Artificial patients in anesthesia: more than a passing fad?

Fabrice Ferré, Stéphanie Allassonnière, Clément Chadebec, Vincent Minville

Submitted to: Journal of Medical Internet Research
on: June 11, 2024

Table of Contents

Artificial patients in anesthesia: more than a passing fad?

Fabrice Ferré[1] MD, PhD; Stéphanie Allassonnière[2] PhD; Clément Chadebec[2] PhD; Vincent Minville[1] MD, PhD

[1]Department of Anesthesia, Intensive Care and Perioperative Medicine Purpan University Hospital Toulouse FR
[2]Université Paris Cité, UMR S1138, INRIA, INSERM, Sorbonne University Paris FR

Corresponding Author:
Fabrice Ferré MD, PhD
Department of Anesthesia, Intensive Care and Perioperative Medicine
Purpan University Hospital
place du Dr Baylac
Toulouse
FR

Abstract

Artificial patients' technology has the potential to transform healthcare by improving and potentially accelerating diagnosis and treatment, and by mapping clinical pathways or patient care processes. Deep learning methods for generating artificial data in healthcare include data augmentation by variational autoencoders (VAE) technology. The aim of our study was to test the feasibility of generating artificial patients with reliable clinical characteristics by using a geometry-based VAE applied, for the first time, on high dimension low sample size tabular data. Real patients' tabular data were extracted from the "MAX" digital conversational agent created for preparing patients for anesthesia (BOTdesign®, Toulouse, France). A 3-stage methodological approach was implemented to generate up to 10,000 artificial patients: training the model and generating artificial data, assessing the consistency and confidentiality of artificial data, and validating the plausibility of the newly created artificial patients. We demonstrated for the first time the feasibility to transpose the VAE technique from imaging to tabular data for the generation of a large number of artificial patients. Our digital patients' cohort was highly consistent. Moreover, artificial patients could not be matched with real patients, thus guaranteeing the essential ethical concern of confidentiality. Further studies integrating dynamic changes (and their variability) are needed to map trends and identify patient trajectories.

(JMIR Preprints 11/06/2024:63130)
DOI: https://doi.org/10.2196/preprints.63130

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Research Letter

Artificial patients in anesthesia: more than a passing fad?

Fabrice Ferré[1], Stéphanie Allassonnière[2], Clément Chadebec[2] and Vincent Minville[1]

[1]Department of Anesthesia, Intensive Care and Perioperative Medicine, Purpan University Hospital, Toulouse, France

[2]Université Paris Cité, UMR S1138, INRIA, INSERM, Sorbonne University, Paris, France

Corresponding Author:

- Fabrice Ferré, MD, PhD
- fabriceferre31@gmail.com, Phone: +33 (0)5 61 77 99 88
- Hôpital Pierre-Paul Riquet, CHU Purpan, place du Dr Baylac, 31059 Toulouse, France

ABSTRACT

Artificial patients' technology has the potential to transform healthcare by improving and potentially accelerating diagnosis and treatment, and by mapping clinical pathways or patient care processes. Deep learning methods for generating artificial data in healthcare include data augmentation by variational autoencoders (VAE) technology. The aim of our study was to test the feasibility of generating artificial patients with reliable clinical characteristics by using a geometry-based VAE applied, for the first time, on high dimension low sample size tabular data. Real patients' tabular data were extracted from the "MAX" digital conversational agent created for preparing patients for anesthesia (BOTdesign®, Toulouse, France). A 3-stage methodological approach was implemented to generate up to 10,000 artificial patients: training the model and generating artificial data, assessing the consistency and confidentiality of artificial data, and validating the plausibility of the newly created artificial patients. We demonstrated for the first time the feasibility to transpose the VAE technique from imaging to tabular data for the generation of a large number of artificial patients. Our digital patients' cohort was highly consistent. Moreover, artificial patients could not be matched with real patients, thus guaranteeing the essential ethical concern of confidentiality. Further studies integrating dynamic changes (and their variability) are needed to map trends and identify patient trajectories.

INTRODUCTION

Along with the growing impact of data science technologies like artificial intelligence, novel healthcare data ecosystems centered around digital patient are developing. For instance, the creation of digital twins paves the way for improved health monitoring and facilitation of personalized therapeutics based on management, analysis, and interpretation of medical data. However, the utility and feasibility of patients' digital twins in routine medical processes are still limited[1]. Another data science-based approach for generating artificial patients involves augmenting real data[2]. In a nutshell, data augmentation is the art of increasing the size of a given data set by creating artificial labeled data. In this way, new artificial data are created with characteristics similar to those of the original population of interest. Such data could be particularly useful in the field of clinical research, holding out the promise of studies that are less expensive, but above all fairer, more inclusive and more effective, especially when patient recruitment is difficult. For example, this approach could deal with the lack of data that is usually observed when subgroups of patients are under-represented in a study (pregnant women, children, rare diseases…) and which makes it difficult, if not impossible, to reliably predict their outcome.

Deep learning methods for generating artificial data in healthcare include generative adversarial networks (GAN), variational autoencoders (VAE), and denoising diffusion -based technologies [1]. Using a VAE, Chadebec *et al.* have recently demonstrated that the artificial augmentation of medical imaging data (MRI extracted from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database) significantly improved classification accuracy[2]. In this setting, the balanced accuracy increases from 66 to 74% for a convolutional neural network classifier trained with small datasets (50 MRI of cognitively normal and 50 MRI of Alzheimer disease patients), while improving greatly the sensitivity and specificity of the classification metrics[2]. In other words, geometry-based VAE was able to produce meaningful samples especially in the context of high dimension low sample size (HDLSS) imaging datasets. This method, validated for an image classification task (voxels), deserves to be tested on tabular data. Given the multiplicity and complexity of the data obtained in anesthesia, an approach based on the artificial augmentation of data could be of major interest. For example, we could identify predictive clinical factors of poor outcome (e.g., postoperative complications) with accuracy and reliability in various surgical populations. Thus, the aim of our study was to test the feasibility of generating artificial patients with reliable clinical characteristics by using a geometry-based VAE applied on HDLSS tabular data.

METHODS

Real patients' tabular data were extracted from the "MAX" database. MAX (BOTdesign®, Toulouse, France) is a digital conversational agent for preparing patients for anesthesia[3,4]. Collected data included demographics characteristics, past medical history, medication and other relevant medical items allowing the calculation of perioperative risk scores (e.g., simplified Lee, Apfel or STOP-BANG scores). In this setting, we have recently demonstrated the predictive value on patient outcome of the MyRISK digital risk score established by MAX (according to a predefined algorithm)

---

[1] For a review, see the White Paper "*Données de santé artificielles : analyse et pistes de réflexion*" available at https://static.botdesign.net/docs/VF_Livre_blanc_Données_de_santé_artificielles-250424.pdf

before the preanesthetic consultation[4]. To generate artificial patients, a 3-stage methodological approach was implemented:

i) Training the model and generating artificial data
ii) Assessing the consistency and confidentiality of artificial data
iii) Validating the plausibility of the artificial patients

RESULTS

#First stage: training the model and generating artificial data

The dataset included 521 patients about to undergo anesthesia, each with 85 clinical features. Once data pre-processing has been completed (filtering, binarization, normalization, ordinal and one-hot encoding), the model was trained on a (521, 103) dimension dataset using a modified Pyraug's training pipeline [2]. Two datasets of 5,000 and 10,000 artificial patients were generated representing respectively a data increase rate of 10 and 20 artificial patients for 1 real patient.

#Second stage: assessing the consistency and confidentiality of artificial data

To assess the consistency of newly generated artificial data, fidelity scores (including statistical analyses of numerical data stability, categorical data distribution stability and numerical value correlation stability) were calculated. With values of 97.6% and 94.6%, the digital data of the 5,000 and 10,000 artificial patients were considered realistic and representative of the real data. To assess the confidentiality of newly generated artificial data, filter similarity scores (proportion of data not similar to the initial dataset) and degree of anonymization (euclidean distance) were calculated. The results strongly confirm the non-similarity of the artificial data with the initial real data (filter similarity scores > 99.9%), and a high degree of anonymization (the artificial data were further away from the initial dataset than the initial data were from each other).

#Third stage: validating the plausibility of the artificial patients

While assessing the relevance of augmented data may be quite straightforward for simple data sets, it reveals very challenging for complex data and may require the intervention of an expert assessing the degree of relevance of the proposed transformations. Thus, a categorization task was performed by

[2[]] Training parameters were set to 1000 epochs, a batch size of 32 and a learning rate of 0.001

three experienced anesthetists who were blinded with the distribution of a balanced sample of 100 real and artificial patients generated using the VAE. Anesthetists were asked to determine whether each patient was real or artificial. The kappa coefficients of agreement were -0.12 [$CI_{95\%}$ from -0.31 to 0.07], 0.15 [$CI_{95\%}$ from 0.1 to 0.26] and 0.09 [$CI_{95\%}$ from -0.14 to 0.15]. Given the very low agreement coefficients (< 0.2), none of the 3 experts was able to differentiate between real and artificial patients, arguing for the medical plausibility of the artificial patients.

DISCUSSION

The use of artificial intelligence in healthcare presents particularly complex challenges inherent in the types of data it relies on (sensitive, sparse, heterogeneous, limited…). However, the research efforts of the last few years, which have focused on meeting these specific challenges, now offer a glimpse of dizzying potential, particularly in the field of clinical studies. In this setting, artificial patients' technology has the potential to transform healthcare by improving and potentially accelerating diagnosis and treatment on the one hand, and by mapping clinical pathways or patient care processes on the other hand[5]. Here, we demonstrated for the first time the feasibility to transpose the VAE technique from imaging to tabular data for the generation of a large number of artificial patients. Our digital patients' cohort was highly consistent. Moreover, artificial patients could not be matched with real patients, thus guaranteeing an essential ethical concern: the confidentiality of sensitive data. Further studies integrating dynamic changes (and their variability) are needed to map trends and identify patient trajectories. In this context, the recent update of MAX with the implementation of a postoperative digital conversational agent for the collection of recovery data based on patient-reported outcome measures could be of major interest. We believe that artificial cohorts of patients can be used to track a variety of health indicators and generate key insights. Artificial patients will revolutionize healthcare paving the way for a more precise, personalized and predictive medicine.

Acknowledgment

We would like to thank Open Studio (IA-Lab, Lyon, France) for their data analyses.

The authors would also like to thank Olivier Thuillart and Jean-Louis Fraysse from BOTdesign® (Toulouse, France), the company that created the digital companion MAX (Medical Assistant eXperience).

Funding Statement

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6).

Conflicts of Interest

None

REFERENCES

1. Fischer RP, Volpert A, Antonino P, Ahrens TD: Digital patient twins for personalized therapeutics and pharmaceutical manufacturing. Front Digit Health 2023; 5: 1302338

2. Chadebec C, Thibeau-Sutre E, Burgos N, Allassonniere S: Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder. IEEE Trans Pattern Anal Mach Intell 2023; 45: 2879-2896

3. Ferre F, Boeschlin N, Bastiani B, Castel A, Ferrier A, Bosch L, Muscari F, Kurrek M, Fourcade O, Piau A, Minville V: Improving Provision of Preanesthetic Information Through Use of the Digital Conversational Agent "MyAnesth": Prospective Observational Trial. J Med Internet Res 2020; 22: e20455

4. Ferre F, Laurent R, Furelau P, Doumard E, Ferrier A, Bosch L, Ba C, Menut R, Kurrek M, Geeraerts T, Piau A, Minville V: Perioperative Risk Assessment of Patients Using the MyRISK Digital Score Completed Before the Preanesthetic Consultation: Prospective Observational Study. JMIR Perioper Med 2023; 6: e39044

5. Elkefi S, Asan O: Digital Twins for Managing Health Care Systems: Rapid Literature Review. J Med Internet Res 2022; 24: e37641

Supplementary Files

TOC/Feature image for homepages

Artificial patients' technology has the potential to transform healthcare by improving and potentially accelerating diagnosis and treatment, and by mapping clinical pathways or patient care processes.