Leukemia detection system using convolutional neural networks by means of microscopic pictures

Pathan Mohd Shafi¹, Vijaykumar Bidve², Haribhau Bhapkar¹, Prashant Dhotre¹,
Veer Bhadra Pratap Singh²
¹MIT Art, Design and Technology University, Pune, India
²School of CSIT, Symbiosis Skills and Professional University, Pune, India

ABSTRACT

All over the world, there are a significant number of patients suffering every year from blood cancer. Most of the people are unaware of the risk involved in such a disease. A majority of these diseases are dangerous and may cause death. The patient who has been diagnosed with such a disease, feels very afraid. The patient may feel that the disease is very uncontrolled. Such diseases are very uncommon, and the patient may get very less assistance and information available about this disease. This symptom is called as acute lymphocytic leukemia (ALL) in medical science. In such a kind of cancer, white blood cells are mostly affected. In case of children, this disease is mainly detected i.e. children are more prone to this disease. If the disease is diagnosed in the early stage, the chances of recovery are maximum. Hence, there should be an accurate and guaranteed mechanism available to detect such type of blood cancers in the patients. This work proposes a system to distinguish the three different types of ALL using a convolutional neural network (CNN) by means of microscopic pictures of peripheral blood smears (PBS) and obtain accuracy levels that surpass those of practicing physicians.

Keywords: Blood cancer, Convolutional neural network, Classification, Detection, Leukemia

This is an open access article under the CC BY-SA license.

Corresponding Author:
Vijaykumar Bidve
School of CSIT, Symbiosis Skills and Professional University
Kiwale, Pune, India
Email: vijay.bidve@gmail.com

1. INTRODUCTION

The bone marrow of leukemia patients exhibits rapid, unchecked proliferation of abnormal cells [1]. Unlike other diseases, leukemia usually does not form a mass (tumour) that can be detected by imaging tests such as X-Rays [2]. Hematopoietic stem cells are progenitor cells of all blood cells, and undergo several developmental stages before they become adults [3], [4].

When a person suffers from leukemia, one of these types of blood cells begins to develop rapidly, and because of this are out of control. These abnormal cells, known as leukemia cells, arrogate the space inside the bone marrow [5]. Leukemia can now be identified by automatic specialised tests, such as cytogenetics, immunophenotyping, and morphological cell categorization [6]. Still, the drawback is that they necessitate expert operators to scrutinise microscopic pictures of the blood or bone marrow, which also leads to a substantial delay in the treatment procedure [7], [8]. These approaches are not employed in cases with regular symptoms. Early detection of acute lymphoblastic leukemia (ALL) symptoms in individuals can improve their chances of survival [9]. The approach of blood cell observation employing cytogenetics and immunophenotyping diagnostic procedures is now recommended for its high accuracy. The difficulty with these procedures depends on the operator’s capabilities and enervation [10], [11].

Journal homepage: http://ijeecs.iaescore.com
The use of microscopic pictures to identify leukemia in human blood samples is only fitting for low-cost and remote diagnosis systems. That’s where new-age solutions come into play [12]. Using deep learning, machine learning and neural networks, various researchers have developed systems that provide a smooth and exceptionally accurate way to detect and classify different types of blood cancer [13]. Leukemia detection system using CNN. Out of the total cases of blood cancer ALL is found in 74% of the cases. Because of common symptoms, the normal diagnosis is delayed, and so therapy is delayed [14]. Early detection is the most important parameter that leads to a faster recovery rate and survival [15]. As the cost of diagnosis of this disease is high, many people of the society, cannot reach up to the level of diagnosis in the most of the cases [16]. There is a need to find a solution, which can help in the diagnosis of such diseases with minimum cost, and will be available at the grassroot level.

In case of this disease if the treatment is started earlier, then the chances of survival of the patient are more. Therefore, the early detection of the disease is the most important way to start the treatment quickly [17]. To detect such a disease manually, by using a blood sample is time and resource consuming [18]. Typically, individuals in pain, do not have the luxury of wasting time since they require quick attention. To obtain the results faster, a more advanced and artificial intelligence-based system is required. A system with advanced technologies will be more accurate and faster, than the traditional systems.

The literature review is conducted to get the information of the existing systems available for cancer detection. A total of 10 papers are shortlisted for the review, which are more relevant to the topic of the proposed work. The subsequent section gives the details of the literature survey, carried out in the study. Bibi et al. [19] suggested the concept of internet of medical things (IoMT). The purpose of their framework is to provide fast and secure detection of the disease. They linked the system with cloud and networks. The system proposed is time efficient, and requires less efforts. The system is also helpful to resolve problems of the patients suffering from Covid-19. The convolutional neural network (CNN) is used for the identification of leukemia. The authors have used databases like ALL-IDB and ASH image bank. The model is efficient for leukemia identification as compared to the existing models. Ghaderzadeh et al. [20] stated that the use of machine learning in the field leukemia detection is more than 97%. The machine learning and deep learning techniques lead to the highest precision and accuracy. The machine learning techniques are helpful to reduce time, cost and provide a safe and faster diagnosis service. The machine learning services can be easily adoptable at a laboratory and a clinic.

Salah et al. [21] described that the machine learning techniques are better for diagnosis in case of microscopic image input. The study categorised the leukemia into four types. The study suggested to develop a field of machine learning to transform the design algorithm to practical use. The discussed multiple articles are related to machine learning for leukemia detection. Its an observation of the authors, that no study of the literature is applied by means of the machine learning algorithm in the real world scenario. Hosain et al. [22] focused on ALL, a most common type of leukemia. The authors stated that cancer is more easier to treat if detected in the early stage. The authors proposed a solution by detecting irregular blood cells. They detected 14 attributes and used 4 main common attributes, for the detection of leukemia. They used the machine learning algorithm to predict formation of odd cancer cells. Two loss functions like CNN and the classification model are used to detect similar blood objects. After identifying the object, the corresponding objects are detected and finally, the leukemia is detected. The precision and epochs are observed by the authors.

Litjens et al. [23] used the deep learning approach, to increase efficiency and objectivity of histopathologic analysis. The technique used reduces work load, and increases time efficiency of the diagnosis. The deep learning techniques can be used for automation in the field of cancer detection, and reduce human intervention. The concept of deep leaning can be used to improve the efficacy of cancer diagnosis. Jagadev et al. [24] explained the techniques for detection of leukemia using image processing and proposed automation in the detection process. The authors suggested classification techniques on the basis of features identified in the blood cells. The authors used a dataset containing blood smear images of normal and cancer patients. Ratley et al. [25] classified leukemia into two types, and further each type is categorised into two types. The authors analysed various image processing and machine learning techniques used for classification of leukemia. The authors focused on metrics, and the limitations of the techniques used in the literature. Raje and Ragole [26] proposed a system based on examining microscopic images, for the detection of leukemia. The work does an initial segmentation using statistical parameters, which differs the white blood cells from the other blood components. Geometric parameters of white blood cells are used for diagnostic prediction of leukemia. The authors applied this system based approach to a large number of images, and checked the results. The system uses various image processing algorithms for the detection of the disease.

Zhao et al. [27] proposed a system of automatic detection and classification of white blood cells on the basis of peripheral blood images. The system proposed an algorithm to detect white blood cells on the basis of color and morphological operator. Then, the granularity feature and SVM algorithm is applied for the classification of white blood cells. The CNN is used for automatic feature extraction of white blood cells. The random forest algorithm is used to find three kinds of white blood cells. The authors claimed that their system

Leukemia detection system using convolutional neural networks ... (Pathan Mohd Shafi)
gives better results with respect to time, cost and accuracy. Pansombut et al. [28] proposed a classification of all subtypes of leukemia. The convolutional neural network approach is used for the classification. The authors identified a good potential in image classification, without the multistep featuring approach. The above mentioned approaches surveyed in the literature have used different techniques for the detection of leukemia, but none of the techniques have an accuracy greater than 85%. This work proposes a technique to distinguish three different types of ALL with the help of microscopic images of peripheral blood smears to achieve a high accuracy level than the existing systems.

The literature review work of all the above authors is summarized in the form of a table. The Table 1 shows the summary. The method, findings and limitations are used to summarize the work described by various authors in the literature.

Table 1. Summary of literature review

<table>
<thead>
<tr>
<th>Sr. no.</th>
<th>Title</th>
<th>Author</th>
<th>Method</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoMT-based automated detection and classification of leukemia using deep learning</td>
<td>Bibi et al. [19]</td>
<td>CNN</td>
<td>Exact percentage accuracy is not mentioned</td>
</tr>
<tr>
<td>2</td>
<td>Machine learning in detection and classification of leukemia using smear blood images: a systematic review</td>
<td>Ghadezadeh et al. [20]</td>
<td>Machine learning and deep learning</td>
<td>Time and cost reduction percentage is not mentioned</td>
</tr>
<tr>
<td>4</td>
<td>Leukemia detection mechanism through microscopic image and ML techniques</td>
<td>Hossain et al. [22]</td>
<td>CNN and classification</td>
<td>The parameters of blood cell classification are not mentioned</td>
</tr>
<tr>
<td>5</td>
<td>Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis</td>
<td>Litjens et al. [23]</td>
<td>Deep learning</td>
<td>The parameters for automation and efficacy are not mentioned in detail</td>
</tr>
<tr>
<td>6</td>
<td>Detection of leukemia and its types using image processing and machine learning</td>
<td>Jagadev et al. [24]</td>
<td>Classification</td>
<td>The classification process is unclear</td>
</tr>
<tr>
<td>7</td>
<td>Leukemia disease detection and classification using machine learning approaches: a review</td>
<td>Ratley et al. [25]</td>
<td>Machine learning</td>
<td>Metrics for classification is mentioned by the authors, but not explained in detail</td>
</tr>
<tr>
<td>8</td>
<td>Detection of leukemia in microscopic images using image processing</td>
<td>Raje and Rangole et al. [26]</td>
<td>Image processing</td>
<td>Segmentation of images is not mentioned in detail</td>
</tr>
<tr>
<td>9</td>
<td>Automatic detection and classification of leukocytes using CNN</td>
<td>Zhao et al. [27]</td>
<td>Granularity, SVM</td>
<td>The granularity levels and SVM inputs are mentioned in detail</td>
</tr>
<tr>
<td>10</td>
<td>Phon-leukemia detection system using CNN</td>
<td>Pansombut et al. [28]</td>
<td>CNN</td>
<td>The multifeature approach of image classification is not mentioned in detail</td>
</tr>
</tbody>
</table>

The summary described in the Table 1 shows that there are limitations in the papers described in the literature. No author has given the implementation details, architecture and results clearly. Also, the percentage accuracy is not clearly defined. In order to overcome all these issues, a system is proposed in this work using the convolution neural network. The section 2, describes method of the implementation, section 3 describes the results and section 4, gives the conclusion of the paper.

2. METHOD

This section describes a detailed method of the system implementation, dataset used and processing. The proposed system uses the kaggle dataset for ALL images. The kaggle dataset comprises of around 3256 peripheral blood smear images of 89 patients suspected of ALL.

The Figure 1 shows a detailed system architecture of the proposed system. The architecture contains multiple stages including training, testing, feature extraction, input, classification and result. The model is trained using CNN to gain the highest accuracy levels. Segregating the dataset using a 70:30 ratio for training and testing, system trained the model employing the VGG16 architecture. Also, an adaptive optimization algorithm called ADAM is used to train the model. ADAM is optimization technique used for training model in deep learning. ADAM combines the best properties of an image to provide optimization which handles
sparse gradients. ADAM implements adaptive changes to learning rate of the optimizer adding power exponential correction factor. The algorithm helps to diagnose the disease more accurately. The system processes the 224×224-pixel images in the dataset by segregating them according to the subtype of ALL, ranging from early precursor B-Cell ALL to precursor B-Cell ALL and Pro B-Cell ALL. An option for benign blood samples that contained hematogones resembling ALL is also added in the system. The CNN is used to extract the different features of every image, like the size of the displayed leukocytes and the frequency of them occurring in the blood, and segregate them into three subtypes. The model is tested using the testing dataset, and has achieved accuracy levels of 85%, with much room for improvement if the data is augmented further.

Figure 1. System architecture

3. RESULTS AND DISCUSSION

The result and discussion section describes the detailed results of the proposed system. The section is divided into three parts including symptoms, diagnostics and treatment. Each of the sections are discussed as in (3.1)-(3.3).

3.1. Symptoms section

This section contains a list of symptoms that can be found with the disease. It also has a visual image demonstrating the symptoms. The Figure 2, gives the symptoms of leukemia suspected blood samples. The Figure 2 is the initial part of the implemented system.

Figure 2. Symptoms section

Some symptoms can be vague. They include, fatigue, fever, loss of appetite or weight loss, night sweats, tiny red spots just under the skin (petechiae), stomach pain, etc. 

- **A lack of red blood cells** may cause symptoms of anaemia, including fatigue or weakness, dizziness or light-headedness, feeling cold, shortness of breath, etc.
- **A lack of healthy white blood cells** may cause fevers, more infections than usual, etc.
- **A lack of platelets** may cause lots of bruising for no clear reason, frequent or severe nosebleeds, bleeding gums, or other unusual bleeding, such as from minor cuts, etc.
- Depending on where the leukemia cells are, the patient might experience a full or swollen belly from cancer cells in the liver or spleen, enlarged lymph nodes such as in the neck or groin, under the arms, or above the collarbone, bone or joint pain, headache, trouble with balance, vomiting, seizures, or blurred vision if the cancer has spread to the brain, trouble breathing if it’s spread to the chest, etc.
3.2. Diagnostic section

This section is the heart of the entire system. This is where artificial intelligence comes into play. This page of the system, has the feature to upload an image of a peripheral blood smear to exact a diagnosis of whether the disease is present in the blood sample. It also gives information about the particular subtype of ALL that has been diagnosed. This page of the system, contains a procedure for medical diagnosis, result of diagnosis, the last part of the page gives detailed information of the status of the disease. The stage of the disease is also detected and informed in the last part of this page of the system. Figure 3, shows the diagnostic section of the system.

![Diagnostic section](image)

Figure 3. Diagnostic section

3.3. Treatment phases

This phase of the system, provides detailed guidance regarding treatment in case the disease is detected. The treatment is to be carried out in three different stages including induction, consolidation and maintenance. The information of all these treatment stages is provided by the system in this phase. Figure 4, shows in detail how the information is provided by the system regarding the various stages of treatment.
4. CONCLUSION

The proposed system facilitates healthcare systems for early detection and diagnosis of ALL. The system is easy to use for healthcare personnel. The system provides detailed diagnosis of the disease. Also, information is provided regarding the symptoms and various phases of the treatment. The system could achieve 85% and more accuracy in the detection of ALL by means of a CNN. This study aims to have a significant contribution to the fight against leukemia. In future, a more complete dataset with a more accessible user interface and complete information available on the homepage is targeted.

ACKNOWLEDGEMENTS

We acknowledge all the people who directly or indirectly guided us in the development of this work. All the authors contributed in the development of the work. We are grateful to family members, friends and colleagues who helped and guided us directly and indirectly for the development of this system, and preparation of this paper. Our special thanks to our employers for their support from the very beginning, that has come a long way, for the publication of this paper.

REFERENCES


**BIographies of Authors**

Dr. Pathan Mohd Shafi is Professor at MIT art, design and technology University, Lonli Kalbhor, Pune, India. He holds a Ph.D. degree in computer science and engineering with specialization in public key cryptography for cross-realm authentication in Kerberos. He has worked as the resource person for workshops and seminars. He has published more than 55 research articles in national and international journals. He is a life member of ISTE and CSI. His research areas are cryptography, machine learning. Dr Shafi has published number of patents. He is working as an expert for various subjects. Also, he has worked as a reviewer for various conferences and journals. He can be contacted at email: shafipathan@gmail.com.
Dr. Vijaykumar Bidve is Associate Professor at symbiosis skills and professional University, Kiwale, Pune, Maharashtra, India. He holds a Ph.D. degree in computer science and engineering with specialization in software engineering. His research areas are software engineering, machine learning. Dr. Vijaykumar has published number of patents. He has published more than 40 research articles in national and international journals. He is a life member of ISTE. He is working as an expert for various subjects. Also, he has worked as a reviewer for various conferences and journals. He can be contacted at email: vijay.bidve@gmail.com.

Dr. Haribhau Bhapkar is working as a Professor of mathematics at MIT ADT University’s School of engineering and science, Pune. He received masters in mathematics in 2000, and Ph.D. (graph theory) in 2016. He has published 42 research papers in national and international reviewed journals, mostly in the areas of graph coloring, graph isomorphism, mathematical modelling, COVID-19 and applications. He has authored/co-authored more than 140 chapters/books of engineering mathematics and general books published by Springer Publication, Elsevier Publication, De Gruyter, and Technical Publication. He has registered 30 copyrights and published 27 patents. He can be contacted at email: hrbhapkar@gmail.com.

Dr. Prashant Dhotre is working as professor in the department of information technology at MIT art, design and technology (MIT ADT) University, Pune, India. He has been awarded with a doctorate degree in computer science and engineering from Aalborg University, Denmark in 2017. His Ph.D. thesis title was “systematic analysis and visualization of privacy policies of online services”. He has more than 18 years of teaching and research experience. He has carried out funded research projects on “location and identity privacy in internet of things” and “drowsy driver detection using image processing” by Savitribai Phule Pune University, Pune. He has 40 publications in international journals and conferences. He can be contacted at email: prashantsdhotre@gmail.com.

Mr. Veer Bhadra Pratap Singh is working as assistant professor in the school of CSIT at symbiosis skill and professional University, Kiwale, Pune, Maharashtra. He had completed his B.Tech. from UPTU, Lucknow, India, master of science in Web Information Systems from University of Ulster, United Kingdom, and M.Tech. from Dr. A P J University, Indore, M.P. and is currently pursuing his Ph.D. from Maha Kaushal University, Jabalpur, M.P. He has published more than 12 papers in various journals and conferences of repute and also published 10 patents. Area of interest is machine learning, artificial intelligence, algorithms. He can be contacted at email: er.veerpratap@gmail.com.

Leukemia detection system using convolutional neural networks ... (Pathan Mohd Shafi)