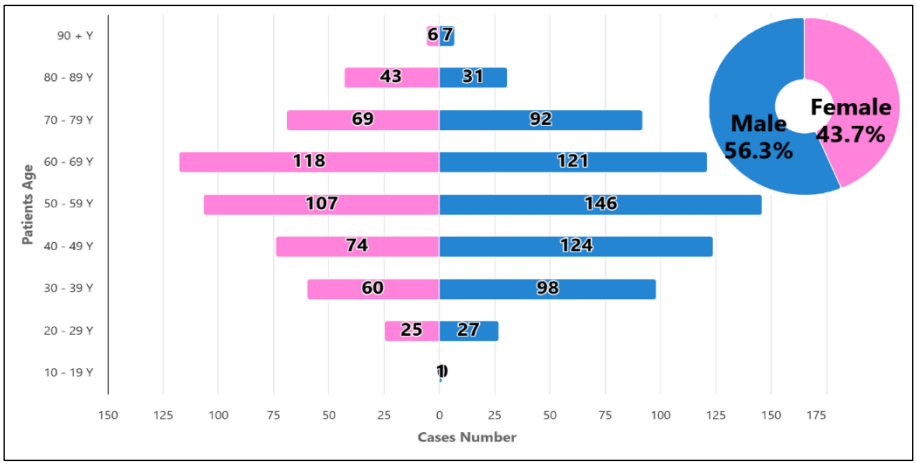


Figure 4: Visualization of gender distribution of COVID-19 patients.

Although younger individuals typically make up a larger proportion of any youthful society, Table 4 and its visual representation in Figure 5 demonstrate that the majority of COVID-19 cases occur in individuals over 40 years old.

- 80% of cases occur in individuals over 40 years old, while only 20% affect young people.
- There are no reported cases among children under 15 years old.
- These findings align with the lesion boxplot depicted in Figure 3 and confirm that positive cases are primarily distributed among individuals over 40 years old.

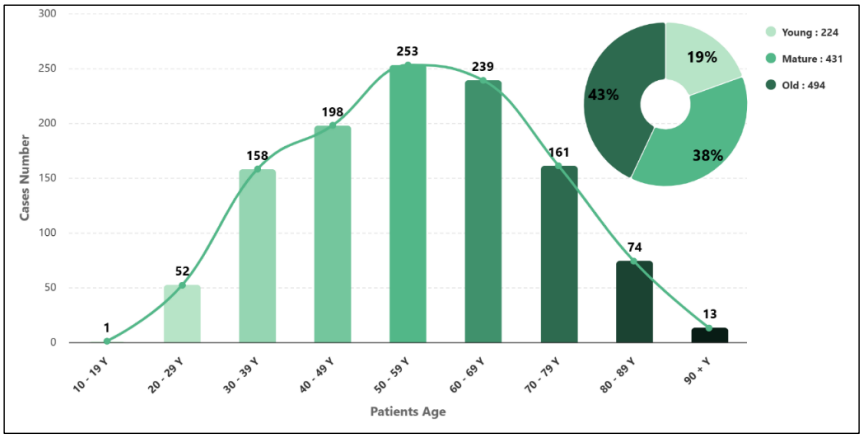


Figure 5: Visualization of age distribution of COVID-19 patients

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Based on this analysis, we can draw the following conclusions:

- As noted in previous studies [45], men appear to be more susceptible to COVID-19 compared to women.
- COVID-19 infection is significantly less prevalent among infants, chil-dren, and young adults.

5.2 Lesion Rate

Table 5 and Figure 6 provide information on the distribution of COVID-19 cases based on injury rate.

Table 4: Age and Gender Distribution				Table 5: Lesion Rate Distribution	
Age (Years)	**Female**	**Male**		**Lesion Rate (%)**	**Patients**
01–10	0	0		**01–10**	344
11–20	0	1		**11–20**	226
21–30	25	27		**21–30**	209
31–40	60	98		**31–40**	28
41–50	74	124		**41–50**	207
51–60	107	146		**51–60**	36
61–70	118	121		**61–70**	32
71–80	69	92		**71–80**	56
81–90	43	31		**81–90**	7
90+	6	7		**91–100**	4
Total	**502**	**647**		**Total**	**1,149**

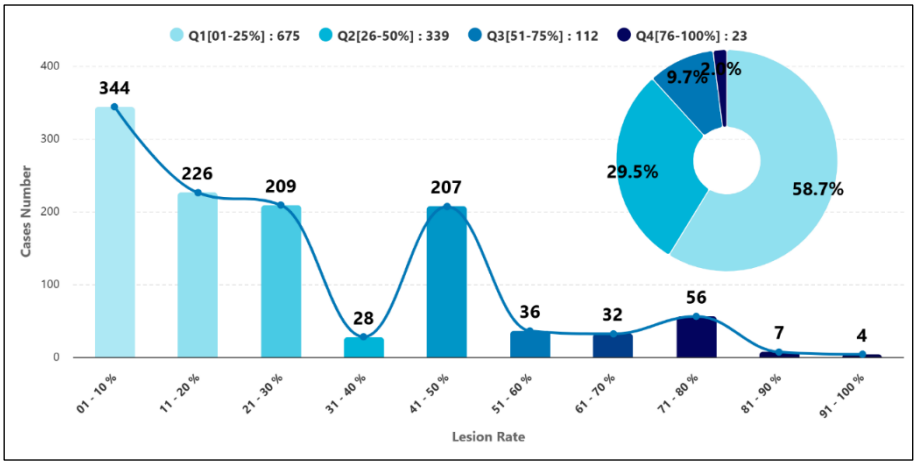


Figure 6: Visualization, Lesion Rate distribution of COVID-19 patients

These statistics reveal that most cases occur in the early stages of the disease:

- 14% of cases have a lesion rate less than 6%.
- 60% of cases have a lesion rate less than 25%.
- 90% of cases have a lesion rate less than 50%. This concords with the lesion boxplot of Figure 7 reveal that the positive cases are mainly distributed in a lower percentage.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- What we can deduct from this analysis is that:
- The symptoms appear on the patient in the early stages of the infection, prompting him to have a CT scan.
- CT scan is a reliable way of detecting COVID-19 even in the early stages.

5.3 Variants

Table 6 displays statistics on the lesion rate per variant of COVID-19. The columns provide details on the average lesion rate, number of cases, and lesion distribution by range (1-25%, 26-50%, 51-75%, and 76-100%). The table highlights that the average lesion rate for each variant ranges from 21% to 47%, with the Alpha variant registering the highest average lesion rate and the Beta variant the lowest. Additionally, the table shows the number of cases for each variant, ranging from 7 for the Alpha variant to 801 for the Lambda variant. The distribution of the lesion range for each variant is displayed in the last four columns of the table. For instance, in the Alpha variant, 43% of cases had lesions in the range of 1-25%, 14% had lesions in the range of 26-50%, 43% had lesions in the range of 51-75%, and 0% had lesions in the range of 76-100%.

Figure 7 presents a stacked bar chart illustrating the lesion rate distribution per variant, which complements the information provided in the table and enhances data comprehension. Each variant is represented by a distinct shade of blue, and the chart shows the lesion distribution ranges (1-25%, 26-50%, 51-75%, and 76-100%). The figure clearly depicts the variation in lesion rate distribution across different COVID-19 variants. For instance, the Alpha variant has an almost uniform distribution of lesions across all ranges, whereas the Lambda variant has a greater proportion of cases with lesions in the 51-75% range.

Table 6: Lesion rate distribution across SARS-CoV-2 variants.

General Statistics			Lesion Rate Distribution			
Variant	**Avg. Lesion (%)**	**Cases**	**01–25 (%)**	**26–50 (%)**	**51–75 (%)**	**76–100 (%)**
Alpha	47	7	43	14	43	—
Beta	21	12	83	8	8	—
Delta	28	64	53	39	8	—
Lambda	28	801	60	29	9	2
Omicron	30	265	56	29	12	3
Total	**29%**	**1,149**	**59%**	**30%**	**10%**	**2%**

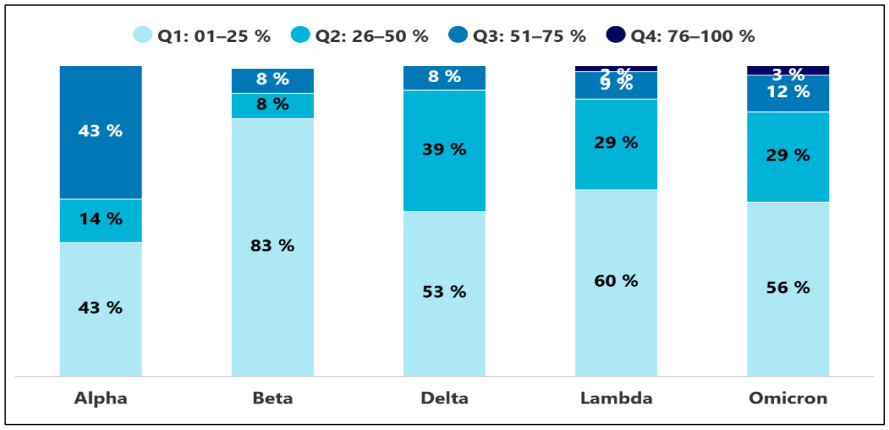


Figure 7: Lesion Rate distribution per Variant

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. COVID-19 DATASETS COMPARISON

Despite the potential value of CT in COVID-19 research, the publicly available datasets are typically limited in terms of the number of cases and time interval, and may involve mixed modalities such as X-ray, CT scan, and MR. Moreover, the scans may originate from various sources with distinct imaging protocols and resolutions, which can constrain their analysis. Furthermore, it is crucial to consider the number of slices, whether infected or not, as it contributes to 3D reconstruction, and the multi-test, which is essential in studying the spread of infection. However, no available data collection mentions the multi-test of examinations. A summary of the available datasets and related COVID-19 information is provided in Table 1.

6.1 Time interval and COVID-19 variants

Like all viruses, SARS-CoV-2, the virus responsible for COVID-19, undergoes genetic changes over time. However, most of these changes have negligible impact on the virus's properties [46]. Table 7 presents the primary COVID-19 variants identified by the World Health Organization (WHO).

In Figure 8, we compare our dataset with other datasets according to the date of CT exams and the coverage of COVID-19 variants.

Table 7: Major SARS-CoV-2 Variants and Emergence [46]

WHO Label	Scientific Name	First Reported	Country/Region
COVID-19	SARS-CoV-2	Dec 2019	China
Alpha	B.1.1.7	Sep 2020	United Kingdom
Beta	B.1.351	May 2020	South Africa
Gamma	P.1	Nov 2020	Brazil
Delta	B.1.617.2a	Oct 2020	India
Lambda	C.37	Dec 2020	Peru
Omicron[5]	B.1.1.529	Nov 2021	Multiple countries

We have observed that our dataset differs from others in that it encompasses all existing variants of COVID-19. This particular aspect yields two significant advantages. Firstly, it mitigates the issue of overfitting, which occurs when the model learns from only one variant. Secondly, it enhances the accuracy of detection, segmentation, and prediction algorithms for COVID-19 by leveraging deep learning techniques, which will be useful for any potential variants that may emerge in the future.

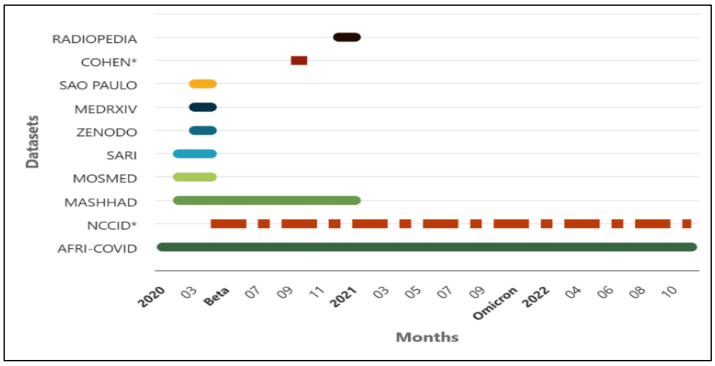


Figure 8: Comparing datasets by date of CT exams

[5] Includes BA.1, BA.2, BA.3, BA.4, BA.5, and descendant lineages. It also includes BA.1/BA.2 circulating recombinant forms such as XE.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6.2 Patients, Studies and Slices number

The richness of a dataset is commonly evaluated by assessing the number of patients, studies, and slices. As such, these metrics are crucial in evaluating the quality of any dataset. Figure 9 presents a ranking of our dataset in comparison to others based on the number of patients, which includes negative cases.

The number of slices helps in the precision of the learning and especially the 3D reconstruction.

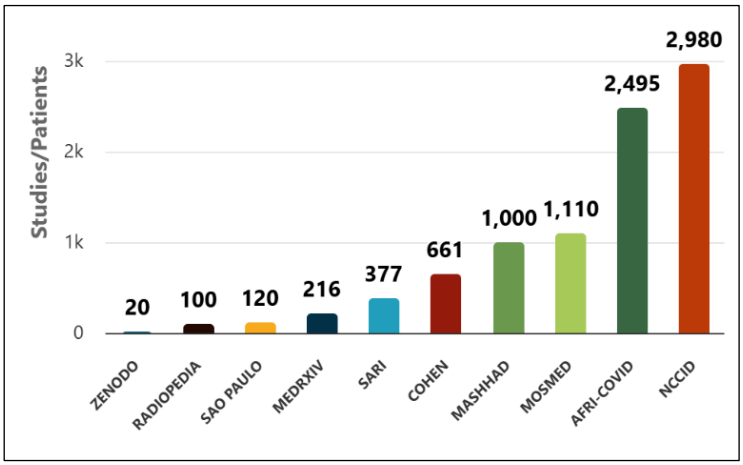


Figure 9: Datasets comparison by patients' number

6.3 Multi-test

The datasets we have reviewed thus far do not provide information on patients undergoing multiple scans. While this information may not contribute to the classification or segmentation processes, it is crucial in predicting the growth rate of lung lesions in patients. Table 8 displays the number of CT scans conducted for each patient.

Table 8: Distribution of patients with multiple CT examinations.

Number of Examinations	Number of Patients
2	99
3	21
4	1
5	4
6	1
Total	**126**

7. RESULTS

7.1 Statistical analysis

Figure 10 provide important statistical information about the spread of COVID-19 in Algeria from 2020 to 2022, focusing on the different variants and their monthly peak. It highlights the dates, number of cases and observations for each variant. Below are some key observations that can be made based on the data presented:

- The first variant mentioned is Alpha, which was first detected in April 2020 and had a monthly peak of 5,388 cases in May 2020. It is interesting to note that the number of cases increased by 360% in just one month, indicating the rapid spread of the virus.
- The next variant, Beta, was detected in July 2020 and had a monthly peak of 16,487 cases, which is a significant increase of 265% compared to the previous month. This suggests that the spread of the virus was accelerating.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- The Delta variant, which is known to be highly transmissible, was detected in November 2020 and had a monthly peak of 25,257 cases. The number of cases increased by 294% compared to the previous month, indicating a sharp rise in transmission.
- The Lambda variant was detected in July 2021 and had a monthly peak of 31,766 cases. Although the increase in cases was lower than some of the earlier variants, it still resulted in a 197% increase in cases.
- Finally, the Omicron variant, which was first detected in January 2022, had a monthly peak of 33,685 cases, representing a 326% increase compared to the previous month. The data also shows that the number of cases increased by 573% in July 2022, indicating a significant surge in transmission.

Table 9 shows the number of cases and the percentage of each SARS-CoV-2 variant in our Afri-COVID. Here is a statistical comment on the given data: The most prevalent variant is the Lambda variant, which accounts for 34.39% of all cases. It is followed by the Delta variant, which accounts for 22.68% of all cases, and the Omicron variant, which accounts for 23.94% of all cases. The least prevalent variant is the SARS-CoV-2 variant, which accounts for only 0.26% of all cases. The Gamma variant has zero cases in the given data, indicating that it is not present in the population under study. It is important to note that the table only represents a snapshot of the prevalence of variants in the population under study and may not be representative of the global prevalence of the variants.

Table 9: Statistics, Variants Cases

Variant	Cases	Percentage
SARS-CoV-2	716	0.26 %
Alpha	8,678	3.20 %
Beta	42,136	15.52 %
Gamma	—	— %
Delta	61,562	22.68 %
Lambda	93,360	34.39 %
Omicron	64,996	23.94 %
Total	**271,448**	**100 %**

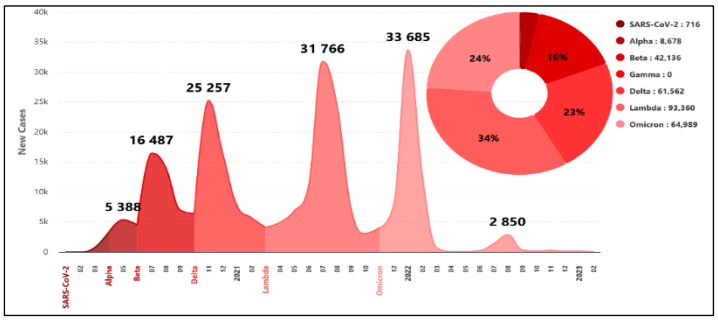


Figure 10: Monthly confirmed cases per Variant in algeria

7.2 Lesion Growth Analysis

Table 10 presents data on the lesion growth of eleven COVID-19 cases caused by different variants[6]. Each row represents a case, and the columns provide information on the ID, sex, age, and variant of the virus. The table is well-organized, and each column has clear headings, making it easy to read and understand. The next four columns in the table provide information on the first study conducted on each case. These columns include the date of the study, the

[6] The DICOM files of these eleven cases are freely available for download at *github.com/m-zemmouri/Afri-COVID*.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

percentage of the lesion on the right and left side of the body, and the total percentage of the lesion. The following four columns provide information on the second study conducted on each case, including the lesion percentage on the right and left side of the body and the total percentage of the lesion. The last column of the table shows the growth of the lesion percentage between the first and second studies. The data in the table shows that lesion growth varies significantly between cases and variants. For example, cases A and B are both caused by the Beta variant, but the lesion growth in case A is much higher than in case B. In contrast, cases C, D, and E are all caused by the Delta variant, but the lesion growth in these cases is very different. Similarly, the lesion growth in cases H, I, J, and K, all caused by the Omicron variant, also shows variation. Overall, the table provides a clear overview of the lesion growth in COVID-19 cases caused by different variants and allows for easy comparison between cases. The data can be used to analyze and understand the differences in lesion growth between variants and inform clinical decision-making.

Table 10: Lesion Growth – Data

Case				First Study				Second Study				Observation	
ID	**Sex**	**Age**	**Variant**	**Date**	**Lesion (%)**	**Right (%)**	**Left (%)**	**Interval (Days)**	**Lesion (%)**	**Right (%)**	**Left (%)**	**Growth (%)**	**Stage**
A	M	38	*Beta*	2020-07-03	16	19	14	11	39	46	33	23	**Spread**
B	F	85	*Beta*	2020-09-11	52	48	55	17	56	51	62	4	**≈**
C	M	47	*Delta*	2020-11-05	9	10	8	5	24	25	23	15	**Spread**
D	M	58	*Delta*	2020-11-25	11	10	13	4	14	12	15	3	**≈**
E	M	76	*Delta*	2020-12-22	18	18	19	9	16	13	20	-2	**≈**
F	M	72	*Lambda*	2021-04-14	24	25	24	15	24	25	24	—	**Stable**
G	M	73	*Lambda*	2021-06-17	7	7	7	7	29	34	24	22	**Spread**
H	M	31	*Omicron*	2021-11-07	15	16	15	4	15	17	13	—	**Stable**
I	F	55	*Omicron*	2021-11-10	12	10	13	16	15	11	20	3	**Spread-R**
J	F	58	*Omicron*	2021-12-07	23	24	21	14	7	8	7	-15	**Recovery**
K	M	45	*Omicron*	2022-01-14	9	10	9	6	8	9	8	-1	**≈**

The algorithm U-net (R231CovidWeb) described in [47] [13] was used to visualize lung segmentation in our data. The algorithm includes placing markups on the lung or using a deep-learning lung and lobe segmentation algorithm. The algorithm identifies five regions of interest: "Bulla/emphysema," "Inflated," "Infiltrated," "Collapsed," and "Lung

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

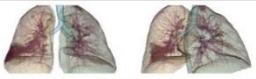
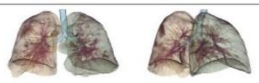
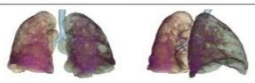
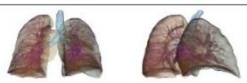
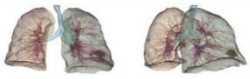
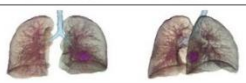
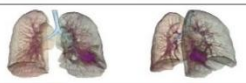
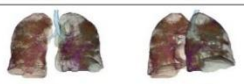
Vessel." The volume of each segment is calculated using "Segment statistics." The results of the spatial reconstruction of the diseased lung segments are displayed in 3D views in standard colors, with "Bulla" shown in black, "Inflated" in blue, "Infiltrated" in yellow, "Collapsed" in pink, and "Vessel" in red. The total results of the segmentation include information on all five regions of interest. In our case we are interested in affected areas that could be calculated as follows:

$$\text{Affected total lung volume} = \frac{V_{inf} + V_{coll}}{V_{total}} \times 100 \quad (1)$$

Table 11 shows the visual growth of the lesion between the first and second studies. Take note that the pink color represents the lesion. The data in the table from (A to K) shows that lesion growth varies significantly between cases and variants, and can be used to inform clinical decision-making.

Based on this visualization, it appears that lesion growth is variable across the different cases. Cases (A, C, D, and G) all experienced some degree of growth, while Cases (B, E, H, J, and K) all shared some degree of shrinkage of the lesion. Only Cases (I and F) had a very slight growth. It's important to note that the duration between the first and second studies was different for each case, so it's possible that the degree of lesion growth could be influenced by the duration of time between the studies. Additionally, there may be other factors such as the variant or patient-specific characteristics that could contribute to differences in lesion growth.

Table 11: Lesion Growth - Visualization

Case	First Study	Second Study	Growth
A			⇈
B			≈
C			⇈
D			≈
E			≈
F			—
G			⇈
H			—
I			⇈
J			⇊
K			≈

8. DISCUSSION AND FUTURE EXPECTATIONS

The strength of our work lies in building huge data of patients affected by COVID-19 covering almost all variants. Furthermore, the data collected has the potential to quantify the COVID-19 lesion, visualize the infected area, and

track the disease changes. As future work, we consider extending this work by collecting chest CT imagery from several institutions and countries. Furthermore, we are planning to develop an algorithm to predict efficiently way the spread of the infection inside patients' lungs. We are also planning to 3D visualize the imaging data with clinical manifestations for a better examination, detection, and diagnosis of COVID-19. Till now, COVID-19 continues to spread across the world following a trajectory that is not easy to predict. We are hoping that:

- The collected COVID-19 data can be used as a tool for large-scale clinical applications and training;
- The advanced tools will be helpful to other health systems facing similar challenges, including abnormalities caused by other viruses and diseases.

Despite deep learning being considered a reliable tool to fight COVID-19 [15, 16], it needs a rich dataset to give good results. In this paper, we present a comparative study of different COVID-19 datasets used by deep learning while focusing on our dataset (Afri-COVID), which introduced a large-scale dataset containing 2,500 anonymized CT images with their results collected from 2,000 reports. Our dataset with its technical characteristics (number of slices per patient and 16-bit image quality) and its dimension represented by the number of patients (2,495 persons with 1,149 positive cases of COVID-19), the number of CT Exams (2,715) and slices ($1.4 * 10^6$ images) and more than 2 000 studies with reports, can be considered the

first dataset in our country - Algeria -, and among the richest datasets at the world level. Moreover, the time taken to collect our dataset has allowed us to collect data on all the variants passed by our country to the present day, which constitutes a solid basis for carrying out an in-depth study on the points of divergence between the different variants of COVID-19 and understanding the virus development.

9. CODE AND DATA AVAILABILITY

The codes used to perform statistical analysis presented in this paper can be downloaded at *github.com/m-zemmouri/Afri-COVID*. The metadata and the DICOM files of eleven cases are is also available at the same repository. Additionally, the full dataset can be shared upon request.

REFERENCES

[1] WHO, Situation Reports Coronavirus Disease (COVID-19) Update (2026).

URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[2] A. De, COVID Live - Coronavirus Statistics - Worldometer.

URL https://www.worldometers.info/coronavirus/

[3] M. R. Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, Coronavirus Pandemic (COVID-19), Our World in Data (2020).

URL https://ourworldindata.org/coronavirus

[4] Hasell, J., Mathieu, E., Beltekian, D., Macdonald, B., Giattino, C., Ortiz-Ospina, E., ... & Ritchie, H. (2020). A cross-country database of COVID-19 testing. Scientific data, 7(1), 345.

[5] Z. Z. Jiang, C. He, D. Q. Wang, H. L. Shen, J. L. Sun, W. N. Gan, J. Y. Lu, X. T. Liu, The Role of Imaging Techniques in Management of COVID-19 in China: From Diagnosis to Monitoring and Follow-Up, Medical Science Monitor: International Medical Journal of Experimental and Clinical Research 26 (2020) e924582–1. doi:10.12659/MSM.924582.

[6] D. M. Ali, L. G. Zake, N. K. El Kady, Role of Chest Computed Tomography versus Real Time Reverse Transcription Polymerase Chain Reaction for Diagnosis of COVID-19: A Systematic Review and Meta-Analysis, Interdisciplinary Perspectives on Infectious Diseases 2021 (2021). doi:10.1155/2021/8798575.

[7] D. P. Fan, T. Zhou, G. P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-Net: Automatic COVID-19 Lung

Infection Segmentation from CT Images, IEEE Transactions on Medical Imaging 39 (8) (2020) 2626–2637. doi:10.1109/TMI.2020.2996645.

[8] Q. Zhang, X. Ren, B. Wei, Segmentation of infected region in CT images of COVID-19 patients based on QC-HC U-net, Scientific Reports 11 (1) (2021) 1–11. doi:10.1038/s41598-021-01502-0.

[9] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coron-avirus–Infected Pneumonia in Wuhan, China, JAMA 323 (11) (2020) 1061–1069. doi:10.1001/JAMA.2020.1585.

[10] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... & Xia, L. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, 296(2), E32-E40.

[11] Ye, Z., Zhang, Y., Wang, Y., Huang, Z., & Song, B. (2020). Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. European radiology, 30(8), 4381-4389.

[12] A. Al-Kababji, F. Bensaali, S. P. Dakua, Y. Himeur, Automated liver tissues delineation techniques: A systematic survey on machine learn-ing current trends and future orientations, Engineering Applications of Artificial Intelligence 117 (2023) 105532.

[13] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, G. Langs, Automatic lung segmentation in routine imaging is primarily a data di-versity problem, not a methodology problem, European Radiology Ex-perimental 4 (1) (2020) 1–13.

[14] Hamza, A., Lekouaghet, B., & Himeur, Y. (2023, November). Hybrid whale-mud-ring optimization for precise color skin cancer image segmentation. In 2023 6th International Conference on Signal Processing and Information Security (ICSPIS) (pp. 87-92). IEEE.

[15] S. Bhattacharya, P. K. Reddy Maddikunta, Q. V. Pham, T. R. Gadekallu, S. R. Krishnan S, C. L. Chowdhary, M. Alazab, M. Jalil Piran, Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey, Sustainable Cities and Society 65 (2021) 102589. doi:10.1016/j.scs.2020.102589.

[16] O. Shahid, M. Nasajpour, S. Pouriyeh, R. M. Parizi, M. Han, M. Valero, F. Li, M. Aledhari, Q. Z. Sheng, Machine learning research towards combating COVID-19: Virus detection, spread prevention, and medical assistance, Journal of Biomedical Informatics 117 (2021) 103751, doi:10.1016/J.JBI.2021.103751.

[17] H. Rohmetra, N. Raghunath, P. Narang, V. Chamola, M. Guizani, N. R. Lakkaniga, AI-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges, Computing (2021) 1doi:10.1007/S00607-021-00937-7.

[18] T. Zhou, S. Canu, S. Ruan, Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism, International Journal of Imaging Systems and Technology 31 (1) (2020) 16–27. doi:10.1002/IMA.22527.

[19] A. Bechar, Y. Elmir, R. Medjoudj, Y. Himeur, A. Amira, Harnessing transformers: A leap forward in lung cancer image detection, arXiv preprint arXiv:2311.09942 (2023).

[20] C. Shorten, T. M. Khoshgoftaar, A survey on Image Data Augmen-tation for Deep Learning, Journal of Big Data 6 (1) (2019) 1–48. doi:10.1186/s40537-019-0197-0.

[21] I. Shiri, Y. Salimi, M. Pakbin, G. Hajianfar, A. H. Avval, A. Sanaat, S. Mostafaei, A. Akhavanallaf, A. Saberi, Z. Mansouri, et al., Covid-19 prognostic modeling using ct radiomic features and machine learning algorithms: Analysis of a multi-institutional dataset of 14,339 patients, Computers in biology and medicine 145 (2022) 105467.

[22] Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). COVID-CT-Dataset: A CT Scan Dataset

about COVID-19. arXiv preprint arXiv:2003.13865.

[23] Y. Himeur, S. Al-Maadeed, I. Varlamis, N. Al-Maadeed, K. Abualsaud, A. Mohamed, Face mask detection in smart cities using deep and transfer learning: Lessons learned from the covid-19 pandemic, Systems 11 (2) (2023) 107.

[24] Y. Himeur, S. Al-Maadeed, N. Almadeed, K. Abualsaud, A. Mohamed, T. Khattab, O. Elharrouss, Deep visual social distancing monitoring to combat covid-19: A comprehensive survey, Sustainable Cities and Society (2022) 104064.

[25] A. I. V. Sørensen, L. Spiliopoulos, P. Bager, N. M. Nielsen, J. V. Hansen, A. Koch, I. K. Meder, S. Ethelberg, A. Hviid, A nationwide question-naire study of post-acute symptoms and health problems after sars-cov-2 infection in denmark, Nature Communications 13 (1) (2022) 4213.

[26] Y. Artik, A. B. Coşğun, N. P. Cesur, N. Hızel, Y. Uyar, H. Sur, A. Ayan, Comparison of covid-19 laboratory diagnosis by commercial kits: Effec-tivity of rt-pcr to the rt-lamp, Journal of medical virology 94 (5) (2022) 1998–2007.

[27] J. Ahmadi, F. Kahkeshpour, H. Farahmand, A. E. Nadimi, K. Ghadimi, G. Bazmandegan, Z. Kamiab, Evaluation of chest ct scan finding in the patients with acute respiratory symptoms following positive results of rt-pcr-covid19, International Journal of Physiology, Pathophysiology and Pharmacology 14 (1) (2022) 48.

[28] N. Padhye, Bayesian measurement of diagnostic accuracy of the rt-pcr test for covid-19, Metrology 2 (4) (2022) 414–426.

[29] Abumalloh, R. A., Nilashi, M., Ismail, M. Y., Alhargan, A., Alghamdi, A., Alzahrani, A. O., ... & Asadi, S. (2022). Medical image processing and COVID-19: a literature review and bibliometric analysis. Journal of infection and public health, 15(1), 75-93.

[30] Sevinç, E. (2022). An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. Computers & Industrial Engineering, 165, 107912.

[31] Z. Kou, L. Shang, Y. Zhang, D. Wang, Hc-covid: A hierarchical crowd-source knowledge graph approach to explainable covid-19 misinforma-tion detection, Proceedings of the ACM on Human-Computer Interac-tion 6 (GROUP) (2022) 1–25.

[32] NHSX. (2026). National COVID-19 Chest Imaging Database (NCCID). Retrieved July 12, 2026, URL nhsx.github.io/covid-chest-imaging-database/

[33] J. Jacob, J., Alexander, D., Baillie, J. K., Berka, R., Bertolli, O., Blackwood, J., ... & Joshi, I. (2020). Using imaging to combat a pandemic: rationale for developing the UK National COVID-19 Chest Imaging Database. European Respiratory Journal, 56(2). doi:10.1183/13993003.01809-2020.

[34] S Shakouri, S., Bakhshali, M. A., Layegh, P., Kiani, B., Masoumi, F., Ataei Nakhaei, S., & Mostafavi, S. M. (2021). COVID19-CT-dataset: an open-access chest CT image repository of 1000+ patients with confirmed COVID-19 diagnosis. BMC research notes, 14(1), 178. doi:10.1186/s13104-021-05592-x.

[35] S. P. Morozov, A. E. Andreychenko, N. A. Pavlov, A. V. Vladzymyrskyy, N. V. Ledikhova, V. A. Gombolevskiy, I. A. Blokhin, P. B. Gelezhe, A. V. Gonchar, V. Y. Chernina, MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset, arXiv preprint arXiv:2005.06465 (may 2020), doi:10.1101/2020.05.20.20100362.

[36] R. Katz, K. Toole, H. Robertson, A. Case, J. Kerr, S. Robinson-Marshall, J. Schermerhorn, S. Orsborn, M. Van Maele, R. Zimmerman, T. Stevens, , A. Phelan, C. Carlson, E. Graeden, COVID AMP: An Open Access Dataset of COVID-19 Response Policies, medRxiv (2023). doi:10.1101/2023.05.01.23289163.

[37] M. Rahimzadeh, A. Attar, S. M. Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset, Biomedical Signal Processing and Control 68 (2021) 102588. doi:10.1016/j.bspc.2021.102588.

[38] Bell, D., Sciacca, F., Campos, A., et al. (2026). COVID-19. Radiopaedia.org. Retrieved July 12, 2026, URL radiopaedia.org/articles/covid-19, doi:10.53347/rID-73913.

[39] Jenssen, H. B., & Sakinis, T. (2021). MedSeg Covid Dataset 2 (Version 2). figshare. Doi:10.6084/m9.figshare.13521509.v2

[40] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, Others, COVID-19 CT lung and infection segmentation dataset, Zenodo, Apr 20 (2020).

[41] J. P. Cohen, P. Morrison, L. Dao, COVID-19 Image Data Collection (mar 2020). arXiv:2003.11597, doi:10.1016/s0140-6736(20) 30211-7.

[42] E Soares, E., Angelov, P., Biaso, S., Froes, M. H., & Abe, D. K. (2020). SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. medRxiv. doi:10.1101/2020.04.24.20078584

[43] Afshar, P., Heidarian, S., Enshaei, N., Naderkhani, F., Rafiee, M. J., Oikonomou, A., ... & Mohammadi, A. (2021). COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. Scientific Data, 8(1), 121. doi:10.1038/ s41597-021-00900-3.

[44] Sun, Z., Zhang, N., Li, Y., & Xu, X. (2020). A systematic review of chest imaging findings in COVID-19. Quantitative imaging in medicine and surgery, 10(5), 1058. doi:10.21037/QIMS-20-564.

[45] G. Bwire, G. M. (2020). Coronavirus: why men are more vulnerable to Covid-19 than women? SN comprehensive clinical medicine, 2(7), 874-876. doi:10.1007/S42399-020-00341-W.

[46] WHO, Tracking SARS-CoV-2 variants.

URL.who.int/activities/tracking-SARS-CoV-2-variants/tracking-SARS-CoV-2-variants

[47] S. Heidarian, P. Afshar, A. Mohammadi, M. J. Rafiee, A. Oikonomou, K. N. Plataniotis, F. Naderkhani, Ct-caps: Feature extraction-based automated framework for covid-19 disease identification from chest ct scans using capsule networks, 2021, pp. 1040–1044.