Objectives: We aimed to evaluate the feasibility of ML algorithm generation using a no-coding platform by clinicians without ML experience. By using such a platform, we developed a model for automated recognition and grading of radiographic distal interphalangeal osteoarthritis (DIP-OA) with subsequent user experience (UX) testing of the algorithm.

Methods: 13690 hands x-rays from 2863 patients within the Swiss Cohort of Quality Management (SCQM) and an external control dataset of 346 non-SCQM patients were collected and scored for DIP-OA by the modified Kellgren and Lawrence Score (KL). 46892 DIP joints were extracted and classified according to the KL score. GIOITTO (L2F, Lausanne) was used as a no-coding platform for training of 2 convolutional neural networks, the first one for segmentation of hand joints and subsequent DIP-joints extraction, and the second one for classification of KL scores according to the presence of osteophytes and joint space narrowing on the previously extracted DIP-joints. The classification model’s performance was tested by an internal test set (SCQM database) and an external test set (non-SCQM). User experience (UX) of a web app developed from the platform as a provisory user interface (UI) was investigated in rheumatologists and radiologists. The usability of heat maps was also investigated.

Results: The sensitivity and specificity of this model for detecting DIP-OA, was 79% and 86%, respectively. The accuracy for grading the KL score was 76% with a kappa score of 0.76. A similar sensitivity (79%) and specificity (80%) for detecting DIP-OA was found in an independent external test set. The accuracy per DIP-OA class differed with 79% for no OA (defined as KL 0 and 1), 65% for KL2, 40% for KL3, and 70% for KL4. On the external test set overall accuracy was 66% and the kappa score 0.75. The platform was intuitive and easy to use, but support from data scientists was still required for upload of the data set exceeding the drag and drop process. DIP-joints extraction, heatmap generation and episodic debugging of the platform. UX testing of the web app revealed a moderate demand by rheumatologists (notably for differential diagnosis and evolution over time) and a low demand by radiologists for the automated scoring of DIP-OA. Conversely, adding heat maps to the UI was considered of high added value.

Conclusion: No-coding platforms are an opportunity to develop end-to-end AI-prototypes for further testing. Their use by clinicians without experience in programming and machine learning is possible. Here, automated radiographic DIP-OA detection is both feasible and usable, whereas grading between individual KL scores, e.g. for clinical trials, remains challenging, partly because of the inter-observer variability inherent to the scoring method.

REFERENCES:

Figure 1. Making of an algorithm for detection and grading of distal hand osteoarthritis using a no-coding platform: end-to-end process.