Machine Learning in Stem Cells Research: Application for Biosafety and Bioefficacy Assessment

Wan Safwani Wan Kamarul Zaman, Salmah Karman, Effirul Ikhwan Ramlan, Siti-Nurainie Tukimin, Mohd Yazed Ahmad

Abstract—The applicability of machine learning-based analysis in the field of biomedical field has been very beneficial in determining the biological mechanism and validation for a wide range of biological scenarios. This approach is also gaining momentum in various stem cells research activities, specifically for stem cells characterization and differentiation pattern. The adoption of similar computational approaches to study and assess biosafety and bioefficacy risks of stem cells for clinical application is the next progression. In particular where tumorigenicity has been one of the major concerns in stem cells therapy. There are many factors influencing tumorigenicity in stem cells which may be difficult to capture under conventional laboratory settings. In addition, given the possible multifactorial etiology of tumorigenicity, defining a one-size-fits-all strategy to test such risk in stem cells might not be feasible and may compromise stem cells safety and effectiveness in therapy. Given the increase in biological datasets (which is no longer limited to genomic data) and the advancement of health informatics powered by state-of-the-art machine learning algorithms, there exists a potential for practical application in biosafety and bioefficacy of stem cells therapy. Here, we identified relevant machine learning approaches and suggested protocols intended for stem cells research focusing on the possibility of its usage for stem cells biosafety and bioefficacy assessment. Ultimately, generating models that may assist healthcare professionals to make a better-informed decision in stem cell therapy.

I. INTRODUCTION

Stem cells are undifferentiated cells found in all multicellular organisms which possess a unique self-renewal ability and multi-potential differentiation [1-3]. Stem cells have been associated with the fields of regenerative medicine and tissue engineering with the goal to improve health and quality of life, especially patients with debilitating diseases. Stem cells can be divided into three categories: (1) embryonic stem cells (ESC) derived (ESC) from early-stage embryos; (2) adult stem cells (ASC) and (3) induced pluripotent stem cells (iPSC). These cells owe its regenerative capacity to its ability to migrate to the injured part of the body, to divide and produce daughter cells, which have the ability to differentiate into other lineages of cells in order to repair the damaged tissue under appropriate conditions [4]. The ability for stem cells to induce regeneration can be influenced by culture condition and the type of secretomes released [5]. Stem cells have been studied and even applied for the treatment of various clinical conditions. However, there are risks which needed for further evaluation (before clinical application), such as miss-differentiation of cells, mis-targeting of cells, immune rejection and the biggest concern is genomic instability or tumor formation [6, 7].

While the application of stem cells for treatment is on the rise, their overall quantity in the body is scarce. Generally, cell therapy protocols require hundreds of millions of MSC per treatment and this would require cell expansion in vitro for about 10 weeks before implantation [8]. In this regard, long-term expansion or manipulation of stem cells may contribute to cellular senescence or even tumorigenesis in vitro, which may cause them to be non-viable for clinical usage. This has led to concerns of biosafety and bioefficacy of stem cells in clinical application [9, 10]. The aforementioned concerns are mainly due to poor understanding of stem cells biological mechanism, which has prevented it from being used widely in research in clinical application. Experimental approaches based on phenotypic and genotypic profiling are limited, whereby, they can also be expensive and time-consuming [11-13]. Furthermore, these approaches would also require subsequent validation assays to confirm its accuracy, which apart from the small sample size, can also lead to misinterpretation of data.

The recent development in stem cells research has shown that machine learning application can be used to overcome some of these limitations, particularly in phenotypic profiling of stem cells [14, 15]. Other potential applications that could be explored are annotation of stem cell genome [16], predictions of protein binding, identification of specific markers [17] or key transcriptional factors of stem cells and characterization of stem cells transcriptional regulatory networks [18].

There are datasets generated from experiments to quantify molecular variables related to stem cells biosafety and bioefficacy, such as the gene and proteins interactions. However, these datasets are complex with an intricate network of molecular interactions and analysis [19, 20]. To address this complexity, machine learning could provide next-level analyses.
that would allow better insights and the generation of new information for better biosafety and bioefficacy assessment [18]. As such, this would also allow medical practitioners to be better informed in offering personalized treatment to patients in stem cells therapy. A similar approach is seen in cancer research, whereby, machine learning has been widely applied in the identification and classification of cancer cells. Similar machine learning models and approaches can be applied in stem cells research [21-24] that could assist in accelerating the evaluation of safety and efficacy of stem cells. This could potentially bring stem cells to the forefront of personalized medicine. The current and future research trends in stem cells research are presented as an overview in Figure 1. In this review, the advantages and limitations of machine learning in stem cells research were presented, including on how next-generation machine learning methods could be used to expand our understanding of stem cells biology and their biosafety and bioefficacy risks. We anticipate that machine learning could have substantial impacts on stem cells research and therapy, providing a supporting tool in making a personalized clinical decision. This includes tailoring treatments for optimization in individual patients.

II. STEM CELLS BIOSAFETY AND BIOEFFICACY PROFILES

In general, stem cells are unspecialized cells with self-renewal and differentiation (into specialized cell) abilities [25]. However, each type of stem cells has different characteristics, which are attributed by their origin, biological characteristics and functionality. The ESC cells are pluripotent stem cells originated from the inner mass of blastocyst of the embryo and can give rise to the entire body tissue organs except for placenta and umbilical cord [26-28]. Meanwhile, ASC cells are somatic cells-derived from a certain part of the adult body, which can only give rise to stem cell progeny of the original site [29] in which they are found. They are known to be multipotent with limited differentiation capacity [30, 31]. Adult stem cells are also known as mesenchymal stem cells (MSCs) [32-35].

[Fig. 1 about here.]

The iPSCs are capable in giving rise to all kinds of cell types in the body but the difference is that iPSCs are reprogrammed stem cells, whereby, somatic or primary cells are biologically reprogrammed, giving rise to stem cells similar in characteristics as ESCs in culture [36-38]. Due to this technique, iPSCs have been widely used as it reduced the dependency on ESCs and ASCs, which are limited in the cell population. The iPSCs have been profoundly utilized not only for repair, replenishment and replace the damaged cell, tissue and organ but they have also been employed for drug-response therapy [39]. Due to the regenerative capacity of stem cells, they have been regarded as a powerful tool in regenerative medicine, particularly in the treatment of debilitating diseases. However, such potential and capability have given rise to other clinical concerns, such as adverse effects associated with biosafety and bioefficacy issues. These effects may not materialize immediately after receiving stem cell therapy. Post therapy monitoring may be difficult as there are no established pre- or post-parameters and further, there is no ‘one-size-fits-all’ protocol to enable such monitoring procedure. To develop a stem cell-based therapy, we must first ensure the safety and efficacy of stem cells. High efficiency of stem cells is needed to have effective homing, engraftment and persistence in damaged tissues, which would enable a stable interaction between the transplanted and the injured tissues. This is important to maximize the therapeutic capacity of stem cells.

Studies on biosafety and bioefficacy of stem cells have been carried out for many years. There have been few reports addressing the biosafety and bioefficacy profiles of stem cells [10, 40, 41], which showed the importance of addressing these issues. As of now, there is no standard or conclusive data that can be used to establish a proper protocol for biosafety and bioefficacy assessment. Further, any protocols and guidelines established would need to be internationally accepted and harmonized [42]. The proposed minimal criteria to define human mesenchymal stem cells (MSC) was established by the Mesenchymal and Tissue Stem Cell Committee of the International Society for Cellular Therapy (ISCT). These criteria are: 1) MSC must be plastic-adherent when cultured in standard culture conditions, 2) MSC must express CD105, CD73 and CD90 and lack in the expressions of CD45, CD34, CD14 or CD11b, CD79a or CD19 and HLA-DR surface molecules and 3) MSC must be able to differentiate to osteoblasts, adipocytes and chondroblasts in vitro [43].

Since then, further investigations on stem cells characteristics were carried out, which can be (directly or indirectly) used to evaluate the biosafety and bioefficacy profiles of stem cells. With regards to adult stem cells (ASC) or MSC, studies have shown that these stem cells were reported to have low risks of tumorigenicity in long-term culture [4, 44] and low risks of abnormalities following long-term cryopreservation [45]. While there are no significant changes in stem cells differentiation ability, cryopreservation caused stem cells to appear less fibroblastic in appearance [46]. Long-term culture of stem cells was also reported to alter its stemness and differentiation ability [46, 47]. On the other hand, tumorigenicity risks of embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) have been reported, which pose a hurdle in stem cell therapy [48, 49]. Furthermore, the differences in cell microenvironment and culture conditions contributed by biophysical and biochemical cues can affect stem cells response, for example, cell culture in hypoxia [50-53], use of serum [54] and fluid shear forces [55]. This evidence also showed that stem cells response is affected by both static and dynamic interventions.

Despite the successful clinical application of stem cells, the sample size is rather small, which may not be sufficient to determine the safety and efficacy of the treatment. The application of human allogeneic adipose-derived MSC showed feasibility in a pediatric patient with no adverse effects of up to 12 months following treatment [56]. While Lennmyr et al. [57] showed that all adult patients affected by lymphoblastic
leukemia was successfully treated with allogeneic hematopoietic stem cell transplantation (alloHSCT) with an increased overall survival rate of more than 10% in 5 years [57]. Meanwhile, Schlenk et al. [58] evaluated alloHSCT among patients having acute myeloid leukemia (AML) showed a significant beneficial impact after treatment. Similarly, Cornelissen et al. [59] reported that AML patients that were treated using alloHSCT have a significant beneficial impact with the overall increased survival rate of 12% and was successfully validated using cytogenetic profiling. However, there have been variations in terms of treatment outcomes and responses, which can be due to multiple factors that can be difficult to ascertain. Examples of these factors include different donors [60], age of donors [61] and different culture protocol [62], which may contribute to non-standardized outcomes and potentially adverse effects in stem cells therapy that can be irreversible. This has made it difficult to assess and to establish a standardized protocol to evaluate the biosafety and bioefficacy risks of stem cells. Therefore, machine learning-based predictive analytical methods are desirable to accelerate the discovery of new stem cell markers for safety assessment and to forecast stem cell therapy efficacy in order to minimize the potential adverse effects and to maximize the success of treatment.

III. LIMITATIONS IN STEM CELL THERAPY WHICH POSE EFFICACY AND SAFETY ISSUES

Although suitable stem cells safety and efficacy profiles and assessment are still not well established, stem cells have been used for various disease treatments. Stem cell therapies have been applied for the treatment of anemia [63], multiple myeloma [64-66], arthritis [67, 68] and even stroke [69, 70]. Stem cells therapies have also been applied for blood-related cancers, whereby, patients have undergone allogeneic hematopoietic stem cell transplantation (alloHSCT) for thalassemia [71] and acute myeloid leukemia (AML) [72-75]. In some cases, determining post-treatment efficacy and safety of stem cell therapy may be restricted due to the difficulty in following up after treatment. It might also be due to stem cells’ dynamic responses in different individual recipient and disease models, which lead to variations in the outcome of the treatment. In this regard, there have been reports addressing post-treatment complications of stem cells therapy within a year following the treatment.

Rovo and Tichelli [76] reported cardiovascular complication risk following allogeneic hematopoietic stem cell transplantation leading to considerable morbidity and mortality, including patients having critical diseases, such as dyslipidemia, arterial hypertension, diabetes mellitus and kidney disease. There have been reports of undesired differentiation and malignant transformation [77] as well as the ability to promote tumor growth and metastases, which has been a major concern in stem cells therapy [78]. Patients who have undergone autologous stem cell transplantation (ASCT) for lymphoma have a significant risk of developing therapy-related acute myeloid leukemia [74, 79, 80] and myelodysplasia (t-AML/MDS) [81]. This may be attributed by the ASCT procedure that includes priming chemotherapy, total body irradiation and the extensive cellular proliferative, which occur during engraftment, leading to the development of t-AML/MDS. Graft-Versus-Host Disease (GVHD) is another adverse effect that occurs following stem cell treatment [82-84], which can be overcome by the use of a mismatched allograft that necessitates T cell depletion. Surprisingly, a greater HLA mismatch was associated with a lower risk of GVHD [85]. The mismatched donor lymphocyte infusion (DLI) was specifically created for prophylactic treatment of T cell depleted mismatched allograft recipient [85].

All of these stem cells donors and recipients’ responses may be important factors needed to be considered carefully. Biological assays and genetic molecular expression data profiling may be able to overcome such limitations and challenges, but they may be cost-prohibitive and time-consuming. Antibiotic matching, biomarker details and signaling pathways are all essential information needed but may require subsequent validation assays for accuracy. In this regard, dependency on biological assays may lead to the misinterpretation of data, particularly in terms of the similarity in biomarkers and molecular signaling pathways of various microenvironment and disease models. Hence, machine learning-based characterization and classification profiling techniques may be able to capture the genotypic and phenotypic differences as well as the changes that occur in a shorter period with more accuracy in terms of safety and efficacy of the stem cells.

IV. MACHINE LEARNING APPLICATION OPPORTUNITY IN STEM CELLS BIOSAFETY AND BIOEFFICACY EVALUATION

Understanding stem cells behavioral response and changes have been mainly carried out through biological assays that employed time-consuming and laborious methods [86, 87]. Furthermore, just like any biomedical datasets, stem cells datasets are generally limited by sample size [88]. To increase data for analysis, some investigators used 3D printing to create bio-scaffolds to mimic the natural environment of stem cells, but this approach was not always successful, whereby, stem cells often migrated away from the printed scaffolds or locations. Hence, the analysis of stem cells’ intrinsic ability and response were not always accurate. To overcome this limitation, the machine learning approach to study stem cells’ complexity is gaining momentum, particularly on the aspect of molecular and genomic changes in pluripotent stem cells.

Machine learning is a common method in data analytics for identification and recognition of patterns, which when applied to stem cell biology, will enable the discoveries of new insights with reasonable accuracy in a shorter amount of time. This approach would also be of advantage as stem cells are known to form predictable patterns in their natural environment as they mature into tissues. Such analytics are difficult to duplicate in the lab, costly to perform, laborious and time consuming to execute. For instance, Libby et al. [89] used extended cellular Potts model to capture pluripotent stem cell organization.
dynamics that enabled them to demonstrate morphogenetic dynamics through a model-driven exploration of stem cells behaviors, which is a vital step in organ modelling [89]. Understanding that the individual stem cell is different even if they are genetically cloned has led to the Allen Cell Explorer produced by the Allen Institute for Cell Science in Seattle, USA. The Allen Cell Explorer, which complements various ongoing projects, is an online catalogue, including the 3D images of stem cells as well as the iPSC that were produced using deep learning analysis and cell lines altered with the gene-editing tool, CRISPR [90]. It is also a growing library that charts the uniqueness of single cells at DNA, RNA and protein levels [90]. This gives a more holistic and unbiased approach to predict and understand multiple aspects of cellular structure and behaviors.

Although machine learning application in stem cells research is not a commonality at this stage, a proof-of-concept study has been presented previously [91], Zhang et al. [91] employed machine learning and microscopic image analysis to identify iPSC progenitor cells in their effort to understand the origin and underlying mechanism of iPSC particularly at the early stage of cell reprogramming, including the biomarkers involved. However, the proposed model by Zhang et al. [91] showed inconsistencies in their prediction with large fluctuation. The model can only predict iPSC progenitor cells with a minimum precision of 52%. The model is incapable of handling additional iPSC features and phases to achieve higher accuracy of the prediction performance.

Machine learning models have been applied in cancer diagnostics and prognostics [15, 92-94], whereby, similar predictions and interpretation models can be applied in understanding stem cells, specifically for its biosafety and bioefficacy evaluation. Similar machine learning techniques can be used in predicting and studying the dynamic changes of stem cells behavior in a particular environment, which should be directed towards understanding its impact on biosafety and bioefficacy for clinical application. The combination of the dataset from cancer research and stem cells research for various machine learning models should be considered as an approach in understanding stem cells behavior and interactions as well as the risks of developing adverse effects following therapy.

1. Image-based Dataset for Machine Learning in Stem Cells Research

The common workflow of image processing consists of the preparation of image input data, pre-processing, segmentation, feature extraction and classification steps. The microscopic image should be prepared by capturing the cell image from cells or tissue samples using the microscopic digital camera or software. Technical expertise for manual classification may be required for image labelling, especially for images that will be used for supervised training. For automated application, the microscopic images provided by authorized databases can be used for training and testing. For simple and efficient methods of segmentation and feature extraction steps, techniques such as Convolutional Neural Networks (CNN), K-Means and Mean Shift should be adopted. The SVM, Naive Bayes and Fully Convolutional Neural Network (FCNN) are among the techniques that have been used for cancer cell classification, which has the potential to be applied in stem cell research. Image processing or analysis has been beneficial in cancer research. Computer vision software based on machine learning and deep learning algorithms is making automated analysis possible in delivering fast and accurate results. In this regard, image processing plays a crucial role in the diagnosis and detection of cancers as well as in monitoring cancer progression patients [23].

Similarly, machine learning is capable in overcoming limitations in stem cells research, whereby, lab-based characterization and classification using chemical reagents and biological assays can be labor-intensive, expensive, and time-consuming as well as less accurate. Most implementations of supervised machine learning relied on extensive training data using extracts from large and high-throughput biological data and features, such as cellular images and genome analysis. From the perspectives of cancer diagnosis and stem cells therapy, the machine learning approach is useful to understand the regulatory genomics. This includes the identification of regulatory variants, the effects and origins of mutation using DNA sequence, analyzing whole cells, the population of cells and tissues through detecting features that can be difficult or impossible to uncover in conventional laboratory settings [95]. Pattern recognitions and classifications of such biological data are important in identifying factors, which pose biosafety and bioefficacy risks of stem cells in clinical application.

While different techniques have been developed for analysis, deep learning method provide a more effective strategy due to the diversity of the data. It has been used to classify lesions and nodules; localize organs, regions, landmarks and lesions; segment organs, organ substructures and lesions; by retrieving images based on content; generating and enhancing images; and combining images with clinical reports [13, 96, 97]. The application of deep learning in analyzing images has been widely used in cancer stem cell phenotype research. Ke Fan et al. [98] demonstrated that the combination of SVM, RF and CNN was able to measure morphological dynamic and colony formation of iPSCs within 7 days. The application of SVM by Tanaka et al. [99] enabled automated classification of adipogenic and osteogenic differentiation as well as undifferentiated features of human mesenchymal stem cells (hMSC) in RGB color image. Meanwhile, Theagarajan and Bir Bhanu [100] have developed and proposed new automated detection and classification of human embryonic stem cell (hESCs) with an accuracy of 94.46% using the application of CNN for phase contrast hESC image analysis.

2. Stem Cells and Cancer Cells Feature Engineering for Machine Learning

Several studies have shown that stem cells and cancer cells share some similarities. These similarities can be attributed to their functional capability that is conceptually similar in terms of their ability to self-renew and to proliferate [101]. However,
they are also fundamentally different, whereby, these cells can be distinguished by different regulatory mechanisms reflected in at least three characteristics; 1) propagation and proliferation ability, 2) morphology and 3) cell surface markers. These characteristics may be considered for risks evaluation associated with the safety and efficacy of stem cells. Image-based high-content screening has also become increasingly important in stem cells research in monitoring the changes in phenotype, such as cell morphology and differentiation [102, 103].

2.1. Proliferation and propagation ability

In terms of proliferation ability, it is important to understand that normal stem cells are regularly more vigilance in controlling their proliferation, but this ability is lacking in cancer cells [101]. Cells population doubling time is one of the distinguishing factors that can be included as a feature used to train the machine learning model. However, it is important to also take into account the different models [101] used to understand cancer propagation. Cancer stem cells (CSCs) model is rare and is a phenotypically distinct group of cells, which may hierarchically induce the stable generation of non-tumorigenic and tumorigenic cells. They can probably be generated from normal stem cells or precursor cells within tissues after mutations and resistant to conventional chemotherapy occurred [104].

Although they are rare, certain markers have been suggested for the identification of CSCs. In the clonal evolution model, cancer cells are distinctive in phenotypes with malignant potential and ability for disease propagation by undergoing additional genetic mutations. While in the interconversion model, cells can interconvert between being actively malignant and relatively quiescent, which is associated with the phenotypic differences between these cells. Although there are distinctive differences in the cells in each model, they are not mutually exclusive, whereby, tumorigenic cells are able to undergo further genetic and epigenetic alterations, depending on their microenvironment regardless of which model the cells follow [105]. Based on the propagation and proliferation ability, specifically associated biomarkers, cell count or numbers and time-period depicting cells population growth could be the features that can be included in the training models to distinguish the characteristics of normal stem cells and cancer-associated cells.

2.2 Cell Morphology

Morphologically, stem cells and cancer cells may show different cell features, which can be viewed microscopically (Figure 2). Microscopic images of cells are crucial to extract information for the machine learning models to distinguish the different features of normal stem cells and cancer cells based on their sizes and shapes. Generally, the appearance of normal stem cells is more consistent in their shapes and smaller in size while cancer cells can be abnormal and vary in shapes and sizes, which may be contributed by their heterogeneous nature. Although they may exhibit unique differences in their morphology, cell features, and motions require laboratory experimental approaches to create sample dataset prior to machine learning modeling. Based on a suitable model, the identification of cancer progenitor cells can be confirmed based on the morphology and motion pattern, which may be different from normal cells.

Zhang et al. [91] used time-lapse microscopic images of iPS forming cells in early stage reprogramming and selected 11 types of cell morphological and motion features, which included the area of coverage and speed for modelling to perform feature selection. Further analysis of cell motion showed that migratory motions for progenitor cells can be distinguished by the direction and distance to bring distant progenitor cells together. However, the input of cell features described by Zhang et al. [91] and Meygola et al. [106] would require high-resolution time-lapse imaging to allow the detection or tracing of cellular events. With regards to cell segmentation and tracking, Dzyubachyk et al. [107] used coupled active surfaces algorithm and time-lapse fluorescence microscopy images. While Türetken et al. [108] proposed an integer programming to track elliptical cell populations in time-lapse image sequences. In the case of image segmentation, the challenge with live-cell imaging is in determining which parts of images correspond to which individual cells. Van Valen et al. [83] showed that this can be solved by applying CNN that can robustly segment fluorescent images of cell nuclei and phase contrast images of cells without the use of a fluorescent cytoplasmic marker.

[Fig. 2 about here.]

2.3. Cell surface markers

Cell surface markers are associated with features and changes in cell morphology and progression. Some require deep epigenetic experimental approaches for input. It is challenging in determining specific markers for normal stem cells, cancer cells or CSCs as most of these markers can be presented in all types of cells, making them non-specific. Otherwise, these aspects would require machine learning approach in identifying specific cell surface markers. In comparing between normal stem cells and cancer cell progression, it is mostly discussed within the context of CSCs, as stem cells have also been shown to be involved not just in cancer initiation and progression but also in CSCs generation [109]. Nevertheless, contradictory results on CSCs and stem cells relationships are still very much debatable.

Although both cells share some similarities in terms of self-renewal and differentiation ability, there are studies Some studies showed different characteristics between the two cell types and this can be delineated by the existence of specific cell markers [101, 110]. This feature can be used for machine learning approach to classify and track stem cells progression in a particular environment for the risks of them conforming or inclining towards cancer-associated cells. An important attribute of CSCs is that they have the ability to trans-differentiate into different phenotypes [111, 112], whereby, they can express angiogenic and vasculogenic markers and also
be able to organize a pseudo vascular network. Several studies have also associated these characteristics to the expression of potentially CSC markers in several types of cancer cell lines, such as breast cancer cell lines, MDA-MB 453 and MDA-MB 231 [113-115], non-small cell lung cancers (NSCLCs) [116], renal cell carcinoma Cell [117], nasopharyngeal carcinoma cell (NPC) [118], colon cancer cells mucosseptoid carcinoma cell lines (YD15) and its derivative (YD15M) [119]. From these studies, it can be summarized that the characteristics of malignancy and cancer progression were typically associated with a panel of surface markers, which are CD133+, CD44+, CD24−, OCT3/4+, or NANOG [120, 121]. On the other hand, the CD44hi ESAlo or CD44hi ESA hi expressions indicated the presence of CSCs population in squamous cell carcinoma in breast through a comprehensive analysis of data obtained from flow cytometry, immunohistochemical and real-time polymerase chain reaction (RT-PCR) [112].

Surface markers regulation leading to the induction of epithelial-mesenchymal transition (EMT), which resulted in the acquisition of invasive and metastatic properties is also one of the characteristics found in CSCs [119, 122]. EMT phenomenon in CSCs has been reported as metastasis precursor, which enable the cells to acquire invasiveness and become extremely resistant to conventional therapies [112, 117]. The down-regulation of E-cadherin and upregulation of N-cadherin, which are termed as cadherin switching cascade, is a major hallmark of EMT. The cells which are undergoing EMT can be accurately identified through intensive genomic profiling for downregulation of cytokeratin (CK) and upregulation of vimentin, N-cadherin and fibronectin. This may be important markers for the characterization and identification of CSCs [119].

Currently, image analysis has been employed in the study of stem cells reprogramming and its progression using iPSCs. Kusumoto et al. [123] employed CNN to identify endothelial cells derived from iPSCs, whereby, the networks were trained using phase contrast images of endothelial cells based on morphology only. The network performance was then assessed by K-fold cross-validation, which confirmed that CNN was able to identify endothelial cells based on morphology with high performance. On the other hand, computer vision-based deep learning was also used to study the progress of stem cells differentiation. CNN was also able to be trained with transmitted light microscopy images to identify pluripotent stem cells from early differentiating cells and its ability to recognize the features with more than 99% accuracy [124]. Similarly, the classification of light microscopic images was used to predict lineage choice and cellular movement of primary hematopoietic progenitors during differentiation [125].

Despite limited machine learning application for stem cell biosafety and bioefficacy, comparative and classification analysis of stem cells can be carried out by comparing the cell images of stem cells and cancer cells without having the dependency on molecular and biological assays. Depending on a particular niche (i.e., whether in vitro or in vivo conditions), stem cells can initiate or acquire senescence or cancer characteristics. This has been demonstrated in glioblastoma multiforme study by Adamski et al. [126], which reported that, there is a putative link between cellular dormancy of malignancies and stem cell-like characteristics in cancer that could be due to the co-expression stem cells markers. Based on these studies, the prediction in the risks of stem cells to acquire cancer characteristics prior to clinical applications is possible.

V. TECHNICAL RECOMMENDATION OF MACHINE LEARNING IN STEM CELLS RESEARCH

Machine learning classification techniques have been applied in cancer research to identify and classify the types of cancer cells with relatively high accuracy, sensitivity and specificity. Some popular applications involved Support Vector Machine (SVM), K-Nearest Neighbors (KNNs), Artificial Neural Networks (ANNs), Decision Tree (DT), Random Forest (RF) and Bayesian Networks (BNs) [121, 122, 127-131]. In addition to cancer, the classification of microscopic red blood cells images from hematological disorder, such as sickle cell disease using deep-CNNs were able to reveal a diverse and any alteration in the cell shapes related to their biomechanical and bio-rheological characteristics. The deep-CNN employed showed good performance, high accuracy and robust predictions that enabled clinicians to assess the severity of the disease [132]. Similar techniques can be used to assess or profile stem cells biosafety and bioefficacy based on image analysis.

[Fig. 3 about here.]
clusters, n is the number of points and T is the number of iterations [133]. The results of segmentation can then be used for the classification process to differentiate between the normal and abnormal cell. The typical process of image processing shown in Figure 3 has been commonly used in most nuclei-based cancer image processing. For example, by applying deep learning, patterns from several types of data, such as from cancer cell dataset and stem cells dataset can be automatically extracted [133]. This includes the detection, segmentation and recognition of cell images that can be used to predict the risk of cell irregularities that could jeopardize stem cells clinical application. The detection of cell irregularities is a multistage process, which also includes pre-processing tasks, such as segmentation and feature extraction from microscopic cell images before the application of CNN [134]. This approach can be used to observe and evaluate stem cells progression, particularly during the expansion phase to detect and predict the risk of abnormalities prior to clinical application. The expansion of cells is required to increase the number of cells to ensure sufficient cells can be used in stem cells therapy. It is a cell manipulation procedure, whereby, technical manipulation can increase the possibility of genotypic and phenotypic alterations [135].

Naik & Dixit [134] reported detailed technical steps in detecting cancer from microscopic biopsy images, comprising of the training and testing of the algorithm model. The machine learning architecture of cancer detection by Naik & Dixit [134] is shown in Figure 4. For both training and testing tasks, the sequence of step started by taking image samples using a microscope, followed by the segmentation and features extraction step using the CNN based image processing and finalized by a classification step using the Naive Bayes Algorithm. By using the CNN based image processing, the microscopic image that contains nuclei, cytoplasm and other features are segmented into 12 smaller bricks. In each segmented brick, the CNN based interpretation on types of cancer was done according to the features of grey level, color, texture, Law’s Texture Energy (LTE), wavelet and Tamura’s features. This interpretation, which was given in percentage, was then subjected to Naive Bayes algorithm to classify whether the image indicates the cells to be cancerous or not. In this regard, the CNN algorithm may be applied as the basic principle of deep learning-based cell identification. As reported by Kusumoto et al. [123], the deep learning identification is more straightforward and achieves higher accuracies compared to the other machine learning techniques without the requirement of image labelling. The technical steps implemented by Naik & Dixit [134] can be adopted for the detection and evaluation of stem cells biosafety and bioefficacy risks.

Wang et al. [137] conducted the nuclei segmentation process on cervical cancer morphological cell image using Mean-Shift clustering algorithm. Figure 6 shows the nuclei image processing and machine learning classification architecture by Wang et al. [137]. The classification was carried out based on the shape and textural features of the segmented images. The color space and Gabor features were extracted from the segmented image and put together to obtain a better classification performance. The nuclei segmentation-based analysis was also conducted by Rawat et. al [138] on leukaemia morphological cell image based on the global thresholding and histogram equalization. The details of nuclei image processing and machine learning classification architecture are described in Figure 7. The normal and abnormal classes were classified using the support vector machine (SVM) classifier. On the other hand, Negm et. al [139] conducted the K-Means clustering-based segmentation process on nucleus, cytoplasm and whole-cell of leukaemia morphological cell image to classify the normal and abnormal classes based on the decision support system classification.

In cancer research, the focus on early detection is important to stop or slow down the progression of tumor growth. Similar
Figure 8 shows the architecture of nuclei image processing and machine learning classification by Negm et al. [139]. As described in this figure, K-means clustering segmentation process started by segmenting the nuclei or whole cell and the images were classified by a representation of three-color components, RGB (red, green, blue). The histogram of the color components indicates the contrast of the images. The most contrasted images were selected for K-mean clustering segmentation step. The K-means clustering-based segmentation was performed by partitioning the pre-processed image into K-mean clusters, classifying and grouping items into k groups (k is the number of pre-selected groups), minimizing the sum of squared distances between the items and the corresponding centroid used in grouping [140]. For example, if the grouping items are; background, other non-target cells and the cells to be extracted, thus, the K number is 3 (K1: background, K2: other non-targeted cells and K3: cells to be extracted).

Through the K-Means algorithm, the desired region of cells (nucleus, cytoplasm and whole-cell) can be separated from the unwanted region (background and other non-targeted cells). The segmented desired region can then proceed for features extraction step, which based on geometry, statistics, textures and size ratio. The analysis of these features was then performed to differentiate the regions for the classification step. Taken together, the techniques and algorithms used in cancer research are recommended to be used in stem cells research, particularly for biosafety and bioefficacy evaluation as summarized in Figure 9. Taking cues from the summary of the image processing pipeline in Figure 9, we proposed a framework-specific for biosafety and bioefficacy assessment, as depicted in Figure 10.

Ideally, this framework will be applicable to identify stem cells abnormality, particularly during the cell expansion phase. Following the proven studies in the similar domain, supervised learning will be employed where images with known normality level will be used as a training data. By utilizing CNN based image processing algorithm, the image of stem cells from microscope was segmented into smaller sub-image of a single cell that contains nuclei, cytoplasm and other features. In each segmented sub-image, the CNN based interpretation on the type of stem cell normality can be carried out according to the cell and nuclei features as exemplified previously [134]. The recommended features that should be considered are size, shape, grey level, color, texture, Law’s Texture Energy (LTE), wavelet and Tamura’s features. The performance metric for the model will be based on the features’ percentage of stem cell from each single cell sub-image that showed normal or abnormal conditions. The sum of average for each feature will then become a metric for the classification using various models, such as Naive Bayes, Decision Tree or Random Forest. The summary of current machine learning application in stem cell research and cancer cell research is shown in Table 1.

VI. THE CHALLENGES AND FUTURE OF PERSPECTIVES

In summary, the possibilities to adopt the aforementioned technical steps in stem cell research, particularly for risk evaluation in biosafety and bioefficacy are immense. The overlapping aspects of stem cell biology and cancer cell biology have led to the increase of large and highly complex datasets being generated from biological experiments from quantifying molecular variables, such as gene, protein and metabolites associated with different cancer and stem cells types. This has given insights into further understanding of the biological systems. Taken together, their involvement in disease progression and mechanism can be realized using machine learning and deep learning approaches. These approaches are able to address the complexity and heterogeneity of these datasets, providing new perspectives and generate novel hypotheses, particularly with regards to biosafety and bioefficacy risks and concerns in stem cells therapy. However, just as in any biomedical datasets, some of the challenges identified that may occur in stem cells research datasets are; 1) data requirements, which require large, labeled data to make deep learning successful, 2) overfitting in data training may inaccurately reflect underlying relationships, particularly in the heterogenous dataset and 3) interpretability of deep learning models may require better interpreting methods of its output [97].

Although the size of these datasets is increasing, there is still a need for massive, large datasets to reach meaningful perspectives and outcomes. Just as any biological system, data from stem cells biology can be incredibly complex with thousands of variables from different facets of physiological conditions. With suitable machine learning and deep learning models, we can assess the aspects of biosafety and bioefficacy of stem cells for clinical application. The generated model could also be used to identify fundamental design principles to create a suitable microenvironment for stem cells growth without jeopardizing their mortality and without altering their epigenetic components that may lead to cellular abnormality. However, to create such large and well-annotated datasets to study such complex network would require multi-omics datasets, which can be very expensive.

One of the options that could be utilized to take on this challenge is to use imaging data and analysis to characterize morphological and phenotypic changes of stem cells. This could be carried out by comparing the data from cancer and stem cells in various conditions and environmental perturbations as well as coupling it with deep learning algorithms. The data obtained would present interesting input in addressing biosafety and bioefficacy risks in stem cells therapy.

Nevertheless, we still have a long way to go to uncover and harnessing the potential of stem cells for therapy and to play a bigger role in the clinical settings. Machine learning approaches
ACKNOWLEDGMENTS

We would like to thank University of Malaya for providing the fund under the Research University (RU) Grant - Faculty Program (Project No: GPFO12A-2018 and GPFO39A-2019) and by the Ministry of Higher Education Malaysia (FRGS-FP114-2020). We thank reviewers and associate editor for their comments which improved this manuscript.

REFERENCES


plastic transplantation in first remission acute myeloid leukemia in young

I. Kang et al., “Donor-dependent variation of human umbilical cord
blood mesenchymal stem cells in response to hypoxic preconditioning

stem cells from older osteoarthritic patients have lesser differentiation

C. A. Herbergs, M. S. Kwa, and H. P. Hermens, “Risk factors in the

G. E. Georges, K. Doney, and R. Storb, “Severe aplastic anemia:
allogeneic bone marrow transplantation as first-line treatment,” Blood

J. M. Byun, D. H. Kim, D. Y. Shin, I. Kim, Y. Koh, and S. S. Yoon,
“Combination of genetic aberration with international staging system
classification for stratification of Asian multiple myeloma patients

D. S. Jevotovsky, A. R. Alfonso, T. A. Einhorn, and E. S. Chiu,
“Osteoarthritis and stem cell therapy in humans: a systematic review,”

C. Goswami, S. Poonia, L. Kumar, and D. Sengupta, “Staging system
to predict the risk of relapse in multiple myeloma patients undergoing

T. J. Myers, F. Granero-Molto, L. Longobardi, T. Li, Y. Yan, and A. Spagnoli,

J. C. Ra et al., “Stem cell treatment for patients with autoimmune disease by systemic infusion of culture-expanded autologous adipose
tissue derived mesenchymal stem cells,” J Transl Med., vol. 9, no. 181, pp. 1-11, 2011.

H. Fernández-Susavila, A. Bugallo-Casal, J. Castillo, and F. Campos,
“Adult stem cells and induced pluripotent stem cells for stroke


S. Shenoy et al., “Current Results and Future Research Priorities in Late
Effects after Hematopoietic Stem Cell Transplantation for Children with
Sickle Cell Disease and Thalassemia: A Consensus Statement from the Second Pediatric Blood and Marrow Transplant Consortium
International Conference on Late Effects after Pediatric Hematopoietic

E. Meijer and J. J. Cornelissen, “Ilogenic stem cell transplantation in
acute myeloid leukemia in first or subsequent remission: weighing
prognostic markers predicting relapse and risk factors for non-relapse

A. Grigg and T. Hughes, “Role of Allogeneic Stem Cell Transplantation


U. M. Saarinen-Pihkala et al., “Pathways Through Relapses and Deaths
of Children with Acute Lymphoblastic Leukemia: Role of Allogeneic Stem

A. Rovó and A. Tichelli, “Cardiovascular Complications in Long-Term
Survivors After Allogeneic Hematopoietic Stem Cell Transplantation,”

V. Volarovic et al., “Ethical and safety issues of stem cell-based

Umbilical Cord Mesenchymal Stem Cells: A New Era for Stem Cell


SALMAH BT. KARMAN (Member, IEEE) received her Bachelor’s degree in Electrical and Electronic Engineering from Oita University, Japan in 2003. She then received Master in Engineering Science (M.Eng.Sc) from University of Malaya, Kuala Lumpur, Malaysia in 2011. In 2016, she received her PhD from Universiti Kebangsaan Malaysia (National University of Malaysia). Currently, she is working as a senior lecturer in the Department of Biomedical Engineering, Faculty of Engineering, University of Malaysia. Her research interest mostly focuses on biosensor and bioinspired devices.

MOHD YAZED B. AHMAD (Member, IEEE) received the B.E degree in Electrical Engineering from the Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia in 2003. He then received M. Eng. Sc (Instrumentation) from the Department of Biomedical Engineering, University of Malaya in 2006 and PhD in Electrical Engineering Science from the University of Technology Sydney, NSW, Australia in 2013. He is currently a senior lecturer and a programme coordinator at the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya. His current research interests include instrumentation & real time machine learning system, wireless power transfer, smart space and IOT system.

EFFIRUL IKHWAN RAMLAN received his B. Sc. (Hons) degree in Intelligent Systems from Universiti Teknologi Mara (UiTM), Malaysia. He earned a PhD in Computer Science from University of Southampton, UK (in 2009) for computational and theoretical development of nucleic acid computers as information processors. He is currently a lecturer at the School of Computing, Engineering and Intelligent Systems (SCeIS), Ulster University, UK. His research focuses on the development of synthetic devices for biological computing, machine learning approaches in genomics and unconventional bio-inspired algorithms for multi-objective optimisation problems.

SITI NURAINIE BT TUKIMIN received her Bachelor’s degree (B. Sc.) in Science of Biology from Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia in 2014. She is currently pursuing her Master’s degree in Engineering of Science from Department of Biomedical Engineering, Faculty of Engineering, University of Malaya (UM), Kuala Lumpur, Malaysia. She is currently doing research on tumor/cancer cell morphology and characteristics for development of tumor tissue mimicking phantom.
FIGURE 1: An overview of the current research trend in stem cells research. Collection of datasets particularly on characterization of stem and cancer cells would enable machine learning approach to be applied in stem cells research and subsequently in stem cell therapy.

FIGURE 2: Representative images of human adult stem cells and selected cancer cell lines (10X magnification). A) ADSC (Human adipose-derived stem cells), B) MCF-7 (Human breast cancer cell line), C) HGT-1 (Human gastric cancer cell line), D) U937 (Human lymphoma cell line) and E) HEPG2 (Human liver cancer cell line). Differences in cell morphology may enable machine learning approach in evaluating the risks of biosafety and bioefficacy in stem cells therapy.
FIGURE 3: Overview of a typical pipeline of image processing steps for machine learning classification. Adapted from Jyoti Rawat et. al [138].

FIGURE 4: The machine learning architecture of cancer detection process which can be applied in stem cells biosafety and bioefficacy assessment. Adapted from Naik & Dixit [134].

FIGURE 6: Nuclei image processing and machine learning classification architecture using mean-shift clustering algorithm. Reproduced from Wang et al. [137] with permission.
A total of 240 blood images taken from American Society of Hematology.

<table>
<thead>
<tr>
<th>Healthy Leukocyte (100)</th>
<th>Acute Myeloblast (80)</th>
<th>Acute Lymphoblast (60)</th>
</tr>
</thead>
</table>

**Sub-image selection module**

Healthy Leukocyte (140) | Acute Myeloblast (140) | Acute Lymphoblast (140)

**Segmentation module**

Nucleus segmentation using Otsu’s method

**Feature extraction module (total feature vector length = 331)**

- Geometrical features (total = 11)
- Chromatic features (total = 15)
- Statistical model (45)
  - Laws’ texture model (210)
  - Transform domain model (50)

**Feature selection Module (Feature Pruning)**

GA-SVM

**Classification Module**

SVM Classifier

- Class decision for an image
- End

**Input ROIs**

- Classifier 1 (HLC/AMC/ALC)

**Input ALC ROIs**

- Classifier 2 (L1/L2/L3)

**Input AMC ROIs**

- Classifier 3 (M2/M3/M5)

**FIGURE 7:** Nuclei image processing and machine learning classification architecture using support vector machine (SVM) classifier module. Reproduced from Jyoti Rawat et. al [138] with permission.
FIGURE 8: Nuclei image processing and machine learning classification architecture using K-Means clustering-based segmentation. Reproduced from Negm et al. [139] with permission.
FIGURE 9: Overview of commonly used image processing-based machine learning techniques in cancer research which can be applied in stem cell research for stem cells biosafety and bioefficacy assessment.

FUNCTION
1. Noise reduction/removal
2. Image enhancing/smoothing.

Region of interest (ROI)
1. Cell Nuclei region
2. Morphology-based feature
3. Shape-based feature
4. Texture-based feature

Accuracy, Precision, Sensitivity and Specificity

Benign/Malignant

Area, Perimeter, Eccentricity
Figure 10: Image processing machine learning architecture proposed for stem cells research particularly for biosafety and bioefficacy assessment using microscopy images of human adipose derived stem cells as an example.
### TABLE 1
A SUMMARY OF CURRENT MACHINE LEARNING APPLICATION IN STEM CELL RESEARCH AND CANCER RESEARCH.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Type of Cell</th>
<th>Task</th>
<th>Learning Method</th>
<th>Feature Selection</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>[91]</td>
<td>iPSC</td>
<td>Live-cell microscopic imaging dataset for iPSC progenitor cells detection.</td>
<td>XGBoost algorithm</td>
<td>11 types of features are extracted (volume, area, sphericity, ellipsoid-prolate, ellipsoid-oblate, nucleus-cytoplasm volume ratio, displacement, speed, Intensity-StdDev, Intensity-Max, Intensity-Min) using time window and two-step feature selection.</td>
<td>XGBoost setting (learning rate = 0.01, n_estimators = 385, gamma = 0, and 5-Kfold cross validation).</td>
</tr>
<tr>
<td>[98]</td>
<td>iPSC</td>
<td>Time-lapse bright field image analysis for iPSC colony formation detection and prediction</td>
<td>CNN for iPSC colonies recognition, HMM for growth curve modelling (Baum-Welch and Viterbi algorithms), and AlexNet algorithm as classifier.</td>
<td>2 main types of features selected (Colony texture and area of growth) using the sliding window method with Autolevels (AL) algorithm, Gaussian filter (GFP) and Random Walker algorithm for dataset pre-processing.</td>
<td>Batch normalized layer training for AlexNet with manual human validation.</td>
</tr>
<tr>
<td>[123]</td>
<td>iPSC</td>
<td>Phase contrast and immunofluorescence image analysis of differentiated endothelial cells</td>
<td>LeNet and AlexNet algorithm</td>
<td>2 types of feature selected (Differentiated feature of endothelial cells – vascular tube and CD31 expression)</td>
<td>LeNet and AlexNet architectures are available in [123] (Figure 3).</td>
</tr>
<tr>
<td></td>
<td>Human ESC</td>
<td>Phase contrast microscopic ESC images classification</td>
<td>DeephESC 2.0 algorithm consisting of CNN, Triplet CNNs and Fused CNN-Triplet.</td>
<td>6 types of features selected (cell clusters, debris, unattached cells, attached cells, dynamically blebbing cells and apoptically blebbing cells) following the image pre-processing with intensity modeling of a mixture of two Gaussians.</td>
<td>DeephESC 2.0 setting (learning rate = 1.2x10^{-5}, momentum = 0.9 and weight decay = 1x10^{-5}), variation of architectures to improve hESC classification is available in [100] (Table 4).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Generative Multi Adversarial Networks (GMAN) is implemented for synthetic hESC images dataset.</td>
<td>Two triplets CNN were introduced to perform fine-grained classification using the architecture presented in Figure 9 [100] with hyper-parameters setting in Table 6.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GMAN discriminators setting is available in Table 7 [100].</td>
</tr>
<tr>
<td>[99]</td>
<td>MSC</td>
<td>Microphotograph of MSC image for differentiation analysis</td>
<td>SVM algorithm</td>
<td>3 main types of features selected (Adipogenic differentiation, osteogenic differentiation and non-differentiated) based on the intensity balance of RGB channels.</td>
<td>RGB pixel intensities with maximum value of R = 255, G = 255, and B = 255.</td>
</tr>
<tr>
<td>[107]</td>
<td>Mouse ESC</td>
<td>Time-lapse fluorescence microscopic image analysis of cell foci and cell motion</td>
<td>Level-set-based cell segmentation and tracking algorithm</td>
<td>1 type of feature selected (time course in cell phase transition) using cell segmentation and tracking with cell motion correction and foci segmentation and pattern recognition.</td>
<td>Level-set-based cell segmentation and tracking hyperparameters (motion compensation algorithm, $\alpha = 0.5$ and $\beta = 0.95$; $\lambda$ of the smoothness energy term of the foci segmentation algorithm = 10%; focus size = 3-5 voxels; expected contrast = 0.05 &amp; 0.2; cell-phase classification threshold = 30% of foci at boundary)</td>
</tr>
<tr>
<td>[129]</td>
<td>Breast cancer</td>
<td>Digital mammography images classification</td>
<td>CNN algorithm (fully connected (FC) layers involving VGG network or Resnet network as a classifier)</td>
<td>8 types of feature selected (regional area, major axis length, mean intensity, background, malignant mass, benign mass, malignant calcification and benign)</td>
<td>CNN setting (First step with 3 training strategy: learning rate = 10^{-3}, 10^{-4}, 10^{-5}. Second step with 2 training strategy: learning rate = to 10^{-3}, weight decay = 1x10^{-5})</td>
</tr>
</tbody>
</table>
patch classifier layers). The complete network design is presented in Figure 1 [129].

<table>
<thead>
<tr>
<th>Year</th>
<th>Disease</th>
<th>Dataset Type</th>
<th>Methodology</th>
<th>Features selected</th>
<th>hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>Breast Cancer</td>
<td>Magnetic Resonance Images (MRI) dataset classification</td>
<td>CNN algorithm for image processing and Naïve Bayes as classifier</td>
<td>4 types of colour-based feature selected (grey-level texture feature, Law’s Texture energy, Tamura feature and wavelet feature)</td>
<td>decay $= 0.001$ and learning rate to $10^{-3}$; weight decay $= 0.01$</td>
</tr>
<tr>
<td>2016</td>
<td>Cervical Cancer</td>
<td>Pap smear images classification</td>
<td>FCN algorithm with U-net variant comprising of residual blocks from Deep Residual Networks and dense blocks from Densely Connected Networks within the shape representation model (SRM) encoder.</td>
<td>3 types of feature selected (cell nuclei irregularity size, shape and texture). Stacked auto-encoder (SAE) is implemented to construct SRM for segmentation.</td>
<td>FCN network architecture is presented in Figure 4 [136]. Optimization of the network was performed with Adam optimizer. FCN algorithm hyperparameters (stochastic gradient descent, learning rate $= 0.01$; reduced by a factor of 10 every 5 epochs, and momentum $= 0.88$ for 200 epochs)</td>
</tr>
<tr>
<td>2017</td>
<td>Cervical Cancer</td>
<td>Pap smear images classification</td>
<td>SVM algorithm as classifier and Mean-shift clustering algorithm with mathematical morphology for segmentation.</td>
<td>5 types of selected features (cell nuclei area, perimeter, eccentricity, roundness, circularity)</td>
<td>N/A</td>
</tr>
<tr>
<td>2018</td>
<td>Lymphoblastic and Myeloblastic leukemia</td>
<td>Microscopic blood images classification</td>
<td>Genetic algorithm (GA) for feature selection with SVM algorithm as classifier.</td>
<td>3 main types of selected features (colour, statistical texture &amp; geometric - area, perimeter, diameter, Euler's no, major axis, minor axis, solidity, eccentricity, roundness, convex area and extent). The complete list of the extracted features and methods for extraction is available in Table 6 [138].</td>
<td>GA setting (size population = 20, number of generations = 10, replacement rate = 0.8, crossover fraction = 0.5, mutation fraction = 0.01, fitness scaling = rank, selection function = roulette, no of variables = 331), SVM hyperparameters (Gamma $g = 0.0057/\kappa$, Cost $C = 15,334$).</td>
</tr>
<tr>
<td>2019</td>
<td>Leukaemia</td>
<td>Acute leukemia blast cells in colored microscopic images classification</td>
<td>Artificial Neural Network (ANN) and Decision Tree as classifier with K-means clustering algorithm</td>
<td>4 main types of selected features (geometry, statistics, textures, and size ratio from selected regions in nucleus, cytoplasm and whole cell)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Notes: iPSC - induced Pluripotent Stem Cells, MSC - Mesenchymal Stem Cells, ESC - Embryonic Stem Cells, Min - Minimum, Max - Maximum