Studies in Big Data 109

Tanzila Saba
Amjad Rehman
Sudipta Roy *Editors*

Prognostic Models in Healthcare: AI and Statistical Approaches

Springer

Studies in Big Data

Volume 109

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are reviewed in a single blind peer review process.

Indexed by SCOPUS, EI Compendex, SCIMAGO and zbMATH.

All books published in the series are submitted for consideration in Web of Science.

Tanzila Saba · Amjad Rehman · Sudipta Roy
Editors

Prognostic Models in Healthcare: AI and Statistical Approaches

Springer

Editors
Tanzila Saba
Artificial Intelligence and Data Analytics Lab
College of Computer and Information Sciences
Prince Sultan University
Riyadh, Saudi Arabia

Amjad Rehman
Artificial Intelligence and Data Analytics Lab
College of Computer and Information Sciences
Prince Sultan University
Riyadh, Saudi Arabia

Sudipta Roy
Artificial Intelligence and Data Science
Jio Institute
Ulwe, Navi Mumbai, India

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-981-19-2056-1 ISBN 978-981-19-2057-8 (eBook)
https://doi.org/10.1007/978-981-19-2057-8

Preface

Medical imaging issues are so complex owing to high importance of correct diagnosis and treatment of diseases in healthcare systems. Recent research efforts have been devoted to processing and analyzing medical images to extract meaningful information such as volume, shape, motion of organs, to detect abnormalities, and to quantify changes in follow-up studies. Medical image analysis is diverse, and the large amount of information introduced through the hybrid systems requires next generation of image quantification that need to be addressed. This book addresses the issues and describes current advanced method in interactive medical image analysis.

The book is organized in 18 chapters.

The chapter "Segmentation of White Blood Cells in Acute Myeloid Leukemia Microscopic Images: A Review" presents state-of-the-art computer-aided diagnosis (CAD) systems as an accurate diagnostic tool for AML and assist pathologists during the diagnosis process. Segmentation of WBC is the first step toward developing an accurate CAD system for AML. To date, WBC segmentation has several challenges due to several reasons such as different staining conditions, complex nature of microscopic blood images, and morphological diversity of WBCs. Current WBC segmentation techniques vary from conventional image processing methods to advanced machine learning and deep learning methods. This chapter discusses current segmentation methods as well as the potential solutions for improving automated WBC segmentation accuracy.

The chapter "Computer Vision-Based Prognostic Modelling of COVID-19 from Medical Imaging" examines prognostic models for COVID-19 patients' survival prediction based on clinical data and lung/lesion radiometric characteristics retrieved from chest imaging. While it seems that there are various early indicators of prognosis, we will discuss prognostic models or scoring systems that are useful exclusively to individuals who have received confirmation of their cancer diagnosis. A summary of some of the research work and strategies based on machine learning and computer vision that have been applied for the identification of COVID-19 have been presented in this chapter. Some strategies based on preprocessing, segmentation, handmade features, deep features, and classification have been discussed, as well as some other techniques.

The chapter "An Accurate Skin Lesion Classification Using Fused Pigmented Deep Feature Extraction Method" handles challenge for skin lesion classification for low contrast and over-segmented images. According to the literature surveyed, available hand-crafted features could not generate better results when the skin lesion images contain low contrast, under and over-segmented images. The hand-crafted features for skin lesions did not discriminate well between the two significantly different densities. The pigmented network feature vector and deep feature vector have been fused using a parallel fusion method to increase classification accuracy. This optimized fused feature vector has been fed to machine learning classifiers that accurately classify the dermoscopic images into two categories as benign and malignant melanoma. The statistical performance measures were used to assess the proposed fused feature vector on three skin lesion datasets (ISBI 2016, ISIC 2017, and PH2). The proposed fused feature vector accurately classified the skin lesion with the highest accuracy of 99.8% for the ISBI 2016, an accuracy of 99.3% for the ISIC 2017 dataset, 98.6% for the PH2 dataset.

The chapter "COVID-19 Prediction, Diagnosis and Prevention Through Computer Vision" presents various computer vision (CV) technologies along with other artificial intelligence (AI) subsets have significant potential to fight in frontline of this turbulent war. Normally radiologists and other clinicians are using reverse transcript polymerase chain reaction (RT-PCR) for diagnosing COVID-19, which requires strict examination environment and a set of resources. Further, this method is also prone to false-negative errors. One of the potential solutions for effective and fast screening of doubtful cases is the intervention of computer vision-based support decision systems in healthcare. CT-scans, X-rays, and ultra-sound images are being widely used for detection, segmentation, and classification of COVID-1. Computer vision is using these modalities and is providing the fast, optimal diagnosis at the early stage controlling mortality rate. Computer vision-based surveillance technologies are also being used for monitoring physical distance, detecting people with or without face masks, screening infected persons, measuring their temperature, tracing body movements, and detecting hand washing. In addition to these, it is also assisting in production of vaccine and contributing in administrative tasks and clinical management. This chapter presents an extensive study of some computer vision-based technologies for detection, diagnosis, prediction, and prevention of COVID. Our main goal here is to draw a bigger picture and provide the role of computer vision in fight against COVID-19 pandemic.

The chapter "Health Monitoring Methods in Heart Diseases Based on Data Mining Approach: A Directional Review" explores data mining techniques for identifying and diagnosing diseases, categorizing patients in disease management, and finding patterns to diagnose patients more quickly and prevent complications. Increasing the accuracy of diagnosis, reducing costs, and reducing human resources in the medical sector have been proven by researchers as the benefits of introducing data mining in medical analysis. Heart disease is evaluated to make the study more comprehensive, including fetal health diagnosis, arrhythmias, and machine learning data mining angiography. Attempts are made to introduce the relevant database in each disease and to evaluate the desired methods in health monitoring.

The chapter "Machine Learning-Based Brain Diseases Diagnosing in Electroencephalogram Signals, Alzheimer's, and Parkinson's" emphases that brain monitoring tools are used to detect these diseases early. An inexpensive and useful tool, as well as low-risk brain signals, are electroencephalograms. In order to analyze brain signals, the use of machine learning-based methods has been able to show its superiority. In order to diagnose Alzheimer's and Parkinson's in machine learning, there are preprocessing steps, feature extraction, feature selection, classification, and evaluation. Since electroencephalogram data have high repetition and correlation in different channels recorded on the head, feature extraction techniques will be of great importance. Feature selection methods seek to select the most effective features to classify and identify disease status. Finally, the selected features will be categorized using different categories. In this chapter, a complete overview of the stages of diagnosis of these diseases with the help of machine learning will be provided.

The chapter "Skin Lesion Detection Using Recent Machine Learning Approaches" states skin cancer is the highest popular form of cancer. Sunlight, ultraviolet rays, moles, and many other reasons cause skin cancer. Skin cancer can be treated if it is diagnosed at the premature stage. Manually diagnosing skin cancer is a time-consuming procedure, requiring a lot of human power while it is a grueling procedure. Various approaches for automatically detecting skin cancer have now been developed in recent years as technology has improved. In this chapter, skin lesion detection steps like preprocessing (to remove noise from images), segmentation (to get skin lesion location), feature extraction, feature selection, and classification methods have been discussed in detail. Furthermore, limitation and gaps in the domain of skin lesions are also discussed that provide help for the researchers.

The chapter "Improving Monitoring and Controlling Parameters for Alzheimer's Patients Based on IoMT" states that currently the Internet has become an integral part of people's lives. With the spread of the Internet, and the diversity of Internet applications, a new type of Internet use called the Internet of Things (IoT) has emerged. In the Internet of Things, information is collected, managed, and communicated in the daily life of man through the Internet. In this chapter, an improved method, low power and lossy network, is proposed to control and monitor the Alzheimer's patient in the cloud robot on the Internet of Things in smart homes. In the proposed method with load balancing in the routing protocol in LLN networks based on RPL is presented. The proposed method improves the structure of the pair-to-pair (P2P) path. Data packets are sent as RPL sorted and irregular. Paths sent in P2P mode have been improved to reduce computational overhead and balance load on the network. Elimination of control messages and load balancing in routing are among the advantages of the proposed method.

The chapter "A Novel Method for Lung Segmentation of Chest with Convolutional Neural Network" states that medical images have made a high impact on medicine, diagnosis, and treatment. The most important part of image processing is image segmentation. This chapter presents a novel X-ray of lungs segmentation method using the U-Net model. First, we construct the U-Net which combine the lungs and mask. Then, we convert to problem of positive and negative TB lungs into the segmentation of lungs and extract the lungs by subtracting the chest from

the radiography. In experiment, the proposed model achieves 97.62% on the public dataset of collection by Shenzhen Hospital, China, and Montgomery County X-ray Set.

The chapter "Leukemia Detection Using Machine and Deep Learning Through Microscopic Images—A Review" presents that there are numerous studies for the detection of acute leukemia, but there are only a few studies to detect chronic leukemia. Additionally, microscopic-based methods can be used to analyze microscopic smear images and detect the incidence of leukemia automatically and quickly. It also discusses the benefits, drawbacks, and limitations of a variety of traditional artificial intelligence-based approaches for detecting leukemia, such as machine learning and deep learning. Hence, this chapter aims to review the existing literature in the field of medical image processing of blood smear images, with a focus on automated leukemia detection. The analysis of various studies shows that deep learning techniques provide the best results compared to machine learning techniques. Hence, the major drawback in recent studies is that most of the research has been done on locally available datasets.

The chapter "A Review on Machine Learning-Based WBCs Analysis in Blood Smear Images: Key Challenges, Datasets, and Future Directions" presents that manual detection, counting, and classification of WBCs are very slow, challenging, and boring task due to complex overlapping and morphological uneven structure. In this chapter, we provide a concise analysis of available ML techniques to use these techniques for leucocytes analysis in microscopic images. The main aim of this chapter is to identify high-performance and suitable ML algorithms for WBCs analysis using blood microscopic smear images. In the proposed review study, the recent and most relevant research papers are collected from IEEE, Science Direct, Springer, and Web of Science (WoS) with the following keywords: "leucocytes detection" or "leucocytes classification." This study gives an extensive review of MIA, but the research focuses more on the ML-based leucocytes/WBCs analysis in smear images. These techniques include traditional machine learning (TML), deep learning (DL), convolutional neural network (CNN) models, hybrid learning, and attention learning-based techniques to analyze medical image modalities to detect and classify cells in smear images.

The chapter "Automatic Detection of Liver Cancer Using Artificial Intelligence and Imaging Techniques—A Review" is a systematic review that evaluates several types of researches and advanced technologies that can help to diagnose liver cancer automatically. Through this review of 26 relevant articles, the following syntheses of liver cancer detection techniques have been produced: a) the use of machine learning (ML) and deep learning (DL) methods and b) the use of classical imaging technologies. Finally, it is found that the latest deep learning (DL) classifiers are capable of detecting liver cancer accurately, fastly, and reliably. However, a major problem with existing relevant articles is that there is a lack of publicly available datasets for the detection of liver cancer and the drawback of these datasets is that almost all have few images. Hence, further research should be performed on large publicly available datasets to improve the complexity of computation for reliable

diagnosis of liver cancer. As a result, it serves mankind much better in efficiency and cost-effectiveness.

The chapter "Spot Filtering Adaptive Thresholding (SFAT) Method for Early Pigment Spot Detection on Iris Surface" states that iris pigment spot is a discrete pigmentation on the iris surface and can detect eye cancer. There are two types of iris spots, freckles and nevi. While freckles are usually harmless, nevi distort the stromal layer, and therefore, its existence is considered high potential for Uveal Melanoma, a type of cancer that can cause blindness. The features used to detect the Uveal Melanoma are size, shape, number of existences, spot of existence, and the color of the pigment spot on the iris surface. In image processing, feature extraction method typically extracts size, shape, and color. However, it is still challenging to produce an accurate extraction result for iris pigment spot. In this study, a threshold intensity value of color is identified as the pigment spot feature used in the feature extraction process.

The chapter "A Big Survey on Biometrics for Human Identification" emphases that biometric authentication systems developed in recent years are better than other traditional authentication methods such as passwords or signatures. All human biological traits are unique as biometrics such as fingerprints, palms, irises, palm blood vessels and fingerprint blood vessels, and other biometrics. Biometric identification systems basically have a complex structure that consists of different parts. Biometric-based authentication systems and authentication methods, along with other authentication systems, can improve the security aspects of authentication systems. Identification methods and tools are used in many important and essential applications such as surveillance processes, security investigations, fraud detection technologies, and access controls. Biometric-based identification methods in machine learning consist mainly of preprocessing, feature extraction, feature selection, classification, and finally evaluation. These systems can also be based on one biometric or based on several biometrics together.

The chapter "Computer-Aided Diagnosis of Pneumothorax Through X-Ray Images Using Deep Learning—A Review" emphases on automated diagnosis of pneumothorax in health surveillance is difficult for radiologists. Early detection of a pneumothorax is crucial for improving treatment outcomes and patient survival. In the medical field, the identification of pneumothorax through image processing is a tricky task. Recently, a rise of interest has been noticed in employing deep learning algorithms to aid pneumothorax detection. Nowadays, different medical Imaging tools are available to detect specific diseases. Chest radiographs are widely used to diagnose pneumothorax. Detection of pneumothorax at early stages can overcome the treatment difficulties. This chapter evaluates several innovative technologies and research that could help detect pneumothorax automatically. Artificial intelligence (AI) provides a significant result for automated pneumothorax (PTX) detection. Research has been done to see pneumothorax disease automatically through the chest radiograph. This article abstracts previous articles for detecting PTX from CXRs through machine and deep learning and also discusses different publicly available datasets. This study provides a detailed overview and discusses the existing literature's goodness and limitation.

In chapter "ML and DL Architectures Comparisons for the Classification of COVID-19 Using Chest X-ray Images," automated and AI-based prediction models for COVID- 19 are the main attraction for the scientist hoping to support some good medical decisions at this difficult time. However, mostly classical image processing methods have been implemented to detect COVID cases resultant in low accuracy. In this chapter, multiple naïve machine and deep learning architectures are implied to evaluate the performance of the models for the classification of COVID-19 using a dataset comprising of chest X-ray images of, i.e., COVID-19 patients and normal (non-infected) individuals. The analysis looks at three machine learning architectures including logistic regression, decision tree (DT) classifier, and support vector machine (SVM), and four deep learning architectures, namely convolutional neural networks (CNNs), VGG19, ResNet50, and AlexNet. The dataset has been divided into train, test, and validation set, and the same data have been used for the training, testing, and validation of all the architectures. The result analysis shows that AlexNet provides the best performance out of all the architectures. It can be seen that the AlexNet model achieved 98.05% accuracy (ACC), 97.40% recall, 98.03% F1 score, 98.68% precision, and 98.05% area under the curve (AUC) score.

The chapter "Data Mining in Medical Laboratory Service Improves Disease Surveillance and Quality Healthcare" examines origin, basic principles, advantages and disadvantages, uses and challenges of data mining in relation to Medical Laboratory Information Management System (MLIMS) while looking at data management from hard to soft copies, possible applications, ethico-legal perspectives, implications of data mining, disease surveillance, and data mining toward quality improvement as used in medical laboratories. It is evident that most of decisions taken in health care and public health are based on information provided by data mining from medical laboratory services based on the diseases of interest. Data mining in medical laboratory services is a tool that aids in monitoring trends in the diagnosis of cancer, HIV, COVID-19, malaria, diabetes, and other diseases based on various parameters of assessment with all demographic variables well documented and analyzed. The interested Agencies or Ministries may apply data mining techniques based on medical laboratory results to find trends in disease outbreaks or deaths, per hospital, state, region, or country through which policies could be formulated and implemented toward surveillance and quality healthcare improvement.

The chapter "Deep Learning-Based Lung Infection Detection Using Radiology Modalities and Comparisons on Benchmark Datasets in COVID-19 Pandemic" presents deep learning techniques for microscopic COVID-19 infection diagnosis, prevention and treatment on public datasets. Additionally, a general CAD architecture for COVID-19 detection is presented and each stage is discussed in detail. Based on radiology images analysis, several lungs treatment strategies for COVID-19 infected patients are suggested. Finally, evidence-based methodologies and modalities were explored in the analysis and findings, leading to a conclusion and possible future healthcare planning.

The editors are very thankful to all the members of Springer (India) Private Limited, especially Aninda Bose, for the given opportunity to edit this book.

Riyadh, Saudi Arabia — Dr. Tanzila Saba
Riyadh, Saudi Arabia — Dr. Amjad Rehman
Navi Mumbai, India — Dr. Sudipta Roy

Contents

Editors and Contributors

About the Editors

Prof. Tanzila Saba earned her Ph.D. in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She won the best student award in the Faculty of Computing UTM for 2012. Currently, she serves as Research Professor and Associate Chair of the Information Systems Department in the College of Computer and Information Sciences, Prince Sultan University, Riyadh, KSA. Her primary research focus in recent years is medical imaging, pattern recognition, data mining, MRI analysis, and soft computing. She led more than 15 research-funded projects. She has above two hundred research publications that have around 10,000 plus citations with h-index 63. Her most publications are in biomedical research published in ISI/SCIE indexed. Due to her excellent research achievement, she is included in Marquis Who's Who (S & T) 2012. She is Editor and Reviewer of reputed journals and on the panel of TPC of international conferences. She has full command of various subjects and taught several courses at the graduate and postgraduate levels. On the accreditation side, she is a skilled lady with ABET & NCAAA quality assurance. She is Senior Member of IEEE. Dr. Tanzila is Leader of Artificial Intelligence & Data Analytics Research Lab at PSU and Active Professional Member of ACM, AIS, and IAENG organizations. She is PSU WiDS (Women in Data Science) Ambassador at Stanford University.

Dr. Amjad Rehman is Senior Researcher in the Artificial Intelligence & Data Analytics Lab CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He received his Ph.D. & postdoc from Faculty of Computing Universiti Teknologi Malaysia with a specialization in Forensic Documents Analysis and Security with honor in 2010 and 2011, respectively. He received rector award for 2010 best student in the university. Currently, he is PI in several funded projects and also completed projects funded from MOHE, Malaysia, and Saudi Arabia. His keen interests are in data mining, health

informatics, and pattern recognition. He is Author of more than 200 ISI journal papers and conferences and is Senior Member of IEEE.

Dr. Sudipta Roy is working as Assistant Professor in Artificial Intelligence & Data Science department at JIO institute, Navi Mumbai. Prior to that, he was Postdoctoral Research Associate at Washington University in St. Louis, MO, USA. He has received his Ph.D. in Computer Science & Engineering from the Department of Computer Science and Engineering, University of Calcutta. He is Author of more than 50 publications in refereed international journals and conferences including IEEE, Springer, Elsevier, and many others. He has authored/edited four books and many chapters. He holds a US patent in medical image processing and filed an Indian patent in the smart agricultural system. He has served as Regular Reviewer of many international journals including IEEE, Springer, Elsevier, IET, and many others, and international conferences. He has served as International Advisory Committee Member and Program Committee Member of INDIAcom-2020, AICAE-2019, INDIACom-2019, CAAI 2018, ICAITA-2018, ICSESS-2018, INDIACom-2018, ISICO-2017, AICE-2017, and many more conferences. Currently, he is serving as Associate Editor of *IEEE Access* (IEEE), *International Journal of Computer Vision and Image Processing* (IJCVIP-IGI Global), and *Topic Editor of Journal of Imaging* (MDPI). In recognition of his exceptional contributions to the IEEE access journal as Associate Editor, the IEEE Access Editorial Board and Editorial Office honored him as Outstanding Associate Editor of 2020. He has more than five years of experience in teaching and research. His fields of research interests are biomedical image analysis, image processing, steganography, artificial intelligence, big data analysis, machine learning, and big data technologies.

Contributors

Ab Jabal Mohamad Faizal Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Johor Kampus Pasir Gudang, Masai, Johor, Malaysia

Abbas Naveed Digital Image Processing Lab, Department of Computer Science, Islamia College University Peshawar KPK, Peshawar, Pakistan

Afzal Erum National University of Sciences and Technology, Islamabad, Pakistan

Ahmadi Farnaz Department of Electrical Engineering, Tabriz University, Azerbaijan, Iran

Akbar Shahzad Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Ali Haris National University of Sciences and Technology, Islamabad, Pakistan

Aljurf Mahmoud King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

Almas Anjum Muhammad National University of Technology (NEUTCH), Islamabad, Pakistan

Alyami Jaber Department of Diagnostic Radiology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia;
Imaging Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia;
Smart Medical Imaging Research Group, King Abdulaziz University, Jeddah, Saudi Arabia

Amin Javeria Department of Computer Science, University of Wah, Wah Cantt, Pakistan

Arab Najafabadi Zakieh Department of Engineering, South of Tehran Branch, Islamic Azad University, Tehran, Iran

Asadi,Jouzani Somayeh School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Awan Mazhar Javed Department of Software Engineering, School of System and Technology, Lahore, Pakistan;
Department of Computer Sciences, Karakoram International University, Gilgit-Baltistan, Pakistan

Ayesha Hareem Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

Ayesha Noor School of Clinical Medicine, Zhengzhou University, Zhengzhou, Henan, PR China

Azizi Nastaran Sepahan Institute of Higher Education, Science and Technology, Isfahan, Iran

Bashir Muhammad Jawad School of Electrical Engineering and Computer Science, National University of Science and Technology (NUST), Islamabad, Pakistan

Ejinaka Obiora Reginald Department of Medical Laboratory Management, Federal School of Medical Laboratory Science, Jos, Nigeria

Elhassan Tusneem A. School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Etukudoh Nkereuwem Sunday Department of Medical Laboratory Management, Federal School of Medical Laboratory Science, Jos, Nigeria

Farooq Khan Niazi Muhammad Consultant Radiologist, Diagnostic Imaging Department King Salman North West Armed Forces Hospital (NWAFH), Tabuk, Saudi Arabia

Firdous Saba Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Gull Sahar Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Hamid Suhardi Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Kedah, Merbok, Kedah, Malaysia

Harouni Majid Department of Computer Engineering, Doalatabad Branch, Islamic Azad University, Isfahan, Iran

Hashim Siti Z. Mohd Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu, Malaysia

Hassan Syed Ale Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Hussain Muzammil School of Systems and Technology, University of Management and Technology, Lahore, Pakistan

Iqbal Sajid Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

Javed Kashif National University of Sciences and Technology, Islamabad, Pakistan

Javed Rabia School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor Bahru, CO, Malaysia;
Department of Computer Science, Lahore College for Women University, Lahore, CO, Pakistan

Jazi Elaheh Izadi Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Karim Adnan National University of Sciences and Technology, Islamabad, Pakistan

Karimi Mohsen Department of Bioelectrics and Biomedical Engineering, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Karimi Zahra Department of Bioelectrics and Biomedical Engineering, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Khalid Aima Riphah Collage of Computing, Riphah International University, Faisalabad, Pakistan

Khan Siraj Digital Image Processing Lab, Department of Computer Science, Islamia College University Peshawar KPK, Peshawar, Pakistan

Kolivand Mahyar Department of Medicine, University of Liverpool, Liverpool, UK;
School of Computing and Digital Technologies, Staffordshire University, Staffordshire, UK

Mahmoudi Helia MD, MPH Candidate in Family Medicine, Department of Community Medicine, Medical University of Isfahan, Isfahan, Iran

Malik Muhammad Awais Department of Software Engineering, School of System and Technology, Lahore, Pakistan

Muiz Fayyaz Abdul Department of Computer Science, University of Wah, Wah, Pakistan

Naaqvi Zunaira Riphah Collage of Computing, Riphah International University, Faisalabad, Pakistan

Najafabadi Somayeh Arab Department of Computer Engineering, Islamic Azad University, Najafabad Branch, Isfahan, Iran

Nasr Afrooz Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Nezhad Aliakbar Rabbani Department of Electrical Engineering, Isfahan University of Technology, Isfahan, Iran

Obeta Uchejeso Mark Department of Medical Laboratory Management, Federal School of Medical Laboratory Science, Jos, Nigeria

Othman Nur Zuraifah Syazrah School of Computing, Faculty of Engineering, University Technology Malaysia, Skudai, Johor, Malaysia

Qurat Ul Ain Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Rahim Mohd Shafry Mohd School of Computing, Faculty of Engineering, University Technology Malaysia, Skudai, Johor Bahru, Malaysia

Rehman Amjad Artificial Intelligence and Data Analytics Lab (AIDA), CCIS Prince Sultan University, Riyadh, Saudi Arabia

Rezakhani Sara Department of Electrical Engineering, Najaf Abad Branch, Islamic Azad University, Isfahan, Iran

Roy Sudipta Artificial Intelligence & Data Science, Jio Institute, Navi Mumbai, India

Saba Tanzila Artificial Intelligence and Data Analytics Lab CCIS, Prince Sultan University, Riyadh, Saudi Arabia;
College of Computing and Information Sciences, Prince Sultan University, Riyadh, CO, Saudi Arabia

Sahar Gul Department of Software Engineering, University of Management and Technology, Lahore, Pakistan

Sajjad Muhammad Digital Image Processing Lab, Department of Computer Science, Islamia College University Peshawar KPK, Peshawar, Pakistan

Saleem Muhammad Rehan Department of Computer Science, School of System and Technology, Lahore, Pakistan

Sharif Muhammad Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan

Swee Tan Tian Bioinspired Device and Tissue Engineering Research Group, Faculty of Engineering, School of Biomedical Engineering and Health Sciences, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Tavakoli Nakisa Department of Computer Engineering, Farsan Branch, Islamic Azad University, Farsan, Iran

Tehseen Ahmad Khawaja Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

Tobeiha Atefeh Department of Electrical Engineering, Najaf Abad Branch, Islamic Azad University, Isfahan, Iran

ul haq Ihtishaam Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan

Urooj Farwa Riphah College of Computing, Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

Segmentation of White Blood Cells in Acute Myeloid Leukemia Microscopic Images: A Review

Tusneem A. Elhassan, Mohd Shafry Mohd Rahim, Tan Tian Swee, Siti Z. Mohd Hashim, and Mahmoud Aljurf

Abstract Acute Myeloid Leukemia (AML) is a fast-growing leukemia caused by the rapid proliferation of immature myeloid cells. AML is a life-threatening disease if left untreated. Therefore, early detection of AML is crucial, maximizes the cure opportunities, and saves patients' lives. Initial AML diagnosis is done by expert pathologists where blood smear images are utilized to detect abnormalities in WBCs. However, manual detection of AML is subjective and prone to errors. On the contrary, computer-aided diagnosis (CAD) systems can be an accurate diagnostic tool for AML and assist pathologists during the diagnosis process. Segmentation of White Blood Cells is the first step toward developing an accurate CAD system for AML. To date, WBC segmentation has several challenges due to several reasons such as different staining conditions, complex nature of microscopic blood images, and morphological diversity of WBCs. Current WBC segmentation techniques vary from conventional image processing methods to advanced machine learning and deep learning methods. This chapter discusses current segmentation methods as well as the potential solutions for improving automated WBC segmentation accuracy.

T. A. Elhassan · M. S. M. Rahim (✉)
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
e-mail: shafry@utm.my

T. A. Elhassan
e-mail: amtusneem@graduate.utm.my

T. T. Swee
Bioinspired Device and Tissue Engineering Research Group, Faculty of Engineering, School of Biomedical Engineering and Health Sciences, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia
e-mail: tantswee@biomedical.utm.my

S. Z. M. Hashim
Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu, Malaysia
e-mail: sitizaiton@umk.edu.my

M. Aljurf
King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia
e-mail: maljurf@kfshrc.edu.sa

T. Saba et al. (eds.), *Prognostic Models in Healthcare: AI and Statistical Approaches*, Studies in Big Data 109, https://doi.org/10.1007/978-981-19-2057-8_1

Keywords While blood cells · Segmentation · Acute myeloid leukemia · Computer-aided diagnosis (CAD)

1 Introduction

Segmentation is the process of partitioning the image space into non-overlapping regions where each region can be characterized by unique features. White blood cell (WBC) segmentation is the process of dividing the blood image into two regions, namely WBCs and other blood components. WBC segmentation is a curtail step in the automatic diagnosis of leukemia. More precisely, it is a prerequisite step for other image processing tasks, including feature extraction, WBC classification, and WBC counting. Segmentation is further used for nucleus and cytoplasm segmentation which has an important role in situations where the nucleus: cytoplasm ratio is a key factor in diagnosis process [1]. To date, WBC segmentation has been performed manually by pathologists or using quantitative methods. However, advanced qualitative methods utilizing advance image processing and pattern recognition is not yet implemented [2]. This makes the process subject to human error and produces inaccurate results. The process is tedious, time-consuming, and subject to inter- and intra-class variation among pathologists. Only 76.6% of cases showed agreement between pathologists during the leukemia diagnosis [3]. Therefore, accurate segmentation techniques for WBCs and their regions using computer-aided systems are needed. These systems can implement advanced image processing and pattern recognition methods, such as deep learning, to extract complex features of WBC and its regions with minimal human intervention. Several methods have been proposed to detect and segment the WBCs and their nuclei and cytoplasm. Nevertheless, automated WBC segmentation is a difficult task and encompasses several challenges due to the noisy nature of microscopic images. Irregular boundaries and textural similarities between WBCs and other blood components are some of the challenges that make it difficult to differentiate between WBCs and other blood components [4–6]. Moreover, WBCs have a complex structure in term of shapes, textures, and colors [7–9]. Section 3 summarizes different segmentation methods implemented in WBC segmentation including its nucleus and cytoplasm.

This chapter is organized as follows: Section 2 presents an overview of AML; Section 3 discusses some of the challenges facing automated segmentation methods; Section 4 introduces the publicly available datasets on AML WBC segmentation; Section 5 describes the different types of automated WBC segmentation methods presented in the literature. Section 6 discusses WBC limitations and future solutions. Finally, the topic is summarized in Sect. 7.

2 Acute Myeloid Leukemia

2.1 An Overview

Leukemia is a blood cancer that originates from cells that normally develop into different types of blood cells. Usually, it develops from immature white blood cells called leukocytes, but some types of leukemia can develop from other blood cells. Leukemia can be divided into two types: acute and chronic. Acute leukemia is a fast-growing cancer that is fatal if left untreated, while chronic leukemia is a slow-growing cancer that patients can live with for a long time.

Acute leukemia can be divided into ALL, which originates from lymphocytes, and AML, which develops from myelocytes. In both cases, acute leukemia starts in the bone marrow, where leukemic cells (i.e., blasts) proliferate and replace normal blood cells. They can also spread to other organs in the body. Early detection of acute leukemia is crucial to ensure that appropriate treatment modalities are provided to save patients' lives [10].

2.2 AML Subtypes

AML has several subtypes, and it is important to classify them to understand patients' prognosis and the appropriate treatment modality (e.g., Acute Promyelocytic Leukemia (APL) is treated by a different type of drug than other subtypes). Classification is based on the level of cell maturation at the time of diagnosis. Two systems have been used for classification: the French–American–British (FAB) classification system and the World Health Organization (WHO) classification system.

FAB System
The FAB classification system was developed by seven hematologists who formed an international cooperative group. It is based on the morphological features of leukemic cells identified under a microscope after staining and the type of cells from which leukemia develops (Table 1).

Subtypes M0–M5 all start in immature forms of white blood cells. However, M6 starts in very immature forms of red blood cells, while M7 starts in immature forms of cells that make platelets.

WHO Classification
The WHO system considers various factors that can affect AML prognosis, mainly based on cytogenetics. The following are WHO AML classes based on genetics abnormities [11, 12]:

- AML with a translocation between chromosomes 8 and 21 [t (8:21)].
- AML with a translocation or inversion in chromosome 16 [t (16:16) or inv (16)].

Table 1 AML FAB classification system

FAB subtype	Name
M0	Undifferentiated acute myeloblastic leukemia
M1	Acute myeloblastic leukemia with minimal maturation
M2	Acute myeloblastic leukemia with maturation
M3	Acute promyelocytic leukemia
M4	Acute myelomonocytic leukemia
M4 eos	Acute myelomonocytic leukemia with eosinophilia
M5	Acute monocytic leukemia
M6	Acute erythroid leukemia
M7	Acute megakaryoblastic leukemia

- APL with the PML-RARA fusion gene.
- AML with a translocation between chromosomes 9 and 11 [t (9:11)].
- AML with a translocation between chromosomes 6 and 9 [t (6:9)].
- AML with a translocation or inversion in chromosome 3 [t (3:3) or inv (3)].
- AML (megakaryoblast) with translocation between chromosomes 1 and 22 [t (1:22)].
- AML with the BCR-ABL1 (BCR-ABL) fusion gene.
- AML with mutated NPM1 gene.
- AML with biallelic mutations of the CEBPA gene (that is, mutations in both copies of the gene).
- AML with a mutated RUNX1 gene.

2.3 Diagnosis of AML

Leukemia diagnosis starts by looking at abnormal white cells, red cells, and platelets counts in a blood sample using a CBC test. If there is evidence of abnormal cells in blood samples, a blood smear is furtherly examined to determine the percentage of blast cells. However, a percentage of more than 20% is considered as evidence of leukemia. Differentiation-based classification methods involving flow cytometry are used to determine the subtype of myeloid leukemia where different patterns of antigen acquisition exist (Fig. 1).

2.4 Morphology of Acute Leukemia

Morphological differences of WBC in acute leukemia are based on two factors: the cell line and the level of cell differentiation (i.e., maturation) [13].

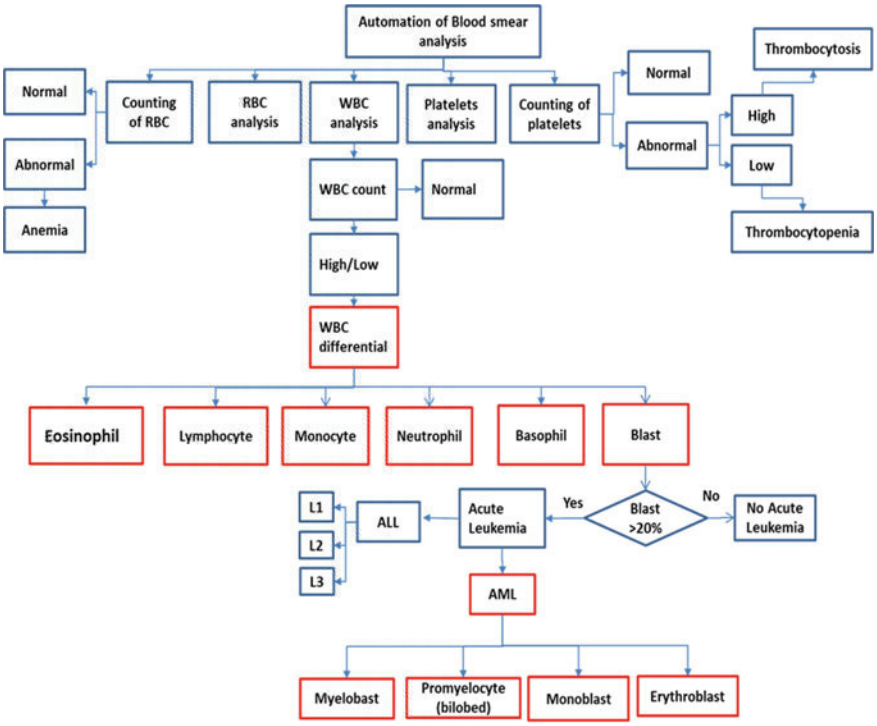


Fig. 1 Diagnosis of AML based on blood smear analysis

Based on FAB system, acute leukemia is classified into myeloid (AML) and lymphoblastic (ALL) which can only be utilized for untreated patients. This is because treatments such as chemotherapy can change the nature of normal and leukemic cells [14]. ALL is classified into three different subtypes (L1, L2, and L3), while AML is divided into eight subtypes (M0, M1, M2, M3, M4, M5, M6, and M7).

Types of WBCs Presented in AML

There are several types of cells in AML blood smears. These cells are divided into immature and mature cells:

Mature Cells

- Lymphocyte: These are rounded cells, containing a single, large round nucleus and a clear cytoplasm. It is classified into two types: *B* and *T* cells.
- Neutrophil: This type of cells has an irregularly shaped nucleus that contains multiple lobes and dotted granular cytoplasm. Neutrophil can be further classified into two types:

(a) Neutrophil (banded): The banded neuropil cell is derived from metamyelocytes. It called "banded" because all the nuclear sections of the nucleus are the same width (band).
(b) Neutrophil (segmented): segmented neutrophils represent the final stage in the lineage that starts with myeloblasts, forming gradually, without any clear transition or further cell divisions, by increasing the contraction of their nuclei. The nuclear segments are connected only by narrow chromatin bridges, which should be no thicker than one-third of the average diameter of the nucleus. The chromatin in each segment forms coarse bands or patches and is denser than the chromatin in band neutrophils. The cytoplasm of segmented neutrophilic granulocytes varies after staining from nearly colorless to soft pink or violet. The abundant granules are often barely visible dots.

Immature Cells

Blasts

- Erythroblasts: Nucleated cells derived from red marrow or red blood cells.
- Monoblasts: Progenitor cells that are differentiated from myeloid stem cell. They are presented in bone marrow rather than in normal peripheral blood. Later, they mature into monocytes.
- Promyelocytes (bilobed): these cells are characterized by the presence of bilobed nuclei that have been likened to butterfly wings. Usually, the cells present abundant cytoplasmic granules; numerous Auer rods and some ermed faggot cells. However, some cases do not have obvious cytoplasmic granules, may have basophilic cytoplasm, and have folded bilobed nuclei that may be mistaken for monocytes.
- Myeloblasts: Myeloblasts are the least mature cells in the granulocyte lineage. Mononuclear, round-to-ovoid cells may be distinguished from proerythroblasts by the finer, "grainy" reticular structure of their nuclei and the faintly basophilic cytoplasm. On first impression, they may look like large or even small lymphocytes (micro myeloblasts), but the delicate structure of their nuclei always distinguishes from micro myeloblasts. The cytoplasm contains azurophilic granules.

Non-blasts

- Promyelocytes: Promyelocytes are the product of myeloblast division and usually grow larger than their progenitor cells. During maturation, their nuclei show an increasingly coarse chromatin structure. The nucleus is eccentric; the lighter zone over its bay-like indentation corresponds to the Golgi apparatus. The wide layer of basophilic cytoplasm contains copious large azurophilic granules containing peroxidases, hydrolases, and other enzymes. These granulations also exist scattered all around the nucleus, as may be seen by focusing on different planes of the preparation using the micrometer adjustment on the microscope.

Non-blasts (Semi-mature Cells)

- Metamyelocytes: Metamyelocytes (young granulocytes) are derived from the final myelocyte. They show further maturation of the nucleus with an increasing number of stripes and points of density that give the nuclei a spotted appearance. The nuclei slowly take on a kidney bean shape and have some plasticity.
- Myelocytes: Myelocytes are the direct product of promyelocyte mitosis and are always clearly smaller than their progenitors. They are characterized with an ovoid nucleus with a banded structure; the cytoplasm is becoming lighter with maturation and sometimes acquiring a pink tinge. A special type of granules, which no longer stain red like the granules in promyelocytes ("specific granules," peroxidase-negative), are evenly distributed in the cytoplasm. Myelocyte morphology is wide-ranging because myelocytes cover three different varieties of dividing cells [15].

AML blood smear samples include normal and abnormal WBCs. Figure 2 illustrates different types of WBCs presented in the blood smear sample. Table 2 presents the five normal types of WBCs and their relative percentages in the blood.

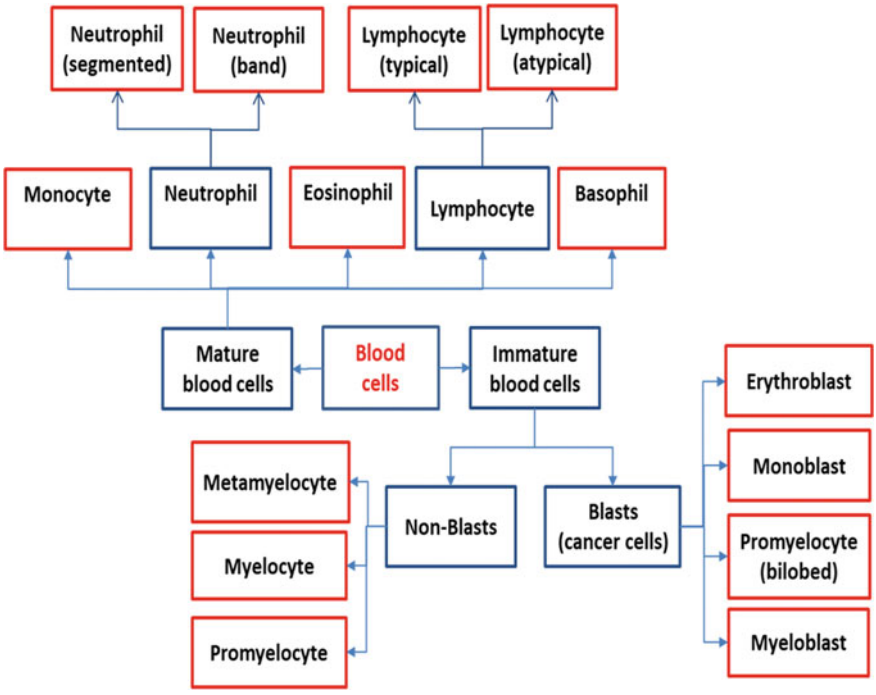
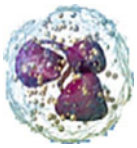
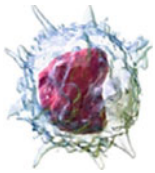
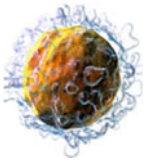
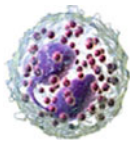
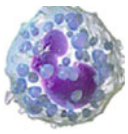


Fig. 2 Classification of WBCs presented in the AML blood smear samples

Table 2 Different types of white blood cells

Cell image	Name of the cell	Percentage in the blood (%)
	Neutrophils	50–70
	Monocytes	3–8
	Lymphocytes	24–45
	Eosinophils	2–4
	Basophils	0.5–1

3 Challenges of WBC Segmentation

WBC segmentation faces several challenges, including differences in the staining processes, illumination variation, complex WBC structure, and the similarities in color and texture between different components of the blood image [7–9].

- Staining variation: Different staining techniques have various effects on the degree of WBC staining. Some parts show different degrees of staining compared to other parts. For example, the nucleus is presented with darker colors compared to the cytoplasm. Moreover, some parts of the nucleus show darker colors compared to other nucleus parts. Different degrees of staining make it difficult to differentiate between the boundaries of the nucleus, cytoplasm, and WBC [6].
- Illumination variation: Illumination methods used by different types of microscopes vary, and they result in different color distributions for the nucleus and

cytoplasm. Moreover, the utilization of new technologies such multicolor light-emitting diodes (LEDs), which allow for multiple color illumination sources, make it difficult to discern between these WBC and other blood components [3, 16]. Imbalanced illumination affects the contrast between cell boundaries, and the background exhibits difficulties in differentiating between nucleus and cytoplasm boundaries [17].

- Complex structure of WBC: WBCs comprise diverse morphological variations in terms of shapes, colors, textures, sizes, and nucleus to cytoplasm ratios. Furthermore, during different WBC maturation stages, a single type of WBC exhibits variation in terms of shapes, sizes, and cytoplasm characteristics, which makes it difficult to differentiate between various WBC components and between WBC and other blood components, such RBCs.
- Complex microscopic image background: The blood smear image background is relatively complex due to the presence of overlapping objects such as WBCs and RBCs. This problem makes it difficult for segmentation algorithms to achieve accurate results. Moreover, these algorithms must apply WBC segmentation in two separate steps. The first is to differentiate between WBCs and RBCs, and the second is to apply the subsequent segmentation algorithm [18].
- Texture and color similarities: Some types of WBCs, such as neutrophils and monoblasts, exhibit similarities in their cytoplasm and image background color and texture; this makes it difficult to differentiate between these regions [19].

4 Datasets

Most of datasets available on AML and utilized by several research papers are private datasets and were used for the purpose of AML classification. The only publicly available dataset for AML classification of WBCs is the (AML_Cytomorphology_LMU) dataset, which was published by Matek et al. in 2019 [20]. However, this dataset is annotated for WBC classification and not for WBC segmentation. The dataset is a single-cell morphological dataset of 18,365 WBCs obtained from 100 patients with AML and 100 non-malignant controls at Munich University Hospital between 2014 and 2017. The dataset consisted of 15 different types of single-cell images labeled by expert pathologists. Four of these were leukemic cells, and the other 11 were normal blood cells. Among the 11 types, seven were mature leukocytes and four were immature. Cancerous and noncancerous WBCs were classified by expert pathologists based on standard morphological classifications. To our knowledge, there are no datasets on AML annotated for segmentation purposes.