# Unit 3

# Peer review activity

**Peer Review of 2020 ACR Data Science Institute Artificial Intelligence Survey**

The purpose of this study is to understand how radiologists are using AI in clinical practice and to establish a baseline for monitoring AI usage trends over time. The problem identified is the expected rapid growth of AI in medical imaging over the next decade. There is a need to understand how radiologists currently use AI and track these trends to aid the development of AI algorithms that improve medical care. Previous surveys have primarily focused on AI’s future use, making this study significant as it assesses AI's present application in radiology.

The study aims to evaluate two key aspects: the impact of implementing an AI algorithm for intracranial haemorrhage (ICH) detection on turnaround times in noncontract CT scans and whether the impact on turnaround time is influenced by how information is presented within the radiologist’s workflow. While the research question itself is not explicitly stated, the study attempts to measure the potential benefits of integrating AI algorithms into radiology practice by analysing their effects on workflow efficiency and turnaround times.

The study aligns with the broader understanding that AI has the potential to revolutionise medical imaging. However, there is a gap in studies assessing its real-time impact rather than its anticipated future use. The research is relevant as it helps bridge that knowledge gap, but certain aspects of its execution raise concerns.

The study utilised a survey method, conducted between April 13 – May 20, 2020. An electronic invitation was sent to all American College of Radiology (ACR) members via email with two reminders. The survey was conducted on SurveyMonkey (San Mateo, California) and included follow-up questions for respondents who did not use AI, asking why and whether they planned to invest in AI tools in the future. For predictive modelling, the study employed the Multi-Objective Optimization on the Basis of Ratio Analysis (MOORA) method, originally introduced by Brauers and Zavadskas in 2006. MOORA is typically used for decision-making across multiple objectives, focusing on evaluating alternatives based on predefined criteria. The study applies this approach by ranking different AI applications in radiology.

The survey method effectively reached a broad audience, and the email reminders ensured a higher response rate. The follow-up questioning for those who did not use AI was a valuable addition, as it provided insights into barriers to AI adoption. However, using MOORA as an evaluation tool is questionable. MOORA is typically applied in comparative efficiency studies, but in this case, the study does not focus on ranking efficiency across different AI tools. If the study aimed to determine which AI subcategory required more investment or fine-tuning, then MOORA would be more justified. However, for assessing AI’s integration in radiology workflows, this method seems somewhat misaligned with the study’s stated objectives.

The study categorised AI use in radiology into five groups: self-developed AI, mammography screening, CT Chest (Embolism detection), MR Brain Analytics, and CT Brain (Haemorrhage detection). Each category was equally weighted (0.25), indicating equal representation in the survey. However, the ranking of these categories based on undisclosed criteria showed mammography screening ranking first, followed by self-developed AI, CT Chest (Embolism), MR Brain Analytics, and CT Brain (Haemorrhage) ranking last. The ranking lacks clarity as the criteria used to assess performance, effectiveness, or impact are not explicitly stated.

The study collected quantitative data, as it involved numerical values, normalised rankings, and comparative metrics, such as total surveyed use, estimated total market use, and estimated total AI sales. The ranking system is based on undisclosed criteria, making it difficult to interpret the significance of the rankings. The survey questions were not disclosed, making replication or validation difficult. There is a disconnect between the introduction and results, as the survey’s objectives are not clearly mapped to the final findings. The study does not specify whether the assessment of AI tools was based on subjective opinions from radiologists or objective performance metrics like accuracy and speed. It is unclear whether all radiologists surveyed had experience with all AI tools, raising concerns about potential bias in responses. The definition of what qualifies as AI in the study is vague, leading to ambiguity.

The title does not accurately reflect the study’s main findings. The abstract does not summarise the study well, focusing more on justifying why the research was conducted rather than summarising results. The conclusion does not sufficiently review the objectives outlined at the beginning. The results appear more comparative between AI subcategories rather than an in-depth assessment of AI’s benefits in radiology.

The study could be improved by providing a brief educational video before the survey to ensure respondents fully understand the definitions and objectives. Clearly defining measurable metrics for assessing AI effectiveness and separating surveys into user experience assessment versus performance measurement would avoid confusion between subjective and objective evaluations.

For future research, expanding the survey to a larger sample size across multiple hospitals would improve generalizability. Clearly defining AI categories and performance metrics in future studies and providing more structured criteria for ranking AI tools would ensure transparency.

Due to the niche nature of this topic, this study is one of the few of its kind with this specific focus at the time. While numerous articles review AI in medical imaging and discuss its deployment, there is a significant lack of research assessing the real-world effectiveness of AI in medical imaging through a retrospective lens. This scarcity of comparable studies made it challenging to find relevant literature for direct comparison.

Allen, B., Agarwal, S., Coombs, L., Wald, C., & Dreyer, K. (2021). 2020 ACR Data Science Institute Artificial Intelligence Survey. Journal of the American College of Radiology, 18(8), 1153-1159. <https://doi.org/10.1016/j.jacr.2021.04.002>

**Peer Review of Epileptic Seizure Prediction Using Big Data and Deep Learning**

The purpose of this study is to create a seizure prediction system that is accurate, fully automated, patient-specific, and tuneable to an individual’s needs. The goal is to positively impact the quality of life of epilepsy patients by providing warnings when there is an increased risk of seizures. Epilepsy is a disease that varies greatly among patients, with seizures being brief and infrequent for 99% of the time. Despite their rarity, their unpredictability significantly impairs quality of life. A survey has indicated that unpredictability is the most debilitating aspect of epilepsy, highlighting the need for a system that provides early warnings for seizures.

The study focuses on the development, implementation, and evaluation of a clinically relevant seizure prediction system. The system must perform well and reliably across different patients, operate autonomously over long periods without requiring frequent maintenance, allow patients to set personal sensitivity preferences, and run in real-time on a low-power platform. The research method used involved recording iEEG signals using intracranial electrodes. These annotated signals were processed by a deep neural network trained to distinguish between preictal and interictal states. The resulting deep learning model was deployed onto the neuromorphic TrueNorth chip.

The research methodology used in this study is appropriate. The objective was clearly defined, and using iEEG signals to identify seizure occurrences is scientifically valid. This is the gold standard current for epilepsy (Jobst et al., 2020). The study detailed each step thoroughly, ensuring transparency. To avoid biases, an equal number of preictal and interictal samples were used during training. The data was split into training and testing sets based on time periods for each patient. Two months of iEEG data, containing at least one seizure for each patient, were used for training and calibration. A new model was trained after each month of incoming data, discarding older data to ensure that inference was conducted chronologically after training. The system's performance was evaluated effectively.

The study used data collected from a previous clinical trial involving an implanted seizure advisory system. Patients were continuously recorded for up to two years using an implanted 16-electrode iEEG system. The data was reviewed and annotated by expert investigators. The study included data from ten patients, collecting a total of 16.29 years of iEEG signal data and 2,817 seizures. The dataset is publicly available, making the study reproducible. The study also incorporated the hour of the day to improve prediction performance.

Since the study focuses on measurable, numerical data and statistical evaluation, it follows a quantitative research approach. However, one notable weakness is the small dataset. Important patient demographics such as age, gender, and ethnicity were not provided, making the dataset unrepresentative of the entire population. Despite this limitation, the study supports its claims with strong evidence and a solid methodology. The results section clearly outlines findings, followed by a detailed discussion. The conclusion effectively ties back to the study’s aims. The researchers demonstrated that the TrueNorth chip could be used as a wearable seizure prediction or intervention system, providing strong evidence for their claims.

Biases could be present in the dataset. Additionally, a larger dataset would enhance the study’s validity. A follow-up study involving patients using the TrueNorth chip alongside an iEEG signal machine would help assess the accuracy of the system in real-time. Future research should focus on expanding the dataset to include a broader range of patients across different ethnicities, ages, and genders to ensure the model's accuracy and applicability for diverse populations.

Kiral-Kornek, Isabell et al. (2018). Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System. eBioMedicine, 27, 103-111.

Jobst BC, Bartolomei F, Diehl B, Frauscher B, Kahane P, Minotti L, Sharan A, Tardy N, Worrell G, Gotman J. Intracranial EEG in the 21st Century. Epilepsy Curr. 2020 Jul;20(4):180-188. doi: 10.1177/1535759720934852. Epub 2020 Jul 17. PMID: 32677484; PMCID: PMC7427159.