

# A Frugal Technological Model for Leukemia Detection Using Digital Microscopy

Juan López Prieto<sup>1</sup>, Miguel Purizaca Pérez<sup>2</sup>, Jimmy Armas Aguirre<sup>3</sup>, Paola A. Gonzalez<sup>4</sup>

<sup>1</sup>Universidad Peruana de Ciencias Aplicadas (UPC), Perú, u201312827@upc.edu.pe

<sup>2</sup>Universidad Peruana de Ciencias Aplicadas (UPC), Perú, u201314483@upc.edu.pe

<sup>3</sup>Universidad Peruana de Ciencias Aplicadas (UPC), Perú, jimmy.armas@upc.pe

<sup>4</sup>Dalhousie University, Canada, paola.gonzalez@dal.ca

*Abstract— In this paper, we propose a low-cost technological model for optimizing the detection process of leukemia using digital microscopy. The model is targeted to be used in remote areas in developing countries where availability of resources is an issue. This model applies the canny algorithm on a bank of images of normal and abnormal microscopic cells. The proposed model includes the capture, digitization, and analysis of microscopic samples. Fives phases are included in this model: 1. Data collection; 2. Data capture; 3. Image processing; 4. Cell classification; 5. Display of results. The model was preliminary validated with five blood samples from three men and two women in different age categories. All these samples were validated by the Head of Clinical Pathology at a public hospital in Callao, Lima, Peru. The results showed that a 90.5% effectiveness rate of white blood cell identification was obtained. This results aims to provide an additional and accurate tool to detection of potential leukemia.*

**Keywords—**frugal technology; leukemia; digital microscopy; image; processing.

## I. INTRODUCTION

Cancer is one of the major leading cause of death recently and has become a public health concern worldwide.. According to (refs), one in six deaths is caused by this disease. Detection and treatment of cancer are often an economical concern for health institutions and government.. In 2017, only 26% of low-income countries reported that public health had pathology services to serve the general population whereas more than 90% of high-income countries provide treatment for patients with cancer [1]. In Peru, statistics indicate that each year about 50.000 new cases of cancer are diagnosed, of which about 30.000 end in death. This is due to the fact that 75% of cancer cases in Peru are diagnosed in advanced stages and current treatments are often unable to cure the disease. [2].

Another area of public concern is the cases of cancer in children. In Peru, for instance, it is estimated that out of 1200 new cases of childhood cancer detected annually, 350 end in death. These estimates are constantly increasing due to the lack of economical and reliable solutions to detect cancer. In Peru, for instance, 60% of the neoplasms used to detect cancer are analyzed in advanced stages. Therefore, the process of cancer diagnosis in Peru takes almost three times more than in

developed countries. A patient can take three months to develop the first symptoms until a diagnosis is achieved. This is in part due to the low educational and socioeconomic level of parents, the poor training of primary care physicians, delays in the patient referral system and lack of equipment to conduct the testing [3].

Leukemia is the most common cancer in children and adolescents, accounting for nearly one in three cancers. About three out of four cases of leukemia in children and adolescents are of Acute Lymphocytic Leukemia (ALL), and most of the other cases are of Acute Myeloid Leukemia (AML) [4].

This paper proposes a technological model that digitizes or optimizes the current diagnostic process with low resources. This model is expected to serve as a tool for both the specialist medical staff and the primary care medical staff. For one hand, the model will provide specialists with more organized data to improve their process of diagnoses. On another hand, the model will help primary care medical staff improve their decision-making process in referral and prioritization of alarming cases. ..

This paper is structured as follows: first, we will analyze some existing solutions regarding the detection of Leukemia, then we will describe the details of the proposed model, and finally we will give some conclusions and discuss the results obtained in our pilot scenario.

## II. LITERATURE REVIEW

The identification of abnormal cells is a critical step in the process of diagnosing leukemia. This process is usually automated through the use of image processing techniques. Below, we briefly discuss some of the existing algorithms that help identifying such abnormal cells as well as the technologies that support this process of identification and classification of cells.

### A. Cloud Platform Evaluation

Cloud platforms allow the hosting of data and services which can be accessed from any device with just an Internet connection. Cloud solutions are often used to process large volumes of data due to its strong processing power as well as its low accessibility cost due to economy of scales.

Digital Object Identifier (DOI):  
<http://dx.doi.org/10.18687/LACCEI2019.1.1.345>  
ISBN: 978-0-9993443-6-1 ISSN: 2414-6390

Table I shows the three widely used cloud platforms worldwide

TABLE I. CLOUD PLATFORMS

Cloud Platforms	Description
MICROSOFT AZURE	Microsoft Azure provides a range of cloud services to meet the business needs of organizations, and allows users to build, manage and deploy applications on the network. Microsoft Azure is a hybrid network that provides coverage in accordance with global regulations, and it is also in compliance with the requirements of the General Data Protection Regulation (GDPR). Microsoft Azure offers 12 months of select free services, as well as 5 GB of memory [5].
SCP - SAP CLOUD PLATFORM	The SAP Cloud Platform is a platform as a Service (PaaS) that provides memory capabilities, core platform services, and unique microservices for building and extending intelligent, mobile-enabled cloud applications. The platform is designed to accelerate digital transformation by helping you quickly, easily and cost-effectively to develop the exact application you need, without investing in local infrastructure. Based on open standards, the SAP Cloud Platform offers complete flexibility and control over your choice of clouds, frames, and applications. The SAP Cloud Platform offers 10 GB of free storage, and the development time is indefinite [6].
IBM - BLUEMIX	Bluemix is a platform that supports various programming languages, as well as different services in the cloud. Bluemix is based on the open technology of Foundry Cloud, and runs on SoftLayer infrastructure. Its main products and services are: storage, mobile devices, Watson, data and analysis, Internet of things, security, application service and integration. Bluemix provides development and testing for a period of 30 days and offers 2 GB of memory for user use [7].

For our proposed technological model, we decided to use Microsoft Azure, because of the University alliance agreement that our university has with them. In addition to accessing these services for free, this cloud platform offers high processing power and large memory.

*B. Evaluation of microscope types for the capture of digital samples*

Digital microscopes are the best tool to take digital images of peripheral blood samples. This equipment are just variations from the traditional microscopes. These microscopes use a camera connected to an LCD or a computer monitor.

Table II compares the types of digital microscopes to capture digital samples. Our comparison is made across four important characteristics: zoom level, parts, and accessibility characteristics. Two references are also included as a medical validation of each tool.

TABLE II. TYPES OF DIGITAL MICROSCOPES

Characteristics	Optic Microscope + Digital Camare	Digital Microscope
Zoom level	> 200x	> 200x
Component parts	Microscope + Digital Camera + Graphic Card + Video Wires	Microscope + Video Wires
Medical validation	[8]	[9]
Accesibility	> S/. 14,000.00	> S/. 2000.00

We opted for the Digital Microscope “LabSystem DM-55 with LCD” due to the higher quality in analyzing digital microscopic samples as well as for its lower cost. .

III. THE PROPOSED FRUGAL MODEL FOR LEUKEMIA

*A. Grounds for the Contribution*

There are different technological solutions that facilitate the identification of abnormalities in microscopic cells indicating a carcinogenic characteristic. Most of these solutions, however, are very expensive and usually purchased by clinical centers specialized in oncology. Our proposed model proposes a low-cost solution that could be used in primary care centers where initial diagnoses can be detected. These centers are usually located in low-resource areas without appropriate equipment to carry out these medical investigations. In addition to offering a low-cost technology solution, the model is intended to be used by non-specialized medical staff (e.g., laboratory technicians, family doctors). If successful, this model has the potential to optimize the process of leukemia detection at early stages, therefore reducing the high fatality rates from this illness.

## B. General view

Edge detection is an image processing technique for finding the boundaries of objects within images. This technique is a critical issue when extracting objects because it reduces image data making it easier to detect these objects.

Table III shows a comparative analysis of the main algorithms used in edge detection. We compared them based on computational complexity, noise sensitivity, and detection of false edges in the images [10]. Computational complexity refers to resolution, missing of true edges and problems due to noise, noise sensitivity is how sensitive it is to identify random variations of the real image, and false edges refers to how many edges are mistakenly identified.

TABLE III. COMPARATIVE TABLE OF EDGE DETECTION ALGORITHMS

Nº	Algorithms	Complexity		Noise sensitivity	False Edges
		Time	Space		
1	Sobel	Lowest	High	Low sensitivity	More
2	Canny	High	High	Very Low sensitivity	Far less
3	Robert	High	High	Sensibility	More
4	Prewit	Low	Lowest	Very Low sensitivity	More
5	Laplacian of Gaussian	Low	Lowest	Very Low sensitivity	More
6	Zero crossing	Low	Low	Very Low sensitivity	More

Canny algorithm seems to be widely used in this field because has a lower percentage of false edge detection than the other algorithms. Although Canny rated high in computational complexity, we decided to use this algorithm in our proposed model as our cloud solution was able to handle high computing processing requirements.

## C. Specific view

A simple graphical representation of the Canny algorithm is presented in Figure 1. This algorithm first allows us to identify the edges, limits of the object and detect through abrupt changes the gray-level in order to detect the object.

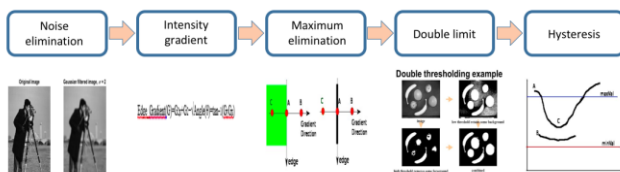


Fig 1. Canny Algorithm process

5 steps briefly describe how this algorithm works:

1. Applies the Gauss filter to soften the image, thereby removing noise from the image.
2. Identifies the intensity gradients of the image. In an image, a border always has one direction (either horizontal, vertical, or diagonal), so Canny uses four filters in four different directions to determine the borders of the image previously blurred by the Gauss filter.
3. Applies maximum suppression to the image, a technique of thinning axes, since the images previously generated by the gradients show very thick edges.
4. Uses the double-limit technique in which two images are generated by comparing each pixel with an intensity (each grayscale pixel has an intensity) and with a constant.
5. Removes the weak axes, hysteresis is applied. There, the pixels of each image are crossed and compared with its 8 neighboring pixels. If at least one of the neighbors is part of one of the previously found axes, it is considered that the pixel is part of a strong axis. Therefore, the pixel is highlighted or painted which allows us to identify the object of the image.

## D. Phases of the Proposed Model

The model proposed in Figure 2 consists of five phases: data collection, data capture, image processing, cell classification, and display of results. Our model intends to cover all the steps in the process of leukemia (or another type cancer) detection. Each phase is explained below.

### 1. Data collection

Two types of data are collected from the patient. The first one is the blood sample acquired through a complete blood count (CBC). The second one is the basic data of the patient: full name, age, and gender. This data is essential to complete the next phases because the microscopic image to be processed will be taken from the blood sample, and the date of the patient is needed to support medical diagnosis. Furthermore, this data can also be stored in a database and further analyzed and aggregated for medical research and health policy developments.

### 2. Data capture

In this phase, the microscopic image product taken from the patient's blood sample is digitalized. The digitalization is done through a digital microscope that captures the image and sends it through the MTP to a computer, where it will be stored in an Azure SQL Database along with the patient information.

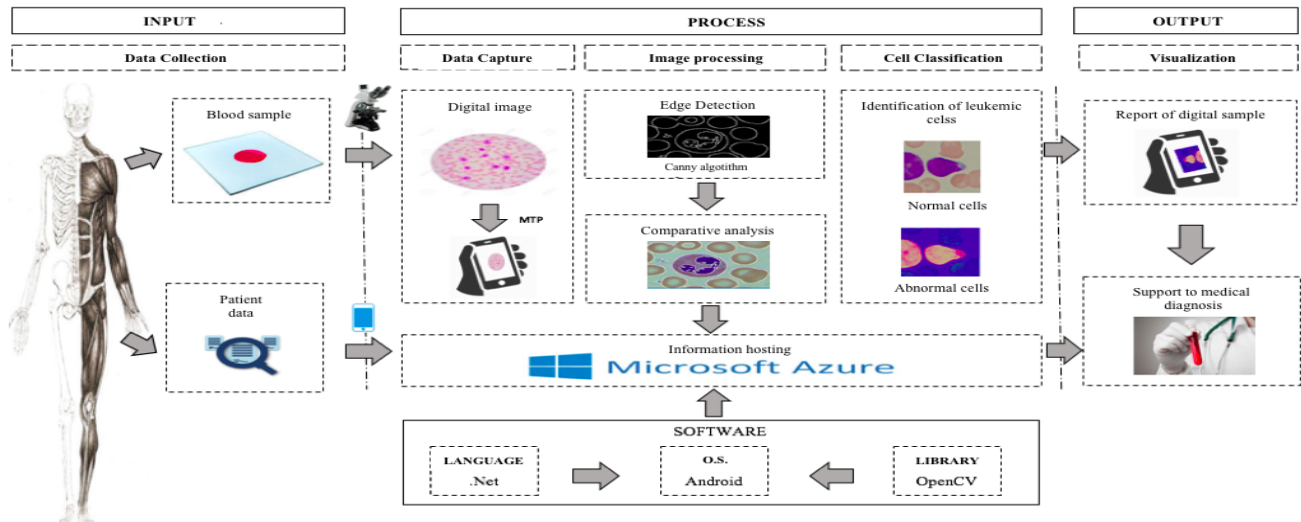


Fig 2. Technological Model

### 3. Data processing

This phase has two parts. First, an edge detection analysis of the images of the cells to be processed is made through the Canny algorithm. Once the edges of those images are identified, a comparative analysis against the images of cell stored in Azure SQL Database will be held. This will allow the identification of the leukocytes.

### 4. Cell classification

Once the comparative analysis of the images is done, the cells are classified into two groups starts: leukocytes and the rest of the blood cells. Only the leukocytes are painted in black to facilitate easy examination by a medical staff who will determine any anomalies in the sample.

### 5. Display of results

In this phase, all the analyses from the previous phases are shown to the medical specialist who will diagnose the patient based on all the findings or, if needed, he or she will request more information in order to clarify and support the diagnosis. It should be noted that the objective of the model is not to diagnose, but to work as a technological tool that will provide the doctor with more information when giving a diagnosis of this sort.

### E. Indicator view

A mobile application that runs the Canny algorithm supports this model. This algorithm will allow us to identify the white blood cells of the digital samples. Likewise, the model was measured according to the following indicators:

- Percentage of effectiveness of identified white blood cells: This indicator helped us to verify the effectiveness of the algorithm.
- Number of red blood cells, platelets or any other identified object. This indicator helped us to define a margin of error of the model.

## IV. EXPERIMENTATION

### A. Case study scenery

For the validation of the proposed model, we worked with five peripheral blood samples, provided by the Head of Clinical Pathology of the Hospital San José del Callao. The mission of this institution is to be a public hospital of the health sector that provides preventive promotional, recuperative and physical therapy to people, families and the population in a healthy environment, and to contribute to the sustained development of the country (Peru). The samples obtained are from five different patients, of which three had leukemia, and the other two did not. Furthermore, three of these patients were men and two were women, with age ranges of 10 to 15 years old, 35 to 40 years old, and 55 to 60 years old.

### B. Mobile application

This application described in section E was developed in .net and kotlin language designed for the Android Operating System. The OpenCV library was used to simplify the development of image processing algorithms. The main functionalities of the mobile application are the login of the medical specialist, the loading of the microscopic samples images and the data of the patient (full name, ID, gender and

age), the display of the processed image besides the original image, and the option which allows to find each sample previously registered per patient.

### C. Implementation

#### 1. Data collection

For the implementation of the proposed model, blood samples from patients were taken, as well as their demographic data. The blood samples were taken from the vein, located in the forearm. A drop was placed on a peripheral plate. Then it was combined with some dyes through the Wright Stain, which facilitates the differentiation of the types of blood cells, to be analyzed through the Digital Microscope. The data collected from the patient is: full name, ID, age, gender and date of birth.

#### 2. Data capture

In this stage, the microscopic image of the blood samples gathered in the previous stage was digitized. By using the Digital Microscope “LabSystem DM-25”, the images are captured and sent to a computer in order to store them in the cloud. In this case, the images of the five microscopic samples analyzed were stored for its later processing. The loading of these images of the samples was done by the mobile application as shown in Figure 3.

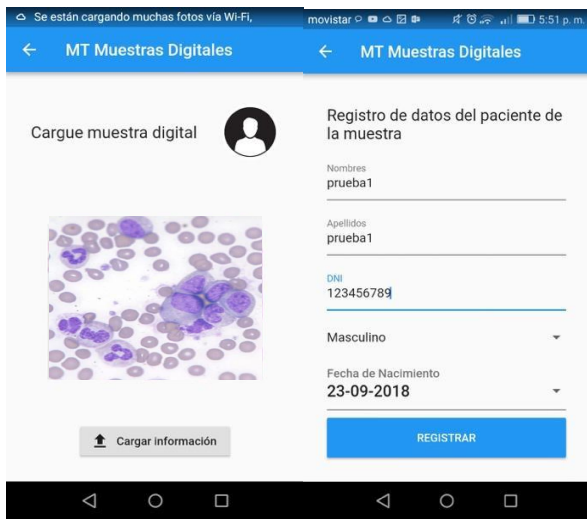


Fig 3. Data capture in the Mobile Application

#### 3. Image processing

For image processing, the five samples are processed in the mobile application. The result of the processing is the identification and staining of black color of the white blood cells nuclei. This graphical distinction is intended to aim the medical staff in his/her analysis.

#### 4. Cells classification

Once the images are processed, the cells are classified into two: white blood cell nuclei and others. Table IV shows the number of nuclei of white blood cells identified by each sample analyzed:

TABLE IV. WHITE BLOOD CELLS IDENTIFIED

# Sample	1	2	3	4	5
Quantity	14	7	4	4	1

#### 5. Visualization

Once leukocytes nuclei have been classified, both the original image and the processed image are displayed in the mobile application, as well as the information collected from the patient. Figure 4 shows the five different records of the samples studied in this validation.

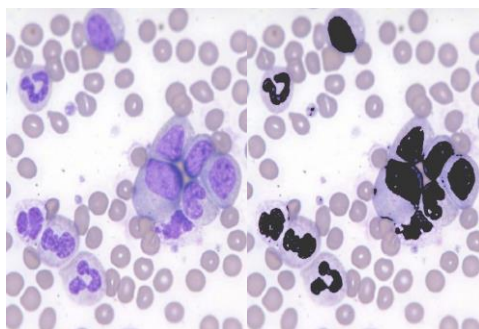


Fig 4. Identification of White Blood Cells

### D. Results

Based on the data collected, it is observed that the identification of the nuclei of the leukocytes is essential for a future diagnosis of any type of leukemia. The digital microscope is also crucial, because it is the tool that will allow us to take the images of the microscopic samples and digitize them for later analysis. For the validation of the model, the quantity of leukocytes identified by the specialist in hematology was compared to the quantity of leukocytes identified by the technological model through the mobile application. Based on this, the effectivity rate of each sample was calculated as shown in Table V:

TABLE V. % OF EFFECTIVITY OF SAMPLES

# Sample	1	2	3	4	5
% Effectiveness	81.8%	87.5%	100.0%	83.3%	100.0%

In average, a 90.5% effectivity rate and a 9.5% error rate were obtained. This error rate was caused because the model identified some figures that weren't leukocytes but that had some similar characteristics (color and shape). Accordingly, future studies must aim to reduce this error rate based on the additional parameter that will grant a better precision when identifying leukocytes.

### E. Conclusions

We proposed a technological model for optimizing the detection process of leukemia using the Canny algorithm through digital microscopy. This model allows the identification of leukocytes based on a bank of images of microscopic cells, and the parameter established for the detection of cells (size and color). Also, the model was evaluated through all its components (PC, microscope and camera) resulting in \$4,146, which is a low-cost compared to another solution such as the ThinPrep image processor that costs \$44,480.

The model was tested at a public hospital in Callao, Lima, Peru. The validation results showed a 90.5% precision rate for the identification of leukocytes, and a 9.5% error rate. Both rates were calculated based on the quantity of leukocytes identified by a medical specialist in regard to the quantity of leukocytes identified by the proposed model. The medical staff pointed out that the model has the potential to significantly reduce the detection times of leukemia by 64%. This percentage was calculated based on the average time for Leukemia detection of the 3 leukemia samples of the experimentation, the average was 32 days, when actually, in Peru the average is 90 days. This proposed model is intended to be used at primary care centers usually located in zones of poverty of Peru, lacking personnel specialized in the analysis of these samples, infrastructure, and tools used for a proper diagnosis. Finally, the medical specialist validated the positive social and economic impact that such a model can have by contributing to Peru's development in terms of the use of technology in the health sector.

## V. REFERENCES

[1] Organización Mundial de la Salud. OMS – Cancer. [online] Available: <http://www.who.int/es/news-room/fact-sheets/detail/cancer>

[2] Ministerio de Salud del Perú (2018) - Ministerio de Salud cuenta este año con presupuesto de 733 millones para desarrollar acciones orientadas a combatir el cáncer. [online] Available: <http://www.minsa.gob.pe/index.asp?op=51&nota=26897>

[3] Perú21 (2018) - En el Perú cada año mueren 350 niños a consecuencia del cáncer [INFORME] [online] Available: <https://peru21.pe/peru/cifras-cancer-peru-ano-mueren-350-ninos-consecuencia-enfermedad-395940>

[4] American Cancer Society (2016) - ¿Cuáles son las estadísticas importantes de la leucemia en niños? [online] Available: <https://www.cancer.org/es/cancer/leucemia-en-ninos/acerca/estadisticas-clave.html>

[5] MICROSOFT AZURE (2018) – Información general, ¿Qué es Azure? [online] Available: <https://azure.microsoft.com/es-es/overview/what-is-azure/>

[6] SAP CLOUD PLATFORM (2018). SAP Cloud Platform: Overview. [online] Available: <https://cloudplatform.sap.com/index.html>

[7] IBM BLUEMIX (2018). Plataforma de cloud: infraestructura de cloud. [online] Available: <https://www.ibm.com/cloud-computing/bluemix/es>

[8] Oscanoa, C, & Mena, M. (2015). Desarrollo de un aplicativo de software, con acceso remoto vía web, orientado a mejorar la calidad del diagnóstico de las pruebas de Papanicolau, utilizando algoritmos computacionales de procesamiento digital de imágenes. UPC, Perú.

[9] Mehedi Hasan, Mohammad Wajih, Khan Wahid, Sayern Miah & Jiven Eriqe. (2016). A Low-Cost Digital Microscope with Real-Time Fluorescent Imaging Capability, in PLOS ONE. [online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167863>

[10] IEEE TGRS Vol.50 no.11b (2014) - Comparative analysis of common edge detection techniques in context of object extraction [online] Available: <https://arxiv.org/pdf/1405.6132.pdf>