Eximia
journal

Eximia Journal

(ISSN 2784-0735)

Vol. 5

2022

Eximia Journal
Vol. 5, 113-145, July, 2022
ISSN: 2784-0735
www.eximiajournal.com

Multiple Myeloma. Detection from Histological Images Using Deep Learning

Reem Al mulhim, Asrar. A. Haque

Department of Computer Science, King Faisal University, Saudi Arabia, Al-Hasa

rooro108@hotmail.com

Abstract. Multiple myeloma (MM)is a blood cancer that develops in the bone marrow. which is the soft spongy tissue found in the center of bones. It is the second-most common form of blood cancer in the KSA, comprising about 1% of all cancers. It necessitates an early diagnosis to lower the mortality risk. The conventional diagnostic tools are resource intensive. As a result, these solutions are difficult to scale in order to reach a wider audience. There are many ways to diagnose multiple myeloma (MM) in bone marrow such as complete blood count test (CBC) or counting myeloma plasma cells in aspirate slide images. However, using these manual detection techniques are very time-consuming and the result will depend upon the experience of the pathologist. So, computer-aided techniques are used for fast processing and accuracy. That's beneficial to consultant and especially to pathologist who plays the role of an assistant in the diagnosis of multiple myeloma by a state of art method developed with deep learning. The Deep learning model uses the Convolutional Neural Network (CNN) that automatically extracts the features and classifies the image using the fully connected network. In this study, we proposed an ensemble deep learning-based approach for automatic binary classification of myeloma histology images, which can improve the early diagnosis of multiple myeloma patients. Also, we explored the capability of pre-trained VGG-16, VGG19, and Resnet-50 models with transfer learning for Multiple Myeloma image categorization and obtained a prediction accuracy of 99.77%, 97.92%, and 97.40%, respectively, and explained reasons that helped the accuracy increase.

Keywords. Multiple Myeloma (MM), convolutional neural network (CNN), deep learning

[1] Introduction

Multiple myeloma is a cancer of plasma cells. It is a malignant disease characterized by plasmacytosis, paraprotein production, and many symptoms. There are many risk factors, including obesity, radiation exposure, family history, and certain chemicals. In multiple myeloma (MM), cancerous plasma cells collect in the bone marrow and cause problems with healthy blood cells. Instead of the reproduction of helpful antibodies, cancer cells produce abnormal proteins that can cause complications. It is the second most diagnosed haematological malignancy with an annual incidence and prevalence in the United States, with an increasing rate of about 15,000 to about 45,000 [1]. According to the Saudi Cancer Registry report from 2015, MM represented 1% of all cancers and about 5% of hematologic malignancies [2].

In Saudi Arabia, the exact incidence of MM is largely underreported. According to the WHO country profile, MM and lymphomas accounted for 9.6%–11% of cancer-related deaths in the Kingdom in 2014. In King Faisal Specialist Hospital and Research Centre (KFSHRC) in Riyadh, they indicate the low proportion of MM in the total hematologic malignancies pool in the Saudi population is probably multifactorial. First, the majority (72%) of the Saudi population is between 15 and 65 years of age. According to the 2014 Saudi cancer registry, MM accounts for 1% of total cancer cases, with an age-standardized incidence of 1% in males and 0.7% in females [3]. It has been observed that Saudi patients with MM are relatively younger and present with poorer performance status and higher international staging system stage disease than western patients with MM [4]. Although the 2014 Saudi Arabian registry shows that the male ratios are higher than those of females with MM in Saudi Arabia, it was suggested that this does not reflect the real pattern because a considerable proportion of cases are underreported, as shown in Figure 1.

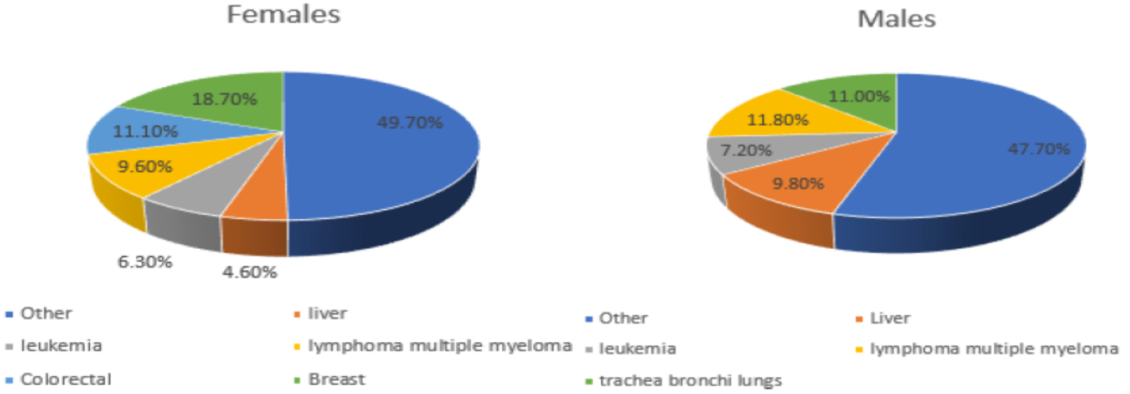


Figure 1 Statistics of Cancer Mortality in Saudi Arabia [5].

Mitotic features (arrows) are seen among plasma cells, as shown in Figure 2. The most crucial factor is that lytic bone disease is a hallmark of MM, being present in about 80% of patients with newly diagnosed MM [6]. The current work explores the effectiveness of deep learning-based algorithms such as the VGG-16, VGG-19, and ResNet-50 models for the classification of multiple myeloma.

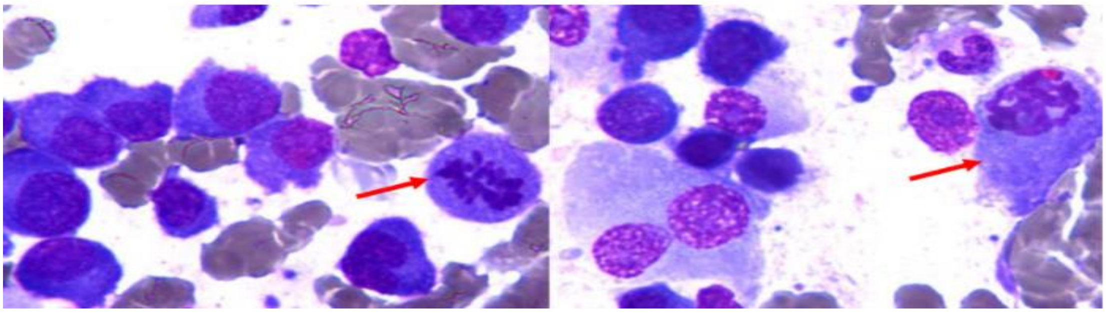


Figure 2 Bone marrow smears from a newly diagnosed multiple myeloma patient [7].

The rest of the report is arranged as follows. In section 2, background and motivation are discussed, followed by the problem statement in section 3. Section 4 Scope of the project, in Section 5 related work; Section 6 novelty of the project; Section 7 project requirements Section 8 expected outcomes; Section 9 solution identification and justification; Section 10 project solution design; Section 11 implementation models; Section 12 Tools and Techniques Used During Implementation; Section 13 Discussion and Future Work.

[2] Background and motivation

The malignant condition known as multiple myeloma (MM) affects plasma cells. This type of disease leads to eventual mortality, hypercalcemia, and many other symptoms. The risk of infection is highest in the first 3 months after diagnosis and decreases with response to treatment [49]. Family physicians play a role in evaluating these patients for infection [14]. Workup of plasma cell disorders involves establishing the presence of a monoclonal paraprotein and radiological investigations [59].

Early detection and early intervention in high-risk soldering myeloma may not only help to delay disease advancement into active disease but also increase the cure rates. Histology analysis is usually performed by pathologist review, but this procedure is time-consuming and error prone. Thus, the primary goal of this project is to contribute to the search for the best and fastest ways to detect multiple myeloma to preserve the health of the patient and help him with early diagnosis. This is done using deep learning, which has proven effective in discovering many diseases and making sure of their presence with high accuracy [34].

The project has been proposed to provide a healthy life opportunity for patients through early detection so that they can receive appropriate treatment, as well as to allow the creation of more accurate and clear programs in the future to determine the stages of this type of disease and conduct more research around it using the developed algorithms. It was also recommended that medical detection be developed by combining it with artificial intelligence. Recently, work has been undertaken to develop computer-assisted diagnostic tools for cancer diagnosis via image processing. Computer-assisted tools need capturing of images, stain color normalization of images, segmentation of cells of interest, and analysis to count malignant versus healthy cells. In the beginning, blood samples will be aspirated from bone marrow by using the injection of the needle onto the bone. The derived blood sample will be conveyed to a slide and stained using haematoxylin and eosin. [19]

This project is required for a variety of reasons, including developing the traditional detection process and updating it using deep learning tools. It is also necessary in the early detection process of patients. When early detection is made, we maintain the patient's survival if possible and find the appropriate treatment for him. To accurately know the type of disease, you need an experienced consultant or specialist to verify its presence. This program helps doctors determine whether the disease is present. It also saves a lot of time and effort to verify it. This program is done by using the VGG16, VGG-19, and ResNet-50 models to predict by training them on images of normal and abnormal cells. Recently, intelligent systems that offer artificial intelligence capabilities often rely on machine learning. Machine learning is the ability of systems to learn from training data that is relevant to a given problem in order to automate the construction of analytical models and complete related activities. A machine learning concept called "deep learning" is based on artificial neural networks. Deep learning models outperform shallow machine learning models and conventional data analysis techniques for many applications. Figure 3 shows the relationship between artificial intelligence (AI), machine learning (ML), and deep learning (DL).

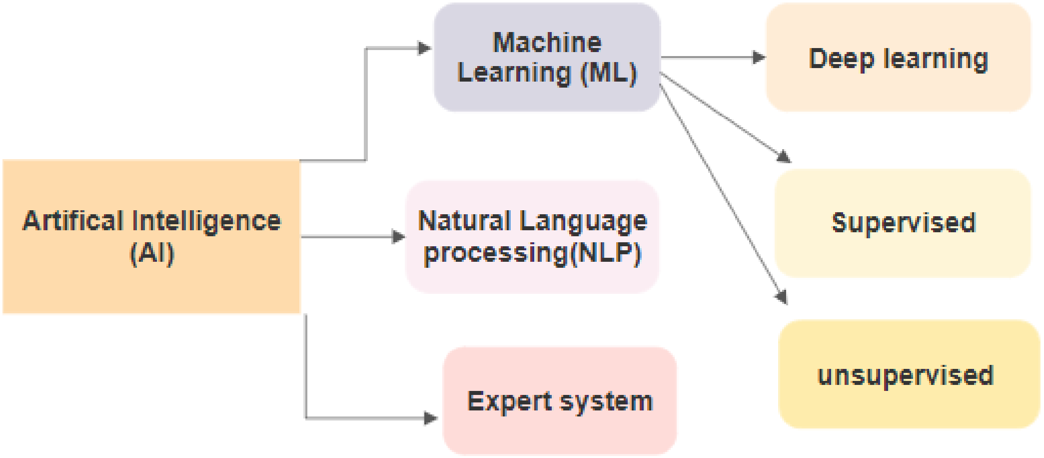


Figure 3 Relationship between artificial intelligence (AI), machine learning (ML) and deep learning (DL)[57].

Machine learning describes a set of techniques that are commonly used to solve a variety of real-world problems with the help of computer systems that can learn to solve a problem instead of being explicitly programmed. As shown in Figure 4, there are three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the user teaches the software to produce a solution based on a set of labeled and known data For supervised learning applications, classification and regression methods, such as random forests, decision trees, and support vector machines, are frequently utilized. In unsupervised machine learning, the algorithms generate answers from unknown and unlabeled data. Unsupervised methods are frequently used by data scientists to find patterns in fresh data sets. Unsupervised machine learning frequently makes use of clustering algorithms like K-means. Reinforcement learning uses a system of reward and punishment to train its algorithms. In this learning, the model will receive rewards for performing correctly and penalties for doing incorrectly, and so it will learn to maximize rewards while minimizing penalties.

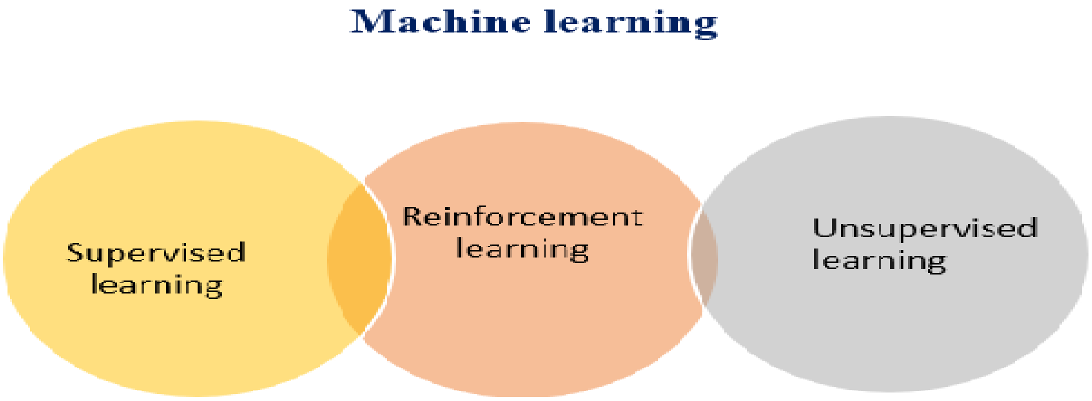


Figure 4 Types of Machine Learning Algorithms [8].

Machine learning and artificial intelligence have the potential add much value to the pharma and healthcare industry state by facilitating informed decision-making, reducing the time to complete the trial, and the overall drug development process. Sixty percent of respondents from the pharma and healthcare industries state that deployment of artificial intelligence helps improve quality control. According to 42 percent of respondents [56], monitoring and diagnosis is another important use case for AI. AI technology helps diagnose diseases and its algorithms can select treatments accordingly. Additionally, it might soon be feasible for this sector to apply AI to provide patients with individualized preventive risk screenings. This opens the possibility of finding the most suitable options for patients based on AI-enabled technology. As is shown in Figure 5.

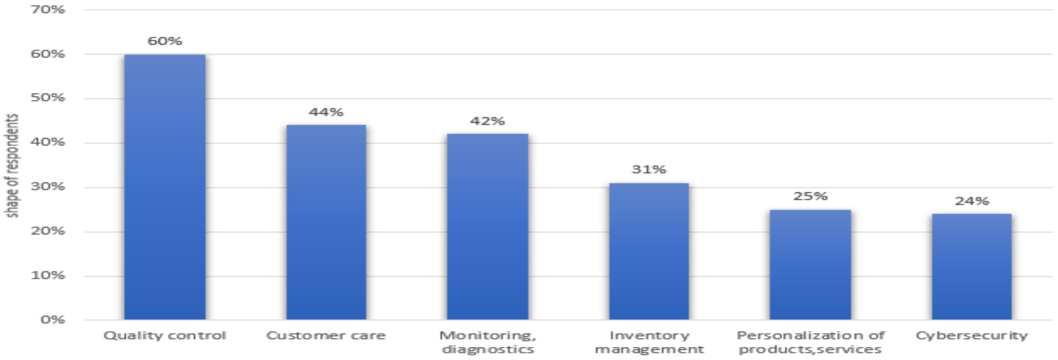


Figure 5 Global AI use cases for pharma and healthcare 2020 [56].

Deep learning is a set of machine learning methods that are inspired by the processing of information and distributed communication in a network of biological neurons. Artificial Neural Networks should be trained in deep learning. ANNs are networks of artificial neurons. Each ANN has a minimum of three layers; one is an input layer that will take the input. A hidden layer that trains on the dataset fed to the input layer, the output layer, gives output depending on the input that is fed. Deep learning has been gaining a lot more interest in medicine than any other machine learning technique. Convolutional networks, which are a type of artificial neural network, are extensively used for image-based applications and are proved to have achieved better results than humans in determining and classifying objects. The different deep learning architectures applied in healthcare are: Neural Network, CNN (Convolutional Neural Network), and RNN (Recurrent Neural Network). A neural network is a series of algorithms that recognize underlying relationships in a set of data through a process that mimics the human brain. Neural networks, also known as artificial neural networks, are a subset of machine learning and are at the heart of deep learning algorithms. A synthetic deep neural network is the CNN (Convolutional Neural Network). CNN is used for image classification, segmentation, and recognition. A recurrent neural network (RNN) is a special type of artificial neural network adapted to work with time series data or data that involves sequences.

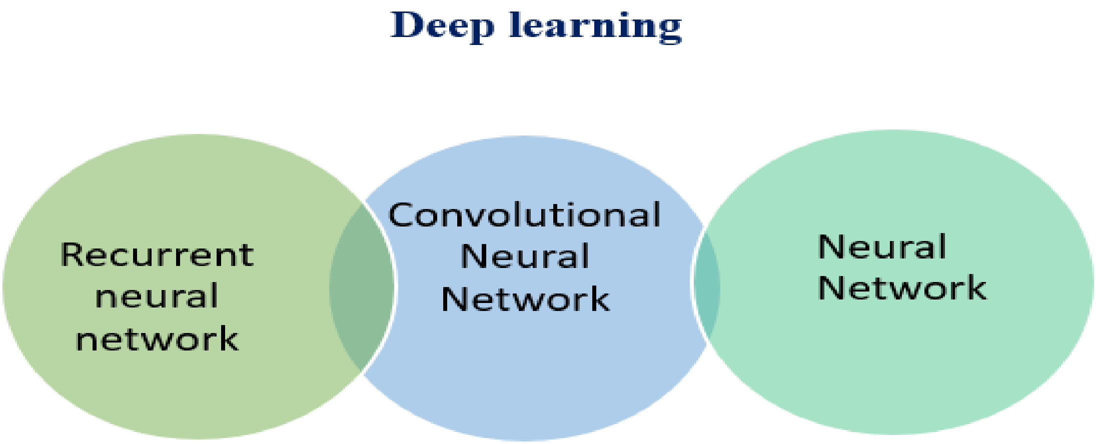


Figure 6 The different type of deep learning [9].

Deep learning is one of the major contributions to artificial intelligence, which is frequently used in various disciplines. The rapid advancement of machine learning, especially deep learning, continues to fuel the medical imaging community's interest in applying these techniques to improve the accuracy of cancer screening [8]. The application of deep learning-based artificial intelligence methods in different domains increases the reliability of automated systems. The efficiency of deep learning methods makes them suitable to be presented in different domains where high accuracy is required and computational time is restricted. Biomedical image analysis is one of the important domains where the processing time is stipulated and errors in the diagnostic outcome can be very costly and dangerous. Artificial intelligence techniques that are supported by deep learning are particularly effective in this. In recent years, biomedical image analysis has become a prominent field where deep learning techniques are frequently applied as is shown in Figure 6 [9].

A convolutional neural network is employed in order to detect multiple myeloma while using microscopic images of blood, a type of machine-learning algorithm that is gaining attention as AI has a significant role to play in the field of medical imaging. CNN are a class of deep learning techniques that are considered applicable to image analyses because they recognize complex visual patterns in a manner like the processes of human perception [10]. Convolution Neural Networks (CNN) are majorly utilized for analysing visual representations. CNN are image classification algorithms. They work fast and efficiently for image classification. When we compare this algorithm with other image classification algorithms, we notice they use a little pre-processing. A CNN model is comprised of an input layer, an output layer, and multiple hidden layers. Early detection of the disease allows for quicker action and saves precious time. For that reason, we are going to detection outcomes as soon as possible whether our cells are normal or abnormal. The proposed model takes an image as input extract feature then classifier as shown in Figure 7.

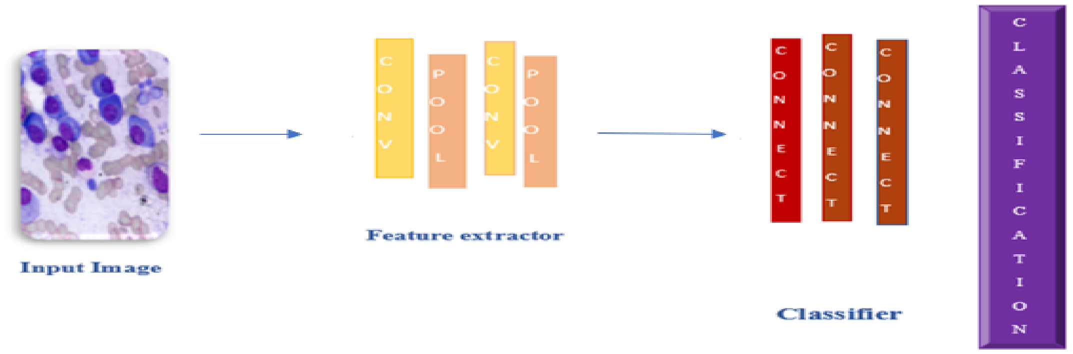


Figure 7 Proposed CNN model architecture [20].

[3] Problem statement

Manual detection of cells, whether they are normal or abnormal, can take a very long time and may be subject to human errors. Therefore, relying on advanced technologies to perform this process has become an urgent necessity due to the massive proliferation of these systems. However, there are usually concerns about the accuracy of these systems and their ability to detect cells in different conditions. In this project, we mainly want to answer the following questions:

1. Is it possible to detect normal and abnormal cells with high accuracy through advanced algorithms like VGG-16, VGG-19, and Resnet50?
2. What are the reasons that help raise the accuracy of advanced algorithms compared to previous experiences?
3. How do we take advantage of the massive development of deep learning to detect difficult diseases like multiple myeloma?
4. What are the differences and advantages between using manual diagnosis and diagnosis through modern algorithms?
5. What is the best performing algorithm among VGG16, VGG19, and Resnet50?

[4] Scope of the project

Early diagnosis of multiple myeloma is critical for the improvement of patient survival. The diagnosis will depend on the size of the sample obtained [15]. Tissue analysis is generally performed by a pathologist on a sample, but this execution is time-consuming and error prone. We will improve that method by using deep learning to detect multiple myeloma, which includes a convolution neural network that has proven effective in detecting complex and rare diseases. Classification is applied at the image patch level using a convolutional neural network (CNN) [13].

We classify multiple myeloma, which is known for being difficult to identify and predict normal or abnormal cells. We'll detect it using deep learning algorithms, which have been shown to be effective in detecting cancers and are among the best in high accuracy and precise prediction. VGG16, VGG19, and Resnet50.Specific models were chosen due to the advantages available in them. As one of the biggest advantages of the Resnet is that, while increasing network depth, it avoids negative outcomes. So, we can increase the depth, but we also have fast training and higher accuracy. The form that can operate with 50 neural network layers is referred known as Resnet50. We also chose the VGG16 model because it is a kind of Deep Convolutional Networks for Large-Scale Image Recognition. One of the most widely used pre-trained models for image categorization is VGG-16.

The goal of our project is to detect myeloma with high accuracy using the VGG-16 algorithm that was used before, but the results for this model were less than 95% and there was little data for this model. We will use the same model and the same data, but the result will be more accurate than previous studies by several factors, which in turn help to significantly raise the accuracy. One such factor is the number of images used, as we know that the more data used, the higher the accuracy, and this is what we will do. First, the model will classify patients' data. normal images in Kaggle- Bone Marrow Cell Classification [53], and abnormal images are found in the SegPC-2021 Challenge [33]. This data was processed and generated only abnormal cells into 3196 images of cells and 1598 images of normal cells. It is already available.
Also, among the factors that help in increasing the accuracy is the increase in the number of layers. When the number of layers of the model increases, the accuracy also increases. When the number of layers of VGG16 increases, the VGG19 model is formed here, as this model is better than VGG16. We will also apply the ResNet-50, it is distinguished by the network has been pre-trained to categorize images into 1000 different item categories.

[5] Related work

Imaging examinations play an important role in the early diagnosis of myeloma, which is very important because research has shown that the incidence of multiple myeloma has significantly exceeded cases of acute leukemia and ranks second among blood diseases. Also, Deep learning algorithms only need the data in the proper format and some network parameters appropriate to the problem [27].

Many studies have varied in detecting and identifying myeloma cells through many algorithms. One of the most prominent recent studies looking at the detection of multiple myeloma was to classify the images as containing cancerous cells or not. The number of healthy images was 755, and 755 images of the affected tissue. They used convolutional neural network models, namely VGG (Visual Geometry Group), VGG-16, VGG-19, MobileNetV1, and MobileNetV2. The last results they achieved gave an accuracy of 94.00%, 95.50%, 99.90%, and 97%, respectively [43].

Other studies about cancer blood, including leukemia, which is also considered a serious and malignant disease, have been conducted. They classified acute lymphoblastic leukemia into healthy and malignant cells, where the number of healthy cells was 3389 and the number of malignant cells 7272, and they used a ResNeXt convolutional neural network with Squeeze-and-Excitation units. They achieved an F1-weighted score of 88.91% in the test group [44].

There have been some recent studies on this disease, but the data is inconsistent, as the data used in the research was 5938 patch images aimed at detecting bone lesions in the tomography data for patients with multiple myeloma to classify patches into lesion and non-lesion. Using the VGG-16 algorithm, an accuracy of 0.92% was reached [45].

In this related work, we summarize data related to this disease using different data types such as x-rays and histological images. Also, many different models were used with varying degrees of accuracy, resulting in the Table 1 consisting of the main ideas are: the most important previous studies; the kind of algorithms they have been used; the kind of data which could be images or registered MM patients; findings about the accuracy they have reached; and limitations about the most difficult studies previously conducted.

Table 1 The summary of related work

Main idea	**Kind of algorithm**	**Kind of Dataset**	**Finding**	**Limitation**
Focusing on the segmentation effects of deep learning-based models on CT images for myeloma and the influence chemotherapy treatments on the prognosis of patients.	The U-Net model (RCNN) model.	Segment the CT images.	*Accuracy 99%.*	Do not look at the relationship between different prognostic indicators and patient survival [18].
Effectiveness of deep learning-based object detection/segmentation algorithms for the detection of multiple myeloma.	VGG Image Annotation software Mask-RCNN and Unet.	segmentation of myeloma cells images.	Mask-RCNN (%) accuracy 93% Unet (%) was 89%.	it failed to detect some of the under-stained cells [35].
a deep automatic Learning approach for multiple myeloma diagnosis and segmentation has being investigated.	Deeplabv3+ model U-net structure to learn the segment.	segmentation of multiple myeloma plasma cell Images.	They got high score 0.93 in bm deep team [19].	-
Deep learning approaches were used to automatically merge PET and CT characteristics for whole-body MM bone lesion diagnosis in 3D.	It was decided to use V-Net and W-Net to segment find the lesions	68Ga-Pentixafor PET/CT Imaging scans of MM patients.	Accuracy (Dice score 72.98%).	The current study is constrained by the tiny patient data set [36].
provides a reliable technique for classifying acute lymphoblastic leukemia. Using the SN-AM dataset, leukemia (ALL) and multiple myeloma (MM) were studied.	DCNN model's-baseline contrast with several cutting-edge techniques, including (SVMs), DT, Random	B-ALL and MM cancer dataset in white blood for stain normalizing.	Accuracy 97.2%.	admitted that a more extensive experimental research taking. The reliance on the databases' size has not

	Forests, Naïve Bayes.			been presented and performed [20].
Utilizing computer-aided detection, better osteosarcoma detection and diagnosis (CAD) and diagnosis (CADx).	VGG19 and Inception V3 are two examples of transfer learning models that are employed.	Public dataset on osteosarcoma histological images.	Accuracy 96% [37].	-
Increase the rate of multiple myeloma identification, carry out early and accurate illness care, and use artificial intelligence for diagnosis a system is created.	GBDT, SVM, DNN, and Random Forest.	Include 2,446 cases of non-myeloma and 1,741 records of multiple myeloma (MM).	Precision (92.9%), recall (90.0%) and F1 score (0.915) [38].	-
Multiple myeloma diagnosis and staging by combining machine learning techniques with serum-based laser-induced breakdown spectroscopy.	Machine learning and multivariate statistics techniques, such as PCA, KNN, SVM, and ANN classifiers.	130 cases of serum samples were taken from MM patients who were registered.	Accuracies 90%.	Applications to other types of malignancies in under investigation [39].
They provide an asymmetric cascade network for the longitudinal prediction of future bone lesions for T1 weighted whole body MR images.	The proposed cascaded architecture, consisting of two distinct configured UNet.	Bone lesions for body MR images.	(AUC) 0.8023.	They observed that anomalous bone regions are the main triggers for false positives

				predictions, which do not progress to lesions [40].
Creating a decision-support tool to monitor the multiple myeloma (MM) disease treatment response in spinal MRI scan pictures.	They combined a CNN with a RNN, referred to as C-RNN model.	(MM) disease in spinal MRI scan images.	AUC of 0.801±0.037 [41].	-
Present PCSeg Tool for plasma cell segmentation from microscopic medical images.	Nuclei mask based on k-means in the segmentation of cluster cells	Images of bone marrow aspirate slides of patients with MM.	Implemented PCSeg Tool-1 score 79.03% [42].	-
Classify five types of white blood cell: Basophil, Eosinophil, Neutrophil, Lymphocyte, Monocyte.	VGG-16 and ResNet-50	Five types of white blood cell: Basophil, Eosinophil, Neutrophil, Lymphocyte, Monocyte images.	The results show that ResNet-50 is the best and can achieve 88.29 % accuracy [54].	-
Classify between different types of white blood cells, namely, eosinophil, lymphocyte, monocyte and neutrophil by using Gaussian and median filter.	The CNN was coupled with Alexnet, Resnet50, Densenet201 and GoogleNet	Images of different types of white blood cells in the database.	DenseNet201 architecture has the highest accuracy after applying the Gaussian filter, with 83.44 percent.DenseNet201, median filter was applied, the highest accuracy rate was obtained by AlexNet architecture	-

			with 80.77% [55].	

[6] Novelty of the project

Several studies have suggested that myeloma risk is associated with a past history of infections, inflammatory conditions, and allergy-related disorders [16]. The poor prognoses that are associated with most cancers create a sense of urgency for the brains behind healthcare Artificial Intelligence (AI) research. Through AI, it is able to detect cancer and other diseases earlier than possible through standard diagnostic methods, which could be lifesaving for future patients [11].

Multiple myeloma (MM) is an incurable but treatable malignancy involving uncontrolled proliferation of malignant plasma cells in the bone marrow [17]. The analysis of histopathology by a pathologist plays an important role in the diagnosis and prognosis of multiple myeloma [28]. It is a difficult disease to diagnose because it is so complex. There have been very few studies and research to detect this disease by deep learning. That is why we chose this disease, as there hasn't been much research done on it. We presented some of them to demonstrate the importance of this condition and to try to diagnose it as soon as possible due to the emergence of some consequences for patients that generated many symptoms. There have been numerous studies to classify this disease, but the majority of them have relied on different data such as imaging rays, blood protein analyses, and comparative analyses between multiple myeloma and leukemia. This is why this type of data was chosen due to its lack of use in previous studies. Some of the model's studies were looked at in a different way, such as VGG-16, VGG-19, MobileNetV1, and MobileNetV2. But on small data, although some models got high accuracy, the VGG16 model was the lowest of them at 94%, so we aim to do the investigation differently to raise the accuracy from previous studies.

Predominately, early diagnosis helps to concentrate on detecting symptomatic patients as early as possible, while screening consists of testing healthy individuals to recognize those having cancers before any symptoms emerge. Survival rates inform us what percentage of people with the same type and stage of cancer are still alive a certain amount of time (usually 5 years) after they were diagnosed. Early detection is important as it is one of the first actions taken to prevent disease. Besides, patients whose cancer is detected early on have a higher chance of healing their disease and survival. The main novelty of the project appears to be effective deep learning using an algorithm to detect multiple myeloma by extracting features from images and identifying if they are normal or abnormal cells. This can be very effective for keeping the patient's life as normal as possible. The novelty of this project is choosing accurate models with several advantages. VGG-16, the model that has proven to be effective among the models with high accuracy and correct prediction, was used. Previous studies of this disease used different datasets and different models. When they used the same model, they obtained an acceptable accuracy, but not a high one with fewer datasets. As a result, we aim to demonstrate the causes that affect the high accuracy of the test. The work of generating images, where the available data are 755 images of abnormal cells and 1598 images of normal cells, to reach 3000 images Increasing the number of images helps the model increase its accuracy. This type of work seeks to develop its mechanism as well as use the increase in the number of layers in the VGG16 model to obtain high accuracy as well. We will also apply the Resnet50 models. This is a new kind of algorithm that has never been used before. It is most commonly applied to analysing visual imagery.

[7] Project Requirements

Nowadays, many developers strive to develop new systems or machines depend entirely on technologies in order to make life easy. These systems should have a set of specified requirements in order to fulfil the user needs. Accordingly, the requirements of the proposed project are as follows:

A. Software Requirements

As a part of the development process, the software requirements should be specified. The software requirements of the project and their uses as following in Table 2:

Table 2 Software requirements of this project.

Library	**Uses**
Numpy	Perform a wide variety of mathematical operations on arrays.
Pandas	Used to perform data manipulation and analysis.
Operating Systems	Contain and manage all the programs and applications that a computer or mobile device is able to run
OpenCV-Python	Used for all sorts of image and video analysis.
Tensorflow	It is an open-source artificial intelligence library, using data flow graphs to build models
Sklearn.Model_Selection	Provides train/test indices to split data in train/test sets
Matplotlib	its ability to play well with many operating systems and graphics backends.
Keras_Preprocessing	provides utilities for working with image data, text data, and sequence data

B. Hardware Requirements

The project is implemented as an automated system. Accordingly, the hardware requirement that is used to develop the project is a computational platform with GPU, which is Colaboratory (Colab) for training and testing and a Google-based service or local computational environment for deployment and Jupiter notebook. They are especially well suited to machine learning, data analysis, and education.

C. Dataset

1) Collection Datasets

Data is collected from the histological images of plasma cells that are normal and abnormal. Then the software is trained to predict the patient's condition, whereby the histological images can be measured in such a system in order to improve pathological procedures. The system digitizes glass slides with stained tissue sections into extremely high-resolution images, making computerized image analysis applicable [29].

As shown in Figure 8, bone marrow aspirates from patients with multiple myeloma, illustrating a preponderance of mostly mature-appearing plasma cells (purple cells) with eccentrically placed nuclei and a few scattered lymphocytes (blue cells).

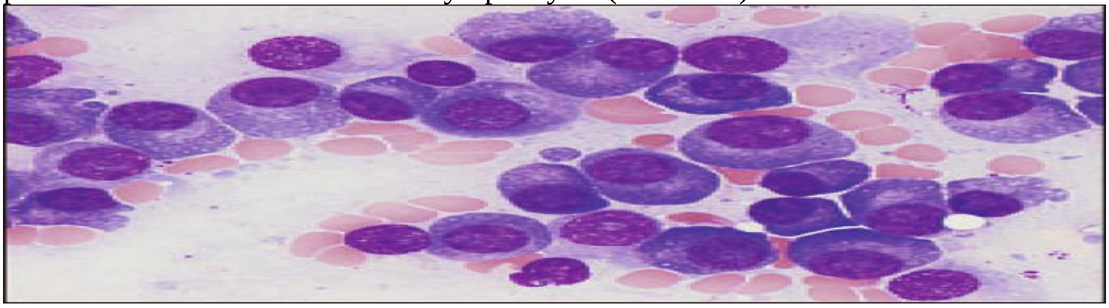


Figure 8 Bone marrow in multiple myeloma showing plasma-cytosis [30].

Data is collected from the histological images of plasma cells that are normal and abnormal. Then the software is trained to predict the patient's condition, whereby the histological images can be measured in such a system in order to improve pathological procedures. The system digitizes glass slides with stained tissue sections into extremely high-resolution images, making computerized image analysis applicable [29]. As shown in Figure 8, bone marrow aspirates from patients with multiple myeloma, illustrating a preponderance of mostly mature-appearing plasma cells (purple cells) with eccentrically placed nuclei and a few scattered lymphocytes (blue cells).

As for the normal cell’s images, they were found on the Kaggle site in the subject (Bone Marrow Cell Classification), where there is a group of images of abnormal and normal cells of bone marrow, where the total number of those images was 171k. Normal cells of the types (Basophils and Monocytes) were used, and the number of cells of Basophils was 441, while Monocytes was 4440. 1598 images were used in this project, including 441 Basophils and 1157 Monocytes. We performed pre-processing of images, feature extraction from the convolutional neural layers of the proposed model, classification, and finally evaluation of the model. Before being fed to the CNN, the data had to undergo various preprocessing steps like resizing and label binarizing.

1) Process of Dataset

In this research, we used a set of data that included images of normal and abnormal cells. We collected abnormal images from the SegPC-2021 and the normal cells from Kaggle [53]. We performed preprocessing of images as feature extraction from the convolutional neural layers and splitting images into classes. Before being fed to the CNN, the data had to undergo various preprocessing steps like resizing, label binarizing, and data augmentation. In general, the idea of the program is to enter the images and extract the features from them by resizing them in a specific way (224,224,3) as it needs to generate them. We generated abnormal images from 750 to 1500 images in a separate code where the number of normal images is available at 1500 images. As an alternative dataset, we will use bone marrow (blood cell images) that are found in Kaggle [58]. The data was then augmented using the standard method of reflecting more training samples. Before we fed the images to the network, we normalized them between values [0,1] using the formula given below by Equation 5.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (5)$$

where x is the un-normalized pixel intensity and $x_{norm}$is the normalized pixel intensity and are minimum and maximum intensity values of the whole image.

We generated images of abnormal cells from 755 to 1500 images through the keras library, the generation was chosen based on the "reflect".

D. Functional Requirement

The following is a list of the project's functional requirements:
-Ability of the models (VGG-16, VGG-19, and Resnet-50) to learn important features from the datasets.
- Ability of the convolution neural network model to classify medical images with the best performance in accuracy.

E. Non-functional Requirement

The non-functional requirements specify the criteria that must be taken into consideration during the planning process. These criteria judge the operation and performance of the project. The following are the non-functional requirements of the project:
-Usability: using an application or programming to detect cells as normal or abnormal is a simple and straightforward interface that allows different doctors and pathologists to easily and efficiently achieve their goals.
-Performance: The programming process can be executed rapidly and efficiently.
-Maintainability: The programming can be updated to add new functions. The maintainability will help to reduce defects and improve the system's functionalities.

[8] Expected outcomes

The aim of the project is to develop traditional cell detection methods by using deep learning, which has many advantages over traditional methods. In addition, the proposed system will reduce the time and effort required for manual searching. Identifying the best approach to the project, which will lead us to achieve the project's results. Therefore, the following are the project's anticipated results:
- Obtaining a high accuracy more than 95% of deep learning algorithms such as (VGG16-VGG19) and ResNet50, after data processing, classification and training of the model. then analyze the test results through the confusion matrix.
- Improved performance and discover the reasons that help increase accuracy by making a comparison between the number of images used and the number of epochs after applying it to the algorithms (VGG16, VGG19, Resnet50).

[9] Solution Identification and Justification

There are different techniques that will be used to help us develop the proposed program and achieve our goal. In this section, we will present several techniques that can be used to develop the project. The project requires images of cells and algorithms that help us detect these cells. The basis of these algorithms is the convolution neural network (CNN). We use it to solve the problem of detecting whether cells are normal or abnormal in microscopic biopsy images by using the Convolutional Neural Network (CNN). One interesting feature of the convolutional neural network compared to its predecessors is that it automatically detects the essential features without depending on humans with high precision [21]. There are three types of layers that make up the CNN [24], which are the convolutional layers where filters are applied to the original image, or to other feature maps in a deep CNN; pooling layers are used to reduce the dimensions of the feature maps; and fully-connected (FC) layers operate on a flattened input where each input is connected to all neurons when these layers are stacked [22]. Convolution having two-dimensional image, I and array K with size h×w (also called as convolution kernel) convolute image I*K is calculated by putting the kernel on image [23]. which is given by Equation 1.

$$(\mathrm{I} * \mathrm{K})_{xy} = \sum_{i=1}^{h} \sum_{i=1}^{w} K_{ij} \times I_{x+i-1,y+j-1} \qquad (1)$$

(CNN) are similar to common ANNs in that they are composed of neurons that self-optimize through learning. Every neuron will still get input and execute an operation (such as a scalar output followed by a non-linear function)—the basis of countless ANNs. From the input raw image vectors to the final output of the class score, the entire network will still display a single perceptive score function (the weight). The last layer will contain loss functions associated with the classes, and all of the regular tips and tricks developed for traditional ANNs still apply [31]. The CNN, one of the deep learning architectures, belongs to the class of feed-forward neural networks. The two significant steps involved are learning features and classifying data, performed by the input layer, hidden layer, and output layer. The images with a pre-defined size are stacked onto the input layer. The size of the image is in the format of height, width, and depth. For RGB, the depth is 3 and 1 for grayscale images. The first segment makes up the feature learning layer, where the convolution and pooling operations are the two primary tasks carried out. The other layers that are under consideration are the normalization layer and the activation layer. The convolution layer is shown by the stride, padding, filter size, and number of filters. The aftereffect of the convolution layer is a feature map, the size of which is given by Equation 2.

$$\boldsymbol{O} = \frac{\mathbf{W-K+2P}}{\boldsymbol{S}} + \mathbf{1} \quad (2)$$

where W denotes the input height/length, K is the filter size, P means the number of zero padding, and S is the stride. When there is an increase in the number of convolution layers, more complex features can be learned. The goal of batch normalization followed by a convolution layer is regularization. ReLU is the activation function utilized (rectified linear unit). Equation 3 gives the relation for ReLU.

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{max}\,(\mathbf{0};\ \boldsymbol{x}) \quad (3)$$

The max pooling operation performed here is for down sampling, which helps in minimizing the size of the feature map. The output layer is the one where the classification is performed. which consists of a fully connected layer, a softmax layer and a classification layer. The input to the fully connected layer is given by the relation is given by Equation 4.

$$\mathbf{y} = \mathbf{WT}\,\boldsymbol{\varkappa} + \mathbf{b} \quad (4)$$

Here, the output class (y) is obtained by multiplying the input (x) from the hidden layer by the weight matrix W, which is added with a bias. The softmax layer converts the raw values of the output classes into normalized scores. The result of the prediction is a probability value of class occurrence. Finally, the classification layer provides a class label according to the probability [54].

After building CNN models from scratch, we will try to fine-tune the model by using the image augmentation technique. Consequently, we will employ one of the pre-trained models—VGG-16—to classify images and check accuracy for training data and testing data. The VGG essential

model is used to obtain the feature map. The essential construction of the VGG model is displayed in Figure 10. The convolutional kernel measurements in the VGG model are 1 × 1 and 3 × 3. The activation function is a softmax function, and the reshaping process is executed before and after treatment to enhance the classification impacts [25].
The architecture of VGG-16 uses 13 convolutional layers and 3 fully connected layers. The convolutional layers in VGG-16 are all 3×3 convolutional layers with a stride size of 1 and the same padding, and the pooling layers are all 2×2 pooling layers with a stride size of 2. The default input image dimension of VGG-16 is 224×224. After each pooling layer, the size of the feature map is decreased by half. The last feature map before the fully connected layers is 7×7 with 512 channels and it is extended into a vector with 25,088 (7×7×512) channels [26].

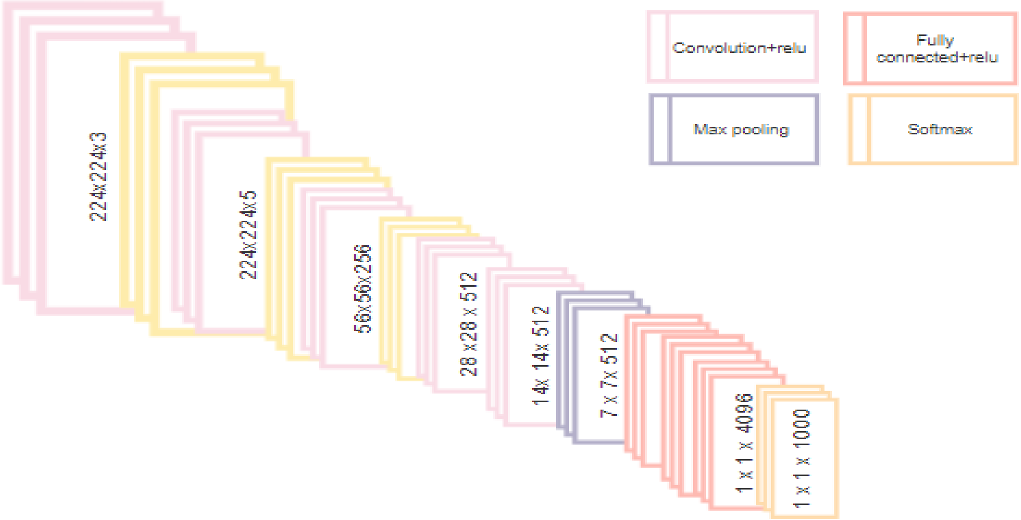


Figure 9 Basic structure of the VGG model [25].
The precise structure of the VGG-16 networks [32]. is as follows:
1- The first and second convolutional layers are comprised of 64 feature kernel filters and the size of the filter is 3×3. As the input image (RGB image with depth 3) reached into a first and second convolutional layer, dimensions changes to 224x224x64. Then the resulting output is passed to the max-pooling layer with a stride of 2.

2- The third and fourth convolutional layers are of 124 feature kernel filters and the size of the filter is 3×3. These two layers are obtained by a max-pooling layer with stride 2 and the resulting output will be decreased to 56x56x128.

3-The fifth, sixth, and seventh layers are convolutional layers with kernel size 3×3. All three use 256 feature maps. These layers are obtained by a max-pooling layer with stride 2.

4-Eighth to thirteen are two sets of convolutional layers with kernel size 3×3. All convolutional layers have 512 kernel filters. These layers are followed by a max-pooling layer with a stride of 1.

5- Fourteen and fifteen layers are fully connected hidden layers of 4096 units followed by a softmax output layer (Sixteenth layer) of 1000 units.

A variation of the VGG model called VGG19 has 19 layers in total (16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer). There are further VGG variations, including VGG11, VGG16, and others. VGG19 has 19.6 billion FLOPs. We notice the difference as shown in Figure 11.

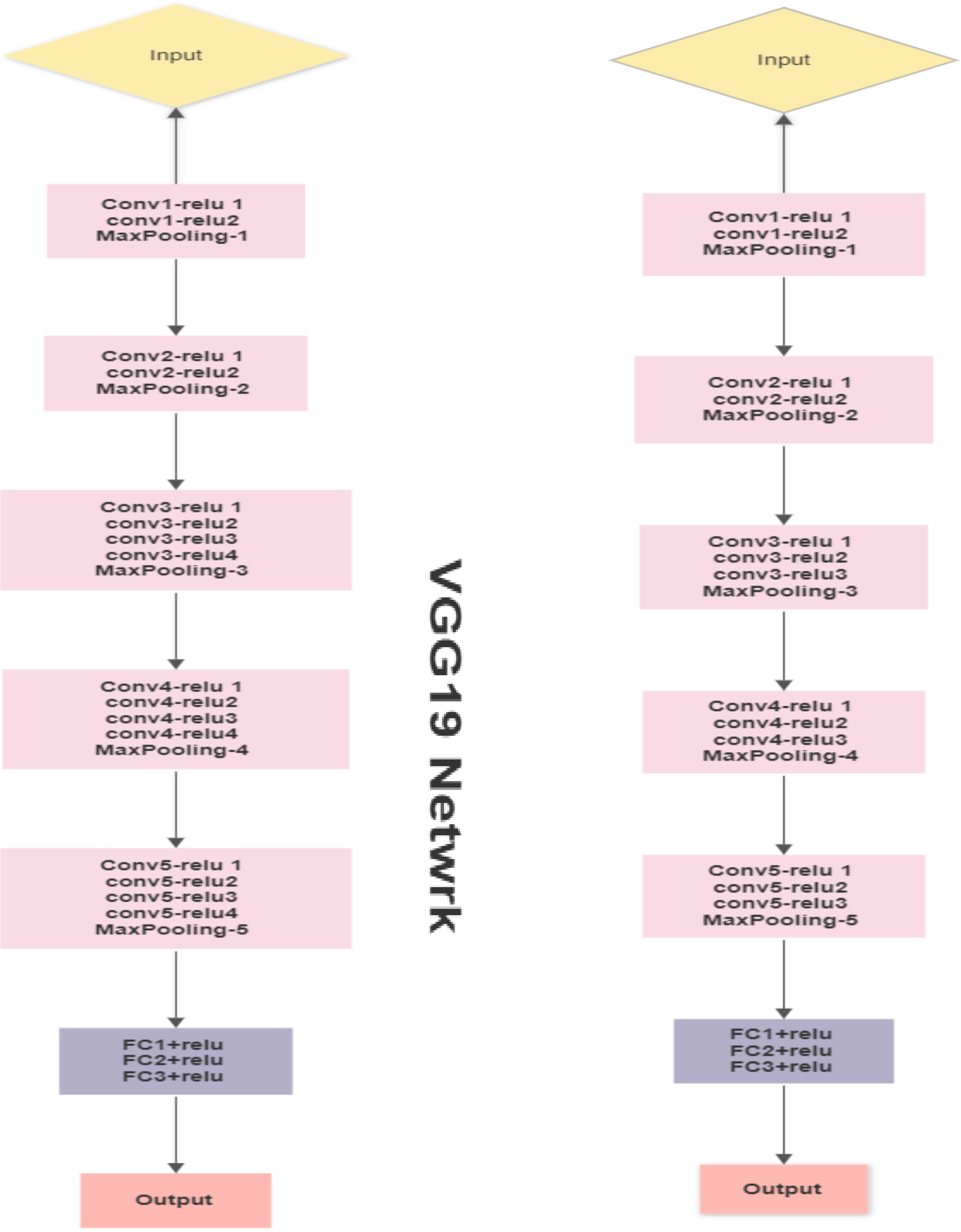
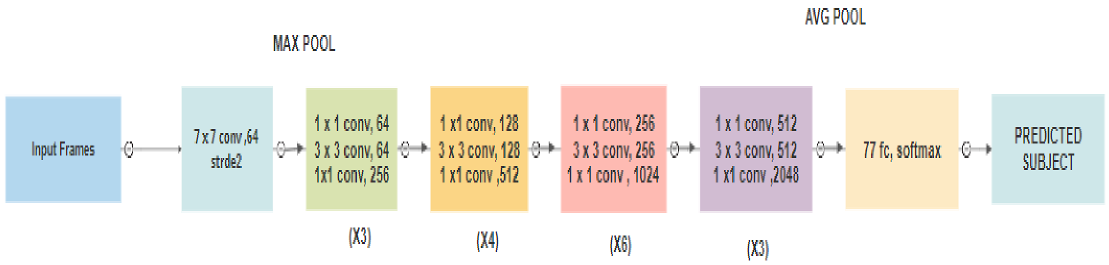


Figure 10 Architectures VGG19 and VGG16 [47].

A multi-layered deep neural network called the visual geometry group network (VGGNet) was developed. The simplicity of the VGG-19, which has 3 3 convolutional layers stacked on top to expand with depth level, makes it handy. Max pooling layers were utilized as a handler in VGG-19 to decrease the volume size. 4096 neurons were employed with two FC layers.

Convolutional layers were used during the training phase for the feature extraction and some of the convolutional layers have max-pooling layers attached to them to lessen the dimensionality of the features. 64 kernels (3× 3 filter size) were used in the first convolutional layer to extract features from the input pictures. The feature vector was created using fully connected layers [47].

ResNet is another representative deep network. Prior to ResNet, the cutting-edge DNN was continuously delving deeper. Deep networks are challenging to train due to the well-known vanishing gradient problem, in which the gradient can become infinitesimally small as it is repeatedly multiplied and propagated to older layers. ResNet can yield more accurate results by using a skip connection (also known as a shortcut connection) to fit the input from the previous layer to the subsequent layer without changing the input.a very deep network of up to 152 layers. There are two kinds of shortcut modules in the implementation of ResNet. The first is a block called an identity that lacks a convolution layer at the shortcut level. In this instance, input and output are of equal size. The other is the convolution block, which has a convolution layer at a shortcut. The input dimensions in this instance are smaller than the output dimensions. In both blocks, 1 × 1 convolution layers are added at the start and end of the network. This is a technique called bottleneck design, which reduces the number of parameters while not degrading the performance of the network so much [52].



It is a widely used ResNet model, and we have explored the ResNet50 architecture in depth. We have found it convenient to carry out our experiments using a typical ResNet-50 neural network design because the dataset under consideration can be directly related to a classification task. as shown in Figure 12.

Figure 11 ResNet-50 neural network architecture [51].

[10] Project solution design

Through analysing the related works and alternative tools and techniques, this section discusses our solution how the algorithms work in order to improve the performance and the accuracy of all results. In this project, we will develop models to increase the accuracy of the results and classify images of cells, whether normal or abnormal. Accordingly, different accuracy models can be compared by classifying the cell-based attributes such as colour, size, and types of cells.

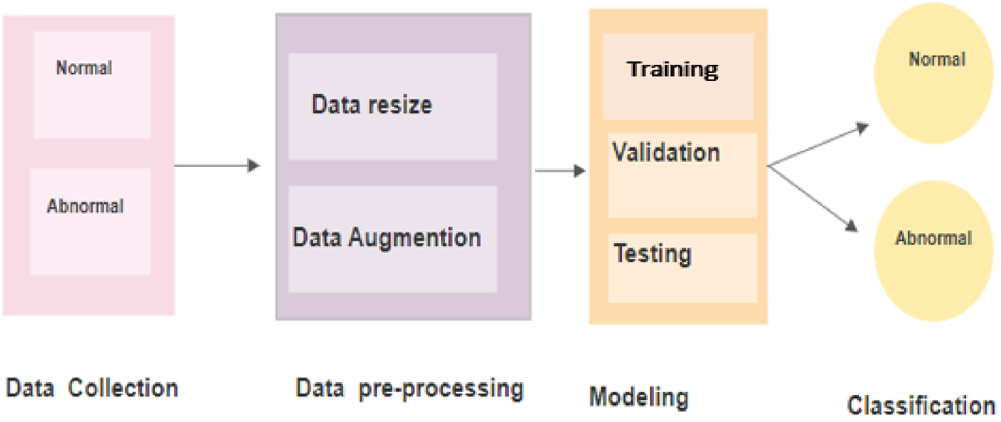


Figure 12 Flowchart the steps of passing data for classification [53].

In the beginning, reading and labeling images of normal and abnormal cells as input to make preprocessing and apply different models. The processing includes creating an image generator for training to apply data augmentation during the training and defining some parameters such as preprocess-input to adequately the images to the format the models Resnet50, VGG-16, and VGG-19 require. The function that properly transforms a standard image into an appropriate input These models then rotate the images by 25 degrees, randomly shift 25% of them horizontally and vertically, reserve 35% for validation data, and add some reflect fill mode, which reflects some pixels in the empty area of the images. also, a generator for validation data with appropriate input images. After that, splitting these images for training and testing can apply models like VGG16, VGG19, and Resnet 50, then make a prediction if those cells are normal or abnormal. As is shown in Figure 13.

F. **Techniques used in Implementation**

In this project, we will use three techniques to detect the cells, whether normal or abnormal. VGG-16 is one of the state-of-the-art techniques that can detect cells and classify 1000 images into 1000 different categories with 92.7% accuracy. It is regarded as one of the best architectures for vision model. The VGG-16 is a deep CNN model that has had a lot of success in the computer vision industry. Also, we look for an increase in accuracy in the VGG16 model, and one of the reasons that helps in increasing the accuracy is the increase in the number of layers. When the number of layers increases. We have VGG19. When we compare with VGG16 and VGG19, we find that VGG19 is a little better, but it requires more memory. The VGG16 model consists of convolutional layers, maximum aggregation layers, and fully connected layers. There are a total of 16 layers with 5 blocks, and each block has a maximum grouping layer. In this technique, detection is evaluated using a cell image dataset by using another model, VGG19. As a result of the experiments, this technique provides high performance in cell detection. Five epochs are used to train the technique. Moreover, it can use a classification technique to use them with VGG-19 to get a higher result. The difference between VGG16 and VGG19 is shown in Table 3 [46].

Table 3 Comparison of VGG16 and VGG19 Layers [46].

Layer	**VGG16**	**VGG19**
Size of layer	41	47
Image Input Size	224 x 224 pixel	224 x 224 pixel
Convolution Layer	13	16
Filter Size	64 & 128	64 & 128 ,256 &512
ReLU	5	18
Max Pooling	5	5
FCL	3	3
Drop Out	0.5	0.5
Sofemax	1	1

The Visual Geometry Group (VGG) created the VGG16 network architecture with 41 layers and the VGG19 with 47 layers. Each layer's 3 × 3 filters are created by VGG to streamline the process. When compared to AlexNet, VGG can build more complex features with less computing by using uniform and smaller filter sizes.

We will also apply ResNet50, a variant of the ResNet model that has 48 convolution layers along with one MaxPool and one Average Pool layer. It is 3.8 x Floating Point operations. Due to its skip connection topology, ResNet-50 has been shown to perform better and use less computational resources than traditional classification architectures like VGG-16. For this purpose, we have constrained the input data (frames) to a 224 × 224 image size, batch size of 32 and output classes to 77 (the number of eligible subjects) during the training phase [51].

The VGG-16 model helps to provide high accuracy in the classification process. The imbalanced dataset is reducing the performance of the classification model. Accordingly, while training the model for 5 epochs, we faced over-fitting, the training accuracy was increasing, and the loss was decreasing. In the test, the loss was increasing, and the accuracy was fixed. We enhanced the performance of the model using one of the most popular ways to prevent over-fitting, which is callback. It is a tool that can execute tasks at different training echelons. This method can help to prevent the model from being over-fitted. Through this comparison, we find the best model that has the highest accuracy and lowest loss, the VGG-16 model, which has proven effective in learning data and classifying normal and abnormal cells with a very simple error rate. We find that the Resnet-50 model has reached its highest accuracy, but the error rate is greater than the VGG-16 model. For this reason, we cannot say that it is the best. The VGG-19 model is e

G. **Evaluation Metrics**

Performance metrics are used to evaluate the performance of ML algorithms. This step is essential to be applied after the model has been implemented. It finds how accurate the model is by calculating the percentage of error.

- **Confusion matrix**

One of those important tools is finding the confusion matrix, which is a summary of the results of predicting the classification problem [52]. The numbers of correct and incorrect predictions are summed up by the count values and broken down by each category. This is the confusion matrix's secret. which is given by Equation 6.

$$\textbf{Accuracy}\ (\text{all correct}/\text{all}) = \text{TP} + \text{TN}/\text{TP} + \text{TN} + \text{FP} + \text{FN} \quad (6)$$

- **Precision**

An important tool for evaluation that we will use is precision. It is the fraction of all positive predictions that are true positive, it helps when the costs of false positives are high and when false positives are too high. After receiving a constant stream of false alarms, individuals who monitor the outcomes will learn to disregard them. which is given by Equation 7.

$$\textbf{Precision}\ (\text{true positives} / \text{predicted positives}) = \text{TP} / \text{TP} + \text{FP} \ (7)$$

- **Recall**

Recall is the fraction of all actual positives that are predicted positive. When the cost of false negatives is high, it helps. which is given by Equation 8.

$$\textbf{Sensitivity or Recall}\ (\text{true positives} / \text{all actual positives}) = \text{TP} / \text{TP} + \text{FN} \ (8)$$

- **F1-score**

F1-score is the harmonic mean of precision and recall, in that weird way that addition and multiplication just mix two ingredients to make a separate dish altogether. That is, having low false positives and false negatives indicates that you are accurately recognizing real threats and are not bothered by false alarms, which is what a good F1 score indicates. An F1 score is considered perfect when it’s 1, while the model is a total failure when it’s 0[48]. which is given by Equation 9.

$$\textbf{F1} - \textbf{Score} = \quad 2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall})) \ (9)$$

[11] Implementation models

When implementing the model, we first prepared the steps that help us to clarify the implementation stages in an orderly manner, and these steps contain 12 steps. As shown in Figure 14, The first step is to import libraries and the weights, which means importing all the necessary libraries for loading our data and the weight for VGG-19. The second step is to download the data, which contains a training file divided into two parts: normal cells (1598 images) and abnormal cells (1598 images), so their total is 3196 images. In the third step, we will count the images in the training file of the VGG-16 model, VGG-19, and Resnet-50, which have a total number of 3598 images. In the fourth step, splitting validation data from train data, In the "Visualize some images" step, we displayed a set of images in the data file that contained images of normal cells and abnormal cells. Then step to load all the models, which use VGG-16, VGG19, and Resnet-50. Next, we compile the model and move on to the training model step, which is the training of images with different epochs to find the best suitable resolution. Then, we evaluate each model step after training. After that, by plotting the training and

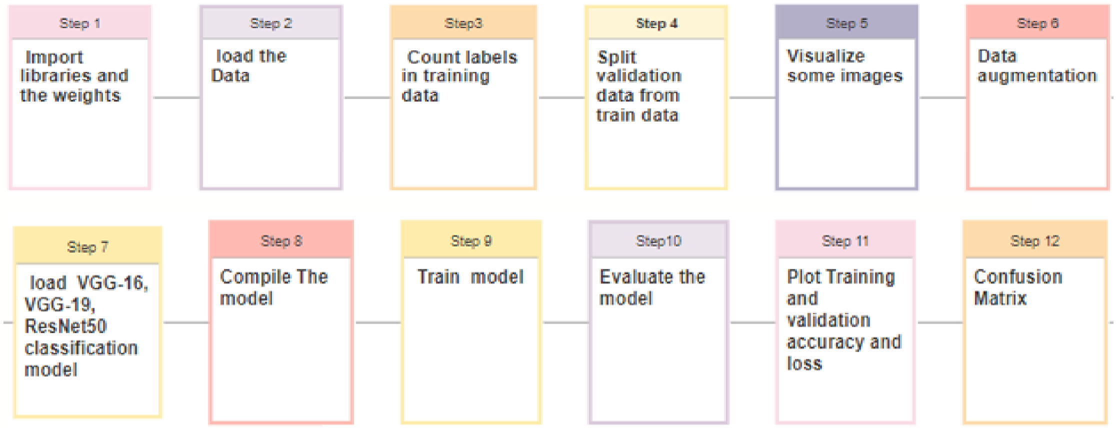


Figure 13 steps of coding work [50][53].

validation accuracy and training and validation loss steps, different forms of model results are found as a result of comparing the loss and accuracy of the training data. Finally, the confusion matrix step to find the difference in confusion matrix values and calculate several important analyses to know the correctness of the program's work. Implemented models discusses about three section (12.1) Dataset Preparation (12.2) Detection Model Training, and (12.3) Experimental Results and Discussions.

H. **Dataset Preparation**

We decided to use sample image cell datasets, which are abnormal images from the SegPC-2021 and the normal cells from Kaggle [53], datasets in order to generate a generalized model that can work in different situations and provide more accurate and reliable results. Each of these datasets has a limitation, like it does not have the full-frame images, or it does not have the blurred images. Based on that, we decide to increase the diversity on the training model through generation datasets with various conditions.

At the beginning, we used images from the kaggle [53] dataset, which contains normal cells identified as expert-annotated cells from the bone marrow smears of 945 patients stained using the May-Grünwald-Giemsa/Pappenheim stain. The images in the SegPC-2021 [33] dataset include abnormal cells; they used stain color normalization. Figure 15 shows a sample of abnormal cells (above) in which the nucleus and cytoplasm are different in shape and number, and normal cells (bottom).

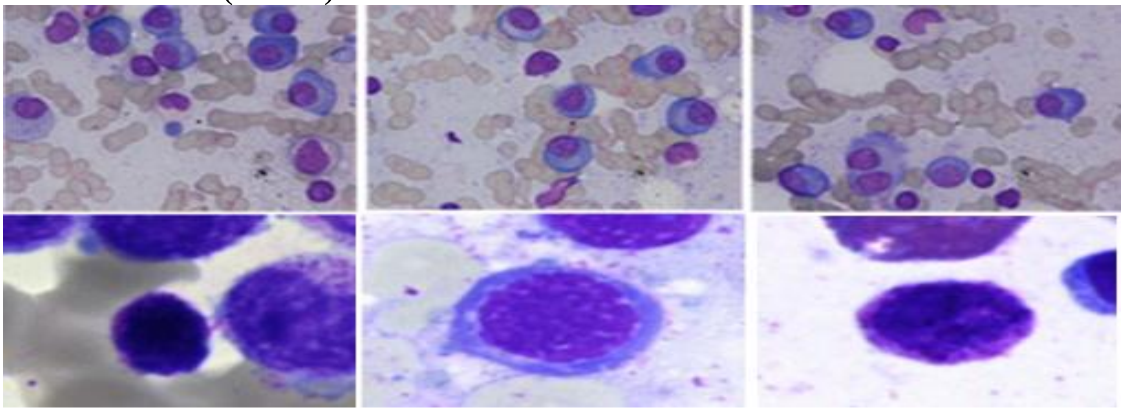


Figure 14 A sample of abnormal cells (above) and normal cells (bottom) [33][53].

We used a set of images, around 1598 images for each of them, and created labels for two files, normal and abnormal, for these images. The sample images were selected based on their different colors and sizes to train the model with several situations in images. The labeling file contains information image size 224*224 with 2 different types of images: bmp and jpg.

This process was performed using ImageDataGenerator in the VGG16, VGG19, and Resnet50 models, which is used to preprocess images before feeding them into networks. A small amount of data is not enough if we want to make our model more robust. It should instead be used with data augmentation, which generates random data. Also, there are some features of the parameters that we used as follows:

1. Apply rotation: rotate the image by 45°.
2. Shift the image to the left or right (horizontal shifts by 20% from the original).
3. High shift range: shift vertically (up or down by 20%).
4. Zoom out the image by 20%.
5. Shear: the image will be twisted along a hub, or an axis, to see the image from different angles.

6. Horizontal flip: it flips the image horizontally.
7. Fill mode: fill the image with its reflection to avoid black frames around it.

Moreover, to avoid over-fitting and increase the accuracy of the detection model, we perform the data augmentation technique. We applied the data augmentation process by performing the horizontal flip, fill mode, and high shift range, as is shown in figure 16. This process provides three images for each image in the combined dataset. The total number of images increased and became 3196 images for the detection task. The figure shows the images of different techniques that are used for data augmentation, which significantly impacts detection performance. A dataset was used, and three images were formed using different data augmentation techniques.

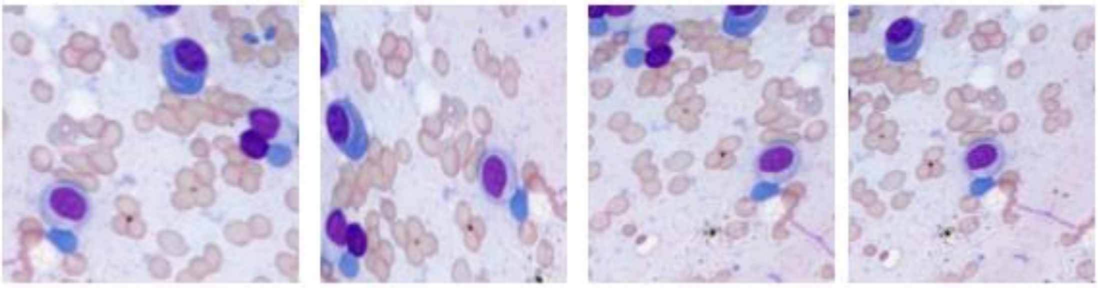


Figure 15 On the left is an original image, and on the right are many augmented images created from that original image by reflect [50].

A. Detection Model Training

For the detection task, VGG-16, VGG-19, and Resnet50 were used. These models are based on training and testing datasets to detect the cells, whether normal or abnormal, with high accuracy. We trained the model on the Google Colab with a graphics processing unit (GPU) and 12 GB of RAM.

Therefore, we implemented the VGG-16, VGG-19, and Resnet50 models, which classify well-known techniques to detect the cells of humans. The model was implemented on Google Colab to allow us to set the runtime as a GPU, which accelerates the training process. In addition, start importing warnings to avoid warnings, and import os to provide functions for interacting with the operating system.

In addition, we used a pre-trained model on Imagenet by downloading the model weights VGG16 and VGG19. These models are the most popular models used by researchers to initialize the neural network. Files are used to store the images that will be used in training and testing, which contain the names of classes. In our case, we have two classes: normal and abnormal.

In addition, Keras provides both the 16-layer and 19-layer versions via the VGG16 and VGG19 classes. It also downloads the weight files from the Internet and stores them. The model expects images as input with a size of 224 x 224 pixels with 3 channels. As a result, we first used the load_img () function to load the image and resize it to the required size of 224×224 pixels. Next, converted the pixels to a NumPy array with Keras. Use the img_to_array () function to accomplish this. Given that the network anticipates receiving one or more images as input, the input array must be four-dimensional (samples, rows, columns, and channels). But we have only one sample (one image). So, we reshape the array by calling reshape () and adding the extra dimension. The image pixels must then be ready. So, preprocess_input should be used to prepare new input for the network.

Models	**Dataset**	**Training Set**	**Testing Set**	**Total Number of Samples**
VGG-16	Normal	958	636	3196
	Abnormal	959	643	
VGG-19	Normal	1497	96	3186
	Abnormal	1497	96	
Resnet-50	Normal	1497	96	3186
	Abnormal	1497	96	

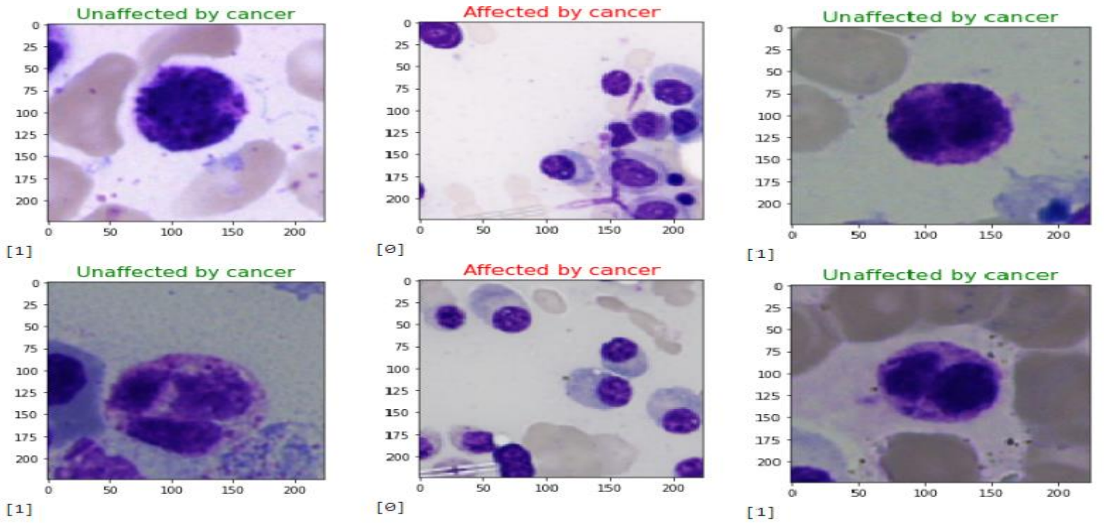
Table 4 The summary of training and testing set for models.

At the end, we prepared the training and testing files, which contained the normal and abnormal. We split the images into 60% for training, and 40% for testing sets by splitting function train_test_split () in VGG-16 while VGG-19 and Resnet-50 splitting manually. The summary of the training and testing data for each experiment is shown in Table 4.

We used a batch size of 64, which means the number of images that will be used for every training step, and set the subdivisions to 5, which means splitting the batch into 5 mini-batches. Thus, it uses 5 images per mini-batch and that will be repeated 5 times until the batch is completed, at which point a new iteration will start with 64 new images. The number of iterations depends on the max batch parameter. We set the maximum batch size to 50. For each epoch, all data samples are passed through the network. The learning rate starts at 0.001 and remains fixed for 6400 iterations. The learning rate demands a decrease over time. Accordingly, it will multiply the learning rate by the scale factor, which is 0.1, to get the new learning rate. We upload the images to Google Drive and mount them on the Google Colab to be used for the training and testing.

[12] Experimental results and discussions

Figure 17 shows sample detection results. Most of the classes for each normal and abnormal cell in the image are predicted correctly. The model mostly provides high performance in detecting the cells and classifying which of them are normal or abnormal.



Model	Testing Set		Total Testing Samples	Precision %	Recall %	F1 - score	support
VGG-16	Normal	636	1279	100%	99%	100%	636
	Abnormal	643		99%	100%	100%	643
VGG-19	Normal	96	192	96%	99%	97%	96
	Abnormal	96		98%	95%	97%	96
ResNet-50	Normal	96	192	99%	99%	99%	96
	Abnormal	96		100%	99%	99%	96

Figure 16 Sample of detection model results.

We used different metrics to evaluate the model, which included confusion matrix values to calculate several important analyses in knowing the correctness of the program's work. A confusion matrix can obtain four different combinations from the predicted and actual values of a classifier, which are true positive, false positive, true negative, and false negative. which means the number of times our actual positive values are equal to the predicted positive. That means the predicted value has a positive value and is correct. which means the number of times our model wrongly predicts negative values as positive. That means the predicted negative value is actually positive. True Negative, which means the number of times our actual negative values are equal to predicted negative values, that means we predicted a negative value, and it is actually negative. which means the number of times our model wrongly predicts negative values as positive. The predicted negative value is actually positive. One of the most important of these important analyses is precision, recall, f1-score, and support, which are also found in the following forms for the various models: VGG-16, VGG-19, and ResNet-50.

As well, both precision and recall are computed using TP, FN, and FP. Table 5 shows the comparison between the three models using the evaluation metrics. This table shows the results from each model using the test set dataset. All models produce high results.

Table 5 Shows the comparison between the three models using the evaluation metrics.

In machine learning, learning curves are a common diagnostic tool for algorithms that take incremental learning cues from a training dataset. After each update during training, the model can be tested on the training dataset and a holdout validation dataset, and graphs of the measured performance can be made to display learning curves. Train Learning Curve and Validation Learning Curve are common to create dual learning curves for a machine learning model during training on both the training and validation datasets.

- **Train Learning Curve**: A learning curve calculated from the training dataset that gives an idea of how well the model is learning.
- **Validation Learning Curve:** A learning curve is calculated from a hold-out validation dataset that gives an idea of how well the model is generalizing.

Also, we find different forms of model results as a result of comparing the loss and accuracy of the training data. This shows us the amount of error during the image training and validation work. In Figures 18. The training accuracy curve increases over epochs from 97% to 100%. It was below validation, but after 10 epochs, the curve increases. The loss curve of training decreases from 0.17 to almost 0 after 10 epochs. And the validation loss curve was almost static at 0.02, which is a pretty good result. Because we want to minimize the loss and maximize the accuracy.

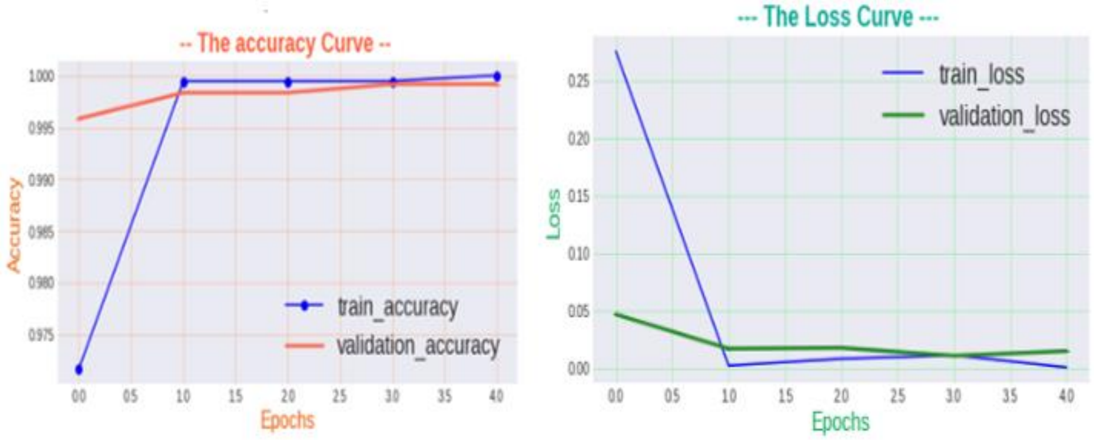


Figure 17 Training and validation accuracy and Training and validation loss in VGG-16.

In Figure 19, the accuracy curve shows:

The validation curve is above my training curve. And the two curves increase from 60% to more than 90% in just 14 epochs.

The loss curve shows:

The validation curve is below training, and the two curves decrease over 14 epochs, from 0.6 to 0.3 for validation and from 0.8 to about 0.4 for training loss.

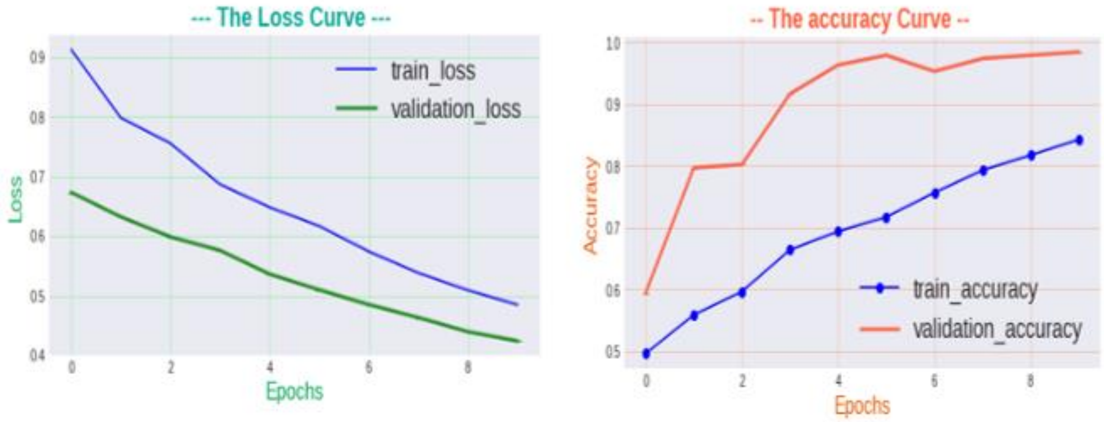


Figure 18 Training and validation accuracy and Training and validation loss in VGG-19.

In Figure 20, the accuracy curve shows

The training curve is oscillatory, almost static. However, the result is pretty good since it reaches above 95%, which means the model is approximately 95% accurate.

The validation curve is oscillating as well, reaching above 90%, which is also a good result.

The loss curve demonstrates that the training loss oscillates to below 0.2 while the validated loss, represented by the green curve on the left, ranges between 0.01 and below 0.2, which is logical when compared to the accuracy result. The validation and training loss learning curves show noisy movement.

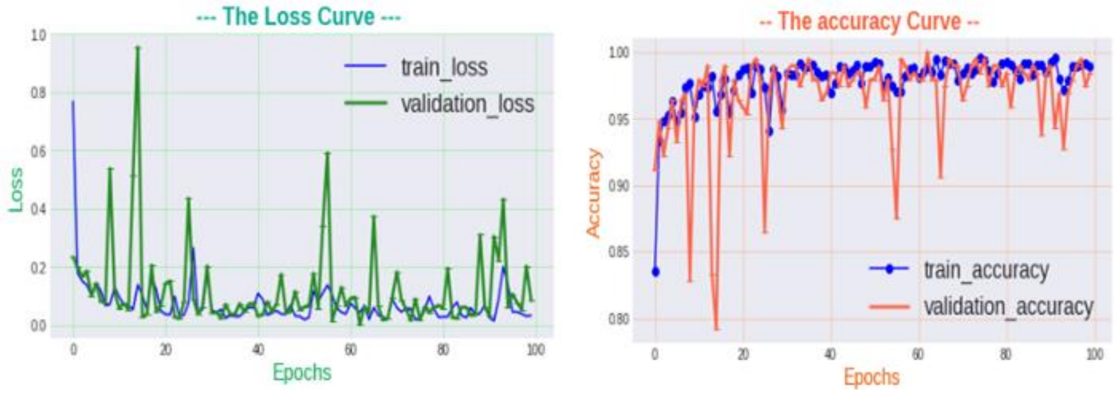


Figure 19 Training and validation accuracy and Training and validation loss in ResNet-50.

Table 6 Comparison between the algorithms VGG-16, VGG-19, Resnet-50.

Model	**number images=1500**		**number images=3000**	
	Accuracy	**epochs**	**Accuracy**	**epochs**
VGG-16	99.67%	epochs=5	99.77%	epochs=5
VGG-19	96.67%	epochs=10	97.92%	epochs=10
ResNet50	95.02%	epochs =70	97.40 %	epochs=100

A comparison was made between two groups of the number of images as shown in Table 6, the first group containing 1500 images, and the second group containing 3000 images, to find out the effect of the number of images in increasing the accuracy of the various algorithms. We note in the VGG-16 algorithm that it reached an accuracy of 99.77% at epoch = 5, more than compared to the other group, which has a small percentage of accuracy.

We also find that the accuracy of the VGG-19 models when the number of images is 3000 is more accurate than the group that contains 1500 images and with the same number of at epoch = 10, where the accuracy reached 97.92%. We note that accuracy is good to know the work of the program correctly.

When using the ResNet-50 model, we found a significant difference compared to the other models. We note that when the group contains 3000 images, it has a different resolution than the other resolution because of the difference in epochs. For example, when the number of epochs is increased to 100, we find that the accuracy reaches 97.40%, but when epochs = 70, the accuracy is 95%. For this reason, we conclude that the greater the number of epochs, the greater the accuracy. When comparing the algorithms, we find the best algorithm in terms of accuracy is VGG-16, which has an accuracy of 99%. Through this, we prove that this algorithm is effective and correct in detecting myeloma disease as well as having the ability to detect in the future other diseases that were not previously detected.

[13] Discussion and future work

Early diagnosis improves cancer outcomes by providing care at the earliest possible stage, therefore, an important public health strategy in all settings. So, the purpose of the project is to detect cells using useful technology and advance a method to improve the medical field by combining it with deep learning through the use of advanced algorithms.

However, the performance of the developed program to detect cells depends basically on the performance of the detection and classification models. Therefore, the detection model was trained using samples from two different datasets to increase diversity and make the model more generalizable. We used an augmentation dataset that helps to increase the performance of the classification model. One of the important factors that increases the accuracy and performance of detection cells is the high number of epochs. These epochs can achieve higher accuracy where the models start detecting the cells, whether normal or abnormal. In the future, the detection and classification models can be improved to handle further challenges related to domain tasks. Furthermore, this model can be used with different diseases and is rare. Finally, we would like to continue optimizing the performance and increasing the accuracy of the results. Additionally, the system can be extended by including other related diseases such as leukaemia and myeloma.

[14] CONCLUSION

The goal of this study is to use deep learning technology to classify myeloma images and to clarify the importance of early detection of multiple myeloma for a variety of reasons that benefit both the patient and the specialist, as well as provide an opportunity to further develop the program in the future. We classified cells with multiple myeloma with high accuracy by using a general dataset with complicated algorithms and then assessing several prediction models for specificity, sensitivity, and accuracy evaluation. We applied the VGG16, VGG19, and ResNet50 models to detect cells, whether normal or abnormal, analyze the results using deep learning tools, and then compare the results between them.This milestone report presented the problem, statement, motivation, scope, methodology, and expected outcome. We also provided related work, problem analysis and solution design, and the tools and techniques to be used during the project. We also provided the implementation details of techniques that were used in the project. The limitation of this project is that it is difficult to find more images of abnormal cells that lead us to generate images by special techniques.

References

[1] J. M. S. Singhal, "Multiple myeloma", 2006 .

[2] F. G. K. S. L. F. D. A. P. R. D. W. K. R. I. K. N. Z. N. M. B. H. H. L. Arnold Bolomsky and P. Validation, "Studies Support Causal Machine Learning Based Identification of Novel Drug Targets for High-Risk Multiple Myeloma", 2018 .

[3] F. A. S. R. E. F. S. H. Y. M. K. S. A. H. A. Z. S. A. F. A. F. M. S. W. R. N. A. C. F. A. M. H. K. M. A. iDAhmed Kotb Abdrabou, "Outcomes of autologous stem cell transplantation for multiple myeloma in Saudi Arabia", 22 Aug 2021.

[4] M. A. A. H. H. Abduljalil OZ, " Retrospective observational study on multiple myeloma cases admitted to Kfsh", June 2006 till the end of December 2013.

[5] W. H. O. C. C. Profiles, 2014, accessed 6 December 2021,< https://www.who.int/cancer/country-profiles/sau_en.pdf>.

[6] T. L. J. T. A. A. L. N. R. F. L. ,. M. O. A. a. N. A. Stine Rasch, " Multiple Myeloma associated Bone Disease", 2020.

[7] V. B. X. P. E. K. G. G. M. D. T. T. P. P. Marie-Christine Kyrtsonis, ", Genetic and molecular mechanisms in multiple myeloma: a route to better understand diseas pathogenesis and heterogeneity", 2010.

[8] L. R. M. J. H. R. E. F. R. M. &. S , Li Shen, " Deep Learning to Improve Breast Cancer Detection on Screening Mammography", 2019.

[9] Muralikrishna Puttagunta & S. Ravi, "Medical image analysis based on deep learning", 2021.

[10] K. H. C. K. I. M. K. K. S. W. S. F. &. S. ,. Keisuke Kawauchi, " A convolutional neural network-based system to prevent patient misidentification in FDG-PET examinations", 2019.

[11] A. Mishra., "Artificial Neural Network (ANN) and regression model for predicting the Albumin to Globulin (A/G) ratio in a serum protein electrophoresis test", 2019.

[12] M. R. N.Mohamed Loey, "A Survey on Blood Image Diseases Detection Using Deep Learning", 2020.

[13] M. Saric, M. Russo, M. Stella and Marjan Sikora, "CNN-based Method for Lung Cancer Detection in Whole Slide Histopathology Images", 2019.

[14] M. M. a. K. E. P. D. THOMAS C. MICHELS, "Madigan Army Medical Center Family Medicine. Multiple Myeloma: Diagnosis and Treatment" ,2017.

[15] M. L.P. b. M. G.R. J. C.J. M. C.-P. d. C. R.P. R.A. Trejo-Ayalaa, "Bone marrow aspiration and biopsy. Technique and considerations", 2015.

[16] R. A. G. K. P. D. T. Meenakshi Guptaa, " Multiple myeloma: the disease and its treatment" ,2013.

[17] L. F. N. K. M. I. G. D. N.C. María-Victoria Mateos, " Bone complications in patients with multiple myeloma in five European countries: a retrospective patient chart review", 2020.

[18] X. S.X. Y. J. R.a. X.D. Jinzhou Wang, "Deep Learning-Based CT Imaging in Diagnosing Myeloma and Its Prognosis Evaluation" ,2021.

[19] R. A. E. S. ,. a. A. S. Afshin Bozorgpour, "Multi-scale Regional Attention Deeplab3+: Multiple Myeloma Plasma Cells Segmentation in Microscopic Images" ,2021

[20] N. J. A. K. S. M. S. C. S. R. S. J. D. H. DEEPIKA KUMAR1, " Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks" ,2020.

[21] G. G. S. N. C. a. F. A. Gopi Battineni, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis" ,2020.

[22] J. Brownlee, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks" ,2019.

[23] V. Y. A. P. D. Hlavcheva, " Application of Convolutional Neural Network for Histopathological Analysis" ,2019.

[24] W. Y. H. S. a. G. G. GU. HAO, " Blind Channel Identification Aided Generalized Automatic Modulation Recognition based on Deep Learning", 2017.

[25] Srikanth Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images" ,2019.

[26] Y. W. B. P. D. L. J. D. Y. Q. H. L. X. W. J. X. Qing Guan, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study", 2019.

[27] M. M. K. S. I. Sumaiya Dabeer, "Cancer diagnosis in histopathological image: CNN based approach", 2019.

[28] G. A. C. R. A. E. A. P. C. Teresa Araújo,"Classification of breast cancer histology images using convolutional neural networks" ,2017.

[29] S. I. Daisuke Komura, "Machine Learning Methods for Histopathological Image Analysis", 2018.

[30] S. S. a. J. Mehta, "Multiple Myeloma" ,2006.

[31] R. N. Keiron Teilo O'Shea, "An Introduction to Convolutional Neural Networks" ,2015.

[32] Srikanth Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images" ,2019.

[33] I. Dates and G. Challenge, " SegPC-2021 - Grand Challenge", 2021.

[34] J. F.M. PaulaRodriguez-Otero BrunoPaiva, " Roadmap to cure multiple myeloma" ,2021.

[35] M. Vyshnav, V. Sowmya, E. Gopalakrishnan, S. V. V.V., V. K. Menon and P. S. K, "Deep Learning Based Approach for Multiple Myeloma Detection" ,2020.

[36] 1. G. T. J. L. Y. Z. H. L. P. C. M. P. ,. A. B. K. S. ,. a. B. H. M. Lina Xu, "Automated Whole-Body Bone Lesion Detection for Multiple Myeloma on 68Ga-Pentixafor PET/CT Imaging Using Deep Learning Methods", 2020.

[37] H. B. L. T. J. L. Z. Y. D M Anisuzzaman, " A Deep Learning Study on Osteosarcoma Detection from Histological Images" ,2020.

[38] H. S. 1. T. H. J. C. C. W. A. L.W. Y. a. H. W. Wei Yan1, "Employment of Artificial Intelligence Based on Routine Laboratory Results for the Early Diagnosis of Multiple Myeloma" ,2021.

[39] Y. Z. X. L. Z. Y. A. L. a. X. Y. Xue Chen, "Diagnosis and staging of multiple myeloma using serum-based laser-induced breakdown spectroscopy combined with machine learning methods", 2021.

[40] J. H. M. P. S. R. M.-A. W. M. W. L. K. M. P.B. M. G. L. Roxane Licandro, "Asymmetric Cascade Networks for Focal Bone Lesion Prediction in Multiple Myeloma", 2019.

[41] C. Zhou, H.P. Chan, L. M. Hadjiiski and Qian Dong, "Deep learning-based risk stratification for treatment management of multiple myeloma with sequential MRI scans", 2021.

[42] P. M. O. S. ,. R. G. R. D. Anubha GuptaID, "PCSeg: Color model driven probabilistic multiphase level set based tool for plasma cell segmentation in multiple myeloma", 2018.

[43] Sanju and . Ashok Kumar, "Classification of Multiple Myeloma Cancer Cells Using Convolutional Neural Networks and Transfer Learning" ,2021.

[44] Jonas Prellberg and Oliver Kramer, "Acute Lymphoblastic Leukemia Classification from Microscopic Images using Convolutional Neural Networks" ,2020.

[45] J. H. B. M. A. W. Matthias Perkonigg, " Detecting Bone Lesions in Multiple Myeloma Patients Using Transfer Learning: First International Workshop, DATRA 2018 and Third International Workshop", 2018.

[46] Wahyudi Setiawan and Fitri Damayanti, " Layers Modification of Convolutional Neural Network for Pneumonia Detection", 2019.

[47] J. W. N. S. S. a. Z. H. Muhammad Mateen, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD" ,2018.

[48] C. E. a. B. N. Zachary C. Lipton, "Thresholding Classifiers to Maximize F1 Score" ,2014.

[49] M. K. Stankovikj Svetlana, " ACUTE COMPLICATIONS IN MULTIPLE MYELOMA" ,2017.

[50] P. B.-C. ,. E. ,. T. B. M. a. G. A. Mohammad N. S. Jahromi, "Privacy-Constrained Biometric System for Non-Cooperative Users" ,2019.

[51] J. H. W. H. a. Y. S. Qingge Ji, "Optimized Deep Convolutional Neural Networks for Identification of Macular Diseases from OpticalCoherence Tomography Images", 2019.

[52] V. R. G. HEREN CHELLAM1, "ANALYSIS OF SENTIMENTAL IMAGES USING DEEP LEARNING APPROACH", 2021.

[53] "Bone Marrow Cell Classification," 2020, accessed 16 December 2021,< https://www.kaggle.com/andrewmvd/bone-marrow-cell-classification/discussion/302679>. [Online].

[54] S. T. P. Supawit Vatathanavaro, "White Blood Cell Classification: A Comparison between VGG-16 and ResNet-50 Models", 2018.

[55] M. Y. A. Cinar, "Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease" ,30 November 2019.

[56] S. R. Department, "Global AI use cases for pharma and healthcare 2020", 2020.

[57] z. mohmad, "Artificial Intelligence Definition, Ethics and Standards" ,2019.

[58] Ismahan, "kaggle," [Online]. Available: https://www.kaggle.com/fanouna/microscopic-bone-marrow-images.

[59] R. Eslick., "Multiple myeloma: from diagnosis to treatment", 2013.