Evaluating the Realism of ANA HEp-2 Cell Images Synthesized Using Latent Diffusion Models—A Multi-center Visual Turing Test

Principal Investigator: Dr. Junxiang Zeng

Co-Investigators: Dr. Xiujun Pan
Dr. Lisong Shen
Dr. Yingxia Zheng

Version Date: 22 July 2024

Table of Contents

1.0 INTRODUCTION

The use of the indirect immunofluorescence (IIF) assay on human epithelial cells (HEp-2) to screen for antinuclear antibodies (ANAs) is considered a fundamental diagnostic technique and serves as the gold-standard method for evaluating rheumatic diseases, including systemic lupus erythematosus (SLE), Sjögren syndrome (SS), systemic sclerosis (SSc), and mixed connective tissue disease (MCTD)[1]. A variety of distinct ANA patterns are observed on HEp-2 cells, and some of which are correlated with specific diseases[2-3]. To establish standardized nomenclature and categorization of the ANA-IIF patterns, the International Consensus on ANA Patterns (ICAP) has developed a comprehensive taxonomy, which in 2021 has classified 29 distinct ANA patterns, each assigned an AC (anti-cell) code ranging from AC-1 to AC-29 [4]. Furthermore, ICAP has established a classification system that differentiates between two levels of reporting: competent-level and expert-level[5], with the former encompassing the most frequently reported ANA patterns. The expert-level, on the other hand, mainly includes rare ANA patterns (occurring in less than 1% of patients testing positive for ANA). For further elucidation, kindly refer to the official ICAP website (https://www.anapatterns.org).

It should also be noted that some rare ANA patterns should be determined by specific unique or distinct features[6-7]. The process of determining ANA patterns was highly subjective, time-consuming, and prone to errors, often influenced by individual bias and clinical experiences[8-11]. Actually, even among highly skilled medical technologists, inter- and intra-grader inconsistencies also occurred[12, 13]. These difficulties of poor detection rate, intricate interpretation, and subtle characteristics imply unacceptable rates of false negative and/or positive detection, further generating its scarcity. Hence, enhancing laboratory testing for rare patterns will contribute to the investigation of their potential associations with specific autoantibodies, which will help to support more reliable and effective clinical diagnoses. Also, the development of automatic rare ANA patterns detection approaches is of great significance to help medical technologists to address staffing shortages, and more importantly, reduce the time required for manual interpretation, consequently enhancing laboratory efficiency, which is of vital importance in clinical practice[11].

Recently, remarkable advancements have been witnessed in the significant role of deep learning (DL) in the recognition of ANA patterns. Our recent research [14] has proposed an attention-based enhancement framework (ABEF) for recognizing rare ANA patterns in ANA-IIF images. However, our model did not exhibit exceptional performance in detecting all categories of the rare ANA patterns, two categories (AC-24, AC-27) of the pattern detection

task in our dataset unmet expectations. This phenomenon may be attributed to insufficient number of samples and the small size of the objects.

We propose to improve upon this, by means of a prospective multi-center cohort study to analyze the Implementing automated detection of ANA through deep learning methods. We intend to make use of the latent diffusion models to synthesize more high-quality images, as well as the integrity of computer-generated data to achieve data augmentation, thereby enhancing the robustness of the model, yet strangely neglected in the literature. We will then seek to develop an multi-center visual Turing test (https://turing.rednoble.net/) , inviting more experts in ANA interpretation to participate in evaluating the authenticity and quality of images generated by our diffusion model.

This project, besides having the scientific goals described in the paragraph above, will help medical technologists to address staffing shortages, and more importantly, reduce the time required for manual interpretation, consequently enhancing laboratory efficiency, which is of vital importance in clinical practice.

2.0 AIM

To rigorously examine the feasibility and efficacy of utilizing latent diffusion models for data augmentation in anti-nuclear antibody (ANA) indirect immunofluorescence (IIF) images.

3.0 HYPOTHESIS

The generation of ANA-IIF images will be realistic enough that even expert cytopathologists will find it difficult to differentiate them (fakes) from real (genuine) ones, and the AI model can be an efficient and reliable tool to assist cytopathologists in their clinical practice.

4.0 OBJECTIVES

4.1 Primary objectives

1. To evaluate the feasibility and efectiveness of using diffusion models to generate synthetic ANA-IIF images and evaluate their performance in image quality and prediction outcomes.
2. By leveraging the multi-center visual Turing test, we invite more experts in anti-nuclear antibody interpretation to participate in evaluating the authenticity and quality of images generated by our diffusion model.

4.2. Secondary objectives

1. Through the visual Turing test, we aim to understand the current level of expertise among ANA

interpreters regarding rare patterns, as well as the potential of our model to serve as an educational or training tool.

5.0 ENDPOINTS

5.1 Primary endpoints

1) To determine the predictive power (precision, recall, F1 score, and finally to mean average precision) of the diffusion model.

2) To determine quality of images generated by diffusion model, Frechet inception distance (FID) serves as a metric for evaluating the quality.

6.0 STUDY POPULATION

6.1 Inclusion criteria

1. Originating from reputable medical institutions.
2. Participants with relevant certification and qualifications.
3. Participants Having over one year of experience in interpreting anti-nuclear antibody (ANA) patterns within a laboratory setting.

6.2 Exclusion criteria

1. Participants lacking relevant professional certification and qualifications.
2. ParticipantsWithout experience in interpreting ANA patterns.
3. Unwilling to accept the rules and informed consent of the visual Turing test.

7.0 STUDY DESIGN

7.1 Study schema

This is a observational prospective cohort study. Participants engage in our research through our Turing test platform (https://turing.rednoble.net/) on the premise of informed consent.The detailed research process is depicted in the diagram below, and further textual description is omitted here.

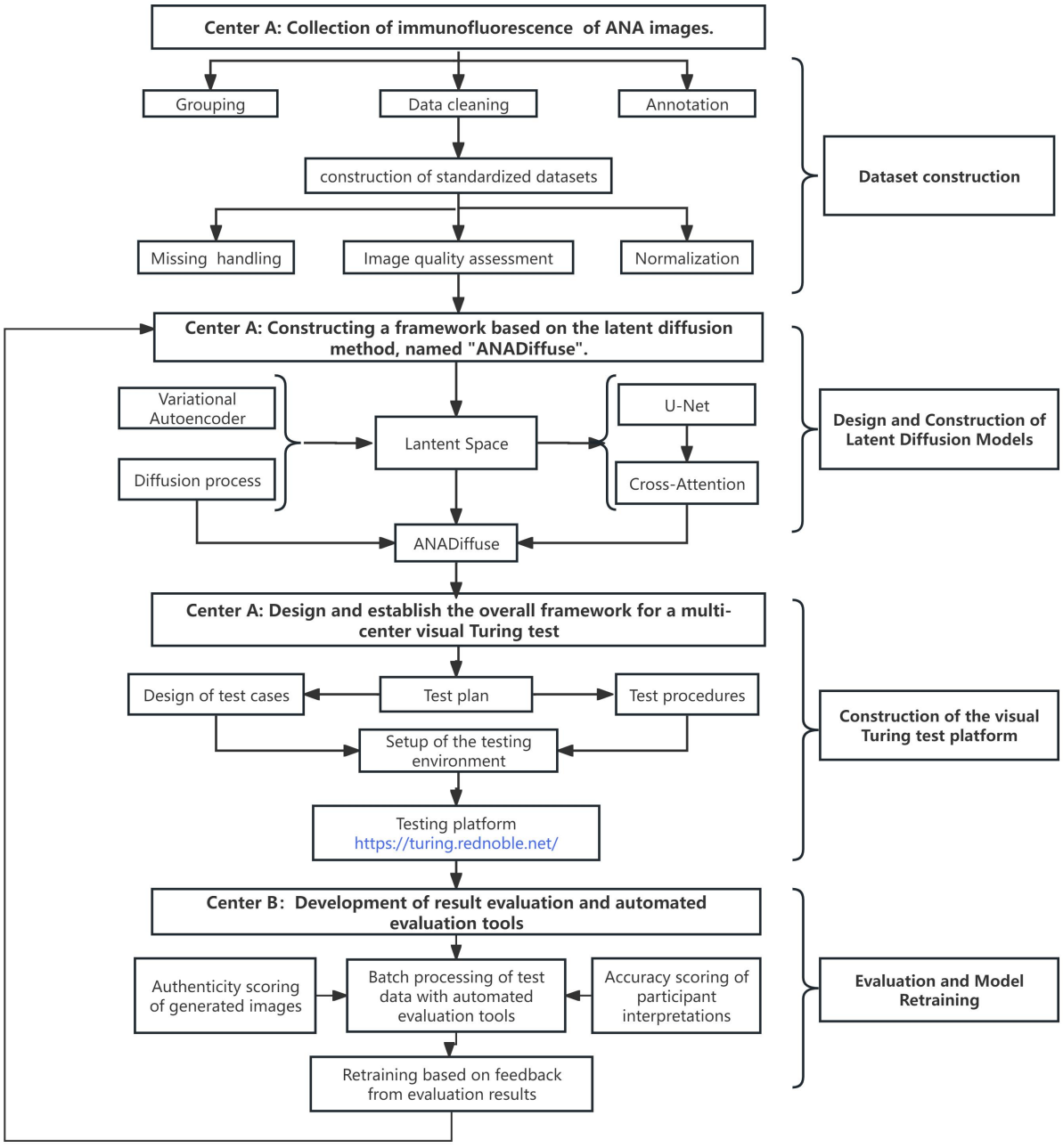

Figure 4. Study flow diagram.

7.2 Data collection

Study data will be stored in an electronic laboratory records database hosted by Shanghai Jiao Tong University Affiliated Xinhua Hospital. The collection of data will be oriented around a thorough assessment of the ANA test reports.

7.3 Network architecture

The architecture diagram of the model is presented below, with no further textual description provided here.

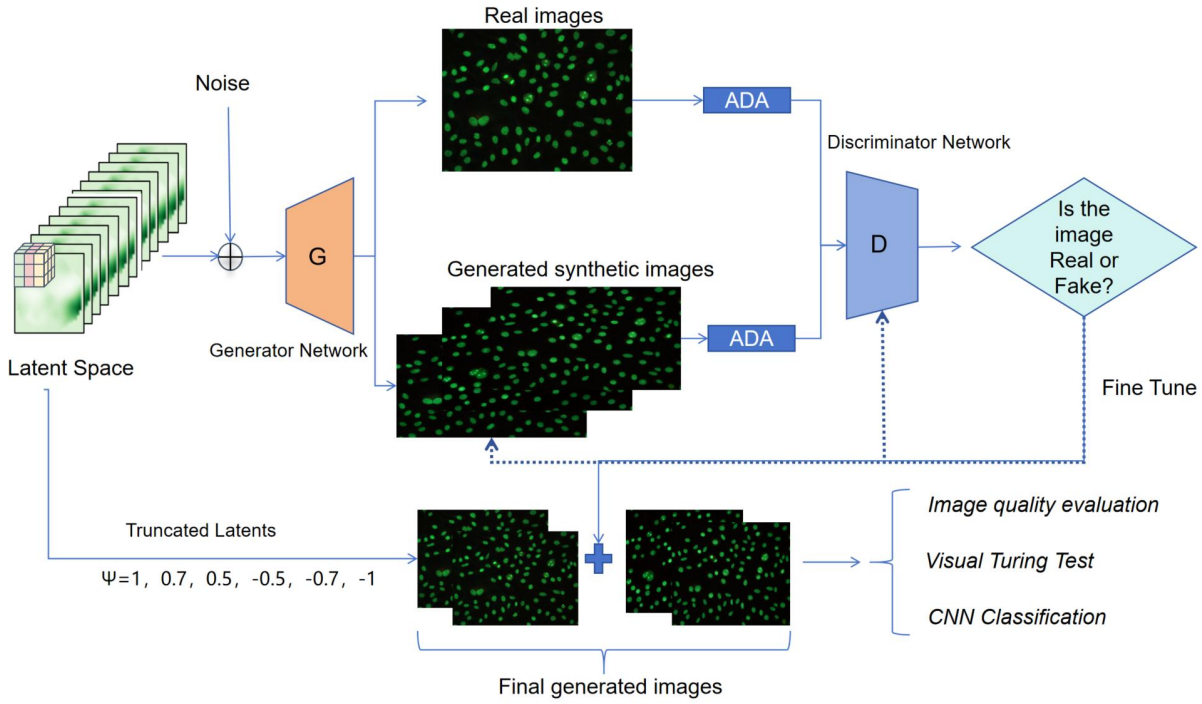

Figure 5. Network architecture diagram of the proposed latent diffusion model**.**

7.4 Data analysis

To implement a more conservative discriminatory approach, we calculated the subsequent evaluation criteria, including precision, recall, F1 score, and finally to mean average precision (mAP). Additionally, the confusion matrix for each category of rare patterns was also calculated. The two-sided McNemar test was used to compare the precision disparities between medical technologists and the AI model. Additionally, Cohen's kappa coefficient was calculated to evaluate inter-observer agreement.

7.5 Data safety monitoring committee

This trial does not include a data safety monitoring committee, as it utilized an epidemiological approach to analyze data collected during routine laboratory testing, without using any patient-identifiable information

8.0 NOVELTY

The aspiration to develop diffusion models for synthesizing images is not inherently novel, and we have mentioned the partially successful yet simplistic and insufficient generative adversarial networks as a precedent. However, the substantially broadened range of parameters and their advanced sophistication that we propose to incorporate will elevate this endeavor to a level that may satisfy both the critical (and currently unaddressed) demands for practical applications in various fields and, more fundamentally, shed light on the intricate visual distinctions between the synthesized images and their real-world counterparts. Through this process of refinement, the pivotal and fundamental cause-and-effect relationships governing image synthesis can be uncovered. This understanding could lead to the identification of new directions and/or innovative avenues for image generation techniques, enabling a significant leap forward in the quality and diversity of synthesized images. In summary, the vastly expanded breadth and depth of parameters utilized in model construction, the significantly enhanced potential for practical and scientific value, the complementary yet essential attention given to both technical limitations and visual fidelity, and particularly the utilization of an in-depth analysis of the underlying data structures and patterns, are all individually groundbreaking and collectively unprecedented in the context of image synthesis.

9.0 ETHICAL CONSIDERATIONS

The study protocol was determined to be exempt by the Institutional Review Board of the Xinhua Hospital affiliated to Shanghai Jiao Tong University School of Medicine (XHEC-C-2024-123). The Principal Investigator will obtain ethical approval and clinical trial authorization by competent authorities according to local laws and regulations.

9.1 Informed Consent

Patient consent was waived due to the design of the study.

9.2 Confidentiality of Participant Records

The names and personal information of study participants will be held in strict confidence and restricted to members of the study team. The data coordinator will maintain a confidential participant identification list (i.e. master list) during the study. Access to confidential information (i.e., source documents and patient records) is only permitted for direct participant management and for those involved in monitoring the conduct of the study (i.e., Sponsors, CRO's, representatives of the IRB/REB, and regulatory agencies). The participant's name will not be used in any public report of the study.

10 REFERENCES

[1] Pisetsky DS. Antinuclear antibody testing - misunderstood or misbegotten. Nat Rev Rheumatol. 2017. 13(8): 495-502.

[2] Hobson P, Lovell BC, Percannella G, Saggese A, Vento M, Wiliem A. Computer Aided Diagnosis for Anti-Nuclear Antibodies HEp-2 Images: Progress and Challenges. PATTERN RECOGNITION LETTERS. 2016. 82(oct.15): 3-11.

[3] Wiik AS, Høier-Madsen M, Forslid J, Charles P, Meyrowitsch J. Antinuclear antibodies: a contemporary nomenclature using HEp-2 cells. J Autoimmun. 2010. 35(3): 276-90.

[4] von Mühlen CA, Garcia-De La Torre I, Infantino M, et al. How to report the antinuclear antibodies (anti-cell antibodies) test on HEp-2 cells: guidelines from the ICAP initiative. Immunol Res. 2021. 69(6): 594-608.

[5] Damoiseaux J, Andrade L, Carballo OG, et al. Clinical relevance of HEp-2 indirect immunofluorescent patterns: the International Consensus on ANA patterns (ICAP) perspective. Ann Rheum Dis. 2019. 78(7): 879-889.

[6] Vermeersch P, Bossuyt X. Prevalence and clinical significance of rare antinuclear antibody patterns. Autoimmun Rev. 2013. 12(10): 998-1003.

[7] Tomić Sremec N, Kozmar A, Sremec J, Anić B, Batinić D. Properties of Uncommon Indirect Immunofluorescence Staining Patterns Determined during Antinuclear Antibody Detection on HEp-2 Cells. J Clin Med. 2021. 10(17): 3866.

[8] Nanda R, Gupta P, Patel S, Shah S, Mohapatra E. Uncommon antinuclear antibody patterns as diagnostic indicators. Clin Biochem. 2021. 90: 28-33.

[9] Bonroy C, Vercammen M, Fierz W, et al. Detection of antinuclear antibodies: recommendations from EFLM, EASI and ICAP. Clin Chem Lab Med. 2023. 61(7): 1167-1198.

[10] Chan E, von Mühlen CA, Fritzler MJ, et al. The International Consensus on ANA Patterns (ICAP) in 2021-The 6th Workshop and Current Perspectives. J Appl Lab Med. 2022. 7(1): 322-330.

[11] Wu YD, Sheu RK, Chung CW, et al. Application of Supervised Machine Learning to Recognize Competent Level and Mixed Antinuclear Antibody Patterns Based on ICAP International Consensus. Diagnostics (Basel). 2021. 11(4): 642.

[12] Rigon A, Infantino M, Merone M, et al. The inter-observer reading variability in anti-nuclear antibodies indirect (ANA) immunofluorescence test: A multicenter evaluation and a review of the literature. Autoimmun Rev. 2017. 16(12): 1224-1229.

[13] Naides SJ, Genzen JR, Abel G, Bashleben C, Ansari MQ. Antinuclear Antibodies Testing Method Variability: A Survey of Participants in the College of American Pathologists' Proficiency Testing Program. J Rheumatol. 2020. 47(12): 1768-1773.

[14] Zeng J, Gao X, Gao L, Yu Y, Shen L, Pan X. Recognition of rare antinuclear antibody patterns based on a novel attention-based enhancement framework. Brief Bioinform. 2024 Jan 22;25(2):bbad531. doi: 10.1093/bib/bbad531. PubMed ID: 38279651