PERSPECTIVE ARTICLE

Application of Machine Learning in 3D Bioprinting: Focus on Development of Big Data and Digital Twin

Jia An[1], Chee Kai Chua[2], Vladimir Mironov[3]*

[1]Singapore Centre for 3D Printing, School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
[2]Engineering Product Development, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372
[3]3D Bioprinting Solutions, 68/2 Kashirskoe Highway, Moscow, Russian Federation 115409

Abstract: The application of machine learning (ML) in bioprinting has attracted considerable attention recently. Many have focused on the benefits and potential of ML, but a clear overview of how ML shapes the future of three-dimensional (3D) bioprinting is still lacking. Here, it is proposed that two missing links, Big Data and Digital Twin, are the key to articulate the vision of future 3D bioprinting. Creating training databases from Big Data curation and building digital twins of human organs with cellular resolution and properties are the most important and urgent challenges. With these missing links, it is envisioned that future 3D bioprinting will become more digital and *in silico*, and eventually strike a balance between virtual and physical experiments toward the most efficient utilization of bioprinting resources. Furthermore, the virtual component of bioprinting and biofabrication, namely, digital bioprinting, will become a new growth point for digital industry and information technology in future.

Keywords: 3D bioprinting; Complexity; Machine learning; Big data; Digital twin

***Correspondence to:** Vladimir Mironov, 3D Bioprinting Solutions, 68/2 Kashirskoe highway, Moscow, Russian Federation 115409; vladimir.mironov54@gmail.com

Received: January 8, 2021; **Accepted:** January 18, 2021; **Published Online:** January 29, 2021

Citation: An J, Chua CK, Mironov V., 2021, Application of Machine Learning in 3D Bioprinting: Focus on Development of Big Data and Digital Twin. *Int J Bioprint*, 7(1):342. http://doi.org/10.18063/ijb.v7i1.342

1. Introduction

Recently, there is surge in scientific publications regarding the application of machine learning (ML) to bioprinting-relevant researches such as medical imaging and segmentation, optimization of bioinks or bioprinting process as well as *in vitro* parametric studies, which are well reviewed in Yu and Jiang[1], Ng *et al*.[2]. Both recent articles focused on the benefits and potential of ML but missed a clear portrait of what future bioprinting looks like. This perspective article is, therefore, written as an extension of previous reviews, focusing on a vision of future three-dimensional (3D) bioprinting enabled by ML.

ML is a collection of computational methods of discovering approximate mathematical functions of the real world based on historical data (**Figure 1**). Given a set of input (X) and output (Y) data, humans are usually able to find a relationship between X and Y as a function of Y(X). If the input has multiple variables ranging from X_0 to X_n and the output Y_0 to Y_n, humans will be easily overwhelmed by the complexities. However, computer algorithms can replace human to "inspect" the input and output and "guess" an approximate function among them. This approximate function generated by the algorithms is called ML model. The more input and output data there are, the more accurate the ML model becomes. This approach is known as mapping or supervised learning. In contrast, in grouping or unsupervised learning, the output (Y) is not given, the computer algorithms must figure out the output on its own, such as a pattern, a cluster, or a relationship in the input data (X_0, X_1,… X_n). Therefore, this approach is best for uncovering hidden patterns or relationships in data. Another approach in ML is reinforcement learning, in which both input (X_0, X_1 … X_n) and output (Y_0, Y_1… Y_n) are known, the algorithms (or agent) are

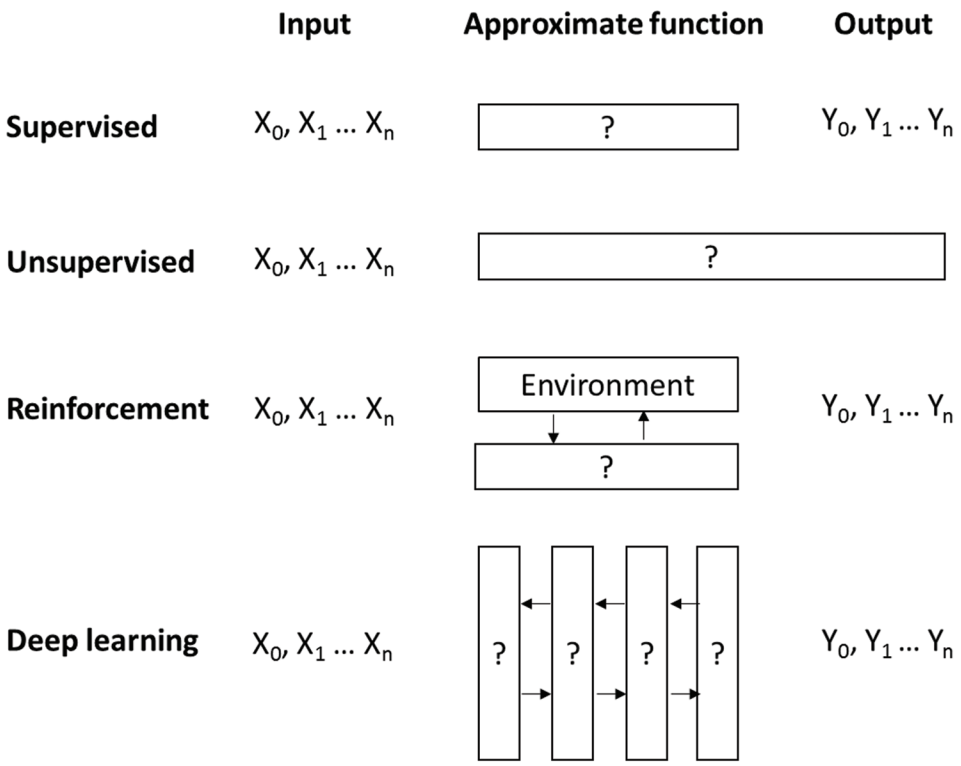

Figure 1. Example methods in machine learning.

to find the functions between X and Y like in supervised ML, but through a dynamic interaction with another test algorithm called environment. The environment rewards or punishes the agent's trial and error learning so that the ML model becomes more and more accurate in predication by maxing the reward. Another similar method is deep learning, in which the trained learning algorithms have multiple hidden layers and are always applied to new datasets instead of dynamically adjusting agent's actions from the continuous feedback. Above are example methods in ML. In fact, learning is quite a broad topic and several techniques have already been used in 3D printing in general[3,4].

2. Complexities in bioprinting

In general, ML models are preferred when complexities arise, because they can account for factors or conditions not considered in traditional mathematical models, that is, they tend to be more robust in the real-world context in terms of predication. Bioprinting coming across ML is inevitable for the reason of complexities. The complexities of bioprinting span across the entire process chain, namely, pre-processing, processing, and post-processing (**Figure 2**). In pre-processing, it is challenging to perform segmentation of tissue images at a single cell level and reconstruct them into a 3D tissue model with cross-scale cellular resolution and tissue properties. ML is, therefore, needed because of multiscale complexities of representing biological tissue models. In addition, ML can also help predict the compatibility of dissimilar materials used in bioprinting[5,6]. In processing and post-processing, it is almost impossible to perform wet experiments when the number of changing parameters exceeds a certain number, for example, ten parameters. ML is, therefore, needed because of multiparameter complexity of finding optimal protocol of bioprinting. Here, it is envisioned that ML coupled with Big Data, will solve the multiscale and multiparameter complexities and transform present 3D bioprinting into future 3D bioprinting, which is "heavily virtual" in nature.

3. *In silico* experiment and big data

In silico, experiment such as digital fabrication and computational study is believed to be a key innovation driver and play a major role in the new era of tissue engineering[7,8]. Although the current challenge is creating more realistic virtual experiment with satisfactory accuracy, there are various methods such as statistical tools and techniques that can be combined with first-principles simulations to solve it[9]. In the case of bioprinting, ML has already been used to optimize the printability of bioinks and drastically reduced the number of experiments from thousands of possible combinations[10]. Moreover, various mathematical models on bioink printability have been developed as

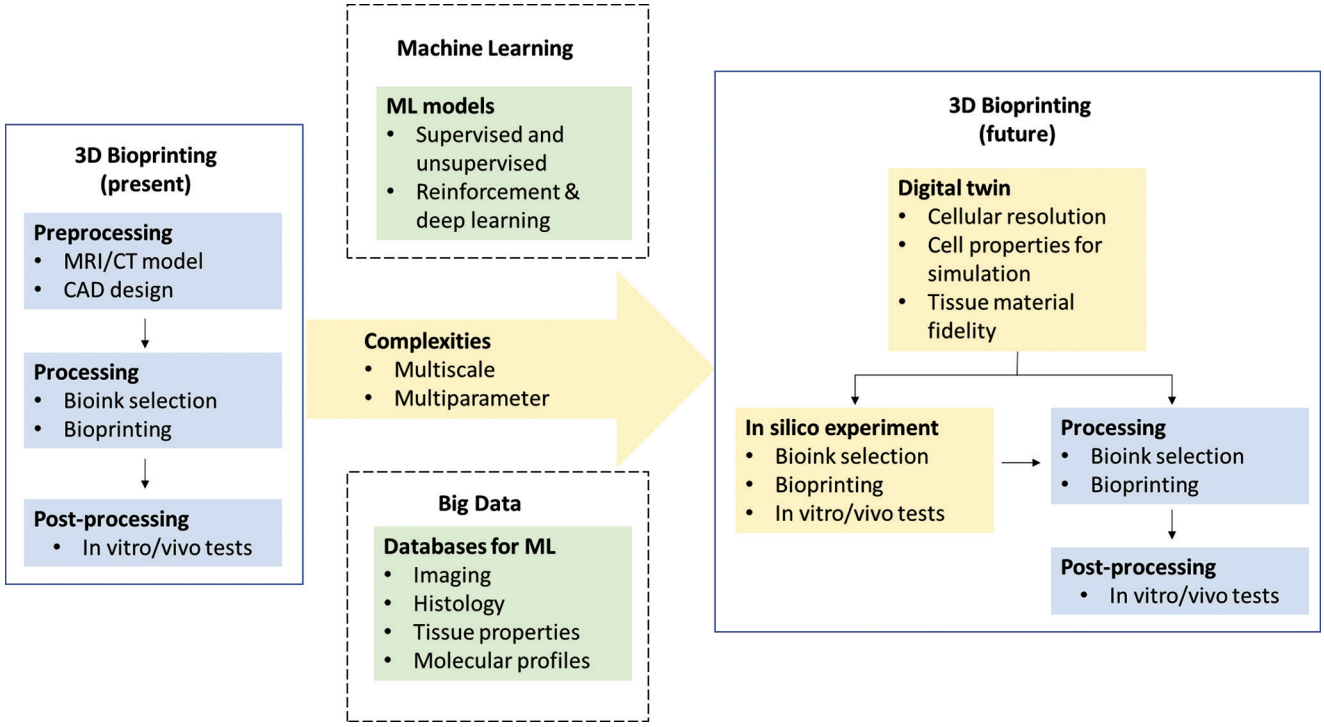

Figure 2. A vision for future bioprinting.

recently reviewed in Zhang *et al*.[11], Schwab *et al*.[12]. These mathematical models are useful for construction of virtual bioprinting process. ML models have also been used in *in vitro* study for identification of cell signature genes out of complex gene expression profiles among different cell groups[13,14]. Another *in vitro* example is virtual histological staining, which bypasses the lengthy and laborious process for tissue preparation. Researchers used deep learning to transform autofluorescence images of tissue into images equivalent to histologically stained tissue[15], and achieved blending of multiple stains by assigning each stain at the pixel level[16,17]. Furthermore, mathematical models and ML models, which help us understand the complexities of biological systems and extract new biological knowledge from complex experimental datasets, are expected to bring tissue engineering much closer to clinical reality[18]. Collectively, the above evidence of virtual experiments in either processing or post-processing of bioprinting suggests that we will see more *in silico* experiments with ML in bioprinting in future.

However, ML cannot be performed without Big Data about modern clinical imaging of organs, histology, immunohistochemistry, biomechanical properties of tissue and organs, molecular profiles of cell, tissue, and organs (genomics and proteomics) function and so on. Big Data can be structured, unstructured or semi-structured and it is much more than traditional databases[19]. For ML purposes, the first step is collection of Big Data or Big Data curation. The sources of big data for bioprinting are huge and diverse, it could be all types of diagnostic images stored in hospital databases, all types of experimental data in worldwide laboratories and research centers, all the "omics" databases already established in past years, or simply the vast scientific literature. Standard and open-access databases with meaningful and valuable training datasets specifically targeting for bioprinting must be created from Big Data curation. A recent example is the construction of a web-based nanomaterial database through Big Data curation, which contains 705 unique nanomaterials, and the annotation of nanostructures generates 2142 nanodescriptors for modeling and ML, but more importantly, the database is publicly available[20]. Another example is geoscience databases, which is large and ideal for ML and automated geoscience analysis[21]. In fact, numerous experimental data and various materials directly related to bioprinting have been generated over past years, making bioprinting a potentially a data-driven research, but so far there is limited database created specifically for bioprinting. In future, we hope to see more developments in this area, in line with the development of databases for 3D printing. Perhaps it may be even possible to predict new bioprinting discoveries by exploiting the current literature alone without relying on experts' opinions[22].

4. Digital twin of human organ

On the other hand, design process in 3D bioprinting must be organized around the concept of digital twins of organs

and virtual shadow. Creating such cell-level twins of organs requires high quality tissue specimens and advanced imaging and 3D reconstruction methods. Fortunately, the Human BioMolecular Atlas Program from Institute of Health in the United States[23], which aims to develop an open and global framework to create 3D molecular and cellular atlas of the human body, may enable the building of an integrated tissue map across scales. However, as pointed out in Campos and De Laporte[8], digital tissues should not only enable architectural replication of native tissues but also be biologically functional. This would require the capability of assigning fidelitous tissue materials to the digital twin and a profound understanding of individual and collective behaviors of cells. Cell-based mathematical models and software[24], which have been extensively used in computational biology, may be useful tools for modeling cell and tissue properties and behaviors to enable the simulation of biological functions of the virtual tissue and organs. In fact, from the economic point of view, an alternative but efficient method should be one that directly converts current magnetic resonance imaging (MRI)/confidence interval-based 3D models into cell-based models, that is, cells and tissue properties are intelligently assigned to a virtual organ model with spatial accuracy and material diversity by artificial intelligent algorithms. Slicing of the digital twin for layer-by-layer bioprinting should also be intact cell-based and matching extrusion layer thickness, which is very different from common slicing in 3D printing. Another alternative further empowering our imagination is *in vivo* cellular imaging such as MRI[25], which can map the anatomic locations of specific cells within living tissue. Given that ML has been successfully used for recognition of cell phenotype[26], it might be reasonable to imagine "*in vivo* 3D scan" of a patent-specific live tissue model into a digital twin with cellular resolution. Nevertheless, the immediate impact of the digital twins of organs on bioprinting is that the *in vivo* performance of physical bioprinting such as preclinical as well as clinical studies must collect information with specially designed assay, biosensor, and so on for updating original model in the form of digital twin. Furthermore, the cell-level digital twins together with physical 3D bioprinting could also revolutionize biology fundamentally by building tissues from scratch to explore entirely new cell configurations for cell cross-talks and cellular morphogenesis[27]. This would help provide new insights into the challenging question: "Print me an organ! Why we are not there yet?" which was recently raised in Ng *et al*.[28].

5. Other aspects of the future

In future, it is necessary to include the development of correspondent infrastructures such as education and training specialists and development and adaptation of software and computational power and so on. Interestingly, ML had been applied to nanotechnology to develop nanocomputing hardware that can boost artificial-intelligence-based applications[29]. It could be a reciprocal advancement to expect in future. Another topic worth watching is ML-based programmable design for 4D printing[30], as it is relevant to 4D bioprinting, a method in which bioprinted tissues transform shape, size, or pattern over time[31]. Aside from academy, the industry also expects a bright future for use of artificial intelligence in 3D bioprinting. For example, in 2019 Procter and Gamble partnered with a biotechnology company Aether to develop AI 3D Bioprinter[32].

6. Toward digital bioprinting

Application of ML in bioprinting and biofabrication will induce dramatic transformation and bioprinting will became a part of digital industry and information technology[33,34]. What could be done to implement these forthcoming transformations? First, bioprinting community must attract experts in computer sciences, mathematical modeling, computer simulations, and ML. Second, special efforts must be done for generation, assembly and maintaining of desirable Big Data. Maintaining and up-dating of such databases are essential. Third, digital organ twins based on sophisticated mathematical modeling and advanced software will become a new type of knowledge presentation, accumulation, and compaction in bioprinting. Finally, during transition from empiric to digital approach bioprinting will enter in digital era and it will become not descriptive but rather predictive technology increasingly based on virtual or *in silico* experiments.

7. Conclusion

In our opinion, when applying ML to bioprinting the most important and urgent challenges are: (1) To build training databases for ML from Big Data curation and (2) to build digital twins of human tissue/organs. The goal is to achieve a predictive power of digital twin of human tissue/organ based on Big Data which is close to virtual crash test in automobile industry. Ultimately, we hope to see a standard bioprinting simulation practice in future to reduce or replace present 3D bioprinting studies. We envision that future 3D bioprinting will become more digital and *in silico*, and eventually strike a balance between virtual and physical experiments to maximize the efficiency of bioprinting resource utilization. In future, digital bioprinting will become a new growth point for digital industry and information technology.

References

1. Yu C, Jiang J, 2020, A Perspective on Using Machine Learning in 3D Bioprinting. *Int J Bioprinting*, 6:95.

https://doi.org/10.18063/ijb.v6i1.253

2. Ng WL, Chan A, Ong YS, *et al*., 2020, Deep Learning for Fabrication and Maturation of 3D Bioprinted Tissues and Organs. *Virtual Phys Prototyp*, 15:340–58.
3. Meng L, McWilliams B, Jarosinski W, *et al*., 2020, Machine Learning in Additive Manufacturing: A Review. *JOM*, 72:1–15.
4. Goh G, Sing S, Yeong W, 2020, A Review on Machine Learning in 3D Printing: Applications, Potential, and Challenges. *Artif Intell Rev*, 54:63–94. https://doi.org/10.1007/s10462-020-09876-9
5. Hamid OA, Eltaher HM, Sottile V, *et al*., 2020, 3D Bioprinting of a Stem Cell-laden, Multi-material Tubular Composite: An Approach for Spinal Cord Repair. *Mater Sci Eng C*, 2020:111707. https://doi.org/10.1016/j.msec.2020.111707
6. Lee JM, Sing SL, Yeong WY, 2020, Bioprinting of Multimaterials with Computer-aided Design/Computer-aided Manufacturing. *Int J Bioprint*, 2020:245. https://doi.org/10.18063/ijb.v6i1.245
7. Geris L, Papantoniou I, 2019, The Third Era of Tissue Engineering: Reversing the Innovation Drivers. *Tissue Eng Part A*, 25:821–6. https://doi.org/10.1089/ten.tea.2019.0064
8. Campos DF, De Laporte L, 2020, Digitally Fabricated and Naturally Augmented *In Vitro* Tissues. *Adv Healthc Mater*, 2020:2001253. https://doi.org/10.1002/adhm.202001253
9. Barnard AS, 2014, *In silico* veritas. *ACS Nano*, 8:6520–5.
10. Ruberu K, Senadeera M, Rana S, *et al*., Coupling Machine Learning with 3D Bioprinting to Fast Track Optimisation of Extrusion Printing. *Appl Mater Today*, 22:100914. https://doi.org/10.1016/j.apmt.2020.100914
11. Zhang Z, Jin Y, Xu C, *et al*., 2018, Evaluation of Bioink Printability for Bioprinting Applications. *Appl Phys Rev*, 5:041304.
12. Schwab A, Levato R, D'Este M, *et al*., 2020, Printability and Shape Fidelity of Bioinks in 3D Bioprinting. *Chem Rev*, 120:11028–55. https://doi.org/10.1021/acs.chemrev.0c00084
13. Arai F, Stumpf PS, Ikushima YM, *et al*., 2020, Machine Learning of Hematopoietic Stem Cell Divisions from Paired Daughter Cell Expression Profiles Reveals Effects of aging on Self-Renewal. *Cell Syst*, 11:640–52. https://doi.org/10.1016/j.cels.2020.11.004
14. Cilloni D, Petiti J, Campia V, *et al*., 2020, Transplantation Induces Profound Changes in the Transcriptional Asset of Hematopoietic Stem Cells: Identification of Specific Signatures Using Machine Learning Techniques. *J Clin Med*, 9:1670. https://doi.org/10.3390/jcm9061670
15. Rivenson Y, Wang H, Wei Z, *et al*., 2019, Virtual Histological Staining of Unlabelled Tissue-Autofluorescence Images via Deep Learning. *Nat Biomed Eng*, 3:466. https://doi.org/10.1038/s41551-019-0362-y
16. Zhang Y, de Haan K, Rivenson Y, *et al*., 2020, Digital Synthesis of Histological Stains Using Micro-Structured and Multiplexed Virtual Staining of Label-free Tissue. *Light*, 9:1–13. https://doi.org/10.1038/s41377-020-0315-y
17. Rivenson Y, de Haan K, Wallace A, *et al*., 2020, Emerging Advances to Transform Histopathology Using Virtual Staining. *BME Front*, 2020:9647163. https://doi.org/10.34133/2020/9647163
18. MacArthur BD, Stumpf PS, Oreffo RO, 2020, From Mathematical Modeling and Machine Learning to Clinical Reality. In: Lanza R, Langer R, Vacanti J, editors. Principles of Tissue Engineering. 5th ed., Ch. 2. Academic Press, Cambridge, Massachusetts, pp. 37–51. https://doi.org/10.1016/b978-0-12-818422-6.00001-0
19. What is the Structure of Big Data? 2019. Available from: https://www.magnimindacademy.com/blog/what-is-the-structure-of-big-data#:~:text=Big%20data%20structures%20can%20be,look%20at%20them%20in%20detail. [Last accessed on 2020 Dec 25].
20. Yan X, Sedykh A, Wang W, *et al*., 2020, Construction of a Web-based Nanomaterial Database by Big Data Curation and Modeling Friendly Nanostructure Annotations. *Nat Commun*, 11:1–10. https://doi.org/10.1038/s41467-020-16413-3
21. Bergen KJ, Johnson PA, de Hoop MV, *et al*., 2019, Machine Learning for Data-driven Discovery in Solid Earth Geoscience. *Science*, 363:eaau0323. https://doi.org/10.1126/science.aau0323
22. Clauset A, Larremore DB, Sinatra R, 2017, Data-Driven Predictions in the Science of Science. *Science*, 355:477–80. https://doi.org/10.1126/science.aal4217
23. Consortium H, 2019, The Human Body at Cellular Resolution: The NIH Human Biomolecular Atlas Program. *Nature*, 574:187. https://doi.org/10.1038/s41586-019-1629-x
24. Hoehme S, Drasdo D, 2010, A Cell-based Simulation Software for Multi-Cellular Systems. *Bioinformatics*, 26:2641–2. https://doi.org/10.1093/bioinformatics/btq437
25. Rogers WJ, Meyer CH, Kramer CM, 2006, Technology Insight: *In Vivo* Cell Tracking by Use of MRI. *Nat Clin Pract*

Cardiovasc Med, 3:554–62.

26. Sommer C, Gerlich DW, 2013, Machine Learning in Cell Biology-Teaching Computers to Recognize Phenotypes. *J Cell Sci*, 126:5529–39. https://doi.org/10.1242/jcs.123604
27. Daly AC, Prendergast ME, Hughes AJ, *et al*., 2021, Bioprinting for the Biologist. *Cell*, 184:18–32. https://doi.org/10.1016/j.cell.2020.12.002
28. Ng WL, Chua CK, Shen YF, 2019, Print me an Organ! Why we are not there yet. *Prog Polym Sci*, 97:101145. https://doi.org/10.1016/j.progpolymsci.2019.101145
29. Sacha GM, Varona P, 2013, Artificial Intelligence in Nanotechnology. *Nanotechnology*, 24:452002. https://doi.org/10.1088/0957-4484/24/45/452002
30. Hamel CM, Roach DJ, Long KN, *et al*., 2019, Machine-learning Based Design of Active Composite Structures for 4D Printing. *Smart Mater Struct*, 28:065005. https://doi.org/10.1088/1361-665x/ab1439
31. An J, Chua CK, Mironov V, 2016, A Perspective on 4D Bioprinting. *Int J Bioprint*, 2:02003.
32. Jackson B, 2019, Aether to Develop Ai 3D Bioprinter in Agreement with Procter and Gamble. Available from: https://www.3dprintingindustry.com/news/aether-to-develop-ai-3d-bioprinter-in-agreement-with-procter-gamble-149405. [Last accessed on 2020 Dec 20].
33. Rezende RA, Kasyanov V, Mironov V, *et al*., 2015, Organ Printing as an Information Technology. *Proc Eng*, 110:151–8.
34. Dernowsek JD, Rezende RA, Lopes da Silva JV, 2-17, The Role of Information Technology in the Future of 3D Biofabrication. *J 3D Print Med*, 1:63–74.