

# Machine Learning Demonstrates High Accuracy for Disease Diagnosis and Prognosis in Plastic Surgery

Angelos Mantelakis, BSc (Hons),  
MBBS\*

Yannis Assael, BSc (Hons), MSc,  
MRes, DPhil†

Parviz Sorooshian, MRES, BMBS,  
MRCS (Eng)‡

Ankur Khajuria, BSc (Hons),  
MBBS (Dist.), PhD, FHEA,  
FRSPH, MRCS (Eng),  
MAcadMEd. MSc (Oxon.)§¶

**Introduction:** Machine learning (ML) is a set of models and methods that can detect patterns in vast amounts of data and use this information to perform various kinds of decision-making under uncertain conditions. This review explores the current role of this technology in plastic surgery by outlining the applications in clinical practice, diagnostic and prognostic accuracies, and proposed future direction for clinical applications and research.

**Methods:** EMBASE, MEDLINE, CENTRAL and ClinicalTrials.gov were searched from 1990 to 2020. Any clinical studies (including case reports) which present the diagnostic and prognostic accuracies of machine learning models in the clinical setting of plastic surgery were included. Data collected were clinical indication, model utilised, reported accuracies, and comparison with clinical evaluation.

**Results:** The database identified 1181 articles, of which 51 articles were included in this review. The clinical utility of these algorithms was to assist clinicians in diagnosis prediction (n=22), outcome prediction (n=21) and pre-operative planning (n=8). The mean accuracy is 88.80%, 86.11% and 80.28% respectively. The most commonly used models were neural networks (n=31), support vector machines (n=13), decision trees/random forests (n=10) and logistic regression (n=9).

**Conclusions:** ML has demonstrated high accuracies in diagnosis and prognostication of burn patients, congenital or acquired facial deformities, and in cosmetic surgery. There are no studies comparing ML to clinician's performance. Future research can be enhanced using larger datasets or utilising data augmentation, employing novel deep learning models, and applying these to other subspecialties of plastic surgery. (*Plast Reconstr Surg Glob Open* 2021;9:e3638; doi: [10.1097/GOX.0000000000003638](https://doi.org/10.1097/GOX.0000000000003638); Published online 24 June 2021.)

## INTRODUCTION

An expanding population in the United States has resulted in an increasing demand for plastic surgery services, which, coupled with static number of residents and increasing number of retiring surgeons, is increasing the pressure for the delivery of high-quality care.<sup>1</sup> It is now estimated that there is a workforce shortage of 800 attending physicians in the United States, reducing the availability of care.<sup>1</sup> Artificial Intelligence (AI) could have a major impact on addressing challenges that healthcare

systems face. Digital technologies are predicted to affect more than 80% of the healthcare workforce in the next 2 decades, changing the way physicians practice medicine and meeting the increasing demand for services.<sup>2</sup> AI can help drive this change by automating repetitive tasks to free up time from clinicians, improving the diagnostic accuracy of diseases and predicting patient outcomes.<sup>2</sup>

Machine learning (ML), a subfield of AI, is a set of models able to learn from past cases (data) to make future predictions. A wide variety of such algorithms are in use today, such as in the automated, individualized suggestions generated during a Google Search, based on ones' previous searches. These models can be classified into two broad categories: supervised learning and unsupervised learning. The difference between these two categories of learning models lies in the existence of labeled data. In supervised learning, the models are trained

From the \*Department of Surgery and Cancer, Imperial College London, UK; †University of Oxford, UK; ‡Queen Victoria Hospital, UK; §Kellogg College, University of Oxford; and ¶Department of Surgery and Cancer, Imperial College London, UK.

Received for publication November 4, 2020; accepted April 14, 2021.

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: [10.1097/GOX.0000000000003638](https://doi.org/10.1097/GOX.0000000000003638)

**Disclosure:** All the authors have no financial interest to declare in relation to the content of this article.

Related Digital Media are available in the full-text version of the article on [www.PRSGlobalOpen.com](http://www.PRSGlobalOpen.com).

using examples of data with known labels, labeled data, and after training, they aim to predict outcomes utilizing new data.<sup>3,4</sup> This function has been utilized in healthcare to assist in both making a diagnosis and for disease outcome prediction. Authors have utilized supervised learning to successfully classify whether a skin lesion is benign (eg, benign nevi) or malignant (malignant melanoma), outperforming the accuracy of 21 board-certified dermatologists (accuracy 72% versus 66%,  $P < 0.05$ ).<sup>5</sup> Similarly, supervised learning has also been utilized in predicting the risk of developing a condition such as breast cancer based on epidemiological data, and the risk of recurrence after treatment.<sup>6,7</sup>

In contrast, unsupervised learning models are trained using unlabeled data, and after training, aim to discover underlying groupings or patterns from the data themselves.<sup>3,8</sup> These algorithms can be particularly useful in identifying previously unknown patterns in vast amounts of unprocessed data, which may then be used in clinical practice. Examples include novel classification of diseases into various subtypes and identifying subgroups of patients with increased risk of certain conditions based on various characteristics (for example, their genome).<sup>9,10</sup>

In addition to meeting demand for plastic surgery services, this technology has the potential to revolutionize how plastic surgery is practiced and enhance surgeon's diagnosis prediction, preoperative planning, and outcome prediction, leading to improved patient care. In burn surgery, even the most experienced surgeons have a clinical estimation of 64%–76% accuracy in the diagnosis of burn depth.<sup>11,12</sup> ML models may outperform this, achieving correct burn depth identification from 2D photographs up to 87%, potentially leading to more appropriate clinical management at presentation.<sup>13</sup> Further, in the prognostication of whether a burn injury will heal within 14 days of presentation, ML models have demonstrated an accuracy of 86%, again surpassing the accuracy of prognostication by clinicians.<sup>4</sup> In the field of microsurgery, postoperative monitoring via 2D image analysis achieves a 95% accuracy in classifying a flap as normal, presence of venous obstruction, or presence of arterial occlusion, leading to potential early identification of flap failure and increased salvage rates.<sup>4</sup> However, the evidence of applications of ML is abstract, with no systematic reviews that summarize the clinical accuracy of such models in practice. This could act as a starting point of developing clinical practice guidelines and to guide future research.<sup>14–17</sup> The aim of this study was to systematically synthesize and report the current literature in the clinical applications of ML in plastic surgery.

## METHODS

### Search Strategy

The protocol for this systematic review was registered with PROSPERO international prospective registration of systematic reviews registration number: CRD42019140924. The full protocol was published a priori, and there were no deviations from the original protocol.<sup>18</sup> This systematic

review was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.<sup>19</sup>

A systematic literature search was performed in MEDLINE (OVID SP), EMBASE (OVID SP), CENTRAL, and ClinicalTrials.gov databases to identify relevant studies for review. The reference lists of all included studies were also screened, and relevant studies were included in the search. Lastly, manual searches of bibliographies, citations, and related articles (Pubmed function) were also performed to identify missed relevant studies. Medical Subject Headings (MeSH) terms were used in combination with free text to construct our search strategy. A sample search strategy used in MEDLINE (OVID SP) is shown in Table 1.<sup>20–70</sup>

### Selection Criteria

All eligible studies between January 1990 and June 2020 were included in this review. We included any primary studies (including case reports) that present clinical data on the application of ML in plastic surgery. Only articles in the English language were included. Our exclusion criteria included descriptions of ML in plastic surgery without clinical data, review articles, conference abstracts, animal studies, and articles pertaining to the use of ML outside the remit of the specialty (as defined by the Intercollegiate Surgical Curriculum Program in Plastic Surgery).

After the library preparation, two independent reviewers (AM and PS) screened the search results for inclusion based on the title and abstracts. Subsequently, a full-text review was performed independently by the same two researchers (AM and PS) for all included studies. At each step, any discrepancy of opinion was resolved with consensus, and if not resolved, was referred to a third reviewer (AK). If any doubt remained, the article proceeded to the next step of the review. The search results of all included articles, abstracts, full-text articles, and records of the reviewers' decisions, including reasons for exclusion, were recorded.

### Outcome Measures

The primary outcome was the ML algorithm statistical accuracy in performing a prespecified clinical task (eg, prediction of a clinical diagnosis or postoperative outcome). Secondary outcomes include the reported

**Table 1. Example Search Strategy Used for MEDLINE<sup>20–70</sup>**

1	("deep learning" OR "artificial intelligence" OR "machine learning" OR "decision trees" OR "random forests" OR SVM OR "support vector machine")
2	exp "NEURAL NETWORKS (COMPUTER)"/ OR exp "DEEP LEARNING"/
3	exp "ARTIFICIAL INTELLIGENCE"/
4	(1 OR 2 OR 3)
5	(microsurgery OR (surgery AND (plastic OR reconstructive OR esthetic OR aesthetic OR burns OR hand OR craniofacial OR "peripheral nerve")))
6	exp "SURGERY, PLASTIC"/ OR exp "RECONSTRUCTIVE SURGICAL PROCEDURES"/
7	(5 OR 6)
8	(4 AND 7)

specificity, sensitivity, area under the curve, and technical characteristics of the algorithms.

#### Data Extraction and Analysis

The data from all full-text articles accepted for the final analysis were independently retrieved by AM and PS, using a standardized data extraction form. Any disagreements were resolved by discussion or referred to the third researcher (AK). The following data (where available) were extracted:

- a) Study details (year of publication, country), patient demographics, study setting, clinical condition examined.
- b) ML algorithm characteristics (intended function, whether the model was supervised or unsupervised, function via classification or outcome prediction, usage of real or synthetic data, and which type of ML model was used)
- c) Primary and secondary outcomes, as above.

Statistical meta-analysis could not be performed because of the heterogeneity of the studies in the conditions examined and software models utilized. Instead, a narrative review was performed, with a subgroup analysis of the mean accuracy of the models, calculated by measuring the number of correct predictions over the total predictions made.

The subgroup analyses are based on the model function (diagnosis prediction, preoperative planning and outcome prediction) and type of models (NNs, SVMs, decision tree/random forest, and linear regression). This subgroup classification was utilized based on the objectives set for AI models in clinical practice by NHS England.<sup>2</sup>

#### Quality Assessment

The quality of the included studies was assessed based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2), performed by two independent reviewers (AM and PS).<sup>71</sup> There were no disagreements between the authors. The QUADAS-2 tool allows for risk of bias assessment and applicability concern assessment of primary diagnostic accuracy studies. Risk of bias was assessed based on the patient selection, index test (in this review, this is the ML algorithm), reference standard (comparator), and flow and timing. Concerns regarding applicability were assessed on the first three terms alone.

## RESULTS

#### Literature Search Results

From a total of 1536 studies, after removal of duplicates, 1181 articles were eligible for a title and abstract review. Of these, 1074 articles did not meet the inclusion criteria and were excluded. Following full-text review of the remaining 107 articles, 56 articles were excluded because the inclusion criteria were not met. A total of 51 articles were included and formed the basis of this

systematic review (Fig. 1). Details of the included studies are summarized in Table 2.<sup>20–70</sup>

#### Breakdown of the Applications of ML Models in Diagnosis Prediction, Outcome Prediction, and Preoperative Planning

In total, 51 studies were included in the review, which evaluated the accuracy of 103 ML algorithms. Of these, 27 were on burns surgery and 24 on general reconstructive surgery. The publication years ranged from 1996 to 2020, with 25 studies published in the past year alone (2019–2020). The clinical utility of these algorithms was to assist clinicians in diagnosis prediction (n = 22), outcome prediction (n = 21), and preoperative planning (n = 8).

In diagnosis prediction, algorithms were created to assist in automated burn depth diagnosis from 2D photography (n = 9) and total burn surface area (n = 1), automated diagnosis of craniosynostosis (n = 5), wound identification in 2D photography (n = 2), diagnosis and severity assessment of facial palsy (n = 1), diagnosis of congenital auricular deformities (n = 1), identification of emotional responses to plastic surgery on Twitter (n = 1), automated age estimation after rhinoplasty (n = 1), and identifying the correct answer to frequently asked questions (n = 1).

In outcome prediction, the ML algorithms created predicted mortality in burn patients (n = 5), the occurrence of AKI in burn and trauma patients (n = 4), occurrence of postoperative complications in breast and head and neck free flap reconstruction (n = 3), concentration and response of aminoglycosides in burn patients (n = 2), postoperative faces after oculoplastic and craniosynostosis surgery (n = 2), burn healing time (n = 1), mortality in patients with necrotizing soft tissue infection (n = 1), delay in radiotherapy following cancer excision (n = 1), posttraumatic stress disorder following burns (n = 1), and factors predicting the occurrence of burns in the pediatric population (n = 1).

In preoperative planning, ML was used to predict which wounds will need grafting (n = 2), which patients will need orthognathic or cleft palate operations (n = 2), planning orthognathic and mandibular resections (n = 2), predicting open wound size (n = 1), and complexion of reconstruction following head and neck cancer excision (n = 2).

#### ML Models Demonstrate High Accuracy, Sensitivity, and Specificity That May Enhance Clinical Decision-making

The 51 studies evaluated 103 ML algorithms (Table 2). The pooled mean of accuracy of ML algorithms was 86.84% (range 60.00–100%). The pooled mean sensitivity and specificity is 81.88% (range 5.00–99.30%) and 86.38% (range 60.00–100%), respectively, as reported in 39 models.

A subgroup analysis was performed based on the clinical utility of the algorithms. For diagnosis prediction, the pooled accuracy, sensitivity, and specificity of ML algorithms was 88.80% (range 66.20–97.60%), 90.62% (range 75.80–97.90%), and 86.81% (range 60.00–99.60%). In outcome prediction, this was 86.11% (range

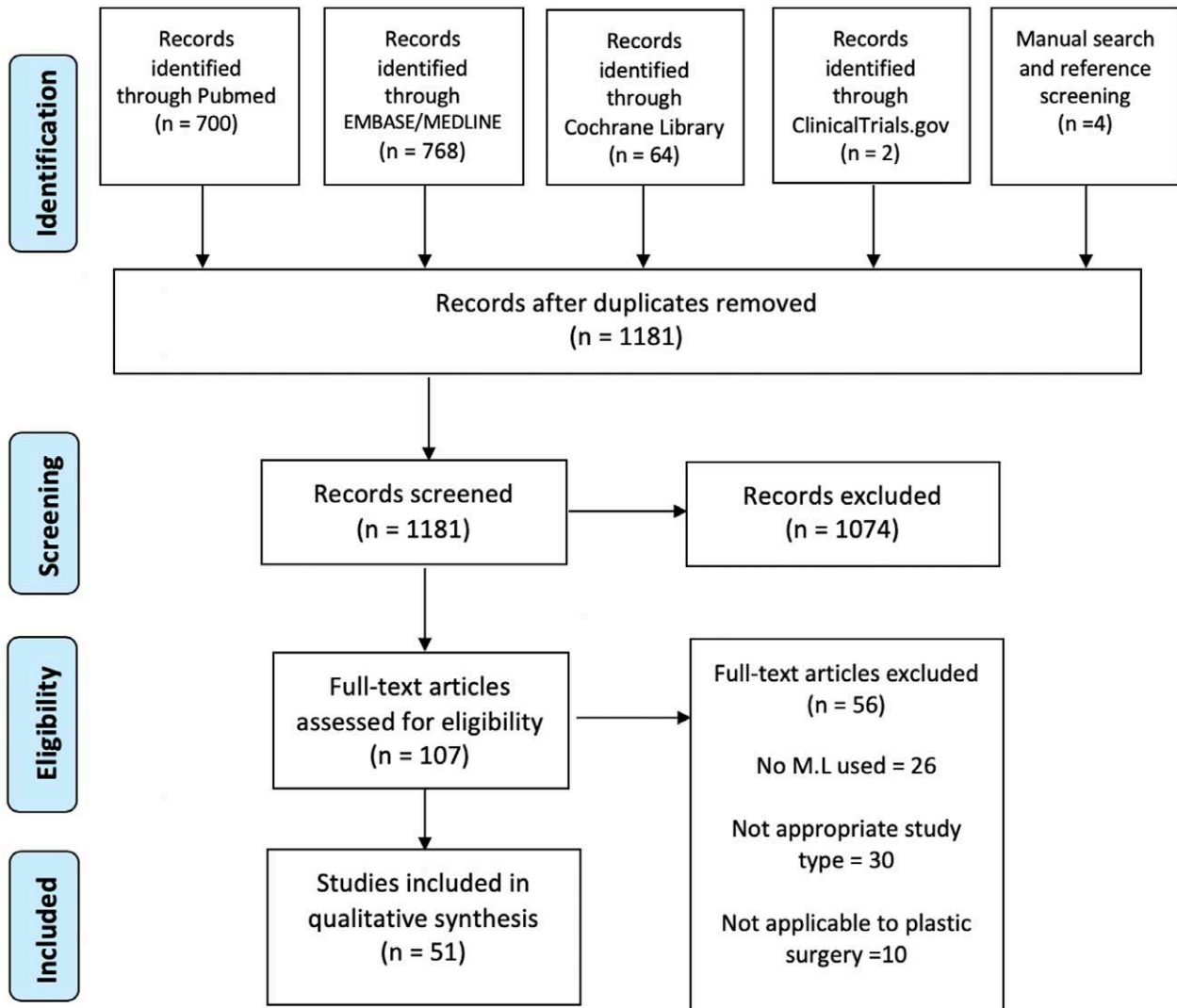


Fig. 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

66.20–97.60%), 69.67% (range 5.00–100%), and 85.94% (range 60.00–100%), respectively. In preoperative planning, two studies reported the accuracy, sensitivity, and specificity, which were 80.28% (range 77.30–83.80%), 98.00% (range 97.00–99.00%), and 67.05% (range 60.00–74.10%).

A second subgroup analysis on the reported accuracy was performed based on the type of model utilized. The mean accuracy for NNs was 88.25% (range 73.80–100%), SVMs 88.02% (range 67.20–100%), decision trees/random forest 78.75% (range 60.00–96.12%), and linear regression 76.85% (range 66.40–95.00%).

#### Breakdown and Analysis of the Supervised and Unsupervised ML Models Utilized

Supervised ML was utilized in 50 of the included studies and unsupervised learning in three studies (two studies employed both supervised and unsupervised learning). The supervised ML algorithms identified are summarized in Table 3. The most commonly used ones were NNs

( $n = 34$ ), SVMs ( $n = 13$ ), decision trees/random forests (DT/RF,  $n = 10$ ), and LR ( $n = 9$ ). The unsupervised ML models utilized were K-means clustering, a shapeswork software with principal component analysis and the algorithm was not reported in one study.

#### Lack of Data Augmentation and Validation during Training

Data augmentation is often used in small datasets, to artificially create more data samples and increase the effective dataset size, and as a result the statistical performance of a model. Data augmentation was used in only six of the 51 included studies. The remaining articles relied only on real data. For diagnostic predictions, the majority of studies utilized 2D photographs ( $n = 15$ ) and CT scans ( $n = 4$ ). For clinical outcome prediction, patient risk factors and laboratory measurements on admission was utilized in most models ( $n = 17$ ). In preoperative planning, CT scans ( $n = 3$ ) and 2D photographs ( $n = 2$ ) comprised the majority of inputs utilized.

**Table 2. Primary Outcomes of Accuracy, Sensitivity, and Specificity for Reconstructive and Burns Surgery**

Study	Author, Year	Function	Model	Accuracy	Sensitivity	Specificity	AUC
1	Abubakar et al, 2020 <sup>20</sup>	DP	CNN	White: 99.3% Afro-Caribbean: 97.1%	NR	NR	NR
2	Chauhan J et al, 2020 <sup>21</sup>	DP	BPBSAM (CNN + SVM)	91.70%	NR	NR	NR
3	Desbois et al, 2020 <sup>22</sup>	DP	DNN with 3 measures	91.98%	NA	NA	NR
			DNN with 4 measures	92.45%	NA	NA	NR
			Boost with 3 measures	97.89%	NA	NA	NR
			Boost with 4 measures	98.08%	NA	NA	NR
			avNN with 3 measures	97.45%	NA	NA	NR
			avNN with 4 measures	98.30%	NA	NA	NR
4	Rashidi et al, 2020 <sup>23</sup>	OP	DNN	100%	92%	93%	0.880
			LR	95%	91%	90%	0.940
			SVM	98%	NR	NR	0.780
			RF	93%	NR	NR	1.000
			k-NN	98%	91%	82%	0.960
5	Bhalodia et al, 2020 <sup>24</sup>	DP	Shapeswork software with PCA	NR	NR	NR	NR
6	Guarin et al, 2020 <sup>25</sup>	DP	NR	NR	NR	NR	NR
7	Formeister et al, 2020 <sup>26</sup>	OP	Gradient Boosted Decision Tree	60.00%	62.00%	60.00%	NR
8	Boczar et al, 2020 <sup>27</sup>	Intervention	IBM Watson	92.30%	NR	NR	NR
9	O'Neil et al, 2020 <sup>28</sup>	OP	Decision Tree	NR	5.00%	86.80%	0.672
10	Yoo et al, 2020 <sup>29</sup>	OP	Deep Learning (Generative adversarial network- GAN)	NR	NR	NR	NR
			Pix2pix	NR	NR	NR	NR
			Lightweight CycleGAN	NR	NR	NR	NR
		DP	Deep Learning + No data augmentation	74.20%	75.80%	72.70%	0.824
			Deep Learning + Std data augmentation	83.33%	78.80%	87.90%	0.872
			Deep Learning + GAN data augmentation	90.90%	87.80%	93.90%	0.957
11	Angullia et al, 2020 <sup>30</sup>	OP	Least squares radial basis function	NA	NA	NA	NA
12	Eguia et al, 2020 <sup>31</sup>	OP	Decision Tree	NA	NA	NA	0.690
			Stepwise Logistic Regression	NA	NA	NA	0.800
			LR	NA	NA	NA	0.830
			k-NN	NA	NA	NA	0.840
13	Ohura et al, 2019 <sup>32</sup>	DP	SegNet	97.60%	90.90%	98.20%	0.994
			LinkNet	97.20%	98.90%	98.90%	0.987
			U-Net	98.80%	99.30%	99.30%	0.997
			Unet_VGG16	98.90%	99.20%	99.20%	0.998
14	Porras et al, 2019 <sup>33</sup>	DP	SVM	95.30%	94.70%	96%	NR
15	Knoops et al, 2019 <sup>34</sup>	DP	SVM	95.40%	95.50%	95.20%	NR
		OP	LRRRLARLASSO	NR	NR	NR	NR
16	Hallac et al, 2019 <sup>35</sup>	DP	Pretrained Google-Net	94.10%	97.80%	86%	NR
17	Levites et al, 2019 <sup>36</sup>	DP	Text-based emotion analysis	NR	NR	NR	NR
18	Shew et al, 2019 <sup>37</sup>	OP	2-class Decision Forest	64.40%	NR	NR	NR
19	Dorfman et al, 2019 <sup>38</sup>	DP	Neural Nets	NR	NR	NR	NR
20	Qiu et al, 2019 <sup>39</sup>	PP	U-Net CNN	NR	NR	NR	NR
21	Aghei et al, 2019 <sup>40</sup>	OP	ANN-MLP	73.3%	76.20%	70.2	0.762
			SVM	67.20%	66.10%	68.40%	0.731
			RF	67.20%	61%	73.70%	0.751
			LR (FS)	67.20%	61%	73.70%	0.711
			LR (BS)	66.40%	64.40%	67.70%	0.718
22	Cirillo et al, 2019 <sup>41</sup>	DP	VGG-16	77.53%	NR	NR	NR
			Google-Net	73.80%	NR	NR	NR
			Res-Net 50	77.79%	NR	NR	NR
			Res-Net 101 without data aug	90.54%	74.35%	94.25%	NR
			Res-Net 101 with data aug	82.72%	NR	NR	NR
23	Tran et al, 2019 <sup>42</sup>	OP	k-NN with k = 1-6 or 8-20	100%	NA	NA	NR
24	Yadav et al, 2019 <sup>43</sup>	DP	MDS modeling	80%	97.00%	60.00%	NR
			SVM	82.43%	87.80%	83.33%	NR
25	Jiao et al, 2019 <sup>44</sup>	DP	R101A CNN	82.04%	NA	NA	NR
			IV2RA CNN	83.02%	NA	NA	NR
			R101FA CNN	84.51%	NA	NA	NR
26	Liu et al, 2018 <sup>45</sup>	PP	Least Squares Regression	NR	NR	NR	NR
			Decision tree	NR	NR	NR	NR
			Sigmoid Neural Nets	NR	NR	NR	NR
			Hyperbolic Tangent Neural Net	NR	NR	NR	NR
			Combined Model (Tree +NN)	NR	NR	NR	NR
27	Martinez-Jemenez et al, 2018 <sup>46</sup>	OP	Recurrent Partitioning	85.35%	NR	NR	NR
			Random Forest				
28	Su et al, 2018 <sup>47</sup>	OP	Random Forest	NA	NA	NA	NR
29	Tang et al, 2018 <sup>48</sup>	OP	L.R	80.50%	84.40%	77.70%	0.875
			XGBoost	85.40%	82.0%	89.7%	0.920
30	Cobb et al, 2018 <sup>49</sup>	OP	Random Forest	NA	NA	NA	NR
			Stochastic Gradient Boosting				NR
31	Cho MJ et al, 2018 <sup>50</sup>	DP	K-means	96%	NR	NR	NR
32	Kuo et al, 2018 <sup>51</sup>	OP	MLR	72.70%	22.10%	93.30%	NR
33	Tan et al, 2017 <sup>52</sup>	PP	NR	NR	NR	NR	NR
34	Huang et al, 2016 <sup>53</sup>	OP	SVM	100%	NA	NA	NR
35	Park et al, 2015 <sup>54</sup>	PP	Feature wrapping	77.30%	99%	74.10%	NR

(Continued)

**Table 2. (Continued)**

Study	Author, Year	Function	Model	Accuracy	Sensitivity	Specificity	AUC
36	Serrano et al, 2015 <sup>55</sup>	PP	SVM	79.73	97%	60%	NR
37	Mukherjee et al, 2014 <sup>56</sup>	DP	SVM with 3rd polynomial kernel	86.13%	NA	NA	NR
38	Mendoza et al, 2014 <sup>57</sup>	DP	Bayesian classifier	81.15%	NA	NA	NR
			LDA	95.70%	97.90%	99.60%	NR
			Random Forest	87.90%	NR	NR	NR
39	Acha et al, 2013 <sup>58</sup>	DP	SVM	90.80%	NR	NR	NR
			k-NN	66.2%	NR	NR	NR
			SVM	75.7%	NR	NR	NR
40	Schneider et al, 2012 <sup>59</sup>	OP	k-NN	83.8%	NR	NR	NR
			SVM	82.4%	NR	NR	NR
			CART Decision Tree with Gini splitting function	73.30%	NA	NA	NR
41	Patil et al, 2009 <sup>60</sup>	OP	Bayesian classifier	97.78%	100%	95.50%	0.978
			Decision Tree	96.12%	96.60%	95.51%	0.961
			SVM	96.12%	98.60%	93.26%	0.961
			Back propagation	95%	96.71%	93.26%	0.949
42	Yamamura et al, 2008 <sup>61</sup>	OP	ANN	100%	NA	NA	NR
			LR	72%	NA	NA	NR
43	Correa et al, 2008 <sup>62</sup>	DP	SVM	95.05%	NR	NR	NR
44	Acha et al, 2005 <sup>63</sup>	DP	Fuzzy-ArtMap Neural Network	82.26%	83.01%	NA	NR
45	Yeong et al, 2005 <sup>64</sup>	OP	ANN	86%	75%	97%	NR
46	Serrano et al, 2005 <sup>65</sup>	DP	Fuzzy-ArtMap Neural Network	88.57%	83.01%	NA	NR
47	Yamamura et al, 2004 <sup>66</sup>	OP	ANN	100%	100%	100%	NR
			LR	80%	66.70%	85.70%	NR
			ANN with leave-one-out crossvalidation	86.60%	66.70%	95.20%	NR
48	Acha et al, 2003 <sup>67</sup>	OP	Fuzzy-ArtMap Neural Network	82.60%	NR	NR	NR
49	Estahbanati et al, 2002 <sup>68</sup>	OP	ANN	90%	80%	NA	NR
			Shallow Neural Net	NA	NA	NA	NR
50	Hsu et al, 2000 <sup>69</sup>	PP	Feed forward, back propagation error adjustment model	98%	NA	NA	NR
51	Fyre et al, 1996 <sup>70</sup>	OP	Feed forward, back propagation error adjustment model	77%	NA	NA	NR

ADTree, alternating decision tree; AUC, area under the curve; CNN, convoluted NNs; DNN, deep neural network; DP, diagnosis prediction; k-NN, k-nearest neighbor; LASSO, least absolute shrinkage and selection operator; LDA, liner discriminant analysis; MLR, multiple logistic regression; NA, not applicable; NB classifier, Naive Bayes classifier; NR, not reported; OP, outcome prediction; PP, preoperative planning; RF, random forest .

Training ML models requires splitting the data set in training, validation, and test sets, where the validation set is used for hyperparameter tuning during training to prevent “overfitting” of the model to the given data. Only 10 of the 35 studies utilized a validation set during training. In total, 35 studies report their data training and testing splits, with an 80%–20% split between the training and testing set being the most common methodology presented (n = 9).

In terms of output, ML algorithms functioned primarily via classification in 45 studies and via regression in six studies. Classification was utilized for the allocation of a new subject to a specific outcome (for example, burn patient needing a grafting versus healing via secondary intention). Regression was used in studies aiming to recreate a prediction of a postoperative outcome (postoperative CT scan, postoperative 2D photograph, and predicted wound size).

**Risk of Bias Assessment**

The risk of bias was assessed via the QUADAS-2 tool for risk of bias assessment and concerns over applicability (Fig. 2). The majority of studies had an unclear risk of bias (RoB) in the patient selection (n = 20) and index test domains (n = 24). Most had a low RoB by the reference standard (n = 39) and flow and timing domains (n = 35). For applicability concern, more than half of the studies had a low risk of RoB regarding the patient selection, index test, and reference standard domains (n = 32, n = 33, and n = 38 respectively).

**DISCUSSION**

This is the first systematic review focusing on the application of ML in plastic surgery, adding to previous reviews on AI in the specialty.<sup>72</sup> After careful selection of studies that demonstrated the clinical application of these algorithms, we identified 51 articles describing the application of 103 ML algorithms. In our review, the mean accuracy for diagnosis prediction, outcome prediction, and preoperative planning was 88.80%, 86.11%, and 80.28%, respectively. The model with the highest mean accuracy was NNs (88.25%), followed by SVMs (88.02%), decision trees/random forest (78.75%), and linear regression (76.85%).

Similar findings have been reported in systematic reviews of other surgical specialties. In orthopedic surgery and neurosurgery, the most common models utilized have been Neural Networks (NNs), followed by support vector machines (SVMs) and logistic regression (LR).<sup>3,73</sup> Outcome prediction of ML models in these specialties ranged from 70% to 97%, which is in line with the findings of this report.<sup>8,72</sup> Nonsurgical specialties have also utilized NNs and SVMs the most frequent, with accuracies approaching 96% depending on the specialty and model intent.<sup>74,75</sup> The reason behind this preference is potentially that NN, SVM, and DT most closely resemble the cognition behind clinical judgment, where clinicians aim to derive outcome classifications based on multiple, nonlinear inputs. In plastic surgery, ML demonstrated potentially superior accuracy in diagnosis and outcome prediction when compared with clinician judgment. In burn surgery,

**Table 3. Technical Characteristics of ML Algorithms Utilized in Burns and Reconstructive Surgery**

Study No.	Author	Function	Purpose	Input	Output	Supervised or Unsupervised	Modeling (Classification or Regression)	Real or Synthetic Data		
								Training	Test	
1	Abubakar et al, 2020 <sup>30</sup>	DP	Differentiate healthy versus burned skin in both white and black skin	2D photographs	Differentiate healthy versus burned skin in both white and black skin	Supervised	Classification	Data augmentation	80% NA	20%
2	Chauthan, J et al, 2020 <sup>31</sup>	DP	Diagnose depth of burns	2D photographs	Differentiate body part + severity of burn	Supervised	Classification	Data augmentation	80%	20%
3	Desbois et al, 2020 <sup>32</sup>	DP	Automated assessment of TBSA	Anthropometric measurements	Automated assessment of TBSA	Supervised	Regression	Real data	80%	NA
4	Rashidi et al, 2020 <sup>33</sup>	OP	Prediction of AKI in burn and trauma patients	Renal injury biomarkers and urine output	Prediction of AKI in burn and trauma patients	Supervised	Classification	Real data	59%	NA
5	Bhalodia et al, 2020 <sup>34</sup>	DP	Measuring severity of craniosynostosis	CT images	Measuring severity of craniosynostosis	Unsupervised	NA	Real data	NR	NR
6	Guarin et al, 2020 <sup>35</sup>	DP	Diagnosis and severity assessment of facial palsy	2D photographs	Automatic localization of 68 facial features in healthy and patients	Unsupervised	N/A	Real data	90%	5%
7	Formeister et al, 2020 <sup>36</sup>	OP	Predicting any type of complications following free flap reconstruction	14 patient characteristics	Prediction of complications in microvascular free flaps	Supervised	Classification	Real data	80%	20%
8	Boczar et al, 2020 <sup>37</sup>	DP	Answering frequently asked questions	Participant question	Correct answer to FAQs	Supervised	Classification	Real data	NR	NR
9	O'Neil et al, 2020 <sup>38</sup>	OP	Predicting flap failure in microvascular breast free flap reconstruction	7 patient characteristics	Flap failure (yes/no)	Supervised	Classification	Data augmentation	50%–70%	NA
10	Yoo et al, 2020 <sup>39</sup>	OP	Postoperative appearance following oculoplastic surgery for thyroid-associated ophthalmopathy	Preoperative photograph	Postoperative photograph	Supervised	Regression	Data augmentation	NR	NR
11	Angullia et al, 2020 <sup>30</sup>	OP	Prediction of changes in face shape from craniosynostosis surgery	High resolution CT	Predict changes in face shape from craniosynostosis surgery	Supervised	Regression	Real data	NR	NR
12	Eguia et al, 2019 <sup>31</sup>	OP	Prediction of in-hospital mortality in patients with necrotizing skin and soft tissue infection	Patient demographics, co-morbidities, and hospital characteristics (73 parameters in total)	Prediction of in-hospital mortality in patients with necrotizing skin and soft tissue infection	Supervised	Classification	Real data	80%	20%
13	Ohura et al, 2019 <sup>32</sup>	DP	Diagnosis of wound ulcer	2D photographs	Differentiation of healthy tissue from ulcer region	Supervised	Classification	Real data	90%	10%
14	Porras et al, 2019 <sup>33</sup>	DP	Diagnosis of craniosynostosis from 3D photographs	3D photographs	Diagnosis of craniosynostosis from 3D photographs	Supervised	Classification	Real data	NR	NR
15	Knoops et al, 2019 <sup>34</sup>	PP	Orthognathic surgery	CT	Need for orthognathic surgery (yes/no)	Supervised	Classification	Real data	80%	20%
16	Hallac et al, 2019 <sup>35</sup>	DP	Diagnosis of congenital auricular deformities	2D photographs	Identify presence of congenital auricular deformities (yes/no)	Supervised	Classification	Real data	NR	NR
17	Levites et al, 2019 <sup>36</sup>	DP	Identify emotional responses to plastic surgery	Twitter key words	Analyze emotional responses to plastic surgery procedures	Supervised	Classification	Real data	60%	20%
18	Shew et al, 2019 <sup>37</sup>	OP	Prediction of delay in radiotherapy	Variable inpatient patient data	Prediction of delay of radiotherapy (more or less than 50 days to treatment)	Supervised	Classification	Real data	NR	NR

(Continued)

**Table 3. (Continued)**

Study No.	Author	Function	Purpose	Input	Output	Supervised or Unsupervised	Modeling (Classification or Regression)	Real or Synthetic Data	Training		
									Training	Validation	Test
19	Dorfman et al, 2019 <sup>38</sup>	DP	Identification of age perception following rhinoplasty	2D photographs	Automated age prediction	Supervised	Classification	Real data	NR	NR	NR
20	Qiu et al, 2019 <sup>39</sup>	PP	Plan mandibular resections	CT	Automated 3D mandibular segmentation preoperatively	Supervised	Regression	Real data	48%	7%	45%
21	Aghaei et al, 2019 <sup>40</sup>	OP	Elaboration of factors predicting pediatric burns	Various health, social, and demographic risk factors	Most important factors in predicting burn occurrence	Supervised	Classification	Real data	70%	NA	30%
22	Cirillo et al, 2019 <sup>41</sup>	DP	Diagnose depth of burns	2D photographs	Classification of burn depth	Supervised	Classification	Data augmentation	NR	NR	NR
23	Tran et al, 2019 <sup>42</sup>	OP	Prediction of AKI in burn and trauma patients	Renal injury biomarkers and urine output	Prediction of AKI in burn and trauma patients	Supervised	Classification	Real data	80%	NA	20%
24	Yadav et al, 2019 <sup>43</sup>	DP	Diagnose depth of burns	2D photographs	Classify burns by depth and surface area	Supervised	Classification	Real data	NR	NR	NR
25	Jiao et al, 2019 <sup>44</sup>	DP	Diagnose depth of burns	2D photographs	Classify burns by depth and surface area	Supervised	Classification	Real data	87%	NA	13%
26	Liu et al, 2018 <sup>45</sup>	PP	Explore whether ML can predict open wound size	Fluid resus volume and other patient factors	Predict open wound size	Supervised	Regression	Real data	90%	NA	10%
27	Marinez-Jimenez et al, 2018 <sup>46</sup>	PP	Predicting which wounds need grafting	Infrared thermography	Prediction of treatment modality required for burn wound	Supervised	Classification	Real data	61%	NA	39.00%
28	Su et al, 2018 <sup>47</sup>	OP	Prediction of PTSD & major depressive disorder in burn patients	Burn-related variables, empirically-derived risk factors from previous meta-analysis & theory-derived cognitive variables	Prediction of PTSD & major depressive disorder in burn patients	NR	NR	NR	NR	NR	NR
29	Tang et al, 2018 <sup>48</sup>	OP	Prediction of AKI in burn patients	Patient risk factors and laboratory measurements	Prediction of AKI in burn patients	Supervised	Classification	Real data	NR	NR	NR
30	Cobb et al, 2018 <sup>49</sup>	OP	Prediction of mortality of burn patients	Patient risk factors and laboratory measurements	Predict whether a patient would (1) live versus (2) die	Supervised	Classification	Real data	66%	NA	34%
31	Cho Mj et al, 2018 <sup>50</sup>	DP	Diagnosis of craniomynostosis	CT images	Automated differentiation of craniomynostosis from benign metopic ridge from CT	Unsupervised	Classification	Real data	NR	NR	NR
32	Kuo et al, 2018 <sup>51</sup>	OP	Predicting surgical site infection	Patient risk factors	Prediction of SSI (yes/no)	Supervised	Classification	Real data	70%	NA	30%
33	Tan et al, 2017 <sup>52</sup>	PP	Complexion of reconstruction following basal cell cancer excision	Patient risk factors	Prediction of intraoperative surgical complexity	Supervised	Classification	Real data	NR	NR	NR
34	Huang et al, 2016 <sup>53</sup>	OP	Prediction of mortality of burn patients	Patient risk factors and laboratory measurements	Prediction of whether a patient would (1) live versus (2) die	Supervised	Classification	Real data	21%	66%	13%

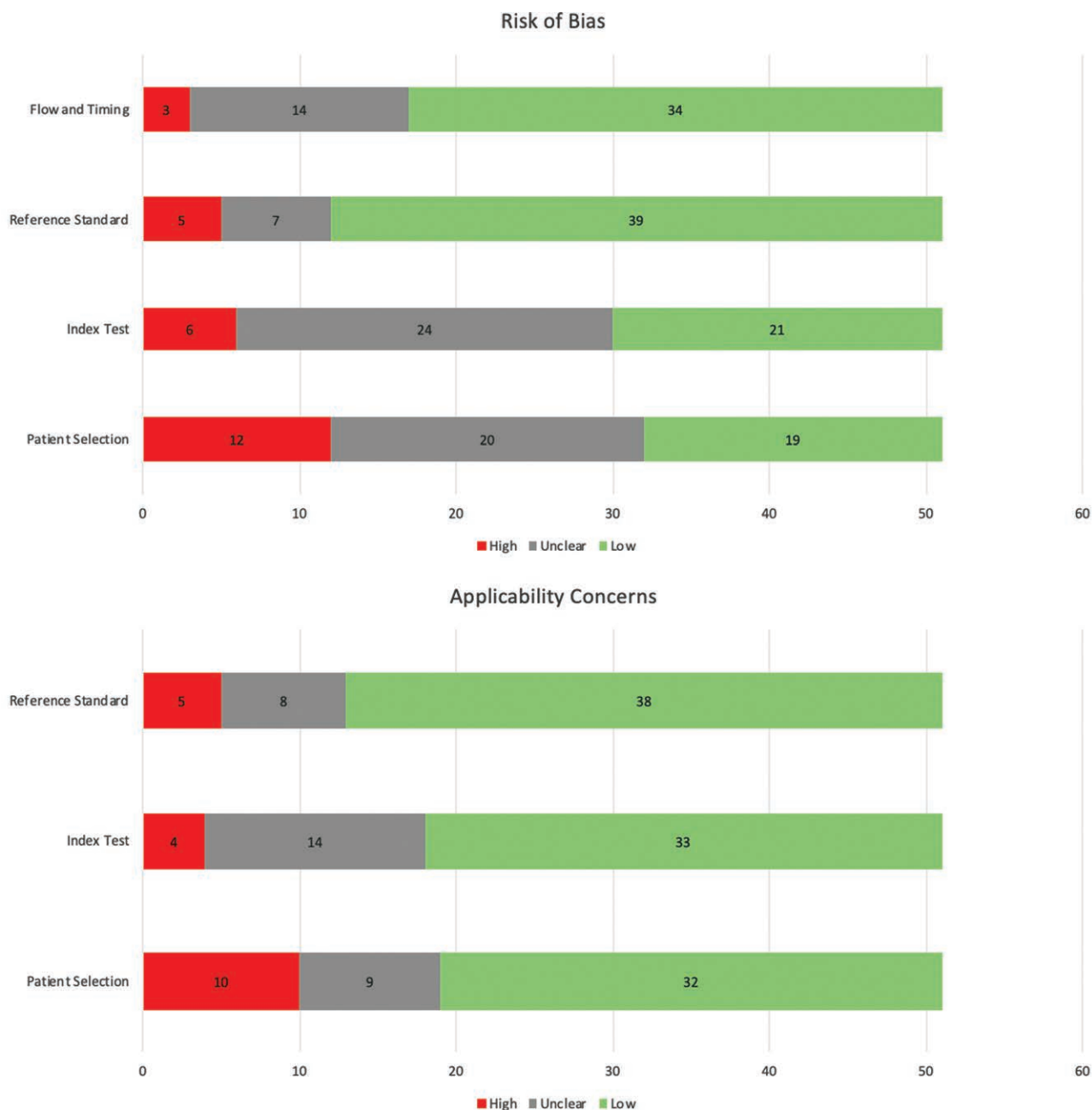
(Continued)



**Table 3. (Continued)**

Study No.	Author	Function	Purpose	Input	Output	Supervised or Unsupervised	Modeling (Classification or Regression)	Real or Synthetic Data	Training		
									Training	Validation	Test
35	Park et al, 2015 <sup>44</sup>	PP	Prediction of need for surgery in patients with cleft lip/palate	Lateral cephalograms	Prediction of need for surgery in patients with cleft lip/palate	Supervised	Classification	Real data	NR	NR	NR
36	Serrano et al, 2015 <sup>35</sup>	PP	Predicting which wounds need grafting	2D photographs	Predicting which wounds need grafting (yes/no)	Supervised	Classification	Real data	21%	NA	79%
37	Mukherjee et al, 2014 <sup>36</sup>	DP	Wound recognition and classification	2D photographs	Automated assessment of wound classification	Supervised	Classification	Real data	NR	NR	NR
38	Mendoza et al, 2014 <sup>37</sup>	DP	Diagnosis of craniosynostosis	CT images	Automated craniosynostosis diagnosis from CT	Supervised	Classification	Real data	NR	NR	NR
39	Acha et al, 2013 <sup>38</sup>	DP PP	Diagnose depth of burns Predicting which wounds need grafting	2D photographs 2D photographs	Classify burns by depth Predict whether a burn will need grafting	Supervised Supervised	Classification Classification	Real data Real data	21% 21%	NA NA	79% 79%
40	Schneider et al, 2012 <sup>39</sup>	OP	Prediction of AKI in burn patients	Patient risk factors and laboratory measurements	Prediction of AKI in burn patients	Supervised	Classification	Real data	71%	NA	29,00%
41	Patil et al, 2009 <sup>40</sup>	OP	Prediction of mortality of burn patients	Patient risk factors and laboratory measurements	Prediction of mortality in burn patients	Supervised	Classification	Real data	K-cross validation	K-cross validation	K-cross validation
42	Yamamura et al, 2008 <sup>41</sup>	OP	Prediction of response of aminoglycosides against MRSA infection in burn patients	Patient risk factors and laboratory measurements	Prediction of response of aminoglycosides against MRSA infection in burn patients	Supervised	Classification	Real data	K-cross validation	K-cross validation	K-cross validation
43	Ruiz-Correa et al, 2008 <sup>42</sup>	DP	Diagnosis of craniosynostosis	CT images	Classification of craniosynostosis	Supervised	Classification	Real data			
44	Acha et al, 2005 <sup>43</sup>	DP	Diagnose depth of burns	2D photographs	Automated assessment of burn wound depth	Supervised	Classification	Real data	56%	NA	44%
45	Yeong et al, 2005 <sup>44</sup>	OP	Prediction of burn healing time	Reflectance spectrometer measurements	Prediction of burn healing time	Supervised	Classification	Real data	NR	NR	NR
46	Serrano et al, 2005 <sup>45</sup>	DP	Diagnose depth of burns	2D photographs	Automated assessment of burn wound depth	Supervised	Classification	Real data	NR	NR	NR
47	Yamamura et al, 2004 <sup>46</sup>	OP	Prediction of aminoglycoside/ab × concentration in burn patients	Patient risk factors and laboratory measurements	Prediction of aminoglycoside/ab × concentration in burn patients	Supervised	Classification	Real data	100% 80%	100% 66.70%	100% 85.70%
48	Acha et al, 2003 <sup>47</sup>	DP	Identify burn tissue from healthy, and classify depth of burn	2D photographs	Identify burn tissue from healthy, and classify depth of burn	Supervised	Classification	Real data	80%	NA	20%
49	Estabhanati et al, 2002 <sup>48</sup>	OP	Prediction of mortality of burn patients	Patient risk factors and laboratory measurements	Prediction of mortality of burn patients	Supervised	Classification	Real data	75%	NA	25%
50	Hsu et al, 2000 <sup>49</sup>	PP	Skull reconstruction of areas needing an operation	CT	Skull reconstruction in CT for preoperative planning	Supervised	Regression	Real data	NA	NA	NA
51	Fryc et al, 1996 <sup>50</sup>	OP	Prediction of mortality of burn patients	Patient risk factors and laboratory measurements	Prediction of mortality of burn patients	Supervised	Classification	Real data	90%	NA	10%
			Prediction of hospital stay of burn patients		Prediction of hospital stay of burn patients	Supervised	Classification	Real data	90%	NA	10%

NA, not applicable; NR, not reported.



**Fig. 2.** Summary of the QUADAS-2 (Quality Assessment on Diagnostic Accuracy Studies-2) analysis.

models included in this review were able to classify burn thickness with an accuracy of up to 99.3%, in contrast to the 60%–70% achieved by surgeons.<sup>21,76</sup> Models have also demonstrated the ability to predict mortality rates with an accuracy of 93%, outperforming commonly used predictive models such as the Belgian score, Boston score, and APACHE II with a sensitivity of 72%, 66%, and 81%, respectively.<sup>50</sup> In microsurgery, models produced high accuracy in prognosis of free flap failure (66%), whereas commonly used prognostic surgical risk calculators have been deemed unreliable for head and neck and breast microsurgical reconstruction (Brier score <0.01 and 0.09–0.44, respectively).<sup>77,78</sup> In addition, ML models demonstrated

a predictive capacity for outcomes for which predictive models have not yet been developed but may assist the surgeon in the clinical workplace. Examples include prediction of AKI in burn patients, mortality from necrotizing infections, and postoperative surgical outcomes in craniosynostosis surgery and reconstructive surgery following craniosynostosis correction.<sup>29,31,48,59</sup>

ML in plastic surgery has an incredible potential to advance patient care, but it is still in its infancy. This review has highlighted several patterns in successful application. Whenever a diagnosis is solely reliant upon a visual stimulus, for example 2D photography or CT, ML has consistently and reliably outperformed surgeons' diagnostic

accuracy.<sup>18,37,39,40,46,51,53,59,63</sup> Further, in conditions in which there are well-established correlations between certain risk markers and an outcome of interest, such as deranged blood tests on admission and AKI in burn patients, ML yielded highly accurate predictive algorithms.<sup>38,44,55</sup> However, attempts to include weakly related risk markers resulted in algorithms that had an overall lower predictive accuracy, rendering them unsafe for clinical practice. This review further identified that some plastic surgery subspecialties, such as hand surgery, have yet to incorporate this technology. This may be due to the challenging nature of classifying potential outcomes (eg, classification of hand function outcomes), or lack of data, yet future studies should aim to harvest the potential of this technology.

From a technological standpoint, this review identified three key areas to improve future algorithms, that is by tapping into the potential of expanding the dataset size using data augmentation, utilizing novel deep learning models, and making proper use of algorithm validation in research. Data augmentation can be invaluable in the creation of future algorithms, solving the main obstacle of accessibility to large amounts of data needed to train these models. It is a process by which one can artificially enhance the diversity of a patient database without actually collecting new data. (See figure, Supplemental Digital Content 1, which displays data augmentation utilizing random cropping, random rotation, and mirroring (horizontal flipping). A single datapoint has now been augmented to seven novel datapoints. <http://links.lww.com/PRSGO/B676>.)

This was utilized in only five studies in this review. O’Neil et al utilized data augmentation to enhance a database of 11 patients to 269, allowing the creation of an algorithm to predict the probability of total free flap failure in microvascular breast reconstruction.<sup>24</sup> Until large-scale anonymized medical datasets become more readily available, such as the OpenSAFELY platform, by tapping to this potential of data augmentation, clinicians can overcome the challenges of limited patient datasets. Secondly, future research could substantially benefit from utilizing more recent advances in the field of NNs and deep learning. Compared with traditional ML, deep NNs can process vast amounts of data efficiently and discover complex underlying patterns in the data at scale. A limitation here is the large volume of appropriately structured data needed to train these models. Lastly, future research should ensure that all algorithms created are validated before testing. Separating the validation and test sets is crucial because it prevents overfitting of an algorithm to a set of given data and reports a misleading higher performance. Our review identified that only 10 of the 51 studies utilized validation, indicating that there is a high risk of bias in the remaining studies, as the high accuracies of the algorithms could be the result of overfitting.

The evidence in this study is limited by the lack of high-quality level I evidence. The existing studies are mostly small retrospective case series that are inherently at the risk of bias. There are no prospective, randomized controlled trials evaluating these technologies in the clinical setting comparing them with clinician acumen, which limits our comparison on the safety and utility of the technologies.

Further, the mean accuracy, sensitivity, and specificity of included algorithms were reported collectively for all algorithms, rather than performing subgroup analysis based on the condition examined because of insufficient studies in the specialty. This pooling of results is not an indication of the accuracy of any individual model, where each algorithm should be examined in isolation. However, this still provided an invaluable insight into the accuracy of these algorithms in plastic surgery. Lastly, because of the limited MeSH terms currently utilized in ML and medicine, potentially important studies on the topic may have been missed. These are expected to be minimal, as we performed a wide library search, which was also completed by extensive reference checking to provide an accurate, up-to-date review.

## CONCLUSIONS

ML has the potential to enhance clinical decision-making in plastic surgery by making highly accurate diagnostic and outcome predictions; however, the technology is still in its infancy. There is vast heterogeneity between published studies in regard to the clinical task the algorithms are designed on and the model utilized, thus not allowing for data synthesis and meta-analysis. There is a pressing need for larger prospective, randomized control trials for level I and II data, where these algorithms are utilized in the clinical setting. Future research could benefit from larger datasets, data augmentation, state-of-the-art deep learning models, and more rigorous validation during design.

Ankur Khajuria, BSc (Hons), MBBS (Dist.), FHEA,  
FRSPH, MRCS (Eng), MAcadMed. MSc (Oxon.)

Department of Surgery and Cancer  
Imperial College London  
Ayrton Road, SW7 5NH  
London, UK

E-mail: [ankur.khajuria09@imperial.ac.uk](mailto:ankur.khajuria09@imperial.ac.uk)

## REFERENCES

1. Yang J, Jayanti MK, Taylor A, et al. The impending shortage and cost of training the future plastic surgical workforce. *Ann Plast Surg.* 2014;72:200–203.
2. Topol E. Preparing the healthcare workforce to deliver the digital future. *Health Educ Engl.* 2019;1:1–48.
3. Celtiki E. A systematic review on machine learning in neurosurgery: the future of decision-making in patient care. *Turk Neurosurg.* 2018;28:167–173.
4. Kanevsky J, Corban J, Gaster R, et al. Big data and machine learning in plastic surgery: a new frontier in surgical innovation. *Plast Reconstr Surg.* 2016;137:890e–897e.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–118.
6. Ahmad LG, Eshlaghy AT, Poorebrahimi A, et al. Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform.* 2013;4:3.
7. Ayer T, Chhatwal J, Alagoz O, et al. Informatics in radiology: comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics.* 2010;30:13–22.
8. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg.* 2018;109:476–486.e1.

9. Folweiler KA, Sandsmark DK, Diaz-Arrastia R, et al. Unsupervised machine learning reveals novel traumatic brain injury patient phenotypes with distinct acute injury profiles and long-term outcomes. *J Neurotrauma*. 2020;37:1431–1444.
10. Lopez C, Tucker S, Salameh T, et al. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *J Biomed Inform*. 2018;85:30–39.
11. Heimbach DM, Afromowitz MA, Engrav LH, et al. Burn depth estimation—man or machine. *J Trauma*. 1984;24:373–378.
12. Brown RF, Rice P, Bennett NJ. The use of laser Doppler imaging as an aid in clinical management decision making in the treatment of vesicant burns. *Burns*. 1998;24:692–698.
13. Liu NT, Salinas J. Machine learning in burn care and research: a systematic review of the literature. *Burns*. 2015;41:1636–1641.
14. Brinker TJ, Hekler A, Utikal JS, et al. Skin cancer classification using convolutional neural networks: systematic review. *J Med Internet Res*. 2018;20:e11936.
15. Gardezi SJS, Elazab A, Lei B, Wang T. Breast cancer detection and diagnosis using mammographic data: systematic review. *J Med Internet Res*. 2019;21:e14464.
16. Nindrea RD, Aryandono T, Lazuardi L, et al. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. *Asian Pac J Cancer Prev*. 2018;19:1747–1752.
17. Thomsen K, Iversen L, Titlestad TL, Winther O. Systematic review of machine learning for diagnosis and prognosis in dermatology. *J Dermatol Treat*. 2019;29:1–5.
18. Mantelakis A, Khajuria A. The applications of machine learning in plastic and reconstructive surgery: protocol of a systematic review. *Syst Rev*. 2020;9:44.
19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;6:e1000100.
20. Abubakar A, Ugail H, Bukar AM. Assessment of human skin burns: a deep transfer learning approach. *J Med Biol Eng*. 2020;40:321–333.
21. Chauhan J, Goyal P. BPBSAM: body part-specific burn severity assessment model. *Burns*. 2020;46:1407–1423.
22. Desbois A, Beguet F, Leclerc Y, et al. Predictive modeling for personalized three-dimensional burn injury assessments. *J Burn Care Res*. 2020;41:121–130.
23. Rashidi HH, Sen S, Palmieri TL, et al. Early recognition of burn- and trauma-related acute kidney injury: a pilot comparison of machine learning techniques. *Sci Rep*. 2020;10:205.
24. Bhalodia R, Dvoracek LA, Ayyash AM, et al. Quantifying the severity of metopic craniosynostosis: a pilot study application of machine learning in craniofacial surgery. *J Craniofac Surg*. 2020;31:697–701.
25. Guarin DL, Yunusova Y, Taati B, et al. Toward an automatic system for computer-aided assessment in facial palsy. *Facial Plast Surg Aesthet Med*. 2020;22:42–49.
26. Formeister EJ, Baum R, Knott PD, et al. Machine learning for predicting complications in head and neck microvascular free tissue transfer. *Laryngoscope*. 2020;130:E843–E849.
27. Boczar D, Sisti A, Oliver JD, et al. Artificial intelligent virtual assistant for plastic surgery patient's frequently asked questions: a pilot study. *Ann Plast Surg*. 2020;84:e16–e21.
28. O'Neill AC, Yang D, Roy M, et al. Development and evaluation of a machine learning prediction model for flap failure in microvascular breast reconstruction. *Ann Surg Oncol*. 2020;27:3466–3475.
29. Yoo TK, Choi JY, Kim HK. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput Biol Med*. 2020;118:103628.
30. Angullia F, Fright WR, Richards R, et al. A novel RBF-based predictive tool for facial distraction surgery in growing children with syndromic craniosynostosis. *Int J Comput Assist Radiol Surg*. 2020;15:351–367.
31. Eguia E, Vivirito V, Cobb AN, et al. Predictors of death in necrotizing skin and soft tissue infection. *World J Surg*. 2019;43:2734–2739.
32. Ohura N, Mitsuno R, Sakisaka M, et al. Convolutional neural networks for wound detection: the role of artificial intelligence in wound care. *J Wound Care*. 2019;28(Sup10):S13–S24.
33. Porras AR, Tu L, Tsering D, et al. Quantification of head shape from three-dimensional photography for presurgical and post-surgical evaluation of craniosynostosis. *Plast Reconstr Surg*. 2019;144:1051e–1060e.
34. Knoop PGM, Papaioannou A, Borghi A, et al. A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. *Sci Rep*. 2019;9:13597.
35. Hallac RR, Lee J, Pressler M, et al. Identifying ear abnormality from 2D photographs using convolutional neural networks. *Sci Rep*. 2019;9:499–504.
36. Levites HA, Thomas AB, Levites JB, et al. The use of emotional artificial intelligence in plastic surgery. *Plast Reconstr Surg*. 2019;144:499–504.
37. Shew M, NJ, Bur AM. Segmentation and classification of burn images by color and texture information. *Otolaryngol Head Neck Surg*. 2019;160:1058–1064.
38. Dorfman R, Chang I, Saadat S, et al. Making the subjective objective: machine learning and rhinoplasty. *Aesthet Surg J*. 2020;40:493–498.
39. Qiu B, Guo J, Kraeima J, et al. Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network. *Phys Med Biol*. 2019;64:175020.
40. Aghaei A, Soori H, Ramezankhani A, et al. Factors related to pediatric unintentional burns: the comparison of logistic regression and data mining algorithms. *J Burn Care Res*. 2019;40:606–612.
41. Cirillo MD, Mirdell R, Sjöberg F, et al. Time-independent prediction of burn depth using deep convolutional neural networks. *J Burn Care Res*. 2019;40:857–863.
42. Tran NK, Sen S, Palmieri TL, et al. Artificial intelligence and machine learning for predicting acute kidney injury in severely burned patients: A proof of concept. *Burns*. 2019;45:1350–1358.
43. Yadav DP, Sharma A, Singh M, et al. Feature extraction based machine learning for human burn diagnosis from burn images. *IEEE J Transl Eng Health Med*. 2019;7:1800507.
44. Jiao C, Su K, Xie W, et al. Burn image segmentation based on mask regions with convolutional neural network deep learning framework: more accurate and more convenient. *Burns Trauma*. 2019;7:6.
45. Liu NT, Rizzo JA, Shields BA, et al. Predicting the ability of wounds to heal given any burn size and fluid volume: an analytical approach. *J Burn Care Res*. 2018;39:661–669.
46. Martínez-Jiménez MA, Ramirez-García Luna JL, Kolosovas-Machuca ES, et al. Development and validation of an algorithm to predict the treatment modality of burn wounds using thermographic scans: prospective cohort study. *PLoS One*. 2018;13:e0206477.
47. Su YJ. Prevalence and predictors of posttraumatic stress disorder and depressive symptoms among burn survivors two years after the 2015 Formosa Fun Coast Water Park explosion in Taiwan. *Eur J Psychotraumatol*. 2018;9:1512263.
48. Tang CQ, Li JQ, Xu DY, et al. [Comparison of machine learning method and logistic regression model in prediction of acute kidney injury in severely burned patients]. *Zhonghua Shao Shang Za Zhi*. 2018;34:343–348.

49. Cobb AN, Daungjaiboon W, Brownlee SA, et al. Seeing the forest beyond the trees: Predicting survival in burn patients with machine learning. *Am J Surg*. 2018;215:411–416.
50. Cho MJ, Hallac RR, Effendi M, et al. Comparison of an unsupervised machine learning algorithm and surgeon diagnosis in the clinical differentiation of metopic craniosynostosis and benign metopic ridge. *Sci Rep*. 2018;8:6312.
51. Kuo PJ, Wu SC, Chien PC, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. *Oncotarget*. 2018;9:13768–13782.
52. Tan E, Lin F, Sheck L, et al. A practical decision-tree model to predict complexity of reconstructive surgery after periocular basal cell carcinoma excision. *J Eur Acad Dermatol Venerol*. 2017;31:717–723.
53. Huang Y, Zhang L, Lian G, et al. A novel mathematical model to predict prognosis of burnt patients based on logistic regression and support vector machine. *Burns*. 2016;42:291–299.
54. Park HM, Kim PJ, Kim HG, et al. Prediction of the need for orthognathic surgery in patients with cleft lip and/or palate. *J Craniofac Surg*. 2015;26:1159–1162.
55. Serrano C, Boloix-Tortosa R, Gómez-Cía T, et al. Features identification for automatic burn classification. *Burns*. 2015;41:1883–1890.
56. Mukherjee R, Manohar DD, Das DK, et al. Automated tissue classification framework for reproducible chronic wound assessment. *Biomed Res Int*. 2014;2014:851582.
57. Mendoza CS, Safdar N, Okada K, et al. Personalized assessment of craniosynostosis via statistical shape modeling. *Med Image Anal*. 2014;18:635–646.
58. Acha B, Serrano C, Fondón I, et al. Burn depth analysis using multidimensional scaling applied to psychophysical experiment data. *IEEE Trans Med Imaging*. 2013;32:1111–1120.
59. Schneider DF, Dobrowolsky A, Shakir IA, et al. Predicting acute kidney injury among burn patients in the 21st century: a classification and regression tree analysis. *J Burn Care Res*. 2012;33:242–251.
60. Patil BM, Joshi RC, Toshniwal D, et al. A new approach: role of data mining in prediction of survival of burn patients. *J Med Syst*. 2011;35:1531–1542.
61. Yamamura S, Kawada K, Takehira R, et al. Prediction of aminoglycoside response against methicillin-resistant *Staphylococcus aureus* infection in burn patients by artificial neural network modeling. *Biomed Pharmacother*. 2008;62:53–58.
62. Ruiz-Correa S, Gatica-Perez D, Lin HJ, et al. A Bayesian hierarchical model for classifying craniofacial malformations from CT imaging. *Annu Int Conf IEEE Eng Med Biol Soc*. 2008;2008:4063–4069.
63. Acha B, Serrano C, Acha JI, et al. Segmentation and classification of burn images by color and texture information. *J Biomed Opt*. 2005;10:034014.
64. Yeong EK, Hsiao TC, Chiang HK, et al. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. *Burns*. 2005;31:415–420.
65. Serrano C, Acha B, Gómez-Cía T, et al. A computer assisted diagnosis tool for the classification of burns by depth of injury. *Burns*. 2005;31:275–281.
66. Yamamura S, Kawada K, Takehira R, et al. Artificial neural network modeling to predict the plasma concentration of aminoglycosides in burn patients. *Biomed Pharmacother*. 2004;58:239–244.
67. Acha B, Serrano C, Acha JI, et al. CAD tool for burn diagnosis. *Inf Process Med Imaging*. 2003;18:294–305.
68. Estahbanati HK, Bouduhi N. Role of artificial neural networks in prediction of survival of burn patients—a new approach. *Burns*. 2002;28:579–586.
69. Hsu JH, Tseng CS. Application of orthogonal neural network to craniomaxillary reconstruction. *J Med Eng Technol*. 2000;24:262–266.
70. Frye KE, Izenberg SD, Williams MD, et al. Simulated biologic intelligence used to predict length of stay and survival of burns. *J Burn Care Rehabil*. 1996;17(6 Pt 1):540–546.
71. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–536.
72. Jarvis T, Thornburg D, Rebecca AM, et al. Artificial intelligence in plastic surgery: current applications, future directions, and ethical implications. *Plast Reconstr Surg Glob Open*. 2020;8:e3200.
73. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol*. 2018;6:75.
74. Krittanawong C, Virk HUH, Bangalore S, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Sci Rep*. 2020;10:16057.
75. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. *Radiology*. 2018;288:318–328.
76. Thatcher JE, Squiers JJ, Kanick SC, et al. Imaging techniques for clinical burn assessment with a focus on multispectral imaging. *Adv Wound Care (New Rochelle)*. 2016;5:360–378.
77. Tierney W, Shah J, Clancy K, et al. Predictive value of the ACS NSQIP calculator for head and neck reconstruction free tissue transfer. *Laryngoscope*. 2020;130:679–684.
78. O’Neill AC, Murphy AM, Sebastiaipillai S, et al. Predicting complications in immediate microvascular breast reconstruction: validity of the breast reconstruction assessment (BRA) surgical risk calculator. *J Plast Reconstr Aesthet Surg*. 2019;72:1285–1291.