A Visual Perception of Robotic Skills Using Domain-Adapted Machine Learning Technique

S.Ajaykrishna,
PG Scholar, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India,

Dr., N. Pughazendi,
Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India,

Abstract: In this research, I offer a domain-adapted end-to-end learning approach for executing and assessing surgical skills by various contributors using automated systems. It detects smoothness and retrieves training knowledge within a specified trajectory. The trajectories approach is more accurate in similar stochastic or machine learning modeling techniques, making it effective for extracting and understanding data features related to abilities. A JIGSAWS data set should be utilized to assess the suggested strategy. The t-SNE methodology is used in this approach, which visual mediums a low-dimensional representations of a variety of trials while emphasising sensitive information and returning odd or flawed trials as outliers separate from more typical skill or participant groups. The information trajectory can be utilised for analysis and learning in surgical assessments and training programmes.

Index Terms: Stochastic Neighbor Embedding with t-distribution, Contrastive Principal Component Analysis (c PCA), Ensemble Models, and Surgical Skills Evaluation (t -SNE).

INTRODUCTION

Ensemble modelling is the creation of several distinct models to forecast a result, either by utilising a variety of modelling techniques or by utilising various training data sets. An ensemble model then aggregates the result of the each base model to create a single, definitive prediction for such unobserved data. The purpose of using ensemble models would be to reduce the forecast's prediction error. Whenever the base models were diverse and autonomous, the ensemble approach lowers the model's forecast error. The approach uses people's collective wisdom to generate a forecast. The ensemble model behaves and functions as a single model even if it comprises several basis models. Principal Component Analysis (PCA), one of the most frequently utilized unsupervised machine-learning techniques, is used in a variety of applications, include exploratory analysis, dimension reduction, information compressing, information de-noising, and many more. an unsupervised, random method that is only utilized for visualisation In order to keep the extremely similar datasets near to each other in lower-dimensional space, use a non-linear dimension reduction technique.

utilises a student t-distribution to figure out the similarity between two points in lower-dimensional space while keeping the local structure of the data. t-SNE overcomes overcrowding and optimisation issues by using a heavy-tailed Student-t probabilities rather than a Gaussian distribution to analyse the similarities between two points in low-dimensional space. The t-SNE is unaffected by outliers.

RELATED WORKS

Robotic-assisted, minimally-invasive surgery is becoming increasingly widespread in current clinical practise (RAMIS). Surgeons need a number of abilities in order to execute RAMIS properly and professionally [1]. It could be beneficial for surgery trainees to have accurate clinical evaluation processes with instructive and enlightening remarks. For many years, outcome-based assessments, standardised checklist, and evaluations have been utilized to evaluate RAMIS proficiency [2]. Due to human bias and variances in how people interpret comparable occurrences, these qualitative evaluation approaches need significant expert supervision, time, and manual assessments, which making them less effective and less dependable. These observation grading methods might also be unresponsive to tiny but significant modifications in the trainee's level of proficiency and fall short of identifying the fundamental causes of surgical failures. On the other side, the automated RAMIS skills evaluation systems close these gaps, save time and money, and give beginning surgeons professional input during this learning period [3]. The application of artificially intelligent (AI) based programs, including such deep computing and machine learning technology, in the assessment of surgical skills may well be made possible by the fact that surgical data are becoming more readily available as a consequence of surgical robot capabilities. Information (or inductive) model and functionality (or domain experience and understanding) modeling are the two principal types of AI models used in self-determining robotic surgery assessment procedures [4]. Data-driven models are based on the core idea that user bias should not be introduced into the learning experience by using end-to-end models that have little to no domain experience. These methods allow the model to select its structure, change its hyperparameters, and add new attributes mostly based on the incoming data. In contrast, feature-based approaches do not even use a model to find traits that are already known owing to human perception or the system's dynamic equation. There are frequently modelling uncertainty and unmodeled dynamic in the operator's ability model, making it challenging to properly describe it with a limited supply of training data. By using domain expertise as prior offenses, which reduces uncertainty, it is possible to solve modelling problems with fewer data points for training [5].

A considerable collection of literature, including pieces from my research program, addresses the subject of self-determining robot skill evaluation utilising data-driven models. In order to discover skill-related temporal variations in participant behaviours while robotic surgery, [6], [7] utilize convolutional neural networks (CNNs). Recurrent neural networks and artificial neural network have both been used in several studies to consistently identify different degrees of surgical competence [8]. Spatial attention methods based on CNN or Recurrent neural networks are used in other publications, such as [3], [9], and [10], to identify elements in an endoscopic video sequence that are associated to abilities and evaluate the surgical expertise of users. Despite the fact that the majority of these experiments have shown a relatively low incidence of skill misclassification, both extraction of features and decision-making operations in such black-box models are murky and frequently questionable. Furthermore, the sheer volume of these quality model them constantly vulnerable to overfitting on tiny sets of data, particularly in the area of robot surgeries, where the lack of suitable human patients and the high expense of research restrict access to conventional and substantial datasets. The aforementioned restrictions make any performance assessments less precise since such trainee's performance rating is based on the model's learned parameters and its level of confidence in the predicted result (see [6]). In order to categorise participants according to their degree of experience, hidden Markov models (HMMs), a sort of data-driven research, are used to dissect a surgical action into its pre-defined building parts, or so-called gestures [11]. These approaches suffer with low identification rates, determining the appropriate amount of concealed states, and the time-consuming human annotation of a huge number of movements. Additionally, HMMs increase the chance of losing crucial temporal information by expanding trajectories into a distinct space characterised by static descriptors.

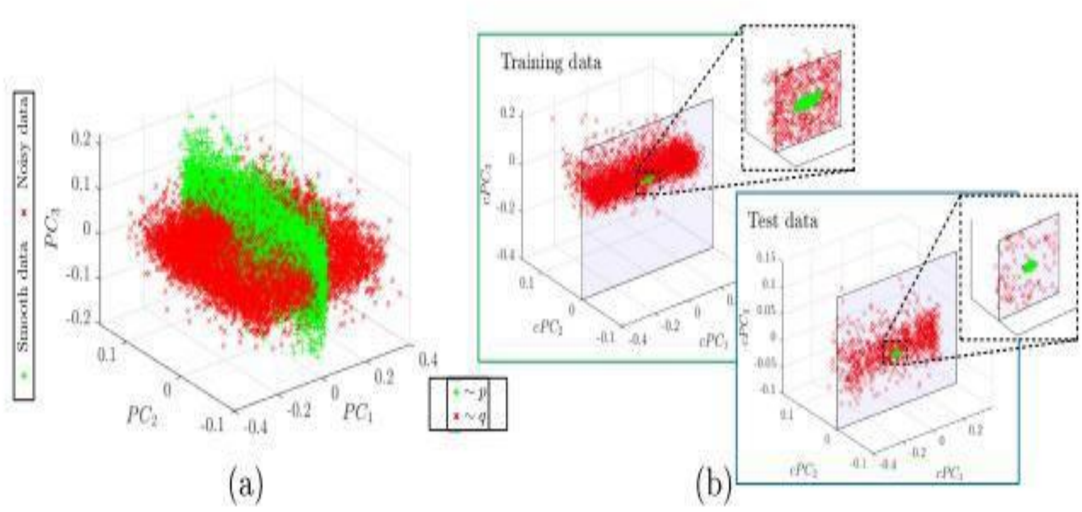

Fig 1 Using (a) PCA (8,000 training data), and (b) c PCA in the three dimensional sub-space (with 8,000 training data and 2,000 test data), adaption of [16], we visualise data points in X s and X n (n = 10,000) in the three dimensional space.

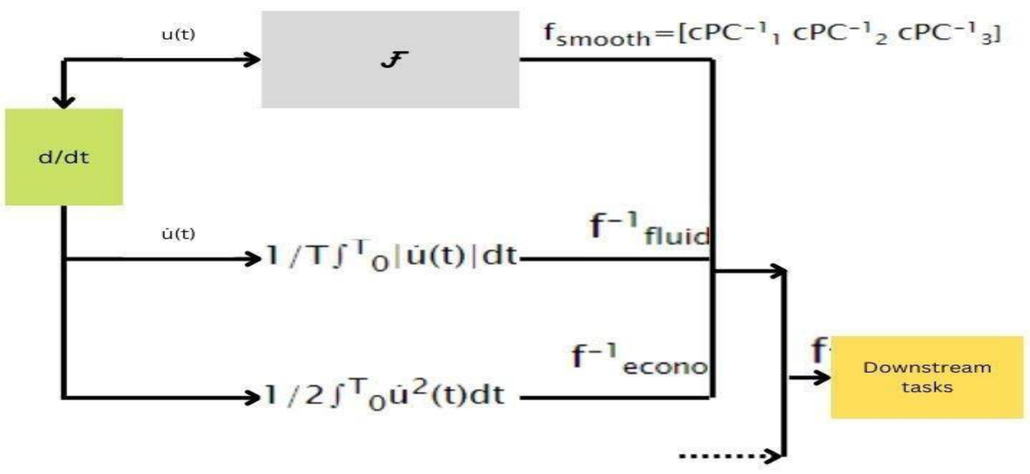

Fig 2 An example of the ensemble model utilised in this study.

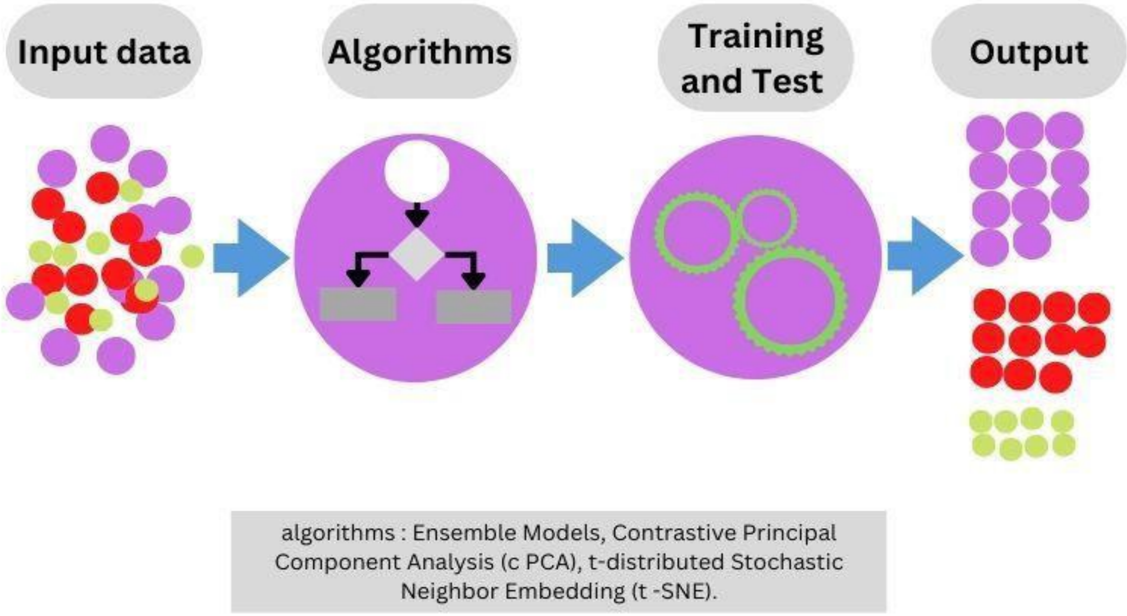

Fig 3 Using ensemble models, contrastive principal component analysis (c PCA), and t-distributed stochastic neighbour embedding, training and test dataset data flow architecture is shown (t -SNE).

METHODOLOGY

Features are created as assessment metrics to measure skills using feature-based methodologies, including total road length [12], movements jerks [13], computing time [12], etc. Finally, using descriptive statistical evaluation, each indication for human participants is compared separately. Because there is high statistical heterogeneity among as well as between participants in such measurements, there are usually evident overlaps among skill categories and very few statistically significant distinctions among various human users. This happens when certain features, such as motion jerk, are too noisy, while others, such as total pathways or processing time, are insufficient indications of the user's skill on their own. Furthermore, these publications need not address specific actions at the operational subtask level; rather, they only analyse global indicators throughout the whole process. For instance, most research neglect to consider problematic temporal metrics such trajectories smoothness (i.e., the absence of unpredictable motions such hand tremor and uncontrolled quick actions) as a crucial element in evaluating a particular trajectory's capabilities. Numerous measures that complement each other might be utilised to provide an accurate and descriptive depiction for surgical performance which can be compared. These tactics remained ad hoc & non-generalizable to new trajectories because they ignore temporal variations and do not establish a framework for skill evaluation.

In this study, we provide a domain-adapted technique for skills evaluation in surgical training programmes, where overall trajectory-based skills of surgeons are prioritised above patient surgical results, bridging the gap between data-driven models and feature-based models. This model mixes temporal properties learned from incoming data with measurements generated manually. In the framework of data-driven learning, my novel method utilises global measures like smoothness and motion effectiveness as acceptable elements for determining the level of skill of the performed trajectory and derives smoothness properties from the input data [14]. To overcome the challenging problem of identifying smoothness/noise inside trajectories, I simply apply contrastive principal component analysis (cPCA) approach [15]. These therapeutically significant properties will be combined to create a three-dimensional, expressive feature space that attracts attention to minute details of the operation while accurately reflecting the user's skill level. To display the high-dimensional feature map in such a three-dimensional embedding and assess the effectiveness of our suggested model, we use t-distributed stochastic neighbour embedding (t-SNE) as merely an unsupervised technique [16].

Whenever domain expertise and information are combined in an unsupervised type of learning, my technique outperforms previous black-box models—even ones that are transparency, reliability, and generalizability—deep learning-based. These characteristics make our method potentially useful in robotic surgery, an area where improved reliability and explicability in training and assessment procedures are highly desired. Notably, this letter cannot comment explicitly on how the trainee performed on their various subtasks because it only gives broad recommendations is for criteria to apply when evaluating surgical talents. However, anomalous behaviours will be implicitly reflected in the final general evaluation, and inaccurate trajectory graphs will be produced in the absence of more educated clusters. It has been shown that a casual viewer may recognise and evaluate a surgeon's competence levels by seeing recordings of surgical endoscopy made in advance operations carried out with a level of perfection comparable to that of a skilled surgical instructor [17]. Hypothesis that human talent assessment could be more instinctive and natural than complex assessment [18] lends weight to this fact. The spectator's lack of attention to minute features like translation and spins within the surgery trajectory may explain for this finding.

He or she focuses largely on the patient's ability to utilise their hands and equipment efficiently, fluidly, and smoothly. We conclude that the level of understanding assigned to the users should increase as the smoothly, fluidity, or power efficiency of a specific trajectory improve. These realities motivate us to build logical and open sub-models that are domain-adapted and independently capture every single one of these crucial elements. We'll demonstrate later how combining these sub-models enables us to create a useful feature extraction model appropriate for any ensuing purpose, such classification or data visualisation. The smoothness detection method is the most difficult component of creating our sub-models, thus i will first go over it in depth in the following sections. The problem arises from the fact that smoothness is in reality a temporal feature that can exist at any point along a particular trajectory but can also be obscured by other temporal data, such as seasonality & general trend patterns. Also, there isn't a simple description or method for identifying or locating non-smooth behaviour throughout time series.

High-dimensional data exploration is continuously seen as a tough endeavour across a wide range of application domains. Traditional approaches such as multidimensional scaling (MDS) and principal component analysis (PCA) [19, 20] are utilized to recognise prevalent themes in data. Other techniques, like localised linear embedding (LLE) [21] and isomap [22], are created to locate the preserved sub-manifold internal and external between pieces of data in the original feature space in order to capture different from the conventional inside more complex data sets. However, we are frequently looking for patterns or patterns that have a lot of certain features. For instance, We wish to emphasise cancer-related alterations that only distinguish the two groups of people in a data gathering of gene-expression assessments from healthy or malignant individuals.

If we use PCA or MDS straight away, it's quite probable that the top principal components now represent demographic variances in people, such as gene attributes associated to skin colour, age, or gender, rather than gene features important to disease aetiology. Similar to this, essential features also couldn't emerge as dominant latent variables utilizing nonlinear manifold learning techniques like LLE and Isomap or they could be entangled with the other significant ones in the low-dimensional encoding. In addition, techniques that preserve pairwise similarities in data (like MDS) or localised patterns in data (like LLE and Isomap) only provide mappings for the provided training endpoints and do not provide simple or predictable extensions for out-of-occurrences [23]. In other words, you must use these approaches for the new training collections, which comprises the old training set as well as the additional inquiry points, if I wish to incorporate extra test points. It is challenging to verify the method's generalisation in classification problems since we don't clearly distinguish among training and test sets. This requires extensive computing resources. When trying to extract crucial details from as a component of the categorization of surgical instruments, hand tremor or general noise on the inside of a certain trajectory talents, we run into a similar problem. Consider a made-up data set with two distinct sorts of trajectories, including randomly produced smooth trajectories χ_s and noisy trajectories χ_n, to represent the hand tremor (not sensor or process noises) $\chi = \chi_s \cup \chi_n$

$\chi_s = \{xi\}^n{}_{i=1} \in \mathbb{R}^{d\times n}$, $\chi_n = \{x\hat{}i\}^n{}_{i=1} \in \mathbb{R}^{d\times n}$

However, if these pieces of information are transferred to an embeddings that is distinct to each dimensions and reveals Real-world examples are found in between the two extreme values, which makes the disparity between samples from p and q the most noteworthy. Alternatively put, the equation for r may be created by adding p and q linearly. High-dimensional data called time series usually contain a number of recurrent and related variations. By using dimensionality reduction techniques, we can pick out and concentrate on the information that will have the most impact on our decisions. Explain four-dimensionality reductions and manifolds learning algorithms that attempt to disclose the low-dimensional underlying structure of the each time series. After detecting and removing superfluous correlations and dimensions, every point with in three-dimensional embedded is examined. Space reflects a unique time series with large dimensions. As the mapped measured values of both the two sets don't clearly demarcate their boundaries, applying traditional approaches does not result in a a two-dimensional low-level map improved noise related characteristics for the constructed training sets χ_s and χ_n.

The aforementioned measurements are insufficient to correctly reflect the user's competence levels during trajectory execution. Concatenating all of these variables and giving them to the subsequent data analysis is one potential option. That makes sense given that what a model must be taught to offer a large number of important, uncorrelated attributes in order to fill the feature space and increase accuracy and generalisation. Using an ensemble method in this kind of skills evaluation project has the noticeable benefit of making each factor's impact on the decision-making process evident. As a result, such framework may offer each user insightful information about their performance, shortcomings, and skills in many aspects of surgery. This clear benefit makes our approach apart from existing black box testing learning algorithms or deep learning algorithms, where the feature extraction and decision-making processes is not transparent, intelligible, or reliable in the context of end-to-end learning across limited data sets of surgical operations.

DISCUSSION

Another characteristic of a clustering G is the presence of intermediate users In2 close to experienced users. My method draws attention to the drawbacks of Annotations made by hand with several labels, which often have a gritty texture. In accordance with the standards in [16], inexperienced users have handled the da Vinci Surgery Systems for less than 10 hours, intermediate users with between ten and just one hundred surgeries, and expert users for more than one hundred procedures. There is little doubt the intermediate users are constrained significantly. Our endoscopic video analysis showed that In2 performed suturing tasks better than competent users. In2 might have experience using da Vinci Systems functioning for 100 hours. This benefit of our technique may be utilised to measure a patient's level of flexibility as they advance from of the subpar cluster to a outstanding cluster all through the training session.

The strategy outlined in this letter, as opposed to current end-to-end high-capacity algorithms for the same surgical techniques study must be conducted, first builds a limited number of critical features before further lowering the dimension of the feature space using t-SNE method to aid human assessment plus interpretation. Because there is always a cost to everything among predictive performance (dimensionally high features collection) and understandability (low in dimension representation), contrasting our method with other large-scale systems performance in categorising data is not totally fair because we are not in competition with current systems. on the basis of forecasting classifiers (i.e., crossvalidation accuracy).

CONCLUSION

This letter presents a brand-new, discipline-specific method for assessing, comprehending, and visualising surgical procedures. I drew on my domain expertise as inspiration and emphasised clinically significant characteristics like fluidity & economies for movement as crucial criteria for assessing the skills required for surgical activities. These assessments indicated major components of the patient's skill level when combined with smoothness-related parameters uncovered utilising an end-to-end preparatory. According to experimental results acquired from the Phantom robotic and the JIGSAWS testing dataset using the t-SNE approach, my method may efficiently transmit degree of competence, abnormality, and concealed data inside a certain surgical trajectory. My method beats current state-of-the-art black-box methods in terms of reliability, safety, and interpretability when measuring surgical task competence and trainee user approaches and procedures due to its clarity and explicability.

REFERENCES

[1] John D. Birkmeyer et al., “Surgical skill and complication rates after bariatric surgery,” New England J. Med., vol. 369, no. 15, pp. 1434–1442, 2013.

[2] N. Ahmidi et al., “A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery,” IEEE Trans. Biomed. Eng., vol. 64, no. 9, pp. 2025–2041, Sep. 2017.

[3] I. Funke et al., “Video-based surgical skill assessment using 3D convolutional neural networks,” Int. J. Comput. Assist. Radiol. Surg., vol. 14, no. 7, pp. 1217–1225, 2019.

[4] N. Muralidhar et al., “Incorporating prior domain knowledge into deep neural networks,” in Proc. IEEE Int. Conf. Big Data, 2018, pp. 36–45.

[5] L. von Rueden et al., “Informed machine learning–A taxonomy and survey of integrating knowledge into learning systems,” IEEE Trans. Knowl. Data Eng., early access, May 12, 2021, doi: 10.1109/TKDE.2021.3079836.

[6] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, “Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks,” Int. J. Comput. Assist. Radiol. Surg., vol. 14, pp. 1611–1617, 2019.

[7] A. Soleymani, X. Li and M. Tavakoli, “Deep neural skill assessment and transfer: Application to robotic surgery training,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2021, pp. 8822–8829.

[8] Xuan Anh Nguyen, D. Ljuhar, M. Pacilli, R. M. Nataraja, and S. Chauhana, “Surgical skill levels: Classification and analysis using deep neural network model and motion signals,” Comput. Methods Programs Biomed., vol. 177, pp. 1–8, 2019.

[9] H. Doughty,, W. Mayol-Cuevas, and D. Damen, “The pros and cons: Rankaware temporal attention for skill determination in long videos,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 7862–7871.

[10] A. Soleymani et al., “Surgical skill evaluation from robot-assisted surgery recordings,” in Proc. Int. Symp. Med. Robot., 2021, pp. 1–6.

[11] L. Tao et al., “Sparse hidden Markov models for surgical gesture classification and skill evaluation,” in Proc. Int. Conf. Inf. Process. Comput.-Assist. Interv., 2012, pp. 167–177.

[12] Timothy N. Judkins et al., “Objective evaluation of expert and novice performance during robotic surgical training tasks,” Surg. endoscopy, vol. 23, no. 3, pp. 590–597, 2009.

[13] K. Liang et al., “Motion control skill assessment based on kinematic analysis of robotic end-effector movements,” Int. J. Med. Robot. Comput.Assist. Surg., vol. 14, no. 1, 2018, Art. no. e1845.

[14] A. Ghasemloonia et al., “Surgical skill assessment using motion quality and smoothness,” J. Surg. Educ., vol. 74, no. 2, pp. 295–305, 2017.

[15] A. Abid et al., “Exploring patterns enriched in a dataset with contrastive principal component analysis,” Nature Commun., vol. 9, no. 1, pp. 1–7,2018.

[16] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9.11 (2008).

[17] Chen, Carolyn, et al. "Crowd-sourced assessment of technical skills: a novel method to evaluate surgical performance." Journal of surgical research 187.1 (2014): 65-71.

[18] Ershad, Marzieh, et al. "Meaningful assessment of surgical expertise: Semantic labeling with data and crowds." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part I 19. Springer International Publishing, 2016.

[19] Jolliffe, Ian T., and Jorge Cadima. "Principal component analysis: a review and recent developments." Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016): 20150202.

[20] COX, AMA, and FT COX. "CHAP. MULTIDIMENSIONAL SCALING. INHandbook of Data Visualization." (2008): 315-347.

[21] Roweis, Sam T., and Lawrence K. Saul. "Nonlinear dimensionality reduction by locally linear embedding." science 290.5500 (2000): 2323-2326.

[22] Tenenbaum, Joshua B., Vin de Silva, and John C. Langford. "A global geometric framework for nonlinear dimensionality reduction." science 290.5500 (2000): 2319-2323.

[23] Bengio, Yoshua, et al. "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering." Advances in neural information processing systems 16 (2003).

[24] Soleymani, Abed, Xingyu Li, and Mahdi Tavakoli. "A Domain-Adapted Machine Learning Approach for Visual Evaluation and Interpretation of Robot-Assisted Surgery Skills." IEEE Robotics and Automation Letters 7.3 (2022): 8202-8208.