i-SART: An Intelligent Assistant for Safety Analysis in Radiation Therapy

Natalia Silvis-Cividjian[1][a], Yijing Zhou[1][b], Anastasia Sarchosoglou[2][c] and Evangelos Pappas[2][d]
[1]Vrije Universiteit Amsterdam, Department of Computer Science, Amsterdam, The Netherlands
[2]University of West Attica, Department of Biomedical Sciences, Athens, Greece

Keywords: Radiation Therapy, Safety and Risk Management, Digital Assistive Technology, Failure Modes and Effect Analysis (FMEA), Natural Language Processing (NLP), Generative AI, Synthetic Data.

Abstract: Along with surgery and chemotherapy, radiation therapy (RT) is a very effective method to treat cancer. The process is safety-critical, involving complex machines, human operators and software. A proactive hazard analysis to predict what can go wrong in the process is therefore crucial. Failure Modes and Effect Analysis (FMEA) is one of the methods widely used for risk assessment in healthcare. Unfortunately, the available resources and FMEA expertise strongly vary across different RT organizations worldwide. This paper describes i-SART, an interactive web-application that aims to close the gap by bringing together best practices in conducting a sound RT-FMEA. Central is a database that at present contains approximately 420 FMs collected from existing risk assessments and cleaned from ambiguities and duplicates using NLP techniques. Innovative is that the database is designed to grow, due to both user input and generative AI algorithms. This is work in progress. First experiments demonstrated that using machine learning in building i-START is beneficial. However, further efforts will be needed to search for better solutions that do not require human judgment for validation. We expect to release soon a prototype of i-SART that hopefully will contribute to the global implementation and promotion of safe RT practices.

1 INTRODUCTION

Cancer is the second leading cause of death worldwide. About 40% of world's population will be diagnosed with cancer at some point during their lifetimes (NCI, 2017). Radiation therapy (RT) is a highly effective cancer management approach received by approximately 50% of all patients. One can say that RT is a field where healthcare meets informatics. The process takes place in complex, computer-controlled linear accelerators (linacs), where high-energy ionizing radiation is used to reduce or eliminate the tumor(s) and at the same time sparing the healthy tissue (Fig. 1). The core RT team consists of different healthcare professionals, including radiation oncologists, medical physicists, radiation therapists, dosimetrists and nurses. A generic process RT process is illustrated in Fig. 1. After a patient is referred for radiotherapy and assessed by a radiation oncologist, the next step involves an imaging exam, usually a CT localization scan. On these images, the radiation oncologist delineates the specific regions that have to be irradiated and prescribes the dose in each of these regions. After that, the treatment planning and treatment delivery teams accurately follow this prescription and deliver the needed radiation, by using the linac and various types of software products. During the whole process, Quality Assurance (QA) and patient monitoring activities are mandatory.

A few devastating accidents that occurred in the last decades demonstrate that the RT process is safety-critical - any mistake, be it caused by hardware, software or humans, can have fatal consequences (Leveson and Turner, 1993), (Borras et

[a] https://orcid.org/0009-0004-4668-5946
[b] https://orcid.org/0009-0007-6487-0004
[c] https://orcid.org/0009-0007-6487-0046
[d] https://orcid.org/0000-0001-8052-6392

Silvis-Cividjian, N., Zhou, Y., Sarchosoglou, A. and Pappas, E.
i-SART: An Intelligent Assistant for Safety Analysis in Radiation Therapy.
DOI: 10.5220/0012364400003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 420-427
ISBN: 978-989-758-688-0; ISSN: 2184-4305

al., 2006). Therefore, RT is nowadays a strongly regulated process, with safety standards in place (Council of European Union, 2014), (IAEA, 2018). According to these standards, an RT process needs to be thoroughly assessed for all the risks it poses, before obtaining permission to proceed (Huq et al., 2016). This can be addressed with a proactive risk analysis, which aims to anticipate failure modes (FM) or hazards, defined as conditions that can lead to incidents, or in other words, the various ways a system can fail.

To conduct a safety assessment, the analysts can choose from a range of systematic methods, such as the traditional Failure Mode and Effects Analysis (FMEA), Fault Tree Analysis (FTA) and Hazard and Operability Analysis (HAZOP) (Pawlicki et al., 2011), or the more modern Systems Theoretic Accident Model and Process (STAMP) (Pawlicki et al., 2016), (Silvis-Cividjian et al., 2020). All these methods work in the same way: first, a team identifies the potential hazards in a process, addresses their causes and evaluates their effects, and finally formulates appropriate mitigation measures.

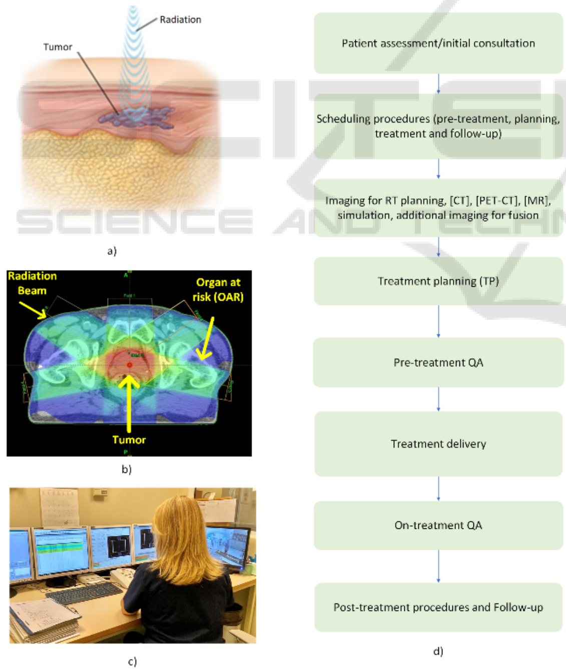

Figure 1: a) The principle of RT; b) The geometry of RT, where one can see that the tumor receives the highest radiation dose (in red) and the healthy tissue the lowest (in blue). From (Kane, 2014); c) A view from a typical treatment facility room, where a radiation therapist needs to distribute their attention over many computer screens. Credits to A. Sarchosoglou; d) The workflow of a generic RT process.

A general problem is that many RT departments lack the time, training, or manpower required to perform an in-depth risk assessment. Another problem is that knowledge tends to remain compartmentalized within departments, with safety analysis results often not being shared widely. For example, currently there is no centralized database with potential RT-specific FMs that could be used as a reference by practitioners who intend to conduct a proactive risk analysis. This is a missed opportunity in our opinion, because despite their diversity, all RT process workflows feature in fact sufficient common FMs.

On the other hand, assistive and data mining software applications, often powered by artificial intelligence (AI), are rapidly emerging in all domains of our daily life, including healthcare and RT. Examples are software systems for electronic patient dossiers, prediction of the response to a treatment, disease risk assessment, or, specific for the RT domain, radiation dose calculation, automatic delineation of tumors and organs at risk on CT scan images, or defacing of CT images of head-and-neck cancer patients for privacy reasons, etc.

In this paper, we will present an attempt to close the gap and improve RT safety worldwide with i-SART, an online platform that assists practitioners in performing an effective proactive FMEA-based safety analysis. Central is a novel database that brings together a large number of possible RT-specific FMs, formulated in English and free of ambiguities or duplicates. Innovative is that the database is designed in such a way that new FM data can be fed not only by safety-aware RT practitioners around the globe, but also by state-of-the-art generative AI (GenAI) algorithms. To the best of our knowledge, this is the first attempt to use GenAI for synthetic FMs. As this is work in progress, synthetic FMs were not included yet in the i-SART database. Nevertheless, a prototype of i-SART will be soon released for all interested RT-practitioners.

The remainder of the paper is organized as follows. In Section 2 we formulate the problem we try to solve with i-SART, in Section 3 we will present the design of i-SART, its database and user interface. Section 4 will present some preliminary results and Section 5 will outline our conclusions and future work plans.

2 PROBLEM STATEMENT

First used by the US Military at the end of 40's, FMEA is a safety assessment method widely adopted in systems engineering in 60's (Arnzen, 1966).

FMEA has been also widely used and recommended for healthcare in general and RT in particular, in order to prevent medical errors propagating and reaching the patient (Ibanez, 2018), (Olch, 2014), (deRosier, 2002).

The general process flow of an FMEA is illustrated in Fig. 2. The method is bottom-up, meaning that for each component in the process, one have to ask the question "*How often would this component fail, and what will happen if it fails?*". The risk of each FM is evaluated using a Risk Priority Number (RPN), calculated as the product of severity, probability of occurrence, and detectability.

The result of an FMEA analysis is a list with all possible FMs, ranked by their RPNs, their causes and their effects, followed by measures to mitigate the most critical ones. In an RT process, some FMs that can occur are readily conceivable, such as "*A wrong patient is invited to the treatment room*" or "*The linac gantry in rotation collides with the treatment couch*". However, to conduct an analysis that will predict all ways a process can fail is challenging. This task demands considerable time, domain knowledge and clinical experience. Unfortunately, these resources may not always be readily available as the main task of RT practitioners is to treat as many patients as possible in a race against the clock with fast-evolving cancer.

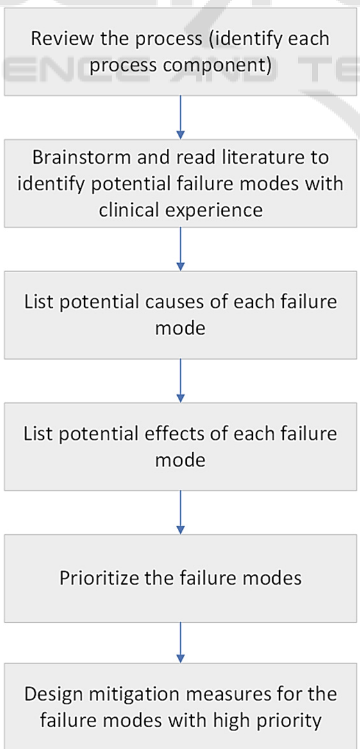

Figure 2: The workflow of an FMEA analysis.

To gain more insights into the current FMEA practices within the RT, we recently conducted a literature review that highlighted the diversity between departments in the way they conduct their FMEA (Sarchosoglou et al., 2022). A challenge we encountered was for example the heterogeneity of the terminology used to formulate FMs. Nonetheless, on a more optimistic note, our findings also revealed noteworthy similarities and common FMs that support the need for a digital tool to aid departments with their proactive risk assessment. Furthermore, the level of safety awareness maturity strongly varies among different RT organizations. In addition, the knowledge about FMs is compartmentalized; if a safety analysis was conducted somewhere, its results usually stay in the department and are not widely disseminated. Moreover, as technology in RT rapidly advances, new, previously unidentified risks are continuously emerging, presenting challenges to professionals tasked with their implementation (Ortiz-Lopez, 2009). Finally, education material and non-proprietary digital tools to assist safety analysts are practically non-existent.

The cumulative effect of all these challenges is that incidents and errors of suboptimal RT treatment still occur on daily basis (Ford and Evans, 2019). Hence, there is an urgent need for assistance in conducting hazard analyses, with the ultimate goal of enhancing the safety of RT patients worldwide. This imperative served as the driving force behind the initiation of the i-SART project, a collaborative effort between the computer science department at the Vrije Universiteit in The Netherlands and the biomedical sciences department at the University of West Attica in Greece. The project addressed the following research questions:

RQ1. Can we build an open-source software tool to assist RT practitioners in conducting an effective FMEA?

RQ2. Can we use machine learning to augment the data obtained by the FMEA studies ?

3 i-SART, AN INTELLIGENT FMEA ASSISTANT

The main goal of i-SART was to engage RT practitioners in a dynamic FMEA learning experience. Given the fact that RT professionals may or may not have experience in FMEA, we expected the usage of this tool to vary accordingly. On the one hand, FMEA beginners can use it as an expert system to guide their analysis. On the other hand, RT experts who are proficient in FMEA will be able to learn about new FMs reported by other departments, or share interesting FMs they have identified in their institutions.

Technically, i-SART is designed as a cloud-hosted web-application with two kinds of users: administrators and RT practitioners (users), each with their permissions and rights. The high-level architecture of i-SART is illustrated in Fig. 3. Central to i-SART is a novel database that aims to contain as many as possible FMs that can occur in various RT techniques, such as Intensity Modulated Radiotherapy (IMRT), Stereotactic Body Radiation Therapy (SBRT), etc. In the database, the FMs are also grouped per sub-process and step in the RT sub-process, such as Treatment planning, Treatment delivery, etc. as illustrated in Fig. 1.

The web-application's back-end was programmed using Python 3.9 and the Django REST [1] framework. Care was taken to ensure a secure transfer of information between client and server, using the JSON Web Token authentication [2]. A relational database MySQL technology was used to achieve persistent storage of both users' and FM information. The front-end user interface of i-SART was developed using a Java-script-based framework called Vue.js[3]. Its main function is to allow an RT user to search in i-SART for FMs that might happen in their particular RT process. There are also searching, filtering and ordering functions available. For example, if a user discovers a new FM in their daily practice, they can add it to the i-SART database, after a preliminary validation by the system administrator, who is an RT expert. The administrator can also visualize on a dashboard all the FMs in different pie-charts, categorized based on their severity or risk (see Fig. 4). A Vue chart component library called Vue-ECharts[4] was used to plot these charts. Finally, all users have the possibility to evaluate the tool and send feedback and suggestions to improve it.

4 FIRST RESULTS

4.1 First FM Data

The first step after the skeleton of the i-SART prototype was ready, was to populate the FM database. First 584 FMs were extracted from scientific papers and un-published safety reports, all written in English (see Table 1). All FMs were classified into subprocesses and their corresponding steps. Interesting is that we found that 32.5% of the collected FMs fall into the subprocess Treatment planning, 25.5% into the subprocess Treatment delivery, and 20.7% into subprocess Imaging.

Table 1: Sources for the first collection of FMs .

Reference type	#FMs	Author/Year/Country
Research paper	16	Gilmore et al. (2021), UK
	284	Pobbiati et al. (2019), Italy
	220	Huq et al. (2016), USA
	7	Bright et al. (2022), USA
	26	Gehad et al. (2021), Egypt
Report from individual RT department	10	Not published, UK, 2022
	21	Not-published, UK, 2021

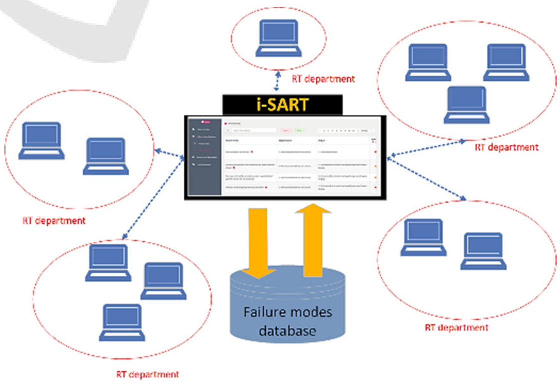

Figure 3: The high-level architecture of i-SART.

[1] https://www.djangoproject.com/
[2] https://django-rest-framework-simplejwt.readthedocs.io/en/latest/
[3] https://vuejs.org/
[4] https://github.com/ecomfe/vue-echarts

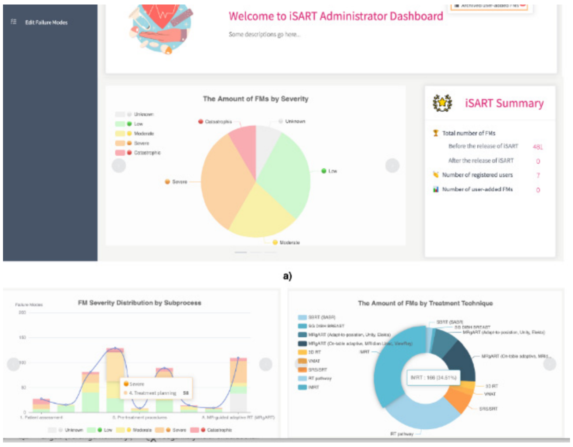

Figure 4: A screenshot of the administrator's dashboard screen in i-SART.

4.2 Cleaning Data to Eliminate Duplicates

Very soon however, we discovered that these "raw" FMs contained many duplicates and ambiguities, inherent to any text formulated in natural language. We distinguished two types of duplicates: *explicit*, defined as two exactly same FMs, which were easy to detect, and *implicit* duplicates, where the semantics was the same, but the syntax was different, which were more difficult to detect. For example, we considered the following two FMs as implicit duplicates.

***Collision** risk due to **gantry** rotation*

***Gantry collision** with visual aid device*

We detected implicit duplicates using both manual review by our RT expert team, and automated Natural Language Processing (NLP) algorithms. For example, first, the RT experts extracted a group of keywords such as *wrong, poor, imperfect*, to help to identify potential duplicates. Next, for each keyword, all kinds of forms (i.e., verb, adjective, noun, adverb, singular, plural) were generated using two NLP libraries called inflect[5] and word-forms[6] and finally were added to the keyword list. While we were aware that words like "poor" and "wrong" may have different causes and effects, we considered two FMs containing these words as candidates for duplicates. We eventually classified them as real duplicates only after a thorough validation by our RT experts.

Next, we inspected FMs that exhibited a tree-like structure. For instance, let us take a look at the following FMs, belonging to the subprocess Imaging and its step "CT image acquisition":

Wrong CT scan for treatment planning: wrong choice of anatomical volumes

Wrong CT scan for treatment planning: fiducial markers not implanted

Wrong CT scan for treatment planning (Vero®): Optoelectronic markers not completely included into the scan

Wrong CT scan for treatment planning (CyberKnife®): scan volume not compliant to the specifics requirements of the TPS

In all these four FMs, the text to the left of the colon (:) describes the same unsafe situation (Wrong CT scan), whereas the text to the right of the colon is an elaboration on the specific causality. We suspected that these FMs might be treated as implicit duplicates, or at least be clustered in the same FM group. Again, this happened in reality only after a validation by the RT experts.

As a result of all these data-cleaning procedures, we were able to eliminate 57 explicit duplicates, 37 implicit duplicates based on keywords and 130 implicit duplicates based on the tree-like structure. Given the fact that an FM can be flagged as duplicate multiple times by different filtering methods, we took action to ensure each FM only appears once. As a result, the total amount of uniquely duplicated FMs was reduced to 166. Eventually, we ended up with 584-166 = 418 unique FMs, which we entered into the database. We have to note that this is an indication of the number of FMs available so far. This work is in progress. Our team is working on fine-tuning the application and improving the database. A prototype of i-SART with approximately this number of FMs will be soon available to be used and tested by interested RT practitioners. To conclude, we would like to emphasize that during the process of eliminating the duplicates, the final decisions need to be taken by our RT experts' team, who will ensure that no critical FM will get excluded by mistake.

4.3 Augmenting the Database with Synthetic FMs

Although we were initially satisfied with the way we populated the i-SART database, we also investigated the possibility of augmenting the database with new, synthetic AI-generated FMs. The reason for this was the consideration that if the database will be used in

[5] https://pypi.org/project/inflect/

[6] https://pypi.org/project/word-forms/

the future for training machine learning algorithms, a few hundred FMs will be definitely not sufficient to achieve a high prediction accuracy. For example, deep learning models used in NLP use typically training datasets containing millions of items (Bailly, 2022). In this section, we will present a few interesting results. We have to note here that these were separate experiments and none of these results have been yet implemented in i-SART.

Generative AI (GenAI) is a modern, powerful technology that can produce new plausible media content from existing content, including text, images, audio, etc trying to mimic human creativity. It originates in the research done at Google in 2017 (Vaswani, 2017) that first analyzed a language trying to discover patterns in it, and then transformed this analysis into a prediction on which word or phrase should come next. Many GenAI algorithms exist, varying from the probabilistic Naïve Bayes Networks and Markov Models, to all kinds of feature-based neural-networks variations, such as recursive neural networks (RNN), convolutional neural networks (CNN), and the GPT-2, -3 and -4 series, where GPT stands for "Generative Pre-trained Transformer". Regardless the algorithm, automated text generation works basically in the same way. In the beginning, all probabilities or adjustable weights in the neural network are unknown; we say that the model is not trained. However, the model can *learn* these parameters if provided with a huge number of training examples. Eventually, when the training is finished, and one starts with one word (also called prompt), the model will be able to accurately predict the most likely next word in a phrase.

Therefore, GenAI seemed a perfect approach suitable for our purpose. We had a rather small collection of training text data (the FMs) and we wanted an AI algorithm to learn how to create more, synthetic FMs. In line with these thoughts, we conducted two preliminary experiments that explored the performance of different GenAI algorithms.

The first experiment, in the context of a MSc CS graduation project (Haddou, 2022), used two different algorithms, Markov Chains and ChatGPT-3 to learn how to create new FMs based on an existing collection. The training database was slightly different, containing around 600 FMs collected from literature and a few RT departments in Europe. From these, eleven FMs that were generated with the Markov Chains algorithm were presented for validation to an RT expert. Out of these six were found useful. There was at least one artificial FM with a high RPN, namely "*Incorrect image data set associated with patient shifts determined"* that was interpreted by the RT expert as "patient shifts determined by incorrect image data set". Another FM was very interesting because the RT expert had seen it a lot of times before, namely '*Patient head's position is not ideal'*. The RT expert noted that this FM would never come spontaneously to her mind. This FM was clearly and correctly generated by the Markov Chain algorithm.

The GPT-3 algorithm generated eleven FMs that were also presented to the same RT expert. Out of these, four of them were found useful. Especially the FM: "*Patient or nurse falls"* and *"Patient falls down due to mobile phone dropping on the floor*" were very interesting. We could conclude from here that synthetic FMs have the potential to raise awareness or reveal unpredicted hazards that might occur during any process, not necessarily RT specific.

The second experiment used a Generative Adversarial Network (GAN) algorithm to generate artificial FMs (Brophy, 2023). As a training dataset we used our most recent FM database. GANs are a branch of GenAI algorithms that consist of two artificial neural networks, called Generator and Discriminator (Goodfellow, 2020). The Generator tries to generate new data as similar to original data as possible, while the Discriminator's role is to determine if the input belongs to the real dataset or not. The optimization process is characterized as a game where the generator successfully learns to "fool" the discriminator in such a way that the discriminator cannot distinguish between the real one and the synthetic one. In particular, our experiment used the seqGAN model (Yu, 2017).

The Bilingual Evaluation Understudy (BLEU) score was one of the metrics used to measure the quality of the FMs produced by the generative algorithm. The basic idea of the BLEU score is straightforward: the closer the synthetic FM is to the human-generated target sentence, the better it is; a score of 1 means a perfect match, and 0 means no match. A BLEU-score has different levels (n), depending on how many n-grams are being compared. For example, if n=1, each word from the original and synthetic text will be compared, and if n=2, each word pair will be matched. As training data we used all the 584 raw FMs initially collected as described in section 4.1, plus 1721 FM taken from the headlines of incidents reported in IAEA SAFRON, (SAFety in Radiation ONcology), an international platform that collects RT incidents and promotes patient safety[7].

The seqGAN model implemented in Pytorch[8] was able to create additional 640 artificial FMs. From these, 230 were identified as useful by an RT expert, with 53 duplicates. A handful of them (only 9) were considered as really novel with respect to the existing FMs in the database (see Table 2). However, so far we found that the synthetic FMs lacked syntactic accuracy and clarity. Fig. 5 shows the performance of the seqGAN algorithm measured using the BLEU scores with levels n = 2, 3, 4, 5. The black dot line splits the training into two phases: 1) Before the divider is the pre-training process where the initial generator was trained, and 2) after the divider is the adversarial training process where the generator continues to update based on the reward from the discriminator. We can see that the more grams the calculation of a BLEU score is based on, the lower the score. In our experiment, the BLEU-2 values were the highest and reached the maximum value of 0.6.

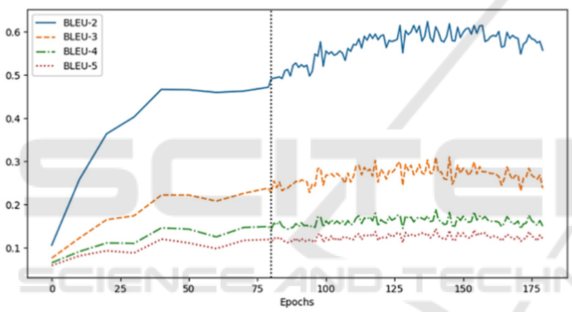

Figure 5: The BLEU-[2, 3, 4, 5] scores of the synthetic FM generation using the seqGAN model.

These first results show that using GenAI algorithms is an interesting idea to generate synthetic FM. However, more efforts will be needed in future to increase their accuracy and eliminate the need of human judgement.

Finally, we also identified a few limitations to this approach. For example, when deciding if the newly generated FM is a valid one, we consulted only one RT practitioner, while the assessment of any FM in a process needs an RT team. Moreover, we didn't include the steps in which an FM could occur. This would bring more clarity to the results. We also expect that a larger FM dataset will also improve the accuracy. This will happen in time, when i-SART will be used by a large community of RT practitioners. Moreover, while AI integration is innovative, there' s a risk of overconfidence in AI-generated FMs without adequate human oversight.

Table 2: Novel, synthetic FMs created using GANs.

Synthetic Failure mode	**Correction/Comments by RT expert**
incorrect isocentre not used	*incorrect isocentre used*
patient positioning with wrong tattoo	*(Nice! I haven't seen this in the 584 FMs!)*
patient was treated with wrong side of tattoo	*(sounds similar to previous one)*
incorrect selection of appointments delivery recorded on verification system	*incorrect selection of appointments on record and verify system*
PTV received higher and treated on the patient	*A higher dose was prescribed for the PTV* and was delivered to the patient*
incorrect collimator angles not imported	*collimator angles not imported (Yes, although I'm not sure if this is technically possible.)*
old treatment protocol	*use of the old treatment protocol instead of the new one*
wrong field size on portal image	*this is fully correct*
planned for the wrong beam angles for one of treatment fields	*wrong beam angles for one of treatment fields*

**PTV means Planning Target Volume and is the region around the tumor that needs to be irradiated*

5 CONCLUSION AND FUTURE WORK

We presented i-SART, a novel web-application that aims to assist RT practitioners in conducting a sound proactive safety assessment using FMEA. i-SART is the result of a successful cooperation between RT and computer science researchers. Central is an FM database that can grow due to contributions from participating users. We also experimented with machine learning techniques, such as NLP for duplicates elimination and generative AI to create synthetic FMs. We conclude that although machine learning can be useful in assisting a safety assessment process, the results need to be always validated by RT experts. Future work includes optimizing the machine learning algorithms, including a variety of other

[7] https://www.iaea.org/resources/rpop/resources/databases-and-learning-systems/safron

[8] https://github.com/williamSYSU/TextGAN-PyTorch

safety analysis methods besides FMEA and investigating the possibilities to offer i-SART as an open-source collaborative tool for the international RT community with the common goal of contributing to a safe and fair global healthcare.

ACKOWLEDGEMENTS

We would like to thank prof. Annette ten Teije for her support in pursuing our research. A prototype of i-SART is currently hosted in the SciCloud, a virtual server service for research offered by the Vrije Universiteit in Amsterdam.

REFERENCES

Arnzen, H.E. (1966). Failure Modes and Effect Analysis: Powerful Engineering Tool for Component and System Optimization. *Reliability and Maintainability Conference*, 355-71.

Bailly, A. et al. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. *Computer Methods and Programs in Biomedicine*, Volume 213, 106504.

Bright, M. et al. (2022). Failure modes and effects analysis for surface-guided DIBH breast radiotherapy. *Journal of applied clinical medical physics,* 23(4), e13541

Borrás, C. et al. (2006). Overexposure of radiation therapy patients in Panama: problem recognition and follow-up measures. *Revista Panamericana de Salud Pública*, 20(2-3), 173-187.

Brophy, E. et al. (2023). Generative Adversarial Networks in Time Series: A Systematic Literature Review. *ACM Comput. Surv*. 55, 10, Article 199.

Council of the European Union. (2014). Directive 2013/59/Euratom on basic safety standards for protection against the dangers arising from exposure to ionising radiation and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom, *Official Journal of EU*. L13; 57:1-73.

deRosier, J. et al. (2002). Using health care failure mode and effect analysis. Joint Commission *J. Qual. Improvement* 27, 248–267.

Ford, E. and Evans, S.B. (2018). Incident learning in radiation oncology: A review, *Medical Physics*, 45: e100-e19.

Gehad, S.M. et al. (2021). Failure Mode and Effect Analysis for Three Dimensional Radiotherapy at Ain Shams University Hospital, *Int. J. of Adv. Res*. 9 (Dec). 971-983.

Gilmore, M. D. and Rowbottom, C. G. (2021). Evaluation of failure modes and effect analysis for routine risk assessment of lung radiotherapy at a UK center. *Journal of applied clinical medical physics*, 22(5), 36-47.

Goodfellow, I. et al. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.

Haddou, S. 2022. iSART: Digital safety assistant for the radiotherapy process. Msc CS thesis, Vrije Universiteit Amsterdam.

Huq, M. S. et al. (2016). The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management, *Medical Physics*, 43: 4209-62.

Ibanez-Rosello, B. et al. (2018). Failure modes and effects analysis of total skin electron irradiation technique. *Clinical and Translational Oncology*, 20(3), 330-365.

International Atomic Energy Agency (IAEA). (2018). Radiation protection and safety in medical uses of ionizing radiation. *Safety Standards Series* No. SSG-46;

Kane, P., The impact of the VERT virtual reality system on teaching and learning associated with radiation therapy planning skills in the second year of the Bachelor of *Radiation Therapy*. 2014

Leveson, N.G. and Turner, C.S. (1993). An investigation of the Therac-25 accidents, in *Computer*, vol. 26, no. 7, 18-41.

National Cancer Institute (NCI). (2017). Cancer Statistics, https://www.cancer.gov/, Accessed online 7-10-2023.

Olch, A. (2014). Quality and Safety in Radiation therapy Learning the New Approaches in Task Group 100 and Beyond, *Medical Physics*, 41: 067301.

Ortiz-Lopez, P. et al. (2009). Preventing accidental exposures from new external beam radiation therapy technologies. *Ann ICRP*, 39(4):1e86.

Pawlicki, T. et al. (2011). *Quality and Safety in Radiation therapy* (CRC Press).

Pawlicki, T. et al. (2016). Application of systems and control theory-based hazard analysis to radiation oncology. *Medical Physics,* 43(3):1514-30.

Pobbiati, C. et al. (2019). Clinical risk analysis of the patient's path in an Advanced Radiotherapy Center through FMEA method. *Journal of Biomedical Practitioners,* 3(1).

Sarchosoglou, A. et al. (2022). Failure modes in stereotactic radiosurgery. A narrative review, *Radiography*, Volume 28, Issue 4, Pages 999-1009.

Silvis-Cividjian, N. et al. (2020). Using a systems-theoretic approach to analyze safety in radiation therapy-first steps and lessons learned. *Safety Science*, *122*, 1-10.

Vaswani, A. et al. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17).* 6000–6010.

Yu, L. et al. (2017). SeqGAN: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence* (31,1).