Classification of B-ALL cancer from non-cancerous cells Based on Convolutional Neural Network

Fariha Tasnim Chowdhury
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

Zakia Sultana
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

Zannatul Ferdous Daliya
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

Sayma Akter
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

Md. Tanvir Rouf Shawon
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

Raihan Tanvir
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology

***Abstract*—Early and accurate detection of B-ALL cancer is crucial for better patient outcomes. This research uses images of blood cells to examine the potential of convolutional neural networks (CNNs) for separating B-ALL cancer stages from non-cancerous cells. 3242 pictures from four classes—Benign, Malignant Early Pre-B, Malignant Pre-B, and Malignant Pro-B—were included in the dataset that we used. The data was split into three groups using stratified splitting: 80% training, 10% validation, and 10% testing. The training data was subjected to cell segmentation prior to modeling. DenseNet, ResNet50, AlexNet, EfficientNet-B0, and GoogLeNet are the five different CNN architectures that were assessed. The highest test accuracy of 99.08% was attained by ResNet50 and EfficientNet-B0, closely followed by DenseNet at 98.77%. All three models demonstrated excellent performance and remarkable classification ability, with F1 scores greater than 0.990. GoogLeNet outperformed AlexNet, which achieved a respectable 96.00% accuracy, with an accuracy of 98.46%. The performance of CNNs, specifically ResNet50, EfficientNet-B0, and DenseNet, in classifying B-ALL cancer stages from blood cell images is demonstrated in this study. The strategy's high accuracy and robust performance across several architectures demonstrate its potential to create computer-aided diagnostic tools for accurate and timely identification of B-ALL cancer.**

Keywords: classification, B-ALL cancer, segmentation, blood cell images, and DenseNet121, Resnet50 , Alexnet , EfficientNet , Googlenet .

I. Introduction

B-cell acute lymphoblastic leukemia (B-ALL) is a type of cancer that affects the immature white blood cells (B-cells) in the bone marrow, causing them to multiply rapidly and crowd out the normal blood cells. B-ALL is the most prevalent subtype of acute lymphoblastic leukemia (ALL), making up approximately 85% of cases in children and75% of cases in adults. B-ALL is an illness that can be fatal and needs to be diagnosed and treated right away. But the diagnostic techniques used today are costly, error-prone, and intrusive. To determine the presence and subtype of B-ALL cells, they usually entail taking a sample of blood or bone marrow and analyzing it under a microscope. This is a laborious, subjective procedure that depends on the expertise of the laboratory staff and the hematologists. Furthermore, patients may experience problems like bleeding, infection, and pain as a result of it. A more dependable, effective, and economical method of diagnosing B-ALL is thus required.

Artificial intelligence (AI) has become a potent instrument for the analysis of medical images, with the capacity to improve and automate the diagnosis of a wide range of illnesses. Deep learning, one of the many varieties of artificial intelligence (AI) techniques, is a subfield of machine learning that leverages many layers of artificial neurons to process and learn from vast amounts of data and carry out intricate tasks. Because they can extract features and patterns from the pixel values, deep convolutional neural networks (CNNs) are a particular kind of deep learning models that are particularly well-suited for processing images. Several computer vision domains, including object detection, face recognition, and scene understanding, have seen the successful application of CNNs. Promising outcomes in biomedical image analysis, including histopathology, radiology, and microscopy, have also been demonstrated by CNNs recently.

The main objective of this research is to develop and evaluate a novel deep CNN model for the classification of B-ALL cells from non-cancerous cells based on blood microscopy images. The goal of the suggested model is to precisely differentiate between the benign cells and the three subtypes of B-ALL cells (Early-Pre B, Pro B-ALL, and Pre B-ALL). This is crucial for figuring out the patients' prognosis and course of treatment. Additionally, the proposed model seeks

to address the drawbacks of traditional diagnostic techniques, including their high cost, high variability, and low accuracy. The suggested model has the potential to help patients, clinicians, and the healthcare system by automating and improving the diagnosis of B-ALL through deep learning, particularly in areas with limited access to cutting-edge diagnostic tools and knowledge.

II. Related Work

The QCResNet model, an acute lymphoblastic leukemia (ALL) classification system based on peripheral blood microscopy samples, was innovatively crafted by Qile Fan[1]. Boasting a dataset of 15,135 images, this groundbreaking model achieved an impressive accuracy of 98.9 percent, surpassing several state-of-the-art techniques in the field.

Aditya Y [2] contributed to the landscape with a CNN-based deep learning system designed for the recognition and classification of ALL. Successfully meeting the initial goals, the system effectively identified and classified the disease, leveraging the power of an image dataset. The accomplishment was notable, with the model demonstrating a commendable accuracy level of 81.3 percent, as reported in the research findings.

Umeer Saeed[4] pioneered an auto-mechanized convolutional neural network system dedicated to diagnosing acute lymphoblastic leukemia (”ALL”). Addressing the challenge of overfitting, Saeed implemented various data augmentation techniques to diversify the image dataset. Qualitative analysis, facilitated by the visualization of activation heatmaps, ConvNet filters, and intermediate layers, contributed to a robust diagnosis capability, with the model achieving an impressive 99.61 percent accuracy.

Angelo Genevese[3] devised a machine learning technique for ALL sample classification, integrating deep learning and image processing. This approach hinged on adaptive image processing techniques to enhance image quality before classifier training. Employing shallow CNNs, specifically VAR-PCANet, Genevese fine-tuned the unsharpening algorithm's parameters, consulting deep CNNs for effective blood sample categorization.

Author Chatap, N embarked on an analysis aiming to quantify leukemia cells in blood samples, employing convolutional neural networks (CNNs) to achieve this objective. The methodology involved utilizing CNNs to accurately count visible blood cells in a microsocopic image of blood, leveraging sorting instructions to locate and count the cells. The study contributed valuable insights into the total count of leukemia cells in blood samples.

Amidst discussions and debates surrounding CNN and Nearest Neighbor concepts, the significance of these two methodologies in identifying and counting cancer cells in blood became evident. Leveraging the capabilities of new classifiers, such as CNNs and nearest neighbor classifiers, facilitated the identification of cancer cells with relative ease in blood samples.

III. Dataset and Data Preprocessing

A. Dataset

A dataset containing information about blood cancer cells was obtained from Kaggle and is used in this paper.The dataset contains 3242 images that serve as the input variable.Included in the input variable are a number of classes that pertain to blood cell conditions. These classes include benign cells

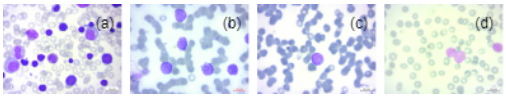

Fig. 1. Class Distribution in the Dataset

[Fig 1(a)]as well as various stages of malignant cells, such as **[malignant] Pro-B[Fig 1(b)]**, **[malignant] Pre-B** [Fig 1(c)], and **[malignant] early Pre-B** [Fig 1(d)]. Quantity of samples in each category: There were 512 benign cases, 979 early pre-B cases, 955 pre-B cases, and 796 pro-B cases of malignant cases.[Fig 1]

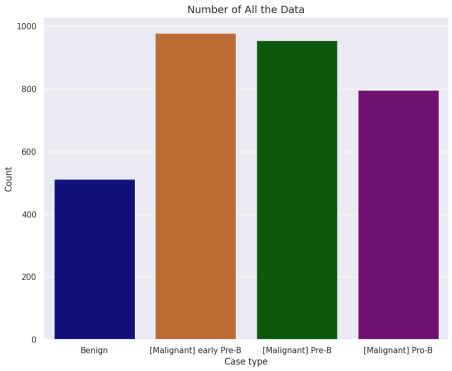

Fig. 2. Dataset

B. Data Preprocessing

Before applying the machine learning algorithms, we perform some data preprocessing steps to clean and prepare the data. The data preprocessing steps are as follows: In this study, our focus is on working with images of blood cells, particularly to train a machine learning model to identify specific cells known as lymphoblasts. The key task at hand is segmentation, which involves isolating the relevant parts of an image while disregarding the rest.

1.At first we normalize the whole dataset.resize it in 224*224 and transform it into tensor

2. Then we split the dataset into training and testing sets, using 80% of the data for training and 10% of the data for testing and 10% for validation test.Train set length: 2593 Validation set length: 324 and Test set length: 325

3. Image Reading and Resizing: The images from the dataset are read and resized to a uniform size of 224x224 pixels.

4. Color Space Conversion: The images are initially in the BGR color space, which is then converted to RGB.

Subsequently, the RGB images are transformed into the LAB color space.

5. K-Means Clustering: The 'a*' channel of the LAB image, which represents the color dimension from green to red, is reshaped and used for K-Means clustering. This unsupervised machine learning algorithm is used to partition the input into K clusters. Each observation belongs to the cluster with the nearest mean.

6. Binary Thresholding: A binary threshold is applied to the clustered image to separate the regions of interest from the background. Pixels with a value greater than the threshold are set to a certain value (maximum value), and those below are set to another (minimum value).

7. Hole Filling and Noise Removal: Binary hole filling is performed to ensure that the regions of interest are solid and continuous. Small objects and holes are removed using morphological operations to reduce noise in the binary image.

8. Masking: The original image and the binary mask are combined using a bitwise 'AND' operation. This operation retains the regions of interest and sets the rest of the image to black.

IV. METHODOLOGY

Our research aims to classify B-ALL cancer cells and non cancerous cell using Convolutional Neural Networks. We employed five different CNN architectures for this task: DenseNet121, Resnet50 , Alexnet , EfficientNet , Googlenet.

DenseNet121:

The DenseNet architecture is known for its efficient use of parameters and its ability to propagate features through its dense blocks. Dense connections are made between layers in this kind of CNN. All layers in DenseNet have direct connections to one another if their feature-map sizes match. Each layer can pass on its own feature-maps to all subsequent layers and receive additional inputs from all previous layers thanks to this architecture. Robust feature reuse, reduced parameter count, and enhanced gradient flow during training are all attributed to this dense connectivity. Effective for image classification tasks, DenseNet architectures, such as DenseNet121, provide efficient parameter sharing.

Resnet50:

ResNet, or Residual Neural Network, is a groundbreaking convolutional neural network architecture designed to address the vanishing gradient problem in deep networks. It utilizes residual blocks with skip connections, allowing information to bypass certain layers during training. This skip connection helps to mitigate the degradation issue that arises when deeper models become more difficult to optimize, which makes it easier to train very deep networks. Remaining blocks in ResNet architectures typically have batch normalization and ReLU activation functions. ResNet's skip connections make it possible to train incredibly deep networks, which improves accuracy across a range of image recognition tasks. Subsequent deep learning architectures have been influenced by its impactful design. Resnet50, a specific variant with 50

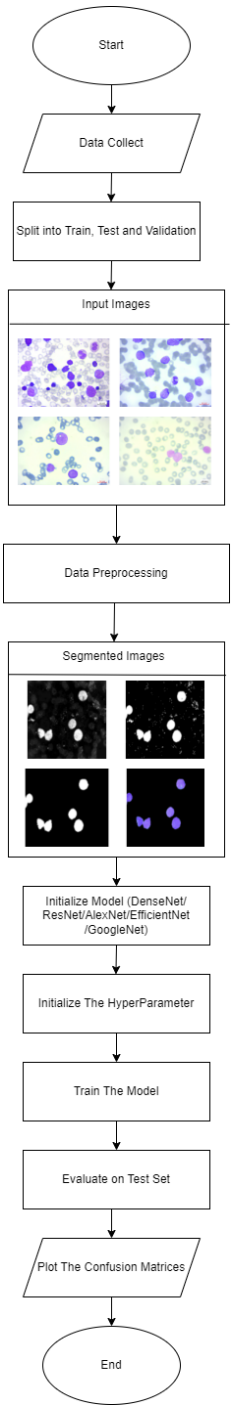

Fig. 3. Methodology Flowchart

layers, addresses the vanishing gradient problem, facilitating the training of deeper networks.

Alexnet:

A groundbreaking convolutional neural network (CNN) architecture is AlexNet. It was instrumental in the widespread adoption of deep learning for image classification applications. Three fully connected layers and five convolutional layers make up AlexNet's eight layers. For feature learning, it integrates dropout and rectified linear unit (ReLU) activations. With its success, deep learning research underwent a sea change and the possibilities of deep neural networks for image recognition were demonstrated. Subsequent deep CNN designs were built upon this architecture.

EfficientNet:

EfficientNet is a state-of-the-art convolutional neural network architecture. It focuses on optimizing model efficiency by simultaneously scaling depth, width, and resolution using a compound scaling method. It achieves a balance between model size and performance by scaling depth, width, and resolution. This approach ensures better performance with fewer parameters, making EfficientNet highly computationally efficient. Based on a baseline network, the architecture scales up to larger variants (B0 to B7) to support a range of model sizes. When it comes to image classification tasks, EfficientNet has achieved elite performance, outperforming many other models of the same era in terms of accuracy and efficiency. Particularly in the quest for ideal resource utilization, its creative design has inspired later neural network architectures.

Googlenet:

GoogleNet is known as Inception, which is a pioneering convolutional neural network (CNN) architecture. It introduced the concept of inception modules, utilizing parallel convolutional filters of different sizes to capture multi-scale features efficiently. GoogleNet's deep structure and computational efficiency were key design elements that increased task accuracy for image classification. The network's original architecture enables a trade-off between computational efficiency and cost, which has an impact on later CNN designs. GoogleNet served as an example of how intricate and varied architectures can be used to achieve high accuracy in deep learning tasks.

All models were compiled and trained on our dataset of B-ALL cancer and non-cancerous cells. The performance of the models was then evaluated and compared to determine the most effective architecture for this classification task.

V. Experimental Results

GoogleNET, RESNET, DENESNET, EFFICIENTNET and ALEXNET are the five models we used for our project.It is **98.7692%** accurate for DENSENET, **96.0000 %** accurate for ALEXNET, and **98.4615%** accurate for GoogleNet. This shows that it works well because RESNET is **99.0769%** accurate and EFFICIENTNET is **99.0769%** accurate. Cutting down on the number of false positives and false negatives helps the DENSENET model get a balanced level of accuracy. You can still use ALEXNET without any problems.With 0.998% of precision, recall, and F1 score, the model EFFICIENTNET is very reliable and useful for the classification task. Its accuracy is 99.07%. Fig[4] A precision and recall score of 98.08% and an F1 score and accuracy of 99.07% show that RESNET does a great job.More accurate results are obtained when the balance between recall and accuracy is just right. Fig[5]

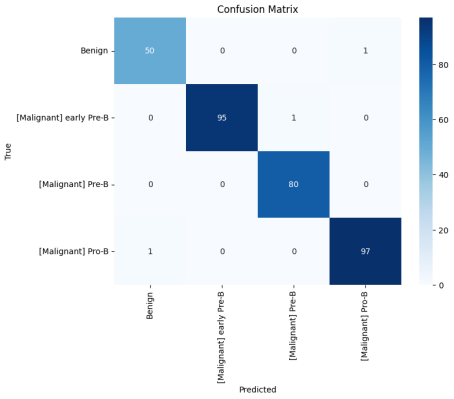

Fig. 4. Test Set Classification Matrix for EfficientNet

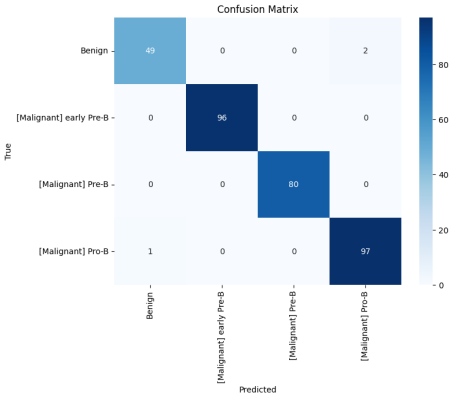

Fig. 5. Test Set Classification Matrix for ResNet.

VI. Result Analysis

A. Performance Evaluation Metrics

In the pursuit of evaluating the performance of neural networks, we employed two fundamental metrics. To assess our model's performance, we used common classification metrics: Accuracy, Recall, Precision, and F1 Score. We generated a confusion matrix from the test set and calculated these metrics using PyTorch and scikit-learn[Fig 6] DENSENET which

MODEL	ACCURACY	F1 SCORE	RECALL	PRECISION
DENSNET	98.7692%	0.9877	0.9877	0.9878
RESNET	99.0769%	0.9907	0.9908	0.9908
ALEXNET	96.0000%	0.9594	0.9600	0.9615
EFFICIENTNET	99.0769%	0.9908	0.9908	0.9908
GOOGLENET	98.4615%	0.9846	0.9846	0.9851

Fig. 6. Model Test Result Comparison

stands for Densely Connected Convolutional Networks, has demonstrated a robust and balanced performance, especially when it comes to precision and recall, showing that it can correctly classify positive instances while also reducing the number of false positives and false negatives.Precision is the model's ability to correctly find positive instances without giving many false positives. Recall, on the other hand, is the model's ability to catch a lot of real positive instances while giving few false negatives. DenseNet is good at handling positive cases with care, as shown by its balanced performance in both precision and recall. This shows that it has a deep understanding of how the data is distributed.

RESNET exhibited exceptional performance, particularly excelling in recall, showcasing its ability to effectively identify positive instances with minimal false negatives. The high precision and accuracy further underline its reliability in accurate image classification.

ALEXNET, while slightly trailing in accuracy compared to other models, maintained competitive precision and recall rates. The F1 score provides a holistic view of its performance, showcasing its ability to balance precision and recall.

EFFICIENTNET emerged as a top-performing model, achieving exceptional precision, recall, and F1 score. Its high accuracy solidifies its effectiveness in accurate image classification across various categories.

GOOGLENET demonstrated a well-balanced performance across precision, recall, and F1 score, maintaining a high level of accuracy. Its ability to accurately classify images is evident, making it a reliable choice for image classification tasks.

B. Evaluation of Model Loss

The second metric, loss, typically influences the modification of the neural network's connection values and is primarily concerned with quantifying the prediction error. The evaluation of model performance was conducted using the test loss metric. The computed test loss values for each model are as follows: DenseNet achieved a test loss of 0.0313, ResNet recorded a test loss of 0.0342, AlexNet exhibited a test loss of 0.0996, EfficientNet demonstrated a test loss of 0.0366, and GoogLeNet also reported a test loss of 0.0366. These values provide insights into the models' generalization capabilities, with lower test loss values indicating better performance in minimizing errors on unseen data.

C. Result Analysis of Hyperparameter Tuning

In our project, we conducted a series of experiments to fine-tune the hyperparameters of our model, aiming to achieve optimal performance. The hyperparameters we focused on were the number of training epochs, the optimizer used, and the learning rate .[Fig 7]

Number of Training Epochs: The number of training epochs is a crucial hyperparameter that determines how many times the learning algorithm will work through the entire training dataset. We experimented with two different settings for this parameter: 5 and 10 epochs. Training for more epochs can allow the model to learn more complex patterns, but there is also a risk of overfitting if the number of epochs is too high.Due to resource constraints and the unavailability of extensive computational capacity for epoch verification, we opted for a limited number of epochs in our training process.

Parameter	Values
Epoch	5, 10
Learning Rate	0.01, 0.001
Optimizer	Stochastic Gradient Descent Adaptive Moment Estimation

Fig. 7. Hyperparameters used for tuning the networks.

Optimizer: The optimizer is responsible for updating the weights of the neural network with the aim of minimizing the loss function. We tested two different optimizers: Adam and Stochastic Gradient Descent (SGD). Adam is known for its efficiency and effective handling of sparse gradients, while SGD is renowned for its robustness and wide usage in the deep learning community.

Learning Rate: The learning rate controls how much to update the model in response to the estimated error each time the model weights are updated. Choosing the right learning rate is essential for both the speed and the final quality of the training. We experimented with two different learning rates: 0.01 and 0.001.

Through a rigorous process of systematic experimentation and thorough validation, we successfully pinpointed the optimal hyperparameter combination that demonstrated superior performance for our specific task. The results obtained from this meticulous hyperparameter tuning procedure played a pivotal role in shaping the final configuration of our model.

Upon comparing our proposed model outcomes with the dataset [Fig 6] we observed that the integration of image segmentation with transfer learning on residual deep neural networks has resulted in improved performance. The inclusion of segmentation proves vital in extracting valuable spatial information from the input dataset by isolating regions of interest. This, in turn, allows Convolutional Neural Networks (CNNs) to concentrate on specific areas, aiding targeted feature extraction. The heightened focus and precision contribute to achieving higher training accuracy.

Our project has successfully developed and trained these models, with the most notable achievements being the stability of training, convergence, and the superior performance of ResNet50,EfficientNet on the test set.

VII. Conclusion

This research successfully unlocked the potential of convolutional neural networks (CNNs) in a vital battle: classifying B-ALL cancer stages directly from blood cell images. Models like ResNet50 and EfficientNet-B0 achieved a remarkable 99.08 percent accuracy, showcasing the immense potential of this approach for automated B-ALL detection. This breakthrough represents a significant leap forward compared to

traditional methods, potentially paving the way for faster, more accessible, and less invasive diagnoses.

Beyond the impressive accuracy, the study further strengthens this promise through two key aspects. Firstly, the use of cell segmentation focused the models on relevant cellular features, potentially contributing to their performance. Secondly, the consistent performance across diverse CNN architectures underscores the generalizability of the approach, suggesting its wider applicability in clinical settings.

However, our journey doesn't end here. While this research lays a strong foundation, further investigation is crucial to unlock the full potential of these models. Optimizing them through techniques like hyperparameter tuning and incorporating larger, more diverse datasets could further enhance their accuracy and real-world effectiveness. Additionally, exploring their performance on similar tasks, such as classifying other leukemia types or cancers, could broaden their impact and accelerate advancements in automated cancer detection.

Ultimately, this research signifies a crucial step towards developing reliable and accurate computer-aided tools for B-ALL cancer diagnosis. Continuous refinement and exploration will pave the way for even more robust and efficient tools, leading to improved diagnoses, better patient outcomes, and potentially saving lives, particularly in regions with limited access to advanced healthcare facilities. This is just the beginning, and future advancements in this field hold immense promise for revolutionizing cancer detection and treatment.

REFERENCES

[1] Q. Fan, B. Qu, J. Teng, A. Hui and T. T. Toe, "Acute lymphoblastic leukemia Classification Based on Convolutional Neural Network," 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2023, pp. 877-883, doi: 10.1109/ICAIBD57115.2023.10206402.

[2] A. Y, D. M, V. S. Naga Sai Teja K, H. Priya J and E. Datta K, "An Improved Leukemia Detection Model using CNN," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 261-264, doi: 10.1109/ICIDCA56705.2023.10099998

[3] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis and F. Scotti, "Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 1205-1209, doi: 10.1109/ICASSP39728.2021.9414362.

[4] Saeed, U., Kumar, K., Khuhro, M.A. et al. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. Multimedia Tools and Application (2023). https://doi.org/10.1007/s11042-023-16191-2

[5] S. Karthikeyan, M. L. Moses, P. Ramya., E. Thrisha., K. Kalarani and R. N. Susheel., "Machine Learning based Algorithmic approach for Detection and Classification of Leukemia," 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2023, pp. 1-6, doi: 10.1109/ICSTSN57873.2023.10151660.

[6] Livieris IE, Pintelas E, Kanavos A, Pintelas P (2018) Identifification of blood cell subtypes from images using an improved SSL algorithm. Biomed J Sci Tech Res. https://doi.org/10.26717/BJSTR.2018.09.001755

[7] Baig R, Rehman A, Almuhaimeed A, Alzahrani A, Rauf HT. Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach. Applied Sciences. 2022; 12(13):6317. https://doi.org/10.3390/app12136317

[8] Antony Vigil, M.S. et al. 'A Combined Neural Network Mechanism for Categorizing the Normal and Cancer Cells'. 1 Jan. 2023 : 7191 – 7203.