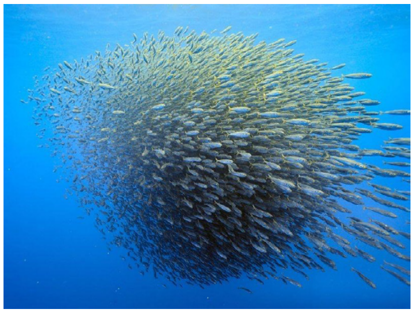
Case Study: Radiologists Diagnose Knee Lesions More Accurately using Swarm AI

Customer: University of California, San Francisco



What is Swarm AI?

Swarm AI® technology provides a powerful combination of real-time human insights and AI algorithms, enabling significantly more accurate results than traditional methods.

Modeled after swarms in nature, which converge on optimal solutions with extreme efficiency, Swarm AI turns networked human groups into AI-optimized systems, enabling accurate insights, decisions, and forecasts to be generated quickly.

Applying the power of Swarm AI is easy using the Swarm platform. It can be accessed from anywhere in the world using standard web browsers and can generate actionable results in a matter of minutes.

For more information, visit the following links:

Unanimous AI
https://unanimous.ai/what-is-si/

Swarm technology
https://unanimous.ai/swarm/

Introduction

In the field of diagnostic radiology, the interpretation of complex medical images such as MRIs can be challenging. There is often a high degree of difference in opinions among experts, even among those with the same level of expertise and experience. For example, high inter-reader disagreement can be seen amongst experienced radiologists reading knee MRIs to detect lesions in the meniscus. This leads to an obvious concern: how can medical professionals give accurate diagnoses if they can't agree amongst themselves? This pervasive problem of subjective diagnostic opinion cuts across Medicine, from mammogram readings in radiology to biopsy diagnosis in pathology.

A new paper (currently in pre-print) written by Rutwik Shah M.D., and colleagues at the Center for Intelligent Imaging at University of California, San Francisco (UCSF), examined this exact problem head-on. They reported significant improvements in diagnostic accuracy among teams of radiologists using Swarm AI® technology. Unlike other forms of AI that aim to replace human practitioners with automated algorithms, Swarm AI is a unique technique for amplifying the collective accuracy of human groups. Modeled on the biological principle of Swarm Intelligence, Swarm AI technology enables networked groups to work together to reach decisions, each connecting from their own workstation while intelligent algorithms monitor their real-time interactions to determine their relative levels of confidence and conviction. This creates a closed-loop system where small groups rapidly deliberate and converge on solutions together.

Specifically, the new paper from UCSF School of Medicine uses the Swarm® software platform from Unanimous AI. As shown in Figure 1, the platform enables small groups to answer questions by collaboratively manipulating a graphical puck to select from among a set of options. Each participant provides input by moving a graphical magnet with a mouse, touchpad, or touchscreen. By positioning their magnet with respect to the puck, participants apply their will on the system. The input from each user is not a discrete vote, but a stream of vectors that varies freely over time. Because all members adjust their intent continuously, the group reaches a solution, not based on the input of any single individual, but based on the dynamics of the system as moderated by the swarming algorithms.

The overall quality of the diagnoses was up to 32% better when using Swarm as compared to the Plurality vote, Most Confident vote, and the AI model. Moreover, the number of correctly identified normal cases (true negative) improved significantly, as opposed to individual readers who were all biased towards over-diagnosing lesions.

Methodology

In this study, two cohorts—one composed of three board certified “attending” radiologists and another of five less-experienced radiology “residents”—diagnosed 36 knee MRI exams for meniscus lesions in one of three classes: 1) normal, 2) focal lesion, 3) widespread lesion. The scans were read at two timepoints, first by private individual survey and then, after a gap of two weeks, collaboratively in real-time on the Swarm platform. Using the survey data, two aggregation methods were tested – a Plurality Voting method and a Confidence Voting method. The plurality method identified the result which received the most votes among the three options. The confidence method used questions from the survey in which the radiologists self-reported their confidence in their answers on a scale of 0% to 100%. The two survey-based diagnoses and the interactive Swarm-based diagnosis were compared to the clinical ground truth for each case as noted by an orthopedic surgeon who visualized the actual meniscus during a procedure called arthroscopy. Finally, the performance of UCSF’s state-of-the-art diagnostic AI model trained on the same task, was evaluated on this same dataset.

The comparisons with ground truth were then tallied as 1) Correctly identified cases, 2) Slightly incorrect (misdiagnosing one class above or below), and 3) Highly incorrect (misdiagnosing normal as widespread lesion or vice versa). For both teams (the group of three experienced attending radiologists and the group of five less-experienced residents) **the diagnoses made using Swarm outperformed the plurality vote and the most confident diagnosis as collected from the individual survey diagnoses. The number of cases that were highly incorrect was halved in both the Resident and Attending teams when using Swarm to diagnose cases. The AI model performed at the same level as the most confident experienced attending radiologist on each question and underperformed the same radiologists using the Swarm platform**.

Figure 1: Radiologist Workstation, with Swarm in left monitor and MRI knee scans in right monitors.

"A key feature of the Swarm platform was in allowing anonymized collaborations between participants. This enabled our radiologists to express their true opinions without peer pressure, which is often not the case in collaborative diagnostic settings."

-Rutwik Shah M.D.
lead author

	Cases Correct	Cases Slightly Incorrect	Cases Highly Incorrect
Plurality Vote	31% (11/36)	**42%** (15/36)	28% (10/36)
Most Confident Vote	36% (13/36)	44% (16/36)	19% (7/36)
AI	36% (13/36)	47% (17/36)	17% (6/36)
Swarm	**47%** (17/36)	50% (18/36)	**3%** (1/36)

Table 1: Attending Radiologist Team Performance

	Cases Correct	Cases Slightly Incorrect	Cases Highly Incorrect
Plurality Vote	26% (9/36)	49% (17/36)	26% (7/36)
Most Confident Vote	37% (13/36)	37% (13/36)	26% (9/36)
AI model	36% (13/36)	47% (17/36)	17% (6/36)
Swarm	**57%** (20/35)	**31%** (11/35)	**11%** (4/35)

Table 2: Resident Radiologist Team Performance

Conclusion

As a result, the UCSF team concluded that Swarm AI is a highly promising tool for generating better diagnostic labels for knee MRIs and other scenarios with high inter-reader disagreement, such as mammograms. The team also noted that Swarm has potential as a tool to generate labels for training AI models, which are currently limited by the availability of high-quality training data and the low accuracy of single-human labels. This result aligns with a study at Stanford University School of Medicine in which the Swarm software platform was tested on small groups of radiologists tasked with diagnosing chest x-rays for the presence of pneumonia, resulting in over 30% reduction in diagnostic errors.