We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Quantitative-Morphological and Cytological Analyses in Leukemia

Cecília Lantos, Steven M. Kornblau and Amina A. Qutub

Abstract

Leukemia, a blood cancer originating in the bone marrow, presents as a heterogeneous disease with highly variable survival rates. Leukemia is classified into major types based on the rate of cancerous cell growth and cell lineage: chronic or acute and myeloid or lymphoid leukemia. Histological and cytological analysis of the peripheral blood and the bone marrow can classify these major leukemia categories. However, histological analyses of patient biopsies and cytological microscopic assessment of blood and bone marrow smears are insufficient to diagnose leukemia subtypes and to direct therapy. Hence, more expensive and time-consuming diagnostic tools routinely complement histological-cytological analysis during a patient’s diagnosis. To extract more accurate and detailed information from patient tissue samples, digital pathology is emerging as a powerful tool to enhance biopsy- and smear-based decisions. Furthermore, digital pathology methods integrated with advances in machine learning enable new diagnostic features from leukemia patients’ histological and cytological slides and optimize patient classification, thus providing a cheaper, more robust, and faster diagnostic tool than current standards. This review summarizes emerging approaches to automatically diagnose leukemia from morphological and cytological-histological analyses.

Keywords: blood cancer, acute leukemia, machine learning, digital pathology, classification, supervised learning

1. Introduction

Leukemia is a very heterogeneous cancer that arises from the combination of many genetic and epigenetic mutation events, all of which alter hematopoiesis [1–3]. Hematopoiesis is the
proliferation of blood cells in the bone marrow (BM). Blood cells differentiate in the BM and, then, when mature, spread out to the peripheral blood (PB) system. In normal circumstances, the multipotent progenitor hematopoietic stem cells in the bone marrow reproduce and commit to differentiate into common myeloid or lymphoid progenitor cells. Myeloid and lymphoid progenitor cells differentiate into two main cell lineages containing unipotential precursor cells. Each precursor matures through multiple stages to become a red blood cell (RBC), a platelet, or a white blood cell (WBC) type. Myeloid cells consist of RBCs, platelets, segmented neutrophils, monocytes, eosinophils, basophils, and mast cells; lymphoid cells are T and B lymphocytes, dendritic cells, or natural killer (NK) cells (Figure 1) [1, 4].

Malignant proliferation in the myeloid or lymphoid cell lineage causes myeloid or lymphoid leukemia. The diseased cells stop maturing, halt differentiation, and then accumulate, hence blocking the development of healthy progenitor cells. Cell maturation in chronic leukemia is blocked at a later stage, and it has a longer course of development compared to acute leukemia, where lineage proliferation is arrested at an early stage of differentiation leading to a very aggressive, fast-growing disease [4, 5].

Based on these two major differences, myeloid or lymphoid and chronic or acute, four major leukemia types are distinguished: acute myeloid leukemia (AML), acute lymphoid leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphoid leukemia (CLL). Each type has a distinguishable morphology, and diagnosis is based on histological analysis of each patient’s bone marrow biopsy and cytological microscopic assessment of bone marrow smear or peripheral blood smear [5, 6].

However, full classification requires more refined categories than the four major leukemia types, and modern classification also includes mutation analysis, cytogenetics, and flow cytometry data. Therefore, older morphological-based classification systems (French-American-British (FAB)) cannot be fully matched with the World Health Organization (WHO) scheme, which utilizes all of these features. The FAB classification system is predominantly used for the 30–40% of AML cases that are not otherwise specified, while in special cases, a morphological

Figure 1. Schematic chart of hematopoiesis.
pattern can be matched to an individual gene mutation or clinical criteria (e.g., AML t(8;21)(q22; q22.1)/RUNX1-RUNX1T1 or AML with myelodysplasia-related changes) [7]. To enable more personalized therapy, diagnosis and therapy selection require the analysis of histology and cytology in combination with all clinical and genetic data of the patient including cytogenetics, gene mutation analysis, and gene expression data obtained from flow cytometry [5, 6, 8, 9].

The ancillary diagnostic tests are costly and time-consuming, currently often requiring a week or more, compared to the histological-cytological analysis, which can typically be performed in a day. The delay in diagnosis can lead to delays in treatment, seriously impacting patients suffering from acute leukemia. A faster, less time-consuming, and more precise automated histology- and cytology-based diagnostic tool would facilitate diagnosis and personalized, rapid treatment [10].

Due to the development of whole slide imaging (WSI), which yields large digital images of an entire tissue sample, patients’ histological-cytological data are stored in an image bank format to allow easy access to study their pathology. A digital tissue bank fosters the computational, automated analysis of the histological-cytological images [11–14]. This review describes possible solutions to integrate morphometrics obtained from images of biopsy and smear samples with standard clinical covariates to optimize diagnosis and direct therapy for leukemia.

2. Morphological diagnosis in leukemia

The current major standard diagnostic test in leukemia is histological-cytological analysis. This includes basic light microscopy of routinely stained bone marrow biopsy, bone marrow smear, and peripheral blood smear.

Diagnosis from the smear is established based on the complete blood count and differential count (the proportion of specific cell types in the specimen). A biopsy can confirm the percentage of the specific cell types in the smear. Normal PB smear contains mature cells and up to 1–2% immature cells. The presence of immature cells at a significantly higher percentage leads to the diagnosis of leukemia. Leukemia in the BM smear is detected based on the irregular proportion of specific immature cells and their morphological alterations [1].

Based on abnormally high proportions of specific blood cells and morphological dysplasia in the biopsy and smear specimen, the French-American-British (FAB) system describes a morphologically based classification for acute leukemia. AML subtypes are divided into eight different groups (M0–M7) and ALL subtypes into three different groups (L1–L3). Such classification system for chronic leukemia is less precise, where the subtypes are overlapping [5, 9].

Although the FAB classification system is based on cellular appearance, some immature cells do not have distinguishable morphological characteristics. Immunophenotyping confirms the diagnosis, especially in ALL T- and B-cell lineage and AML minimally differentiated (M0) and AML megakaryoblastic (M7) subtypes [5].
As a result, histology and cytology are major diagnostic tools; however, their current prognostic potential is limited, as the majority of genetic events do not have known, defining morphological characteristics [5, 10]. Thanks to emerging computer technologies, a pathologist’s qualitative decision can be supported by an automated quantitative decision tool. Morphometrics of the pathological slides can both provide new diagnostic information not visible to the naked eye and improve the prognostic ability of histological-cytological analyses [15, 16].

3. Digital cytology analysis

Statistical analysis of cells and cellular features can guide a pathologist’s diagnosis of leukemia. The number of RBCs, WBCs, and platelets, the proportion of specific immature and mature cells, and more detailed morphological features recognized by automated WSI can all help direct a diagnosis. Digital image analysis of BM biopsy has been applied to study relapse in AML [17, 18]. The focus of most prior studies has been recognition of the acute leukemia FAB subtypes or differentiation of acute and chronic leukemia from the PB and BM smears. While PB and BM smears can differ in the type and maturation of their cells, the quantitative process to recognize the cell types and extract morphometric information is similar. We describe it below:

Following steps a pathologist would take, computer-based digital pathology aims to detect, localize, and recognize the specific cell type under study (Figure 2). An acquired image is preprocessed through image enhancement steps; then, cells are detected, and cell boundaries are traced out using segmentation algorithms for morphometric analysis.

Processing steps to automate the analysis of leukemia smear images are shown in Figure 2. A whole slide image (WSI) of a Wright-Giemsa-stained bone marrow smear is shown in Figure 2a. This is a typical smear image annotated by a pathologist. The Wright-Giemsa-stained image reveals RBCs as smaller pink cells without nuclei (either distributed as single cells or clustered) and a few WBCs of different sizes with dark purple nuclei. Notably, image acquisition techniques, staining methods, and digitization protocols can differ in each laboratory. Furthermore, environmental effects can introduce artifacts and degrade the quality of the image. Image preprocessing steps can improve the image quality and correct for differences in protocols and in illumination. Methods of image correction techniques include illumination.

Figure 2. Image processing pipeline (a–e: Image from Carlos Bueso-Ramos, MD, PhD, MD Anderson Cancer Center).
normalization, color or stain correction for image enhancement, contrast enhancement and smoothing, contrast stretching, and histogram equalization [19–21].

Leukemia is detected based on the number, type, and proportion of various cell types in the blood. Segmentation algorithms enable identification of individual cells from smear images (Figure 2b); these algorithms can distinguish overlapping cells from individual cells in order to extract cell-based features and can also divide each WBC into its components: cell membrane, nucleus, and cytoplasm (Figure 2c) [19, 21–37]. Following segmentation, metrics can be extracted from the WBCs and their subcellular components (cell size, nuclear size, etc.).

4. Quantification of cytology using machine learning

To computationally classify tissue types from smear images, identified cells and tissues in the images have to be transformed into a vector of features. Conventional machine learning algorithms typically utilize a domain-specific approach to classify cell and tissue types based on a series of handcrafted features. These algorithms extract metrics from images based on a human engineering process that requires domain knowledge [38, 39].

Features of the smear sample can be extracted from an individual cell in the image or across the entire slide. Once a WBC is segmented within the image, features are extracted either from the whole WBC or separately from the nucleus and cytoplasm. The major discriminating cellular characteristics to classify WBCs are (a) geometric features such as shape (e.g., roundness) and size (e.g., nucleus-cytoplasm size ratio); (b) color features; (c) texture features such as density, granularity, and Fourier descriptors for texture quantification calculated by the two-dimensional Fourier transform; and (d) irregularity or boundary roughness measured by fractal dimension [10, 23, 33, 35, 40–49]. Although the analysis at the single cell level provides useful information, it is not sufficient for the diagnosis of a very heterogeneous disorder such as leukemia. In addition to single cell data, characteristics of multicellular groups need to be studied [1]. New studies have extended cell-based morphometric analysis to distinguish major leukemia types and subtypes (Table 1).

The common characteristics in these studies are general steps of the image processing pipeline: preprocessing, segmentation, feature engineering, and supervised classification (Table 1). They discriminate cancerous vs. healthy tissue, AML vs. ALL, CL vs. AL, or AML and ALL subtypes. The main differences across the various studies are the choice of the specific engineered features and the choice of the classification method as illustrated below.

Most of the digital pathology studies of leukemia analyze PB. A healthy blood smear is distinguished from a leukemic smear if one or more immature cells are present. This can be determined from the nucleus structure or from whole cell characteristics. Discriminating features that classify healthy tissue, AML and ALL in the PB are extracted from the cell nucleus. BM is more heterogeneous than PB, and features of BM images are extracted from the whole cells or separately from the nuclei and the cytoplasm. Commonly used features include texture-based metrics and morphology. Texture is based on the spatial variation of the
gray-level pixel intensities which can be characterized by their homogeneity, energy, and correlation, among other metrics represented in the gray-level co-occurrence matrix (GLCM). Shape is based on geometrical parameters such as area, perimeter, compactness, minor axis, and major axis.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Leukemia type</th>
<th>Extracted features</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reta et al. 2015 [58]</td>
<td>BM (Bone marrow): AML vs ALL, M2 vs M3 vs M5 vs vs L1 vs L2</td>
<td>Cell + nucleus + cytoplasm: Shape: area, perimeter, circularity, width, length, elongation, major axis, minor axis, eccentricity, extent, equivalent diameter, Euler number, Convex area Size Ratio: Nucleus/Cytoplasm area, Nucleus/Cell area and perimeter Color/ Pixel Intensity Statistics: Mode, Mean, Standard deviation, Variance, Sum Texture: Homogeneity, Contrast, Correlation, Energy, Entropy 10 Eigenvalues (PCA) of R, G, B channel of RGB image and of gray image</td>
<td>k-nearest neighbor (kNN), Random Forest (RF), Simple Logistic (SL), Support Vector Machines (SMV), Random Committee (RC)</td>
</tr>
<tr>
<td>Reference</td>
<td>Leukemia type</td>
<td>Extracted features</td>
<td>Classification</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Gumble et al. 2017 [51]</td>
<td>PB: ALL vs Healthy</td>
<td>Nucleus &amp; cytoplasm (Binary gray): Shape/Texture: Area, Total White Cells, Total Black Pixels, Perimeter, Eccentricity, Solidity, Form Factor, Bounding Box</td>
<td>k-Nearest-Neighbor (kNN)</td>
</tr>
<tr>
<td>Harun et al. 2011 [56]</td>
<td>PB: AML vs ALL</td>
<td>Cell &amp; cytoplasm &amp; nucleus: Shape: Area, Nuclear/ Cytoplasm Size Ratio</td>
<td>Hybrid Multilayer Perceptron Neural Network (HMLP NN)</td>
</tr>
<tr>
<td>Mohapatra et al. 2014 [52]</td>
<td>PB: ALL vs Healthy</td>
<td>Nucleus &amp; cytoplasm&amp; cell: Shape: Area</td>
<td>Ensemble of Classifiers (EOC), Naïve Bayesian (NB), k-nearest neighbor (KNN), Multilayer Perceptron (MLP NN), Radial Basis Functional Network</td>
</tr>
</tbody>
</table>

Quantitative-Morphological and Cytological Analyses in Leukemia
http://dx.doi.org/10.5772/intechopen.73675
<table>
<thead>
<tr>
<th>Reference</th>
<th>Leukemia type</th>
<th>Extracted features</th>
<th>Classification</th>
</tr>
</thead>
</table>
Nucleus: Shape: Form Factor, Roundness, Compactness, Elongation, Perimeter Color: mean intensity of R,G,B and Hue, Saturation, Lightness components Texture: Fourier transform: Mean, variance, skewness, kurtosis of the frequency components  
Boundary roughness: Fractal HD dimension Contour signature: Variance, skewness, kurtosis (center-contour)  
Nucleus & Cytoplasm & Cell: Shape: Area  
Nucleus & Cell: Shape: Size ratio: Nucleus / Cell  
Nucleus & Cytoplasm: Color: RGB and HSV components  
Cytoplasm: Vacuole count  
Nucleus: Shape: Form Factor, Roundness, Compactness, Elongation, Perimeter Nucleus Indentation, Nucleoli count, Texture: Fourier descriptor, Wavelet and Haralick coefficients (GLCM statistics): Contrast, Correlation, Energy, Homogeneity, Entropy  | Ensemble of classifiers (EOC), Naive Bayesian (NB), K-nearest neighbor (KNN), Multilayer Perceptron (MLP NN), Radial Basis Functional Network (RBFN), Support Vector Machine (SVM) |
Cytoplasm: Color: RGB and HSV components Texture: Coarseness Intensity and shape: Auer Rodes as Cytoplasmic Holes  

Table 1. Leukemia subtype classification.
major axis, eccentricity, form factor, elongation, and solidity. Fractal or Hausdorff dimension (HD) represents the nucleus boundary roughness (Jacob and Mundackal) [50].

4.1. Examples of digital pathology for leukemia

To provide examples of digital pathology’s impact in leukemia classification, we summarize here a few of the recent studies. In one study, ALL cells were distinguished from healthy PB cells from shape and texture features extracted from the nucleus and cytoplasm (Gumble and Rode). These features included area, total white blood cells, total black pixels, perimeter, eccentricity, solidity, form factor, and bounding box parameters [51]. In another study, Mohapatra et al. added color and the Fourier descriptor as a cell-based nuclear feature to the shape, fractal, and texture parameters to distinguish ALL from healthy lymphoblasts/lymphocytes [52].

What literally do these features mean? In the Mohapatra et al. study, color features of a cell were calculated from the mean intensity of the nucleus color components in RGB or HSV color space and from a grayscale intensity map. In the case of RGB images, the mean intensity of the red, green, and blue channels and, in the case of HSV images, the mean intensity of the hue, saturation, and lightness components were computed. The same color features were calculated for the cytoplasm. The Fourier descriptors were the mean, variance, skewness, and kurtosis of the texture in the frequency domain. The fractal/HD of the nucleus boundary roughness was considered, as was the variance, skewness, and kurtosis computed between the cell’s center and each contour point. Texture features from the cytoplasm included wavelet coefficients and metrics derived from the GLCM including contrast, correlation, energy, homogeneity, and entropy values. The area was calculated for the nucleus, cytoplasm, and the whole cell [52].

In addition to determining leukemia from cell-based features, AML can be distinguished from healthy tissue by extracting whole tissue—/slide-based features as illustrated in two other studies (Madhukar et al., Agaian et al.) [53, 54]. Furthermore, AML can also be distinguished from ALL through comparing cellular features in patient smears, as shown by Jacob and Mundackal [50], Supardi et al. [55], and Harun et al. [56]. Jacob et al. and Supardi et al. used cellular metrics based on texture, shape, and Hausdorff dimension, while Harun et al. classified the two leukemias by cell and nuclear perimeters, areas of the cytoplasm and white cells, and nucleus-cytoplasm ratio [56].

More specifically, AML and ALL subtypes have been discriminated based on cell-based features in three different studies. To classify AML subtypes, Kazemi et al. predicted five AML groups (M2, M3, M4, M5, and all the remaining subtypes (M0, M1, M6, M7 considered as one group) based on handcrafted morphological features from blood microscopic images. The features used were extracted from cells’ nuclei: irregularity, Hausdorff dimension, shape, color, and texture features complemented by the nucleus-cytoplasm ratio. The same set of features allowed more accurate discrimination of healthy tissue vs. AML tissue than AML tissue vs. ALL tissue [57]. Reta et al. performed a similar analysis which discriminated L1, L2, M3, M3, and M5 subtypes in ALL and AML based on cellular features, with nucleus features proving to be the most discriminative [58]. An earlier study (Escalante et al.) was also able to discriminate
multiple leukemia tissue types from the BM: ALL vs. AML, L1 vs. L2, M2 vs. M3 + M5, M3 vs. M2 + M5, M5 vs. M2 + M3, M1 vs. M3 vs. M5, and L1 vs. L2 vs. M1 vs. M3 vs. M5 [59]. However, in the latter study, there was no significant difference in model performance using features extracted from the nucleus and cytoplasm vs. the whole cell. This contradicts other studies that suggest classification based on subcellular morphometry improves AML [60] and ALL [61] subtype recognition. In particular, these groups found that color and shape information in the cytoplasmic holes, which indicate vacuoles, and color and shape information on the nucleus, which indicate nucleoli, can reveal the presence of Auer rods discriminating AML from ALL where Auer rods are absent [61].

In addition to the large number of publications characterizing acute forms of leukemia, studies (Vaghela et al.) have suggested measurements of WBC roundness and counts can discriminate chronic myeloid vs. chronic lymphoid leukemia [62].

4.2. Classification methods

4.2.1. Traditional machine learning methods

Support vector machine (SVM) is a common classification method in leukemia (Jacob et al. [50]; Agaian et al. [53]; Kazemi et al. [57]; Madhukar et al. [54]). However, other methods (Supardi et al., Gumble and Rode) have been applied with success to classify AML and ALL histology and cytology images including k-nearest neighbor classifiers [51, 55], a hybrid multilayer neural network (HMLNN) (Harun et al.) [56], and an ensemble particle swarm model selection method (EPSMS) (Escalante et al.) [59]. Alternatively, Kumar et al. suggested using a shallow neural network (NN) classifier after the AML slide is processed using wavelet transformation [37]. Other groups (Mohapatra et al.; Reta et al.; Escalante et al.) have compared multiple classifiers on leukemia image datasets and found that depending on the target, different classification methods appear to be the optimal solution [52, 58, 59].

When a small amount of data is available, conventional feature engineering-based machine learning algorithms provide fairly accurate predictions [39]. The accuracy of feature engineering proposed models depends on the distinct leukemia databases studied, the number and quality of the images, and the image acquisition mode; these require different data preprocessing steps. These methods are mainly based on supervised classification of leukemia subtypes. When the set of quantitative morphological features of the leukemia subtype is trained on a labeled dataset, then classifiers have been able to predict the four major leukemia types or the FAB classes applied to a test set. In case of insufficient number of training samples, Kasmin et al. proposed reinforcement learning to classify ALL, AML, CLL, and CML from PB cellular nucleus’ geometrical, texture, color, and statistical parameters [63].

Although these previous studies found new morphological features from the digitalized leukemia patient histology slides, and were successfully able to identify the major leukemia types and M0–M7 and L1–L3 subtypes, morphological features from the leukemia cells were not correlated with non-morphological information such as genetic mutations and clinical data. The morphological classification methods currently are not sufficient to recognize the majority.
of the underlying molecular abnormalities and cannot be used to direct therapy. In addition, the subtype groups’ underlying genetic patterns are not unique per subtype. Should morphological classification match genetic backgrounds, this could help speed up the diagnosis process. One study attempted to correlate morphological quantitative features in order to classify ALL lymphoblasts into the WHO subtypes and compare the results with flow cytometry analysis. To this aim, an unsupervised feature selection method was applied, and an optimal subset of the features was extracted to match the WHO classification [61]. This study and others that follow are helping pave the way for increasingly sophisticated means of classifying leukemia by images that enable incorporation of genetic and epigenetic details. Advances in computational methods are too, as the next section describes.

4.2.2. Deep learning methods

Although engineered feature-based conventional machine learning algorithms provide fairly accurate predictions, they do not reach the capability of human perception. The feature engineering process requires defining a carefully chosen set of features. This is a laborious process, and the feature parameters are very sensitive to the specific training set from where they were extracted. Due to this rigidity, a conventional machine learning algorithm likely could not be applied to a second dataset without parameter tweaking. To overcome these limitations, deep learning algorithms trained on large amounts of data can extract generalized features to perform human-level pattern recognition [64, 65].

When a large amount of data is available, for identifying morphological features in leukemia, a deep learning approach can be applied. Deep learning can self-discover new, hierarchical features in images (feature learning) allowing better pattern recognition for classification. These features are identified without human knowledge, and the learning approach is called “domain-agonistic,” where the computational system alone is able to distinguish distinct tissue types in any type of cancer. Today, with the increasing computing capacity of modern computers and the availability of big data storage, huge amounts of data can now be extracted and analyzed to identify key features for classification. This has enabled deep learning methods to outperform previous conventional machine learning approaches and to achieve higher accuracy [39, 66].

Deep learning is the extension of conventional, artificial neural networks where, instead of a single-layered network, a multilayered connected network processes input data and generates output. The network design is dependent on the input dataset and classification target. For pattern classification problems, convolutional neural networks (CNNs) are the ideally suited network design. The network learns from the example images fed to it and extracts hierarchical features automatically layer by layer (e.g., from low-level features like edges to higher-level features such as the cell, tissue, and then organ) without expert human intervention while retaining highly expressive power (Figure 3) [65–67].

The input of the CNN is a series of images, cropped from the whole slide image, and the images are processed in batch. For WBC classification, one cropped image contains one whole cell. Contrary to the cell-based analysis, for tissue classification, the images are slide-based, so the features are learned directly from the spatial pattern. The image size and the number of images fed to the network should be chosen carefully, and the variety of images should represent the
variability of the tissue type. Grayscale images are two-dimensional: width and height. Color images have a third dimension, depth, representing the RGB color channels [38, 65, 67, 68].

Once the set of images is defined and labeled, feature maps are created by sliding a series of filters representing shapes, textures, or colors over the input image (convolution), thus identifying local dependencies. The filters representing the features are learned during the training process through backpropagation and a gradient descent algorithm. After convolution, an activation process introduces nonlinear properties to the linear convolution to improve the model accuracy and to avoid overfitting. The convolutional layer then is down-sampled (pooling). This is successively repeated as many times as necessary according to the hierarchical complexity of the image. The last feature map is then flattened into a one-dimensional vector to feed a fully connected layer for neural network (NN) classification. The NN classification process can be replaced by a different classification scheme such as an SVM or random forest [38, 65, 67, 68].

Convolutional neural networks are ideally suited for pattern recognition and medical image analysis. In fact, CNNs have been successfully applied to feature learning to detect and diagnose a number of different cancers, including leukemia cells. Deep learning methods have been used for white blood cell detection and classification [68], lymphocyte detection [38], and lymphoma subtype classification [38] by identifying three subtypes of lymphoma: chronic lymphocytic leukemia (CLL), follicular lymphoma (FL), and mantle cell lymphoma (MCL). It also has been applied to the analysis of ALL cellular images to classify ALL subtype histopathology [67, 69].

Although the current research in pattern recognition is dominated by the supervised deep learning approach, the unsupervised approach is expected to provide breakthrough results in the near future, and extensive research is currently ongoing to optimize these algorithms [65, 66].

5. Conclusions and future outlook

Standard leukemia diagnosis and therapy are currently based on morphological classification of patients’ bone marrow smears and biopsies, peripheral blood smears, and molecular and cytogenetic analyses to identify genetic abnormalities. However, morphological and genetic classification analysis is insufficient to fully predict appropriate response to therapy, while

![Figure 3. Feature learning for classification.](image_url)
Emerging nonstandard methods to improve and personalize leukemia classification can be expensive and time-consuming. Digital pathology is emerging as a powerful, inexpensive tool to enhance biopsy- and smear-based decisions.

This review discussed how computational cytology can help improve leukemia diagnosis by enhancing pathologist smear-based decisions and improve leukemia diagnosis with automated, biologically meaningful pattern recognition. Techniques summarized in this review extract quantitative imaging features from stained bone marrow and peripheral blood smear samples to detect and classify leukemia. To identify morphological features, conventional machine learning approaches have been broadly applied to classify leukemia types and subtypes based on feature engineering. However, to acquire a new set of morphological features in leukemia, a deep learning approach would provide higher accuracy.

For most of the cases reviewed in this chapter, the image processing pipeline implements a supervised classification scheme, where the morphometric features are extracted from a set of labeled data (ALL vs. AML, FAB, M1, etc.) and then are validated on a test dataset. In future studies, supervised morphological analysis can be complemented with unsupervised classification schemes such as unbiased clustering. This approach could reveal whether entirely new classification schemes should be implemented for ALL or AML, independent from known acute or chronic leukemia subtype morphological classification. It also could potentially reveal common underlying genetic or proteomic patterns.

Emerging omics analysis methods are determining protein expression signatures for leukemia patients; however, these new processes can be time and labor intensive. To determine genetic information and protein signature membership rapidly and without the time delay required for proteomic-based signature assignment, advances in digital pathology offer potentially exciting, inexpensive, rapid alternatives. If morphological surrogates that reliably correlate with clinical, genetic, or proteomic features, either individually or in combinatorial patterns, can be identified directly from histology images, then this could significantly speed up leukemia diagnosis, reduce the cost of the diagnostic workup, optimize the assignment of patients to a particular therapy, and potentially uncover new pathways for drug targeting.

Cell metrics can be predefined manually, and often metrics are those known to be pertinent to leukemia cells. These algorithms, which together are employed as part of a “feature engineering process,” extract metrics from images based on features of cells (e.g., size or nucleus shape). Using a supervised classification approach, the metrics are extracted from predefined leukemia subtypes. As an example, a set of quantitative morphological features defining a leukemia subtype are trained on a labeled dataset according to the FAB morphological classes, and the resulting developed classifier is then used to predict the leukemia subtypes on a test set.

In the unsupervised classification approach, new clusters of leukemia subtypes are created from the engineered features. Contrary to the feature engineering process, learning algorithms self-discover features representative of leukemia cell types (feature learning) where features are learned from annotated (supervised) or unannotated (unsupervised) data (Figure 2).
Acknowledgements

This work is supported by Gulf Coast Consortia, Computational Cancer Biology Fellowship, and funded by CPRIT (Cancer Prevention and Research Institute of Texas) Grant Number RP170593.

Author details

Cecília Lantos1*, Steven M. Kornblau2 and Amina A. Qutub1

*Address all correspondence to: cecilia.lantos@rice.edu

1 Department of Bioengineering, Rice University, Houston, TX, USA
2 Department of Leukemia and Department of Stem Cell Transplantation, The University of Texas, Houston, TX, USA

References


