Meaningless Applications and Misguided Methodologies in Artificial Intelligence—Related Orthopaedic Research Propagates Hype Over Hope

Prem N. Ramkumar, M.D., M.B.A., Michael Pang, B.S., Teja Polisetty, B.S., J. Matthew Helm, M.D., and Jaret M. Karnuta, M.D., M.S.

Abstract: There exists great hope and hype in the literature surrounding applications of artificial intelligence (AI) to orthopaedic surgery. Between 2018 and 2021, a total of 178 AI-related articles were published in orthopaedics. However, for every 2 original research papers that apply AI to orthopaedics, a commentary or review is published (30.3%). AI-related research in orthopaedics frequently fails to provide use cases that offer the uninitiated an opportunity to appraise the importance of AI by studying meaningful questions, evaluating unknown hypotheses, or analyzing quality data. The hype perpetuates a feed-forward cycle that relegates AI to a meaningless buzzword by rewarding those with nascent understanding and rudimentary technical knowledge into committing several basic errors: (1) inappropriately conflating vernacular (“AI/machine learning”), (2) repackaging registry data, (3) prematurely releasing internally validated algorithms, (4) overstating the “black box phenomenon” by failing to provide weighted analysis, (5) claiming to evaluate AI rather than the data itself, and (6) withholding full model architecture code. Relevant AI-specific guidelines are forthcoming, but forced application of the original Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines designed for regression analyses is irrelevant and misleading. To safeguard meaningful use, AI-related research efforts in orthopaedics should be (1) directed toward administrative support over clinical evaluation and management, (2) require the use of the advanced model, and (3) answer a question that was previously unknown, unanswered, or unquantifiable.

In the last 4 years, the intersection of artificial intelligence (AI) and orthopaedic surgery has been explored and extensively discussed in the peer-reviewed literature. Although the quality of the literature presently favors narrative and systematic reviews over meaningful original research, there is continued interest in publishing on the subject of “artificial intelligence and machine learning.” Rather than providing use cases that offer the uninitiated an opportunity to assess and appreciate the importance and impact of AI itself, a large portion of present-day research instead reduces the technology with low-quality data. In turn, this perpetuates a feed-forward cycle that relegates the topic to a meaningless buzzword and cultivates a false understanding of its potential. Worse, the hype props those with nascent understanding and rudimentary technical knowledge into developing and releasing internally validated algorithms that repack package extant registry data without addressing meaningful problems or answering new questions. Now that AI awareness has been sufficiently established in the orthopaedic literature as a prospective predictive modeling tool, relevant use case application with thoughtful methodologic execution is critical. The original Transparent
Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines do not apply to AI research. Rather than propagating the misconception that orthopaedic applications surrounding AI should be directed toward clinical evaluation and management, attention should instead be redirected to improving the experience of the orthopaedic patient and surgeon to fully realize the expectations of AI.

The Good

Although the related technical aspects may be beyond the scope of the orthopaedic surgeon, AI is not difficult to appreciate or conceptualize, as we have lived in a world that has surreptitiously applied it to our existence for years. From targeted suggestions that offer us our next media segment to consume or the very diction that finishes our thought before we’re even done typing, AI has transformed numerous industries through newfound efficiencies and supportive decision-making. In brief, AI is computer automation. By building complex associations that are reinforced with positive and negative feedback, algorithms can be built to yield predictions that can iteratively improve with newly introduced data. Unlike regression models that statically analyze finite relationships with relatively homogeneous data formats, AI is capable of dynamically analyzing disorganized data and prospectively predicting outcomes.

With the exponential growth of computing power and subsequent decrease in cost to run these automated processes, AI remains ripe to transform health care and, specifically, orthopaedic surgery along all phases of the care continuum as we continue to aggregate more data than what a single individual or team of individuals can process. Recent applications and research efforts implementing AI in the field of orthopaedics have demonstrated great promise in predicting future injury risk, interpreting advanced imaging, identifying implants, projecting patient-reported outcomes, reporting value-based metrics, and augmenting tele-health. Overall, AI remains an exciting prospective tool that applies data in real time to automate tasks that have the potential to fundamentally improve patient care, unburden administrative work, and translate once-qualitative insights into tangible quantitative metrics.

The fundamental question surrounding the emphasis on AI in orthopaedics is “why should we care?” It was not long ago when surgeons felt the business of medicine was outside the scope of our expertise. Today, we find it vital to our survival and increasingly essential in educating trainees. How we understand, leverage, and interact with technology will similarly be critical to our day-to-day workflow. AI has the potential to dictate how we perform clinical activities, communicate in medicolegal terms, and allocate our time for optimal resource use. Consider the clinic encounter wherein documentation of multiple elements within the electronic health record (EHR) are necessary to appropriately bill. Oftentimes, we are faced to choose between documenting during the face-to-face encounter, limit our time with the patient, or run late. All options detract from the experience of both the surgeon and the patient. Through natural language processing, a subset of AI, autocompletion of documentation and clicks for standard encounters can rapidly save time. In one report by Kim et al., billing codes were able to be automatically generated from operative note dictations. Today, administrative burden associated with the clinical encounter on EHRs directly contributes to burnout. For every 15-minute patient encounter, physicians spend over 16 minutes just on the EHR and 11% of that time documenting spills into the “after-hours” (weekends or outside of clinic business hours).

The importance of adopting AI-based techniques is misunderstood. Contrary to popular belief, AI is not intended to—nor should it ever be suggested to—replace routine patient-facing clinical activities of the orthopaedic surgeon. Overall, as orthopaedic surgeons we do a great job of taking care of patients. The actual “doctor” activities of diagnosis and management do not require reinvention. Instead, AI is better served redressing the areas that have contributed to a U.S. health care system many consider to be broken, from patient access and caregiver wellness to financial transparency. Numerous opportunities exist for the orthopaedic surgeon to leverage what many consider to be the fourth industrial revolution. In addition to personalizing medicine and decreasing administrative burnout, AI offers the ability to navigate the present pandemic by enabling remote patient monitoring for the orthopaedic patient. Through pre-existing sensors on smartphones and wearable devices, physical examination metrics like knee range of motion and gait can be remotely ascertained in the perioperative period. Remote patient monitoring offers unique convenience to both the surgeon and the patient while ensuring patient safety. When considering financial transparency and value-based care in orthopaedics, patient-specific payment models have been proposed that stratify patients by their preoperative comorbidities to normalize surgeon reimbursement to the preoperative surgical risk. During a time where reimbursements continue to be devalued for procedural work, AI can be preoperatively leveraged to predict the risk profile of a patient before the incision to better arbitrate reimbursement fees. With more than 16 hours per week spent on haggling with prior authorization approval, this area remains ripe for automation. The use cases for AI in improving the experience of patients, orthopaedic surgeons, and their teams through
automation are limitless and cause for continued hope in this domain of research.

The Bad

Despite the promise of AI in orthopaedics, the exponential growth in this realm of research has created several dilemmas that warrant acknowledgement and redressal. While the early acceptance from the orthopaedic community to learn a foreign concept was initially inspiring, AI has now been couched as a colossal buzzword that conveys emptiness and invites skepticism. Beginning with terminology, several articles refer to the topic as “AI and machine learning (ML)” when this is both redundant and unclear. Conceptually, AI is the umbrella term that enables a machine to simulate human behavior; ML is a narrower term and small subset of AI that performs specific tasks to yield predictions using pattern recognition. AI systems can apply ML, but ML cannot apply AI. When the nomenclature is described as a pair, this undermines the authority of the presented content and creates confusion.

A large portion of the orthopaedic research concerning AI is portrayed and personified as if the readership is evaluating the merits as its judge, jury, and executioner. This is evident with literature entitled “Machine learning algorithms predict...” or “AI predicts...”. This portrayal is closer to marketing than science and underscores a fundamentally misunderstood tenet: AI is only as good as the data that fuel its process. The bottleneck of AI-derived insights is not the analytic process, but rather the quality of the data. As with any peer-reviewed publication containing statistical analysis, the data drive the quality and generalizability of the conclusions. In most cases wherein authors claim to be evaluating AI-based processes, most are in fact simply evaluating the reflection of their data. When certain reports refute the ability of an AI-based process to supersede the predictive power of traditional regression, it occurs when the data are low volume or low quality. If a Ferrari and Prius were asked to race in rush hour traffic on the 101 freeway in Los Angeles, the horsepower of the Ferrari doesn’t help much, and one could conclude the Prius is just as fast as a Ferrari. While this is farcical, this is the exact phenomenon that is occurring when we attempt to compare deep learning models with archaic regression analyses in the setting of poor data.

Not only are these studies failing to actually evaluate AI, an equivalent corollary would be to rename these articles “ANOVA shows the obese and morbidly obese have worse outcomes after knee replacement” or “Student t test demonstrates smokers are more likely to develop nonunion after fracture fixation.” Just as we are not evaluating t tests, we are not evaluating AI. Instead, we are using AI as a tool to evaluate the data and its conclusions as with any other research report. Taken one step further, some authors even attempt to caution or apply unique ethical principles to AI, as if it warrants a new playbook. The ethics related to AI do not differ from the ethics of any other statistical process, which center on the data. The same principles that guide data storage, management, and analysis apply regardless of the statistical modeling process.

With respect to quality control for original research applying AI-based algorithms, there exist three points of emphasis with respect to the methods, results, and discussion. While guidelines surrounding AI-related research are necessary, they should not overshadow the importance of robust data and clear conclusions. The future of AI in orthopaedics lies in its ability to provide meaningful insights into patient care and outcomes, not as a marketing tool.

Fig 1. Year-over-year growth of artificial intelligence research in orthopaedics from 2018 to 2021 with respect to original and nonoriginal research.
research are not yet available and will be discussed in depth in the subsequent section, authors in should always provide the full code (i.e., GitHub URL) to maximize generalizability and transparency. Additionally, the “black box” of this realm of research exists but is presently overstated. With most analyses, there exists a way to cross-sectionally evaluate how the algorithm is performing with SHAP (SHapley Additive exPlanations) analysis\(^1\)\(^,\)\(^2\) for individual predictions or heatmaps\(^8\)\(^,\)\(^9\) for imaging feature identification. In addition, if the study was not evaluated with multiple populations, then the work needs external validation before releasing a makeshift online prediction tool (i.e., https://www.shinyapps.io/) available to the general public. Without external validation, the tool could confer premature generalizability and generate false predictions that compromise patient safety or misguide communication.

**The Ugly**

The potential surrounding AI has in turn left both journals and authors susceptible to abuse as we stand atop its hype cycle. AI confers the distinct advantage of automation to provide insights that were previously unknown, unanswered, or unquantifiable. When evaluating a study that boasts use of AI or ML, ask (1) if the question has ever been answered before and (2) if AI was necessary to answer the question. Although “publish or perish” remains a sad reality, the onus is as much on the authors as it is on the journals to avoid duping editorial boards and readers unfamiliar with AI into believing a study is unique simply because a trendy statistical model was applied or a popular buzzword increased pageviews. For those familiar with ML techniques, it is very easy to repackage previous studies using extant registry data—especially using unoriginal or previously answered questions—and generate a clinical prediction tool. Unless that tool undergoes external validation with multiple population data sources and is prospectively applied for evaluation, the study has done nothing more than to continue hollowing meaningful applications of AI and fortify its stature as a buzzword. Rather than indulging in the gratuitous application of AI, the most meaningful use for this technology is to apply it to clinical activities that are nondiagnostic and treatment-related. Physicians have spent over a decade trained in the art, science, and humanity of medical diagnosis and management; directing efforts to unburdening the clinical workforce from clinical medicine would be a misappropriation of resources. Instead, AI should be directed toward optimizing the health care experience by improving access, decreasing the documentation burden, and optimizing value-based communication. One way to optimize value-based communication is to take advantage of the evidence-based aspects of predictive AI to arbitrate risk-adjusted reimbursement models before elective orthopaedic surgery among surgeons, payers, and hospitals.\(^19\)\(^,\)\(^33\)

Along the same vein of perpetuating hype over substance with regards to AI research in orthopaedics, Figure 1 illustrates the increased interest to publish on the topic. From 2018 to 2021, a total of 178 AI-related articles were published in orthopaedics; of these 178, 54 were narratives or review articles on the subject (30.3%). For every 2 original research papers that develop and apply AI techniques to orthopaedic applications, a commentary or update is published. A recent trend among reviews entails the forced application of the TRIPOD guidelines to grade AI research, which serves as an emblematic microcosm for the most pressing issue in this realm of research.\(^34\) Despite the fact that these guidelines were designed only for traditional multivariate models and specific AI guidelines are imminent,\(^35\) editorial boards and authors either lack the familiarity or interest to question the coarse translation of antiquated guidelines in favor of publishing yet another hollow review article.\(^21\)\(^,\)\(^36\) To many researchers and journals, the orthopaedic literature on AI should focus on meaningful original research instead of misguided content intended to generate clicks.\(^37\)

**Conclusions**

Although AI has already demonstrated great promise to transform orthopaedics, research efforts should be directed toward optimizing the health care experience by improving access, decreasing the documentation burden, and optimizing value-based communication instead of clinical diagnostic and treatment support tools. When evaluating the quality of AI research in orthopaedics, nomenclature matters, and most studies are evaluating the data rather than the actual analytic technique. Fundamental principles of sharing the full code instead of prematurely releasing publicly available and internally validated prediction tools will be more useful for researchers. The hype surrounding AI in orthopaedic research has generated a review article for every 2 original reports; forced application of the original TRIPOD guidelines are not applicable to AI research and relevant guidelines are forthcoming. To ensure quality AI research in orthopaedics, the question must be previously unknown, unanswered, or unquantifiable and the advanced model must be necessary.

**References**


2. Makhni EC, Makhni S, Ramkumar PN. Artificial intelligence for the orthopaedic surgeon: An overview of...


