

# Artificial intelligence: A rapid case for advancement in the personalization of Gynaecology/ Obstetric and Mental Health care

Gayathri Delanerolle<sup>1</sup> , Xuzhi Yang<sup>2</sup>, Suchith Shetty<sup>3</sup> ,  
Vanessa Raymont<sup>1</sup>, Ashish Shetty<sup>4,5</sup>, Peter Phiri<sup>6,7</sup>,  
Dharani K Hapangama<sup>8</sup>, Nicola Tempest<sup>8</sup>,  
Kingshuk Majumder<sup>9</sup> and Jian Qing Shi<sup>2,10</sup>

## Abstract

To evaluate and holistically treat the mental health sequelae and potential psychiatric comorbidities associated with obstetric and gynaecological conditions, it is important to optimize patient care, ensure efficient use of limited resources and improve health-economic models. Artificial intelligence applications could assist in achieving the above. The World Health Organization and global healthcare systems have already recognized the use of artificial intelligence technologies to address 'system gaps' and automate some of the more cumbersome tasks to optimize clinical services and reduce health inequalities. Currently, both mental health and obstetric and gynaecological services independently use artificial intelligence applications. Thus, suitable solutions are shared between mental health and obstetric and gynaecological clinical practices, independent of one another. Although, to address complexities with some patients who may have often interchanging sequelae with mental health and obstetric and gynaecological illnesses, 'holistically' developed artificial intelligence applications could be useful. Therefore, we present a rapid review to understand the currently available artificial intelligence applications and research into multi-morbid conditions, including clinical trial-based validations. Most artificial intelligence applications are intrinsically data-driven tools, and their validation in healthcare can be challenging as they require large-scale clinical trials. Furthermore, most artificial intelligence applications use rate-limiting mock data sets, which restrict their applicability to a clinical population. Some researchers may fail to recognize the randomness in the data generating processes in clinical care from a statistical perspective with a potentially minimal representation of a population, limiting their applicability within a real-world setting. However, novel, innovative trial designs could pave the way to generate better data sets that are generalizable to the entire global population. A collaboration between artificial intelligence and statistical models could be developed and deployed with algorithmic and domain interpretability to achieve this. In addition, acquiring big data sets is vital to ensure these artificial intelligence applications provide the highest accuracy within a real-world setting, especially when used as part of a clinical diagnosis or treatment.

<sup>1</sup>University of Oxford, Oxford, UK

<sup>2</sup>Southern University of Science and Technology, Shenzhen, China

<sup>3</sup>Eötvös Loránd University, Budapest, Hungary

<sup>4</sup>University College London, London, UK

<sup>5</sup>University College London NHS Foundation Trust, London, UK

<sup>6</sup>Southern Health NHS Foundation Trust, Southampton, UK

<sup>7</sup>Primary Care, Population Sciences and Medical Education, University of Southampton, Southampton, UK

<sup>8</sup>University of Liverpool, Liverpool, UK

<sup>9</sup>University of Manchester Hospitals NHS Foundation Trust, Manchester, UK

<sup>10</sup>The Alan Turing Institute, London, UK

## Corresponding author:

Gayathri Delanerolle, University of Oxford, Warneford Lane, Oxford OX1 2JD, UK.

Email: gkaush@outlook.com



## Keywords

artificial intelligence, disease sequelae, gynaecology, machine learning, mental health, obstetrics, women's health

Date received: 19 March 2021; revised: 26 April 2021; accepted: 28 April 2021

## Introduction

Artificial intelligence (AI) is a generic term referred to the use of a machine to model intelligent behaviour with minimal human intervention. Since 1956, AI has developed into a powerful multi-mode tool across many different areas within modern society including healthcare. AI is considered as a composite of the fourth industrial revolution and demonstrates much promise to resolve key health issues (World Economic Forum). These AI tools could be classified based on their applicability and the scientific methods used to create them (see Figure 1). Machine learning (ML) is a commonly used AI method to develop many AI applications.

AI could be a vital component in the transformation of medicine, particularly digital medicine.<sup>1</sup> It enables the development of precision methodologies applicable to multiple healthcare domains. AI-based advanced algorithms and improvements in computing power could assist in meeting healthcare challenges such as efficiency and performance of clinical services in the forthcoming years. They could also be valuable to model healthcare data to improve the alertness of clinical services for infection control responses, managing increasing population and complex disease requirements and the challenges posed by the world's ageing population.

Artificial narrow intelligence (ANI) excels at solving a single specific task which is well-defined and has no scope for subjective ambiguities. Most of the current usage of AI in healthcare falls under this category.<sup>2</sup> For example, AI-based advancements in robotic surgery demonstrate optimizing minimally invasive precision surgery instead of traditional open surgical methods.<sup>3</sup> Artificial general intelligence (AGI) takes a step further by seeking to emulate human intelligence and capabilities. This aspect of AI is under extensive study and research. It could help facilitate the shift from traditional knowledge into practice models to knowledge into evidence-based practice (EBP) models that rely on extensive data and may even provide intelligent solutions within a fraction of the time taken by the traditional translational research pathways. Automation and formal reasoning-based AI applications are used across various clinical areas such as Radiology, Neurology, Orthopaedics, Surgery and Oncology. These AI tools could become decision support systems that could guide practitioners to aid with their clinical decision-making process. It may be in the form of early detection, diagnosis, treatment or improving long-term disease management. Generating clinical trial-based big data sets would also be useful for clinical population-based research.

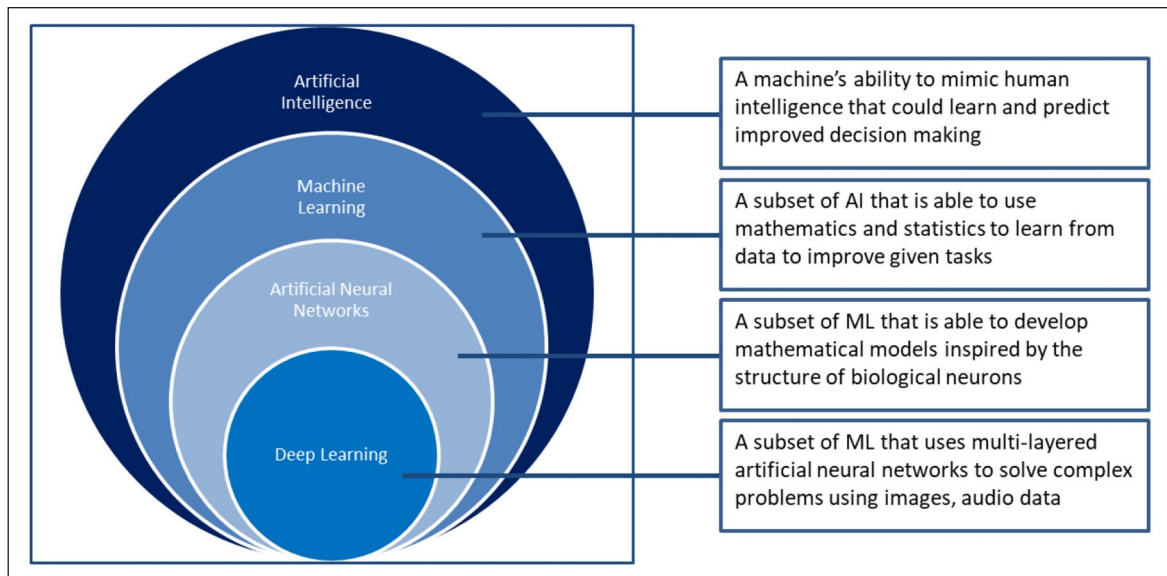
While the EBP may sound logical and simplistic, the logistical aspects require further advancement and refinement to take advantage of existing data sets and develop new data sets to address any practice gaps using more structured AI applications.<sup>4</sup> To successfully utilize EBP, a barrier in many global healthcare systems could be Information Technology (IT) infrastructure in terms of security, timely accessibility and the collaboration between the clinicians and the IT personnel. Furthermore, challenges like heterogeneous and unstandardized health data, diverse patient population, biased decision-makers and so on need to be appropriately addressed to avoid the risk of unintended and negative consequences. Thus, to truly achieve optimal gain from AI in combination with EBP, various factors would be required, including the introduction of structured methods of collating and analysing the data through existing IT systems. These could support clinical trials with sample sizes that could generate big data sets required to test the efficacy of AI applications and further support the development of novel 'real-world' research designs and methods.

While AI remains a challenging subject matter to evaluate and apply within healthcare service, especially within the context of obstetrics and gynaecology (OB/GYN), or women's health in general where primary and secondary disease sometimes shares sequelae,<sup>5–9</sup> the diffusion of AI knowledge and its application use have become an increasingly important topic to further research. It could be argued that this is an important facet for OB/GYN and mental health (MH) *sequelae*. Therefore, this rapid review aims to highlight the potential usefulness of AI methods in OB/GYN and MH domains and to demonstrate the major deficiencies in the current evidence base, hampering the development and use of AI technology to investigate the shared clinical *sequelae* of these important conditions relevant to women's health (Figure 2).

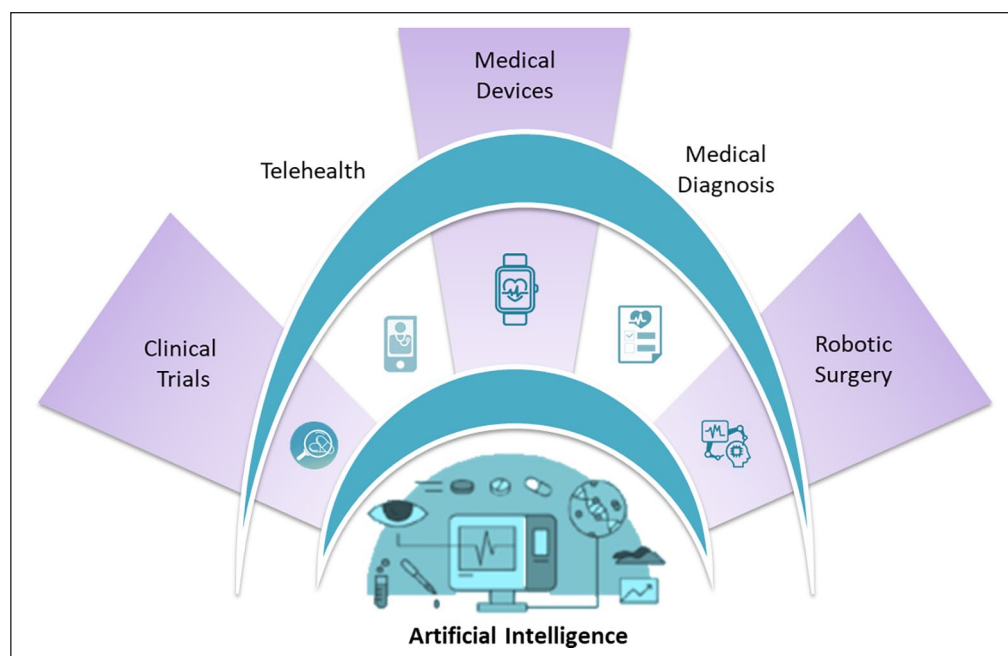
## Data analysis framework

It is vital to understand key concepts of AI to develop suitable applications that are patient-centric and also address clinical *sequelae* between OB/GYN and MH which is important in women's health.

The objective of a data analysis procedure in medicine is to help clinicians diagnose or test a new treatment by extracting the potential information unobserved behind the raw data. Thus, raw data and statistical methodology are two indispensable parts during the data analysis procedure. With the improvement of data storage, collection



**Figure 1.** A schematic representation of the classification of AI-based methods



**Figure 2.** Multi-faceted application on AI in healthcare

and analysis ability, various forms of data are available, including imaging data, acoustic data and questionnaires, which are different from traditional data. Thus, it is crucial to extract valuable features and variables from the raw data, and the data pre-processing becomes as important as the statistical methodology.

In this section, we would like to review some of the data sources in modern MH and gynaecology. We then discuss some data pre-processing methods and look at the statistical methodology applied to analyse the data set. In the next

section, some use-cases of AI applications in the OB/GYN and MH domain are discussed.

### *Input (data source)*

AI-based solutions are powered by data. Availability of data and leveraging it to extract useful information are critical to improve patient outcomes. Recent estimates show that a single patient generates up to 80 megabytes of data yearly through medical imaging and electronic medical

records<sup>10</sup> and the amount of patient health data is only growing exponentially. Therefore, a variety of data types – such as imaging data, electronic health records and questionnaires – can be collected to extract important features/predictors through some computer science or statistical methods and then carry out an AI-based or statistical analysis.

Electronic Health Record (EHR) plays a critical role in modern health information management and service innovations. A patient's EHR data contain his or her visit history, diagnosis, treatment, medication and so on. The EHR system is usually established or managed by an authorized institute in digital form (e.g. *International Statistical Classification of Diseases and Related Health Problems–Tenth Edition* (ICD-10)) so that the researchers could utilize these well-developed data to provide health services, such as early stage disease detection, treatment advice. However, different healthcare data systems capture different levels of data, requiring AI systems to compile data across systems.<sup>11</sup>

In gynaecological disease, imaging data are vital as is. Furthermore, AI-enabled usage of imaging data is emerging in diagnosis, precision medicine, treatment and monitoring. The most challenging task in imaging data analysis is object recognition.<sup>12</sup> Computer science has evolved a long way in object recognition in images from the handwritten digit recognition achieved in the early 1990s.<sup>13</sup> For instance, the ImageNet project<sup>14</sup> was designed for researchers to train their classification algorithms to achieve an error rate as low as possible. By April 2010, this large database had more than 14 million images, with each image labelled into 21,000 categories manually. In 2015, ResNet (Residual Network) from Microsoft's Research team achieved an error rate of 3.5%,<sup>15</sup> while the human benchmark for the same data set is at 5.1%. Many neural networks have been able to achieve sub 5% error rate since 2015. Thus, many deep learning (DL) architectures perform particularly well in image pattern recognition and can be of great benefit to doctors who rely heavily on images in their diagnosis or treatment.

The rapid integration of AI into the field of healthcare has occurred at an exponential rate over the recent years. Innovative technology has created a multitude of opportunities for millions of women who are affected by various MH conditions. There have been several cases whereby ML algorithms have been implemented in primary care to improve women's MH. For example, suicide prevention is an imperative public health priority; however, the risk factors for suicidal behaviour/ideation after hospitalization are significantly under researched. Edgcomb et al.<sup>16</sup> developed an algorithm to differentiate the risk of self-harm and suicide attempts, post-medical hospitalization, among women diagnosed with depression, bipolar disorder and chronic psychosis. They concluded that screening women for sex-specific predictors has the potential to improve the prevention of self-harm and suicide attempt. By assessing

these sex-specific predictors, the algorithm has significant clinical implications in the prevention of suicidal behaviour/ideation in women.

The transition to use AI systems in lower- and middle-income countries has become of increased interest over the years due to the substantial lack of trained MH professionals. Due to the staff shortages, common mental disorders that develop during pregnancy and the post-partum period often go undetected and untreated, even though there is a multitude of evidence-based interventions currently implemented for use. Green et al.<sup>17</sup> used an existing AI system that used mobile phones to deliver the 'Thinking Healthy Programme' to users. Results indicated those women who received the intervention, attributed positive life changes to it and recorded an overall improvement in their mood.

In many MH studies, factors/predictors could be put into predesigned or pretested questionnaires, and raw data could be collected by asking patients to finish these questionnaires. Although a questionnaire is a powerful and easy-to-use tool in MH issue diagnosis, it also suffers from heterogeneity. Newson et al.<sup>18</sup> analysed 126 different questionnaires and interviews for 10 different MH disorders, including depression, anxiety and so on. They observed that the similarity score across different assessment tools ranges from only 29% to 58% for a given disorder. This inconsistency will lead to a bias on clinical diagnosis and treatment. Elsworth et al.<sup>19</sup> highlighted the importance of factoring in subjectivity while inferring responses from data collected through questionnaires. Thus, we need standardized assessment tools that leave no room for subjective ambiguities in understanding underlying aetiologies and improve the precision of modern statistical algorithms based on the data from questionnaires.

In addition, with the substantial progress in natural language processing (NLP) technologies, researchers and clinicians in MH can now analyse linguistic and acoustic data. One way to collect this type of data is to carry out a predesigned interview with patients and evaluate their characteristics from the interview based on NLP and ML methods.<sup>20</sup> Furthermore, non-clinical linguistic data – such as comments written on Twitter, Facebook and blogs – are also a valuable data source for clinicians to make inferences about patient's thoughts and opinions. Patient-reported outcomes are valuable to improving clinical practice and acceptability of treatments. Wearable devices are another source of acoustic data that can provide personalized clinical services.<sup>21</sup> Therefore, exploring key AI-based scientific methods which are also technical components would further demonstrate their use in various applications.

## Methodology

AI methodologies and methods vary, although their technical composites remain unified using a few key principles such as data pre-processing, extraction, classification, selection, statistical and ML models.

**Table 1.** Summarizes vital procedures associated with data pre-processing methods.

Quality assurance	Data sampling	Encoding	Train/test split
Missing data values, inconsistent values, duplicate values are some of the data quality issues addressed in this step. The use of the ‘garbage in – garbage out’ model is one of the standard concepts used in ML to refer to poor output obtained by processing flawed data.	Varying sampling methods and techniques raise data classification issues. Therefore, classification issues should be resolved prior to training an ML model that each class is equally represented in the training data. Mock data could be useful to train some of these ML models. Therefore, data sampling is one of the techniques used to ensure equal representation across classes.	Encoding is an important step to use differently labelled data that could still be part of a disease population. Certain features with text values, for instance, cannot be directly interpreted by a machine. However, through encoding, such features are transformed into a format that retains the original meaning of the feature, but still could be accepted as input by an ML algorithm.	An ML model is built using retrospective data, although it is practically useful when it is processed correctly. To test the generalization ability of an ML model, the retrospective data could be split into training and test data sets. While the algorithm is exposed to only the training data, the finished model is evaluated on the test data. Master test plans could be put together to demonstrate the outputs of these ML models.

ML: machine learning.

**Data pre-processing.** Data collated and collected from different sources like EHR, questionnaires, social networks and so on are not likely to be in a standard format,<sup>22,23</sup> such that they could be directly used within statistical models or an ML algorithm for analysis. Therefore, data pre-processing would be a useful step to complete. Data pre-processing refers to the first step in the data analysis pipeline involving data-cleansing, transforming, encoding and formatting (Table 1) which could then be processed by an analytical framework.<sup>22</sup> A few standard procedures followed in data pre-processing are indicated below.

**Feature extraction/feature selection.** Modern statistical/AI models could be developed using pre-processed data to do prediction or classification. Output data gathered across multiple data systems in multiple formats may not be used in its original form or act as variables to assess predictive values. Data transformations are required to arrive at valuable predictors. With the predictors that have been extracted, several ML and/or statistical methods could be applied, such as logistical regression, random forest and deep neural networks (DNNs).

Although the representation and the quality of the data are improved following the data pre-processing stage, statistical algorithms may still have poor performance due to the high correlations that may exist among the features/predictors. Feature selection could eliminate the irrelevant and redundant from the set of original predictors; thus, it is an important step within the data mining process. Some of the traditional statistical methods used for feature selection include unpaired *t*-test,<sup>20</sup> recursive feature elimination method,<sup>24</sup> the Boruta algorithm<sup>25</sup> and logistic regression (LR) with  $L_1$  penalty.<sup>11</sup> However, the traditional methods work well when there are relatively fewer features to explore. In the presence of a high-dimensional data set for example, many traditional feature selection methods could become invalid; thus, a series of novel algorithms

would be required. One such method commonly used is the ‘penalized likelihood approach’ (PLA). PLA forces the model to select important features by following the *likelihood function* with penalty terms. For instance, LASSO<sup>26</sup> and the Elastic Net<sup>27</sup> are the two feature selection methods that have been widely applied to mental healthcare AI applications.<sup>25</sup> Another selection feature-based method is called ‘sequential approach’ (SA), that uses a sequential procedure to select features one-by-one in a sequence. Another similar method is Least Angle Regression<sup>28</sup> and an application example of this is L2Boost.<sup>29</sup> However, PLA and SA are developed from different statistical logic models focusing on minimizing the prediction error and the identification of relevant features, respectively.

**Statistical model – classification.** Data mining is an important statistical model-based classification. In data mining, classification means mapping of input observations to a decision class variable. Statistical classifiers do this job by learning from a data set of input variables and outcome variables. Therefore, a classifier can be expressed by a function

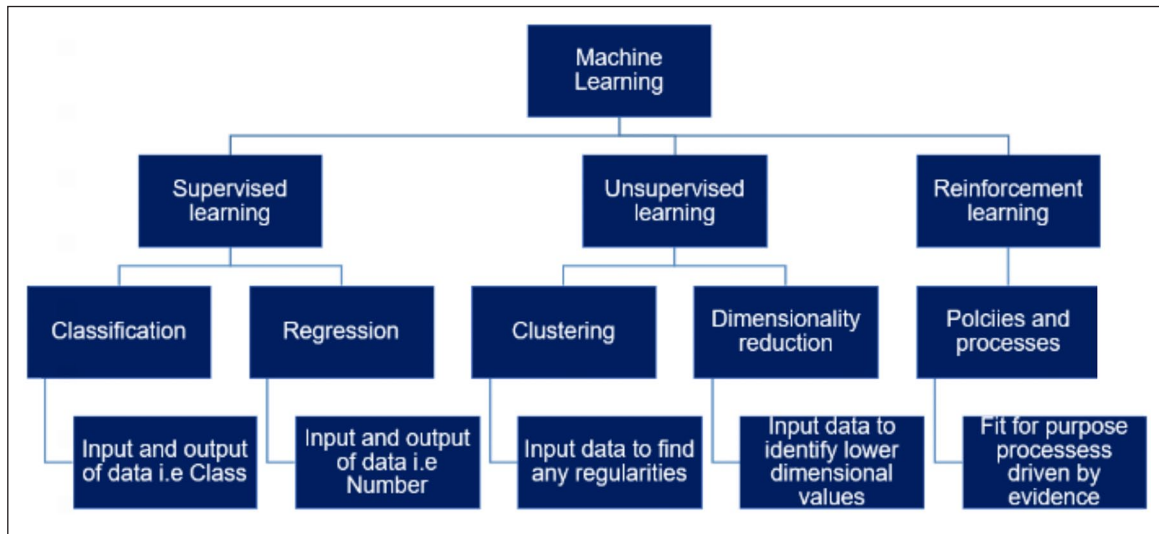
$$f : D \rightarrow C$$

where  $D$  denotes a set of variables and  $C$  denotes a set of class sets. Thus,  $f$  is a classification function which maps a set of variables to a decision set. For example, traditional statistical classification methods such as LR have been widely used in MH and OB/GYN disease diagnosis. However, given the complexity of using aggregated data sets which includes both research study and *real-world* data, linear relationships between variables and outcomes could be challenging to finalize without the involvement of a clinician. Therefore, these applications are currently acting as diagnostic-support tools for clinicians. Equally, the simple linear models may also lead to some bias which is equivalent to diagnosing clinical conditions. To better address this

**Table 2.** Summarizes the three primary methods of ML.

Supervised learning	Unsupervised learning	Reinforcement learning
This includes techniques such as logistic regression, support vector machine (SVM), and DNN involving learning from a training data set, where each data point is an input–output pair, and the objective is to find the best function that maps the input and the output data sets.	This includes techniques such as learning algorithms, in contrast, deal with the data that contain inputs, which means that the data need not be labelled by a human. The objective in unsupervised learning is to identify patterns in the input data which is not otherwise apparent. For instance, the Principal Component Analysis (PCA) is a well-known unsupervised learning method that aims at reducing the dimension of a data set.	This includes techniques such as learning is a relatively new learning framework that also does not need the labelled input–output pairs, but lets the agent take actions repetitively while rewarding positive actions and penalizing negative ones. It tries to let the agents take actions that maximize the notion of cumulative reward. <sup>32</sup>

ML: machine learning; DNN: deep neural network.



**Figure 3.** A schematic representation of multiple ML algorithms, which are a subset of AI methods that are commonly used in the development of healthcare AI applications. This hierarchy of ML algorithms is composed of three primary techniques of supervised, unsupervised and reinforcement learning. Supervised and unsupervised techniques are primary categories that use classification and regression models that could focus on qualitative and quantitative data sets, respectively, to provide clear outputs.

limitation though, non-linear classification methods – such as Support Vector Machines (SVMs), Decision Trees (DT) and Artificial Neural Networks (ANNs) – have been applied to the development of AI-based software and hardware to resolve clinical problems. Although some of these algorithms have demonstrated high performance and accuracy, limitations around interpretability are an issue. For instance, it is challenging to identify the relationship between the original predictors and the outcome using DNN architecture which could be a limitation when used in AI applications for clinical purposes.<sup>30,31</sup> However, combining these methods with statistical learning could improve the development of any applications.

### Statistical learning methods

Statistical learning methods are consistent of various methods, although a key component is ML. The core objective of ML is to apply one or more statistical algorithms on *big*

*data sets* to demonstrate an automatic prediction or classification. The three vital statistical learning paradigms include supervised learning, unsupervised learning and reinforcement learning as summarized in Table 2.

Similarly, the three primary components of ML could be further classified as shown in Figure 3.

Clinical management starts with a diagnosis and most patients would have ongoing treatments leading to the gathering of large data sets with multiple data points consisting of *input–output* pairs of data. For MH and/or OB/GYN applications, supervised learning algorithms would be a useful aspect to explore. Key techniques used within AI applications are shown in Table 3.

**LR.** The usual LR model demonstrates key points to estate coefficients as demonstrated in the equation below.

Let  $\{(y_i, x_i)\}_{i=1}^n$  denote the  $n$  observations, where  $x_{i1}, x_{i2}, \dots, x_{ip}$  are the  $p$  features of the  $i$ th observation, and  $y_i$  is the binary output

**Table 3.** Machine learning algorithm–based AI application methods.

Supervised learning	Unsupervised Learning	Reinforcement Learning
Logistic regression	Principal component analysis (PCA)	Monte Carlo simulation
Support vector machine (SVM)	K-means clustering	Q-learning
Deep neural network	Expectation maximization (EM) algorithm	State–action–reward–state–action
Naïve Bayes	Hierarchical clustering	Double Q-learning
K-nearest neighbour (KNN) algorithm		
Random forest		

$$\pi_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

$\beta_0, \beta_1, \dots, \beta_p$  being the coefficients or the mathematical weights.

Then, we consider  $y_i = \sigma(\pi_i)$ , where  $\sigma(x) = 1 / (1 + e^{-x})$  is called the sigmoid function that binds the output to the interval (0,1). This forms the mathematical setup of LR model with the objective of estimating the coefficients  $\beta_0, \beta_1, \dots, \beta_p$ .

Within this method, a number of input variables are linearly combined with appropriate weights and regressed to logical units (i.e. 0 or 1). It is a traditional yet powerful statistical tool to develop a binary (or multi-class) classification algorithm. However, traditional LR would be ill-posed when the dimension of the data  $p$  is larger than the sample size  $n$ . Tikhonov and Arsenin<sup>33</sup> added an  $L_2$  penalty that follows the likelihood function, which is called ridge regression, that guarantees the existence of a solution.<sup>33</sup> But the  $L_2$  penalty raised a new problem due to the strict convexity of the regularization function. Every feature of this mathematical equation would receive a non-zero weight, as a result fail to carry out the variable selection. In 1996, Tibshirani<sup>26</sup> substituted the  $L_2$  penalty in ridge regression with  $L_1$  penalty instead, that is, LASSO. The LASSO singularity of the  $L_1$  regularization at origin point guarantees that this could be a variable selection-based algorithm, for example. In addition, in 2005, Zou and Hastie<sup>27</sup> proposed the Elastic Net regularization, which is a compromise between  $L_1$  and  $L_2$  regularizations. Thus, it is not only a variable selection algorithm but also could be applied to the data with highly correlated variables.

Use of LR technique has been tested in some MH studies for classification purposes to confirm the presence or absence of depression.<sup>24</sup> Although LR is a simplistic method, it could be an effective method within the context of a fitting clinical problem.<sup>24</sup> However, the presumed linear relationship between the input variables and the outcome could be problematic,<sup>25</sup> creating the need for more novel methods like SVM, Decision Trees and the Gaussian process classifier.<sup>34</sup>

**SVMs.** SVM is one of the most theoretically driven classification algorithms in modern ML. It was first established by Vapnik et al.,<sup>35</sup> and then was widely used in disease diagnosis,<sup>36</sup> handwriting recognition<sup>37</sup> as well as facial

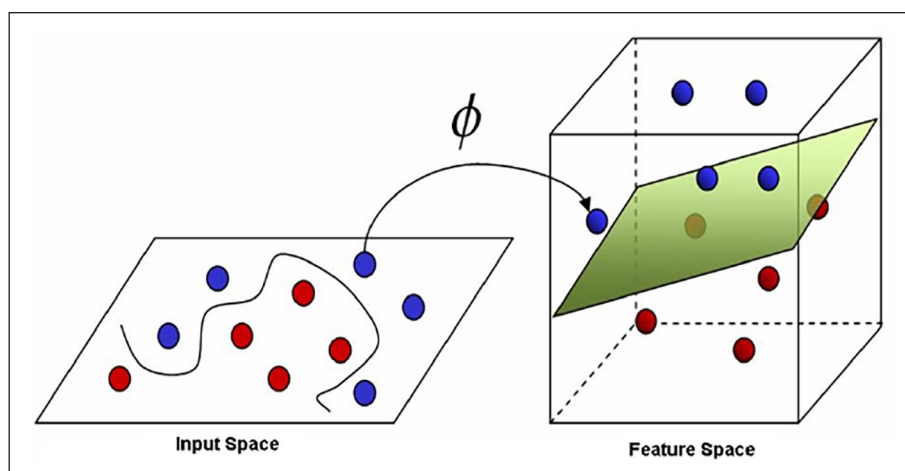
recognition.<sup>38</sup> Linear SVM attempts to modulate an optimal hyperplane in a multi-dimensional data set to classify testing data sets. The hyperplane border that separates objects across classes is optimal in the sense of maximizing the margin of separation, thereby increasing the generalization of the model. This key characteristic of SVM models makes it an attractive technique to develop AI applications.

However, it is often not possible to find a linear separation in practice, especially in the context of *real-world data* or aggregated data. However, using the kernel trick, a non-linear mapping  $\Phi$  from the input space  $X$  to a high-dimensional space, where finding a linear separation could be possible (Figure 4). Although, there is no set standard for the choice of kernel for a given problem, leading to possible bias in the model and hence poorer performance.<sup>11</sup>

Both linear and non-linear SVM classifiers are widely used in MH studies for diagnosing purposes. Key examples of applications of SVM in MH care include use of warning markers extracted from NLP to detect risk of violence in schools<sup>20</sup> and detect psychological distress of university students.<sup>39</sup> However, these require to be further developed and tested if they are to be unilaterally used in various clinical contexts.

**Ensemble learning.** Ensemble methods are a group of general techniques in ML where several predictors are combined to develop a single entity. It is often difficult to directly design an accurate algorithm with high prediction accuracy. Therefore, a mixed-methods approach is widely suitable for developing AI applications for the purpose of clinical consumption. The underpinning concept of ensemble learning would be to boost techniques, and to develop a series of weak learners to build a strong learner.

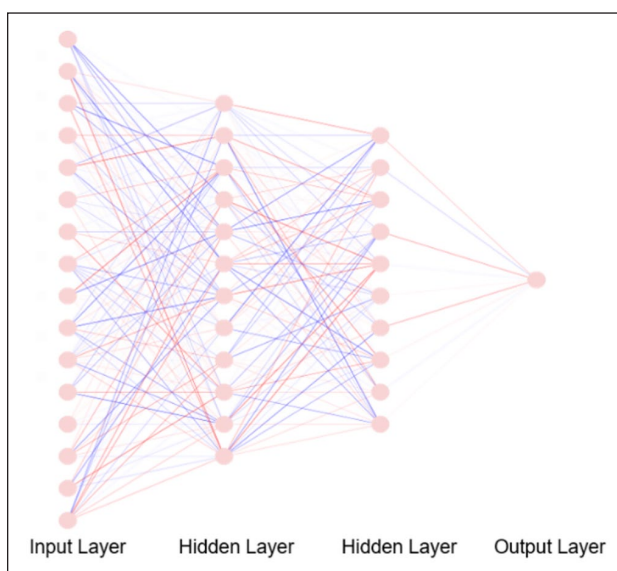
Techniques to create an ensemble model can be broadly classified into sequential and parallel. Sequential techniques such as boosting create a sequence of base learners. Base learners are assumed to be weak in the beginning of the sequence, but with higher weights assigned to misrepresented cases in the subsequent learners although eventually, it is hypothesized stronger models would be obtained. Base learners in an ensemble model are usually homogeneous, that is, of the same type. However, it is also possible to ensemble different types of base learning algorithms resulting in heterogeneity, which is particularly explored



**Figure 4.** The non-linear SVM classifier with the kernel trick.

in neural network architectures.<sup>40</sup> Parallel learner techniques, however, generates independent base learners in a parallel format. Although individual base learners are likely to be weak, voting is done across all the learners that significantly reduces error of the system. For instance, random forest is a parallel ensemble of multiple decision trees. These ensemble learning methods have been involved in MH and OB/GYN studies. Some examples include predicting late-onset pre-eclampsia among low-risk patient populations in their early trimester,<sup>6</sup> identifying high-quality embryos in in vitro fertilization (IVF),<sup>40</sup> interpreting cardiotocography (CTG) reports to diagnose foetal compromise during pregnancy and labour.<sup>41</sup>

**DL.** DL is a complex ML technique and is a useful tool to utilize for epidemiological data sets in particular. Consider the problem of deciding whether a patient is suffering from both depression and anxiety when given a number of predictors which are reported as symptomatology. One way to determine a primary diagnosis and/or secondary diagnosis, in other words a *sequelae*, would be to perform LR on the data set so that the weighted sum of the variables which in this case would be symptoms, would aid to indicate the primary causation of the reported symptomatology and thereby confirm a diagnosis. Similarly, if there was a *disease sequelae* involved, this could be determined by evaluating the weight-in. However, given the complexities of clinical and/or clinical trial data, the interaction terms, which depict the link between variables, may increase dramatically and lead to the failure of the traditional generalized linear model. The neural networks of DL, however, handles this issue well by adding multiple hidden layers of the data (Figure 5) to perform various non-linear transformations on the original data set. This is followed by the application of a *back-propagation algorithm*<sup>42</sup> on the training data set to obtain a good set of evidence by way of hidden factors. Although the concept of neural networks



**Figure 5.** A representation of the ANN with a 16-dimensional input layer and two hidden layers; each one with 12 and 10 neurons. Each of the two hidden layers may represent a specific type of features that need to be detected. The interaction between two nodes is represented by coloured edges, where positive interaction is shown by red and negative interaction is shown by blue. The edge width and edge opacity are proportional to edge weights.

was introduced as early as the 1960s<sup>43</sup> and technique for practical implementation using back-propagation algorithms dating back to 1986,<sup>42</sup> application development was limited due to the requirement of high computational power and large volumes of data. However, since 2012 with the availability of fast processors and big data, neural networks have been able to achieve high levels of prediction and classification precision. Subsequently, many neural network architectures have been widely used within health care including MH and OB/GYN.

**Table 4.** Confusion matrix.

Confusion matrix		Actual		
		Positive	Negative	
Predicted	Positive	<i>a</i>	<i>b</i>	Positive predictive value = $a / (a + b)$
	Negative	<i>c</i>	<i>d</i>	Negative predictive value = $d / (c + d)$
		Sensitivity $a/(a+c)$	Specificity $d/(b+d)$	Accuracy = $(a + b) / (a + b + c + d)$

ANNs from the AI world are rough analogies to the functionality of the brain as described by Dr Sussillo, a Computational Neuroscientist within the Google Brain Team. The ANNs comprise multiple nodal layers analogous to neurons, where the layers are connected through a series of mathematical weights that behaves like a synapses between neurons. This hidden layer's connection to the output layer allows the data to be analysed by an algorithm that is embedded within an AI application. One of the most successful applications of DL has been image recognition. Thus, researchers and clinicians who rely on radiological methods to make diagnosis or manage treatment optimization will benefit further through automation-based algorithms. Predicting clinical outcomes, for example, of uterine fibroid embolization (UFE) based on pre-procedural magnetic resonance imaging (MRI) scans,<sup>44</sup> using time-lapse images to identify high-quality embryos in IVF<sup>45</sup> and evaluating myometrial invasion (MI) depth based on endometrial cancer MRI<sup>46</sup> are a few examples of DL model applicability within the OB/GYN domain. DL techniques are also widely used in cases other than analysing images. For example, sentiment analysis from tweets<sup>40</sup> and providing personalized recommendations using data gathered from wearable devices are a few wellbeing-based AI applications currently in use.

### Model evaluation

Based on the evidence thus far, there are many algorithms, models and techniques to address classification problems. As per the 'no free lunch' theorem,<sup>47</sup> a single ML algorithm cannot be universally applied. Therefore, AI techniques and methods used to develop application to solve clinical problems need further exploration within themselves and mixed-methods approaches should be utilized to construct fit-for-purpose models. Thus, post-construction of any statistical or ML models, it is vital to evaluate its performance. Similarly, in order to generalize models, testing of these within a clinical population, for example, is important. As a result, model evaluation procedures would be required first in an experimental setting. Depending on the objective, the model evaluation could be divided into two sections: classification evaluation (for classification) and regression evaluation (for prediction).

In terms of regression evaluation, there is a wealth of criteria by which the models can be evaluated and compared, such as mean square error, relative square error, standardized residuals and coefficient of determination ( $R^2$ ).<sup>48</sup> However, these still require validation within a clinical trial so that 'real-world' applicability could be tested.

Since most models in MH and OB/GYN are constructed for diagnosis, the primary focus on reviewing classification evaluation methods would be to consider a confusion matrix. A confusion matrix showcases the number of correct and incorrect predictions made by the classification model compared to the true outcomes (Table 4). Apart from model accuracy (number of correctly classified cases to the total number of cases), it provides researchers a clear insight into the sensitivity (the true positive rate) and specificity (the true negative rate) of the classifier, which are the two most commonly used metrics used to measure the performance of a model.

Within the confusion matrix, Kolmogorov–Smirnov (KS) chart measures the degree of separation of a classification algorithm.<sup>49</sup> The K–S scores 100 if the classifier separates the positive data points and negative data points perfectly, while the K–S score is 0 if the model selects the cases randomly. This is followed by assessing the receiver operating characteristic (ROC) curve developed by plotting the ratio of sensitivity and (1 – specificity) at various thresholding settings. The ROC curve plays a similar role as the K–S score. A simple area under the ROC curve (AUC) is equal to the probability that a classifier will rank a randomly chosen positive observation higher than a randomly chosen negative one.<sup>50</sup> At this point, an assumption is made about positive observations, in that they rank higher than the negative ones. Therefore, a 'perfect' classifier should achieve an AUC of 100% ideally.

While interpretability of a metric is important to accept the results of a model, it is also vital to choose an appropriate metric to optimize for, depending on the objective of the model. Accuracy, for instance, is straight-forward in its definition and interpretation. But, in cases where there are missing data or lack of a positive case is more important than failing to properly classify a negative case. A model's performance particularly in terms of sensitivity and specificity should be preferred over overall accuracy, similarly, to reaching an accurate diagnosis for a patient.

## OB/GYN-MH sequelae

Equipped with the brief understanding of the AI/ML pipeline in a broader context, the next step would be to evaluate potential applications to evaluate MH and OB/GYN individually as well as in the context of the sequelae of OB/GYN and MH. Interestingly, while physical and MH share a complex symbiotic relationship, there is a dearth of evidence for multi-morbidity research globally. This limits the ability of AI researchers and clinical researchers alike to develop cross-specialist applications. Theoretically, multi-purpose and generalizable AI applications would be cost-effective to many healthcare systems, thus allowing multiple clinicians to access their patient's clinical information more conveniently. Currently, communication between clinicians in acute healthcare and mental healthcare services are quite disconnected. This is further complicated in healthcare systems where patients may access primary, secondary and tertiary care centres or a combination of national services and private healthcare as the data are maintained in separate systems. As a result of this, patients may receive suboptimal care. This is a significant problem for patients needing both OB/GYN and MH care, since symptomatology and patient-reported outcome as well as differing clinician approaches even, in some cases, could misinterpret a potential mechanistic relationship between the two conditions. However, current research data demonstrate the need for a better understanding of OB/GYN sequelae with MH, for example, as is the case in post-partum depression, gestational diabetes or endometriosis. This is furthered by patients using social media to discuss health issues and recurring symptoms, for example. De Choudury et al.<sup>51</sup> noted this observation and considered developing a predictive model to forecast significant post-partum changes in mothers using their Twitter posts. They analysed 36,948 tweets from 376 mothers who were in their prenatal period and the tweets from the same Twitter users were then reviewed during their post-partum period. A total of 40,426 tweets were analysed through an SVM method that showed a better prediction accuracy post-partum changes compared to other ML models with an improvement range from 10% to 35%.

Another project that used social media platforms like Twitter, Facebook and Instagram examined the possibility of using social media as a public health tool, in order to ensemble a recurrent neural network (RNN) and convolutional neural network (CNN) to identify pregnant women. Warikoo et al.<sup>40</sup> analysed a small corpus of 3000 tweets and could identify pregnancy cases with F1 score of 95%. Furthermore, among those identified as accounts with legitimate pregnancy case, they could classify 62 tweets as ones discussing health issues or expressing health concerns (like nausea, swollen feet, etc.) and 157 tweets with negative behavioural emotions (sad, depressed, exhausted, etc.). This was an interesting finding and could be further developed to improve the understanding of the psychosocial or

negative effects often discussed in relation to the use of social media as well as cognitive behavioural interventions that could aid digital interphases to minimize distress. Through these projects, AI demonstrates its versatility in terms of design via conceptual development for 'real-world' problems.

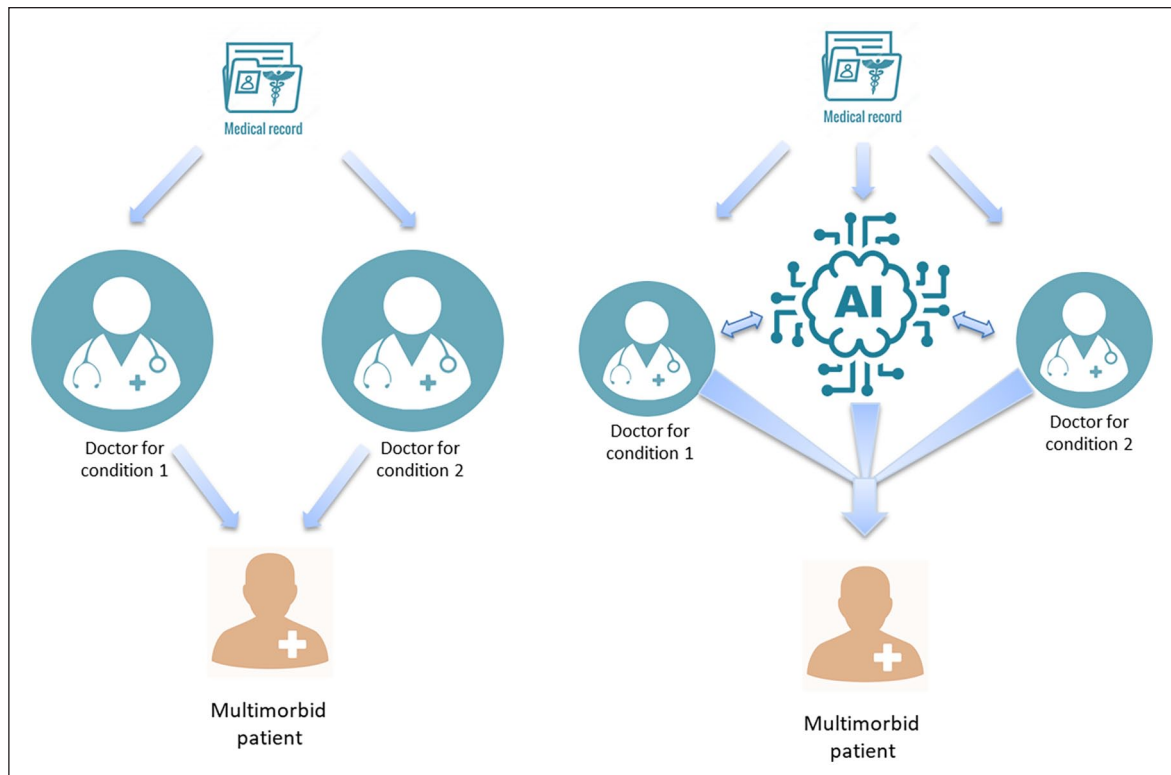
Tasks such as facial recognition could be assembled within the ANN as a series of numbers to describe the pixel of the images.<sup>52</sup> Tulchinsky et al.<sup>53</sup> further report that ANN interacts to deliberately mimic cognitive functions to provide probabilistic responses using dependable and trained mathematical systems interpreting multi-factorial data from multi-morbid conditions such as the sequelae of pre-term labour and MH. This gives rise to complex algorithm-based AI systems known to most as Medical Technologies. These medical technologies could give rise to important AI software applications in OB/GYN especially for refining genomics. Most of the recent advances in genomic medicine can be attributed to use of AI techniques to some extent.<sup>54</sup> With this growth, obstetricians-gynaecologists are encouraged to leverage genetic testing and support their patients though proper counselling during the process.<sup>55</sup>

Apart from biotechnological innovations, AI could be used to leverage data collected from wearable devices. These could be used by OB/GYN patients to demonstrate a variety of data that could assist clinicians, especially to highlight any ongoing symptomatology either physical or MH related, as symptoms and potential psychiatric comorbidities, for example, need to be managed differently (Figure 6).

## AI and personalized medicine

AI applications have already demonstrated their applicability in various other clinical specialties and the directionality of these covers the development of patient-centric clinical care pathways for the future, in areas such as the utilization of data analytics, which drives the novel treatment landscape and personalized medicine (PM).<sup>56</sup> PM is a comprehensive approach to treat patients who demonstrate a primary and secondary condition. In order to improve accuracy and performance of PM, AI systems that use patient reports and clinical outcomes are required to adapt and continuously improve the decision outcomes. Equally, it is also important to develop analytical methods, tools and data sets that could augment the analysis of *real-world* patterns of disease progression and prognosis. Harnessing this information in the development of AI applications is fundamental to ensure that they perform the desired tasks effectively.<sup>57</sup> Advanced cloud computing and human machine interfaces are also vital to support DL analytics, since solutions from such AI systems need to be easily interpretable and accessible in a timely manner.

Another vital precept of AI applications, within PM in particular, would be the identification of sub-groups of the



**Figure 6.** Treating multi-morbid conditions: traditional approach (left) versus AI-supported integrated approach (right).

primary disease population that may then influence the characterization of the overall disease risk, pathology, prognosis and subsequent treatment.<sup>58</sup> For example, patients who have had preterm births (PTBs) are sub-categorized to extremely preterm (<28 weeks), very preterm (28–32 weeks) and moderately to late preterm (32–37 weeks). These patients may have a varying degree of risk, prognosis and treatments depending on their disease sub-group. In addition, their physical and MH would differ. In order to better evaluate the clinical management requirements of these patients, multi-dimensional and multi-morbidity data sets would be useful to evaluate clinical outcomes. This would aid in developing better personalized interventions and AI could play a vital role, given its ability to analyse large data sets with a higher precision accuracy by collating and reviewing the evidence.

The use of digital phenotyping methods by AI is a technique that could further aid personalization of treatments.<sup>59</sup> Some of these methods have already been used in areas such as Psychiatry<sup>59</sup> and similar approaches could be used in OB/GYN areas. This is in contrast to traditional healthcare data analytics methods, where long-term epidemiological data are needed to address key issues in healthcare, such as over-treatment, under-treatment, delayed diagnosis, misdiagnoses, low productivity and over-expenditure. In the traditional process, clinical research had to be conducted over a long period to develop statistical models. It may take an average of 10–15 years with additional time

and operational requirements using implementation methods before being able to make any changes to the relevant clinical services. This in turn would delay any clinical care improvements that could benefit patients and healthcare services alike. On the contrary, AI is able to use a variety of data sets, simultaneously even, in different formats and use advanced analytical methods to identify relevant predictors and provide more accurate outcomes, even with heterogeneous data, to almost contemporaneously support healthcare systems.<sup>58,60,61</sup> Principles such as digital phenotyping by way of using scientific and logic models in an AI application could help predict the burden of disease for example, response to treatment, prognosis and any subsequent changes to patient-reported outcomes.<sup>56</sup> In contrast to the current traditional clinical processes for characterizing symptoms, AI would use probabilistic learning within this context. This approach could prevent the disease and/or monitor progression of the disease and/or aid overall long-term management. Thus, as a result of AI, a bespoke clinical decision-making process can be generated to enhance the current ‘fragmented and disintegrated’ system that could be economical for healthcare systems as well.<sup>62,63</sup>

While the AI application development sphere could use multiple methods to identify and report digital phenotypes relevant to clinical outcomes, these are not well-examined complex multi-morbid conditions. For example, endometriosis patients report depression and anxiety, although

research into determining the mechanistic nature remains unknown. Furthermore, the lack of a multi-morbid data set to this effect could further introduce rate-limiting factors. Also, the use of AI, particularly in PM, depends on high performance computing methodologies and big data sets to improve the precision accuracy. Thus, the recent advances of AI applications and its testing requires greater focus on conducting comprehensive clinical research if these are to be introduced to clinical practice. In addition to this, expansive, multi-centre, multi-national clinical trials are required to obtain accurate data sets to develop large data sets that enable the testing and verification of the efficacy of the AI tool to be reliable and patient friendly. This is particularly important for AI applications that are considered as medical devices based on the evolving regulations and legislations. Despite these limitations, a large proportion of AI-based research is broadly defined and published before appropriate conformity assessments are performed.

However, prediction modelling and the use of algorithms have been successfully demonstrated in Radiology, Oncology, Cardiovascular and Orthopaedics surgery.<sup>64–65</sup> In these clinical specialties, AI has been reported to enhance the prediction of risk with high levels of precision and allow development of software applications, as well as technical protocols to enhance performance for tailored diagnosis and treatments. Neural network-driven machine learning and evolutionary algorithms have shown much promise in predicting some difficult problems using polynomial algorithms through supervised, semi-supervised or unsupervised imputation. Although this requires investment in high performance computing infrastructure, it could be replaced with the use of algorithms embedded into existing electronic systems as combination or stand-alone software. Therefore, implementation of the PM pathways impact healthcare significantly and could provide substantial cost savings and could be particularly useful for high impact areas such as OB/GYN by reducing costs in high-cost areas such as fertility treatments. This collective evidence warrants a more intimate review of the use of AI currently in key OB/GYN and MH disorders.

### *AI in obstetrics*

**Pregnancy health surveillance.** Pregnancy surveillance is one of the most important issues during the early stages of pregnancy and the rest of the antenatal period, particularly, in high-risk pregnancies. Pregnancy-associated clinical research provides insightful knowledge to improve clinical practices, although research to identify the possible MH sequelae remains limited. In addition, a stronger emphasis is indicated undoubtedly towards maternal mortality and morbidity associated with pregnancy-related complications. To simultaneously address both these aspects effectively and potentially, predictive

models could be developed using AI tools that may reduce both morbidity and mortality rates. Reviewing available literature on use of AI in healthcare for pregnant women, Davidson et al.<sup>66</sup> observed that AI methods have been explored on the entire range of the pregnancy process, from preconception to postnatal health concerns. However, the study found that the current research lacks thorough external validation, thereby limiting the generalizability of the proposed methods. Also, they found limited research on understanding the impact of pharmacological intervention after pregnancy.

Another example is the use of CTG, which is considered to be an effective and reliable method of diagnosis of foetal compromise during pregnancy and labour. However, interpretation of the CTG is subjective. As a result, AI-based methodologies could be introduced to improve the effectiveness and efficiency for the interpretation and support clinicians to improve their decision-making as well as reduce the cost-burden of carrying out high-risk surgical interventions. Fergus et al.<sup>67</sup> explored this further using 13 features that were extracted from CTG reports to ensemble a classification tool comprising of a Fisher linear discrimination analysis (FLDA), Random Forest and SVM to classify different caesarean section and vaginal delivery births. SVM is a classifier that seeks a hyperplane that maximizes the margins between the classes. FLDA aims to project the data to a line that maximizes the distance between the means of classes while minimizes the variance within each class. Thus, SVM and FLDA are relatively easier to separate those observations that are far away from the decision boundary. However, SVM and FLDA perform poorly when used to classify caesarean and vaginal delivery births that are close to the boundary. Random Forest algorithms, however, use data points to vote and classify new observations, which make it useful to classify observations that are located close to the boundary. Therefore, the ensemble model used in Fergus et al.<sup>67</sup> syntheses demonstrated the strength of each model and a possible solution to improve the overall prediction accuracy. This algorithm was then trained on 552 intrapartum recordings that indicated an AUC of 0.96. While interesting, a significant limitation in Fergus and colleagues' study is the size of the data set, which is fairly small for testing and validating ML algorithms. Interestingly, Lu et al.<sup>68</sup> developed a novel framework for intelligent analysis and automatic interpretation of digital CTG signalling. They extracted several features and characteristics from the CTG trace, and incorporated multiple scoring methods of Kreb's, Fischer's and Modified Fischer's for evaluating foetal wellbeing. Their hypothesis was to accurately place the changes in the foetal heart rate alongside uterine contractions, thus making the clinical interpretation more accurate. This tool could either be used in hospitals or at home by the patients themselves. Therefore, this tool has the potential to support obstetricians' function more

effectively. However, like many other AI tools that show initial promise, it requires clinical trials to test and validate its use for a wider population, as well as demonstrate its superiority over other existing methods.

**PTB.** Although antenatal care plays a key role in the birth of a healthy baby, predicting the time of delivery is considered as one of the most challenging problems for both the mother-to-be and for the obstetricians. PTB is considered a major cause of perinatal mortality and is an important health issue, particularly in low- and middle-income countries (LMICs). The psychological impact of PTB is significant and often a sensitive area of discussion among women, which is further complicated by cultural stigmatization. Therefore, it is not a surprise that there is limited research available within the PTB sequelae with MH especially among the Black, Asian and Minority Ethnic (BAME) populations. Furthermore, the financial burden of PTB on the woman, their families and healthcare systems are equally burdensome. Therefore, using non-invasive tools developed via statistics and AI could provide a valuable solution that could include self-reporting methods, for example, to be introduced to women using mobile apps or mobile chat-bots. These applications use conceptual algorithms to operate and could provide women support (non-clinical) in the first instance. Currently, there are various patient diaries that are accessible using either the Apple or Android applications. These could be further developed for use within the clinical domains.

However, further development of the digital sphere using AI would require exploration of the current evidence base. For example, the subjectivity aspect associated with PTB could also be supported by developing prediction models using ML algorithms. Electrohysterography (EHG), an effective and reliable tool to monitor pregnancy and labour, could also be used to predict PTB.<sup>67</sup> Many studies have demonstrated that peak frequency, medium frequency, root mean square and sample entropy of uterine contractions that are extracted from EHG to be significantly different between preterm and term deliveries. Fergus et al.<sup>67</sup> and Idowu et al.<sup>69</sup> considered combinations of this with the impregnability of the frequency-related parameters in terms of signal quality variation, as core features to perform various ML classifiers on an open data set.<sup>67,70</sup> They showcased that ML classification algorithms could provide a robust method to predict PTB while Idowu et al.<sup>69</sup> suggested that Random Forest Classifier outperforms other algorithms with a specificity of 86%, a sensitivity of 97% and an area under the curve (AUC) of 94%. Encouraged by these results, Despotovic et al.<sup>71</sup> considered a novel set of features to perform a series of ML classifiers, and the Random Forest Classifier still outperformed other methods with an AUC of 99% and a sensitivity of 98.4%. Thus, this level of optimization could be translated to cost-effective software packages that could support

clinicians to deliver optimized care for their patients. While these AI tools have the potential to act as innovative tools, to determine their precise viability and effectiveness among a wider population of patients, global clinical trials with significant sample sizes are needed. This would also assist healthcare organizations and policy-makers to produce informed decisions around care pathways that are cost-effective and sustainable.

### *AI in gynaecology*

**IVF.** AI application use in reproductive medicine is another exciting partnership between humans and machines. The execution of emulating or exceeding human intelligence is at the forefront of using AI concepts, especially within the field of laboratory science associated with IVF. Within the last decade or so, several technological advancements have emerged, including time-lapse microscopy that became a 'revolving door' to automatization of embryo culture techniques. Formerly, the classic morphological evaluation is the most widespread approach for blastocyst selection during the embryo fertilization process.<sup>96</sup> However, with the use of time-lapse imaging, the morphological appearance and morphokinetic events can be monitored and evaluated for selection of viable embryos. But the process of embryo evaluation and transfer is determined by embryologists, which could be subjective like most other clinician-based reviews. AI could facilitate optimization and standardization of this process, by providing improved image reconstruction for recognition and selection of viable embryos in an automated manner.

As a result of this, continuity of the development of observations could be applied in the form of an ANN architecture to evaluate quality of embryos using morphokinetic and morphological events that achieved an 83% accuracy as shown by Zaninovic et al.<sup>72</sup> also showed the application of an Inception-V1 algorithm to improve the parameters to obtain a 97.6% accuracy to discriminate between groups of poor and good blastocysts by way of using 50,392 images and selecting only 10,148 embryos using the time-lapse system. Despite this evidence, time-lapse hardware use is prohibited in most laboratories due to the lack of statistically significant clinical trial data to show the possible superiority of this software in comparison to conventional methods.<sup>73</sup> As a result, Khosravi et al.<sup>45</sup> developed a single static two-dimensional image capture optical light microscopy using an AI-based ensemble model. Eight final prediction models based on ResNet or DenseNet were selected to form the ensemble model, and individual model selection relied on diversity and contrasting criteria, which made this model to be more robust and accurate. Although this showed a further improvement, a larger data set is required to test its use within clinical practice. A significant limitation appears to be the existing poor accessibility to clinical trial data sets with

sufficiently large sample sizes, since successful testing of any AI application requires a considerably large data set to showcase precision accuracy.

**Management of uterine fibroids.** Fibroids (leiomyoma) are non-cancerous growths of the uterine muscle that often occur in pre-menopausal women. Fibroids can vary in size from tiny seedling growths to large pelvic masses, which can occupy most of the pelvis and the abdominal cavity. Many women with fibroids are asymptomatic, yet depending on the size, number and location of fibroids, others suffer with multiple symptoms caused by a fibroid such as heavy periods, pressure symptoms, or discomfort and sub-fertility. UFE is an effective treatment for the management of fibroids and alleviating the associated symptoms. However, currently there are no means of predicting the response to treatment. MRI is used before and after UFE to determine response to treatment. Luo et al.<sup>44</sup> have used a DL model to predict clinical outcomes of UFE based on pre-procedure MRI scans. They trained and tested a Residual Convolutional Neural Network (ResNet) to predict fibroid volume reduction and clinical outcome using a retrospective cohort of 409 patients with 727 fibroids at a single institution. All AI models achieved higher accuracy in predicting outcome when compared with experienced radiologists. The limitations to the study included reliance on data from a single institution and the lack of an advanced MR sequence. Cancerous change in a fibroid (sarcoma) is rare and uterine sarcomas account for 8% of uterine malignancies.<sup>74</sup> The prognosis of malignant leiomyosarcoma is poor, and the incidence is low. For example, leiomyosarcoma in women under 40 is less than 1:1000, while up to one in three women will have a fibroid. There is no single non-invasive investigation that can reliably identify rare and malignant leiomyosarcomas from benign and common fibroids.<sup>75</sup> Diagnosis and screening of women with uterine masses, using AI offers promise. Malek et al.<sup>76</sup> developed an AI method based on an ensemble ML approach (includes *K*-nearest neighbour (KNN), Random Forest and SVM) to distinguish uterine sarcoma from benign leiomyomas based on perfusion-weighted MRI parameters. This study demonstrated that the Random Forest classifier achieves the best result with a 100% sensitivity and a 90% specificity in 42 women with 60 fibroids.

**Gynaecological surgery.** In addition to the improvements demonstrated in diagnosis and predicting prognosis in a variety of gynaecological conditions thus far, gynaecological surgery could also significantly benefit from AI applications. This has already demonstrated cost reductions in complex Orthopaedics and Oncology surgery with the use of robotics surgical methods, another AI tool that often consists of multiple AI applications. Another good example of AI applications in surgery is the virtual reality-AI used to identify patient factors, repetitive

patterns and treatment algorithms to project outcomes. This is particularly important as AI is unaffected by many variables associated with traditional surgical approaches, for example, the surgical skills of the surgeon or differences between patients. While it is challenging to create a precise algorithm to address these aspects built on the current AI knowledge base, it has enabled the introduction of robotics and AI-assisted software use in imaging spatial awareness. These systems could aid surgeons by producing better images and by developing three-dimensional printing (3DP) that could duplicate a surgical site, which is superior to a standard two-dimensional image. These superior images are representative of a higher degree of precision accuracy in a virtual context. As a result, a more accurate preoperative plan with minimal errors could be constructed. Since this method would increase awareness of the surgeon and thereby assist precision navigation and protection of surrounding structures, it is particularly useful for complex surgical cases. For example, this methodological protocol was applied to a patient with deep infiltrating endometriosis (DIE) by Ajao et al.<sup>77</sup>

A similar approach could also be used when introducing simulation studies for trainee surgeons to advance their skills during their clinical training programmes. Augmented reality is another common AI application used in simulation studies to reconstruct objects in a real-world setting. This step generates more informative images for trainees to focus on and will reduce the risk mitigation protocols needed during their surgical training. Yet another contribution from the AI world in this niche is computer-generated holography (CGH) that can aid in providing continuous depth sensation. Naturally, such adaptations in a surgical training programme, will reduce the training time, allow targeted improvement of skills of trainees and facilitate objective appraisal and selection of the trainees, while reducing risks to the patients. As a result of these types of training programmes, indemnity insurance, which healthcare organizations are required to put in place, could cost less as the individual organizations can produce credible data on the quality of the training provided according to a robust strategy for minimizing surgical risk. The cost savings made from such training programmes could be reverted into creating better infrastructure for safe and efficient delivery of surgeries related to women's health.<sup>1,59,60,78–84</sup>

The use of AI in the operating room set up will need strong AI systems that are more complex and complicated akin to self-driving cars. This form of general AI does not currently exist for clinical application. Strong AI systems should be capable of demonstrating human intelligence coupled with soft skills possessed by physicians. Ability to reason, solve problems, make judgements and communicate while demonstrating consciousness and self-awareness is the Holy Grail to developing strong AI.

**Endometrial cancer.** Endometrial cancer is the most common gynaecological cancer and there are clinical and pathological features such as pelvic and para-aortic node metastasis, MI which reduces the survival rate of the women.<sup>85</sup> Accurate clinical staging of endometrial cancer will aid in efficient and accurate stratification of patients for the best surgical treatment or for adjuvant therapy. Therefore, identification of the high-risk patients is important, since only those can be allocated to have more extensive surgery or further adjuvant treatment, and thus, their survival chances will be improved while not disadvantaging the low-risk patients with unnecessary invasive treatments and the associated risks.

Lymph node metastasis reduces survival in women with endometrial cancer and that risk is ameliorated by the addition of lymph node resection to the hysterectomy as their surgical treatment. However, it is a procedure associated with significant additional surgical morbidity; therefore, efficient and accurate stratification of patients for the more extensive surgery (lymph node dissection) with preoperative identification of women who have lymph node involvement will improve survival while reducing the surgical risk and morbidity. This is particularly true for up to 22% of women who have low-grade primary cancers but have lymph node metastasis. Ahsen et al.<sup>86</sup> developed a  $\ell_1$ -,  $\ell_2$ -norm SVM algorithm to carry on an ML classifier on a high-dimensional data set (213 micro-RNA features with only 86 samples) to predict the metastasis in endometrial cancer. Note that the traditional  $\ell_2$ -norm does not offer a sparse solution and  $\ell_1$ -norm and Elastic Net regularization both have a poor performance while the features are far more than the sample size, the SVM with convex linear combination of  $\ell_1$ - and  $\ell_2$ -norm as penalty provides a classification method that could not only handle the high-dimensional data set but also guarantee the sparsity of the solution. Using this novel algorithm, 18 discriminatory micro-RNAs are detected from 213 micro-RNAs and achieved 100% accuracy on the training cohort. When applied to an independent testing cohort, the classifier correctly predicted 90% of node-positive cases, and 80% of node-negative cases with a 6.25% false discovery rate (FDR).

MI also increases the clinical stage of the endometrial cancer, which may be associated with distant spread and poor survival. Chen et al.<sup>46</sup> pointed out that women with more than 50% MI have an increased risk of lymph node involvement therefore may also require lymph node resection in addition to hysterectomy as their surgical therapy. This study included a training set of 500 MRI images of women with endometrial cancer, to establish a two-stage DL diagnostic model to facilitate the clinicians to diagnose MI. Combined accuracy of the radiologists and trained network model was 86.2% in determining deep MI, with authors presenting it to be a time-efficient diagnostic pathway.

## AI in MH

MH is a highly complex field and psychiatric disorders can be challenging to diagnose and treat. As a result of this, the sequential relationship of MH symptoms in OB/GYN conditions can be especially difficult. Therefore, a simpler, delineated method to ‘unpick’ any symbiotic relationship would help understand the mechanisms, compounding and perpetuating factors. To do this, the use of self-reporting tools that could aid clinical decision-making to diagnose, monitor and treat any ongoing issues in a more personalized manner could be naturally useful. Such applications are becoming more common now with the advancement of digital medicine. The acceptance and tolerability of treatments as a result having more personalized approaches could provide improved patient-reported outcomes. Therefore, in this regard, developing AI tools could offer a pragmatic and cost-effective solution.

MH diagnosis usually involves binary classification problems (such as high risk of violence and low risk of violence) or multi-class problems (such as highly stressed level, stressed level and normal level). Both can be managed in ML models or DL algorithms. The diagnosis of psychiatric disorders is often supported using criteria from the *Diagnostic and Statistical Manual of Mental Disorders Association*<sup>87</sup> or the *International Classification of Diseases*.<sup>88</sup> In order to address the shortage of psychiatrists and psychologists and enhance the efficiency of the decision-making, some ML methodologies are applied to formulate a model to predict anxiety and depression.<sup>24</sup> Based on literature review, 16 factors were considered as predictors and added to an interview questionnaire. Five kinds of ML strategies were applied on the data among 470 sailors with these features to identify the at-risk seafarers for early referral to psychological treatment. For this binary classification problem, Catboost, an ensemble ML model, appeared to be the best one with accuracy and precision of 82.6% and 84.1%, respectively. In addition, Morel et al.<sup>25</sup> and Sanderson et al.<sup>89</sup> focused on comparing the performance of LR and XGBoost on binary classification problems. The AUC was applied to evaluate the reliability of the ML methods and it shows that XGBoost outperforms the LR on hospital re-admission data and death-by-suicide data. Another study combined the ML mechanism and NLP to carry out an assessment scheme to evaluate the risk for school violence. By NLP technologies, different types of linguistic features were extracted from the interview that was fed to ML classifiers to make a prediction on the risk of school violence for each student. The use of linguistic features significantly improved ML classifiers’ predictive performance with 94.6% AUC on test data.<sup>20</sup> For multi-nomial classification problem, Ahuja et al.<sup>39</sup> extracted the features from Perceived Stress Scale (PSS) test which includes 14 questions overall and involves Random Forest, Naive Bayes, SVM and KNN to analyse

stress in the college students. The students were divided into three categories: highly stressed, stressed and normal. The result showed that SVM outperforms the other four methods with increased accuracy (85.71%).

According to a World Health Organization (WHO) report, wide arrays of MH disorders are present but often underdiagnosed in women. The global MH disease burden is approximately 790 million, with 41% attributed to depressive disorders among women.<sup>90</sup> Currently, the proportion of this sample that has existing gynaecological and/or obstetric conditions is unclear due to limitations in multi-morbidity research and study design (e.g. limited sample sizes and cross-sectional methods) within these specialties. Another rate-limiting factor could be the continued stigmatization around MH disorder which may prevent patients with these issues from presenting to services. Digital medicine and PM approaches combined with AI could provide methods to identify prevalence data more comprehensively<sup>91</sup> and provide interventions at a fraction of the cost. Using PM methods, patients would be able to self-report and transmit this data to a clinician using digital solutions such as a simple mobile application. These applications could also be personalized in a manner that allows patients to become more engaged and as a result, proactively seek clinical assistance using digital MH or telepsychiatry methods. Chat bots are a good example of this method where patients could have conversations through a digital interface. This method was used by the creators of the app Woebot, where users could express their emotions by having therapeutic discussions through the digital interface.<sup>92</sup> This was originally developed as a Cognitive Behavioural Therapy (CBT) platform as part of the UK's National Health Service (NHS) to assist patients manage their MH through an independent 'think and behave' concept. Woebot was built using NLP methods with the assistance of clinicians and patients who discussed their intended therapeutic experience. However, Woebot did not complete a Clinical Trial; thus, it has limited usage although over 1000 downloads from the Google app store have been reported.

Another AI-based tool used in MH is Wysa, where Google play store reported over 100,000 downloads. Wysa was developed by researchers in Columbia and Cambridge Universities using an AI-based emotional intelligence concept to help their users with processing emotions and thoughts via a multitude of tools and techniques such as evidence-based CBT, dialectical behavioural therapy and meditation.<sup>93</sup> However, this application is also yet to be tested for routine clinical use, therefore, currently lacks clinical trial data to justify its applicability to a wider audience. These 'telemedicine' methods have been considered by users, clinicians, clinical trialists and academics with much more enthusiasm during the COVID-19 pandemic, as the use of digital techniques has been key to keeping some clinical services open. The COVID-19 pandemic

also encouraged policy-makers and health care systems to streamline their services using innovative methods and to re-think their care delivery models where possible. E-consultations are now commonly used in a variety of specialisms, including MH. There are a variety of telepsychiatry research projects running globally, to increase access to MH care providers and improve patient benefit. However, there are issues that still need to be addressed, such as access to digital tools, especially among the older population, the availability of encrypted software to protect cloud-based systems' data storage, to name a few.

Despite the positive steps, a significant challenge for AI scientists in developing clinically relevant applications remains the lack of large and representative data sets, access to developing clinical trials and NLP for text and speech in multiple languages. For example, current AI application's ability to learn the nuances primarily of the English language is largely dependent on users themselves. This could be problematic when generalizing these tools to global healthcare systems.<sup>94</sup> Straw and Callison-Burch<sup>95</sup> performed a literature review on use of NLP for MH and observed that most of the current solutions have significant bias with respect to religion, race, gender, nationality, sexuality and age. In the development of technology to track mood, tracking body language and facial recognition, as well as verbal communication, maybe vital. These applications are also not usually accessed by clinicians themselves, but these data could be incredibly useful to formulate clinical management plans or to determine if there are any underlying conditions requiring more clinical interventions.

Due to these factors, the use of AI applications within the clinical MH domain still remains a futuristic ambition. However, many clinicians are starting to consider the potential benefits they could bring.

## Conclusion

This rapid review indicates that AI could be extremely valuable within the OB/GYN and MH domains. Although there is available evidence of AI being used in these individual domains in silos, the potential use of AI in the combined OB/GYN-MH sequelae is yet to be explored. Using multi-morbid big data sets would further improve the AI applications and its use in a clinical setting, provided they are tested and validated sufficiently using clinical trials. It is equally important to have multi-morbid data sets comprising clinical trial and real-world data sets as well as, be inclusive of different populations, race and gender, as these could influence patient-reported outcomes and possible personalization of treatments where comorbidities are involved. Failure to take these into account when developing AI applications especially within areas such as OB/GYN and MH may lead to the underestimation of patient risk, adversely influence the ethical and legal implications, limiting its potential application at a global

scale. With the evolving progress made in AI across various industries and the cross-disciplinary approach based in computer science, psychology and medicine, AI is a disruptive technology innovation that leads to ground-breaking tools, thus needs serious consideration. Reassuringly, given that a plethora of AI applications has been developed thus far using various ML methods, it appears to be accepted within the medical and scientific community. However, the dilemma remains that these have had limited testing, including the ones currently used in clinical practice. As such, to use AI applications widely, to its full potential, clinical trials are required to demonstrate the true efficacy and reliability within generalizable populations. The pervasive application of AI-based tools might support physicians with patient care by predicting accurate treatment outcomes while minimizing adverse events. The objectivity of AI systems, propelled by the enhanced data processing capabilities, can also contribute to minimizing bias in clinical care.

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

### ORCID iDs

Gayathri Delanerolle  <https://orcid.org/0000-0002-9628-9245>

Suchith Shetty  <https://orcid.org/0000-0001-5541-0953>

### Supplemental material

Supplemental material for this article is available online.

### References

- Girasa R. AI as a disruptive technology. In: Girasa R (ed.) *Artificial intelligence as a disruptive technology*. New York: Springer. 2020, pp. 3–21.
- Vincent-Lancrin S and van der Vlies R. Trustworthy artificial intelligence (AI) in education: Promises and challenges. *OECD Education Working Papers, No. 218*. Paris: OECD Publishing, 2020, <https://www.oecd.org/education/trustworthy-artificial-intelligence-ai-in-education-a6c90fa9-en.htm#:~:text=Reaching%20the%20full%20potential%20of,biases%20against%20individuals%20or%20groups>.
- Zemmar A, Lozano AM and Nelson BJ. The rise of robots in surgical environments during COVID-19. *Nat Mach Intell* 2020; 2: 566–572.
- Combi C. Editorial from the new editor-in-chief: artificial intelligence in medicine and the forthcoming challenges. *Artif Intell Med* 2017; 76: 37–39.
- Hasanpoor-Azghdy SB, Simbar M and Vedadhir A. The emotional-psychological consequences of infertility among infertile women seeking treatment: Results of a qualitative study. *Iran J Reprod Med* 2014; 12(2): 131–138.
- Bergink V, Laursen TM, Johannsen BM, et al. Pre-eclampsia and first-onset postpartum psychiatric episodes: a Danish population-based cohort study. *Psychol Med* 2015; 45(16): 3481–3489.
- Pope CJ, Sharma V, Sharma S, et al. A systematic review of the association between psychiatric disturbances and endometriosis. *J Obstetric Gynaecol Canada* 2015; 37(11): 1006–1015.
- Das P, Naylor C and Majeed A. Bringing together physical and mental health within primary care: a new frontier for integrated care. *J R Soc Med* 2016; 109(10): 364–366.
- Mbarak B, Kilewo C, Kuganda S, et al. Postpartum depression among women with pre-eclampsia and eclampsia in Tanzania; a call for integrative intervention. *BMC Pregnant Childbirth* 2019; 19: 270.
- Clemens S. Better patient outcomes through mining of biomedical big data. *Front ICT* 2018; 5: 30.
- Ye C, Li J, Hao S, et al. Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. *Int J Med Inform* 2020; 137: 104105.
- Prasad DK. Survey of the problem of object detection in real images. *Int J Image Proces* 2012; 6: 441–466.
- Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE* 1998; 86(11): 2278–2324.
- Russakovsky O, Deng J, Krause M, et al. ImageNet Large Scale Visual Recognition Challenge. *ICJCV*, 2015, <https://www.image-net.org/challenges/LSVRC/2015/index.php>
- He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27–30 June 2016, pp. 770–778. New York: IEEE.
- Edgcomb JB, Thiruvalluru R, Pathak J, et al. Machine learning to differentiate risk of suicide attempt and self-harm after general medical hospitalization of women with mental illness. *Med Care* 2021; 59: S58–S64.
- Green EP, Lai Y, Pearson N, et al. Expanding access to perinatal depression treatment in Kenya through automated psychological support: Development and usability study. *JMIR Form Res* 2020; 4(10): e1789.
- Newson JJ, Hunter D and Thiagarajan TC. The heterogeneity of mental health assessment. *Front Psychiatry* 2020; 11: 76.
- Elsworth GR, Nolte S and Osborne RH. Factor structure and measurement invariance of the health education impact questionnaire: does the subjectivity of the response perspective threaten the contextual validity of inferences. *SAGE Open Med* 2015; 3: 2050312115585041.
- Ni Y, Barzman D, Bachtel A, et al. Finding warning markers: leveraging natural language processing and machine learning technologies to detect risk of school violence. *Int J Med Inform* 2020; 139: 104137.
- Little B, Alshabrawy O, Stow D, et al. Deep learning-based automated speech detection as a marker of social functioning in late-life depression. *Psychol Med* 2020; 2020: 1–10.
- Malley B, Ramazzotti D and Wu JT. Data pre-processing. In: MIT Critical Data (ed.) *Secondary analysis of electronic health records*. Cham: Springer, 2016, pp. 115–141.

23. Li S and Ping W. Study on the data preprocessing of the questionnaire based on the combined classification data mining model. In: *2009 international conference on e-learning, e-business, enterprise information systems, and e-government*, Hong Kong, China, 5–6 December 2009, pp. 217–220. New York: IEEE.
24. Sau A and Bhakta I. Screening of anxiety and depression among seafarers using machine learning technology. *Inform Med Unlocked* 2019; 16: 100228.
25. Morel D, Kalvin CY, Liu-Ferrara A, et al. Predicting hospital readmission in patients with mental or substance use disorders: A machine learning approach *Int J Med Inform* 2020; 139: 104–136.
26. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc: Series B* 1996; 58(1): 267–288.
27. Zou H and Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc: Series B* 2005; 67(2): 301–320.
28. Efron B, Hastie T, Johnstone I, et al. Least angle regression. *Ann Stat* 2004; 32(2): 407–499.
29. Hastie T, Friedman J and Tibshirani R. *The elements of statistical learning* (Springer Series in Statistics). New York: Springer, 2001.
30. Samek W, Wiegand T and Müller KR. *Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models*. New York: Springer, 2019.
31. Chakraborty S, Tomsett R, Raghavendra R, et al. Interpretability of deep learning models: a survey of results. In: *IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (Smartworld/scalcom/uic/atc/cbdcom/iop/sci)* San Francisco, CA, 4–8 August 2017, pp. 1–6. New York: IEEE.
32. Bishop CM. *Pattern recognition and machine learning*. New York: Springer, 2006.
33. Tikhonov AN and Arsenin VY. Solutions of ill-posed problems. *New York* 1977; 1: 30.
34. Shi JQ and Choi T. *Gaussian process regression analysis for functional data*. Boca Raton, FL: CRC Press, 2011.
35. Boser BE, Guyon IM and Vapnik VN. A training algorithm for optimal margin classifiers. In: *5th annual ACM workshop on computational learning theory*, Pittsburgh, PA, 27–29 July 1992, pp. 144–152. New York: ACM Press.
36. Alty SR, Millasseau SC, Chowieniczyc PJ, et al. Cardiovascular disease prediction using support vector machines. In: *46th Midwest symposium on circuits and systems*, Cairo, Egypt, 27–30 December 2003, pp. 376–379. New York: IEEE.
37. Ahmad AR, Khalia M, Viard-Gaudin C, et al. Online handwriting recognition using support vector machine. In: *2004 IEEE region 10 conference TENCN 2004*, Chiang Mai, Thailand, 24 November 2004, pp. 311–344. New York: IEEE.
38. Zhang S and Qiao H. Face recognition with support vector machine. In: *IEEE international conference on robotics, intelligent systems and signal processing*, Changsha, China, 8–13 October 2003, pp. 726–730. New York: IEEE.
39. Ahuja R and Banga A. Mental stress detection in university students using machine learning algorithms. *Procedia Comput Sci* 2019; 152: 349–353.
40. Warikoo N, Chang YC, Dai HJ, et al. An ensemble neural network model for benefiting pregnancy health stats from mining social media. In: *Asia information retrieval symposium*, Taipei, Taiwan, 28–30 November 2018.
41. Fergus P, Selvaraj M and Chalmers C. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using cardiotocography traces. *Comput Biol Med* 2018; 93: 7–16.
42. Rumelhart DE, Hinton GE and Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; 323: 533–536.
43. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958; 65(6): 386–408.
44. Luo YH, Xi IL, Wang R, et al. Deep learning based on MR imaging for predicting outcome of uterine fibroid embolization. *J Vasc Interv Radiol* 2020; 31(6): 1010–1017.
45. Khosravi P, Kazemi E, Zhan Q, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit Med* 2019; 2: 21–29.
46. Chen X, Wang Y, Shen M, et al. Deep learning for the determination of myometrial invasion depth and automatic lesion identification in endometrial cancer MR imaging: a preliminary study in a single institution. *Eur Radiol* 2020; 30(9): 4985–4994.
47. Wolpert DH and Macready WG. No free lunch theorems for optimization. *Trans Evol Comp* 1997; 1: 67–82.
48. Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993; 39(3): 424–432.
49. Adeodato PJ and Melo SB. On the equivalence between Kolmogorov–Smirnov and ROC curve metrics for binary classification. *arXiv*, 2016, <https://arxiv.org/abs/1606.00496v1>
50. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 1997; 30(7): 1145–1159.
51. De Choudhury M, Counts S and Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Paris, 27 April 2013, pp. 3267–3276. New York: ACM.
52. Savage N. How AI and neuroscience drive each other forwards. *Nature* 2019; 571(7766): S15–S17.
53. Tulchinsky T, Varavikova E and Bickford J. *The new public health*. 3<sup>rd</sup> ed. New York: Elsevier, 2014.
54. Sobia R. Artificial intelligence for genomic medicine The PHG and Foundation, 2020, <https://www.phgfoundation.org/documents/artificial-intelligence-for-genomic-medicine.pdf>
55. ACOG Technology Assessment in Obstetrics Gynecology No. 14: Modern Genetics in Obstetrics and Gynecology. *Obstet Gynecol* 2018; 132(3): e143–e168.
56. Beam AL and Kohane IS. Big data and machine learning in health care. *JAMA* 2018; 319: 1317–1318.
57. Ahmed Z, Mohamed K, Zeeshan S, et al. Artificial intelligence with multifunctional machine learning platform development for better healthcare and precision medicine. *Database*, 2020, <https://academic.oup.com/database/article/doi/10.1093/database/baaa010/5809229>

58. Rodriguez F, Scheinker D and Harrington RA. Promise and perils of big data and artificial intelligence in clinical medicine and biomedical research. *Circ Res* 2018; 123: 1282–1284.
59. Melcher J, Hays R and Torous J. Digital phenotyping for mental health of college students: a clinical review. *Evidence-Based Mental Health* 2020; 23: 161–166.
60. Makary MA and Daniel M. Medical error – the third leading cause of death in the US. *BMJ* 2016; 353: i2139.
61. Ritchie MD, de Andrade M and Kuivaniemi H. The foundation of precision medicine: integration of electronic health records with genomics through basic, clinical, and translational research. *Front Genet* 2015; 6: 104.
62. Elenko E, Underwood L and Zohar D. Defining digital medicine. *Nat Biotechnol* 2015; 33: 456–461.
63. Collins FS and Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; 372: 793–795.
64. Meskó B, Drobni Z, Bényei É, et al. Digital health is a cultural transformation of traditional healthcare. *Mhealth* 2017; 3: 38.
65. Dongarra J, Meuer M, Simon H, et al. Top 500 list, 2017, <http://www.top500.org/lists/2017/11>
66. Davidson L and Boland MR. Enabling pregnant women and their physicians to make informed medication decisions using artificial intelligence. *J Pharmacokinet Pharmacodyn* 2020; 47(4): 305–318.
67. Fergus P, Cheung P, Hussain A, et al. Prediction of pre-term deliveries from EHG signals using machine learning. *PLoS ONE* 2013; 8(10): e77154.
68. Lu Y, Qi Y and Fu X. A framework for intelligent analysis of digital cardiotocographic signals from IOMT-based foetal monitoring. *Future Gener Comput Syst* 2019; 101: 1130–1141.
69. Idowu IO, Fergus P, Hussain A, et al. Artificial intelligence for detecting preterm uterine activity in gynecology and obstetric care. In: *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*, Liverpool, 26–28 October 2015, pp. 215–220. New York: IEEE.
70. Fele-Zorz G, Kavsek G, Novak-Antolic Z, et al. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Med Biol Eng Comput* 2008; 46(9): 911–922.
71. Despotovic D, Zec A, Mladenovic K, et al. A machine learning approach for an early prediction of preterm delivery. In: *2018 IEEE 16th International symposium on intelligent systems and informatics*, Subotica, 13–15 September 2018.
72. Zaninovic N, Khosravi P, Hajirasouliha I, et al. Assessing human blastocyst quality using artificial intelligence (AI) convolutional neural network (CNN). *Fertil Steril* 2018; 110(4): 89.
73. Chen M, Wei S, Hu J, et al. Does time-lapse imaging have favorable results for embryo incubation and selection compared with conventional methods in clinical in vitro fertilization? A meta-analysis and systematic review of randomized controlled trials. *PLoS ONE* 2017; 12(6): e0178720.
74. Santos P and Cunha TM. Uterine sarcomas: clinical presentation and MRI features. *Diagn Interv Radiol* 2015; 21(1): 4–9.
75. Lawlor H, Ward H, Maclean A, et al. Developing a pre-operative algorithm for the diagnosis of uterine leiomyosarcoma. *Diagnostics* 2020; 10(10): 735.
76. Malek M, Gity M, Alidoosti A, et al. A machine learning approach for distinguishing uterine sarcoma from leiomyomas based on perfusion weighted MRI parameters. *Eur J Radiol* 2019; 110: 203–211.
77. Ajao MO, Clark NV, Kelil T, et al. Case report: three-dimensional printed model for deep infiltrating endometriosis. *J Minim Invasive Gynecol* 2017; 24(7): 1239–1242.
78. Firmin S. It's a no brainer: an introduction to neural networks, 2020, <https://community.alteryx.com/t5/Data-Science/It-s-a-No-Brainer-An-Introduction-to-Neural-Networks/ba-p/300479>
79. Desai GS. Artificial intelligence: the future of obstetrics and gynecology. *J Obstet Gynaecol India* 2018; 68(4): 326–327.
80. Kent J. Artificial intelligence success requires partnership, training, 2020, <https://healthitanalytics.com/news/klas-artificial-intelligence-success-requires-partnership-training>
81. Bresnick J. Artificial intelligence in healthcare spending to hit \$36B, <https://healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b> (26 January 2020).
82. University Hospital Grenoble; TIMC-IMAG; University Grenoble Alps et al. Computer assisted fetal monitoring – obstetrics (surfao-obst), 2020, <https://clinicaltrials.gov/ct2/show/NCT03857126>
83. Moawad G, Tyan P and Louie M. Artificial intelligence and augmented reality in gynecology. *Curr Opin Obstet Gynecol* 2019; 31(5): 345–348.
84. Waran V, Narayanan V, Karupiah R, et al. Utility of multimaterial 3D printers in creating models with pathological entities to enhance the training experience of neurosurgeons. *J Neurosurg* 2014; 120(2): 489–492.
85. Kamel A, Tempest N, Parkes C, et al. Hormones and endometrial carcinogenesis. *Horm Mol Biol Clin Invest* 2016; 25(2): 129–148.
86. Ahsen ME, Boren TP, Singh NK, et al. Sparse feature selection for classification and prediction of metastasis in endometrial cancer. *BMC Genomics* 2017; 18: 233.
87. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. Washington, DC: American Psychiatric, 2013.
88. World Health Organization; International Classification of Diseases. ICD 11 for Mortality and Morbidity Statistics, 2018, <https://icd.who.int/browse11/l-m/en>
89. Sanderson M, Bulloch AG, Wang J, et al. Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning. *Eclinicalmedicine* 2020; 20: 100281.
90. Ritchie H and Roser M. Mental Health Published online at OurWorldInDataorg, 2018, <https://ourworldindata.org/mental-health>

91. Urteaga I, McKillop M and Elhadad N. Learning endometriosis phenotypes from patient-generated data. *Npj Digit Med* 2020; 3: 88.
92. Fitzpatrick KK, Darcy A and Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety Using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* 2017; 4(2): 7785.
93. Inkster B, Sarda S and Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 2018; 6(11): e12106.
94. Chen IY, Szolovits P and Ghassemi M. Can AI help reduce disparities in general medical and mental health care? *AMA J Ethics* 2019; 21(2): 167–179. DOI: 10.1001/amajethics.2019.167
95. Straw I and Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE* 2020; 15(12): e0240376.
96. Zaninovic N, Rocha C, Zhan Q, et al. Application of artificial intelligence technology to increase the efficacy of embryo selection and prediction of live birth using human blastocysts cultured in a time-lapse incubator. *Fertil Steril* 2018; 110(4): 372–373.