Obesity Surgery (2025) 35:638–641
https://doi.org/10.1007/s11695-024-07653-z

BRIEF REPORT

Evaluating AI Capabilities in Bariatric Surgery: A Study on ChatGPT-4 and DALL·E 3's Recognition and Illustration Accuracy

Mohammad Mahjoubi[1] · Shahab Shahabi[1] · Saba Sheikhbahaei[2] · Amir Hossein Davarpanah Jazi[1]

Received: 21 October 2024 / Revised: 22 October 2024 / Accepted: 22 December 2024 / Published online: 29 December 2024

Abstract

Background With the rise of artificial intelligence (AI) in medical education, tools like OpenAI's ChatGPT-4 and DALL·E 3 have potential applications in enhancing learning materials. This study aims to evaluate ChatGPT-4o's proficiency in recognizing bariatric surgical procedures from illustrations and assess DALL·E 3's effectiveness in generating accurate surgical illustrations.

Methods Illustrations of six bariatric surgical procedures (One Anastomosis Gastric Bypass, Roux-en-Y Gastric Bypass, Single Anastomosis Duodeno-Ileal Bypass with Sleeve Gastrectomy, Sleeve Gastrectomy, Biliopancreatic Diversion, and Adjustable Gastric Banding) were sourced from the IFSO Atlas of Metabolic and Bariatric Surgery. ChatGPT-4 was tasked with identifying each procedure based on these illustrations to evaluate its classification accuracy. Simultaneously, DALL·E 3 was prompted with the specific names of each procedure to generate corresponding medical illustrations.

Results ChatGPT-4 correctly identified only the Adjustable Gastric Banding illustration, misclassifying the other five procedures. DALL·E 3 failed to produce accurate illustrations for all six procedures.

Conclusion The study underscores the need for further evaluation of AI in bariatric surgery. Both ChatGPT-4 and DALL·E 3, while promising, have significant limitations in recognizing and generating accurate illustrations of bariatric surgical procedures. These findings call for continued research and development to make AI models suitable for medical education applications in bariatric surgery.

Keywords Bariatric surgery · Generative artificial intelligence · ChatGPT · Image · DALLE

Introduction

Since 1950, the metabolic and bariatric surgery (MBS) field has experienced significant growth. This expansion, from simple restrictive procedures to more complex operations, has increased demand for MBS [1].

In recent years, artificial intelligence (AI) has been increasingly explored for its potential applications in medical education, leading to significant efforts to improve the field [2]. Previous studies have investigated language models such as OpenAI's ChatGPT-3 and ChatGPT-4 to enhance educational resources for patients and practitioners in the field of bariatric surgery [3, 4]. With the ongoing advancements in AI technology, particularly the introduction of OpenAI's text-to-image model DALL·E and the image recognition capabilities of ChatGPT-4o, it is crucial to evaluate these tools before integrating them into clinical practice.

This study aims to assess ChatGPT-4's proficiency in recognizing and interpreting bariatric surgical procedures and the effectiveness of DALL·E in generating accurate illustrations pertinent to bariatric surgery.

Method

In this study, we utilized illustrations from the IFSO Atlas of Metabolic and Bariatric Surgery [5] to visually represent the following surgical procedures: One Anastomosis Gastric Bypass (OAGB), Roux-en-Y Gastric Bypass

✉ Amir Hossein Davarpanah Jazi
davarpanahjazi@gmail.com

Mohammad Mahjoubi
mo.mahjoubi.md@gmail.com

Shahab Shahabi
shshahabi@yahoo.com

Saba Sheikhbahaei
sabash1368@gmail.com

[1] Minimally Invasive Surgery Research Center, Iran University of Medical Sciences, Tehran, Iran

[2] Hunter New England Local Health District, Newcastle, Australia

(RYGB), Single Anastomosis Duodeno-Ileal Bypass with Sleeve Gastrectomy (SADI-S), Sleeve Gastrectomy, Biliopancreatic Diversion (BPD), and Adjustable Gastric Banding (AGB) (Fig. 1). These images served as the reference standard for identifying and categorizing the surgical techniques. We then utilized ChatGPT-4.0 to determine the type of surgical procedure depicted in each image, evaluating the model's ability to classify the procedures accurately.

Additionally, we employed OpenAI's DALL·E 3, an advanced generative AI text-to-image model, to create medical illustrations for the same procedures presented in Fig. 2. By providing the model with the specific names of each bariatric surgery technique, we generated images intended for use in medical education. To assess the accuracy of the generated images, we compared them to the reference illustrations from the IFSO Atlas.

Result

When the IFSO Atlas illustration for the OAGB procedure was provided to ChatGPT-4.0, the model misclassified it as an RYGB. Similarly, the illustration for RYGB was incorrectly identified as a Biliopancreatic Diversion with Duodenal Switch (BPD/DS). Furthermore, the SADI-S procedure

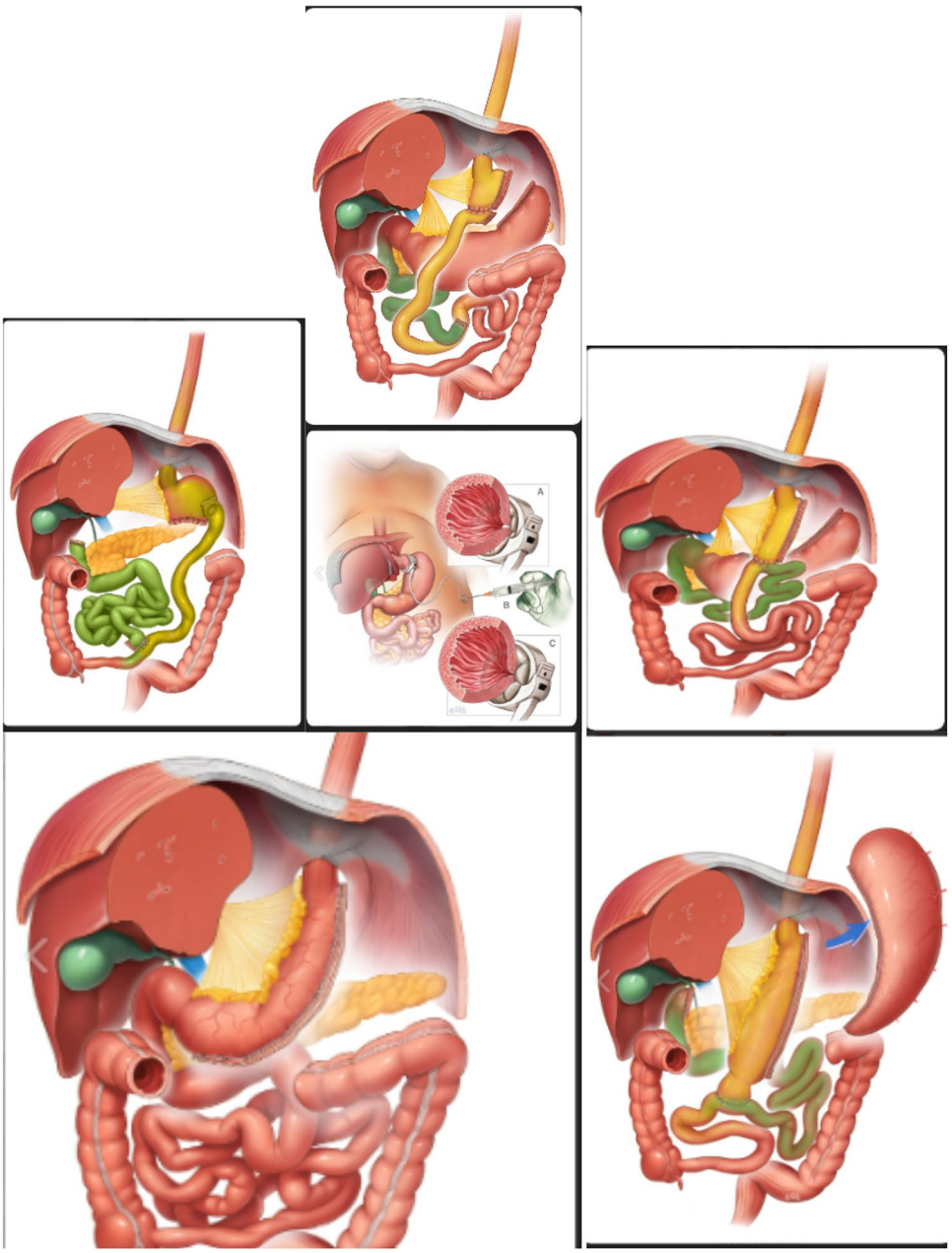

Fig. 1 Illustrations from the IFSO Atlas of Metabolic and Bariatric Surgery

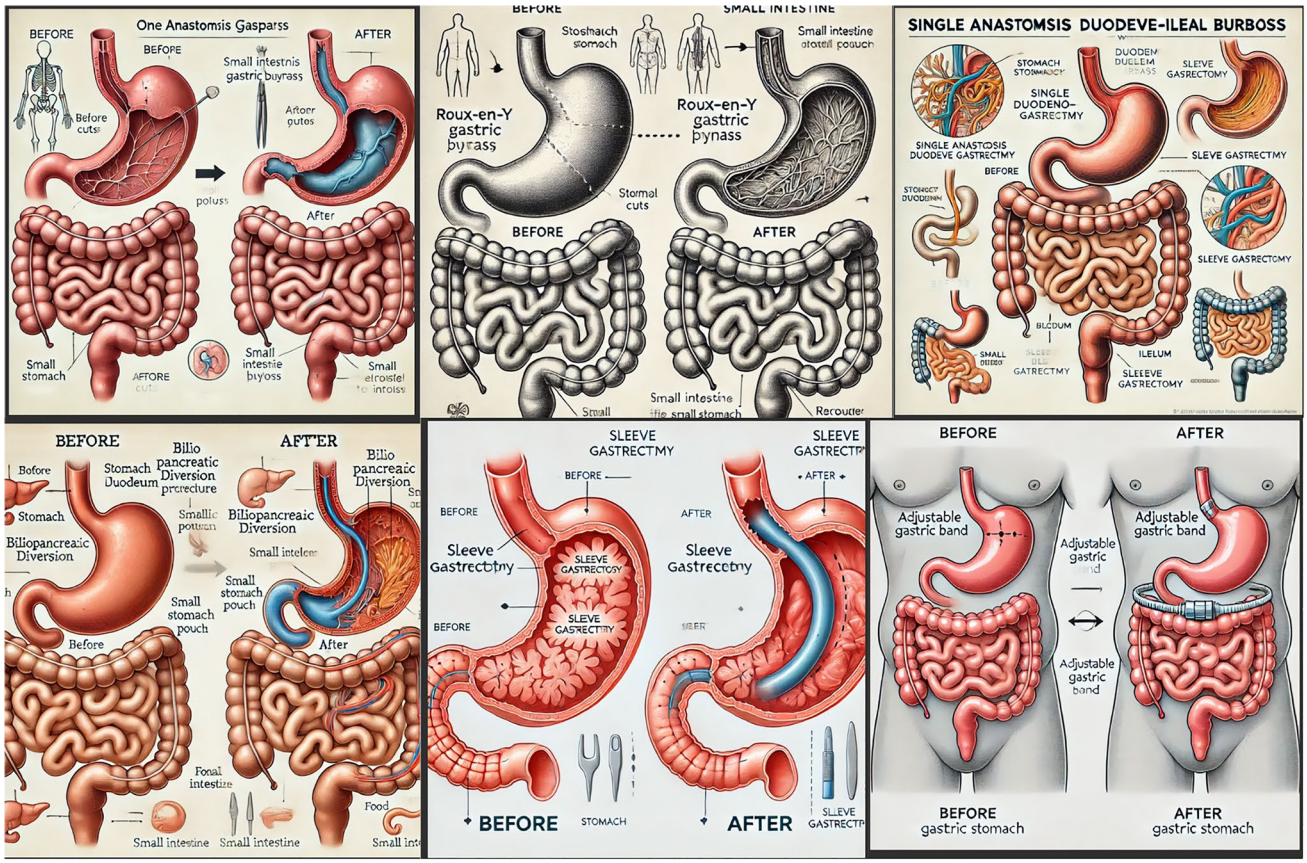

Fig. 2 DALL·E 3-generated illustrations of various bariatric surgeries

illustration was misclassified as a Sleeve Gastrectomy. The Sleeve Gastrectomy illustration was recognized as a Duodenal Switch procedure.

In contrast, the model correctly identified the AGB illustration and comprehensively explained the surgical technique.

Appendix 1 provides detailed transcripts of the interactions between ChatGPT-4.0 and the researchers, including the context of the queries and responses, for further review and analysis.

In the second part of our study, we utilized DALL·E 3 to generate illustrations of the mentioned bariatric procedures.

- OAGB: The image did not depict any specific details of bariatric surgery. It displayed minimal venous and arterial structures without indications of stapling, resection, or anastomosis. Additionally, the accompanying written description contained several grammatical and dictation errors.
- RYGB: The illustration did not depict the bariatric procedure. It featured dashed lines on the stomach unrelated to the operation, with no depiction of pouch creation or anastomosis between the small bowel and gastric wall.
- SADI-S: The image emphasized vascular structures. However, it failed to depict stapling, sleeve gastrectomy, or the anastomosis between the duodenum and small intestine. Misleadingly, it showed an anastomosis between the right colon and pylorus.
- SG: The illustration included an object resembling an orogastric tube used as a sizer and dashed lines that might represent stomach stapling. Inside the stomach, a structure similar to brain tissue was depicted.
- BPD: The image did not portray any aspects of the surgical procedure. The gastric tissue appeared intact, and no anastomosis was shown. There was no representation of gastroileal anastomosis, and elements resembling embryological development of the pancreas were included alongside misleading vessels throughout the colon.
- AGB: The illustration did not show the standard placement of the gastric band or a tube connecting the band to the subcutaneous reservoir. Instead of depicting the procedure, it presented a large circle encompassing the stomach, small intestine, and large intestine.

Appendix 2 contains the images generated by DALL·E 3.

Discussion

In this study, ChatGPT-4 was tested on its ability to recognize bariatric surgery illustrations but failed to identify all except one. Similarly, DALL·E 3 demonstrated clear limitations in generating detailed illustrations of bariatric procedures. The images produced often lacked vital elements essential for surgical illustrations. They did not accurately represent the anatomy of the gastrointestinal tract when compared to authoritative sources like the IFSO Atlas of Bariatric Surgery**.**

The literature on the capabilities of DALL·E and the image recognition abilities of ChatGPT-4 is limited but growing. In one instance, researchers attempted to use DALL·E 3 to generate illustrations of congenital heart diseases, failing to produce accurate representations [6]. Another study exploring the use of DALL·E in dermatology found that while the model struggled to depict various skin diseases accurately, it showed promising capabilities in generating images of normal-appearing skin [7]. These limitations underscore the necessity for further research to determine the appropriate role of this revolutionary technology in medical education. While current models may not be suitable for generating precise medical illustrations or recognizing complex medical images, they hold potential in other areas.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11695-024-07653-z.

Author contribution M.M. and A.H.D.J. conceptualized and designed the study. M.M., A.H.D.J., and S.S. contributed to data acquisition and analysis. A.H.D.J. and M.M. interpreted the data and drafted the manuscript. S.S. and S.S.B. critically revised the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Data availability No datasets were generated or analysed during the current study.

Declarations

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Phillips BT, Shikora SA. The history of metabolic and bariatric surgery: Development of standards for patient safety and efficacy. Metabolism. 2018;79:97–107.
2. Aburumman R, Al Annan K, Mrad R, et al. Assessing ChatGPT vs. standard medical resources for endoscopic sleeve gastroplasty education: a medical professional evaluation study. Obes Surg [Internet]. 2024;34(7):2718–24. https://doi.org/10.1007/s11695-024-07283-5
3. Jazi AHD, Mahjoubi M, Shahabi S, et al. Bariatric evaluation through AI: a survey of expert opinions versus ChatGPT-4 (BETA-SEOV). Obes Surg [Internet]. 2023. https://doi.org/10.1007/s11695-023-06903-w
4. Lopez-Gonzalez R, Sanchez-Cordero S, Pujol-Gebellí J, et al. Evaluation of the impact of ChatGPT on the selection of surgical technique in bariatric surgery. Obes Surg. 2024.
5. Surgery IA of B. No Title [Internet]. 2024. Available from: https://www.ifso.com/atlas-of-bariatric-and-metabolic-surgery/. Accessed 15 Aug 2024.
6. Temsah MH, Alhuzaimi AN, Almansour M, et al. Art or Artifact: evaluating the accuracy, appeal, and educational value of AI-generated imagery in DALL·E 3 for illustrating congenital heart diseases. J Med Syst. 2024;48(1):54.
7. Cheraghlou S. Evaluating dermatologic domain knowledge in DALL-E 2 and potential applications for dermatology-specific algorithms. Int J Dermatol. 2023;62(10):e521–3.