



Automated operative workflow analysis of endoscopic pituitary surgery using machine learning: development and preclinical evaluation (IDEAL stage 0)

*Danyal Z. Khan, MRCS,^{1,2} Imanol Luengo, PhD,³ Santiago Barbarisi, MSc,³ Carole Addis, PhD,³ Lucy Culshaw, PhD,³ Neil L. Dorward, FRCS,¹ Pinja Haikka, PhD,³ Abhiney Jain, MRCS,¹ Karen Kerr, PhD,³ Chan Hee Koh, MBChB,^{1,2} Hugo Layard Horsfall, MBBS,^{1,2} William Muirhead, FRCS,^{1,2} Paolo Palmisciano, MD,⁴ Baptiste Vasey, MMed,⁵ Danail Stoyanov, PhD,^{2,3} and Hani J. Marcus, PhD, FRCS^{1,2}

¹Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; ²Wellcome/EPSCRC Centre for Interventional and Surgical Sciences, University College London; ³Digital Surgery Ltd., Medtronic, London, United Kingdom; ⁴Department of Neurosurgery, Trauma Center, Gamma Knife Center, Cannizzaro Hospital, Catania, Italy; and ⁵Nuffield Department of Surgical Sciences, University of Oxford, United Kingdom

OBJECTIVE Surgical workflow analysis involves systematically breaking down operations into key phases and steps. Automatic analysis of this workflow has potential uses for surgical training, preoperative planning, and outcome prediction. Recent advances in machine learning (ML) and computer vision have allowed accurate automated workflow analysis of operative videos. In this Idea, Development, Exploration, Assessment, Long-term study (IDEAL) stage 0 study, the authors sought to use Touch Surgery for the development and validation of an ML-powered analysis of phases and steps in the endoscopic transsphenoidal approach (eTSA) for pituitary adenoma resection, a first for neurosurgery.

METHODS The surgical phases and steps of 50 anonymized eTSA operative videos were labeled by expert surgeons. Forty videos were used to train a combined convolutional and recurrent neural network model by Touch Surgery. Ten videos were used for model evaluation (accuracy, F1 score), comparing the phase and step recognition of surgeons to the automatic detection of the ML model.

RESULTS The longest phase was the sellar phase (median 28 minutes), followed by the nasal phase (median 22 minutes) and the closure phase (median 14 minutes). The longest steps were step 5 (tumor identification and excision, median 17 minutes); step 3 (posterior septectomy and removal of sphenoid septations, median 14 minutes); and step 4 (anterior sellar wall removal, median 10 minutes). There were substantial variations within the recorded procedures in terms of video appearances, step duration, and step order, with only 50% of videos containing all 7 steps performed sequentially in numerical order. Despite this, the model was able to output accurate recognition of surgical phases (91% accuracy, 90% F1 score) and steps (76% accuracy, 75% F1 score).

CONCLUSIONS In this IDEAL stage 0 study, ML techniques have been developed to automatically analyze operative videos of eTSA pituitary surgery. This technology has previously been shown to be acceptable to neurosurgical teams and patients. ML-based surgical workflow analysis has numerous potential uses—such as education (e.g., automatic indexing of contemporary operative videos for teaching), improved operative efficiency (e.g., orchestrating the entire surgical team to a common workflow), and improved patient outcomes (e.g., comparison of surgical techniques or early detection of adverse events). Future directions include the real-time integration of Touch Surgery into the live operative environment as an IDEAL stage 1 (first-in-human) study, and further development of underpinning ML models using larger data sets.

<https://thejns.org/doi/abs/10.3171/2021.6.JNS21923>

KEYWORDS surgical workflow; machine learning; neural networks; artificial intelligence; computer vision; endoscopic transsphenoidal surgery; pituitary adenoma; pituitary surgery

ABBREVIATIONS AI = artificial intelligence; CNN = convolutional neural network; DNN = deep neural network; eTSA = endoscopic transsphenoidal approach; IDEAL = Idea, Development, Exploration, Assessment, Long-term study; ML = machine learning; RNN = recurrent neural network.

SUBMITTED April 8, 2021. **ACCEPTED** June 15, 2021.

INCLUDE WHEN CITING Published online November 5, 2021; DOI: 10.3171/2021.6.JNS21923.

* D.Z.K. and I.L. share first authorship of this work. D.S. and H.J.M. share senior authorship of this work.

MACHINE learning (ML), a subdomain of artificial intelligence (AI), has already revolutionized many industries and has the potential to disrupt medicine and surgery.¹ There has been rapid growth in the efforts of ML models to interpret the medical data, including natural language documentation and diagnostics.^{2,3} Of significance to surgeons is the potential of ML to interpret videos of events that occur in operations. With advancements in computational power, we will be able to apply ML to surgery in real time.³ An initial step in this process is training ML systems to recognize and analyze the critical components of surgery by using ML techniques.

An established method for this is “operative workflow analysis”—systematically deconstructing operations into steps and phases.⁴ A step refers to the completion of a named surgical objective (e.g., hemostasis), whereas a phase represents a major surgical event that is composed of a series of steps (e.g., closure).⁴ During each step, certain surgical instruments (e.g., forceps) are used to achieve a specific objective, and there is the potential for technical error (lapses in operative technique), which may result in adverse events.⁴ ML-based recognition of these elements will thus allow surgical workflow analysis to be generated automatically and accurately. Such technology has the potential to provide standardized operative skill assessment, automate the generation of operative notes, allow off-line video indexing for education, facilitate the creation of simulations, and augment surgical training programs.^{5–9} By integration with the wider surgical team (such as nursing staff and anesthesiologists), these ML systems may aid orchestration of the team to a common workflow, improving efficiency and resource management.⁶ Additionally, this complements the potential for real-time intraoperative ML guidance for surgeons and facilitates progression through the surgical steps, potentially reducing operative times and errors.^{10,11}

Artificial deep neural networks (DNNs), a type of ML model, have previously achieved automatic accurate phase and instrument recognition in cataract surgery and laparoscopic cholecystectomy.^{12–15} Low-volume surgeries or those with steeper learning curves may especially benefit from augmented training, assessment, simulation, and intraoperative guidance.¹⁴ The endoscopic transsphenoidal approach (eTSA) to resection of pituitary adenomas is an exemplar—being performed at tertiary level care, at a comparatively low volume with a steep learning curve. It is therefore an ideal application of ML-based operative workflow analysis and would represent, to the best of our knowledge, the first neurosurgical operation analyzed in this way. Crucial to the safe integration of such technology is structured and iterative development, best captured by the Idea, Development, Exploration, Assessment, Long-term study (IDEAL) stages—beginning at the preclinical stage 0.^{16,17} In this IDEAL stage 0 study, we sought to use Touch Surgery for the development and evaluation of ML-powered analysis of the phases and steps in eTSA pituitary surgery.

Methods

This paper was generated using multiple reporting guidelines, given that no single guideline comprehensive-

ly captures this preclinical stage of ML technology development yet. We were therefore guided by the relevant sections of the IDEAL framework, and by TRIPOD and CONSORT-AI reporting guidelines.^{16–19}

Study Design

A preclinical development and evaluation (IDEAL stage 0) design was adopted.^{16,17} The study was based at a tertiary neurosurgical center (National Hospital for Neurology and Neurosurgery, London), which acts as a regional referral center for pituitary tumors and performs approximately 150–200 pituitary operations each year.

Data Collection

A library of anonymized operative videos of the eTSA for pituitary adenoma was used for ML model development. Videos were collected from surgical cases treated between August 8, 2018, and October 11, 2020. Cases were included if an operative video was available that was complete or near complete (missing steps but not missing an entire phase). Cases using microscopic surgery, and revision surgery in which the primary surgery was performed within 6 months, were excluded. For each case, patient (age, biological sex, tumor type) and operative (operative time) characteristics were recorded. Informed written patient consent was obtained and the project was registered with the local governance committee.

All eTSAs for pituitary adenoma operations performed at our center are single specialty (neurosurgery), performed by an attending surgeon and a subspecialty fellow. The majority of cases are performed using a mononos-tril technique with an endoscope holder during the sellar phase of the operation. Operative videos were recorded using a high-definition endoscope (Hopkins Telescope with AIDA storage system; Karl Storz Endoscopy). Videos were exported (MPV format) onto an encrypted hard drive and uploaded to Touch Surgery (Medtronic, Inc.; <https://www.touchsurgery.com/professional>), web-based software for surgical video storage and ML-derived surgical analytics.

In total, 50 videos were collected, which allowed for an 80/20 split with 40 videos used for training and 10 videos used for testing and evaluation (selected via a random number table). This 80/20 split between training and testing sets has generally been adopted within the ML literature for this sample size, and considers a minimum of 25 videos as sufficient for model training and 6 videos for testing.^{12,20}

Data Labeling

A workflow of 3 surgical phases and 7 constituent steps (Table 1) was generated through literature review and local expert surgeon consensus (N.L.D., H.J.M.). Operative video labeling of steps and phases was performed using Touch Surgery by at least 2 authors in duplicate (D.Z.K., H.J.M.), with differences settled through discussion and mutual agreement. “Steps” were defined as a sequence of activities used to achieve a surgical objective, and “phases” as a major event occurring during a surgical procedure, composed of several steps.⁴ Through labeling of step

TABLE 1. Phases and steps used for labeling of operative videos in patients undergoing eTSA

Phase	Step
Nasal	Nasal corridor creation (displace turbinates & identify sphenoid ostium)
	Anterior sphenoidotomy
	Posterior septectomy & removal of sphenoid septations
Sellar	Anterior sellar wall removal
	Tumor identification & excision
Closure	Hemostasis
	Skull base reconstruction (including repair of CSF leak)

and phase time stamps, manual video segmentation was achieved. Of note, videos were not submitted to formal analysis if a large portion (an entire phase or more) of the recording was missing.

Model Development and Evaluation

The training video set ($n = 40$) was analyzed by Touch Surgery to develop an ML model capable of recognizing the phases and the steps of the procedure. To develop an ML model to perform surgical workflow analysis, frames were extracted from each of the 40 videos at a constant frame rate (1 frame per second) and associated with a label indicating the phase and step to which they belonged according to the expert annotations. Using these frames as visual input and the associated label as the ground truth target, a 2-stage training pipeline was introduced in which convolutional neural network (CNN) models were first pretrained to recognize steps and phases from a short temporal window (1–5 frames). From a computer vision perspective, due to the ambiguity of the different anatomical and instrument landmarks visible in the dynamic field of view during different steps, a single frame or short sequence may not carry sufficient information to aid correct classification.^{12,21} To compensate for this, once networks were pretrained, a recurrent neural network (RNN) was trained in order to improve temporal resolution and the consistency of the predictions (Fig. 1).^{12,20}

The accuracy of the final model was then evaluated using the testing video set ($n = 10$), comparing the step and phase recognition of the model to the phase or step label assigned by expert surgeons. The evaluation metrics used were accuracy, precision, recall, and F1 score. Accuracy was calculated as the average per-class correct classification ratio of all the frames for each class. F1 score, the harmonic mean between precision (or positive predictive value) and recall (or sensitivity), was calculated per class (as defined in eq. 1) and then averaged across classes. We considered an accuracy of $\geq 90\%$ for phase recognition and 70% for step recognition as sufficient prior to progression to prospective, real-time, first-in-human studies (IDEAL stage 1).^{12,21}

Equations

We used the following equations for our calculations.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}} \quad [\text{eq. 1}]$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad [\text{eq. 2}]$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad [\text{eq. 3}]$$

$$\text{F1 score} = \frac{2 \times (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad [\text{eq. 4}]$$

Results

General Characteristics

A total of 50 cases (49 patients) of eTSA for pituitary surgery were included in the final analysis. The median age of included patients was 52 years (IQR 41–68 years), with a 25:24 male/female ratio. All cases were considered pituitary adenomas at the time of resection, with most of these considered macroadenomas on radiological assessment (46/50, 92%). Histological analysis confirmed pituitary adenomas in the majority of included cases (48/50, 96%); other pathologies included lymphocytic hypophysitis (1/50, 2%) and chronic lymphocytic leukemia (1/50, 2%). Forty-seven cases were primary surgeries and 3 were revision surgeries.

Video Characteristics

The median length of operation was 67 minutes (range 31–146 minutes), with a video resolution of 1280 × 720 pixels. Forty-five extracted operative videos were complete (45/50, 90%), with a minority missing a small number of steps (5/50, 10%). Figure 2 details the average duration of each operative step. The longest phase was the sellar phase (median 28 minutes, IQR 21–33 minutes), followed by the nasal (median 22 minutes, IQR 14–29 minutes) and closure (median 14 minutes, IQR 6–20 minutes) phases. The average duration of steps in descending order was as follows: step 5 (tumor identification and excision; median 17 minutes, IQR 14–21 minutes); step 3 (posterior septectomy and removal of sphenoid septations; median 14 minutes, IQR 8–23 minutes); step 4 (anterior sellar wall removal; median 10 minutes, IQR 6–13 minutes); step 7 (skull base reconstruction; median 10 minutes, IQR 9–12 minutes); step 6 (hemostasis; median 6 minutes, IQR 5–6 minutes); step 2 (anterior sphenoidotomy; median 4 minutes, IQR 3–6 minutes); and step 1 (nasal corridor creation; median 3 minutes, IQR 2–5 minutes).

Figure 3 highlights variations in the temporal relationship of steps, with 50% of videos containing all 7 steps performed sequentially in numerical order, and the other 50% of operations either 1) not containing all 7 steps or 2) containing all steps but in nonnumerical order. For example, 20% of videos did not include a formal step 7 (closure)—either reflecting heterogeneity in practice among local surgeons in certain clinical contexts or being due to incomplete video data. The steps repeated most frequently were steps 3 (posterior septectomy and removal of sphenoid septations) and step 5 (tumor identification and excision).

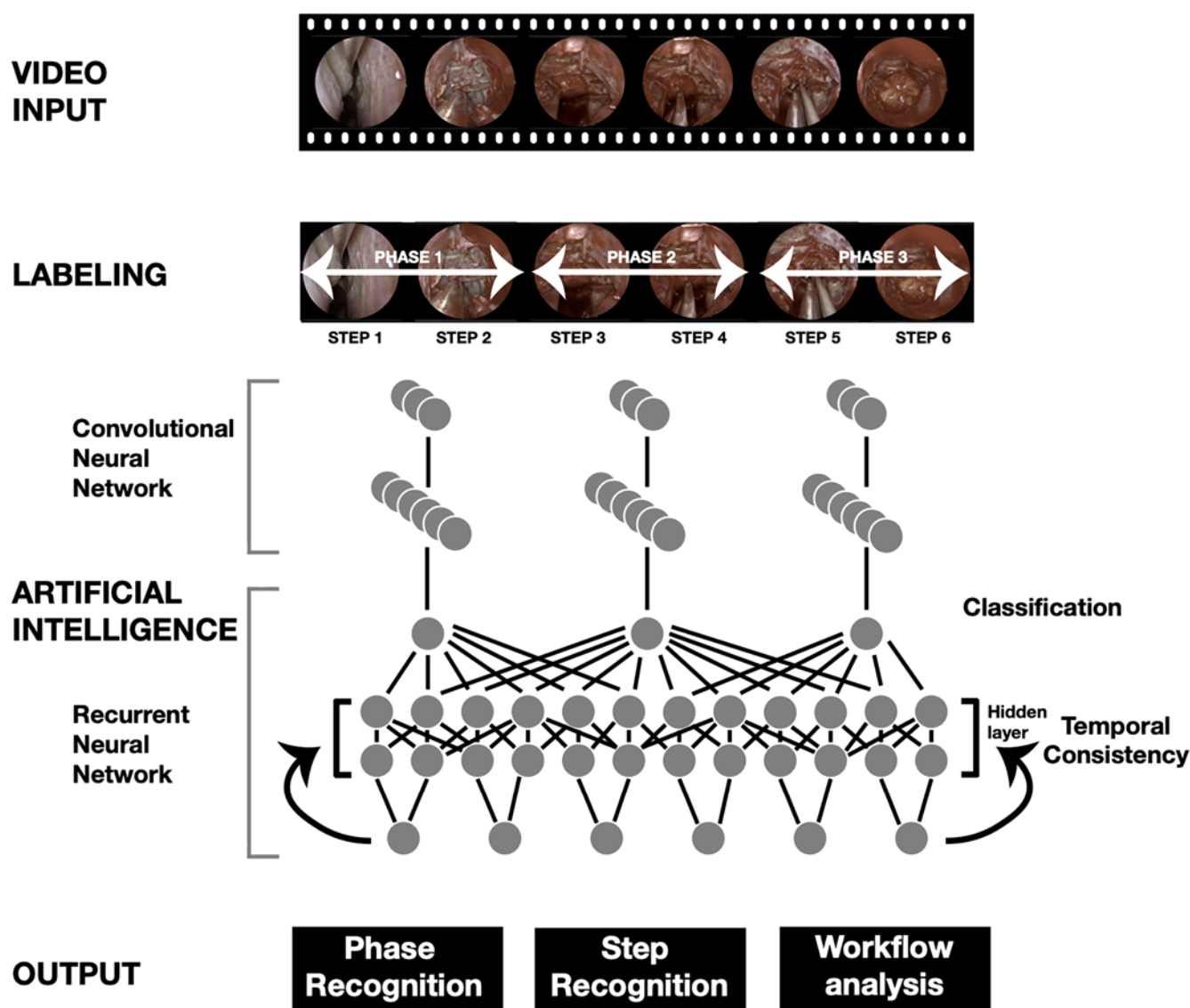


FIG. 1. Overview of operative video processing and analysis. Figure is available in color online only.

noid septations, 7 times); 4 (anterior sellar wall removal, 7 times); and 5 (tumor identification and excision, 2 times). However, steps 1 (nasal corridor creation) and 7 (skull base reconstruction) were never repeated.

Model Performance

During development of the ML model, different approaches were tested iteratively until the final version for this study was achieved. The stand-alone CNN network achieved accuracies of 80% for phase recognition and 65% for step recognition. The addition of the RNN improved accuracies to 86% and 73% for phase and step recognition, respectively. Final postprocessing improvements further boosted the final performance to 91% and 76% for phase and step recognition, respectively. Final model evaluation metrics are displayed in Table 2. F1 scores were of a similar degree—with phase recognition of 90% and step

recognition of 75%. Video 1 displays the model's predictions in action during an illustrative operation.

VIDEO 1. Operative video displaying Touch Surgery step and phase predictions. Steps are included in the *upper left* corner and phases are included in the *upper right*. "GT" in yellow represents the "ground truth" steps or phases labeled by experts. "Pred" represents the algorithm's prediction of steps and phases. The prediction text is *green* when correct (aligning with the ground truth) and *red* when incorrect. For each time point, a prediction certainty percentage is presented. Copyright 2021 Medtronic. All rights reserved. Used with the permission of Medtronic. Click here to view.

Discussion

Principal Findings

We have demonstrated that Touch Surgery developed and evaluated an accurate and automated ML model for surgical workflow recognition that is capable of detecting

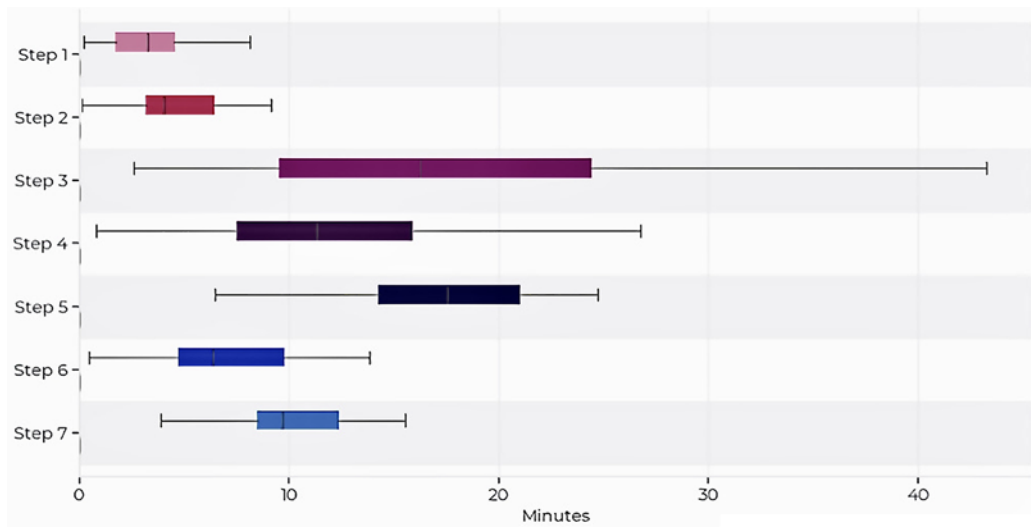


FIG. 2. Average and range of time per step. Copyright 2021 Medtronic. All rights reserved. Used with the permission of Medtronic. Figure is available in color online only.

the steps and phases of an operation. In this IDEAL stage 0 study, our ML model achieves 91% accuracy (F1 score 90%) for phase recognition and 76% accuracy (F1 score 75%) for step recognition in eTSA for pituitary surgery. The analysis of the eTSA for resection of pituitary adenomas makes this the first study of its kind in neu-

rosurgery.²² The eTSA has emerged as the first-line approach for resecting the majority of symptomatic pituitary adenomas.^{23–25} However, there is variation in the ways in which these operations are performed, as evidenced by our analysis of surgical steps, which vary in duration and order.^{26–29} Variations in practice are largely guided by lo-

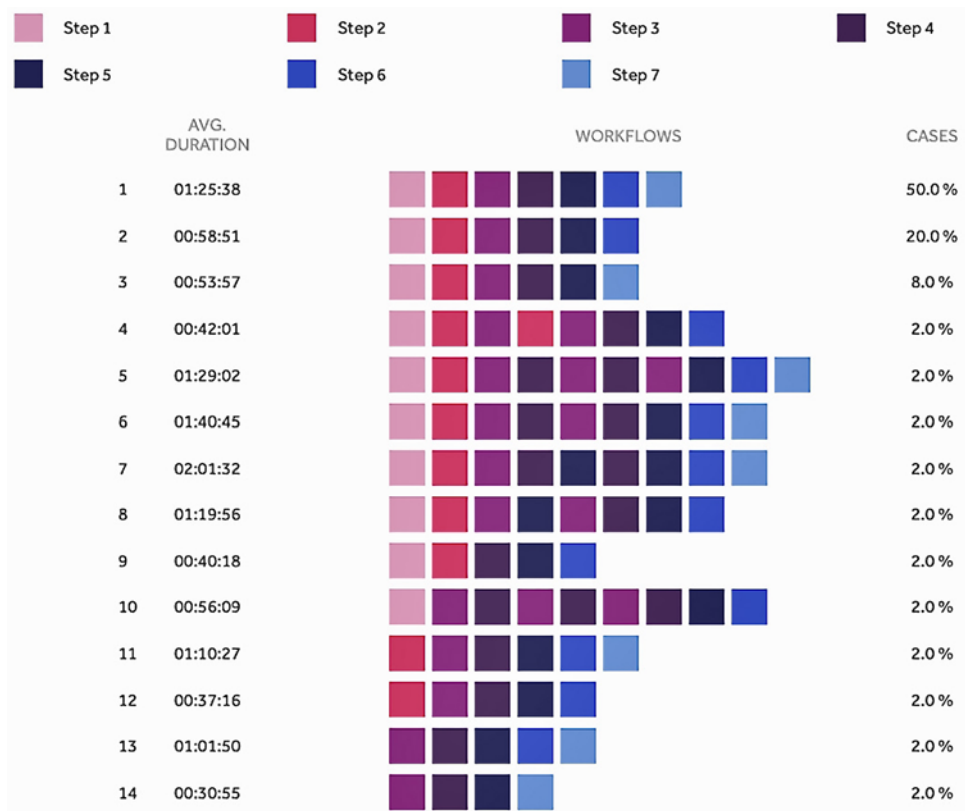


FIG. 3. Step variations in operative video workflow. Copyright 2021 Medtronic. All rights reserved. Used with the permission of Medtronic. Figure is available in color online only.

TABLE 2. Summary of overall recall, precision, F1 score, and accuracy for steps and phases in patients undergoing eTSA

Evaluation Metric	Recall	Precision	F1 Score	Accuracy
Phases	89.23%	91.49%	90.24%	91.25%
Steps	71.98%	82.09%	75.42%	75.69%

cal preference and may affect surgical outcomes.^{26–29} We found that despite variations within our own practice, Touch Surgery generated an ML model capable of accurate phase and step recognition.

There is a steep learning curve for the attainment of the necessary endoscopic skills, and this is compounded by the relatively low-volume nature of pituitary surgery.^{30–32} This is particularly evident during training programs, with the majority of US residents having performed fewer than 10 pituitary surgeries during training, thus requiring dedicated fellowships to gain the necessary skills and competency for these operations.³³ As residency programs move to competency-based frameworks and pituitary services are consolidated into centers of excellence, structured training and objective assessment of pituitary surgery are increasingly relevant.^{32,34}

Automated operative workflow analysis may meet these educational and training demands.^{9,14,21} For example, automatic indexing of contemporary operative videos may supplement teaching of residents and fellows, and may facilitate personal reflection on particular aspects of surgical performance (e.g., a technically challenging step).²¹ Similarly, deconstructing videos into critical operative steps may facilitate comparative analysis of the surgical performance of individual surgeons of various grades—examining step order and durations.¹⁴ Building on this, ML models have been used to analyze operative step performance, allowing assessment of operation-specific competency in a structured, objective, and personalized way.^{14,35,36} The automatic collation of surgical workflow and performance metrics can also provide the foundations of simulation (physical or virtual) development and validation.^{5–9}

Moreover, the integration of ML-based operative step analysis into the live operating room environment may improve surgical team efficiency and resource management. For example, through orientation onto the current proceedings (e.g., phase and step) and anticipation of the next necessary instruments, the entire surgical team is orchestrated to a common workflow and is prepared for the immediate next steps.³⁷ This would require the integration of ML systems into the workflows of the wider surgical team (e.g., scrub technicians and anesthesiologists), a concept that has been found to be generally acceptable to team members.³⁸ For more junior members of the surgical team, live display of workflow metrics (current step, time spent per step, required instruments) may supplement the guidance of senior surgeons, facilitate progress through operative steps, and consolidate in-practice learning.^{5–9} Operations performed in this “smart” operating room may therefore be shorter and more economical (e.g., unnecessary instruments not opened or used).⁶

Furthermore, in this era of personalized medicine, we are moving toward the data-driven analysis of the entire in-

dividual patient pathway. Combining intraoperative phase and step recognition with other ML-based technologies has numerous potential uses—including administrative, patient selection, and outcome prediction. For example, after automated deconstruction of an operation into constituent components, natural language processing techniques may be used for automatic generation of operative notes—which are generally otherwise template-based and may omit up to 50% of essential steps and events.^{2,39} Similarly, incorporating preoperative variables (e.g., automated imaging analysis) into ML models with intraoperative events may highlight characteristics (e.g., age or tumor morphology) that are predictive of successful outcomes or complications. Such information could be used to aid patient selection, preoperative planning, and tailored informed consent—adapted to individual patient factors.²² Finally, the correlation of intraoperative video data to postoperative outcomes may allow prediction of outcomes postoperatively and exploration of the maneuvers, instruments, or materials linked to superior or inferior outcomes. Not only would this allow for targeted refinement of operative techniques, but it may also support early identification of errors and form part of early-warning systems for potential adverse events.^{40,41} Ultimately, this ML-based surgical workflow analysis—through its uses in education, training, intraoperative guidance, and patient pathway integration—aims to make surgery safer, more effective, and even more individualized.^{21,37}

Findings in the Context of the Literature

Surgical phase recognition through ML approaches is a growing field, most prevalent in general surgery (e.g., laparoscopic cholecystectomy) and ophthalmological surgery (e.g., cataract surgery).²¹ There have been no reports of the use of this technology in neurosurgery,^{21,22} although surveys of neurosurgeons and neurosurgical patients indicate receptiveness to AI being integrated into operating rooms.^{38,42}

Multiple ML and statistical models have been explored, including dynamic time warping, hidden Markov models, support vector machines, and DNNs. The DNNs (particularly CNNs and RNNs) are used most commonly and, although they require more data and higher computational power, they have displayed increased phase recognition accuracy in multiple studies across surgical specialties. Indeed, a recent study comparing the use of multiple ML models in cataract surgery phase recognition found a combined CNN-RNN configuration with temporal modeling with accuracies > 90%.¹⁴ Touch Surgery’s ML model is a similar configuration and has displayed accuracies of 93% and 73% for phase and step recognition, respectively, in cataract surgery workflow analysis prior to its application in pituitary surgery.¹² Similar heights of accuracy have been observed in laparoscopic sleeve gastrectomy, with an accuracy of up to 86% for operative step detection using a CNN-RNN temporal model.¹¹ Additionally, accurate phase recognition has been achieved for peroral endoscopic myotomy (87.6% accuracy)⁴³ but has proven to be more difficult in laparoscopic proctocolectomies (67% accuracy), where the operative steps used are less standardized and more complex.⁴⁴

Limitations and Strengths

This study has validated the first ML model that is capable of analyzing neurosurgery videos. However, it has several limitations, as follows: 1) the included sample is small and highly selected (single-center, nonconsecutive, endoscopic endonasal technique); 2) the model may be overfitted to local surgical practice; and 3) prospective validation has yet to be performed and external validity has yet to be determined. The labeled steps and phases were based on local consensus, and therefore further work is needed to create a standardized and generalizable workflow framework for this procedure. Steps at present may contain multiple substeps, which are likely to require a larger sample size before the model can accurately delineate them. Looking ahead, progression through IDEAL stages will facilitate further development of this technology. In particular, analysis of multicenter and prospective data will increase the predictive potential and generalizability of the ML software, facilitating its integration into “smart” operating theaters with real-time operation analysis.⁴¹ Refinement of an ML model on this scale may facilitate transfer learning, such that the trained algorithm is adapted to other surgeries after a period of adjustment—without requiring total de novo model creation and potentially requiring less data to achieve accurate phase and step recognition.

Conclusions

In this IDEAL stage 0 study, ML techniques have been used to automatically analyze operative videos of eTSA for pituitary surgery. Using a combined CNN and RNN model, Touch Surgery achieved phase and step recognition accuracies of 91% and 76%, respectively. ML-based surgical workflow analysis has numerous potential uses—such as education (e.g., automatic indexing of contemporary operative videos for teaching); improved operative efficiency (e.g., orchestrating the entire surgical team to a common workflow); and improved patient outcomes (e.g., comparison of surgical techniques or early detection of adverse events). This technology has previously been shown to be acceptable to neurosurgical teams and patients. Future directions include the real-time integration of Touch Surgery into the live operative environment as an IDEAL stage 1 (first-in-human) study, and further development of underpinning ML models using larger data sets.

Acknowledgments

Figure 1 was produced by Amy Warnock at the Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London. No specific funding was received for this piece of work. Mr. Khan, Mr. Muirhead, Mr. Marcus, Dr. Koh, Dr. Layard Horsfall, and Prof. Stoyanov are supported by the Wellcome (203145Z/16/Z) EPSRC (NS/A000050/1) Centre for Interventional and Surgical Sciences, University College London. Mr. Khan is also funded by an NIHR Academic Clinical Fellowship. Mr. Marcus is also funded by the NIHR Biomedical Research Centre at University College London. Mr. Vasey receives support of a non-study-related clinical or research effort that he oversees from the Berrow Foundation. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any author's accepted manuscript version arising from this submission.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56.
2. Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc*. 2020;27(3):457-470.
3. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018; 268(1):70-76.
4. Lalys F, Jannin P. Surgical process modelling: a review. *Int J CARS*. 2014;9(3):495-511.
5. Sarker SK, Chang A, Albrani T, Vincent C. Constructing hierarchical task analysis in surgery. *Surg Endosc*. 2008;22(1): 107-111.
6. Dijkstra FA, Bosker RJI, Veeger NJGM, van Det MJ, Pierie JP. Procedural key steps in laparoscopic colorectal surgery, consensus through Delphi methodology. *Surg Endosc*. 2015; 29(9):2620-2627.
7. Strauss G, Fischer M, Meixensberger J, Falk V, Trantakis C, Winkler D, et al. Workflow analysis to assess the efficiency of intraoperative technology using the example of functional endoscopic sinus surgery. Article in German. *HNO*. 2006; 54(7):528-535.
8. Krauss A, Muensterer OJ, Neumuth T, Wachowiak R, Donaubauer B, Korb W, Burgert O. Workflow analysis of laparoscopic Nissen fundoplication in infant pigs—a model for surgical feedback and training. *J Laparoendosc Adv Surg Tech A*. 2009;19(suppl 1):S117-S122.
9. Grenda TR, Pradarelli JC, Dimick JB. Using surgical video to improve technique and skill. *Ann Surg*. 2016;264(1):32-33.
10. Maktabi M, Neumuth T. Online time and resource management based on surgical workflow time series analysis. *Int J CARS*. 2017;12(2):325-338.
11. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019; 270(3):414-421.
12. Zisimopoulos O, Flouty E, Luengo I, Giataganas P, Nehme J, Chow A, et al. DeepPhase: surgical phase recognition in CATARACTS videos. In: *Proceeding of 21st International Conference, Medical Image Computing and Computer-Assisted Intervention*; September 16–20, 2018; Granada, Spain. Springer; 2018:265-272.
13. Lecuyer G, Ragot M, Martin N, Launay L, Jannin P. Assisted phase and step annotation for surgical videos. *Int J CARS*. 2020;15(4):673-680.
14. Yu F, Silva Croso G, Kim TS, Song Z, Parker F, Hager GD, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA Netw Open*. 2019;2(4):e191860.
15. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Paday N. Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging*. 2017;36(1):86-97.
16. Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Ann Surg*. 2019;269(2):211-220.
17. Marcus HJ, Bennet A, Chiari A, Day T, Hirst A, Hughes-Hallett A, et al. IDEAL-D framework for device innovation: a consensus statement on the preclinical stage. *Ann Surg*. Published online on April 7, 2021. doi:10.1097/SLA.0000000000004907
18. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020;2(10):e537-e548.
19. Collins GS, Reitsma JB, Altman DG, Moons KG. Transpar-

- ent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation*. 2015;131(2):211-219.
20. Kadhodamohammadi A, Sivanesan Utharaj N, Giataganas P, Gras G, Kerr K, Luengo I, et al. Towards video-based surgical workflow understanding in open orthopaedic surgery. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2020; 9(3):286-293.
 21. Garrow CR, Kowalewski KF, Li L, Wagner M, Schmidt MW, Engelhardt S, et al. Machine learning for surgical phase recognition: a systematic review. *Ann Surg*. 2020;273(4):684-693.
 22. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. 2019;43(5): 1235-1253.
 23. Cappabianca P, Cavallo LM, de Divitiis E. Endoscopic endonasal transsphenoidal surgery. *Neurosurgery*. 2004;55(4): 933-941.
 24. Liu JK, Das K, Weiss MH, Laws ER Jr, Couldwell WT. The history and evolution of transsphenoidal surgery. *J Neurosurg*. 2001;95(6):1083-1096.
 25. Couldwell WT, Weiss MH, Rabb C, Liu JK, Apfelbaum RI, Fukushima T. Variations on the standard transsphenoidal approach to the sellar region, with emphasis on the extended approaches and parasellar approaches: surgical experience in 105 cases. *Neurosurgery*. 2004;55(3):539-550.
 26. Buchfelder M, Schlaffner S. Pituitary surgery for Cushing's disease. *Neuroendocrinology*. 2010;92(suppl 1):102-106.
 27. Lucas JW, Zada G. Endoscopic surgery for pituitary tumors. *Neurosurg Clin N Am*. 2012;23(4):555-569.
 28. Shah NJ, Navnit M, Deopujari CE, Mukerji SS. Endoscopic pituitary surgery—a beginner's guide. *Indian J Otolaryngol Head Neck Surg*. 2004;56(1):71-78.
 29. Cappabianca P, Cavallo LM, de Divitiis O, Solari D, Esposito F, Colao A. Endoscopic pituitary surgery. *Pituitary*. 2008; 11(4):385-390.
 30. Leach P, Abou-Zeid AH, Kearney T, Davis J, Trainer PJ, Gnanalingham KK. Endoscopic transsphenoidal pituitary surgery: evidence of an operative learning curve. *Neurosurgery*. 2010;67(5):1205-1212.
 31. Snyderman C, Kassam A, Carrau R, Mintz A, Gardner P, Prevedello DM. Acquisition of surgical skills for endonasal skull base surgery: a training program. *Laryngoscope*. 2007; 117(4):699-705.
 32. McLaughlin N, Laws ER, Oyesiku NM, Katznelson L, Kelly DF. Pituitary centers of excellence. *Neurosurgery*. 2012;71(5): 916-926.
 33. Jane JA Jr, Sulton LD, Laws ER Jr. Surgery for primary brain tumors at United States academic training centers: results from the Residency Review Committee for neurological surgery. *J Neurosurg*. 2005;103(5):789-793.
 34. Casanueva FF, Barkan AL, Buchfelder M, Klibanski A, Laws ER, Loeffler JS, et al. Criteria for the definition of pituitary tumor centers of excellence (PTCOE): a Pituitary Society statement. *Pituitary*. 2017;20(5):489-498.
 35. Zia A, Sharma Y, Bettadapura V, Sarin EL, Essa I. Video and accelerometer-based motion analysis for automated surgical skills assessment. *Int J CARS*. 2018;13(3):443-455.
 36. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Netw Open*. 2020;3(3):e201664.
 37. Loukas C. Video content analysis of surgical procedures. *Surg Endosc*. 2018;32(2):553-568.
 38. Horsfall HL, Palmisciano P, Khan DZ, Muirhead W, Koh CH, Stoyanov D, et al. Attitudes of the surgical team toward artificial intelligence in neurosurgery: an international two-stage cross-sectional survey. *World Neurosurg*. 2020;146: e724-e730.
 39. van de Graaf FW, Lange MM, Spakman JI, van Grevenstein WMU, Lips D, de Graaf EJR, et al. Comparison of systematic video documentation with narrative operative report in colorectal cancer surgery. *JAMA Surg*. 2019;154(5):381-389.
 40. Padoy N. Machine and deep learning for workflow recognition during surgery. *Minim Invasive Ther Allied Technol*. 2019;28(2):82-90.
 41. Bonrath EM, Gordon LE, Grantcharov TP. Characterising 'near miss' events in complex laparoscopic surgery through video analysis. *BMJ Qual Saf*. 2015;24(8):516-521.
 42. Palmisciano P, Jamjoom AAB, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. *World Neurosurg*. 2020;138: e627-e633.
 43. Ward TM, Hashimoto DA, Ban Y, Rattner DW, Inoue H, Lillemoe KD, et al. Automated operative phase identification in peroral endoscopic myotomy. *Surg Endosc*. 2020;35(7):4008-4015.
 44. Bodenstedt S, Wagner M, Katić D, Mietkowski P, Mayer B, Kenngott H, et al. Unsupervised temporal context learning using convolutional neural networks for laparoscopic workflow analysis. arXiv. Preprint posted online February 13, 2017. doi:1702.03684

Disclosures

Mr. Barbarisi; Drs. Luengo, Addis, Culshaw, Haikka, and Kerr; and Prof. Stoyanov are employees of Digital Surgery, Medtronic, which is developing products related to the research described in this paper. Prof. Stoyanov also has direct stock ownership in Odin Vision, Ltd.

Supplemental Information

Videos

Video 1. <https://vimeo.com/574862056>.

Author Contributions

Conception and design: Khan, Luengo, Barbarisi, Culshaw, Dorward, Kerr, Koh, Layard Horsfall, Muirhead, Vasey, Stoyanov, Marcus. Analysis and interpretation of data: Khan, Luengo, Barbarisi, Addis, Culshaw, Haikka, Kerr, Marcus. Drafting the article: Khan, Luengo, Kerr, Marcus. Critically revising the article: Khan, Addis, Culshaw, Dorward, Haikka, Jain, Kerr, Koh, Layard Horsfall, Muirhead, Palmisciano, Vasey, Stoyanov, Marcus. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Khan. Administrative/technical/material support: Luengo, Barbarisi, Culshaw, Dorward, Jain, Kerr, Koh, Layard Horsfall, Muirhead, Palmisciano, Stoyanov, Marcus. Study supervision: Stoyanov, Marcus.

Correspondence

Danyal Z. Khan: National Hospital for Neurology and Neurosurgery & Wellcome/EPSCRC Centre for Interventional and Surgical Sciences, University College London, United Kingdom. d.khan@ucl.ac.uk.