Critical Challenges and Guidelines in Evaluating Synthetic Tabular Data: A Systematic Review

NAZIA NAFIS, Healthy Lifespan Institute, School of Computer Science, University of Sheffield, UK
IÑAKI ESNAOLA, School of Electrical and Electronic Engineering, University of Sheffield, UK
ALVARO MARTINEZ-PEREZ, Healthy Lifespan Institute, Department of Sociological Studies, University of Sheffield, UK
MARIA-CRUZ VILLA-URIOL, Healthy Lifespan Institute, School of Computer Science, University of Sheffield, UK
VENET OSMANI, Digital Environment Research Institute, Queen Mary University of London, UK

Abstract. Generating synthetic tabular data can be challenging, however evaluation of their quality is just as challenging, if not more. This systematic review sheds light on the critical importance of rigorous evaluation of synthetic health data to ensure reliability, relevance, and their appropriate use. Based on screening of 1766 papers and a detailed review of 101 papers we identified key challenges, including lack of consensus on evaluation methods, improper use of evaluation metrics, limited input from domain experts, inadequate reporting of dataset characteristics, and limited reproducibility of results. In response, we provide several guidelines on the generation and evaluation of synthetic data, to allow the community to unlock and fully harness the transformative potential of synthetic data and accelerate innovation.

CCS Concepts: • **Computing methodologies** → **Machine learning approaches**; • **Applied computing** → **Health informatics**.

Additional Key Words and Phrases: synthetic data, tabular data, time series data

ACM Reference Format:
Nazia Nafis, Iñaki Esnaola, Alvaro Martinez-Perez, Maria-Cruz Villa-Uriol, and Venet Osmani. 2025. Critical Challenges and Guidelines in Evaluating Synthetic Tabular Data: A Systematic Review. 1, 1 (April 2025), 26 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Access to high-quality data is fundamental to advancing scientific research. In disciplines such as healthcare, data is pivotal to enhance patient care, optimise resource management, and enable the discovery of new medical insights, particularly with the rise of artificial intelligence. Structured health data, such as tabular electronic health records, have been recognised as having one of the highest potential to provide timely and relevant information in clinical decision-making [1]. However, complex data-sharing governance rules have resulted in health data being locked away in isolated silos [2, 3], where they generally remain inaccessible except to a few researchers [4]. This inevitably hampers

Authors' addresses: Nazia Nafis, Healthy Lifespan Institute, School of Computer Science, University of Sheffield, Sheffield, UK, nnafis1@sheffield.ac.uk; Iñaki Esnaola, School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, UK, esnaola@sheffield.ac.uk; Alvaro Martinez-Perez, Healthy Lifespan Institute, Department of Sociological Studies, University of Sheffield, Sheffield, UK, a.martinez-perez@sheffield.ac.uk; Maria-Cruz Villa-Uriol, Healthy Lifespan Institute, School of Computer Science, University of Sheffield, Sheffield, UK, m.villa-uriol@sheffield.ac.uk; Venet Osmani, Digital Environment Research Institute, Queen Mary University of London, London, UK, v.osmani@qmul.ac.uk.

reproducible health research, hindering the advancement of patient care and impeding the future potential of clinical artificial intelligence [5].

There is an urgent need to democratise access to health data [4] without losing sight of patient privacy and confidentiality [6]. In this context, synthetic health data emerges as an attractive solution to address this challenge, and institutions worldwide are increasingly recognising their potential. For example, the United States Department of Health and Human Services has made available a Synthetic Health Data Generation Engine[1] to accelerate patient-centred outcomes research and "address the need for research-quality synthetic data". The United Kingdom's National Health Service (NHS) has rolled out an 'Artificial Data Pilot'[2] that aims to "provide users with large volumes of data that share some of the characteristics of real data while protecting patient confidentiality". Similar efforts are recorded in Canada's health economic hub Health City [7] and Germany's Charité Lab[3] for Artificial Intelligence in Medicine. Research-quality synthetic data (the focus of our work) can be used to rapidly develop and test preliminary hypotheses before applying them to real datasets [8]. They can also improve research pipeline by acting as a proxy for real-world data [9]. Furthermore, the controlled generation of synthetic health data can include a balanced representation of different demographic groups [10]. This would ensure that the previously underrepresented socio-demographic groups are adequately represented, thereby mitigating biases in health research that arise from skewed real-world health datasets and, in turn, address model fairness [11–13].

However, despite the above-mentioned advantages of synthetic health data, major challenges remain with their large-scale adoption. One of the major challenge is the lack of consensus on evaluating synthetically generated data vis à vis the corresponding real data [14–16]. This not only makes it difficult to track the state-of-the-art progress of synthetic data generation methods but also poses barriers to trust and adoption, as well as presents regulatory and compliance issues.

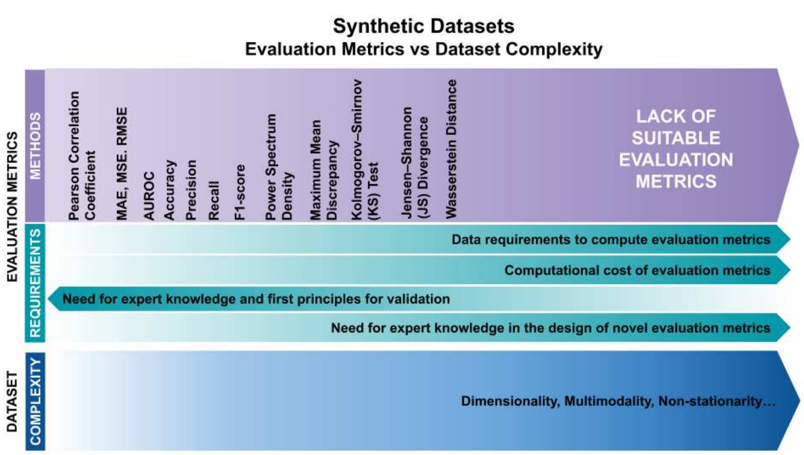

Fig. 1. There is an increasing lack of appropriate evaluation metrics (due to increasing difficulties in computation and increasing difficulties in evaluation of the metrics), with the increase in data complexity of the synthetic datasets.

To shed light on the evaluation approaches of synthetic data we have conducted a systematic review of 1766 research articles published in the last ten years.

This is the first review of this type and size to understand which approaches are being used to evaluate the quality of synthetic data, along with the associated data generation methods and their target application areas. We focused on structured tabular and time-series health data since this is one of the areas with the highest potential in advancing healthcare [17] and present

[1]https://aspe.hhs.gov/synthetic-health-data-generation-engine-accelerate-patient-centered-outcomes-research
[2]https://digital.nhs.uk/services/artificial-data
[3]https://claim.charite.de/en/

unique data challenges, such as dealing with missingness. In addition, there is a higher consensus on the evaluation methods for other data modalities, such as imaging and text, where the respective research communities have developed metrics such as Fréchet Inception Distance (FID) [18] and BERTScore [19] respectively.

We observe that with the increasing complexity of the synthetic datasets (including dimensionality, multimodality, and non-stationarity), there is growing lack of suitable evaluation metrics, as shown in Fig.1. This is manifested in the increasing difficulties in the computation as well as evaluation of the metrics. Therefore, there's a critical need for the use of appropriate statistical evaluation metrics to critically evaluate complex synthetic data, including involvement of expert stakeholders in: i) selection of appropriate evaluation metrics and, ii) interpretation of the resulting outcomes.This type of collaboration between researchers and clinical practitioners can lead to development of methods and metrics that implicitly incorporate domain knowledge, resulting in decreased need for expert knowledge in evaluating future synthetic data. As a result, distilling domain knowledge into operational constraints and guaranteeing that the underlying medical processes that govern the data generation are safeguarded, will open the door to novel machine learning evaluation paradigms.

In the following section we show the results of our analysis, followed by guidelines in evaluating synthetic tabular data.

2 RESULTS

Based on the screening of 1766 papers and a detailed review of 101, we present the following results, grouped into four categories namely, evaluation, generation, purpose and impact of synthetic data, as well as reproducibility of the results.

2.1 Evaluation of Synthetic Data

We categorise the approaches used in the evaluation of synthetic data in: Direct vs Indirect approaches, and Quantitative vs Qualitative methods.

Direct evaluation approaches involve using existing, standardised metrics to assess the quality of the synthetic data. **Indirect evaluation approaches** include non-standardised, domain-specific methods (such as TSTR - Train on Synthetic, Test on Real) to assess synthetic data in real-world applications. Indirect approaches extend beyond the

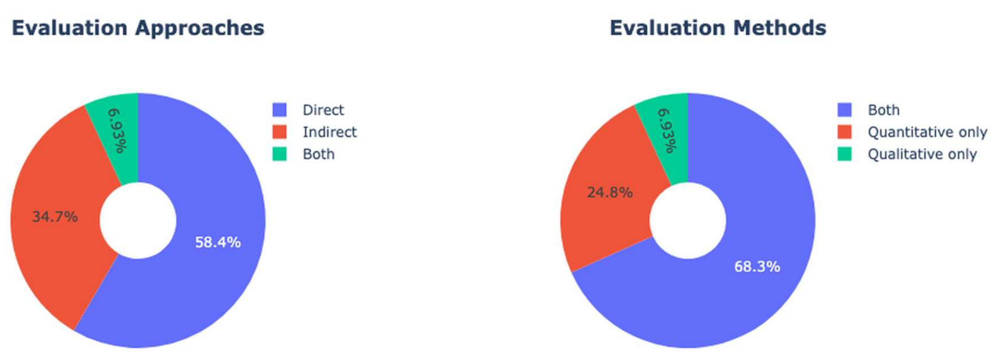

Fig. 2. **(L)** Breakdown of Evaluation Approaches into Direct vs Indirect Evaluation, and **(R)** Breakdown of Evaluation Methods into Quantitative vs Qualitative Methods

standard metrics and often include context-driven or subjective evaluations. We found that Direct approaches are the most common at 58.4%. About 34.7% of the publications use Indirect approaches. Additionally, 6.93% use both Direct and Indirect approaches in conjunction, to perform a holistic evaluation of their synthetic data (Fig. 2).

Quantitative evaluation methods give objective, measurable results and are crucial for ensuring that synthetic data aligns statistically with real data. They include **Statistical techniques** which use quantifiable metrics to compare synthetic data with the original data, and **ML-based techniques** which make use of classification and/or regression to assess how well synthetic data performs when used for specific downstream tasks. We found that Statistical evaluation is the most popular (54.5%), whereas ML-based techniques feature in 13.9% of all publications. About 24.8% publications use both Statistical and ML-based evaluation techniques in conjunction. (Fig. 3).

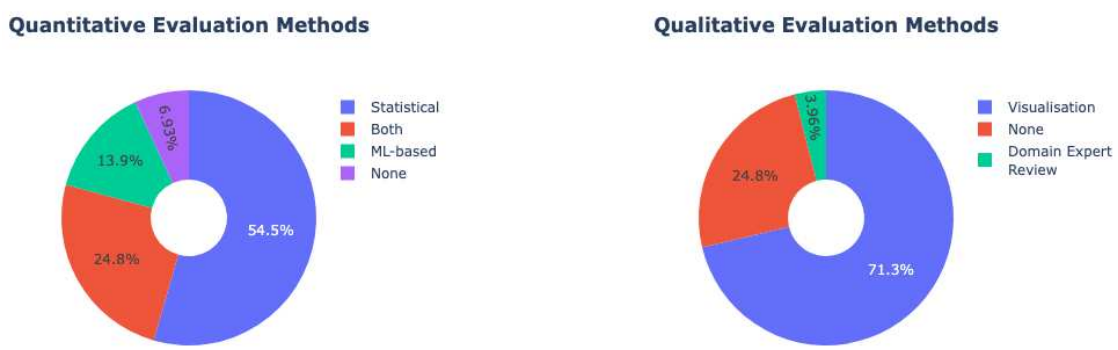

Fig. 3. **(L)** Breakdown of Quantitative Evaluation Methods into Statistical and ML-based methods, and **(R)** Breakdown of Qualitative Evaluation Methods into Visualisation and Domain Expert Review

The popularity of Statistical evaluation techniques remains high, however, a trend can be seen in publications using both Statistical and ML-based techniques together (Fig. 4).

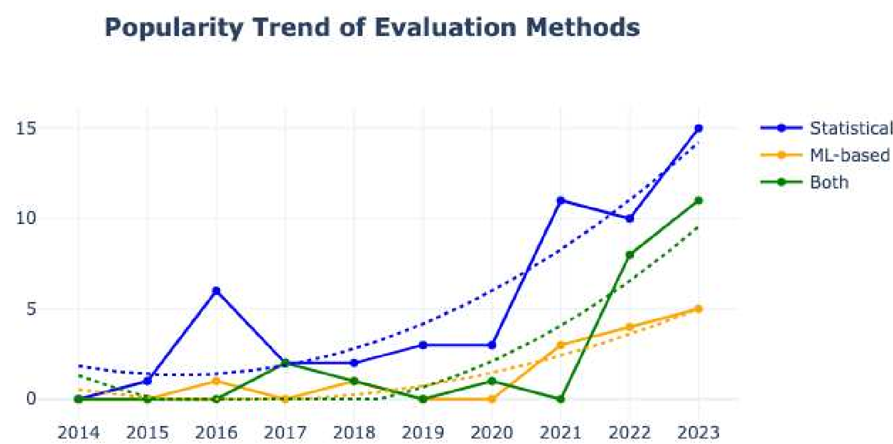

Fig. 4. Popularity trend of Statistical and ML-based Evaluation methods over the last decade, as obtained from publications included in this review. Dotted lines represent the overall polynomial trend.

The most used quantitative evaluation metrics include Jensen-Shannon (JS) distance, Pearson Correlation coefficient, and Maximum Mean Discrepancy (MMD), apart from the popular metrics such as AUC and F1-score for classification tasks, and Mean Square Error (MSE) and Root Mean Square Error (RMSE) for regression tasks (Fig. 5). We also note that the majority of the included papers (89.1%) use existing metrics, and only 10.9% of the publications use their own Author-defined metric for evaluation of the synthetic data.

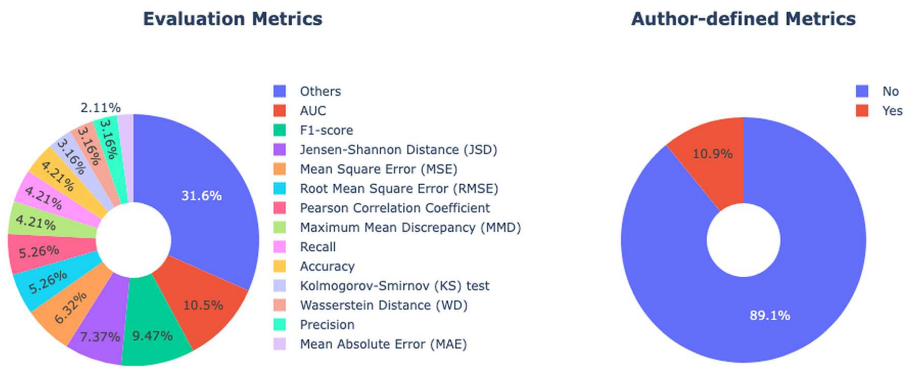

Fig. 5. **(L)** Most popular Evaluation Metrics, and **(R)** Breakdown of whether the metric is author-defined or not

Qualitative evaluation methods rely on subjective judgement and human interpretation to assess the quality of synthetic data. In the majority of the cases (71.3%), they take the form of **Visual Inspection** of the graphical representation of distributions of synthetic data (Fig. 3). We also note that despite its significance, use of **Domain Experts** as a qualitative evaluation method is not yet widely used and present only in a handful of all papers (3.96%).

Quantitative methods by themselves are used in about 24.8% of all publications, whereas for qualitative, this value is 5.93%. Most research (68.3%) uses a combination of both quantitative and qualitative methods. Fig. 6 gives a more detailed depiction of the most popular evaluation metrics and the papers utilising them that have been included in this review.

2.2 Generation of Synthetic Data

We categorise the models used for the generation of synthetic data into probabilistic and mechanistic models.

Probabilistic Models use statistical and probability distribution approaches to capture the statistical properties (such as distribution, correlations, and relationships between variables) of the real data, to generate the synthetic data. **Mechanistic Models**, on the other hand, use explicit rules, equations, or processes to simulate data based on how the underlying systems work. The most popular generation models are based on GANs, SMOTE, VAEs, Markov Chains, and Random Permutations (Fig. 7). Diffusion-based models are seeing a rise in popularity for longitudinal tabular data.

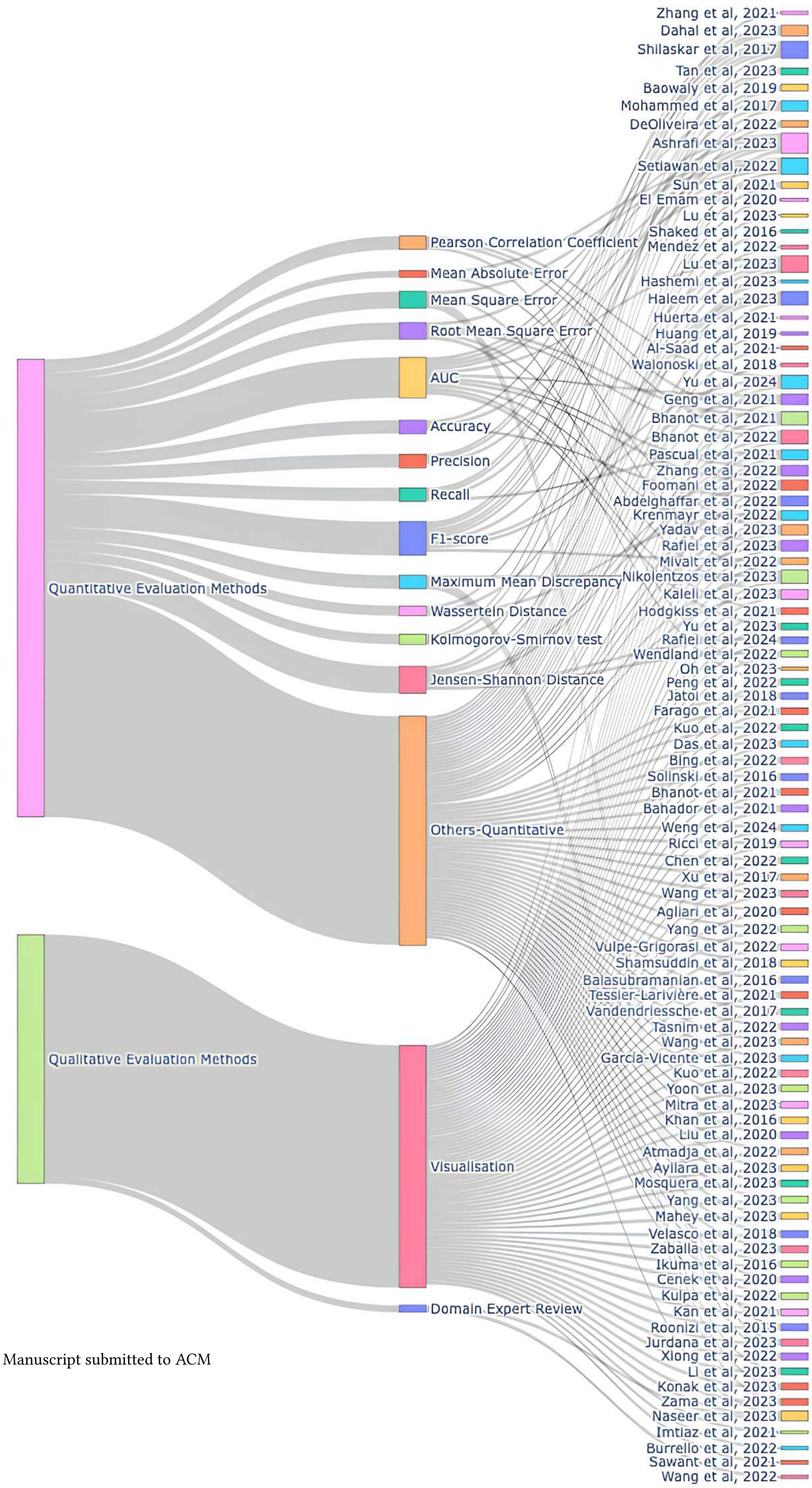

Fig. 6. Sankey depicting the most popular Evaluation Metrics and the papers utilising them, as obtained from the publications included in this review

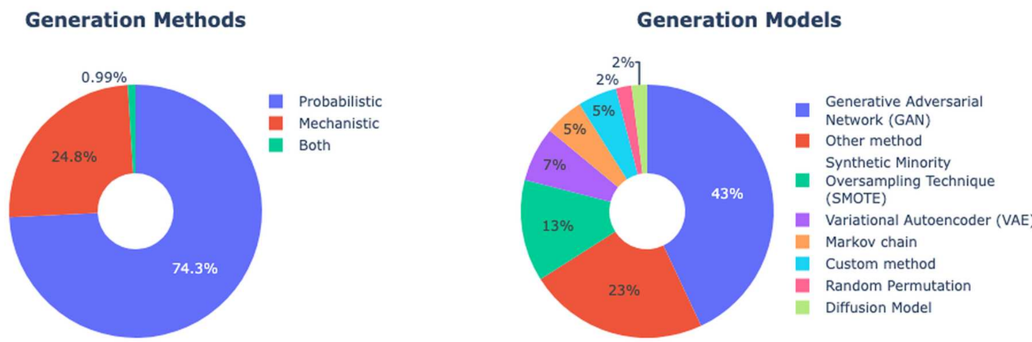

Fig. 7. **(L)** Breakdown of Generation Models into Probabilistic vs Mechanistic Models, and **(R)** Most popular Generation Models

We also observed a growing divergence between Probabilistic and Mechanistic models, with Probabilistic Models increasingly being more frequently used (74.3%) and Mechanistic Models tending to be more referenced in older publications only (Fig. 8).

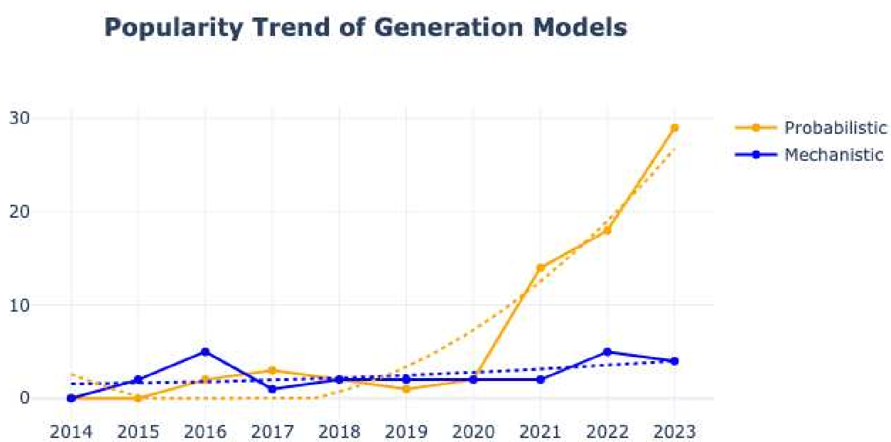

Fig. 8. Popularity trend of Probabilistic and Mechanistic Generation Models over the last decade, as obtained from publications included in this review. Dotted lines represent the overall polynomial trend.

2.3 Purpose and Impact of Synthetic Data

We found that privacy preservation, followed by predictive modelling and data-quality enhancement are the most popular objectives for the use of synthetic tabular data in healthcare. The most common diseases within the set of publications included in this review, for which synthetic health data is used, include those of the circulatory system, the nervous system, and neoplasms (Fig. 9). This may be driven by the popularity of the datasets, with MIMIC III[20] and MIMIC IV[21] being the most popular tabular and time-series health datasets.

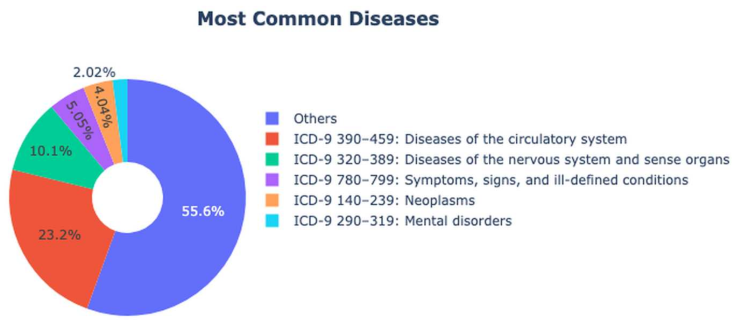

Fig. 9. Most common diseases (grouped by their ICD-9 codes), as seen through the publications included in this review.

The widest-used repository is Physionet[4] where these datasets are held, a testament to the value this repository provides to the research community. At the same time, this also poses a risk of perpetuating possible biases in the Physionet resources itself (for example, underrepresentation bias) to a global user base.

While there is some evidence to suggest that clinically oriented journals are also beginning to consider synthetic health data research, the majority of the articles are published in journals and conferences with a primarily technical focus. Most of the research in synthetic health data is carried out in North America, followed by Europe and Asia, which may be influenced by the availability of health data and data protection regulations necessitating the use of synthetic data.

2.4 Reproducibility of Results

We define reproducibility as a factor of the use of publicly available datasets and the reporting of publicly accessible source code of the research. We note that, despite being a crucial piece of information, reproducibility is often not emphasised, and only 21.78% of all included publications were reproducible according to our definition. The majority of the papers (63.36%) use publicly available real-world health datasets, whereas about 12.87% used paid real-world health datasets for the creation of synthetic data. However, only 24.75% of the publications give details about their code along with a link to the code repository, which affects the overall reproducibility. Reproducibility is essential for ensuring the reliability and impact of scientific research. Particularly in healthcare where decisions can directly affect patient outcomes, reproducibility helps prevent errors, biases, and misleading conclusions. However, we found that it is an often overlooked aspect in most publications dealing with synthetic data.

In response to these results, we devise a set of reporting guidelines on the generation and evaluation of synthetic data, which are henceforth described in Section 3.

3 EVALUATION GUIDELINES FOR SYNTHETIC DATA

(1) **Standardised Evaluation of Synthetic data:** We found that synthetic data are sometimes used without a thorough assessment. When there is an assessment, we found that not only there is no consensus on the evaluation methods, but the chosen evaluation metrics are inconsistently applied. This makes an operational assessment of the entire process unreliable, thereby making it difficult to track state-of-the-art advancements and creates barriers to trust and the adoption of synthetic data.
For example, using Mean Square Error (MSE) metric as a measure of distortion to assess the validity of a synthetically generated waveform signal is appropriate but the validation needs to consider the particular

[4]https://physionet.org/

features of the process. Two synthetic data generation methods with comparable MSE performances can still yield qualitatively very different signal features. Given the averaging that is implicitly performed in the computation of MSE, synthetic signals with uniformly distributed errors or distortions with respect to the real data will obtain similar MSE scores in comparison to signals with localised distortion patterns such as those modelled as impulsive noise. The signal characteristics in these two cases provide a stark contrast, and therefore, they will yield different conclusions in general when applied for validation or predictive tasks.

(2) **Better Reporting of Dataset Characteristics:** Poor reporting of dataset details is a cause for concern since the type of data (such as categorical or continuous) and their distribution, significantly impact the quality of the generated synthetic data. Furthermore, potential biases may propagate in the synthetic data. We recommend improved reporting of dataset characteristics used for the generation of synthetic data.
For example, in developing a novel method, one would expect to assess its performance using real data or previously validated synthetic data. However, using synthetic data that has not been validated prior to the assessment of the proposed method does not provide robust evidence towards the validity of the method. Moreover, any claims about the validity of the synthetic data based on the performance of the proposed method are inherently inconclusive as they incur a circular reference problem that compromises the generated evidence.

(3) **Prioritisation of Reproducibility of Results:** We emphasise the importance of clearly described evaluation metrics used, normalisation and aggregation of real data that were used as training datasets, and the entire experimental setup for the generation of synthetic data, including the chosen hyperparameters and the source code where possible. Reproducibility allows other researchers to validate and verify the claims of a study, and stakeholders, including clinicians and patients to develop trust in synthetic health data.

4 CONCLUSION

The potential of synthetic data to revolutionise Health AI research is immense, offering opportunities to address data scarcity, enhance privacy, and enable more robust model development. However, realising this potential requires a concerted effort to address critical challenges. Ensuring the applicability and fitness of synthetic data through rigorous assessment is paramount for its responsible use. Additionally, transparent reporting of datasets and ensuring reproducibility of results remain a cornerstone requirement of scientific progress.

Most importantly, the evaluation of synthetic health data must be guided by expert knowledge. As discussed earlier, domain expertise is critical for understanding the nuances of healthcare data and ensuring that synthetic datasets are technically sound, clinically relevant, and meaningful.

By adopting these guidelines and committing to ongoing collaboration, the health AI community can ensure that synthetic data is leveraged effectively and ethically, ultimately driving innovation and improving patient outcomes.

5 METHODOLOGY

In this section, we expand on our methodology for carrying out the systematic review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [22] statement provides a standardised framework for reporting systematic reviews. It consists of updated instructions on identifying, selecting, praising, and synthesising publications. Fig.10 outlines our methodology in accordance with the latest PRISMA guidelines.

5.1 Search Strategy

To establish an unambiguous search strategy, we laid out the following: **(i.)** The relevant databases used to search within a time frame, **(ii.)** The search terms to ensure comprehensive coverage of relevant studies, and, **(iii.)** The inclusion and exclusion criteria.

5.1.1 Databases and Search Engines. For this systematic review, we looked for publications on the following five databases: Scopus[5], Web of Science[6], PubMed[7], IEEE Xplore[8], and Association for Computing Machinery (ACM)[9]. Additional publications were also manually selected using Google Scholar.

5.1.2 Search Terms and Additional Limits.

Search Terms. We began by identifying publications dealing with synthetic data generation or augmentation. The focus was on finding publications that dealt with tabular or time-series data using wildcards in the search, and included all similar words including plurals and noun and verb forms of words. Keywords such as 'patient', 'health', and 'clinical' were also used to conduct the search within the health domain. As a result of the aforementioned considerations, the following search string was designed: *(synthe* OR augment*) AND generat* AND (time-series OR time* OR temporal*) AND (tabular OR record*) AND (patient* OR medic* OR health* OR clinic* OR ehr*)*, to be searched in the title-abstract-keywords or the topic field.

Additional Limits. We limited our search to publications in the last ten years, from January 1, 2014 to Jan 31, 2024. We also limited the search to peer-reviewed conferences and journal articles, written in the English language.

5.1.3 Inclusion and Exclusion Criteria.

Inclusion Criteria. The publications that were included in this systematic review met the following conditions:

- Publications that deal with tabular or time-series data.
- Publications that describe a method of generation of synthetic data and its evaluation against real data, or Publications that do not describe a method of generation of synthetic data but describe its evaluation vis a vis real data.
- Publications that deal with the generation of complete new synthetic datasets as well as the ones which deal with the augmentation of existing datasets with synthetic data.
- Peer-reviewed publications from journals and conferences. Strictly no pre-prints.

Exclusion Criteria. All possible publications that would be irrelevant to our study were excluded if they met any one of the following conditions:

- Publications that are not in the health domain.
- Publications that deal with image-, audio-, video-, or text-only modalities of data.
- Publications which themselves are narrative or systematic reviews.
- Publications on synthetic data that neither describe a method of generation of synthetic data nor its evaluation against real data.

[5]https://www.scopus.com
[6]https://www.webofscience.com
[7]https://pubmed.ncbi.nlm.nih.gov
[8]https://ieeexplore.ieee.org
[9]https://www.acm.org

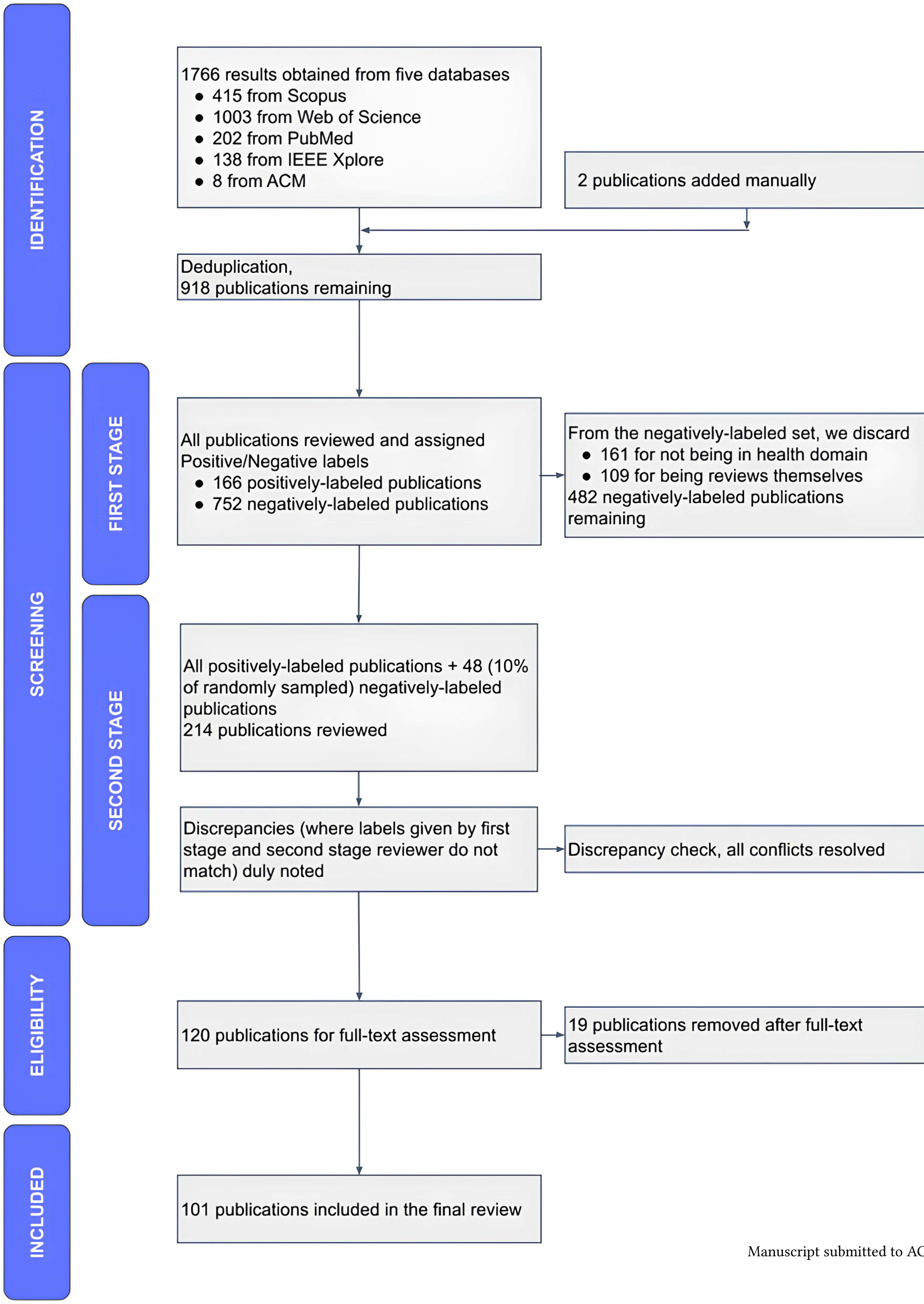

Fig. 10. Popularity trend of synthetic tabular health data over the last decade, as obtained from publications included in this review

- Publications that describe a method of generation of synthetic data, but the synthetic data is not structurally similar (and therefore, not comparable) to real data.

5.2 Search Queries

Table 3 lists the search queries used across five databases respectively, and the number of relevant results obtained from each. Another set of limits on search engines were enforced via their graphical user interfaces (GUIs), hence they may have not been fully captured in the queries themselves. They are being mentioned here for the sake of completeness.

- The search was limited to publications in the last **ten years**, from January 1, 2014, to January 31, 2024.
- The search was limited to **peer-reviewed** conferences and journal articles written in the **English** language.

5.3 Selection Process

Based on the search strategy discussed in Section 5.1, 1766 publications were obtained from five search engines. This included 415 publications from Scopus, 1003 from the Web of Science, 202 from PubMed, 138 from IEEE Xplore, and 8 from ACM. Since many results were duplicated across databases, we carried out a deduplication process based on the DOIs of publications. As a result, 848 results were excluded and 918 publications remained for the first stage of assessment.

First Screening. The 918 publications obtained from search queries after deduplication, were divided among five reviewers with approximately 184 publications per reviewer. Each reviewer labelled the publications assigned to them Yes/No, signifying whether they thought the publication should be included or excluded along with the rationale for their decision. This gave us 166 publications labelled Yes, and 752 publications labelled No. From the 'No' set, we again discarded some publications, the most common reasons for their exclusion being that the research was not in the human health domain (n=161) or that the paper was a systematic review itself (n=109).

Next, we created a smaller subset of approximately 10% of the remaining No-labelled publications (482) and added it to all the Yes-labelled set. This combined set of 214 publications was used to perform a 'spot check' of labels: Publications were grouped by their first-stage reviewers and divided among the rest of the reviewers for a second round of reviewing. The first-stage reviewer's decision to include or exclude any particular publication was preserved but kept hidden from the view of the second-stage reviewers, to ensure that the reviewers' cognitive biases do not creep in during the labelling process, and that every publication gets assigned the correct label irrespective of who it was reviewed by in either of the two reviewing stages.

Second Screening. As with the first stage, each reviewer in the second stage provided a Yes or a No label to each publication in their set. No reviewer got to review the same publication in the second stage which they had already reviewed in the first stage. At the end of this exercise, any 'discrepancies' (cases where the labels given by the first-stage reviewer and the second-stage reviewer did not match) were duly noted. Then, a round of 'discrepancy checks' was carried out, where all reviewers looked at all discrepancies and provided feedback as to which of the two labels they agree with.

After a discussion on each occurrence of conflicting labels and resolving all discrepancies, we got the final labels for each publication. This resulted in a set of 120 publications for which the consensus of the reviewers was to include them in this systematic review.

Category	Data Items
Publication Details	DOI, Authors, Title, Publication Year, Journal/Conference Title
Synthetic Data Generation	Category of Generation Model (Mechanistic, Probabilistic), Name of Generation Method used, Purpose of Generation of synthetic data Disease/Disorder focused on, ICD-9 code
Synthetic Data Evaluation	Category of Quantitative Evaluation Method (ML-based, Statistical, Both, None), Category of Qualitative Evaluation Method (Visualisation, Others, None), Name of Evaluation Method used
Training Dataset Characteristics	Name, Size, Institution of Origin, Country of Origin, Visibility (Public/Private), Cost of Dataset Access
Source Code	Link to Source Code Repository

Table 1. Attributes against which data was collected for every publication

5.4 Data Items

For each of the 120 publications, an in-depth analysis was carried out. An additional 27 publications were excluded from the study upon full-text analysis, based on their relevance to this research.

Then, data was collected for the final 101 publications for 18 attributes which included: **(i.)** details about the publication including DOI, authors' names, title and year of publication and the details about the venue (journal/conference), **(ii.)** details specific to the generation of synthetic data such as the method (eg. WGAN-GP, Graph VAEs), its category (mechanistic/probabilistic), and purpose (eg. privacy preservation, clinical trial simulation), **(iii.)** details about the evaluation methods used, which includes the type of Quantitative evaluation (ML-based, Statistical or a combination), the type of Qualitative evaluation (for eg. Visualisation), the name of the method (eg. Jensen-Shannon divergence, Wasserstein distance) and the specific evaluation metrics used, **(iv.)** details about the dataset used in synthetic data generation, including the dataset name, size, institution and country of origin, and cost of dataset access, and **(v.)** details pertaining to the reproducibility of results including whether the dataset is openly accessible and if the source code has been made available.

A complete list of all data items against which data was captured is available in Table 1.

5.5 Reporting

The reporting of this systematic review adheres to the PRISMA guidelines [22]. We undertook measures to ensure transparency and reproducibility and facilitate critical appraisal and interpretation of the findings.

Risk of Bias Management: To establish the transparency of the findings and the results of this systematic review, we: (i.) used multiple databases to ensure no platform-specific bias creeps in, (ii.) used value-neutral search terms in the search query (iii.) got the publications reviewed by five reviewers in multiple screening stages (iv.) performed spot checks and discrepancy checks to ensure no reviewer-induced bias creeps in.

REFERENCES

[1] Maryam Tayefi, Phuong D. Ngo, Taridzo Chomutare, Hercules Dalianis, Elisa Salvi, Andrius Budrionis, and Fred Godtliebsen. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13, 2021.

[2] Antonio Cruz, Samantha Marshall, Christine Daum, Hector Perez, John Hirdes, and Lili Liu. Data silos undermine efforts to characterize, predict, and mitigate dementia-related missing person incidents. *Healthcare Management Forum*, 35:084047042211061, 06 2022.

[3] Rebecca Asiimwe, Stephanie Lam, Samuel Leung, Shanzhao Wang, Rachel Wan, Anna Tinker, Jessica Mcalpine, Michelle Woo, David Huntsman, and Aline Talhouk. From biobank and data silos into a data commons: convergence to support translational medicine, 08 2021.

[4] Hippolyte Lefebvre, Christine Legner, and Martin Fadler. Data democratization: toward a deeper understanding. 09 2021.

[5] Filipe Bernardi, Domingos Alves, Nathalia Crepaldi, Diego Yamada, Vinícius Lima, and Rui Rijo. Data quality in health research: an integrative literature review (preprint). *Journal of Medical Internet Research*, 25, 08 2022.

[6] Julie-Anne Smit, Menno Mostert, and Johannes Delden. Protecting privacy while optimizing the use of (health)data: The importance of measures and safeguards. *The American Journal of Bioethics*, 22:79–81, 07 2022.

[7] Reg Joseph, Antonio Bruni, and Chris Carvalho. Health city: Transforming health and driving economic development. *Healthcare Management Forum*, 34:084047042094226, 08 2020.

[8] Theodora Kokosi and Katie Harron. Synthetic data in medical research. *BMJ Medicine*, 1, 09 2022.

[9] Mauro Giuffrè and Dennis Shung. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *npj Digital Medicine*, 6, 10 2023.

[10] Nicolo Micheletti, Raffaele Marchesi, Nicholas I-Hsien Kuo, Sebastiano Barbieri, Giuseppe Jurman, and Venet Osmani. Generative ai mitigates representation bias using synthetic health data. *medRxiv*, 2024.

[11] Nicolo Micheletti, Raffaele Marchesi, Nicholas Kuo, Sebastiano Barbieri, Giuseppe Jurman, and Venet Osmani. Generative ai mitigates representation bias using synthetic health data, 09 2023.

[12] Enrico Barbierato, Marco Della Vedova, Daniele Tessera, Daniele Toti, and Nicola Vanoli. A methodology for controlling bias and fairness in synthetic data generation. *Applied Sciences*, 12:4619, 05 2022.

[13] Boris Breugel, Trent Kyono, Jeroen Berrevoets, and Mihaela Schaar. Decaf: Generating fair synthetic data using causally-aware generative networks, 10 2021.

[14] Antonio J. Rodríguez-Almeida, Himar Fabelo, Samuel Ortega, Alejandro Deniz, Francisco J. Balea-Fernandez, Eduardo Quevedo, Cristina Soguero-Ruíz, Ana Maria Wägner, and Gustavo Marrero Callicó. Synthetic patient data generation and evaluation in disease prediction using small and imbalanced datasets. *IEEE Journal of Biomedical and Health Informatics*, 27:2670–2680, 2022.

[15] James Jordon, Jinsung Yoon, and Mihaela Schaar. Measuring the quality of synthetic data for use in competitions, 06 2018.

[16] Ahmed Alaa, Boris van Breugel, Evgeny Saveliev, and Mihaela Schaar. How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models, 02 2021.

[17] Mikel Hernandez, Gorka Epelde, Ane Alberdi, Rodrigo Cilla, and Debbie Rankin. Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 493, 04 2022.

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 12 2017.

[19] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert, 04 2019.

[20] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[21] Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1, 2023.

[22] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, 372, 2021.

[23] M. A. Jatoi and Nidal S. Kamel. Brain source localization using reduced eeg sensors. *Signal, Image and Video Processing*, 12:1447 – 1454, 2018.

[24] Emma Farago and Adrian Chan. Motion artifact synthesis for research in biomedical signal quality analysis. *Biomedical Signal Processing and Control*, 68:102611, 07 2021.

[25] Chih-En Kuo, Tsung-Hua Lu, Guan-Ting Chen, and Po-Yu Liao. Towards precision sleep medicine: Self-attention gan as an innovative data augmentation technique for developing personalized automatic sleep scoring classification. *Computers in Biology and Medicine*, 148:105828, 2022.

[26] Alireza Rafiei, Milad Ghiasi Rad, Andrea Sikora, and Rishikesan Kamaleswaran. Improving irregular temporal modeling by integrating synthetic data to the electronic medical record using conditional gans: a case study of fluid overload prediction in the intensive care unit. *medRxiv*, pages 2023–06, 2023.

[27] Joshua DeOliveira, Walter Gerych, Aruzhan Koshkarova, Elke Rundensteiner, and Emmanuel Agu. Har-ctgan: a mobile sensor data generation tool for human activity recognition. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 5233–5242. IEEE, 2022.

[28] Trisha Das, Zifeng Wang, and Jimeng Sun. Twin: Personalized clinical trial digital twin generation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 402–413, 2023.

[29] Simon Bing, Andrea Dittadi, Stefan Bauer, and Patrick Schwab. Conditional generation of medical time series for extrapolation to underrepresented populations. *PLOS Digital Health*, 1(7):e0000074, 2022.

[30] Karan Bhanot, Saloni Dash, Joseph Pedersen, Isabelle Guyon, and Kristin P Bennett. Quantifying resemblance of synthetic medical time-series. In *ESANN*, 2021.

[31] Khaled El Emam, Lucy Mosquera, and Jason Bass. Evaluating identity disclosure risk in fully synthetic health data: model development and validation. *Journal of medical Internet research*, 22(11):e23139, 2020.

[32] Ya-Ting Lu, Horng-Jiun Chao, Yi-Chun Chiang, and Hsiang-Yin Chen. Explainable machine learning techniques to predict amiodarone-induced thyroid dysfunction risk: multicenter, retrospective study with external validation. *Journal of Medical Internet Research*, 25:e43734, 2023.

[33] Mateusz Soliński, Jan Gierałtowski, and Jan Żebrowski. Modeling heart rate variability including the effect of sleep stages. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(2), 2016.

[34] Karan Bhanot, Miao Qi, John S Erickson, Isabelle Guyon, and Kristin P Bennett. The problem of fairness in synthetic healthcare data. *Entropy*, 23(9):1165, 2021.

[35] Sigal Shaked and Lior Rokach. Publishing differentially private medical events data. In *Availability, Reliability, and Security in Information Systems: IFIP WG 8.4, 8.9, TC 5 International Cross-Domain Conference, CD-ARES 2016, and Workshop on Privacy Aware Machine Learning for Health Data Science, PAML 2016, Salzburg, Austria, August 31-September 2, 2016, Proceedings*, pages 219–235. Springer, 2016.

[36] Nooshin Bahador, Guoying Zhao, Jarno Jokelainen, Seppo Mustola, and Jukka Kortelainen. Morphology-preserving reconstruction of times series with missing data for enhancing deep learning-based classification. *Biomedical Signal Processing and Control*, 70:103052, 2021.

[37] Xutao Weng, Hong Song, Yucong Lin, You Wu, Xi Zhang, Bowen Liu, and Jian Yang. A joint learning method for incomplete and imbalanced data in electronic health record based on generative adversarial networks. *Computers in Biology and Medicine*, 168:107687, 2024.

[38] Karan Bhanot, Joseph Pedersen, Isabelle Guyon, and Kristin P Bennett. Investigating synthetic medical time-series resemblance. *Neurocomputing*, 494:368–378, 2022.

[39] Giannis Nikolentzos, Michalis Vazirgiannis, Christos Xypolopoulos, Markus Lingman, and Erik G Brandt. Synthetic electronic health records generated with variational graph autoencoders. *NPJ Digital Medicine*, 6(1):83, 2023.

[40] Leonardo Ricci, Michele Castelluzzo, Ludovico Minati, and Alessio Perinelli. Generation of surrogate event sequences via joint distribution of successive inter-event intervals. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12), 2019.

[41] Jack Hodgkiss, Soufiene Djahel, and Zonghua Zhang. A new attack method against ecg-based key generation and agreement schemes in body area networks. *IEEE Sensors Journal*, 21(15):17300–17307, 2021.

[42] Cai Chen, Fulai Peng, Yue Sun, Danyang Lv, Ningling Zhang, Xingwei Wang, and Lin Wang. Epileptic seizure prediction based on eeg by auto-machine learning. In *2022 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 710–715. IEEE, 2022.

[43] Hongteng Xu, Weichang Wu, Shamim Nemati, and Hongyuan Zha. Patient flow prediction via discriminative learning of mutually-correcting processes. *IEEE transactions on Knowledge and Data Engineering*, 29(1):157–171, 2016.

[44] Ameer Mohammed, Majid Zamani, Richard Bayford, and Andreas Demosthenous. Toward on-demand deep brain stimulation using online parkinson's disease prediction driven by dynamic detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2441–2452, 2017.

[45] Xianglong Wang, Berkman Sahiner, Christopher G Scully, and Kenny H Cha. Afe-gan: Synthesizing electrocardiograms with atrial fibrillation characteristics using generative adversarial networks. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–5. IEEE, 2023.

[46] Sana Imtiaz, Muhammad Arsalan, Vladimir Vlassov, and Ramin Sadre. Synthetic and private smart health care data generation using gans. In *2021 International Conference on Computer Communications and Networks (ICCCN)*, pages 1–7. IEEE, 2021.

[47] Navid Ashrafi, Vera Schmitt, Robert P Spang, Sebastian Möller, and Jan-Niklas Voigt-Antons. Protect and extend-using gans for synthetic data generation of time-series medical records. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 171–176. IEEE, 2023.

[48] Alireza Rafiei, Milad Ghiasi Rad, Andrea Sikora, and Rishikesan Kamaleswaran. Improving mixed-integer temporal modeling by generating synthetic data using conditional generative adversarial networks: A case study of fluid overload prediction in the intensive care unit. *Computers in Biology and Medicine*, 168:107749, 2024.

[49] Elena Agliari, Francesco Alemanno, Adriano Barra, Orazio Antonio Barra, Alberto Fachechi, Lorenzo Franceschi Vento, and Luciano Moretti. Analysis of temporal correlation in heart rate variability through maximum entropy principle in a minimal pairwise glassy model. *Scientific Reports*, 10(1):15353, 2020.

[50] Jin Li, Benjamin J Cairns, Jingsong Li, and Tingting Zhu. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. *NPJ Digital Medicine*, 6(1):98, 2023.

[51] Ziqi Zhang, Chao Yan, and Bradley A Malin. Keeping synthetic patients on track: feedback mechanisms to mitigate performance drift in longitudinal health data simulation. *Journal of the American Medical Informatics Association*, 29(11):1890–1898, 2022.

[52] In-Sun Oh, Han Eol Jeong, Hyesung Lee, Kristian B Filion, Yunha Noh, and Ju-Young Shin. Validating an approach to overcome the immeasurable time bias in cohort studies: a real-world example and monte carlo simulation study. *International Journal of Epidemiology*, 52(5):1534–1544, 2023.

[53] Alessio Burrello, Daniele Jahier Pagliari, Marzia Bianco, Enrico Macii, Luca Benini, Massimo Poncino, and Simone Benatti. Improving ppg-based heart-rate monitoring with synthetically generated data. In *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 153–157. IEEE, 2022.

[54] Vincent Mendez, Clément Lhoste, and Silvestro Micera. Emg data augmentation for grasp classification using generative adversarial networks. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3619–3622. IEEE, 2022.

[55] Zhenyi Yang, Rebecca Miao, Marina Orlova, Ivan Nechepurenko, and Valeriy Gavrishchaka. Discovery of early-alert indicators using hybrid ensemble learning and generative physics-based models. In *2022 5th International Conference on Information and Computer Technologies (ICICT)*, pages 224–232. IEEE, 2022.

[56] Adrian Vulpe-Grigoraşi and Ovidiu Grigore. Gan based synthetic ecg for psychological stress. In *2022 E-Health and Bioengineering Conference (EHB)*, pages 1–4. IEEE, 2022.

[57] Rittika Shamsuddin, Barbara M Maweu, Ming Li, and Balakrishnan Prabhakaran. Virtual patient model: an approach for generating synthetic healthcare time series data. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 208–218. IEEE, 2018.

[58] Arvind Balasubramanian, Jun Wang, and Balakrishnan Prabhakaran. Discovering multidimensional motifs in physiological signals for personalized healthcare. *IEEE journal of selected topics in signal processing*, 10(5):832–841, 2016.

[59] Olivier Tessier-Larivière, Luke Y Prince, Pascal Fortier-Poisson, Lorenz Wernisch, Oliver Armitage, Emil Hewage, Guillaume Lajoie, and Blake A Richards. Pns-gan: Conditional generation of peripheral nerve signals in the wavelet domain via adversarial networks. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 778–782. IEEE, 2021.

[60] Benjamin Vandendriessche, Mustafa Abas, Thomas E Dick, Kenneth A Loparo, and Frank J Jacono. A framework for patient state tracking by classifying multiscalar physiologic waveform features. *IEEE Transactions on Biomedical Engineering*, 64(12):2890–2900, 2017.

[61] Nabila Tasnim, Joyita Halder, Shahed Ahmed, and Shaikh Anowarul Fattah. An approach for analyzing cognitive behavior of autism spectrum disorder using p300 bci data. In *2022 IEEE Region 10 Symposium (TENSYMP)*, pages 1–6. IEEE, 2022.

[62] Winston Wang and Tun-Wen Pai. enhancing small tabular clinical trial dataset through hybrid data augmentation: combining smote and wcgan-gp. *Data*, 8(9):135, 2023.

[63] Clara García-Vicente, David Chushig-Muzo, Inmaculada Mora-Jiménez, Himar Fabelo, Inger Torhild Gram, Maja-Lisa Løchen, Conceição Granja, and Cristina Soguero-Ruiz. Evaluation of synthetic categorical data generation techniques for predicting cardiovascular diseases and post-hoc interpretability of the risk factors. *Applied Sciences*, 13(7):4119, 2023.

[64] Nicholas I-Hsien Kuo, Mark N Polizzotto, Simon Finfer, Federico Garcia, Anders Sönnerborg, Maurizio Zazzi, Michael Böhm, Rolf Kaiser, Louisa Jorm, and Sebastiano Barbieri. The health gym: synthetic health-related datasets for the development of reinforcement learning algorithms. *Scientific data*, 9(1):693, 2022.

[65] Jinsung Yoon, Michel Mizrahi, Nahid Farhady Ghalaty, Thomas Jarvinen, Ashwin S Ravi, Peter Brune, Fanyu Kong, Dave Anderson, George Lee, Arie Meir, et al. Ehr-safe: generating high-fidelity and privacy-preserving synthetic electronic health records. *NPJ Digital Medicine*, 6(1):141, 2023.

[66] Anumita Mitra, Palash Kumar Kundu, Rajarshi Gupta, Jayanta Saha, and Arunansu Talukdar. Cardiosim: a pc-based cardiac signal simulator using segmental modeling of electrocardiogram. *Computer Methods in Biomechanics and Biomedical Engineering*, 26(13):1532–1548, 2023.

[67] Naveed Khan, Sally McClean, Shuai Zhang, and Chris Nugent. Using genetic algorithms for optimal change point detection in activity monitoring. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 318–323. IEEE, 2016.

[68] Ebadollah Kheirati Roonizi, Massimo W Rivolta, Luca T Mainardi, and Roberto Sassi. A comparison of three methodologies for the computation of v-index. In *2015 Computing in Cardiology Conference (CinC)*, pages 593–596. IEEE, 2015.

[69] Orhan Konak, Lucas Liebe, Kirill Postnov, Franz Sauerwald, Hristijan Gjoreski, Mitja Luštrek, and Bert Arnrich. Overcoming data scarcity in human activity recognition. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–7. IEEE, 2023.

[70] Huaning Tan, Renxing Chen, Meng Qin, Lining Tang, Zhibing Wu, Qianlin Luo, and Yujuan Quan. Tabular gan-based oversampling of imbalanced time-to-event data for survival prediction. In *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pages 376–380. IEEE, 2023.

[71] Ruoqi Liu, Changchang Yin, and Ping Zhang. Estimating individual treatment effects with time-varying confounders. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 382–391. IEEE, 2020.

[72] Hendrico Yehezky Nata Atmadja, Alhadi Bustamam, et al. Generated tabular data with multi-gans for arrhythmia classification based on dnn models. In *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, pages 69–74. IEEE, 2022.

[73] Atiye Sadat Hashemi, Kobra Etminani, Amira Soliman, Omar Hamed, and Jens Lundström. Time-series anonymization of tabular health data using generative adversarial network. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2023.

[74] Xiaoxia Wang, Chun-Pong Tang, Yanming He, Joseph M. Plasek, Yun Xiong, Yangyong Zhu, D. Bates, and Li Zhou. Using an optimized generative model to infer the progression of complications in type 2 diabetes patients. *BMC Medical Informatics and Decision Making*, 22, 2022.

[75] Álvaro Huerta, Arturo Martínez-Rodrigo, José Joaquín Rieta, and Raúl Alcaraz. Ecg quality assessment via deep learning and data augmentation. *2021 Computing in Cardiology (CinC)*, 48:1–4, 2021.

[76] Olawale F. Ayilara, Robert W. Platt, Matt Dahl, Janie Coulombe, Pablo Gonzalez Ginestet, Dan Chateau, and Lisa M. Lix. Generating synthetic data from administrative health records for drug safety and effectiveness studies. *International Journal of Population Data Science*, 8, 2023.

[77] Muhammad Salman Haleem, Audrey Ekuban, Alessio Antonini, Silvio Marcello Pagliara, Leandro Pecchia, and Carlo Allocca. Deep-learning-driven techniques for real-time multimodal health and physical data synthesis. *Electronics*, 2023.

[78] Md Haider Zama and Friedhelm Schwenker. Ecg synthesis via diffusion-based state space augmented transformer. *Sensors (Basel, Switzerland)*, 23, 2023.

[79] Lucy Mosquera, Khaled El Emam, Lei Ding, Vishal Sharma, Xue Hua Zhang, Samer El Kababji, Chris Carvalho, Brian Hamilton, Dan Palfrey, Linglong Kong, Bei Jiang, and Dean T. Eurich. A method for generating synthetic longitudinal health data. *BMC Medical Research Methodology*, 23, 2023.

[80] Ming Huang, Nilay D. Shah, and Lixia Yao. Evaluating global and local sequence alignment methods for comparing patient medical records. *BMC Medical Informatics and Decision Making*, 19, 2019.

[81] Kamana Dahal and Mohd. Hasan Ali. A hybrid gan-based dl approach for the automatic detection of shockable rhythms in aed for solving imbalanced data problems. *Electronics*, 2022.

[82] Febryan Setiawan and Che-Wei Lin. A deep learning framework for automatic sleep apnea classification based on empirical mode decomposition derived from single-lead electrocardiogram. *Life*, 12, 2022.

[83] Vedran Jurdana, Miroslav Vrankic, Nikola Lopac, and Guruprasad Madhale Jadav. Method for automatic estimation of instantaneous frequency and group delay in time–frequency distributions with application in eeg seizure signals analysis. *Sensors (Basel, Switzerland)*, 23, 2023.

[84] Priyanshu Mahey, Nima Toussi, Grace Purnomu, and Anthony T. Herdman. Generative adversarial network (gan) for simulating electroencephalography. *Brain Topography*, 36:661–670, 2023.

[85] José Manuel Velasco, Oscar Garnica, Juan Lanchares, Marta Botella, and José Ignacio Hidalgo. Combining data augmentation, edas and grammatical evolution for blood glucose forecasting. *Memetic Computing*, 10:267 – 277, 2018.

[86] Zhaohan Xiong, Martin K. Stiles, Anne M. Gillis, and Jichao Zhao. Enhancing the detection of atrial fibrillation from wearable sensors with neural style transfer and convolutional recurrent networks. *Computers in biology and medicine*, 146:105551, 2022.

[87] Lucas Krenmayr, Roland Frank, Christina Drobig, Michael Braungart, Jan Seidel, Daniel Schaudt, Reinhold von Schwerin, and Kathrin Stucke-Straub. Ganeraid: Realistic synthetic patient data for clinical trials. *Informatics in Medicine Unlocked*, 2022.

[88] Onintze Zaballa, Aritz Pérez Martínez, Elisa Gómez-Inhiesto, Teresa Acaiturri Ayesta, and José Antonio Lozano. Learning the progression patterns of treatments using a probabilistic generative model. *Journal of biomedical informatics*, 137:104271, 2022.

[89] Philipp Wendland, Colin Birkenbihl, Marc Gomez-Freixa, Meemansa Sood, Maik Kschischo, and Holger Fröhlich. Generation of realistic synthetic data using multimodal neural ordinary differential equations. *NPJ Digital Medicine*, 5, 2021.

[90] Hanamanth S Kaleli and Vasudev Dehalwar. Generation of synthetic ecg signal using generative adversarial network with transformers. *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6, 2023.

[91] Siao Sun, Fusheng Wang, Sina Rashidian, Tahsin M. Kurç, Kayley Abell-Hart, Janos G. Hajagos, Wei wei Zhu, Mary M. Saltz, and Joel H. Saltz. Generating longitudinal synthetic ehr data with recurrent autoencoders and generative adversarial networks. In *Poly/DMAH@VLDB*, 2021.

[92] Swati Shilaskar, Ashok A. Ghatol, and Prashant N. Chatur. Medical decision support system for extremely imbalanced datasets. *Inf. Sci.*, 384:205–219, 2017.

[93] Farnaz H. Foomani, D. M. Anisuzzaman, Jeffrey A. Niezgoda, Jonathan Niezgoda, William M. Guns, Sandeep Gopalakrishnan, and Zeyun Yu. Synthesizing time-series wound prognosis factors from electronic medical records using generative adversarial networks. *Journal of biomedical informatics*, page 103972, 2021.

[94] Tommy Peng, Avinash Malik, Laura R. Bear, and Mark L. Trew. Impulse data models for the inverse problem of electrocardiography. *IEEE Journal of Biomedical and Health Informatics*, 26:1353–1361, 2021.

[95] David Geng and Zhe Sage Chen. Auxiliary classifier generative adversarial network for interictal epileptiform discharge modeling and eeg data augmentation. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1130–1133, 2021.

[96] Filip Mivalt, Vladimir Sladky, Irena Balzekas, Tereza Pridalova, Kai J. Miller, Jamie J. Van Gompel, Timothy J. Denison, Benjamin H. Brinkmann, Václav Kremen, and Gregory A. Worrell. Deep generative networks for algorithm development in implantable neural technology. *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1736–1741, 2022.

[97] Damian Pascual, Ali John Amirshahi, Amir Aminifar, David Atienza Alonso, Philippe Ryvlin, and Roger Wattenhofer. Epilepsygan: Synthetic epileptic brain activities with privacy preservation. *IEEE Transactions on Biomedical Engineering*, 68:2435–2446, 2020.

[98] Chang de Lu, Chandan K. Reddy, Ping Wang, Dong Nie, and Yue Ning. Multi-label clinical time-series generation via conditional gan. *IEEE Transactions on Knowledge and Data Engineering*, 36:1728–1740, 2022.

[99] Takeshi Ikuma, Melda Kunduk, Daniel S Fink, and Andrew J. Mcwhorter. Synthetic multi-line kymographic analysis: A spatiotemporal data reduction technique for high-speed videoendoscopy. *The Journal of the Acoustical Society of America*, 140 4:2703, 2016.

[100] Ziqi Zhang, Chao Yan, Thomas A. Lasko, Jimeng Sun, and Bradley A. Malin. Synteg: a framework for temporal structured electronic health data simulation. *Journal of the American Medical Informatics Association : JAMIA*, 2020.

[101] Mrinal Kanti Baowaly, Chao-Lin Liu, and Kuan-Ta Chen. Realistic data synthesis using enhanced generative adversarial networks. *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 289–292, 2019.

[102] Yasmin Abdelghaffar, Ahmed Hashem, and Seif Eldawlatly. Generative adversarial networks for augmenting eeg data in p300-based applications: A comparative study. *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6, 2022.

[103] Mohammad Al-Saad, Madeleine Lucas, and Lakshmish Ramaswamy. Privacy vulnerabilities of wearable activity monitors: Threat and potential defence. *2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC)*, pages 105–116, 2021.

[104] Tian Yu, Boyuan Cui, Yaqian Xu, and Xilin Liu. Refine eeg spectrogram synthesized by generative adversarial network for improving the prediction of epileptic seizures*. *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1–4, 2023.

[105] Lisa Cenek, Liubou Klindziuk, Cindy Lopez, Eleanor McCartney, Blanca Martin Burgos, Selma Tir, Mary E. Harrington, and Tanya L. Leise. Circada: Shiny apps for exploration of experimental and synthetic circadian time series with an educational emphasis. *Journal of Biological Rhythms*, 35:214 – 222, 2020.

[106] Parul Yadav, Manish Gaur, Nishat Fatima, and Saqib Sarwar. Qualitative and quantitative evaluation of multivariate time-series synthetic data generated using mts-tgan: A novel approach. *Applied Sciences*, 2023.

[107] Zuyi Yu, Amar Kachenoura, Régine Le Bouquin-Jeannés, Huazhong Shu, Paul Berraute, Anca Nica, Isabelle Merlet, Laurent Albera, and Ahmad Karfoul. Electrophysiological brain imaging based on simulation-driven deep learning in the context of epilepsy. *NeuroImage*, 285, 2023.

[108] Jason A. Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, and Scott McLachlan. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association : JAMIA*, 25:230 – 238, 2017.

[109] John Kulpa, Emma Farago, and Adrian D. C. Chan. A toolkit for motion artifact signal generation. *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2022.

[110] Chi Nok Enoch Kan, Richard J. Povinelli, and Dong Hye Ye. Enhancing multi-channel eeg classification with gramian temporal generative adversarial networks. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1260–1264, 2021.

[111] Ahmed Ammar Naseer, Benjamin Walker, Christopher Landon, Andrew Ambrosy, Marat Fudim, Nicholas Wysham, Botros Toro, Sumanth Swaminathan, and Terry Lyons. Scoehr: Generating synthetic electronic health records using continuous-time diffusion models. In *Machine Learning in Health Care*, 2023.

[112] Zhaozhi Qian, Thomas Callender, Bogdan Cebere, Sam M. Janes, Neal Navani, and Mihaela van der Schaar. Synthetic data for privacy-preserving clinical risk prediction. *Scientific Reports*, 14, 2023.

[113] Nidhi Kalidas Sawant, Shivnarayan Patidar, Naimahmed Nesaragi, and U. Rajendra Acharya. Automated detection of abnormal heart sound signals using fano-factor constrained tunable quality wavelet transform. *Biocybernetics and Biomedical Engineering*, 41:111–126, 2021.

[114] Tania Akter, Mohammad Hanif Ali, Md Shahriare Satu, Md Imran Khan, and Mufti Mahmud. Towards autism subtype detection through identification of discriminatory factors using machine learning. In *International Conference on Brain Informatics*, pages 401–410. Springer, 2021.

[115] Sujitra Katekarn, Noppadon Sisuk, Teonchit Nuamchit, Kanjana Jittiporn, and Janjira Payakpate. Studying cardiac electrophysiology via epssim. In *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 458–463. IEEE, 2023.

[116] Anant Suraj Vemuri, Stephane Nicolau, Adrien Sportes, Jacques Marescaux, Luc Soler, and Nicholas Ayache. Interoperative biopsy site relocalization in endoluminal surgery. *IEEE Transactions on Biomedical Engineering*, 63(9):1862–1873, 2015.

[117] Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza. Predicting epidemic risk from past temporal contact data. *PLoS Computational Biology*, 11, 2014.

[118] Tancredi Covioli, Tommaso Dolci, Fabio Azzalini, Davide Piantella, Enrico Barbierato, and Marco Gribaudo. Workflow characterization of a big data system model for healthcare through multiformalism. In *European Performance Engineering Workshop*, 2023.

[119] Jakub Tomek, Rebecca A. B. Burton, and Gil Bub. Ccoffinn: Automated wave tracking in cultured cardiac monolayers. *Biophysical journal*, 111 8:1595–1599, 2016.

[120] Matthew Squires, Xiaohui Tao, Soman Elangovan, Raj Gururajan, Xujuan Zhou, and Udyavara Rajendra Acharya. A novel genetic algorithm based system for the scheduling of medical treatments. *Expert Syst. Appl.*, 195:116464, 2022.

[121] Guilherme C Oliveira, Quoc C Ngo, Leandro A Passos, Joao P Papa, Danilo S Jodas, and Dinesh Kumar. Tabular data augmentation for video-based detection of hypomimia in parkinson's disease. *Computer Methods and Programs in Biomedicine*, 240:107713, 2023.

A LIST OF PAPERS REVIEWED

An abridged version of the papers reviewed and their characteristics is provided in Table 2

Publication	Evaluation Metrics	Generation Methods	Dataset	Purpose
Jatoi et al, 2018 [23]	Custom method	Negative variational free energy, Localization error	Statistical Parametric Mapping - SPM12 software	Predictive modelling
Farago et al, 2021 [24]	Autoregressive modeling, Markov chain, RNN	Morphology, Mean, Variance, Autocorrelation, Power Spectral Density (PSD), Probability Distribution	Custom dataset	Signal quality analysis
Kuo et al, 2022 [25]	SAGAN	Accuracy, Standard Deviation	Custom dataset	Improve personalisation of prediction
Rafiei et al, 2023 [26]	CTGAN and SMOTE	AUC, AUROC, Sensitivity, Specificity, PPV, NPV, Bhattacharyya Distance	North Carolina Health System electronic medical record (EMR)	Fluid overload

DeOliveira et al, 2022 [27]	CTGAN and HAR-CTGAN	Weighted Average F1-score, Ambiguity score	ExtraSensory dataset	Generating discreet synthetic data
Das et al, 2023 [28]	VAE	Dimension-Wise Probability, Bernoulli Success Probability, Counterfactual Digital Twin Evaluation, Presence Disclosure, Attribute Disclosure, Nearest Neighbor Adversarial Accuracy Risk	Phase III breast cancer clinical trial (NCT00174655), Small Cell Lung Carcinoma clinical trial dataset (NCT01439568),	Clinical trials
Bing et al, 2022 [29]	VAE	KNN	MIMIC-III	Mitigating representation bias
Bhanot et al, 2021 [30]	HealthGAN	Root Mean Square Error (RMSE), Pearson's Correlation Coefficient, Directional Symmetry, Short Time-Series Distance	American Time Use Survey (ATUS), Medical claims Autism Spectrum Disorder (ASD)	Privacy preservation, maintaining utility
El Emam et al, 2020 [31]	Conditional trees	Matching real with synthetic samples	Washington State Inpatient Database (SID) and Canadian COVID-19 case dataset	Privacy preservation
Lu et al, 2023 [32]	SMOTE	Precision, Recall, F1-score, Geometric mean, Area under the curve of the receiver operating characteristic curve (AUROC), Area under the precision-recall curve (AUPRC)	Taipei Medical University Hospital and Wan Fang Hospital (derivation), Taipei Medical University Shuang Ho Hospital (validation)	Predictive modelling
Solinski et al, 2016 [33]	Detrended fluctuation analysis (DFA)	Shannon Entropy, Poincare Plots, Multiscale Multifractal Analysis	Holter electrocardiogram (ECG) database and Complete electroencephalogram (EEG) recordings	NA
Bhanot et al, 2021 [34]	HealthGAN	Log Disparity, Time-Series Disparity	MIMIC-III and Average Sleep Time of Americans (ATUS)	Fairness
Shaked et al, 2016 [35]	Markovian Model	Mean Similarity, Intersection	MIMIC-III	Privacy preservation
Bahador et al, 2021 [36]	DE-NLPCA	Accuracy	Activities of daily living (ADL) dataset, and EEG / ECG dataset from Northern Ostrobothnia Hospital	Predictive modelling
Weng et al, 2024 [37]	MVIIL-GAN	Missing Values Reconstruction Error	MIMIC-IV	Dataset balancing
Bhanot et al, 2022[38]	Bootstrapping, Random Permutation and HealthGAN	Root Mean Square Error (RMSE), Pearson's Correlation Coefficient, Short Time-Series Distance (STS), Directional Symmetry (DS)	American Time Use Survey (ATUS) dataset and Autism Spectrum Disorder (ASD) claims dataset.	Addressing Data unavailability, Privacy preservation

Nikolentzos et al, 2023 [39]	Variational Graph Autoencoder (VGAE)	Weisfeiler-Lehman Subtree (WL) Graph Kernel, Shortest Path (SP) Graph Kernel, Graph Kernel-Maximum Mean Discrepancy (GK-MMD), Pearson Correlation Coefficient	MIMIC-IV	Privacy preservation
Ricci et al, 2019 [40]	Custom method	Poisson Sequence, Correlated Heartbeatlike Sequence, Hénon Map Sequence	Electronic oscillator sequence, Heartbeat sequence, and Neural sequence	NA
Hodgkiss et al, 2021 [41]	Custom method	AUC	Normal Sinus Rhythm Dataset	Cybersecurity
Chen et al, 2022 [42]	CTGAN	Classifier	CHB-MIT EEG dataset	Dataset balancing
Xu et al, 2017 [43]	Custom method	Markov Chain, Vector Auto-Regressive Model, Continuous-Time Markov Chain, Logistic Regression, Hawkes Processes, Modulated Poisson Processes, Self-Correcting Process	MIMIC-II	Dataset balancing
Mohammed et al, 2017 [44]	ARMA	F1-score, AUC, NOP	Custom dataset	Addressing Data unavailability
Wang et al, 2023 [45]	AFE-GAN (Atrial Fibrillation-like ECG GAN)	Two of the four winning atrial fibrillation detectors from the 2017 PhysioNet Challenge - Hong detector, Datta detector	training set from the 2017 PhysioNet Challenge	Addressing Data unavailability
Imtiaz et al, 2021 [46]	BGAN (Boundary-seeking GAN)	Visualisation only	Custom dataset - from Fitbit Charge 2 HR smartwatches	Privacy preservation
Ashrafi et al, 2023 [47]	simpleGAN, medGAN, DoppelGANger, DPGAN, and PPGAN	F1-score, Precision, Recall, Root Mean Square Error (RMSE), AUC, Attacker Advantage	(Patient interactions with a tablet game (PflegeTab))	Privacy preservation
Rafiei et al, 2024 [48]	SMOTE, CTGAN	Jensen-Shannon Divergence (JSD), Bhattacharyya Distance, Mann-Whitney U Test, Benjamini-Hochberg (BH) procedure	Custom dataset	Predictive modelling
Agliari et al, 2020 [49]	Custom method	Power Spectrum Density (PSD)	Custom dataset	Method evaluation

Li et al, 2023 [50]	EHR-M-GAN and EHR-M-GANconditional	Maximum Mean Discrepancy (MMD), Dimension-Wise Probability, Discriminative Score, Patient Trajectories, Pearson pairwise correlations, Autocorrelation function, Membership Inference Attack, Differential Privacy	MIMIC-III, eICU and HiRID	Addressing Data unavailability, Privacy preservation
Zhang et al, 2022 [51]	LS-EHR	Jensen–Shannon Divergence (JSD), AUC	Custom datasets (two)	NA
Oh et al, 2023 [52]	Monte Carlo simulations	Relative Bias, Confidence Limit Ratios (CLRs), Mean Square Error (MSE)	Custom dataset (South Korea's patients' healthcare resource utilization database)	Checking bias
Burrello et al, 2022 [53]	DA techniques and DL HR algorithms	Visualisation only	PPGDalia	Health monitoring
Mendez et al, 2022 [54]	GAN	Welch's Test	NA	Health monitoring and Data privacy
Yang et al, 2022 [55]	Physics-based models	DFA	PhysioNet	Predictive modelling
Vulpe-Grigorasi et al, 2022 [56]	GAN	RMSSD, SDNN	PhysioNet	Increased diagnosis accuracy
Shamsuddin et al, 2018 [57]	Virtual Patient Model	NB, SVM and TB	ARem and EEG	Addressing Data unavailability
Balasubramanian et al, 2016 [58]	MDMs (Multidimensional Motifs)	Graph Clustering Method	Electromagnetic Articulography and Motion Capture and Muscle Activity	Personalised diagnosis and therapy
Tessier-Larivière et al, 2021 [59]	PNS-GAN	Power Spectral Density, Euclidean distance	BIOS-IT3 Dataset	Data Augmentation
Vandendriessche et al, 2017 [60]	MSE	Classifier	MIMIC-III (heart and sepsis)	Predictive modelling
Tasnim et al, 2022 [61]	SMOTE	Classifier	BCIAUT-P300	Addressing Health inequality
Wang et al, 2023 [62]	SMOTE and WCGAN-GP	Classifier	ImmPort (Immunology Database and Analysis Portal) data	Enhancing health clinical data
García-Vicente et al, 2023 [63]	SMOTE, CTGAN, TVAE	LASSO, SVM, KNN, DT	Norwegian Centre for E-health Research	Data quality enhancement
Kuo et al, 2022 [64]	GAN	Classifier	MIMIC-III and EuResist23	Data quality enhancement and Privacy preservation

Yoon et al, 2023 [65]	Sequential encoder-decoder methods and GAN	Classifier	MIMIC-III	Privacy preservation
Mitra et al, 2023 [66]	CardioSim PC-based system	Classifier	mitdb	Data quality enhancement
Khan et al, 2016 [67]	NA	MEWMA	accelerometer	Health and well-being assessment
Roonizi et al, 2015 [68]	SHVR	Mean Square Error (MSE), Wilcoxon Rank Test	synthetic ECG	Predictive modelling
Konak et al, 2023 [69]	TimeGAN and Animations	MMD	PAMAP2 and SONAR-LAB	Addressing Data unavailability
Tan et al, 2023 [70]	TabGAN and SMOTE	Jensen–Shannon Divergence (JSD), Wasserstein Distance (WD), Diff Corr	SUPPORT and METABRIC	Predictive modelling
Liu et al, 2020 [71]	Deep Sequential Weighting (DSW)	LR, RF, KNN, PSM, CFR, CF, BART	Custom dataset and MIMIC-III	Predicitve modelling
Atmadja et al, 2022 [72]	GAN	CNN	MIT-BIH	Predictive modelling
Hashemi et al, 2023 [73]	GAN	PCA, t-SNE, pairwise correlations,RNN	Sins, MIMIC-VI	Privacy preservation
Wang et al, 2022 [74]	Markov Jump Process	Domain Expert Review	NA	Predictive modelling
Huerta et al, 2021 [75]	Standard Data Augmentation Transformations	McNemar test	PhysioNet/CinC Challenge 2017 database	Addressing Data unavailability
Ayilara et al, 2023 [76]	OSIM2 and ModOSIM	Concordance Correlation Coefficient	Population Research Data Repository (PRDR)	Method evaluation
Haleem et al, 2023 [77]	TC-Multi GAN and Document Sequence Generator	Wasserstein Distance, Kolmogorov-Smirnov Test, Jensen-Shannon Distance, Distance Pairwise Correlation, Sample Kernel Density Estimations	GATEKEEPER EU project	Addressing Synthetic Data feasibility
Zama et al, 2023 [78]	Diffusion-based model	Dynamic Time Warping, Maximum Mean Discrepancy	PTB-XL	Privacy preservation
Mosquera et al, 2023 [79]	RNN with LTSM and GRU	Hellinger's Distance, Cox Regression Hazard Ratios	Alberta Health's administrative dataset	Addressing Synthetic Data feasibility
Huang et al, 2019 [80]	Delete, update, switch operations	Pairwise Similarity Score	Rochester epidemiology project	Predictive modelling
Dahal et al, 2023 [81]	EC–WCGAN	Precision, Recall, F1-score	AHADB, VFDB and CUDB	Dataset balancing
Setiawan et al, 2022 [82]	SMOTE	Accuracy, Sensitivity, Specificity, ROC, Cross-Validation, MSE, MAE	PhysioNet Apnea-ECG database (PAED)	Dataset balancing

Jurdana et al, 2023 [83]	Custom method	MSE	EEG, Royal Brisbane	Predictive modelling
Yang et al, 2023	TS-GAN	LSTM-based Discriminator, Discriminator loss, Maximum Mean Discrepancy, Principal Component Analysis, t-SNE, Sequence diagrams, Accuracy	ECG_200, NonInvasiveFetalECG_Thorax1, and mHealth	Data augmentation
Mahey et al, 2023 [84]	Simulation	Channel by Channel Covariance, EOG, 1/f Function, Spatial Covariance	NA	Addressing Data unavailability
Velasco et al, 2018 [85]	Evolutionary algorithm	Wilcoxon Rank Sum Test (Mann Whitney Wilcoxon) (MWW)	Principe de Asturias Hospital	Addressing Data unavailability
Xiong et al, 2022 [86]	Custom method	Mean Square Error (MSE), Average Standard Deviation, Frequency Distribution	PhysioNet 2017 challenge dataset	Dataset balancing
Krenmayr et al, 2022 [87]	GAN with bi-LSTM	Euclidean Distance, Wasserstein Distance	NA	Addressing Data unavailability
Zaballa et al, 2023 [88]	Probabilistic generative model (HMM and EM)	Average Log Likelihood	NA	Predictive modelling
Wendland et al, 2022 [89]	Multimodal Neural Ordinary Differential Equations	Jensen-Shannon Divergence	PPMI (Parkinson), and NACC (Alzheimer)	Predictive modelling
Kaleli et al, 2023 [90]	GAN with CNN and Transformer	Percent Root Mean Square Difference (PRD), Root Mean Square Error (RMSE), Frechet Distance (FD)	MIT-BIH dataset	Privacy preservation, Predictive modelling
Sun et al, 2021 [91]	Longitudinal GAN	AUROC, AUPCR, AUC	Cerner Health Facts database	Predictive modelling, Privacy Preservation
Shilaskar et al, 2017 [92]	Resampling, modified Particle Swarm Optimization	Accuracy, Precision, Recall, Sensitivity, F1-score	Vani Dataset, Thyroid Dataset, PdA, Cleveland, Audiology, SVD, Vertigo	Predictive modelling
Foomani et al, 2022 [93]	GAN	Jensen-Shannon Divergence, AUC	EMR data from Vascular Centers, Milwaukee, WI	Predictive modelling
Peng et al, 2022 [94]	Gaussian Kernels	Root Mean Square Error (RMSE)	NA	Predictive modelling
Geng et al, 2021 [95]	GAN	AUC, F1-score	NA	Predictive modelling
Mivalt et al, 2022 [96]	GAN	Cohen's Kappa, F1-score	Multicenter Intracranial EEG Dataset	NA
Pascual et al, 2021 [97]	GAN	Cosine Similarity, Recall	EPILEPSIAE	Privacy preservation

Lu et al, 2023 [98]	GAN	Jensen-Shannon Divergence, Normalised Distance, AUC, F1-score	MIMIC-III and MIMIC-IV	Privacy preservation
Ikuma et al, 2016[99]	Karhunen-Loeve transformation, Time-series model perturbations	Correlation analysis relative vibration power represented by a synthetic waveform	NA	Predictive modelling
Zhang et al, 2021 [100]	GAN	Diagnosis forecast analysis, Kolmogorov-Smirnov Test	Synthetic Derivative at Vanderbilt University Medical Center	Privacy preservation
Baowaly et al, 2019 [101]	GAN	Dimension-Wise Average, Kolmogorov-Smirnov (KS) Test, Association Rule Mining	MIMIC-III and NHIRD	Predictive modelling
Abdelghaffar et al, 2022 [102]	GAN	Relative Entropy, Accuracy	Wadsworth BCI Dataset from the BCI competition III	Predictive modelling
Al-Saad et al, 2021 [103]	DPGAN	Dimension-Wise Average, AUROC, Area under the Precision-Recall Curve, Accuracy	Arizona State's Kinesiology Department	Predictive modelling, Privacy preservation
Yu et al, 2023 [104]	GAN	AUC	CHB-MIT dataset	Predictive modelling
Cenek et al, 2020 [105]	Frequency Domain Model	NA	CIRCADA-S	Predictive modelling
Yadav et al, 2023 [106]	GAN	Mean Absolute Error (MAE), MRLE, PCA, t-SNE	UNIMIB	Predictive modelling
Yu et al, 2024 [107]	Temporal Convolutional Network	Dipole Localization Error (DLE), Normalized Hamming Distance, Sensitivity, Specificity, False Detection Rate, F1-score, Pearson Correlation	NA	Predictive modelling
Walonoski et al, 2018 [108]	Markovian model (PADARSER)	Prevalence Difference Error	Multiple	Predictive modelling
Kulpa et al, 2022 [109]	Autoregressive model, Markov chain, RNN	Power Spectral Density (PSD)	MIT-BIH NSTDB	Predictive modelling
Kan et al, 2021 [110]	GAN	Average Error Rate	Temple University Hospital Abnormal EEG Corpus	Predictive modelling
Naseer et al, 2023 [111]	Continuous-Time Diffusion Models	Dimension-wise distribution, Pairwise Correlation difference, Log-cluster, Synthetic ranking agreement, Membership Inference Attack, Blinded Clinician Evaluation, Domain Expert Review	MIMIC-III and ED-EHR datasets	NA

Qian et al, 2024 [112]	DPGAN, PATEGAN), ADSGAN	Fidelity (Alpha-Precision), Diversity (Beta-Recall), Authenticity, Wasserstein distance, Jensen-Shannon distance, Inverse Kullback-Leibler divergence, Chi-Squared Tes, Kolmogorov-Smirnov test, k-anonymity, DOMIAS AUC	Ever-smokers in UK Biobank Database	Privacy preservation
Lu et al, 2023 [32]	SMOTE	Decision Tree, Random Forest, Logistic Regression, Extreme Gradient Boosting, Support Vector Machines	Custom data	Dataset balancing
Sawant et al, 2021 [113]	SMOTE	Sensitivity, Specificity, Overall score (Average of Sensitivity and Specificity)	PhysioNet / CinC challenge 2016, and PASCAL	Dataset balancing
Akter et al, 2021 [114]	SMOTENC (SMOTE variant)	Classifier	Quantitative Checklist for Autism in Toddlers-10 (Q-CHAT-10), and Autism Spectrum Quotient-10 (AQ-10)	Predictive modelling
Katekarn et al, 2023 [115]	Custom method	Participant satisfaction questionnaire and SPSS	Custom dataset	Method evaluation
Vemuri et al, 2016 [116]	Custom method	Custom evaluation metrics (Measurement of Uncertainty, Measuring Uncertainty in Endoscope Tip)	NA	Predictive modelling
Valdano et al, 2015 [117]	Custom method	Visualisation only	NA	Disease modelling
Covioli et al, 2023 [118]	Simulation	Visualisation only	MIMIC-III and MIMIC-III waveform matched dataset	Method evaluation
Tomek et al, 2016[119]	Cellular automata	Median, p-value	Custom dataset	Method evaluation
Squires et al, 2022 [120]	Python packages (random and fake)	Custom evaluation metrics	Custom dataset	Addressing Data unavailability
Oliveira et al, 2023 [121]	CGAN and CT-GAN	False Positives, False Negatives, Accuracy, Specificity, Sensitivity, AUC	PARK Facial Mimic	Data Augmentation

B SEARCH QUERIES

The following is a list of search queries used across five databases, and the number of relevant results obtained. It should be noted that additional filtering criteria were set on these databases, including the date of publication range (2014-2024) and the language of the publication (English).

Database	Query	#
Scopus	(synthe* OR augment*) AND generat* AND (time-series OR time* OR temporal*) AND (tabular OR record*) AND (patient* OR medic* OR health* OR clinic* OR ehr*)} on title-abstract-keywords	415
Web of Science	(synthe* OR augment*) AND generat* AND (time-series OR time* OR temporal*) AND (tabular OR record*) AND (patient* OR medic* OR health* OR clinic* OR ehr*)} (Topic)	1003
PubMed	(synthe* [Title/Abstract] OR augment* [Title/Abstract]) AND generat*[Title/Abstract] AND (time-series[Title/Abstract] OR time*[Title/Abstract] OR temporal* [Title/Abstract]) AND (tabular[Title/Abstract] OR record* [Title/Abstract]) AND (patient*[Title/Abstract] OR medic*[Title/Abstract] OR health*[Title/Abstract] OR clinic*[Title/Abstract]) on title-abstract}	202
IEEE Xplore	("All Metadata":synthe* OR "All Metadata" :augment*) AND ("All Metadata":time* OR "All Metadata":temporal*) AND ("All Metadata":tabular OR "All Metadata":record*) AND ("All Metadata":patient* OR "All Metadata":medic* OR "All Metadata":health* OR "All Metadata":clinic* OR "All Metadata" :ehr*) AND ("All Metadata":generat*)	138
ACM	[Abstract: synthe*] AND [[Abstract: time*] OR [Abstract: temporal*]] AND [[Abstract: patient*] OR [Abstract: medic*] OR [Abstract: health*] OR [Abstract: clinic*] OR [Abstract: ehr*]] AND [E-Publication Date: (01/01/2014 TO 31/01/2024)]	8

Table 3. Search Queries used, and the number of Results obtained from each Database