

Systematic Review of Advanced AI Methods for Improving Healthcare Data Quality in Post COVID-19 Era

Monica Isgut^{1b}, Student Member, IEEE, Logan Gloster, Student Member, IEEE, Katherine Choi, Janani Venugopalan, and May D. Wang^{2b}, Fellow, IEEE

(Methodological Review)

Abstract—At the beginning of the COVID-19 pandemic, there was significant hype about the potential impact of artificial intelligence (AI) tools in combatting COVID-19 on diagnosis, prognosis, or surveillance. However, AI tools have not yet been widely successful. One of the key reasons is the COVID-19 pandemic has demanded faster real-time development of AI-driven clinical and health support tools, including rapid data collection, algorithm development, validation, and deployment. However, there was not enough time for proper data quality control. Learning from the hard lessons in COVID-19, we summarize the important health data quality challenges during COVID-19 pandemic such as lack of data standardization, missing data, tabulation errors, and noise and artifact. Then we conduct a systematic investigation of computational methods that address these issues, including emerging novel advanced AI data quality control methods that achieve better data quality outcomes and, in some cases, simplify or automate the data cleaning process. We hope this article can assist healthcare community to improve health data quality going forward with novel AI development.

Index Terms—Data quality, COVID-19 pandemic, data harmonization, missing data, tabular errors, noise and artifact, deep learning, quality control, artificial intelligence algorithm development.

Manuscript received 8 February 2022; revised 27 April 2022 and 7 July 2022; accepted 22 September 2022. Date of publication 21 October 2022; date of current version 6 January 2023. This work was supported by the Georgia Tech Bioinformatics and COVID-19 Disruption Graduate Research Awards. (Corresponding author: May D. Wang.)

Monica Isgut and Logan Gloster are with the Department of Bioinformatics, School of Biology, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: misgut@gatech.edu; lgloster@gatech.edu).

Katherine Choi is with the School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30322 USA (e-mail: kchoi99@gatech.edu).

Janani Venugopalan and May D. Wang are with the Wallace H. Coulter School of Biomedical Engineering, Georgia Institute of Technology, Emory University, Atlanta, GA 30322 USA (e-mail: janvenu@gmail.com; maywang@gatech.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/RBME.2022.3216531>, provided by the authors.

Digital Object Identifier 10.1109/RBME.2022.3216531

I. INTRODUCTION

THE Coronavirus Disease 2019 (COVID-19) pandemic has caused significant morbidity and mortality around the world, which prompted artificial intelligence (AI) research to develop software tools to help combat the disease. However, most of these tools have been largely unsuccessful due to the limited availability of high-quality, large-scale, and timely data to investigators [1], [2], [3]. Although large quantities of COVID-19 health data have been collected in real-time since the beginning of the pandemic, data collection was generally siloed and many public repositories contained unreliable or unharmonized datasets [2]. AI tools developed using dirty data can be biased and fail upon validation or deployment (Fig. 1, Suppl. Fig. 1), leading to an opportunity cost in terms of the possibility for enhanced technology-driven COVID-19 surveillance, triaging, diagnosis, and prognosis. In this review, we expand on our previous data quality control reviews [5], [6], [7], [8], [9], [10], our past research on data quality [5], [6], [11], [12], [13], [14], [15] and our previous work discussing state-of-the-art AI methods contextualized in the COVID-19 pandemic [4]. We discuss four common data quality issues: a) lack of data standardization of data from different sources, b) missing data, c) tabulation errors, and d) noise and artifacts. For each data quality issue, we describe how this issue is relevant to COVID-19 health data and provide a systematic review of advanced quality control tools that address the issue with superior performance compared to classical methods. These four categories span several large, national datasets collected for the purpose of COVID-19 public health surveillance that present unique challenges in approaching quality control depending on the data modality and its potential source of error. Accordingly, in each section we discuss the classical and advanced approaches particular to the most common data modalities relevant to the main quality control topic. Other critical aspects of health data quality control for machine learning include such topics as data safety, class imbalance, data duplication and many more general best practices for handling data. To stay within the scope of our literature search, however, our discussion will center on the four main categories and data modalities introduced above. A select non-exhaustive set of available COVID-19 health datasets is

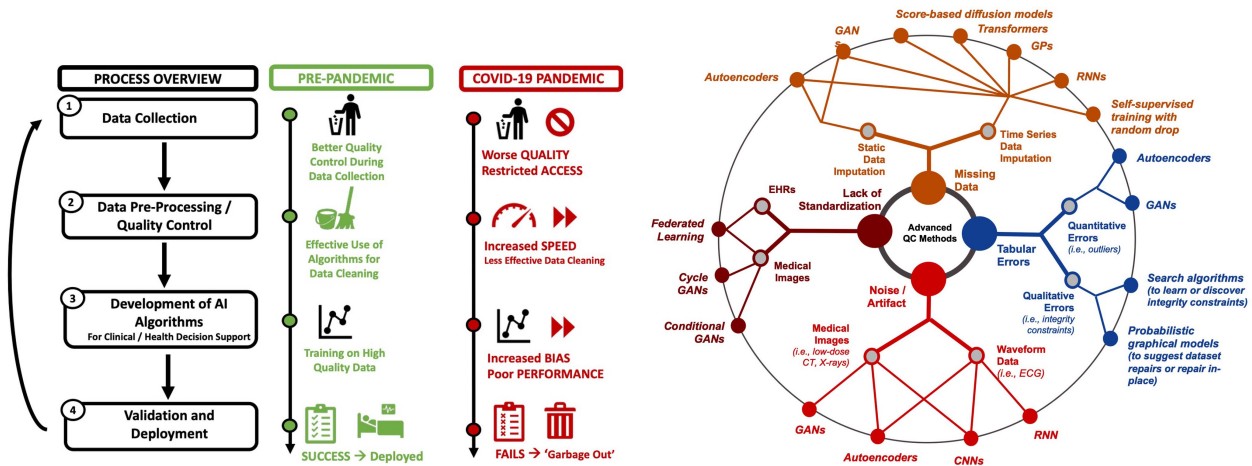


Fig. 1. (Left) Impact of the COVID-19 Pandemic on the Development Cycle of AI-Based Health Informatics Tools. Data collected during the pandemic were of generally poor quality with high-quality data being restricted to select research institutions. The high demand for rapid machine learning algorithm development often conflicted with more time-consuming quality control standards, resulting in poor model performance on real-world data and overall slower AI tool deployment. More details in Suppl. Fig. 1. (Right) Common Deep Learning-Based Concepts for Advanced Data Quality Control Tools. This figure illustrates the four major concepts relevant to different advanced quality control methods. Each concept branches into the data modalities that are commonly affected followed by specific approaches used to address the main quality control issue. Abbreviations: GAN = generative adversarial network, CNN = convolutional neural network, RNN = recurrent neural network, GP = Gaussian process, ECG = electrocardiogram, CT = computed tomography, EHRs = electronic health records.

TABLE I
RELEVANT DATA QUALITY ISSUES FOR COMMON COVID-19 HEALTH DATA SOURCES AND EXAMPLES OF AVAILABLE DATASETS

Data Source	Missing Data	Noise/Artifact	Tabulation	Lack of Standardization	Dataset Examples
Wearables Data	Yes	Yes	Yes (waveform)	N/A	DETECT [16], Stanford Wearables Study [17], TemPredict [18], [19], Achievement App [20]
Medical Image Data	Yes	Yes	N/A	Yes	National COVID-19 Chest Image Database [21]
EHR Data	Yes	N/A	Yes	Yes	UK Biobank [22], N3C [23], NHS-TRE [22]
Audio Data	Yes	Yes	N/A	Yes	COUGHVID [24], Coswara [25], NoCoCoDa [26], Cambridge [27]

described in Table I (short version), with a more detailed version available in Suppl. Table I.

II. PRISMA FRAMEWORK

We used the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) framework (Fig. 2) to systematically identify and review original research articles presenting novel deep learning-based approaches towards data quality control, specifically focusing on the following data quality issues: missing data, noise and artifact, tabulation errors, and lack of data standardization. The systematic database search included Scopus, PubMed, and the Institute for Electrical and Electronics Engineers (IEEE). The search was restricted to articles published on or after 2018 that contained the terms listed in Suppl. Table II. The search included terms listed in a) the abstract, title, or article keywords for Scopus, b) the abstract or title only for PubMed, or c) the abstract only for IEEE. All 3,345 articles found across all three databases for the different search terms were saved into a table. An additional set of 239 articles were found using approaches that did not involve a database search, such as ancestry and descentance approaches. The total number of articles after combining these was 3,584. After removing any duplicated entries, the total number of articles was

2,517. The first round of screening involved screening by title. Exclusion criteria included review papers, articles that were not written in English, and articles not related to deep learning or data quality control. After the first round of screening, a total of 2,199 articles were excluded and 318 were included. Further detailed inclusion and exclusion criteria are described in Suppl. Table III.

III. QUALITY CONTROL FOR LACK OF DATA STANDARDIZATION

Data standardization refers to harmonizing data from different sources into a cohesive dataset for use in downstream analyses. Lack of standardization is a major challenge for EHR data because different healthcare systems and clinics can use different EHR software with different data formats and data field. A key example of the lack of consistent COVID-19 EHR data across institutions is illustrated by the work of the Consortium for Clinical Characterization of COVID-19 by EHR (4CE) [28], which is one of hundreds of efforts to aggregate COVID-19 data. The researchers collected aggregated data from 100 hospitals in several different countries, and found substantial differences between the datasets, such as variations in units for certain data fields, different code systems used for laboratory tests and diagnoses,

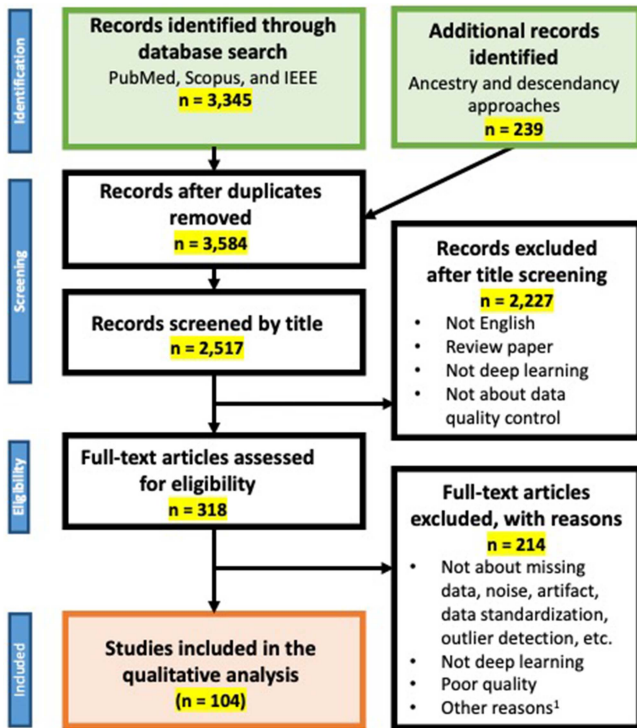


Fig. 2. PRISMA Framework. ¹Inclusion and exclusion criteria are described in more detail in Suppl. Table III.

and the lack of meta data on hospital specific reference ranges. Lack of standardization is also a significant challenge with medical image data, and this has been especially problematic for COVID-19 related research. Publicly available databases were set up that allowed anyone to upload COVID-19 chest X-ray or CT images, and many of the uploaded datasets contained images from heterogeneous sources [2]. Images produced in different healthcare institutions might use different scanning machines, settings, or protocols that influence the image, making it possible for a machine learning algorithm to learn a biased or incorrect classification rule from the data. A summary of how the issue of lack of data standardization is relevant for various types of COVID-19 health data is presented in Table I (short version) Suppl. Table IV (detailed).

The following subsections provide a detailed analysis of classical and advanced approaches to manage standardization issues, with a focus on EHR and medical image data. Specific examples of comparisons between classical and advanced methods, along with an illustration of the issue of lack of standardization, are detailed in Suppl. Fig. 2. Advanced methods are further summarized in Suppl. Table V and Fig. 1 (Right), with a visual timeline presented in Fig. 3. The issue of lack of standardization is further visualized in Fig. 4.

A. Classical Approaches for Handling Unstandardized Data

EHR data standardization challenges are in part being overcome using standards such as HL7's Fast Healthcare Interoperability Resources (FHIR) [29] and SMART-on-FHIR [30]

that harmonize EHR data from different sources to facilitate data sharing. For researchers aiming to harmonize distinct EHR datasets that are not already harmonized through a unified data sharing framework such as FHIR, significant pre-processing needs to be done to transform the data into standardized formats. This can be a generally time-consuming and manual task, requiring syntactic and semantic data transformations. Syntactic data transformations consist of transforming the data from one format to another, such as changing a data table from a long and narrow format to a short and wide format [31], as exemplified in Suppl. Fig. 3. These transformations typically require the user to specify the total scope of the operations allowed in a table, such as splitting or merging columns, or string manipulations. On the other hand, semantic data transformations require the need for external information such as a mapping from ICD-10 codes to disease names for diagnoses, as shown in Suppl. Fig. 4.

For medical image standardization, the Digital Imaging and Communications in Medicine (DICOM) Standard [32] is a framework developed to allow for easy image storage and exchange for medical images from diverse vendors. The DICOM Grayscale Standard Display Function (GSDF) has been shown to increase visual consistency across medical images, but only improves the luminance response, which is just one of many factors that influence the quality of a medical image, including reflection, spatial resolution, noise, geometrical distortions, display chromaticity, veiling glare, and temporal response [33]. Thus, DICOM GSDF is useful for medical image standardization but not the most extensive solution. Another classical approach is using histogram matching [34], [35], which involves transferring a source image into a target domain by setting up an intensity histogram for each image domain and then matching the histograms between two images. This has been used to standardize the luminance value and saturation distributions for different images, but has sub-optimal performance, does not consider all factors that influence image quality, and can sometimes produce artifacts [34], [35].

B. Advanced Approaches for Handling Unstandardized Data

Recent advances have facilitated use of datasets from diverse sources for jointly training machine learning models, without the need to share data, hence having the potential to save researchers significant time and effort that would otherwise be spent on manual data harmonization. In 2017, Google published a blog post introducing an approach called Federated Learning (FL) [36], which allows for a centralized machine learning model to be jointly trained across several distributed clients without the need for data sharing. Because data sharing is avoided, each healthcare institution (i.e., client) utilizes its own siloed dataset for training the joint model, and thus an added advantage is that there is no need to manually combine and harmonize the healthcare data from each independent institution into a standardized dataset. While FL precludes the need to manually combine and harmonize data from different clients for joint model training, an important consideration is that the distribution of data between

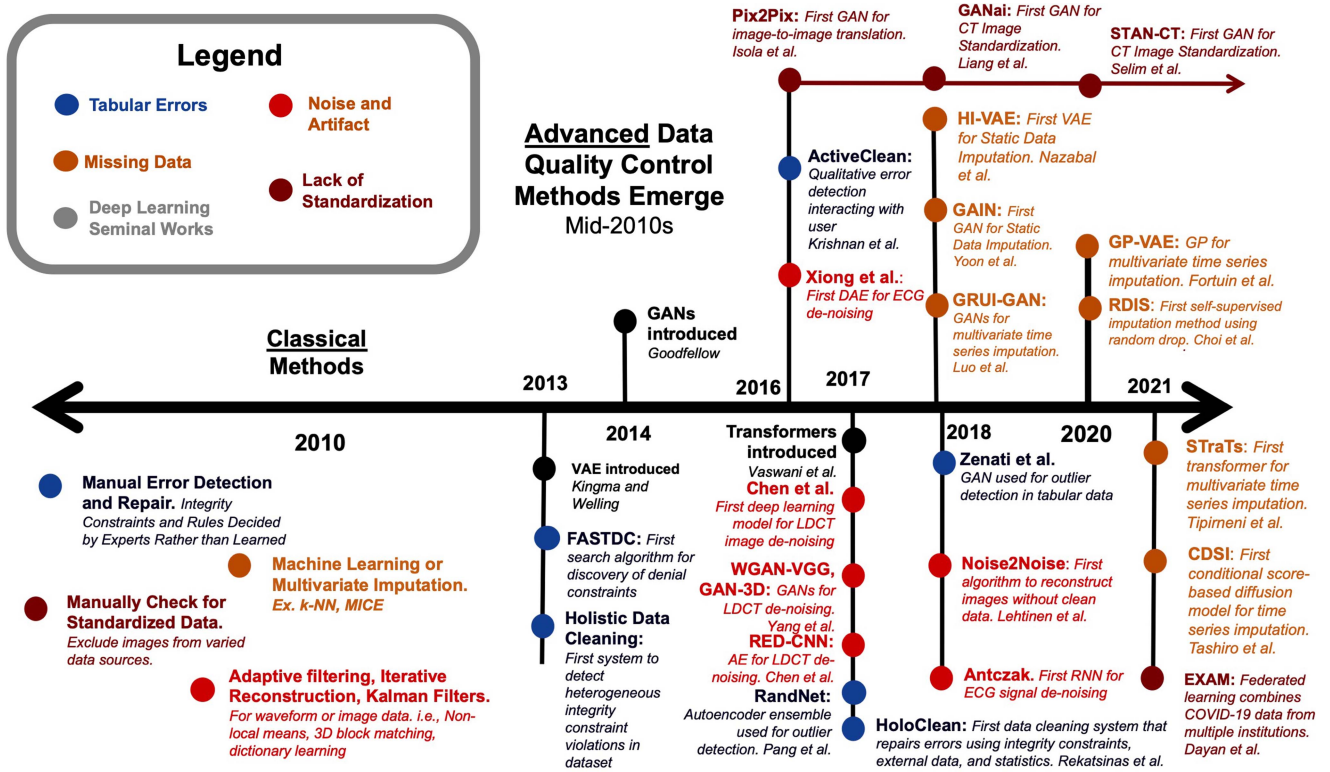


Fig. 3. A brief timeline of major advancements in state-of-the-art methods for data quality control. Abbreviations: VAE = variational autoencoder, GAN = generative adversarial network, CNN = convolutional neural network, RNN = recurrent neural network, GP = Gaussian process, DAE = denoising autoencoder, LDCT = low dose computed tomography, ECG = electrocardiogram, k-NN = k-Nearest Neighbor, MICE = multiple imputation by chained equations, GRU = gated recurrent unit, Wasserstein Generative Adversarial Network – Visual Geometry Group, RED-CNN = residual encoder decoder convolutional neural network, HI-VAE = heterogeneous incomplete variational autoencoder.

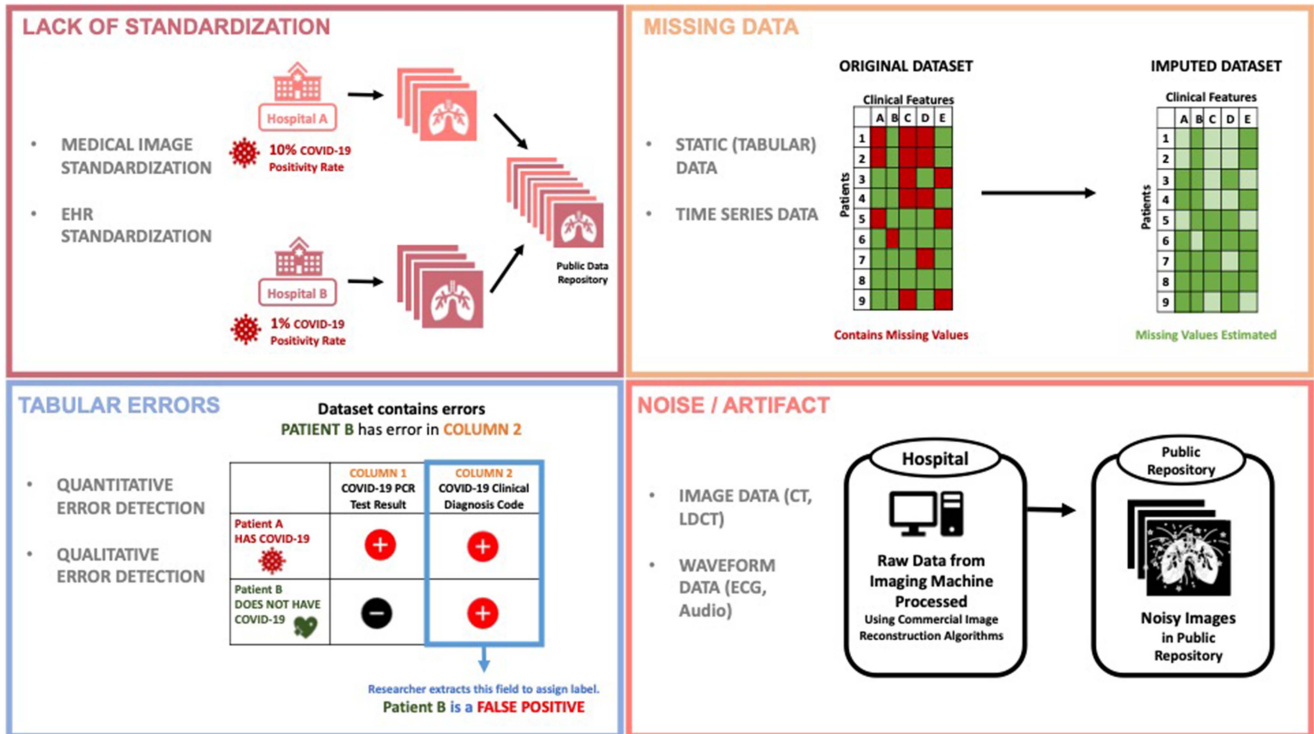


Fig. 4. Visualization of data quality issues discussed in the paper. More details, including examples of classical and advanced quality control methods and comparisons, are available in Suppl. Figs. 2, 5, 6 and 7.

clients is generally statistically heterogeneous and thus non-I.I.D, posing challenges for model convergence [37]. However, methods have recently been developed to address these issues in FL whilst continuing to avoid the need for data sharing [37], [38], [39].

Recently, FL has been successfully applied for training a COVID-19 prognosis clinical decision support algorithm using chest X-ray and electronic health records from a large number of healthcare institutions. A FL multi-modal neural network model called EXAM (Electronic medical record chest X-ray Artificial intelligence Model) [40] was trained to predict the future oxygen requirements and 24 or 72-hour prognosis of COVID-19 patients using data from 20 different healthcare institutions without the need for data sharing or harmonization. Model training consisted of several “rounds”, each of which required training locally using each institution’s private data and servers for one epoch, and then sending the updated model parameters for all the local models back to a centralized server to be aggregated. The FL approach enabled use of large quantities of data, allowing EXAM to generalize and achieve a 16% average improvement in AUC over the models trained individually at each site. It serves as an example of how machine learning models can be used to jointly train machine learning models without the need to share data, thus precluding the need for the time- and effort-intensive process of manually combining and harmonizing the data. While FL is not itself a tool for data harmonization, it enables machine learning models to be trained across large numbers of institutions with private datasets, while precluding the need to harmonize the data to begin with.

Issues with medical image standardization can be addressed using novel image-to-image translation approaches. This field was launched with a seminal work by Isola et al. published in 2016 (Pix2Pix [41]), which uses a conditional generative adversarial network (cGAN) neural architecture to learn a loss function for a mapping from an input image in one domain to an output image in another domain. The model allows input black and white images to be converted into output color images, or input daylight photos to be transferred into the equivalent night-time photo, for example. A standard GAN comprises a generator subnetwork that generates samples from random input data and a discriminator that takes as input both real and synthesized samples and classifies whether a given sample is real. Both the generator and discriminator have contrasting objectives, where the goal of the generator is to produce samples that the discriminator incorrectly classifies as being real. To better adapt standard GANs for image-to-image translation tasks, cGANs use labeled (i.e., paired) training examples, whereby the label image is the version of the input image in the desired domain and the generator learns to produce a new image conditioned on the label image. While cGAN was one of the first models designed for image translation tasks, it is limited by the need for paired training examples, which are not generally readily available in medical image datasets. After the introduction of Pix2Pix, many works in medical imaging followed suit with GAN-based algorithms for image-to-image translation. For example, Wolterink et al. [42] and Emami et al. [43] used GANs to generate synthetic computed tomography (CT) images from

input magnetic resonance (MR) images. Recent works have also used GANs to generate MR [44] or CT [45], [46] images from positron emission tomography (PET) images.

A novel GAN architecture called CycleGAN was introduced in 2017 by Zhu et al. [47] and allows for image-to-image translation without the need for paired training examples. Unlike previous approaches, which might require training samples to be images of the same patient to be taken using MR and CT imaging techniques (i.e., paired samples), for example, CycleGAN uses information about the distribution of images from the two different domains, thus precluding the need for the paired samples. Specifically, to train this model, unpaired training example images from both the source domain and the desired domain are needed. There are two generators, one of which generates synthetic samples of one domain using input samples from the other domain and the other which reconstructs the input from the synthetic samples. In addition to the standard GAN adversarial loss, CycleGAN also uses two cycle consistency loss terms, each of which minimizes the reconstruction loss of translating each synthetic image back to its original source domain. This loss term enables image-to-image translation without the need for paired samples. Recent works in medical image-to-image translation are based on this architecture, such as CyTran [48] and the work by Wolterink et al. [42].

While many of these algorithms have shown success in translating medical images from one image modality to another, GANs (especially CycleGANs) have the potential to also be used to standardize images from different hospital institutions or scanning machines. Even for two datasets with the same type of image (i.e., CT scan), there might be variations in the parameters used by the technician or the manufacturer of the machine. CyTran was developed to generate contrast CT images from non-contrast images, which can be useful for the standardization of images from different sources taken with these different settings. Translating between contrast and non-contrast CT images is a particularly challenging task because it requires a model to effectively recognize tissue types, organ structures, and/or tumors, which can have different radio density measurements between the types of images. To address this challenge, CyTran combines a CycleGAN framework with a convolutional transformer block to generate images, enabling the model to simultaneously recognize large-scale global structural aspects of the images while translating them to the desired CT contrast style. When tested for style transfer with varying pairs of contrast phases (i.e., native to venous, native to arterial), CyTran achieved a consistently higher structural similarity index measure (SSIM) compared to Pix2Pix and CycleGAN, indicating an improved ability to retain structural information between contrast phases. It also achieved a lower RMSE and achieved the best overall subjective evaluation by physicians in terms of translated image quality. While this method was specifically focused on CT contrast style transfer, it may be a promising approach for broader CT image standardization tasks. Recently, STAN-CT [49] was developed as a novel end-to-end framework for CT image standardization. STAN-CT uses both a GAN framework and a DICOM synthesis framework. DICOM CT images are used as input and are then translated into a standardized distribution of

CT image patches using the GAN, after which the standardized image patches undergo basic quality control, integrated back into the full image, and saved in the DICOM format. This end-to-end solution can significantly assist with CT image standardization for the purposes of COVID-19 research. It builds on a previous method (GANai) developed by Liang et al. in 2018 [35]. Both methods have improved standardization for image texture features such as contrast, correlation, homogeneity, energy, and correlation, compared to histogram matching, and once pre-trained, can be useful tools for researchers. Another work by Zunair et al. [50] directly addressed the issue of limited publicly available COVID-19 chest X-ray data by developing a synthetic chest X-ray image dataset using CycleGAN which translates non-COVID-19 patient images to COVID-19 patient images. This approach can potentially be useful when the available COVID-19 image data are heterogeneous and from diverse or unknown original sources. Overall, deep GAN-based image-to-image translation methods are allowing for improved and more versatile CT image standardization than was previously possible. Whereas classical approaches such as Digital Imaging and Communications in Medicine (DICOM) Grayscale Standard Display Function (GSDF) and histogram matching can standardize certain aspects of images such as luminance values, GAN-based methods have dramatically broadened the scope of what is possible for image standardization, allowing for image translation between different medical image styles, CT contrast styles, and more. This can help address some of the most prevalent data quality issues in COVID-19 AI research. Similarly, FL is a breakthrough approach that enables broad, large-scale AI tool development and validation without the need for manual data harmonization, which has been a major roadblock during the pandemic.

IV. QUALITY CONTROL FOR MISSING DATA

The problem of missing data is a significant data quality issue with COVID-19 health data from various sources. For example, a recent study on the quality of a dataset containing EHRs and COVID-19 test results for thousands of patients in Portugal (“SINAVE-Med”) found over 90% of the data were missing for important features such as the date of patients’ first positive laboratory test results or the indication of whether or not a positive case required intensive care unit (ICU) admission [51]. Another study on the use of wearable devices for COVID-19 research found that many patients stop wearing their devices or let the charge expire during the time when they are symptomatic [17], and, despite its prevalence in multiple large datasets, potential audio clipping in crowdsourced COVID-19 cough data can impact the reliability of time-frequency representations in discriminative neural networks and lead to poor diagnostic performance as well [52]. A summary of how the issue of missing data is relevant for various types of COVID-19 health data is presented in Table I (short version) and Suppl. Table IV (detailed).

The following subsections describe classical and advanced approaches towards handling missing data, with a focus on

static (i.e., tabular) and time series data, including a few methods specific to audio waveform data. Examples of classical and advanced methods and their comparisons in performance are illustrated in Suppl. Fig. 5. Advanced methods are further summarized in Suppl. Table V and Fig. 1 (Right), with a visual timeline presented in Fig. 3. The issue of missing data is further visualized in Fig. 4.

A. Classical Approaches for Handling Missing Data

Missing data can be addressed using complete case analysis or imputation. Complete case analysis, also known as case deletion, involves deleting samples that contain incomplete data across all or most features of interest, or deleting features with incomplete data across all or most samples. Although it is a simple and common approach, it is generally not recommended, especially if there is a high quantity of missing data, because removing samples can lead to a loss of statistical power. Imputation involves entering an estimate for an unknown missing value. The simplest imputation method involves imputing missing values for a given feature as the mean, median, or mode of the non-missing data for that feature or as a constant value. However, simple imputation can change the distribution of the data and bias downstream analyses. There are several common imputation algorithms developed for time series datasets, the simplest of which involves imputing the missing values as the value of the known observation that occurred either immediately before or immediately after the missing values (also described as “last observation carried forward; LOCF” or “next observation carried backward; NOCB” [53]). More sophisticated classical multivariate static and time series data imputation frameworks are summarized in Suppl. Tables VI and VII, respectively.

B. Advanced Approaches for Handling Missing Data (Static Data)

Deep learning-based methods have recently been developed that achieve state-of-the-art imputation performance for multivariate static and time series data. For static tabular data imputation, successful models have recently been developed using generative adversarial networks (GANs) [54] and autoencoders. In 2018, Generative Adversarial Imputation Nets (GAIN) was published by Yoon et al. as the first GAN-based algorithm for missing static data imputation [55]. The model uses a generator that imputes the missing values for each sample, and a discriminator that is trained to identify which values for each sample are known observed values and which are imputed by the generator. The model is trained until the discriminator can no longer differentiate between known and imputed values. Specifically, the user inputs the original data matrix (samples by features), where missing values are set to zero. Using this matrix, two other matrices are created - a random matrix, which sets all non-missing values to zero and sets each missing value to a random number, and a mask matrix, which sets all non-missing values to one and missing values to zero. The generator uses all three to produce an imputed samples by features matrix. The model uses another generator subnetwork (“hint generator”) that produces an encoded “hint” matrix based on just the mask

matrix, enabling the discriminator to utilize information about the pattern of missingness in the original dataset. Finally, in contrast to a standard GAN discriminator which classifies an entire sample as real or synthetic, using cross-entropy loss to distinguish between real and synthesized values for each sample. The outcome is an adversarial model that imputes missing values until they can no longer be distinguished from known values in a dataset. When tested on several datasets, GAIN achieved significantly improved imputation performance compared to classical methods such as multivariate imputation by chained equations (MICE) and expectation maximization. When datasets were imputed using various methods and then used as input into a logistic regression algorithm, the classification performance was highest when GAIN was used as the imputation method. In 2019, MisGAN [56], which also uses GANs for imputation, was introduced. Using the Fréchet Inception Distance (FID) as an evaluation metric, MisGAN was found to consistently outperform GAIN across various datasets, especially at higher missing rates over 70%. However, the authors tested MisGAN with only image datasets, and its comparative performance to GAIN for tabular data may differ. Both algorithms are applicable to data missing completely at random (MCAR), which means that there is no relationship between the features or variables in the data and the missing values. In contrast, Multiple Imputation using Denoising Autoencoders (MIDA) was introduced as one of the first autoencoder-based methods for imputation and does not rely on the MCAR assumption, allowing for broader applicability [57]. Denoising autoencoders (DAEs) are similar to standard autoencoders in that input data is reconstructed through a series of encoding and decoding hidden layers, however they differ in that the input is corrupted (i.e., by setting values randomly to zero, adding noise, etc.). MIDA is structured such that the encoding and decoding layers are of sequentially higher dimensions than the input data, enabling better imputation performance. The original dataset is input into the model with missing values initially imputed as the univariate mean (for numerical data) or mode (for categorical data) of each feature column, and the output is the fully imputed representation. When various datasets with data not MCAR were imputed using both MIDA and MICE and then used for downstream classification or regression tasks, model performance was generally better when using MIDA for imputation. However, MIDA was not compared to other state-of-the-art deep learning imputation methods, so its comparative performance to GAIN or MisGAN is unclear. Another disadvantage of GAIN is that it was designed for continuous and binary data types but not for other data such as mixed numerical (i.e., continuous real-value, discrete count data) or nominal (i.e., categorical and ordinal). Nazabal et al. published their work on HI-VAE [58] in 2018 as a variational autoencoder-based imputation framework which can be applied for a broader set of data types under the MCAR assumption and is particularly suitable for datasets with nominal variables. Since then, novel autoencoder-based static imputation methods continue to be introduced [59], [60], [61], [62]. A few of these static deep learning-based imputation methods are further summarized in Suppl. Table VIII.

C. Advanced Approaches for Handling Missing Data (Time Series Data)

The development of multivariate time series imputation algorithms using deep learning is a larger and fast-growing field of research and is centered on the models developed using concepts from recurrent neural networks (RNNs), Gaussian processes (GPs), variational autoencoders (VAEs [63]), and GANs (Suppl. Table IX). While RNNs have been developed for handling missing time series data since the late 1990 s and early 2000 s, more recently in 2016, a model based on gated recurrent units (GRUs) called GRU-D [64] was developed as an end-to-end RNN framework for classification tasks using incomplete input time series data. GRU-D uses information about which values are observed or missing and patterns of missingness over time to improve classification prediction. Specifically, the model comprises a modified GRU structure that contains two trainable decay mechanisms (input and hidden decay) and that uses a mask vector for each sample for each time point that zeroes out missing values. The input decay uses information about the time interval during which an input feature is missing to decay the last observed value towards the mean, and uses these decayed values to impute missing data points. The hidden decay term is applied to the hidden state of the previous time point. Together, these modifications result in more fine-tuned imputation than traditional approaches and enable the RNN to be trained end-to-end with missing data. GRU-D was shown to have state-of-the-art performance compared to classical baseline methods when applied to the MIMIC- III intensive care unit and PhysioNet datasets. For example, when GRU-D was used for classification with MIMIC-III data, model area under the curve (AUC) increased by nearly 3% compared to when the dataset was imputed using LOCF and then trained using a standard GRU-based neural network. Another work by Yoon et al., M-RNN [65], uses a neural network architecture comprising bidirectional RNN layers that interpolate missing values within each data stream (i.e., each univariate time series sequence in a multivariate time series dataset) using known values in the same sequence, followed by a fully connected layer with dropout that uses information across different data streams to impute the missing values in each data stream. The dropout allows the model to capture the uncertainty in the imputed data, allowing for more robust imputation results, and the interpolation and imputation blocks allow for more robust performance for frequently and infrequently sampled data, respectively. When compared to classical methods such as spline or cubic interpolation and to advanced methods including GRU-D, M-RNN achieves consistently better imputation performance across various datasets, as measured by root mean squared error (RMSE) loss. Interestingly, despite having superior imputation performance, datasets imputed with M-RNN did not achieve significantly higher prediction accuracy when the imputed dataset was used for downstream classification tasks. When the model was modified to include binary cross-entropy loss as an end-to-end imputation and classification method like GRU-D, it achieved significantly improved classification accuracy

compared to all baselines including GRU-D, especially at higher missing rates up to 90%. This suggests that end-to-end deep learning models that jointly perform imputation and classification or regression tasks appear to have better overall performance. In 2018, Cao et al. introduced another end-to-end RNN-based model called Bidirectional Recurrent Imputation for Time Series (BRITS) [66], which uses a bidirectional RNN graph to improve the imputation accuracy. BRITS achieves significantly improved classification performance compared to GRU-D and M-RNN, with an average increase in AUC of 0.025 and 0.030 respectively, across two separate datasets. This model was later applied for use in wearables data imputation with a modified regularization term, where the authors were successfully able to use BRITS to impute missing values for features such as steps and heart rate, achieving better performance than baselines such as k-NN or SOFTIMPUTE, even with features that had 50% missing values [67]. In 2020, Variational-Recurrent Imputation Network (V-RIN) [68] was introduced as an end-to-end model comprising a VAE-based imputation subnetwork connected to an RNN-based imputation subnetwork that can be trained for a classification or regression task and that leverages information about imputation uncertainty. In particular, the VAE subnetwork takes as input multi-dimensional vectors for each patient sample for each time stamp, with missing values zeroed out. The decoding layer outputs an imputed representation of each input vector and is used to generate an uncertainty matrix comprising standard deviation values for each imputed feature, where a higher standard deviation implies a greater imputation uncertainty and all non-missing features have zero uncertainty. For each time point, this matrix, along with the representation, a mask matrix, and a time stamp matrix, are fed into the RNN subnetwork, which comprises a novel version of the GRU-D structure slightly modified to incorporate an uncertainty decay term that leverages information from the uncertainty matrix. V-RIN achieves better performance compared to M-RNN, BRITS, and GRU-D for a mortality prediction task using both Physionet and MIMIC-III datasets.

The growth in GAN-based multivariate time series imputation models was spearheaded by the introduction of GRUI-GAN by Luo et al. in 2018 as the first model to use a GAN for time series imputation [69]. They introduced a novel RNN cell (GRUI) which considers time lags incurred by irregularly sampled data and is included in the discriminator and generator of the GAN. This allows the model to learn the relationships between observed and unobserved data, and the temporal information. One of the disadvantages of GRUI-GAN is that it is based on a two-stage framework which can be time-consuming to run [70]. First it trains a GAN to generate samples and then it searches for generated samples that are most similar to each original input sample that contains missing values. In 2019, E²GAN (End-to-End Generative Adversarial Network) [70] was proposed as a one-stage method for multivariate time series imputation using GANs, which improves training efficiency. E²GAN generally achieved improved imputation performance across various missing rates compared to GRUI-GAN and classical methods and achieved significantly improved AUCs for mortality prediction

when imputed datasets were used for classification with various classifiers including support vector machine, logistic regression, and RNN. E²GAN is not an end-to-end joint imputation and classification model but still achieved improved performance over BRITS (AUC increased by 0.02) for mortality prediction using a healthcare dataset. In 2021, Miao et al. expanded on this work by introducing SSGAN [71], which comprises a generator, a discriminator, and a semi-supervised classifier that iteratively classifies unlabeled time series data and is based on a bi-directional RNN model like BRITS [71]. SSGAN achieved significantly improved imputation performance compared to BRITS and E²GAN as measured by RMSE loss. Although SSGAN is not an end-to-end imputation and classification or regression model, datasets imputed using this method and then used for downstream RNN-based learning tasks still achieved better prediction performance than BRITS and all other baselines, with a 17.2% improvement over BRITS for one of the datasets tested.

Wasserstein GANs (WGANs) were also introduced in 2017 to address implementation issues occurring in regular GANs [72]. Originally, GANs often relied on a loss function using Jensen-Shannon Divergence (JSD) that often produced training issues such as a vanishing gradient, which can occur when JSD is locally saturated and the loss function can no longer accurately update the generator, or mode collapse where a generator's settings collapse to one mode and produce the same outputs [72]. WGANs address this issue by replacing JSD with Wasserstein distance (WD) in the loss function thereby avoiding mode collapse and vanishing gradients due to the fact WD is continuous and converges on a linear function that prevents discriminator saturation [72].

Variants of WGANs have since been used to restore missing audio data through 'audio inpainting'. For example, Ebner & Eltelt 2020 approached the issue of long audio gaps (200 ms) through a dual-discriminator WGAN (D2WGAN) that showed improvement over the original WGAN model [73]. This improvement is due to the inclusion of two discriminators, each tasked to discriminate overlapping audio samples with either short (1 s) or long (2.5 s) missing data bordering real audio. The advantage of this approach then is the inclusion of more correlated information about the missing data and as such audio from the D2WGAN was subjectively scored for higher restoration quality than the original method. Likewise, GACELA [74], a long audio gap inpainter, was recently developed based on a conditional GAN (cGAN) with a Wasserstein loss to synthesize even longer, context-dependent audio gaps up to 1500 ms. Older GAN-based [75], [76] and neural network [77], [78], [79], [80] solutions have also been developed in an attempt to fill progressively longer audio gaps over time.

A few multivariate time series imputation algorithms have been developed in recent years that use GPs and VAEs, including GP-VAE [81], SGP-VAE [82], and L-VAE [83]. For example, GP-VAE (2020), the first algorithm that uses both concepts for time series imputation, models the low-dimensional representation of time series data based on a GP and showed improved imputation performance over baselines including HI-VAE and classical methods such as LOCF [81]. When datasets imputed

by GP-VAE were used for classification with logistic regression, it performed better than baselines. L-VAE achieved better predictive performance than BRITS, GRUI-GAN, GP-VAE, and other baselines, making it among the best-performing time series imputation methods published to date.

Other multivariate time series imputation methods published in 2020 and 2021 have been developed using novel approaches, including random drop [84], transformers [85], self-attention [86], and conditional score-based diffusion [87] models. For example, Random Drop Imputation with Self-Training (RDIS) [84] is an ensemble model that takes an incomplete dataset as input and then generates synthetic datasets by randomly selecting and labeling known values as “missing”. The objective of the model is explicitly to learn to impute these “missing” values for each dataset, meanwhile generating values for the actual missing data points in each ensemble component. The entropy of all imputed values is then computed, and imputed values for all data points with entropy below a given threshold are used as pseudo-labels for a subsequent self-training task to impute missing values from the original dataset. RDIS achieved improved imputation performance compared to M-RNN and BRITS. Another model, Global and Local Time Series Imputation with Multi-directional Attention Learning [86] (GLIMA) comprises local and global RNN layers followed by a multi-directional self-attention layer. GLIMA achieves improved performance over GRU-D, M-RNN, BRITS, GRUI-GAN, and E2GAN when the dataset was imputed and then used for various classifiers including RNN, support vector machine (SVM), and logistic regression (LR).

Overall, deep learning-based static and time series imputation models have been widely successful, with significantly better imputation performance and improved robustness to higher missing rates compared to classical methods. Furthermore, many of these are end-to-end methods that combine imputation and downstream learning, allowing users to train their models on incomplete data in one step, which would not otherwise be possible.

V. QUALITY CONTROL FOR TABULATION ERRORS

For structured tabular data, such as electronic health records or public health surveillance datasets, an important data quality consideration is the presence of errors, which can impact the accuracy of downstream data analyses, even when using robust machine learning algorithms [88]. Tabulation errors can be heterogeneous and may comprise incorrect values, uninterpretable values (ex. typos), inconsistent use of diagnostic or medications coding systems (ex. ICD-9 vs. ICD-10) or inputting data with the wrong units. A summary of how the issue of tabulation errors relevant for various types of COVID-19 health data is presented in Table I (short version) and Suppl. Table IV (detailed).

The following subsections describe classical and advanced methods for handling tabulation errors, with a focus on both quantitative and qualitative error detection methods applicable for static (i.e., tabular) data. Suppl. Fig. 6 provides examples of tabulation errors and classical and advanced quality control methods, as well as performance comparisons between these

methods. Advanced methods are further summarized in Suppl. Table V and Fig. 1 (Right), with a visual timeline presented in Fig. 3. The issue of tabulation errors is further visualized in Fig. 4.

A. Classical Approaches for Handling Tabulation Errors

Handling tabulation errors can be a manual and time-consuming task. Errors can be detected using quantitative or qualitative approaches. Quantitative error detection involves identifying outliers in a dataset, as described in Suppl. Table X. Qualitative error detection involves specifying logical patterns or relational constraints for observations or features in the dataset and using those to identify violations in the dataset [89]. These relational constraint rules are typically defined by domain experts and are summarized in Suppl. Table XI.

B. Advanced Approaches for Handling Tabulation Errors (Quantitative Error Detection)

Neural networks have been used for outlier detection from the early 2000s [90]. More advanced deep learning models have been developed in recent years based on progress in the neural networks research field. In 2016, the first ensemble autoencoder model for outlier detection (RandNet) was introduced for tabular data [91]. They set up the ensemble with several autoencoders, each with random connections between neurons dropped to make the models different enough from each other. Each autoencoder of the ensemble is trained independently, and the reconstruction error (“outlier score”) for each sample point for each autoencoder is calculated. The final outlier score for each sample is the median score obtained from all ensemble components and is used to train a supervised classifier to identify labeled outlier points. This model achieved a significantly higher outlier classification accuracy (92.87%) on various datasets compared to classical baselines such as local outlier factor (LOF, 50.63%). Deep Autoencoding Gaussian Mixture Model [92] (DAGMM) was later introduced as an unsupervised end-to-end anomaly detection method combining an autoencoder with a Gaussian Mixture Model (GMM). The latent representation and reconstruction error from the autoencoder are then modeled using the GMM to evaluate the energy or likelihood of each input sample. DAGMM achieved an improved F1 score for outlier classification compared to various baselines across several datasets. In 2017, Schlegl et al. [93] introduced the first deep generative adversarial network (GAN) for anomaly detection which was originally applied to image data but was then shown by Zenati et al. to be applicable to tabular data [94] as well. A standard convolutional GAN is initially trained using normal data without anomalies, with the objective of minimizing residual and discriminator loss terms that are designed to enable the model to better learn the statistical distribution of “normal” data. These loss terms are included as part of an “anomaly score,” which, upon running the trained model with both normal and anomalous samples, will return a high score for samples that do not resemble the normal data. In subsequent years, additional GAN-based outlier detection models were developed and applied to tabular data, including the works by Zenati et al. [94] in 2018 and Liu

et al. [95] in 2019 (“Multiple Objectives Generative Adversarial Active Learning”, MO-GAAL). The model by Zenati et al. outperformed that of Schlegl et al. in terms of precision, recall and F1 score for tabular outlier classification, and performed comparably to DAGMM. MO-GAAL uses multiple generators with different objectives to model the distribution of the dataset and generate potential outlier data points. As the model is trained using both normal and anomalous samples through multiple iterations, the discriminator starts to learn the distribution of the input data and learns the boundary between normal and anomalous data samples. MO-GAAL was compared to many classical outlier detection methods (i.e., distance-based, clustering-based, density-based, angle-based, and classification-based methods) across various datasets using the Friedman test and had the highest average rank in terms of overall performance. In 2019, adVAE (self-adversarial variational autoencoder) [96] was developed by Wang et al. and used a Gaussian transformer network to generate outlier latent variables, as well as encoder and generator subnetworks, both of which can discriminate between normal and outlier latent variables. This model outperformed MO-GAAL for outlier classification performance across almost all datasets tested. Many other deep learning-based works have been published for outlier detection [97], [98], [99] and this continues to be a rapidly growing field of research [100]. Meanwhile, there continues to be growth in development of novel statistical algorithms for outlier detection that do not involve deep learning, including density-based [101], graph-based [102], and ensemble-based [103] methods.

C. Advanced Approaches for Handling Tabulation Errors (Qualitative Error Detection)

In recent years, qualitative error detection tools have emerged that can automatically define integrity constraints, discover integrity constraint violations in a given dataset [104], [105], [106], and either provide suggestions on potential errors to repair or fix the errors automatically [89]. Given how time-consuming, complex, and prone to errors process of qualitative error detection can be, these algorithms have the potential to have a significant impact on data quality for many different research applications. Examples of recently developed cutting edge machine learning algorithms for error repair are described in Suppl. Table XII. In 2013, FASTDC [105] was developed to discover denial constraints from the dataset, which precludes added time and effort needed for a user to do this manually. Also in 2013, Chu et al. developed Holistic Data Cleaning (Holistic) [107], the first approach to data cleaning that integrates data from heterogeneous integrity constraint rules to identify and repair errors based on constraint violations. In this work, the user provides pre-specified data quality or integrity constraint rules, and the algorithm automatically finds and repairs violations. Another model, KATARA [108], uses an external knowledge base to identify qualitative errors in a dataset and then suggests to the user a set of possible repairs. In 2016, ActiveClean [109] was developed as an iterative data cleaning framework developed for convex-loss machine learning applications, such as support vector machines, linear regression, or logistic regression. The

user specifies the model of interest, and the machine learning model is iteratively trained. In each iteration, ActiveClean suggests to the user what data may need to be repaired and the user then manually repairs the data using value transformations or filtering operations. Thus, ActiveClean uses data from the model performance to improve repair suggestions, leading to more successful data cleaning.

In 2017, HoloClean [110] was developed as a probabilistic framework which takes as input a dirty dataset, an external reference knowledgebase, and user-defined integrity constraints, and finds and automatically repairs errors in the dataset. It does this using an error detection module, which utilizes integrity constraints and the external data to detect outlier and identify repairs, a compilation module, which compiles data as features in a probabilistic graphical model, and a repair module, which repairs data using probabilistic inference on the graphical model. To measure data cleaning performance, each value in each original dataset used for benchmarking was manually labeled as correct or incorrect and then precision, recall, and F1 score for the repaired dataset were calculated after data cleaning. HoloClean had better performance compared to previous baselines including Holistic and KATARA. For example, for a healthcare dataset, HoloClean achieved an F1 score of 0.832, compared to 0.435 and 0.379 for Holistic and KATARA, and was successful for all datasets whereas other methods had extremely long runtimes or failed to identify certain types of errors. Other frameworks, such as PIClean [111] and MLNClean [112], followed suit with similar approaches. PIClean (2019) is a probabilistic interactive data cleaning system that uses relations between data columns to identify potential dataset errors and repairs. The suggestions are provided to users, which can then be confirmed or rejected, and user feedback is used to improve the performance of the model. It is more interactive with the user compared to HoloClean. In 2019, another system (MLNClean) was developed using Markov logic networks (MLNs) to clean the dataset by identifying both schema-level and instance-level errors. Like HoloClean, it requires domain experts to specify integrity constraints. However, it achieves consistently improved error detection performance compared to HoloClean even at greater error percentages and is also consistently faster. More recently, Rotom [113] was developed, which is a platform that leverages data augmentation and uses transformer [114], meta-learning, and self-supervised learning concepts to detect qualitative errors. It requires the user to provide a small number of labeled training examples that are known to be incorrect but achieves good performance for error detection. Rotom was compared to other data augmentation frameworks, but its performance compared to previous state-of-the-art data cleaning methods was not discussed.

Overall, these methods are promising in that they not only save researchers significant time and effort compared to manual qualitative data cleaning but also achieve good consistent error detection or repair performance.

VI. QUALITY CONTROL FOR NOISE AND ARTIFACTS

Noise and artifacts can occur in waveform and medical image data (Table I Suppl. Table IV). They can interfere with relevant

physiological signals and thus pose a significant quality issue for researchers working with COVID-19 healthcare data. For waveform and medical image data, we focus the discussion on waveform data including electrocardiogram (ECG) signals, audio data and medical image data including X-ray or computed tomography (CT) images, respectively, which are most relevant for COVID-19 related research. ECG waveforms have been shown to be associated with COVID-19 cardiovascular complications [115], and several deep learning algorithms have recently been developed that can detect [115] or predict [116] these using ECG waveform data. Similarly, respiratory-related audio measurements related to speech or coughing sounds have been used to diagnose positive cases – although reliant on typically noisy and imbalanced datasets. Several publications have used the synthetic minority over-sampling technique (SMOTE) to address the overrepresentation of non-COVID19 subjects in otherwise large datasets (i.e., Coswara and Sarcos datasets) [117], [118], [119]. This common issue and its solutions, however, are out of the topical scope of our discussion. X-ray and CT images have also emerged as important data sources for AI tool development for COVID-19 screening and diagnosis. Low-dose chest CT (LDCT) scans have also been discussed as an alternative to normal-dose CT scans for COVID-19 routine practice to reduce the added risk to the patient [120], [121].

The following subsections detail key classical and advanced methods to handle noise and artifacts, with a focus on ECG waveform data (i.e., ECG, audio) and CT image data. Examples of these methods and comparisons between classical and advanced approaches are visualized in Suppl. Fig. 7. Advanced methods are further summarized in Suppl. Table IV and Fig. 1 (Right), with a visual timeline presented in Fig. 3. The issue of noise and artifacts is further visualized in Fig. 4.

A. Waveform Noise and Artifacts (Classical Approaches)

Waveforms are continuous data points collected over time and are generally obtained from sensors such as photoplethysmograms (PPGs), accelerometers, thermometers, touch sensors, and gyroscopes. They can be highly susceptible to noise and various types of artifacts including motion artifacts and electromagnetic interference (EMI). Classical waveform de-noising methods include empirical mode decomposition, adaptive filtering, and wavelet transforms (WT). Classical signal processing methods are not always optimal for improving ECG signal quality. Methods such as empirical mode decomposition (EMD) are suboptimal because they might remove true signals from ECGs, and adaptive filters such as normalized least mean squares (LMS) require a reference noise signal, which cannot always be obtained. Although WT methods are widely used for ECG de-noising, they work in the frequency domain and cannot always distinguish true signals from artifacts when the artifacts morphologically resemble the true signal, leading to residual noise in some cases.

One of the most common data quality issues addressed in audio analysis is removing background noise from a poor signal-to-noise ratio (SNR) to enhance speech recognition or audio quality. There are a variety of sources that contribute to a low

SNR such as acoustic, environmental or distorted sounds. Multiple classical approaches to audio denoising include Wiener filtering [122], [123], spectral subtraction [124], [125], minimum mean squared error (MMSE) estimation [126] and optimally-modified log-spectral amplitude (OM-LSA) estimation [127]. However, these methods can sometimes introduce additional artifacts, such as the generation of ‘musical noise’ through spectral subtraction due to the flat, short-time noise spectrum estimate that is subtracted from the whole spectrum [124]. Additionally, common among these approaches is the use mel frequency cepstral coefficients (MFCCs) as representational features and an outcome in voice recognition that is a more uniform, but less recognizable speech spectrum [128], [129].

B. Waveform Noise and Artifacts (Advanced Approaches)

Since the mid-2010s, a small but growing research field has developed to investigate how deep learning models can be used to for ECG de-noising, to achieve better performance than traditional approaches. For example, in 2016, Xiong et al. [130] developed the first de-noising autoencoder (DAE) for ECG signal de-noising using wavelet transform. They artificially added baseline wander (BW), muscle artifact (MA), and electrode motion (EM) artifacts to clean signals and then did a wavelet transform as a first de-noising step. The de-noised signal was then input into the encoding and decoding layers of the DAE to remove the residual noise and to output the clean signal. The resulting signal quality was significantly improved compared to WT, Stockwell transform, and other baselines. For example, for muscle artifact-corrupted ECG signals with signal-to-noise ratios (SNRs) of 5 decibels (dB), the signals de-noised using the proposed method achieved SNRs of 1.93 dB higher than those de-noised with WT (18.16 dB and 16.23 dB respectively). In 2019, Wang et al. [131] further developed a first generative adversarial network (GAN) approach towards ECG de-noising, where the generator subnetwork is trained to de-noise ECG waveform signals and the discriminator is trained to differentiate between de-noised signals and the original clean signal. The objective of the model is for the discriminator to be unable to adequately distinguish the de-noised signals from the original clean signals. This model achieved superior performance compared to WT for all common ECG artifacts tested at varying input noise levels. For example, WT improved the SNR of MA-corrupted signals from 5 dB to 18.16 dB, whereas the de-noised signals from the proposed method had a much higher SNR of 37.23 dB. This constitutes a 105% higher SNR compared to WT, which is significantly greater than the 12% improvement in SNR over WT from using the Xiong et al. method.

In 2018, Antczak developed the first recurrent neural network (RNN) for ECG signal de-noising using long short-term memory (LSTM) units and incorporating a de-noising autoencoder [132]. The input into the model was a pre-processed ECG signal normalized to zero mean, separated into samples where each sample comprised 600 time stamped data points in the signal. They pre-trained their model on synthetic data and then achieved good performance on actual ECG signals. While simple bandpass and

WT filters performed best for signals with smaller noise levels (SNR greater than -7 dB), the RNN consistently performed best with more noisy signals. Arsene et al. more recently developed ECG de-noising models using convolutional neural networks (CNNs) [133], [134], and found that CNN-based models achieve better performance than a LSTM model they trained. Chiang et al. [135] developed a convolutional denoising autoencoder and emphasized the superior performance of convolutional layers for ECG de-noising. In 2020, Casas et al. [136] developed an autoencoder-based adversarial network applicable to ECG and other waveforms that uses a noisy signal as input and through encoding and decoding convolutional layers, it outputs a clean signal. Meanwhile, the discriminator of the model is trained to differentiate between clean and noisy signals using the inner hidden layer of the autoencoder. While this was not compared to other deep learning-based ECG de-noising methods, it was able to improve the SNR of noisy signals from -6.72 dB to 5.30 dB, compared to WT which only achieved a SNR of -5.90 dB. Another recent model developed by Zubair et al. [137] was designed specifically to remove motion artifacts from ECGs. The model comprises an RNN which de-noised the input ECG signals, followed by a standard deep neural network that resulted in a clean signal without artifacts. In general, these deep learning-based approaches perform significantly better than classical methods.

Although there is a significant level of interest in automated COVID-19 detection using cough, sound and speech data, multiple authors report the presence of background noise which, despite having been reduced using classical audio denoising methods such as spectral noise gating and data augmentation via time stretching, can still persist [119], [138], [139]. However, many different types and combinations of deep learning methods have been employed to resolve the same issue including CNNs [140], RNNs [141], [142], [143], and GANs [144], [145], [146]. Deep learning methods for denoising audio commonly rely on a long-mel spectrograms for feature representation; however, some authors contend that while this method of compression and normalization can reduce training set size and training time, there may be an advantage to using raw waveform data for synthesizing higher quality sound and for future applications that avoid information loss but rely on more computational resources.

C. Medical Image Noise and Artifacts (Classical Approaches)

LDCT scans reduce patients' exposure to high levels of radiation compared to normal-dose images (NDCT), but simultaneously compromise the quality of the image. The added noise and artifacts in these images can make it challenging to use them for machine learning tasks if not properly managed. Common types of artifacts found in LDCT, NDCT, and X-Ray images are summarized in Suppl. Table XIII. Classical approaches for handling LDCT noise and artifacts in the image domain include non-local means [147], [148], dictionary learning [149], and block matching 3D methods [150], which became popular in the early 2010s (Suppl. Fig. 8).

D. Medical Image Noise and Artifacts (Advanced Approaches)

In recent years, there has been rapid growth in development of state-of-the-art deep learning-based tools to remove noise or artifacts from CT images in the image domain, after the image is reconstructed from the waveform domain. In 2017, Chen et al. [151] developed the first deep learning model for LDCT image de-noising from the image domain. They used a convolutional neural network (CNN) to map LDCT images to normal-dose versions of those images and achieved a peak signal-to-noise ratio (PSNR) and subjective quality rating comparable to that of block matching 3D. Also in 2017, Yang et al. (WGAN-VGG) [152] and Wolterink et al. [153] developed some of the first generative adversarial network-based models for LDCT de-noising. Although WGAN-VGG performed worse than dictionary reconstruction, in terms of PSNR and structural similarity index measure (SSIM), it is advantageous because it does not procure unintentional image blurring or waxy or blocky artifacts and achieved significantly better subjective quality ratings by radiologists. In the same year, Chen et al. introduced Residual Encoder-Decoder Convolutional Neural Network (RED-CNN) [154], which used both autoencoder and convolutional neural network-based architecture concepts. The model comprises a series of convolutional encoding layers followed by deconvolutional decoding layers to remove noise from LDCT images. The encoding layers filter noise from the original image, whereas the decoding layers and shortcut connections between encoding and decoding layers are used to recover structural details from the image. Based on qualitative analysis of image quality by experienced radiologists, images de-noised using RED-CNN had improved artifact reduction, noise suppression, contrast retention, lesion discrimination, and overall quality compared to baseline methods such as dictionary learning, non-local means, and the previously published algorithm by Chen et al. Unlike previous deep learning image de-noising models, RED-CNN also had improved PSNR compared to all baselines. In 2018, Structurally-Similar Multi-Scale Generative Adversarial Network (SMGAN-3D) [155] was also introduced as novel GAN-based model for LDCT image de-noising, but it did not achieve significant improvements over RED-CNN. Dilated Residual Learning (DRL) [156] is a CNN-based model introduced in 2019 which uses an edge detection convolution layer to identify object boundaries and uses dilated convolution layers to capture more contextual information from the input image in fewer layers to make computation less expensive. These modifications enabled it to achieve superior performance compared to the original Chen et al. model [151].

Most recently, in 2021, Eformer [157] was introduced for medical image de-noising and achieved state-of-the-art performance over baselines on the American Association of Physicists in Medicine (AAPM) Low-Dose CT Grand Challenge Dataset, including RED-CNN. Eformer is the first transformer-based model for medical image de-noising and is applicable not only to LDCT but to other types of medical images as well. It uses an encoder-decoder network with transformer blocks and integrates Sobel-Feldman filters for edge enhancement. Eformer surpasses

all baseline state-of-the-art models tested in terms of PSNR and RMSE, with a nearly 3% higher PSNR and 12% lower RMSE than RED-CNN. This continues to be an active area of research with new models building on previous ones and achieving improved performance [158], [159], [160]. A few models have been specifically designed to handle metal [161] or motion [162] artifacts in CT scans or general image modalities. For example, in 2017, DeblurGAN [162] was introduced to correct motion artifacts, and in 2019, Liao et al. introduced the Artifact Disentanglement Network (ADN) [163] for metal artifact reduction from CT images.

Many general image de-noising algorithms can also be applied to medical images. For example, a highly cited general image de-noising algorithm, Noise2Noise [164], was developed in 2018 and was able to de-noise images without access to clean training example images and has been successfully applied to remove noise from magnetic resonance (MR) images. MedGAN [45] is another such GAN-based algorithm that appears to be broadly applicable but that was tested with positron emission tomography (PET) images. This model introduces a novel generator architecture that links together several convolutional encoder-decoder subnetworks with skip connections to enhance the resolution of the generated images. Additionally, a feature extraction subnetwork is introduced that is pre-trained using synthetic samples and input samples of the target domain, with loss terms that minimize the differences between the generated and target domain styles. These novel contributions enable MedGAN to translate noisy input images into de-noised versions of the same images. Suppl. Table XIV provides a summary of some of the deep learning algorithms for medical image noise and artifact removal published in recent years, with a focus on those applicable to CT scans, which are commonly being used for COVID-19 research applications. Suppl. Table XV illustrates the role of deep learning in the reconstruction process for producing high quality images.

Apart from image noise and artifact, low resolution is also a common image quality issue in which the image can appear blurry [165] and thus difficult to interpret. It is important to note that state-of-the-art deep learning algorithms have been developed to improve the resolution of input images using deep convolutional neural networks [166], [167], unfolding networks [168], transformer networks [169], and more.

VII. FUTURE DIRECTIONS

For each of the four data quality issues discussed in this review, we identified key trends in quality control COVID-19 data that are likely to persist moving forward (Suppl. Fig. 9). To address the problem of lack of data standardization, both generative neural network models and federated learning architectures have enabled levels of standardization that were not previously possible. Generative models such as GANs have expanded the scope of what is possible with medical image standardization using classical approaches such as DICOM GSDF and histogram matching, enabling translation between different image styles, CT contrast styles, and more. Federated learning is a powerful methodology that precludes the need to directly harmonize data

from different sources prior to combining the data into a machine learning model for training, enabling data science researchers to train models with large-scale multi-institutional data without the added time and effort needed to manually standardize the data. We anticipate that the use of these tools will continue and expand moving forward. We anticipate that there will be continued growth in the use of advanced deep neural network architectures for static and time series data imputation, for managing noise and artifacts in waveform and medical image data, and for detecting outliers. We also expect that advanced probabilistic graphical models and automated approaches will increasingly be used to detect and correct qualitative errors.

VIII. CONCLUSION

This review summarizes a selection of novel state-of-the-art data quality control methods being developed to address the issues of lack of standardization, missing data, tabular errors, and noise and artifact. Broader leveraging of these tools across the COVID-19 research community can lead to better-performing algorithms powered by better-quality data.

ACKNOWLEDGMENT

The authors thank Dr. Wang's lab for providing valuable insights, and particularly thank Isha Shah, of the Georgia Institute of Technology Department of Computer Science, for her contributions. The authors appreciate the vibrant discussions in The Third NSF Workshop on Predicting Pandemic Emergence of the Predictive Intelligence for Pandemic Prevention (PIPP) initiative on Feb. 2021.

REFERENCES

- [1] L. Wynants et al., "Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal," *BMJ*, vol. 369, 2020, Art. no. m1328, doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328).
- [2] M. Roberts et al., "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Mach. Intell.*, vol. 3, no. 3, pp. 199–217, 2021. [Online]. Available: <https://www.nature.com/articles/s42256-021-00307-0>
- [3] "Data science and AI in the age of COVID-19," The Alan Turing Institute, 2020. [Online]. Available: https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf
- [4] F. Giuste et al., "Explainable artificial intelligence methods in combating pandemics: A systematic review," *IEEE Rev. Biomed. Eng.*, early access, Jun. 23, 2022, doi: [10.1109/RBME.2022.3185953](https://doi.org/10.1109/RBME.2022.3185953).
- [5] S. Kothari, J. H. Phan, T. H. Stokes, A. O. Osunkoya, A. N. Young, and M. D. Wang, "Removing batch effects from histopathological images for enhanced cancer diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 765–772, May 2014.
- [6] S. Kothari, J. H. Phan, and M. D. Wang, "Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade," *J. Pathol. Inform.*, vol. 4, no. 1, 2013, Art. no. 22. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2153353922006435>
- [7] T. H. Stokes, R. A. Moffitt, J. H. Phan, and M. D. Wang, "Chip artifact CORRECTion (caCORRECT): A bioinformatics system for quality assurance of genomics and proteomics array data," *Ann. Biomed. Eng.*, vol. 35, no. 6, pp. 1068–1080, 2007. [Online]. Available: <https://doi.org/10.1007/s10439-007-9313-y>
- [8] L. Tong, H. Wu, and M. D. Wang, "Caesnet: Convolutional AutoEncoder based Semi-supervised Network for improving multiclass classification of endomicroscopic images," *J. Amer. Med. Inform. Assoc.*, vol. 26, no. 11, pp. 1286–1296, 2019. [Online]. Available: <https://doi.org/10.1093/jamia/ocz089>

- [9] L. Tong et al., "Impact of RNA-seq data analysis algorithms on gene expression estimation and downstream prediction," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 17925. [Online]. Available: <https://doi.org/10.1038/s41598-020-74567-y>
- [10] J. Venugopalan, N. Chanani, K. Maher, and M. D. Wang, "Novel data imputation for multiple types of missing data in intensive care units," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1243–1250, May 2019.
- [11] R. A. Hoffman, J. Venugopalan, L. Qu, H. Wu, and M. D. Wang, "Improving validity of cause of death on death certificates," in *Proc. ACM Int. Conf. Bioinf., Comput. Biol., Health Inform.*, 2018, pp. 178–183.
- [12] T. H. Stokes, R. A. Moffitt, J. H. Phan, and M. D. Wang, "Chip artifact CORRECTION (caCORRECT): A bioinformatics system for quality assurance of genomics and proteomics array data," *Ann. Biomed. Eng.*, vol. 35, no. 6, pp. 1068–1080, 2007.
- [13] T. H. Stokes, J. Torrance, H. Li, and M. D. Wang, "ArrayWiki: An enabling technology for sharing public microarray data repositories and meta-analyses," *BMC Bioinf.*, vol. 9, no. 6, pp. 1–14, 2008.
- [14] R. Chandramohan, P.-Y. Wu, J. H. Phan, and M. D. Wang, "Benchmarking RNA-Seq quantification tools," in *Proc. IEEE 35th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2013, pp. 647–650.
- [15] R. A. Moffitt et al., "caCORRECT: Improving the accuracy and reliability of microarray data in the presence of artifacts," *BMC Bioinf.*, vol. 12, no. 1, pp. 1–17, 2011.
- [16] G. Quer et al., "Wearable sensor data and self-reported symptoms for COVID-19 detection," *Nature Med.*, vol. 27, no. 1, pp. 73–77, 2021. [Online]. Available: <https://doi.org/10.1038/s41591-020-1123-x>
- [17] T. Mishra et al., "Pre-symptomatic detection of COVID-19 from smart-watch data," *Nature Biomed. Eng.*, vol. 4, no. 12, pp. 1208–1220, 2020. [Online]. Available: <https://doi.org/10.1038/s41551-020-00640-6>
- [18] B. L. Smarr et al., "Feasibility of continuous fever monitoring using wearable devices," *Sci. Rep.*, vol. 10, no. 1, pp. 21640–21640, 2020. [Online]. Available: <https://doi.org/10.1038/s41598-020-78355-6>
- [19] "UCSF tempredict study," *Pulse Blog*, 2020. [Online]. Available: <https://blog.ouraring.com/ucsf-tempredict-study/>
- [20] A. Shapiro et al., "Characterizing COVID-19 and influenza illnesses in the real world via person-generated health data," *Patterns*, vol. 2, no. 1, pp. 100188–100188, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7815963/pdf/main.pdf>
- [21] "National COVID-19 chest image database (NCCID) – national COVID-19 chest image database documentation," NHSX, 2020. [Online]. Available: <https://nhxs.github.io/covid19-chest-imaging-database/index.html>
- [22] C. Bycroft et al., "The UK biobank resource with deep phenotyping and genomic data," *Nature*, vol. 562, no. 7726, pp. 203–209, 2018. [Online]. Available: <https://doi.org/10.1038/s41586-018-0579-z>
- [23] A. Wood et al., "Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: Data resource," *BMJ*, vol. 373, 2021, Art. no. n826. [Online]. Available: <https://dx.doi.org/10.1136/bmj.n826>
- [24] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Sci. Data*, vol. 8, no. 1, 2021, Art. no. 156. [Online]. Available: <https://doi.org/10.1038/s41597-021-00937-4>
- [25] N. Sharma et al., "Coswara—A database of breathing, cough, and voice sounds for COVID-19 diagnosis," 2020. Accessed: May, 01, 2020. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv200510548S>
- [26] M. Cohen-McFarlane, R. Goubran, and F. Knoefel, "Novel coronavirus cough database: Nococoda," *IEEE Access*, vol. 8, pp. 154087–154094, 2020. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8545298/pdf/access-cohenmcfarlane-3018028.pdf>
- [27] J. Han et al., "Sounds of COVID-19: Exploring realistic performance of audio-based digital testing," *npj Digit. Med.*, vol. 5, no. 1, 2022, Art. no. 16. [Online]. Available: <https://doi.org/10.1038/s41746-021-00553-x>
- [28] G. A. Brat et al., "International electronic health record-derived COVID-19 clinical course profiles: The 4CE consortium," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–9, 2020. [Online]. Available: <https://doi.org/10.1038/s41746-020-00308-0>
- [29] D. Bender and K. Sartipi, "HL7 FHIR: An agile and RESTful approach to healthcare information exchange," in *Proc. IEEE 26th Int. Symp. Comput.-Based Med. Syst.*, 2013, pp. 326–331. [Online]. Available: <https://ieeexplore.ieee.org/document/6627810/>
- [30] J. C. Mandl, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, "Smart on FHIR: A standards-based, interoperable APPs platform for electronic health records," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 5, pp. 899–901, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4997036/>
- [31] X. Chu, "Data cleaning," in *Encyclopedia of Big Data Technologies*, S. Sakr and A. Y. Zomaya, Eds., Cham: Springer, 2019. [Online]. Available: https://doi.org/10.1007/978-3-319-77525-8_3
- [32] M. Mustra, K. Delac, and M. Grgic, "Overview of the DICOM standard," in *Proc. 50th Int. Symp. ELMAR*, 2008, vol. 1, pp. 39–44.
- [33] E. Samei, A. Rowberg, E. Avraham, and C. Cornelius, "Toward clinically relevant standardization of image quality," *J. Digit. Imag.*, vol. 17, no. 4, pp. 271–278, 2004.
- [34] A. Weeks, L. Sartor, and H. Myler, "Histogram specification of 24-bit color images in the color difference (C-Y) color space," *Electron. Imag.*, vol. 3646, pp. 290–300, 1999. [Online]. Available: <https://doi.org/10.1117/12.341098>
- [35] G. Liang, S. Fouladvand, J. Zhang, M. Brooks, N. Jacobs, and J. Chen, "GANai: Standardizing CT images using generative adversarial network with alternative improvement," 2018, pp. 1–11.
- [36] M. Brendan and R. Daniel, "Federated learning: Collaborative machine learning without centralized training data," Apr. 6, 2017. [Online]. Available: <https://ai.googleblog.com/2017/04/federated-learningcollaborative.html>
- [37] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.
- [38] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, 2020, vol. 2, pp. 429–450.
- [39] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with Non-IID data," *arXiv:1806.00582*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180600582Z>
- [40] I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," *Nature Med.*, vol. 27, no. 10, pp. 1735–1743, 2021. [Online]. Available: <https://dx.doi.org/10.1038/s41591-021-01506-3>
- [41] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, and B. A. Research, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134. [Online]. Available: <https://github.com/phillipipix2pix>
- [42] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Sevinck, C. A. T. van den Berg, and I. Išgum, "Deep MR to CT synthesis using unpaired data," in *Simulation and Synthesis in Medical Imaging*, Cham: Springer Int. Publishing, 2017, pp. 14–23, doi: https://doi.org/10.1007/978-3-319-68127-6_2
- [43] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, "Generating synthetic CTs from magnetic resonance images using generative adversarial networks," *Med. Phys.*, vol. 45, pp. 3627–3636, 2018.
- [44] H. Choi and D. S. Lee, "Generation of structural MR images from amyloid PET: Application to MR-less quantification," *J. Nucl. Med.*, vol. 59, no. 7, pp. 1111–1117, 2018.
- [45] K. Armanious et al., "MedGAN: Medical image translation using GANs," *Computerized Med. Imag. Graph.*, vol. 79, pp. 101684–101684, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S089561119300990?via%3Dihub>
- [46] K. Armanious, C. Jiang, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, "Unsupervised medical image translation using Cycle-MedGAN," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sept. 2–6, 2019, pp. 1–5, doi: [10.23919/EUSIPCO.2019.8902799](https://doi.org/10.23919/EUSIPCO.2019.8902799)
- [47] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image translation using cycle-consistent adversarial networks," in *IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251, doi: [10.1109/ICCV.2017.244](https://doi.org/10.1109/ICCV.2017.244)
- [48] N.-C. Ritea et al., "CyTran: Cycle-consistent transformers for non-contrast to contrast CT Translation." [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2021arXiv211006400R>
- [49] M. Selim, J. Zhang, B. Fei, G. Q. Zhang, and J. Chen, "STAN-CT: Standardizing CT image using generative adversarial networks," (in eng), *AMIA Annu. Symp. Proc.*, vol. 2020, pp. 1100–1109, 2020.
- [50] H. Zunair and A. B. Hamza, "Synthesis of COVID-19 chest X-rays using unpaired image-to-image translation," *Social Netw. Anal. Mining*, vol. 11, no. 1, 2021, pp. 1–12. [Online]. Available: <https://dx.doi.org/10.1007/s13278-021-00731-5>
- [51] C. Costa-Santos et al., "COVID-19 surveillance - A descriptive study on data quality issues," 2020. [Online]. Available: <https://www.medrxiv.org/content/medrxiv/early/2020/11/05/2020.11.03.20225565.full.pdf>
- [52] A. Khanzada et al., "Challenges and opportunities in deploying COVID-19 cough AI systems," *J. Voice: Official J. Voice Found.*, vol. 35, no. 6, pp. 811–812, 2021.
- [53] S. Moritz and T. Bartz-Beielstein, "imputeTS: Time series missing value imputation in R," *R. J.*, vol. 9, no. 1, pp. 207–218, 2017.

- [54] I. Goodfellow et al., “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: [10.1145/3422622](https://doi.org/10.1145/3422622).
- [55] J. Yoon, J. Jordon, and M. Van Der Schaar, “Gain: Missing data imputation using generative adversarial nets,” in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 5689–5698. [Online]. Available: <http://proceedings.mlr.press/v80/yoon18a.html>
- [56] S. Cheng-Xian Li, B. Jiang, and B. Marlin, “MisGAN: Learning from incomplete data with generative adversarial networks,” *arXiv:1902.09599*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190209599C>
- [57] L. Gondara and K. Wang, “MIDA: Multiple imputation using denoising autoencoders,” in *Advances in Knowledge Discovery and Data Mining*, Cham: Springer Int. Publishing, 2018, pp. 260–272.
- [58] A. Nazábal, P. M. Olmos, Z. Ghahramani, and I. Valera, “Handling incomplete heterogeneous data using VAEs,” *Pattern Recognit.*, vol. 107, Nov. 01, 2020, Art. no. 107501, doi: <https://doi.org/10.1016/j.patcog.2020.107501>.
- [59] Q. Ma, W.-C. Lee, T.-Y. Fu, Y. Gu, and G. Yu, “MIDIA: Exploring denoising autoencoders for missing data imputation,” *Data Mining Knowl. Discov.*, vol. 34, no. 6, pp. 1859–1897, 2020. [Online]. Available: <https://doi.org/10.1007/s10618-020-00706-8>
- [60] M. Śmieja, M. Kołomycki, M. Struski Juda, and M. A. T. Figueiredo, “Iterative imputation of missing data using auto-encoder dynamics,” in *Proc. Neural Inf. Process.*, 2020, pp. 258–269.
- [61] N. Abiri, B. Linse, P. Ed, and M. Ohlsson, “Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems,” *Neurocomput.*, vol. 365, no. C, pp. 137–146, 2019, doi: [10.1016/j.neucom.2019.07.065](https://doi.org/10.1016/j.neucom.2019.07.065).
- [62] H. Aidos and P. Tomás, “Neighborhood-aware autoencoder for missing value imputation,” in *Proc. IEEE 28th Eur. Signal Process. Conf.*, 2021, pp. 1542–1546.
- [63] D. P. Kingma and M. Welling, “Auto-Encoding variational bayes,” *arXiv:1312.6114*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6114K>
- [64] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, “Recurrent neural networks for multivariate time series with missing values,” *Sci. Rep.*, vol. 8, no. 1, pp. 6085–6085, 2018. [Online]. Available: www.nature.com/scientificreports https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5904216/pdf/41598_2018_Article_24271.pdf
- [65] J. Yoon, W. R. Zame, and M. van der Schaar, “Estimating missing data in temporal data streams using multi-directional recurrent neural networks,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1477–1490, Oct. 8, 2018.
- [66] W. Cao et al., “BRITS: Bidirectional recurrent imputation for time series,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, vol. 31, pp. 6776–6786.
- [67] T. Feng and S. Narayanan, “Imputing missing data in large-scale multivariate biomedical wearable recordings using bidirectional recurrent neural networks with temporal activation regularization,” in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 2529–2534.
- [68] A. W. Mulyadi, E. Jun, and H. I. Suk, “Uncertainty-Aware variational-recurrent imputation network for clinical time series,” *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9684–9694, 2022, doi: [10.1109/TCYB.2021.3053599](https://doi.org/10.1109/TCYB.2021.3053599).
- [69] Y. Luo, X. Cai, Y. Zhang, J. Xu, and X. Yuan, “Multivariate time series imputation with generative adversarial networks,” in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 1603–1614.
- [70] Y. Luo, Y. Zhang, X. Cai, and X. Yuan, “E² GAN: End-to-end generative adversarial network for multivariate time series imputation,” in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 3094–3100. [Online]. Available: <https://doi.org/10.24963/ijcai.2019/429>
- [71] X. Miao, Y. Wu, J. Wang, Y. Gao, X. Mao, and J. Yin, “Generative semi-supervised learning for multivariate time series imputation,” in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 10, pp. 8983–8991. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/17086>
- [72] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 214–223. [Online]. Available: <https://proceedings.mlr.press/v70/arjovsky17a.html>
- [73] P. P. Ebner and A. Eltelt, “Audio inpainting with generative adversarial network,” *arXiv:2003.07704*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv200307704E>
- [74] A. Marafioti, P. Majdak, N. Holighaus, and N. Perraudin, “GACELA: A generative adversarial context encoder for long audio inpainting of music,” *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 1, pp. 120–131, Jan. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9257074/>
- [75] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, “Adversarial generation of time-frequency features with application in audio synthesis,” *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, pp. 4352–4362, 2019. [Online]. Available: <https://proceedings.mlr.press/v97/marafioti19a.html>
- [76] C. Donahue, J. McAuley, and M. Puckette, “Adversarial audio synthesis,” *arXiv:1802.04208*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180204208D>
- [77] N. M. Khan and G. M. Khan, “Audio signal reconstruction using cartesian genetic programming evolved artificial neural network (CGPANN),” in *Proc. IEEE 16th Int. Conf. Mach. Learn. Appl., Conf. Proc.*, 2017, pp. 568–573. [Online]. Available: <https://ieeexplore.ieee.org/document/8260692/>
- [78] S. Hershey et al., “CNN architectures for large-scale audio classification,” *arXiv:1609.09430*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160909430H>
- [79] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, “A context encoder for audio inpainting,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 27, no. 12, pp. 2362–2372, Dec. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8867915/>
- [80] A. van den Oord et al., “WaveNet: A generative model for raw audio,” *arXiv:1609.03499*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160903499V>
- [81] V. Fortuin et al., “GP-VAE: Deep probabilistic multivariate time series imputation,” in *Proc. 23rd Int. Conf. Artif. Intell. Statist.*, 2020, pp. 1651–1661.
- [82] M. Ashman, J. So, W. Tebbutt, V. Fortuin, M. Pearce, and R. E. Turner, “Sparse gaussian process variational autoencoders,” *arXiv:2010.10177*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv201010177A>
- [83] S. Ramchandran, G. Tikhonov, K. Kujanpää, M. Koskinen, and H. Lähdesmäki, “Longitudinal variational autoencoder,” *arXiv:2006.09763*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv200609763R>
- [84] T.-M. Choi, J.-S. Kang, and J.-H. Kim, “RDIS: Random drop imputation with self-training for incomplete time series data,” *arXiv:2010.10075*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv201010075C>
- [85] S. Tipirneni and C. K. Reddy, “Self-Supervised transformer for sparse and irregularly sampled multivariate clinical time-series,” *arXiv:2107.14293*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2021arXiv210714293T>
- [86] Q. Suo, W. Zhong, G. Xun, J. Sun, C. Chen, and A. Zhang, “GLIMA: Global and local time series imputation with multi-directional attention learning,” in *Proc. IEEE Conf. Big Data*, 2020, pp. 798–807. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9378408>
- [87] Y. Tashiro, J. Song, Y. Song, and S. Ermon, “CSDI: Conditional Score-based diffusion models for probabilistic time series imputation,” *arXiv:2107.03502*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2021arXiv210703502T>
- [88] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, “Data cleaning: Overview and emerging challenges,” in *Proc. Int. Conf. Manage. Data*, 2016, pp. 2201–2206. [Online]. Available: <https://doi.org/10.1145/2882903.2912574>
- [89] X. Chu and I. F. Ilyas, “Qualitative data cleaning,” *Proc. VLDB Endow.*, 2016, vol. 9, no. 13, pp. 1605–1608. [Online]. Available: <https://doi.org/10.14778/3007263.3007320>
- [90] G. Williams, R. Baxter, H. Hongxing, S. Hawkins, and G. Lifang, “A comparative study of RNN for outlier detection in data mining,” in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 709–712.
- [91] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga, “Outlier detection with autoencoder ensembles,” in *Proc. 17th SIAM Int. Conf. Data Mining*, 2017, pp. 90–98, doi: [10.1137/1.9781611974973.11](https://doi.org/10.1137/1.9781611974973.11).
- [92] Z. Bo et al., “Deep autoencoding gaussian mixture model for unsupervised anomaly detection,” in *Proc. Int. Conf. Learn. Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=BJJLHbb0-T_Schlegl_P_Seeboeck_S_M_Waldstein_U_Schmidt-Erfurth_G_Langs
- [93] T. Schlegl, P. Seeboeck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” in *Information Processing in Medical Imaging*, Cham: Springer Int. Publishing, 2017, pp. 146–157.
- [94] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, “Efficient GAN-Based anomaly detection,” *arXiv:1802.06222*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180206222Z>
- [95] Y. Liu et al., “Generative adversarial active learning for unsupervised outlier detection,” 2019. [Online]. Available: <https://arxiv.org/abs/1809.10816>

- [96] X. Wang, Y. Du, S. Lin, P. Cui, Y. Shen, and Y. Yang, "adVAE: A self-adversarial variational autoencoder with Gaussian anomaly prior knowledge for anomaly detection," *Knowl.-Based Syst.*, vol. 190, pp. 105187–105187, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0950705119305283?via%3Dihub>
- [97] J. Bian, X. Hui, S. Sun, X. Zhao, and M. Tan, "A novel and efficient CVAE-GAN-based approach with informative manifold for semi-supervised anomaly detection," *IEEE Access*, vol. 7, pp. 88903–88916, 2019.
- [98] M. Munoz-Organero, "Outlier detection in wearable sensor data for human activity recognition (HAR) based on DRNNs," *IEEE Access*, vol. 7, pp. 74422–74436, 2019.
- [99] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang, "Deep structured energy based models for anomaly detection," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, 2016, pp. 1100–1109. [Online]. Available: <https://proceedings.mlr.press/v48/zhai16.html>
- [100] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2021, doi: [10.1145/3439950](https://doi.org/10.1145/3439950).
- [101] M. Riahi-Madvar, A. Akbari Azirani, B. NaserSharif, and B. Raahemi, "A new density-based subspace selection method using mutual information for high dimensional outlier detection," *Knowl.-Based Syst.*, vol. 216, 2021, Art. no. 106733.
- [102] D. Amagata, M. Onizuka, and T. Hara, "Fast and exact outlier detection in metric spaces: A proximity graph-based approach," in *Proc. Int. Conf. Manage. Data*, 2021, pp. 36–48, doi: [10.1145/3448016.3452782](https://doi.org/10.1145/3448016.3452782).
- [103] N. Jayanthi, B. Vijaya Babu, and N. Sambasiva Rao, "An ensemble framework based outlier detection system in high dimensional data using tree technique," *Mater. Today: Proc.*, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2214785320391550>
- [104] X. Chu, I. F. Ilyas, P. Papotti, and Y. Ye, "Ruleminer: Data quality rules discovery," in *Proc. IEEE Int. Conf. Data Eng.*, 2014, pp. 1222–1225. [Online]. Available: <https://ieeexplore.ieee.org/document/6816746/>
- [105] X. Chu, I. F. Ilyas, and P. Papotti, "Discovering denial constraints," *Proc. VLDB Endowment*, vol. 6, pp. 1498–1501, 2013. [Online]. Available: <https://dl.acm.org/doi/10.14778/2536258.2536262>
- [106] M. Farid, A. Roatis, I. F. Ilyas, H.-F. Hoffmann, and X. Chu, "Clams: Bringing quality to data lakes," in *Proc. Int. Conf. Manage. OOF Data*, 2016, pp. 2089–2092. [Online]. Available: <http://dx.doi.org/10.1145/2882903.2889391>
- [107] X. Chu, I. F. Ilyas, and P. Papotti, "Holistic data cleaning: Putting violations into context," in *Proc. IEEE 29th Int. Conf. Data Eng.*, 2013, pp. 458–469.
- [108] X. Chu et al., "Katara: A data cleaning system powered by knowledge bases and crowdsourcing," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2015, vol. 2015, pp. 1247–1261. [Online]. Available: <https://dl.acm.org/doi/10.1145/2723372.2749431>
- [109] S. Krishnan, J. Wang, E. Wu, M. J. Franklin, and K. Goldberg, "ActiveClean: Interactive data cleaning for statistical modeling," *Proc. VLDB Endow.*, vol. 9, no. 12, pp. 948–959, 2016, doi: [10.14778/2994509.2994514](https://doi.org/10.14778/2994509.2994514).
- [110] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré, "HoloClean: Holistic data repairs with probabilistic inference," *Proc. VLDB Endow.*, vol. 10, no. 11, pp. 1190–1201, 2017, doi: [10.14778/3137628.3137631](https://doi.org/10.14778/3137628.3137631).
- [111] Y. Zhuoran and X. Chu, "Piclean: A probabilistic and in-teractive data cleaning system," in *Proc. Int. Conf. Manage. Data*, 2019. [Online]. Available: <https://doi.org/10.1145/3299869.3320214>
- [112] Y. Gao, C. Ge, X. Miao, H. Wang, B. Yao, and Q. Li, "A hybrid data cleaning framework using Markov logic networks," 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/9151362/>
- [113] Z. Miao, Y. Li, and X. Wang, "Rotom: A meta-learned data augmentation framework for entity matching, data cleaning, text classification, and beyond," in *Proc. Int. Conf. Manage. Data Jun. 9, 2021*, pp. 1303–1316.
- [114] A. Vaswani et al., "Attention is all you need," 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [115] F. Angeli et al., "Electrocardiographic features of patients with COVID-19 pneumonia," *Eur. J. Intern. Med.*, 2020, vol. 78, pp. 101–106. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0953620520302533>
- [116] L. Tan et al., "Toward real-time and efficient cardiovascular monitoring for COVID-19 patients by 5G-enabled wearable medical devices: A deep learning approach," *Neural Comput. Appl.*, 2021. [Online]. Available: <https://doi.org/10.1007/s00521-021-06219-9>
- [117] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 cough classification using machine learning and global smartphone recordings," *Comput. Biol. Med.*, vol. 135, 2021, Art. no. 104572. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8213969/pdf/main.pdf>
- [118] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and S. M. S. Islam, "Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method," *Comput. Biol. Med.*, vol. 145, 2022, Art. no. 105405. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8926945/pdf/main.pdf>
- [119] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Comput. Biol. Med.*, vol. 141, 2022, Art. no. 105153. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8679499/pdf/main.pdf>
- [120] S. M. H. Tabatabaei et al., "A low-dose chest CT protocol for the diagnosis of COVID-19 pneumonia: A prospective study," *Emerg. Radiol.*, vol. 27, no. 6, pp. 607–615, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32789807>
- [121] S. Tofighi, S. Najafi, S. K. Johnston, and A. Gholamrezaezhad, "Low-dose CT in COVID-19 outbreak: Radiation safety, image wisely, and image gently pledge," *Emerg. Radiol.*, vol. 27, no. 6, pp. 601–605, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32390122>
- [122] T. V. Sreenivas and P. Kirnapure, "Codebook constrained wiener filtering for speech enhancement," *IEEE Speech Audio Process.*, vol. 4, no. 5, pp. 383–389, Sep. 1996. [Online]. Available: <https://ieeexplore.ieee.org/document/536932/>
- [123] N. Alamdari, S. Yaraganalu, and N. Kehtarnavaz, "A real-time personalized noise reduction smartphone app for hearing enhancement," in *Proc. IEEE Signal Process. Med. Biol. Symp.*, 2018, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/document/8615620/>
- [124] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1979, vol. 4, pp. 208–211. [Online]. Available: <https://ieeexplore.ieee.org/document/1170788/>
- [125] N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study," *Procedia Comput. Sci.*, vol. 54, pp. 574–584, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050915013903>
- [126] Y. Dong, D. Li, D. Jasha, W. Jian, G. Yifan, and A. Acero, "A minimum-mean-square-error noise reduction algorithm on Mel-frequency cepstra for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Conf. Proc.*, 2008, pp. 4041–4044.
- [127] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 113–116, Apr. 2002.
- [128] T. Gao, J. Du, L.-R. Dai, and C.-H. Lee, "A unified DNN approach to speaker-dependent simultaneous speech enhancement and speech separation in low SNR environments," *Speech Commun.*, vol. 95, pp. 28–39, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167639316303843>
- [129] Y. H. Tu, J. Du, T. Gao, and C. H. Lee, "A multi-target SNR-progressive learning approach to regression based speech enhancement," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 28, pp. 1608–1619, 2020.
- [130] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," *Eng. Appl. Artif. Intell.*, vol. 52, no. C, pp. 194–202, 2016. [Online]. Available: <https://doi.org/10.1016/j.engappai.2016.02.015>
- [131] J. Wang et al., "Adversarial de-noising of electrocardiogram," *Neurocomputing*, vol. 349, pp. 212–224, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952523121930534X>
- [132] K. Antczak, "Deep recurrent neural networks for ECG signal denoising," *arXiv:1807.11551*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180711551A>
- [133] C. T. C. Arsene, R. Hankins, and H. Yin, "Deep learning models for denoising ECG signals," in *Proc. IEEE 27th Eur. Signal Process. Conf.*, 2019, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8902833>
- [134] C. Arsene, "Design of deep convolutional neural network architectures for denoising electrocardiographic signals," *IEEE Conf. Comput. Intell. Bioinf. Comput. Biol.*, 2020, pp. 1–8, doi: [10.1109/CIBCB48159.2020.9277718](https://doi.org/10.1109/CIBCB48159.2020.9277718).

- [135] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ECG signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60806–60813, 2019. [Online]. Available: <https://dx.doi.org/10.1109/access.2019.2912036>
- [136] L. Casas, A. Klimmek, N. Navab, and V. Belagiannis, "Adversarial signal denoising with encoder-decoder networks," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, vol. 2021, pp. 1467–1471. [Online]. Available: <https://ieeexplore.ieee.org/document/9287738/>
- [137] M. Zubair, G. N. V. S. C. Mouli, and R. A. Shaik, "Removal of motion artifacts from ECG signals by combination of recurrent neural networks and deep neural networks," in *Proc. IEEE 2nd Int. Conf. Elect., Control Instrum. Eng.*, 2020, pp. 1–7.
- [138] V. Dentamaro, P. Giglio, D. Impedovo, L. Moretti, and G. Pirlo, "AUCO ResNet: An end-to-end network for COVID-19 pre-screening from cough and breath," *Pattern Recognit.*, vol. 127, 2022, Art. no. 108656. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320322001376>
- [139] V. Despotovic, M. Ismael, M. Cornil, R. M. Call, and G. Fagherazzi, "Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results," *Comput. Biol. Med.*, vol. 138, 2021, Art. no. 104944. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8513517/pdf/main.pdf>
- [140] N. Alamdari, A. Azarang, and N. Kehtarnavaz, "Improving deep speech denoising by noisy2noisy signal mapping," *Appl. Acoust.*, vol. 172, 2021, Art. no. 107631. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003682X20307350>
- [141] A. Maas, Q. V. Le, T. M. O'neil, O. Vinyals, P. Nguyen, and A. Y. Ng, "Recurrent neural networks for noise reduction in robust ASR," 2012. [Online]. Available: <https://research.google/pubs/pub45168/>
- [142] T.-A. Hsieh, H.-M. Wang, X. Lu, and Y. Tsao, "WaveCRN: An efficient convolutional recurrent neural network for end-to-end speech enhancement," *IEEE Signal Process. Lett.*, vol. 27, pp. 2149–2153, 2020. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020ISPL...27.2149H>
- [143] K. Osako, R. Singh, and B. Raj, "Complex recurrent neural networks for denoising speech signals," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2015, pp. 1–5.
- [144] S. Abdulatif, K. Armanious, K. Guirguis, J. T. Sajeew, and B. Yang, "AeGAN: Time-frequency speech denoising via generative adversarial networks," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 451–455.
- [145] J. Su, Z. Jin, and A. Finkelstein, "HiFi-GAN: High-fidelity denoising and dereverberation based on speech deep features in adversarial networks," *arXiv:2006.05694*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2020arXiv200605694S>
- [146] P. Santiago, B. Antonio, and S. Joan, "SEGAN: Speech enhancement generative adversarial network," pp. 3642–3646, 2017, doi: [10.21437/Interspeech.2017-1428](https://doi.org/10.21437/Interspeech.2017-1428).
- [147] Y. Chen et al., "Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means," *Phys. Med. Biol.*, vol. 57, no. 9, pp. 2667–2688, 2012.
- [148] J. Ma et al., "Low-dose computed tomography image restoration using previous normal-dose scan," *Med. Phys.*, vol. 38, no. 10, pp. 5713–5731, 2011.
- [149] Y. Chen et al., "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, no. 16, pp. 5803–5820, 2013.
- [150] P. Fumene Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3D random noise filtering for absorption optical projection tomography," *Phys. Med. Biol.*, vol. 55, no. 18, pp. 5401–5415, 2010. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/20736500>
- [151] H. Chen et al., "Low-dose CT via convolutional neural network," *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, 2017. [Online]. Available: <https://www.osapublishing.org/boe/fulltext.cfm?uri=boe-8-2-679&id=357201>
- [152] Q. Yang et al., "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021013/pdf/nihms973044.pdf>
- [153] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/7934380/>
- [154] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
- [155] N. Wang, Z. Cui, Y. Su, C. He, Y. Lan, and A. Li, "SmGAN: A self-modulated generative adversarial network for single image dehazing," *AIP Adv.*, vol. 11, no. 8, 2021, Art. no. 085227. [Online]. Available: <https://aip.scitation.org/doi/abs/10.1063/5.0059424>
- [156] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Deep learning for low-dose CT denoising using perceptual loss and edge detection layer," *J. Digit. Imag.*, vol. 33, no. 2, pp. 504–515, 2020. [Online]. Available: <https://doi.org/10.1007/s10278-019-00274-4>
- [157] A. Luthra, H. Sulakhe, T. Mittal, A. Iyer, and S. Yadav, "Eformer: Edge enhancement based transformer for medical image denoising," *arXiv:2109.08044*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2021arXiv210908044L>
- [158] Z. Huang et al., "CaGAN: A cycle-consistent generative adversarial network with attention for low-dose CT imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1203–1218, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9154550/>
- [159] J. Chi, C. Wu, X. Yu, P. Ji, and H. Chu, "Single low-dose CT image denoising using a generative adversarial network with modified U-Net generator and multi-level discriminator," *IEEE Access*, vol. 8, pp. 133470–133487, 2020.
- [160] K. Choi, M. Vania, and S. Kim, "Semi-supervised learning for low-dose CT image restoration with hierarchical deep generative adversarial network (HD-GAN)," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 2683–2686.
- [161] L. Gjestebj et al., "A dual-stream deep convolutional network for reducing metal streak artifacts in CT images," *Phys. Med. Biol.*, vol. 64, no. 23, 2019, Art. no. 235003.
- [162] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 8183–8192. [Online]. Available: <http://arxiv.org/abs/1711.07064>
- [163] H. Liao, W.-A. Lin, J. Yuan, S. K. Zhou, and J. Luo, "Artifact disentanglement network for unsupervised metal artifact reduction," in *Medical Image Computing and Computer Assisted Intervention*, Cham: Springer Int. Publishing, 2019, pp. 203–211.
- [164] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data - official tensorflow implementation of the ICML 2018 paper," 2018. [Online]. Available: <https://github.com/NVlabs/noise2noise>
- [165] T. Higaki, Y. Nakamura, F. Tatsugami, T. Nakaura, and K. Awai, "Improvement of image quality at CT and MRI using deep learning," *Japanese J. Radiol.*, vol. 37, no. 1, pp. 73–80, 2019. [Online]. Available: <https://doi.org/10.1007/s11604-018-0796-2>
- [166] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016. [Online]. Available: <https://ieeexplore.ieee.org/document/7115171/>
- [167] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep CNN with batch renormalization," *Neural Netw.*, vol. 121, pp. 461–473, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0893608019302394?via%3Dihub>
- [168] K. Z. Luc, V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3217–3226. [Online]. Available: <https://github.com/cszn/USRNet>
- [169] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5791–5800.