



Innovation in Healthcare: Highlights from 2023

The following scientific highlights have been selected to showcase key findings across several therapeutic areas. The research abstracts summarised were presented at various congresses EMJ had the pleasure of attending, and were all award-winning, late-breaking, or highly commended innovative abstracts.

Citation:

EMJ Innov. 2024;8[1]:18-26. DOI/10.33590/emjinnov/10307465.
<https://doi.org/10.33590/emjinnov/10307465>.





ChatGPT Can Improve Imaging Report Readability for Patients

NATURAL language processing models could be used to provide concise summaries of radiology reports so that they are more accessible for patient interpretation, according to award-winning research presented at the European Congress of Radiology (ECR) 2023, which took place in Vienna, Austria, between 1st–5th March 2023.

Researchers explored whether ChatGPT (OpenAI, San Francisco, California, USA) could be used to reword radiology reports to increase readability, reduce complexity, and explain medical jargon in an understandable format for patients. To do this, the team used a mixture of 30 CT-positive and normal scan reports (abdominal CT [n=10], chest CT/CTPA [n=10], trauma pan CT scan [n=10]). Selection took place at random over a 1-week period from a tertiary hospital medical imaging department. Of the selected scans, six were reported by a consultant radiologist, 14 by senior radiology registrars, and 10 by junior registrars.

Following selection, ChatGPT was tasked with completing a revised version of the report conclusions. It was asked to do this without changing the meaning of the original, and to revise the report so that it could be understood by a person aged 14 years. Reports were evaluated using the Flesch Kincaid Reading Ease (FKRE) score. After calculating the mean

and standard deviation of FKRE scores, a paired t-test was used to compare the means of the original radiologist and ChatGPT reports. The simplified reports were reviewed by a consultant radiologist and radiology registrar to ensure accuracy and context retention.

The results revealed that following ChatGPT simplification of original reports, average FKRE score increased from 20.3 (standard deviation: 17.1) to 65.2 (standard deviation: 15.7), equating to a change in reading level from that expected of university educated graduates to a level appropriate for 12-year-old school children. Whilst the overall report readability improved, there were six reports for which no readability improvement occurred. In these instances, two reports were from trauma pan scans with multiple positive findings, and two other reports were short and deemed not suitable for simplification.

The researchers concluded that natural language processing models such as ChatGPT can create simplified summaries of radiology reports to optimise their readability and interpretation by patients. A limitation is that additional inappropriate and potentially incorrect information can be added into reports by these models. Thus, there does remain a requirement for oversight by a radiologist. ●

Deep Learning Models Predict Outcomes Following Liver Transplant

DEEP learning models can efficiently predict major adverse cardiovascular events (MACE) following liver transplantation (LT), according to new research presented at the European Association for the Study of the Liver (EASL) Congress 2023, held in Vienna, Austria. The team, from the University of Texas Health Science Center in Houston, USA, and Mayo Clinic, Jacksonville, Florida, USA, aimed to develop and validate deep learning models' ability to predict post-transplantation MACE among patients undergoing LT.

The authors identified patients who received an LT between January 2007–March 2020, and built multiple predictive models for the risk of developing any of the post-transplantation MACE, including myocardial infarction, atrial fibrillation, pulmonary embolism, heart failure, cardiac arrest, and stroke as an outcome. A MACE is primarily predicted using the Bidirectional Gated Recurrent Units (BiGRU) deep learning sequence processing model in different prediction interval lengths up to 5 years after the LT index date, using patients' demographics and retrospective diagnosis, medications, and procedures to claim data recorded back to 3 years before the LT index date. Performance of

the deep learning model against other machine learning models, such as logistic regression, random forest, and light gradient-boosting machine, was assessed using a cohort of 18,304 LT recipients (mean age: 57.4 years; 60.9% male; 39.1% female). Models' optimisation was done using five-fold cross validation on 80% of the cohort (training set), and the performance of the models was assessed using the remaining 20% (testing set), based on the area under the receiver operating characteristic curve and the area under the precision-recall curve.

Using different prediction intervals (0–30 days, 0–1 year, 0–3 years, and 0–5 years) after the LT index date, and compared to the three machine learning models, the top-performing model was the deep learning model, BiGRU, which achieved an area under the receiver operating characteristic curve of 0.833 (95% confidence interval: 0.8127–0.8522), and area under the precision-recall curve of 0.560 (95% confidence interval: 0.5205–0.6058) for a 30-day prediction interval after LT. The team concluded that this model will help clinicians to identify high-risk candidates for further risk stratification or other management strategies, to improve LT outcomes. ●

"Performance of the deep learning model against other machine learning models was assessed using a cohort of 18,304 LT recipients."





Prognostic Value of Hyperaemic Stenosis Resistance

THE Hyperaemic Stenosis Resistance (HSR) index was shown to have theoretical advantages for the detection of ischaemia-inducing high-risk coronary lesions, according to data presented at EuroPCR 2023, held in Paris, France, between 16th–19th May 2023. The HSR index represents the ratio between the hyperaemic pressure drop across a stenosis, as well as the flow through it. While it aims to indicate stenosis severity to guide revascularisation, its diagnostic and prognostic value is unclear, which is why researchers evaluated this in patients with stable chronic coronary syndromes. This is the first study to determine the HSR's prognostic value, and its potential to identify which lesions would benefit from revascularisation accurately.

A total of 853 patients and 1,107 vessels were included in the study, derived from the Inclusive Invasive Physiological Assessment in Angina Syndromes (ILIAS) Registry. Results showed a higher area under the curve (0.71; 95% confidence interval [CI]: 0.66–0.75) for the presence of ischaemia on non-invasive stress test for the HSR, compared to coronary flow velocity reserve (0.63) and fractional flow

reserve (0.66). They further identified the optimal cut-off value for HSR to be 0.80 mm Hg/cm per second, which is similar to previous study results.

"Researchers evaluated this in patients with stable chronic coronary syndrome."

Further survival analysis showed statistical association between HSR as a continuous variable and target vessel failure at 5-year follow-up (hazard ratio: 1.40; 95% CI: 1.06–1.84; $p=0.016$), as well as an independent association between abnormal HSR based on the cut-off value of 0.80, and 5-year target vessel failure (hazard ratio: 2.48; 95% CI: 1.49–4.13; $p<0.0005$). The team further showed through sensitivity analysis that an HSR-first approach identified increased risk of adverse clinical outcomes more adequately in lesions initially deferred from revascularisation, compared to coronary flow velocity reserve and fractional flow reserve. This led the team to conclude that their study affirmed the theoretical advantages of HSR. ●

Electronic Creatinine Alert System: A First Step Towards Preventing Hospital-Acquired Acute Kidney Injury

ACUTE Kidney Injury (AKI) is a common complication in critically ill and non-critically ill patients, for which there is currently no specific treatment. Early recognition is crucial to prevent progression to advanced stages and the need for kidney replacement therapies. Electric alert systems are emerging as a tool capable of alerting clinicians to potentially harmful situations. This retrospective study aimed to analyse the incidence of AKI in a tertiary hospital using an Electronic Creatinine Alert System (ECAS), which alerts clinicians to serum creatinine changes in real time, enabling early interventions. Results were presented at the European Renal Association (ERA) Congress held in Milan, Italy between 15th–18th June 2023.

The retrospective study analysed 46,149 patients discharged from a tertiary referral hospital between 1st January 2019–31st December 2021. The exclusion criteria were discharges from critical care units, patients admitted to the emergency room, patients with AKI criteria on admission, patients admitted to the nephrology department, and patients with Stage G5 chronic kidney disease or on kidney replacement therapy. The ECAS alerted patients following an increase of ≥ 0.3 mg/dL of serum creatinine or an elevation of ≥ 1.5 times the baseline creatinine value.

A total of 69,0002 discharges were analysed, with the majority of patients included having AKI

Stage 1 (69.7%), followed by Stage 2 (21.3%), and Stage 3 (9.0%). The median age of the included participants was 75 years, and 62% were male. The ECAS was activated in 13.5% of discharges, with an increasing trend over the years (11.8% in 2019, 12.4% in 2020, and 12.1% in 2021). Geriatrics (14.2%), cardiology (11.9%), general surgery (9.9%), infectious diseases (9.2%), and cardiac surgery (7.2%) were the departments with the highest ECAS activations. Results indicated that the ECAS activations were associated with poor outcomes. Patients who activated the ECAS experienced significantly longer hospital stays compared to those who did not (6 days versus 13 days; $p < 0.001$). Moreover, the survival distributions for the ECAS activation were statistically different ($X^2 = 5.522$; $p = 0.019$) and kidney recovery at discharge was significantly lower in the patients with AKI 2 (18.5%) and AKI 3 (8.5%), compared to patients with AKI 1 (73%; $p < 0.001$).

Overall, the results suggest ECAS as a suitable electronic alert system for rapid AKI identification. The findings prompt the adoption of a nephrology rapid response team for early AKI detection before creatinine elevation, incorporating methods such as point-of-care ultrasonography and acute kidney stress biomarkers to enhance early intervention strategies. ●

"Electric alert systems are emerging as a tool capable of alerting clinicians to potentially harmful situations."



Use of Artificial Intelligence in Predicting Rheumatoid Arthritis

USING extremity MRI to predict early rheumatoid arthritis (RA) can lead to more timely treatment, and the possible prevention of chronicity. Recent research presented at the European Alliance of Associations for Rheumatology (EULAR) Congress 2023, held in Milan, Italy, suggested that using artificial intelligence (AI) to interpret images may prove more accurate than visual screening. A team from Leiden University Medical Centre, the Netherlands, developed a deep learning AI method that automatically analyses extremity MRI scans in order to predict RA at an early stage.

The research involved MRI scans of the hands and feet of a total of 1,974 patients, of whom 1,247 had early onset arthritis (EAC) and 727 had clinically suspect arthralgia (CSA). Of the EAC group, 538 patients developed RA in 2 years, and 113 of the patients with CSA also developed RA. MRI scans were pre-processed automatically, and a self-supervised deep learning model was pre-trained to fill in blacked-out parts of the images. The model was then fine-tuned to predict RA development, its accuracy evaluated under the receiver operator curve (AUC).

Results showed that deep learning models have the ability to predict RA development accurately. On the test set, the model obtained a mean AUC of 0.683 in the EAC group, and 0.727 in the CSA group, using MRI scans of the hands (wrist and metacarpophalangeal joints). Models trained separately on the wrists and feet received a mean AUC of 0.679, 0.647, 0.664; and 0.688, 0.669, and 0.715, for the EAC and CSA group, respectively. These accuracies were close to the expert-level using Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS), with reported AUCs of 0.74 and 0.69 in predicting RA in CSA.

The team concluded that automatic RA prediction using an AI interpretation of MRI scans is possible, as this new visualisation method confirms the significance of inflammatory features, but also has the potential to highlight new imaging biomarkers. Researchers added that including the MRI data of healthy controls, as is done in RAMRIS-based prediction, may improve the ability of AI to predict RA development even further. ●

"Including the MRI data of healthy controls may improve the ability of AI to predict RA development even further."

Telemedicine Effective for Multidisciplinary Narcolepsy Care

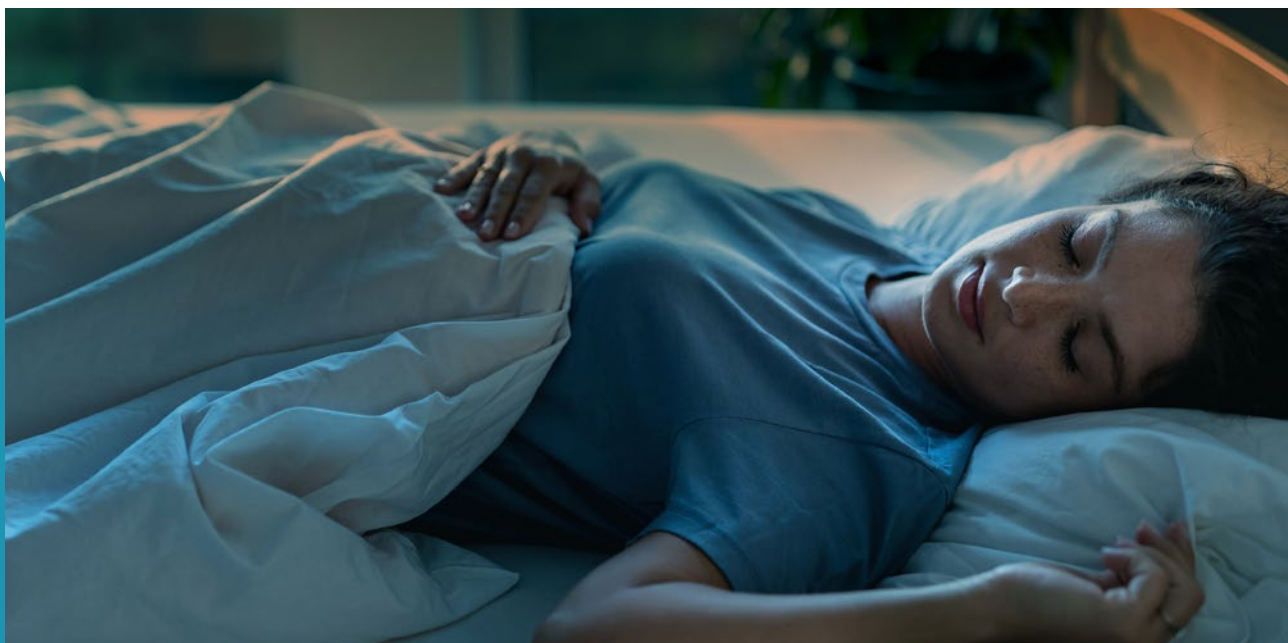
RESULTS from the TENAR randomised controlled trial presented at the European Academy of Neurology (EAN) Congress in Budapest, Hungary, have provided insights into the use of telemedicine for the care of people with narcolepsy. Narcolepsy, a rare central hypersomnia associated with endocrine and psychosocial problems, poses unique challenges that require a comprehensive, multidisciplinary approach. The TENAR trial, conducted by researchers from IRCCS Institute of Neurological Sciences of Bologna, Italy, and the University of Bologna, Italy, sought to assess the efficacy of telemedicine in managing narcolepsy compared to traditional in-office visits.

The trial included individuals with narcolepsy aged over 14 years, and compared a multidisciplinary care approach, encompassing neurological, endocrinological, and psychosocial care, delivered through telemedicine versus in-office visits over a 1-year period. The primary outcome was the control of sleepiness, measured by the Epworth Sleepiness Scale (ESS) at 12 months, with a non-inferiority margin of 1.5 points. Secondary outcomes included: control of other symptoms, treatment compliance, metabolic control, quality of life, feasibility, patient and family satisfaction, safety, and disease-related costs.

Of the 208 individuals selected, 202 successfully completed the study at 12 months. The baseline variables were well balanced between the telemedicine and in-office outcomes. The primary outcome, ESS score improvement, showed a 1.3 mean point improvement in both groups, and there was no statistical difference between the groups. Secondary outcomes, including BMI improvement, demonstrated similar positive trends in both telemedicine and in-office management, with no statistical difference between the groups.

"Of the 208 individuals selected, 202 successfully completed the study at 12 months."

The preliminary results of the TENAR trial indicate the viability in terms of feasibility, effectiveness, and safety of multidisciplinary care procedures for narcolepsy. These findings highlight the possibility of redefining the approach to narcolepsy, and the potential of telemedicine to enhance accessibility, improve patient outcomes, and reduce associated costs. ●



ChatGPT: A Potential New Tool to Answer Patient Questions About Fertility



CHATGPT (OpenAI, San Francisco, California, USA) may be a useful tool for patients seeking factual and unbiased information regarding fertility and fertility treatment, according to new research presented at the 39th Annual Meeting of the European Society of Human Reproduction and Embryology (ESHRE), held in Copenhagen, Denmark.

"Out of the ten answers, six scored 5 or more, and three received a score of 3–4."

ChatGPT is a language model that uses deep learning to generate human-like text. Researchers from Monash University, Melbourne, Australia, examined the quality of information provided by ChatGPT, using ten common patient questions as prompts. Three questions related to fertility awareness (impact of female/male age on fertility and fertile window in the menstrual cycle), one to the chance of success with *in vitro* fertilisation, one to elective egg freezing, one to the benefits of add-ons, one to polycystic ovary syndrome and pregnancy, one to choosing a fertility clinic, and one to how many *in vitro* fertilisation cycles should be attempted.

The two authors scored the quality of the information generated by ChatGPT using a scoring matrix with a range of 0–7, where higher

scores indicate higher quality. Text was rated against humanistic answers based on how well it corresponded (0–3), evidence of commercial bias or controversial claims (no=0, yes=1), use of accurate statistics, and whether it was stated that medical advice should be sought (no=0, yes=1).

The scores returned by the experts were closely aligned, with only one point difference for one of the answers. Out of the ten answers, six scored 5 or more, and three received a score of 3–4. Only one answer, the answer to the question about the benefits of add-ons, scored less than 3. This was also the only question where the response had evidence of commercial bias, and one of only two that made claims that could be considered controversial. While the scoring method used in this study is exploratory in nature, the use of expert evaluation can be used to fine-tune the parameters of machine learning models, and improve their performance.

People seeking fertility-related information rely heavily on online sources, such as clinic websites, consumer advocacy organisations, patient support groups, and social media. Overall, the study concluded that the quality of information generated by ChatGPT was high, with little evidence of commercial bias, suggesting that ChatGPT may be a useful tool to answer patient questions about fertility. ●

Harnessing Generative Artificial Intelligence Data for Skin Disease Classification

RESEARCHERS have successfully employed generative artificial intelligence (AI) techniques using probabilistic models (DPM) in an effort to address the shortage of labelled training data for real-world dermatology applications. Results of this research were presented at the European Academy of Dermatology and Venereology (EADV) Congress in Berlin, Germany, and indicate the potential of this approach in augmenting training datasets and improving skin disease classification. Thus far, research efforts in dermatology have struggled with limited access to training data, due to privacy constraints and strict data sharing policies. Researchers propose using DPMs for image augmentation within supervised machine learning pipelines, with the objective to complement existing validation datasets and overcome challenges posed by the lack of real-world dermatological data.

"High-quality images were isolated through the development of a data curation pipeline."

The study fine-tuned DPMs on six different disease conditions, including basal cell carcinoma and melanoma as malignant classes; actinic keratosis and atypical melanocytic nevus as pre-malignant classes; and lentigo and seborrheic keratosis as benign classes. High-quality images were isolated through the

development of a data curation pipeline in order to address the variation in generated image quality.

The results obtained indicated that the generative data augmentation approach maintains a comparable classification accuracy to visual classifiers, even when trained on fully synthetic skin disease datasets. Inclusion of synthetic images in a hybrid dataset, alongside the original images, improved performance, and yielded a top-3 accuracy of 85.01%. This outperformed datasets comprising of only original images (84.48%), or only synthetic images (84.09%), indicating that DPMs are effective in generating high-quality images and improving augmentation of datasets, without compromising classifier performance.

The findings of the study showcase the generative capabilities of DPMs in creating macroscopic skin disease images. By conditioning probabilistic diffusion-based generation on text prompt inputs, fine-grained synthetic images were possible. Researchers also propose a closed-loop data augmentation pipeline to automatically generate images whilst complementing real-world skin disease datasets. This research emphasises the reliability of synthetic images as data sources for skin disease classification, offering a potential avenue for medical applications and data sharing without compromising confidentiality. ●

