Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Diagnostic and Interventional Imaging 000 (2022) 1–4

Original article

French Imaging Database Against Coronavirus (FIDAC): A large COVID-19 multi-center chest CT database

Loic Boussel[a,b,*], Jean-Michel Bartoli[c], Samy Adnane[a], Jean-François Meder[d,e], Patrick Malléa[f], Jeremy Clech[f], Marc Zins[g], Jean-Paul Bérégi[h]

[a] Department of Radiology, Croix-Rousse Hospital, Hospices Civils de Lyon, 69004 Lyon, France
[b] Univ Lyon, INSA Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, 69100 Villeurbanne, France
[c] Department of Radiology, Aix-Marseille University, University Hospital La Timone, AP-HM, 13385 Marseille, France
[d] Department of Neuroimaging, Sainte-Anne Hospital, 75013 Paris, France
[e] Université Paris Cité, 75006 Paris, France
[f] NEHS DIGITAL, 92240 Malakoff, France
[g] Department of Radiology, Paris Saint-Joseph Hospitals, 75014 Paris, France
[h] Department of Medical Imaging, University Hospital of Nîmes, 30900 Nîmes, France

ARTICLE INFO

Keywords:
Computed tomography
COVID-19
Database
Multicenter studies

ABSTRACT

Purpose: During the first wave of the COVID-19 pandemic, the French Society of Radiology and the French College of Radiology, in partnership with NEHS Digital, have set up a system to collect chest computed tomography (CT) examinations with clinical, virological and radiological metadata, from patients clinically suspected of COVID-19 pneumonia. This allowed the constitution of an anonymized multicenter database, named FIDAC (French Imaging Database Against Coronavirus). The aim of this report was to describe the content of this public database.

Materials and methods: Twenty-two French radiology centers participated to the data collection. The data collected were chest CT examinations in DICOM format associated with the following metadata: patient age and sex, originating facility identifier, originating facility region, time from symptom onset to CT examination, indication for CT examination, reverse transcription-polymerase chain reaction (RT-PCR) results and normalized CT report performed by a senior radiologist. All the data were anonymized and sent through a NEHS Digital system to a centralized data center.

Results: A total of 5944 patients were included from the 22 centers aggregated into 8 regions with a mean number of patients of 743 ± 603.3 [SD] per region (range: 102–1577 patients). Reasons for CT examination and normalized CT reports were provided for all patients. RT-PCR results were provided in 5574 patients (93.77%) with a positive result of RT-PCR in 44.6% of patients.

Conclusion: The FIDAC project allowed the creation of a large database of chest CT images and metadata available, under conditions, in open access through the CERF-SFR website.

© 2022 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Since the onset of the COVID-19 epidemic, chest CT has proven to be an effective tool for disease diagnosis [1–6], patient referral [7, 8], and disease prognosis [9, 10]. Multiple solutions based on artificial intelligence have been developed to establish the diagnosis and quantify lung involvement in order to estimate the risk for the patient to develop a severe form or to die [5, 11, 12]. However, not all of these solutions have been tested on independent datasets, which hinders the interpretation of their results. Moreover, most of these solutions are not integrated into the clinical radiology workflow, either because they are published by companies that are not yet very present in the radiology landscape, or because they are open source solutions whose code is difficult and costly to deploy in current practice. It is therefore necessary to have validation databases collected from a large panel of imaging centers in order to be able to objectively compare the solutions and guide the user in choosing one solution or another.

In this context, the French Society of Radiology (SFR) and the French College of Radiology (CERF), in partnership with NEHS (Nouvelle Entreprise Humaine en Santé) Digital, have set up a system to

Abbreviations: COVID-19, Coronavirus disease 2019; CT, Computed tomography; CERF, French College of Radiology; DICOM, Digital imaging and communications in medicine; FIDAC, French Imaging Database Against Coronavirus; GGO, Ground-glass opacity; RT-PCR, Reverse transcription polymerase chain reaction; SFR, French Society of Radiology

* Corresponding author.
E-mail address: loic.boussel@chu-lyon.fr (L. Boussel).

https://doi.org/10.1016/j.diii.2022.05.006
2211-5684/© 2022 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

collect chest CT scans, with clinical, virological and radiological metadata, from patients clinically suspected of presenting with COVID-19 pneumonia during the first wave of the epidemic. This allowed the constitution of an anonymized multicenter database, named FIDAC (French Imaging Database Against Coronavirus) available to the scientific community.

The purpose of this report was to describe the implementation and the content of the public FIDAC database.

2. Materials and methods

2.1. Data collection

The announcement of the start of the data collection was made by the CERF and the SFR to all French radiology structures on April 15, 2020. The inclusion criteria were the realization of a chest CT scan in a patient with a suspicion of COVID-19 lung involvement at initial admittance or during the follow-up and/or a reverse transcription polymerase chain reaction (RT-PCR) proven COVID-19 infection and/or radiological signs of COVI-19 on the chest CT. The clinical suspicion was based on the presence of one or a combination of the following signs: fever or chills, cough, shortness of breath or difficulty breathing, fatigue, muscle or body aches, headache, anosmia or ear, nose, throat symptoms. Exclusion criteria were patients under 18 years of age or a chest CT ordered outside a context of COVID-19. Data collection was performed from April 1st to August 5th, 2020.

The data collected were chest CT images in digital imaging and communications in medicine (DICOM) format associated with the following metadata: patient age and sex, originating facility identifier, originating facility region (if more than 100 patients, otherwise "other"), time from symptom onset to CT examination (<3 days, 3–7 days, 8–21 days, and >21 days), indication for CT examination, final RT-PCR test result (positive, negative or not performed) and the normalized CT report. Indication for CT examination was labelled as "low symptoms COVID-19" when the patient was presenting with one or a combination of the clinical signs described above but without the need for oxygen therapy, "suspicion of COVID-19 in a oxygen dependent patient" when the patient was presenting with one or a combination of the clinical signs described above but with the need for oxygen therapy, "follow-up" and "other". Normalized CT report was termed as "classic/probable COVID-19" when presenting with peripheral, bilateral ground-glass opacities (GGO) or multifocal GGO of rounded morphology ± consolidation or crazy paving, reversed halo sign or sub-pleural bands of consolidations. It was classified as "intermediate for COVID-19" when presenting with multifocal, diffuse, peripheral, or unilateral GGO ± consolidation lacking a specific distribution and non-rounded or non-peripheral or with only few very small GGO with a non-rounded and non-peripheral distribution or with atypical findings: large pleural effusion, major lymph nodes size increase or bronchiolitis pattern. It was classified as "non-COVID" when demonstrating another pathology and "normal" when no pulmonary disease was detected by the radiologist. This classification was performed by a senior radiologist from the local center.

The transfer of CT images and metadata was performed either using the Nexus platform (NEHS Digital) with a dedicated workflow adapted to the study for the centers owning a Nexus system or through a dedicated web portal developed by NEHS Digital for the other centers.

This study is compliant with the General Data Protection Regulation (GDPR) and the French regulation authority (Commission Nationale de l'Informatique et des Libertés, CNIL) by applying anonymization before sending the data out of the radiology structures. Indeed, direct identifiers (*e.g.*, name and birthdate) have been removed, unique identifiers (*e.g.*, Patient Id or Image Id) have been randomly regenerated, and some metadata have been generalized (*e.g.* radiology structures have been aggregated in regions). During the data collection, several privacy impact assessments have been performed to evaluate the risk of re-identification by individualization, by correlation or by inference. Those evaluations led to the removal of some additional metadata such as manufacturer and CT model's name. Therefore, all the images and the metadata were anonymized in compliance with the CNIL (MR004). All patients were informed about the study.

2.2. Data structure

In order to ease the use of the database, the metadata were include in the DICOM header of the CT images as private tags (Table 1).

Furthermore, patient's age tag (0 × 10,0 × 1010) was replaced by the age range of the patient.

3. Results

A total of 5944 patients were included from 22 centers (Alès, Aubagne, Clermont Ferrand, Coulommiers, Douai, Fécamps, Gap Sisteron, Hospice Civil de Lyon, Lannion, Lens, Le Mans, Marne la Vallée, Martinique, Meaux, Montpellier, Nice, Nîmes, Poitiers, Rennes, Sens, Strasbourg, Voiron). These centers were aggregated into 8 French regions with a mean number of patients of 743 ± 603 (SD) per region (range: 102–1577 patients). The distribution by region was as follows: Grand Est, 1577 patients (26.53%), Auvergne Rhône Alpes,1216 patients (20.46%), Occitanie, 1412 patients (23.76%), Ile-de France, 900 patients (15.14%), Hauts de France, 284 patients (4.78%), Bretagne, 249 patients (4.19%), Provence Alpes Côtes d'Azur, 102 patients (1.72%) and other regions (including Pays de la Loire, Martinique, Bourgogne-Franche-Comté, Nouvelle-Aquitaine and Normandie), 204 patients (3.43%).

Patients' age categories were distributed as follows (number, percentage of the dataset): 18–20 years (17; 0.29%), 20–30 years (912; 15.35%), 30–40 years (407; 6.85%),40–50 years (587; 9.88%), 50–60 years (853; 14.35%), 60–70 years (1011; 17.01%), 70–80 years (987; 16.61%), 80–90 years (848; 14.27%) and >90 years (321; 5.40%). Median age was 60 years. Patient's sex was provided in 5251 patients (88.34%) with a male/female ratio of 0.55/0.45.

Delay between onset of symptoms and CT examination was provided in 3453 patients (58.09%). It was lower than 3 days in 1128 patients (32.66%), between 3 and 7 days in 916 patients (26.53%), between 8 and 21 days in 1158 patients (33.53%) and larger than 21 days in 251patients (7.27%).

Table 1
List of the private tags and the corresponding metadata.

TAGS	Metadata
(0C19.1001:NEHS-COVID 19)	Region's name
(0C19.1010:NEHS-COVID 19)	Indication for CT examination
(0C19.1011:NEHS-COVID 19)	Time between onset of symptoms and CT examination
(0C19.1012:NEHS-COVID 19)	Normalized CT report
(0C19.1013:NEHS-COVID 19)	RT-PCR result

CT: Computed tomography; RT-PCR: Reverse transcription polymerase chain reaction.

Reason for the CT examination was provided for all the 5944 patients and classified as low symptoms COVID-19 in 2329 patients (39.18%), oxygen dependent patient in 667 patients (11.22%), follow-up in 162 patients (2.73%) and other in 2786 patients (46.87%).

Normalized CT report was provided for all the 5944 patients and classified as "classic/probable COVID-19" in 1931 patients (32.49%), "intermediate for COVID-19" in 1156 patients (19.45%), "non-COVID" in 697 patients (11.73%) and "normal" in 2160 patients (36.33%).

RT-PCR results were provided in 5574 patients (93.77%) with a positive test in 2486 patients (44.6%) and a negative test in 3088 patients (55.4%).

4. Discussion

The FIDAC project allowed creating a French multicenter anonymized database of chest CT scans performed in patients with suspected or confirmed COVID-19. Access to this public database can be performed by request to the SFR on the CERF-SFR website (https://cerf.radiologie.fr/contact). The access rights to the database are conditioned by the non-commercialization of the data, the non-resale of the data and their use for research and the development of community solutions.

Beside open access, the FIDAC database has several advantages that make it suitable for training and validation of the algorithms. First, the amount of CT dataset collected is large (5944 patients) with 2996 patients presenting with clinical signs of COVID-19 and 2486 with a positive COVID-19 RT-PCR test. Then, since the dataset was collected throughout France, it includes CT from several vendors and with different acquisition protocols. These two elements are particularly important since the training and the validation of deep learning algorithms necessitate a large amount of data representative of the studied population [13–17]. Furthermore, the rate of completion of the metadata of major interest, including RT-PCR results, patients' age categories and normalized CT report is high, allowing to apply these labels on a large proportion of the CT database.

Another important feature of the FIDAC database is the use of the 16-bits DICOM format for the CT images. This make the review of the images easier for radiologists as DICOM is a standard for storage of medical images and many viewers are available to display these images [18]. Furthermore, this 16-bits format allows preserving the dynamic of the images as the values of the pixels range from minus one thousand to several thousand Hounsfield Units. Finally, the metadata are included in the DICOM header in order to make the database more compact and to avoid consistency errors.

This large collection of data was made possible, in a very short time, by the pre-existing interconnection network of radiology departments based on the Nexus solution (NEHS Digital). For centers that did not have Nexus at the time of the study, NEHS Digital set up a web interface to upload files. Although this latter system was efficient, the transfer was greatly simplified by the use of Nexus, which underlines the interest of having such interconnection systems for multicenter imaging studies on a national scale.

The FIDAC database has several limitation. First some of the metadata are unbalanced. Indeed, there is an over representation of the regions of the north and east part of France that faced a larger number of COVID-19 patients during the first month of the pandemic [19]. Furthermore, older people are overrepresented in the dataset. Indeed, younger people were less likely to develop a severe form of COVID-19 and thus to necessitate a CT examination [20–22]. Another limit is the relatively low level of completion of some of the metadata including the delay between the onset of symptoms. This is related to the lack of information when the CT was performed in aged patient in an emergency setting. Finally, data collection was not carried out consecutively in each center and no biological data was recorded. This limits the ability to perform advanced statistics on the population included and to draw conclusions on the effectiveness of the scanner for the diagnosis of covid-19.

In conclusion, the FIDAC project, developed by the French Society of Radiology and the French College of Radiology allowed the creation of a large database of chest CT images and metadata throughout France during the first wave of the COVID-19 pandemic. This database is available, under conditions, in open access through the CERF-SFR website.

Ethical statement

The FIDAC study is compliant with the GDPR and the French regulation authority (CNIL).

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

The authors declare that this report does not contain any personal information that could lead to the identification.

Disclosure of interest

Patrick Malléa and Jeremy Clech are NEHS Digital employees.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Loic Boussel: Writing- Original draft preparation, Writing- Reviewing and Editing, Final draft approval.

Jean-Michel Bartoli: Writing- Original draft preparation, Final draft approval.

Samy Adnane: Writing- Original draft preparation, Final draft approval.

Jean-François Meder: Writing- Original draft preparation, Final draft approval.

Patrick Malléa: Writing- Original draft preparation, Methodology, Final draft approval.

Jeremy Clech: Writing- Original draft preparation, Methodology, Final draft approval.

Marc Zins: Writing- Original draft preparation, Final draft approval.

Jean-Paul Bérégi: Conceptualization, Writing- Original draft preparation, Writing- Reviewing and Editing, Final draft approval.

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

All the authors had fully participated to the study and approved the final draft.

Acknowledgments

We acknowledge Alexandre Fernandez and Brieuc Tredan from NEHS Digital for their help in data collection.

We acknowledge, for their participation in the creation of the database, the radiologists from the following centers: CHU de Strasbourg, CHU de Nîmes, CHU de Clermont-Ferrand, Hospices Civils de Lyon, CHU de Montpellier, CH de Marne La Vallée, CHU de Rennes, CH de Meaux, CH de Lens, CH de Voiron, Centre hospitalier Intercommunal des Alpes du Sud - Gap et Sisteron, CH de Coulommiers, CH de Valenciennes, CH de Lannion, CH de Douai, CH Le Mans, CHU de Poitiers, CHU de Nice, CH de Sens, CH de Aubagne, CHU de Fort de France, CH de Fecamp.

References

[1] Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America Expert Consensus Document on reporting chest CT findings related to COVID-19: endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiol Cardiothorac Imaging 2020;2:e200152.
[2] Hani C, Trieu NH, Saab I, Dangeard S, Bennani S, Chassagnon G, et al. COVID-19 pneumonia: a review of typical CT findings and differential diagnosis. Diagn Interv Imaging 2020;101:263–8.
[3] Li J, Long X, Wang X, Fang F, Lv X, Zhang D, et al. Radiology indispensable for tracking COVID-19. Diagn Interv Imaging 2021;102:69–75.
[4] Weisberg EM, Chu LC, Rowe SP, Fishman EK. Radiology, COVID-19, and the next pandemic. Diagn Interv Imaging 2021;102:583–5.
[5] Kato S, Ishiwata Y, Aoki R, Iwasawa T, Hagiwara E, Ogura T, et al. Imaging of COVID-19: an update of current evidences. Diagn Interv Imaging 2021;102:493–500.
[6] Chassagnon G, Regard L, Soyer P, Revel M-P. COVID-19 after 18 months: where do we stand? Diagn Interv Imaging 2021;102:491–2.
[7] Herpe G, Lederlin M, Naudin M, Ohana M, Chaumoitre K, Gregory J, et al. Efficacy of chest CT for COVID-19 pneumonia diagnosis in France. Radiology 2021;298:E81–7.
[8] Ducray V, Vlachomitrou AS, Bouscambert-Duchamp M, Si-Mohamed S, Gouttard S, et al. Chest CT for rapid triage of patients in multiple emergency departments during COVID-19 epidemic: experience report from a large French university hospital. Eur Radiol 2021;31:795–803.
[9] Feng Z, Yu Q, Yao S, Luo L, Zhou W, Mao X, et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. Nat Commun 2020;11:4968.
[10] Devie A, Kanagaratnam L, Perotin J-M, Jolly D, Ravey J-N, Djelouah M, et al. COVID-19: a qualitative chest CT model to identify severe form of the disease. Diagn Interv Imaging 2021;102:77–84.
[11] Lassau N, Ammari S, Chouzenoux E, Gortais H, Herent P, Devilder M, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. Nat Commun 2021;12:634.
[12] Jalaber C, Lapotre T, Morcet-Delattre T, Ribet F, Jouneau S, Lederlin M. Chest CT in COVID-19 pneumonia: a review of current knowledge. Diagn Interv Imaging 2020;101:431–7.
[13] Chen J, See KC. Artificial intelligence for COVID-19: rapid review. J Med Internet Res 2020;22:e21476.
[14] Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020;12:165–74.
[15] Al-Najjar H, Al-Rousan N. A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea. Eur Rev Med Pharmacol Sci 2020;24:3400–3.
[16] Tsai EB, Simpson S, Lungren MP, Hershman M, Roshkovan L, Colak E, et al. The RSNA international COVID-19 open radiology database (RICORD). Radiology 2021;299:E204–13.
[17] Revel M-P. COVID-19 pneumonia: the fight must go on. Diagn Interv Imaging 2021;102:61–2.
[18] Haak D, Page CE, Deserno TM. A survey of DICOM viewer software to integrate clinical research and medical imaging. J Digit Imaging 2016;29:206–15.
[19] Lescure F-X, Bouadma L, Nguyen D, Parisey M, Wicky P-H, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis 2020;20:697–706.
[20] Wang Y, Zhou Y, Yang Z, Xia D, Hu Y, Geng S. Clinical characteristics of patients with severe pneumonia caused by the SARS-CoV-2 in Wuhan, China. Respir Int Rev Thorac Dis 2020;99:649–57.
[21] Zhang J-J, Dong X, Cao Y-Y, Yuan Y-D, Yang Y-B, Yan Y-Q, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020;75:1730–41.
[22] Gao Y, Ding M, Dong X, Zhang J, Azkur AK, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: a review. Allergy 2021;76:428–55.