International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

DNA Microarray for Cancer Classification Using Deep Learning Based on RESNET-50 Convolutional Neural Network Technique

B. Shyamala Gowri[1, *], S. Anu H. Nair[2], K. P. Sanal Kumar[3], S. Kamalakkannan[4]

Submitted: 15/05/2024 **Revised:** 27/06/2024 **Accepted:** 07/07/2024

Abstract: In today's world, microarray data structures are significant in the diagnosis and classification of various malignant tissues and diseases. Therefore, to effectively address the difficulties of gene expression and classification, it offers high dimensionality and a limited number of genetic samples. Gene expression datasets are frequently utilized in disease prediction and diagnosis, particularly in cancer treatment. Medical diagnostics is one of the most important fields in which image-processing techniques are utilized effectively. Moreover, image processing is vital for enhancing diagnostic and surgical precision. However, accurately characterizing cancer biologically and identifying disease-causing gene expression can be challenging and time-consuming. To address this issue, we introduce the ResNet-50 Convolutional Neural Network (ResNet-50 CNN) method to increase the accuracy of DNA microarray cancer classification. Furthermore, we pre-process the image using the Adaptive Median Filter (AMF) method to remove background noise and enhance a smooth image. Additionally, the pixel-based color contrast, texture, and local contrast of blood cell cancer images are all enhanced using the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique. After that, we employ the Enhanced Canny Edge Detection (ECDD) algorithm to improve edge detection and reduce the error rate in image segmentation. Finally, we propose a ResNet-50 CNN method using a Deep Learning (DL) algorithm to improve the accuracy of classifying DNA microarrays as either cancerous or non-cancerous. Furthermore, the introduced algorithm can be used to classify microarray cancers through performance evaluations such as precision, recall, precision, time complexity, and F1 score. The results of the proposed model indicate an efficiency rating of 96.8% when assessed using various performance measures.

Keywords*: DNA, Microarray, cancer classification, Image processing, ResNet-50 CNN, and Enhanced Canny Contour Detection.*

1. Introduction

Genetic mutations that produce malignant cells in the human body cause cancer. During development, these cells divide on their own and increase throughout the organ, frequently resulting in death. As a result, gene expression analysis sheds light on how specific genes contribute to the onset and spread of cancer. Gene expression variations are employed as indicators for early cancer detection and to pinpoint potential therapeutic targets. Additionally, several strategies make predictive, preventive, and personalized health possible [1]. The process by which the information included in Deoxyribonucleic Acid (DNA) is converted into proteins and other molecules is known as gene expression. Messenger RNA (mRNA) is also created from DNA and subsequently translated into proteins. Gene expression analysis evaluates the set of genetic alterations that take place in tissues or individual cells under particular circumstances.

The organized sequence of genes in gene expression groups determines an organism's characteristics. This organized arrangement is called a genome and is required for protein synthesis [2]. DNA forms this structure in the absence of moisture deficit, dehydration, or high ion concentrations. B-DNA is the stable form of DNA, characterized by ten bases per revolution, and it exists in an environment conducive to life. C-DNA: In the presence of a catalyst, a particular procedure known as reverse transcription is used to create complementary DNA. Hundreds of genes are present in microarray technology structures based on

*1,*Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu, India.*
2Assistant Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, (Deputed To WPT, Chennai), Chennai, India.
3Assistant Professor, Department of CS, RV Government Arts College, Chengalpattu, India
4Associate Professor, Department of Information Technology, School of Computing Sciences, (VISTAS), Chennai, Tamil Nadu, India
*1,*shyamalagowribalaraman@gmail.com, 2anu_jul@yahoo.co.in, 3sanalprabha@yahoo.co.in , 4kannan.scs@velsuniv.ac.in*

this method [3], and the function of diseased and healthy cells is compared through tissues. Therefore, DNA microarray gene technology is beneficial for researchers to identify genes that cause diseases like cancer. Furthermore, genetically linked cancers aid in the identification of subtypes. Therefore, high-throughput classification methods are of great importance to address microarray gene expression problems.

In contrast to the histopathological findings derived from the microscopic examination of the biopsy images, biopsy images are categorized as either cancerous or non-cancerous tissue based on the size, shape, and distribution of cells and nuclei [4]. Gene expression levels are continually assessed in gene expression-based research to describe tissue samples. Microbial technologies modify gene expression based on the pathogen [5]. Furthermore, using gene expression data can accurately predict different tumor types, providing better treatment to patients. However, retrieving features collected from the dataset is expensive and time-consuming.

Cancer diagnosis is challenging due to the high dimensionality and complexity of gene expression data. Despite generations of research, the clinical categorization and identification of tumor-specific characteristics of cancer are still unknown. However, these tasks are challenging due to the massive complexity and noise of these data [6]. There are several difficulties because of the small sample size and the high number of genes. More research is being done to identify genes and increase accuracy because gene expression data are crucial for cancer diagnosis.

The contribution of this section is to present a RESNET-50-CNN method that improves the accuracy of classifying DNA microarrays as cancerous or noncancerous. Additionally, blood cell cancer (ALL) data obtained from the Kaggle dataset can be used to classify cancers based on DNA microarrays. Further, the AMF method can be used to remove noise from the input image and smooth the image. Additionally, the CLAHE technique can be used to forecast the local variation of the image and improve the color-based characteristics of the blood cell cancer image. Similarly, the ECDD algorithm can be implemented during image segmentation to detect the edges of blood cancer images and reduce the error rate. Performance evaluation of cancer classifiers can be improved by using different measures.

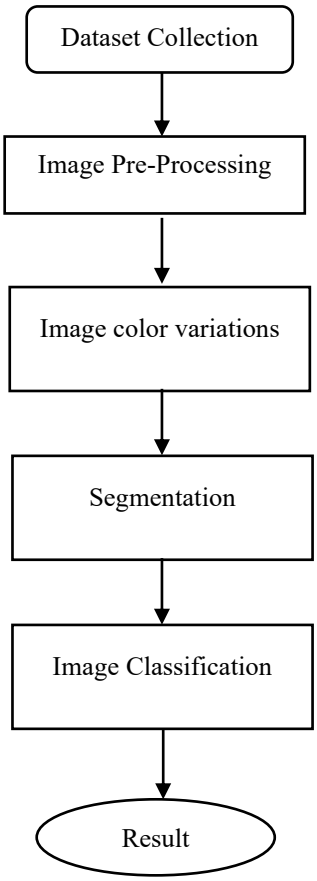


Figure 1. The Basic Architecture Diagram

As shown in Figure 1, blood cell cancer (ALL) data obtained from the Kaggle dataset can be used to classify cancers based on DNA microarray based on the basic architecture diagram. In addition, cancer accuracy can be improved through input images, image pre-processing, color enhancement, image segmentation, and classification.

2. Literature Survey

The author [7] described that the analysis of microarray cancer cell classification is determined by focusing on feature selection methods. In addition, the taxonomy and open analysis of feature selection methods are described in detail based on various techniques to solve the problem of microarray cancer classification. However, achieving cancer classification accuracy when using whole genomes poses significant challenges.

The selection of informative genes is an essential method for predicting the classification accuracy of genome-wide cancer cells, and this process significantly contributes to the classification performance of microarrays [8]. However, genetic selection is required to address the high dimensionality of microarray data caused by noise.

To address the noise issue in microarray data [9], the results of different optimizers were evaluated using the RNA-seq dataset. Current advances are based on DL techniques for assessing optimal feature selection methods in gene expression data analysis.

The novel [10] proposes a Machine Learning (ML) approach for leukemia disease prediction by analyzing feature selection methods. Furthermore, it utilizes Chi-squared (Chi2) feature selection techniques based on the leukemia microarray genomic dataset to overcome imbalances.

The Barnacle Mating Optimizer (BMO) technique, which is optimized by Support Vector Machine (SVM), was utilized to choose prognostic and informative genes for cancer classification based on microarray gene expression. Nonetheless, a significant obstacle in cancer prediction is the identification of genes with limited information [11].

To choose genes for cancer diagnosis, the author described the use of the Genotype-Based Ensemble Classifier System (GECS) to determine genetic subtypes of tumors from the genetic perspective of non-small cell lung cancer. Furthermore, [12] provides a valuable approach to assessing cancer heterogeneity using GECS techniques for molecular pathology definition.

The use of computerized methods to detect cancer cells in five areas, breast, liver, lung, prostate, and kidney, has significantly enhanced accuracy [13]. However, computationally, dimensionality reduction techniques lessen the accuracy of cancer detection.

Additionally, it employs techniques such as ensemble Genetic Bee Colony Optimization (GBCO) and feature selection correlation coefficients to optimize selected genes or characteristics [14]. However, to avoid computational complexity for the entire data set, it is necessary to choose the most critical genes or features.

The novel [15] described a CNN model based on DL methods to classify cancer diseases and their types through gene expression data analysis, select features for the whole dataset, and learn and extract features from the original input data.

The Ebola Optimised Search method (EOSA), a metaheuristic method, was also used to compare the classification results to the suggested model [16]. The CNN framework's hybrid model serves as the foundation for this approach, which aims to enhance breast cancer diagnosis.

Table 1. Microarray for cancer classification Method using deep learning

Author	Year	Methodology	Performance Evaluation	Dataset	Limitation	Accuracy
M S, K.; Rajaguru [17]	2023	SVM, CNN	Mean Square Error (MSE), Accuracy	Lung Harvard 2 Dataset (LH2)	The intrinsic nonlinearity and non-separable noise components of large-scale microarray gene expression make data extraction extremely difficult.	94.47%
Nogueira, A [18]	2023	SVM, Decision Tree (DT)	low false negative rate	DNA microarray datasets	High-dimensional feature vectors provide numerous difficulties for human analysis.	83%
Basavegowda, H. S [19]	2020	Deep Neural Network (DNN)	precision, recall, f-measure	Microarray datasets	Low classification accuracy leads to more relevance problems.	95%
Xie [20]	2023	logistic regression (LR)	Accuracy	Gene Expression Omnibus (GEO) database	Many unrelated and duplicate genes interfere with the correct disease prediction process.	91%
Ke, L [21]	2023	Genetic Algorithm (GA)	Accuracy, precision	DNA microarray dataset	The DNA microarray dataset contains a large number of genes that are not taxonomically significant.	89%
Alromema, N [22]	2023	Gradient Boosting (XGBoost),	Receiver Operator Characteristic (ROC),	Breast_GSE22820 dataset	High-throughput multidimensional microarray	0.976%

			F1-Score		gene expression datasets pose a feature selection problem.	
Razzaque [23]	2024	SVM, K-Nearest Neighbor (KNN)	Precision, Recall	microarray medical datasets	Designing precise and effective classifiers and assessing their effectiveness are crucial tasks in microarray classification.	0.88

Techniques from existing systems improve microarrays for cancer classification, as illustrated in Table 1, using evaluations of performance, datasets, limitations, and accuracy.

The author [24] proposed a linear SVM classification method based on the LUAD dataset, which achieved an improvement of 0.57% on the dataset with a small sample size and number of genes and an improvement of 1.11% on the BRCA dataset.

To identify breast cancer gene expression (BRCA-1), the author presented a portable fluorescent microarray-based imaging device linked to a smartphone. Its foundation is the interaction binding of Cy-3 target DNA and probe DNA [25].

3. Proposed Methodology

To improve the accuracy of DNA microarray cancer classification, we implement the RESNET-50 CNN method, which has shown significant potential in deep learning applications. To enhance the input data for this advanced model, the blood cell cancer (ALL) dataset obtained from Kaggle can be pre-processed; first, the AMF technique should be applied. In addition, it uses AMF methods to improve image quality, remove noise, and enhance smooth images. After the pre-processing step, the CLAHE method effectively predicts local changes in images, significantly in the detection of blood cell cancers. By improving the pixel-based color variant of these images, CLAHE facilitates the visualization of salient features and enables accurate evaluation of clinical conditions. Next, implementing the ECDD algorithm significantly improves edge detection capabilities. It reduces error rates associated with image segmentation, which is a critical factor in accurately classifying cancerous and non-cancerous samples. Finally, these approaches culminated in the deployment of the ResNet-50-CNN architecture, which leverages deep learning algorithms to improve the classification accuracy of DNA microarrays significantly.

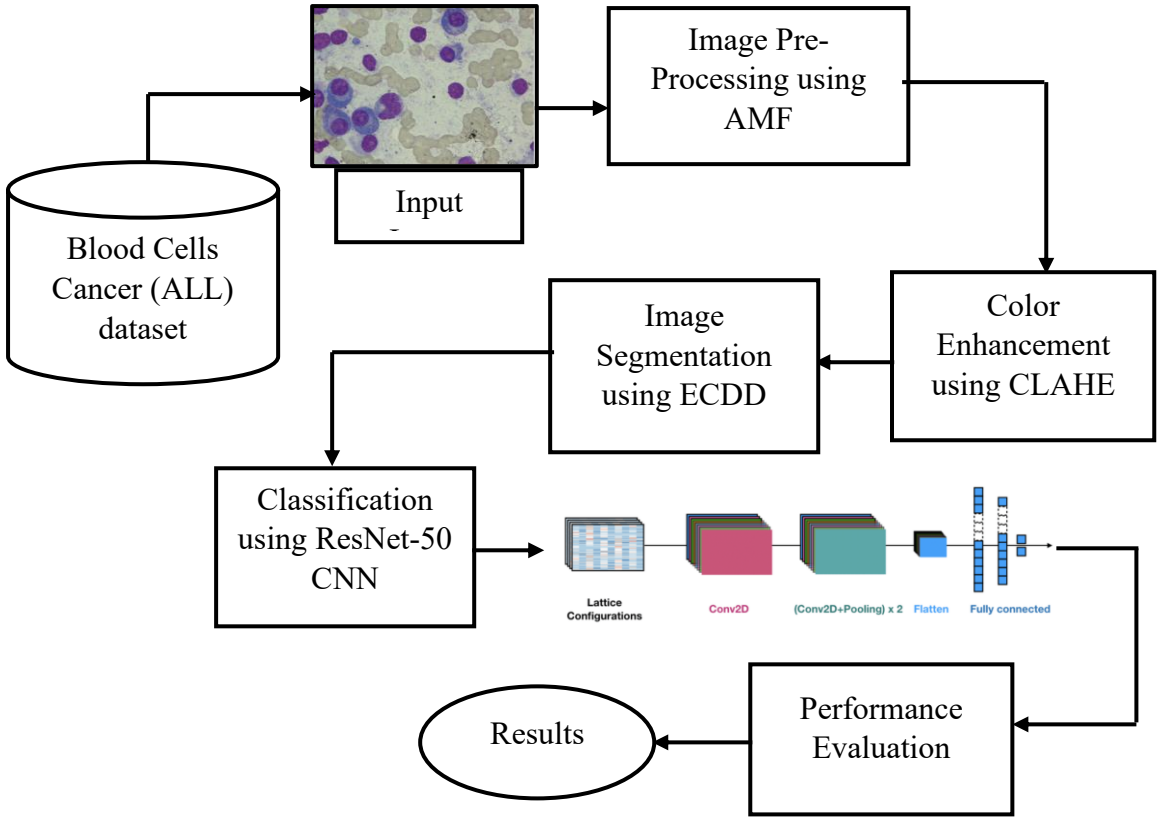
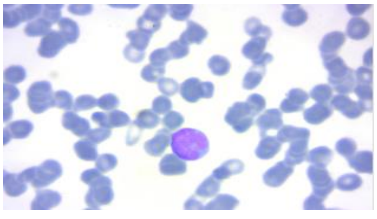


Figure 2. The Proposed ResNet-50 CNN Architecture Diagram

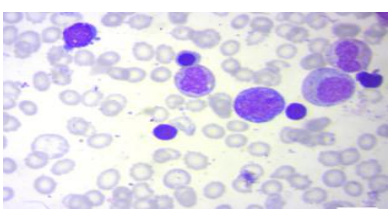
Figure 2 shows the proposed ResNet-50 CNN architecture. It is designed to characterize cancer by analyzing DNA microarray data using deep learning techniques. This architecture uses residual connectivity to simplify the training of deep networks, thereby increasing the model's ability to capture complex patterns in genetic data. By integrating advanced computational techniques, the RESNET-50 framework aims to improve classification accuracy and thereby make a significant contribution to the field of cancer diagnosis.

3.1 Blood Cells Cancer (ALL) Dataset

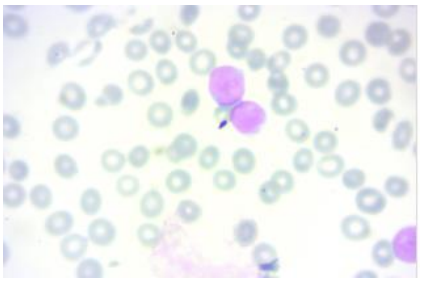
The blood cell cancer ALL dataset consists of 3,242 PBS images from 89 patients and blood samples were prepared for all patients with suspicion. Additionally, the All dataset is accessible through the blood cell cancer website: https://www.kaggle.com/datasets/mohammadamireshraghi/blood-cell-cancer-all-4class. Based on blood cell cancer, the dataset is divided into benign and malignant forms. This dataset has 3,242 photos, of which 2,163 are used for training and 1,079 are utilized for prediction accuracy. The images are categorized as benign, malignant [Pre-B]—955 files, malignant [Pro-B]—796 files, and 979 files. They are employed in testing and measurement.



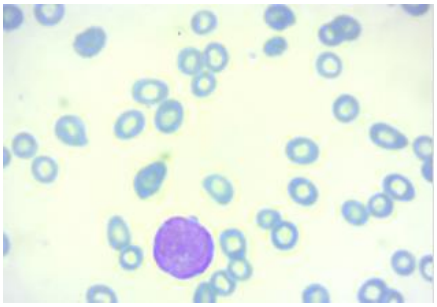
A. Benign



B. Malignant [Pre-B]



C. Malignant [Pro-B]



D. Malignant [Early-B]

Figure 3. Blood Cancer Image Dataset

As illustrated in Figure 3, blood cell cancer image collection can be used to classify microarray cancer into benign and malignant groups by analysing image counts in blood samples.

3.2 Adaptive Median Filter (AMF)

In this section, the AMF method is utilized to remove background noise from blood cancer images during pre-processing, identify a smoother image, and enhance image quality. DNA microarray cancer classification can be improved by introducing the AMF algorithm to smooth the high-frequency interference images and remove background noise in benign and malignant images obtained from the blood cancer dataset. Furthermore, smooth images can be detected from microarray data using nonlinear exponential functions based on the model of robust array data. In addition, it effectively removes noise by replacing a pixel's value with the average of its neighboring values and adapting it to the data's local statistics. The AMF method has been used to reduce noise and enhance smooth images from high-dimensional data from DNA microarrays, enabling accurate cancer classification based on features and edges in gene expression patterns.

$$min\ \underset{a_1 \in z}{arg} \sum_{(m,n)\in\zeta} \left\| a_{(1)} - a_{(m,n)} \right\|_2$$

Use median filters to define the robust performance of different nonlinear vector filters to remove high-frequency noise and smooth blood cancer images. Microarray cancer classification is estimated by minimizing the distance between all other samples in the input vector, as shown in Equation 1. Let's assume $\left\| a_{(1)} - a_{(m,n)} \right\|$ −Euclidean distance, $a_1 \in z$ −lowest ranked vector order, z-minimizing the distance input set. (1)

The proposed AMF algorithm can use the spectral image properties and the minimization image to remove background noise. Instead of reducing the magnitude difference calculated using spatially aligned vector inputs, we reduce the distance metric defined by the color ratio metric. As shown in Equations 2 and 3, each aspect ratio is associated with an integral of the absolute difference defined over the color aspect ratio input. Calculate the set of ordered processing speed measures. Let's assume $\frac{the\ a_{(u,v)1}}{a_{(u,v)2}}$ −ratio quantity, $F_{(u,v)}$ −the color-ratio inputs, and $\frac{a_{(1)1}}{a_{(1)2}}$ −output ratio.

$$F_{(u,v)} = \sum_{(m,n)} \left| \frac{a_{(u,v)1}}{a_{(u,v)2}} - \frac{a_{(m,n)1}}{a_{(m,n)2}} \right| \quad (2)$$

$$\frac{a_{(1)1}}{a_{(1)2}} \le \frac{a_{(2)1}}{a_{(2)2}} \le \cdots \le \frac{a_{(h)1}}{a_{(h)2}} \quad (3)$$

Equation 4 illustrates the estimate of the aspect ratio values of local picture areas by using the outliers that correspond to the maximum cumulative distance to minimize the sum of the absolute differences across color scales.

$$min\ \underset{\left(\frac{x_{(1)1}}{x_{(2)2}}\right)\in w}{arg} \sum_{(g,h)\in\zeta} \left| \frac{x_{(1)1}}{x_{(2)2}} - \frac{x_{(g,h)1}}{x_{(g,h)2}} \right| \quad (4)$$

Operations in the colorimetric domain analyze the structural content of the DNA microarray blood cell cancer image to remove background noise from the high-frequency portion of the colorimetric signal. Equation 5 shows that normalizing the output ratio can smooth the cancer image of blood cells. Let's assume a_2^*, a_1^* −vector, b-output ratio, a^* −nnormalization vector.

$$b = [b_1, b_2] = \left[a_2^* \frac{a_{(1)1}}{a_{(2)2}} -, a_1^* \frac{a_{(1)1}}{a_{(2)2}}\right] \quad (5)$$

Normalizing the output ratio can reduce noise and enhance the image in DNA microarray cancer classification for blood cell cancer.

3.3 Contrast Limited Adaptive Histogram Equalization (CLAHE)

This section utilizes the CLAHE technique to visualize color changes in blood cell images. The CLAHE method can improve clinical image contrast of DNA microarray cancer classification based on improving the quality of blood cell cancer images. This method detects subtle changes in pixel intensity in blood cell cancer images. The intensity of each pixel in the cancer image is translated to nearby pixels based on its assessment. In the color variant analysis, a color histogram presents the color distribution in images of blood cell cancer. The CLAHE method effectively extracts color-based information from blood cancer images and improves the local contrast of these images. By applying the CLAHE method, we can analyze the color distribution across the three RGB channels, which is particularly useful when identifying DNA microarray cancer.

As shown in Equations 6–9, the mean, standard difference, skew, variance, and kurtosis of each channel were used to calculate the local contrast and pixel-based color distribution of the blood cancer images.

$$M(\mu_u) = b \sum_{v=1}^{H} \frac{1}{H} T_{uv} \quad (6)$$

$$S_D(\sigma_u) = \sqrt{\frac{1}{H} \sum_1^H (T_{uv} - \mu_b)^2} \quad (7)$$

$$S(s_i) = \sqrt[3]{\frac{1}{H} \sum_1^H (T_{uv} - \mu_b)^3} \quad (8)$$

$$K(K_i) = \sqrt[4]{\frac{1}{H} \sum_1^H (T_{uv} - \mu_b)^4} \quad (9)$$

Let's assume H-number of images, T_{uv} −pixel value in color channel, u-pixel image, H-pixel count image, $M(\mu_u)$ −average, M-mean, S_D −standard deviation, (σ_u) −color variation distinction, $S(s_i)$ −skewing value, $K(K_i)$ − kurtosis value. The square root of the color distribution difference is used to generate and analyze the pixel value and average measurement of the color channel in a blood cancer image in order to assess the contrast and pixel-based color distribution of the image.

3.4 Enhanced Canny Edge Detection (ECDD)

In this section, we demonstrate the ECDD algorithm's significant performance by minimizing the error rate during image segmentation. Combining DNA microarrays with ECDD algorithms could lead to more accurate cancer prognoses, advances in personalized medicine, and improved patient outcomes. Ganey uses an optimization procedure to identify the edges, and the maximum gradient size of the Gaussian smooth image is roughly suggested as the optimum detection. Edges can be determined through a complex optimization procedure using benign, malignant [pre-B], [pro-B], and [ear-B] images obtained from blood cell cancer datasets. Furthermore, the ECDD method approximates the gradient of Gaussian smooth images for optimal cancer detection in DNA microarrays.

Equation 10 illustrates that the Gaussian template and the original image's neighborhood are weighted. Moreover, edge detection is utilized to analyze and segment the neighborhoods in order to estimate the blood cell image's center point. Let's assume $U(a,b)$ −pixel value original image, $U_x(a,b)$ −pixel value smooth image, G-Gaussian template, $(2+u, 2+v)$ −partial finite difference of neighborhood.

$$U_x(a,b) = T_{uv} \frac{1}{5\times5} \sum_{u=2}^{2} \sum_{v=2}^{2} U\,(a+u)(b+y) \times G(2+u, 2+v) \quad (10)$$

As shown in Equations 11 to 13, calculate the direction and amplitude of the large gradient in the smooth-edged image of the blood cell cancer image. Let's assume $m_b(a,b), m_a(a,b)$ −original image filtered along rows and lines, $\theta -$ Gradient direction, d_a, d_b −blood cancer cell image.

$$G(a,i) = \sqrt{m_a^2(a,b) + m_b^2(a,b)} \quad (11)$$

$$\theta = arctan\left(m_b(a,b)/m_a(a,b)\right) \quad (12)$$

$$d_a = \begin{bmatrix} \frac{-11}{2\ 2} \\ \frac{-11}{2\ 2} \end{bmatrix}, d_b = \left[\frac{1}{2}, \frac{1}{2}\right] \quad (13)$$

Through the development of a smoothed image and the use of the Gaussian template and the original image as a weighted neighborhood, DNA microarray cancer detection can detect and analyze edges in blood cancer cell images in the dataset.

3.5 ResNet-50 Convolutional Neural Network (ResNet-50 CNN)

The proposed RESNET-50 CNN method for classifying DNA microarrays as cancerous or non-cancerous can improve the accuracy of categorizing blood cell cancer images as benign and malignant. A three-layer structure made up of a convolutional layer, a pooling layer, and a fully linked layer is used in the DNA microarray technology, which has been proposed for the categorization of cancer. Furthermore, we automatically extract local characteristics from the input matrix by implementing filters at each convolutional layer. Similarly, we solve the vanishing slope problem by estimating alternate crossings using the ResNet50-

$$\mathrm{f(i)} = \mathrm{L(pi + v)}$$

$$\mathrm{f(i)} = \mathrm{L(i)} + \mathrm{i}$$

The feature map is created by computing a convolution layer over the input data using the weights assigned to the dependencies. The data is then sent into an activation function. Equation 16 also calculates the

$$O_y = L\left(v_y + \sum_{y=1}^{g} w_b u_b\right)$$

The feature computation function is located in the pooling layer, which comes after the convolutional layer. Among the pooling techniques, average pooling and maximal pooling are also examined. DAN classifies cancer for the microarray using the maximum pooling function to calculate the maximum value of the input.

$$s_b = max\left(o_y\right)$$

Convolutional layers can perform block normalization after processing the input time series. ReLU is used to calculate the activation function, as seen in Equation 18-20. Let's assume v_r −bathch

$$S = d * j + u$$

$$v_r = VR_s$$

$$u_l = ReLU(v_r)$$

CNN model. The forward activation map and the maximal activation map of the weights employed in the linear combination of the feature vectors are the slope class of the cumulative average pooling layer. The blood cell cancer image's convolutional layer feature map is enhanced using the ReLU activation function. In order to forecast cancer image signatures and CVD signature maps, pooling layers lower the dimensionality of DNA microarrays. The accuracy can be improved by analysing important blood cancer image features, predicting the dimensions of the output features, and proposing a Resnet50-CNN model in the output layer to classify benign and malignant DNA microarray cancers.

The output layer activation function is described in 14 to 15 as the graph is computed with the blood cancer image to identify weights. Let's assume $the\ s(u)$ −activation function, T-weight, $D(u)$ −output, U-identity map, u-initial input, v-output, and j-bias term.

(14)

(15)

output of the feature map input values for the following layer. Where O_y −output, v_y −bias, i_y −imput vector, P_y −weight vector, s-activation function.

(16)

Using the output of the pooling layer, it is possible to predict the essential features of blood cancer images from fully connected layer class labels. Equation 17 shows the activation function of the SoftMax classifier for cancer DNA microarray, which depends on the input probability distribution.

(17)

normalization, $*$ −convolutional operator, d— convolutional layer, v-input time series, u-bias, S −linear activation unction.

(18)

(19)

(20)

To determine the gradient of the mean pooling weight class, utilize Equation 21. Let's assume m-class, S^v −gradiant of class, s-layer, β_w^v −and weight average pool, r −non-linear normalization.

$$\beta_w^v = \frac{1}{r}\sum_y \sum_x \frac{\delta_s v}{\delta M_{xy}^w} \quad (21)$$

Equation 22 indicates that the weighted average layer can be used to analyze the feature map's target class. ReLU and weights determine the DNA microarray system's feature map performance for cancer classification. Let's assume w-feature map, v-target class, g-map function,

$$g = ReLU(\sum_w \beta_w^v M^w) \quad (22)$$

The DNA microarray blood cell cancer image feature map can be analyzed by examining the ReLU activation function in a convolutional layer. Furthermore, by utilizing the weighted average layer to estimate the target category of the microarray cancer classification feature map, the dimensionality of the output features can be forecast.

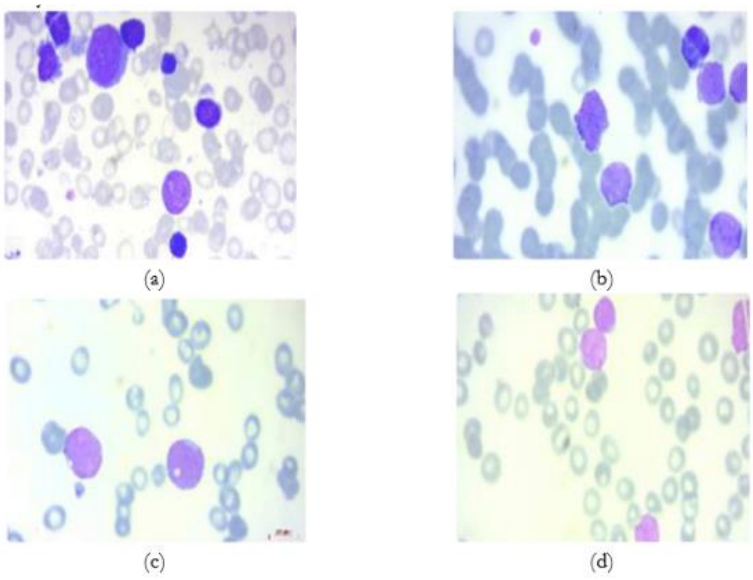


Figure 4. Classifying the images

Figure 4 illustrates that DNA microarray cancer can be predicted using images classified as (a) Benign, (b) early pre-B, (c) pre-B, and (d) pro-B, based on the Blood Cells Cancer (ALL) dataset.

4. Result and Discussion

In this paper, performance indicators such as precision, recall, accuracy, temporal complexity, and F1 score, referred to as the confusion matrix, were used in the context of DL algorithms. It is often used to evaluate and demonstrate that proposed models and algorithms are performing. Data and specifics from a scanned DNA microarray can be used to analyze prognostics and cancer classification. The size of the matrix, where the rows reflect the fundamental categories and the columns correspond to the anticipated categories, depends on the number of categories. Additionally, based on techniques like SVM, DNN, and XGBoost, the suggested ResNet-50CNN method for DNA microarray cancer prediction can increase classification accuracy.

Table 2. Simulation Parameter

Simulation	Value
taset Name	od Cells Cancer (ALL) dataset
of Image	42
ining Image	63
ting Image	79
nguage	hon
ol	yter

sing image processing techniques and simulation parameters using the ResNet-50 CNN approach, the Blood Cells Cancer ALL dataset may be used to classify blood cell cancer images into benign and malignant DNA microarrays for cancer classification. Table 2 shows how the Jupyter Notebook and Python language can be used to evaluate the effectiveness of this classification method.

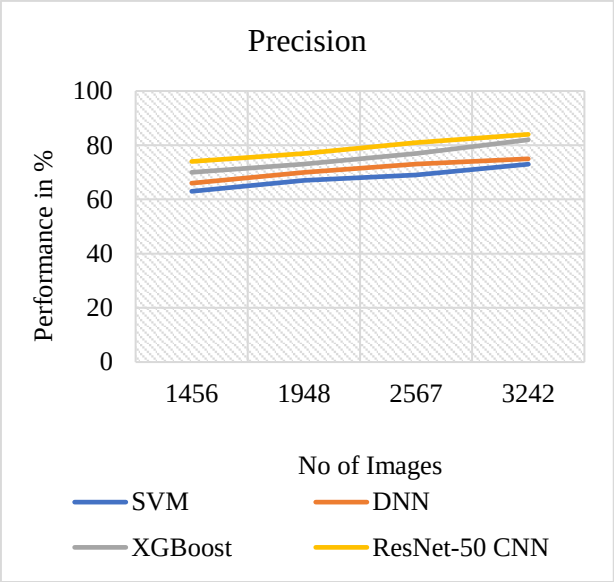


Figure 5. Analysis of Precision

Figure 5 shows the prediction precision achievable by analyzing the increasing number of blood samples used for training and testing using the blood cancer image dataset. Analytical techniques such as SVM, DNN, and XGBoost derived from the existing previous method increase the accuracy range to 73%, 75%, and 82%, respectively. Furthermore, the proposed ResNet-50 CNN method improved the precision of DNA microarray blood cell cancer classification by 84% over the previous process that classified only benign and malignant types.

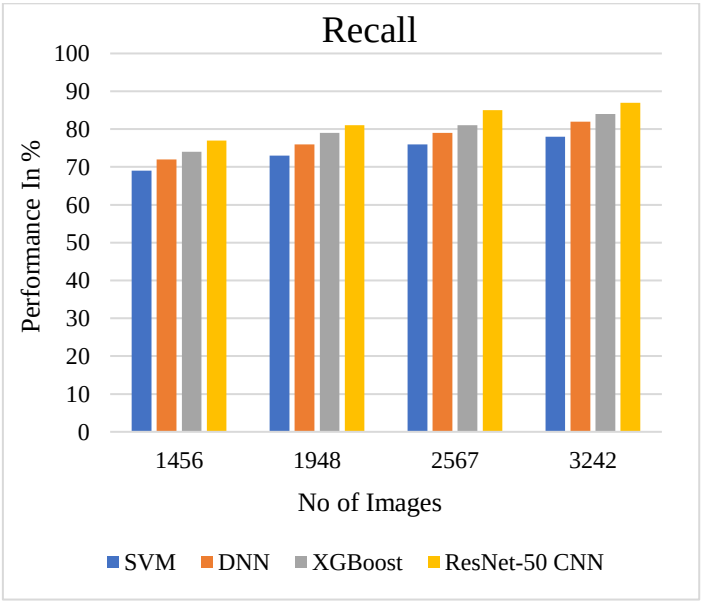


Figure 6. Analysis of Recall

Figure 6 illustrates the prediction recall achievable through the analysis of the cumulative number of blood samples used for training and testing with the blood cancer image dataset. Various analytical techniques, including SVM, DNN, and XGBoost, enhanced the recall rates to 78%, 82%, and 84%, respectively. Additionally, the proposed ResNet-50 CNN method improved the accuracy of DNA microarray blood cell

cancer detection by 87% compared to earlier methods that classified benign and malignant types.

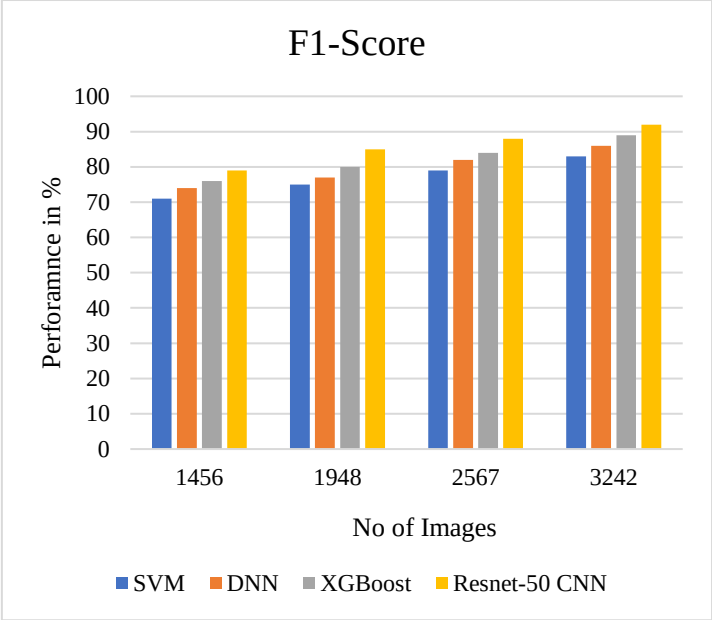


Figure 7. Analysis of F1-Score

Figure 7 displays the F1-score prediction that can be created by analyzing the total number of blood samples used for training and testing using the blood cells cancer image dataset. SVM, DNN, and XGBoost were among the analytical methods that improved the F1-Score rates to 83%, 86%, and 89%, respectively. Furthermore, compared to previous techniques that distinguished between benign and malignant kinds, the suggested ResNet-50 CNN method increased the accuracy of DNA microarray blood cell cancer detection by 92%.

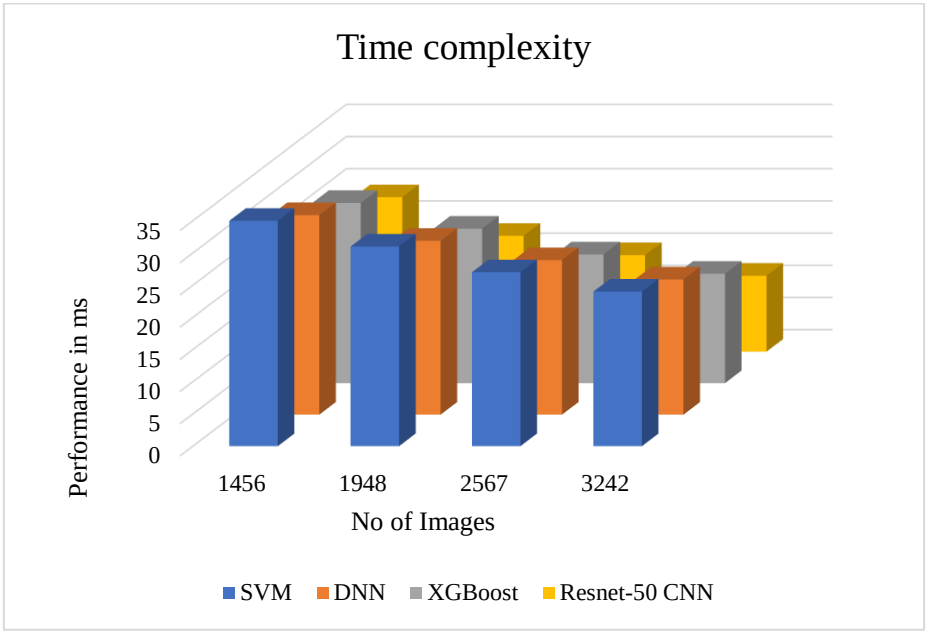


Figure 8. Analysis of Time Complexity

Figure 8 shows the time complexity predictions derived from assessing the total number of blood samples used for training and testing using the blood cells cancer imaging dataset. Analytical methods such as SVM, DNN, and XGBoost show the time complexity by 24ms, 21ms, and 17ms, respectively. Furthermore, the proposed ResNet-50 CNN method has indicated the accuracy of DNA microarray blood cell cancer detection by 11.8ms compared to previous techniques for distinguishing between benign and malignant types.

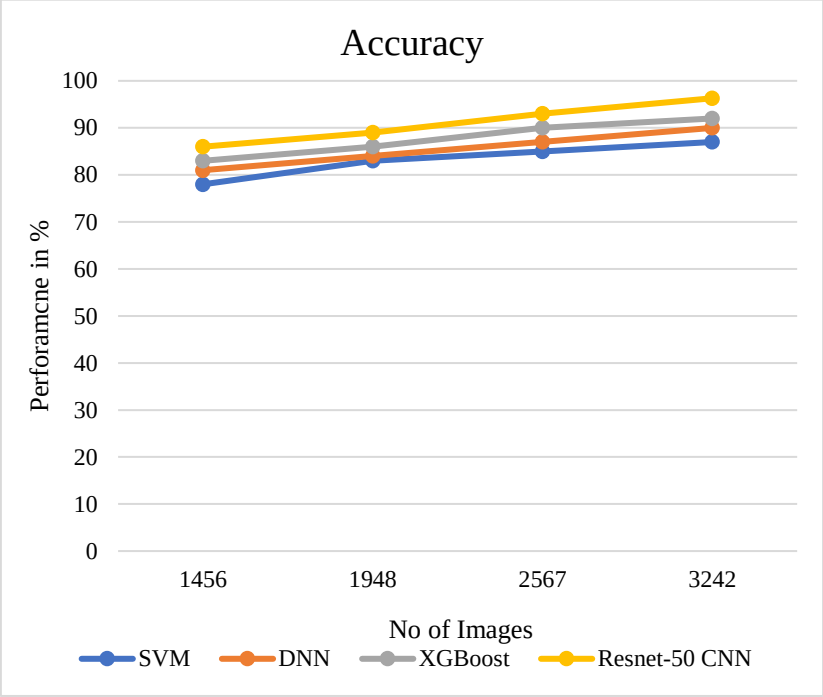


Figure 9. Analysis of Accuracy

As shown in Figure 9, the suggested approach can increase the accuracy of DNA microarray cancer classification when compared to alternative techniques. It is evaluated and trained using the blood cancer picture dataset. Furthermore, compared to earlier techniques like SVM, DNN, and XGBoost, the accuracy is demonstrated to be 87%, 90%, and 92%. Similarly, the suggested strategy outperforms the prior methods with a 96.3% accuracy rate in DNA microarray cancer categorization.

5. Conclusion

In this conclusion, we utilized the proposed ResNet-50 CNN method to enhance the accuracy of DNA microarray analysis for cancer classification through DL. We focused on identifying cancer classifications in microarray data using the Blood Cells Cancer dataset, particularly for Acute Lymphoblastic Leukemia (ALL), through image processing techniques. Additionally, we developed a model for cancer classification using image processing techniques and the blood cell cancer (ALL) dataset and assessed its accuracy, recall, precision, time complexity, and F1 score. The proposed method was implemented for both testing and training with the blood cancer image dataset. The findings indicate that this method significantly improves the accuracy of DNA microarray cancer classification when compared to other techniques, including SVM, DNN, and XGBoost methods. Ultimately, the proposed method achieved an impressive performance rating, attaining an accuracy of 96.3% in DNA microarray cancer classification.

Reference

[1] Alharbi, F.; Vakanski, A. Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. Bioengineering 2023, 10, 173. https://doi.org/10.3390/bioengineering10020173.

[2] Thakur, Tanima, et al. "[Retracted] Gene Expression-Assisted Cancer Prediction Techniques." Journal of Healthcare Engineering 2021.1 (2021): 4242646.

[3] Haznedar, Bulent, Mustafa Turan Arslan, and Adem Kalinli. "Optimizing ANFIS using simulated annealing algorithm for classification of microarray gene expression cancer data." Medical & Biological Engineering & Computing 59 (2021): 497-509.

[4] Devi, T. G., Patil, N., Rai, S., & Philipose, C. S. (2023). Gaussian Blurring Technique for Detecting and Classifying Acute Lymphoblastic Leukemia Cancer Cells from Microscopic Biopsy Images. Life, 13(2), 348. https://doi.org/10.3390/life13020348.

[5] Danaee P, Ghaeini R, Hendrix DA. A Deep Learning Approach for Cancer Detection and

Relevant Gene Identification. Pac Symp Biocomput. 2017; 22:219-229. doi: 10.1142/9789813207813_0022. PMID: 27896977; PMCID: PMC5177447.

[6] Begum, S., Sarkar, R., Chakraborty, D., Sen, S., & Maulik, U. (2021). Application of active learning in DNA microarray data for cancerous gene identification. Expert Systems with Applications, 177, 114914. https://doi.org/10.1016/j.eswa.2021.114914

[7] Hambali, M. A., Oladele, T. O., & Adewole, K. S. (2020). Microarray cancer feature selection: Review, challenges, and research directions. International Journal of Cognitive Computing in Engineering, 1, 78-97. https://doi.org/10.1016/j.ijcce.2020.11.001.

[8] H. Almazrua and H. Alshamlan, "A Comprehensive Survey of Recent Hybrid Feature Selection Methods in Cancer Microarray Gene Expression Data," in IEEE Access, vol. 10, pp. 71427-71449, 2022, doi: 10.1109/ACCESS.2022.3185226.

[9] Gupta, S., Gupta, M. K., Shabaz, M., & Sharma, A. (2022). Deep learning techniques for cancer classification using microarray gene expression data. Frontiers in Physiology, 13, 952709. https://doi.org/10.3389/fphys.2022.952709

[10] Rupapara, V., Rustam, F., Aljedaani, W., Shahzad, H. F., Lee, E., & Ashraf, I. (2022). Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. Scientific Reports, 12(1), 1-15. https://doi.org/10.1038/s41598-022-04835-6

[11] E. H. Houssein, D. S. Abdelminaam, H. N. Hassan, M. M. Al-Sayed and E. Nabil, "A Hybrid Barnacles Mating Optimizer Algorithm with Support Vector Machines for Gene Selection of Microarray Cancer Classification," in IEEE Access, vol. 9, pp. 64895-64905, 2021, doi: 10.1109/ACCESS.2021.3075942.

[12] Y. Ren, Z. -Y. Yang, H. Zhang, Y. Liang, H. -H. Huang and H. Chai, "A Genotype-Based Ensemble Classifier System for Non-Small-Cell Lung Cancer," in IEEE Access, vol. 8, pp. 128509-128518, 2020, doi: 10.1109/ACCESS.2020.3008750.

[13] N. Bharanidharan et al., "Local Entropy Based Remora Optimization and Sparse Autoencoders for Cancer Diagnosis Through Microarray Gene Expression Analysis," in IEEE Access, vol. 12, pp. 39285-39299, 2024, doi: 10.1109/ACCESS.2024.3371887.

[14] S. K. Prabhakar and S. -W. Lee, "An Integrated Approach for Ovarian Cancer Classification with the Application of Stochastic Optimization," in IEEE Access, vol. 8, pp. 127866-127882, 2020, doi: 10.1109/ACCESS.2020.3006154.

[15] Ravindran U, Gunavathi C. A survey on gene expression data analysis using deep learning methods for cancer diagnosis. Prog Biophys Mol Biol. 2023 Jan; 177:1-13. doi: 10.1016/j.pbiomolbio.2022.08.004. Epub 2022 Aug 19. PMID: 35988771.

[16] Mohamed, T. I., Ezugwu, A. E., Vincent, J., Ikotun, A. M., & Mohammed, M. (2023). A bio-inspired convolution neural network architecture for automatic breast cancer detection and classification using RNA-Seq gene expression data. Scientific Reports, 13(1), 1-19. https://doi.org/10.1038/s41598-023-41731-z.

[17] M S, K.; Rajaguru, H.; Nair, A.R. Evaluation and Exploration of Machine Learning and Convolutional Neural Network Classifiers in Detection of Lung Cancer from Microarray Gene—A Paradigm Shift. Bioengineering 2023, 10, 933. https://doi.org/10.3390/bioengineering10080933

[18] Nogueira, A.; Ferreira, A.; Figueiredo, M. A Machine Learning Pipeline for Cancer Detection on Microarray Data: The Role of Feature Discretization and Feature Selection. BioMedInformatics 2023, 3, 585-604. https://doi.org/10.3390/biomedinformatics3030040

[19] Basavegowda, H. S., & Dagnew, G. (2020). Deep learning approach for microarray cancer data classification. CAAI Transactions on Intelligence Technology, 5(1), 22-33. https://doi.org/10.1049/trit.2019.0028

[20] Xie, Weidong, et al. "Improved multi-layer binary firefly algorithm for optimizing feature selection and classification of microarray data." Biomedical Signal Processing and Control 79 (2023): 104080.

[21] Ke, L., Li, M., Wang, L. et al. Improved swarm-optimization-based filter-wrapper gene selection from microarray data for gene expression tumor classification. Pattern Anal

Applic 26, 455–472 (2023). https://doi.org/10.1007/s10044-022-01117-9

[22] Alromema, N.; Syed, A.H.; Khan, T. A Hybrid Machine Learning Approach to Screen Optimal Predictors for the Classification of Primary Breast Tumors from Gene Expression Microarray Data. Diagnostics 2023, 13, 708. https://doi.org/10.3390/diagnostics13040708

[23] Razzaque, Abdul, and Abhishek Badholia. "PCA based feature extraction and MPSO based feature selection for gene expression microarray medical data classification." Measurement: Sensors 31 (2024): 100945.

[24] Hajieskandar, AliReza, et al. "Molecular cancer classification method on microarrays gene expression data using hybrid deep neural network and grey wolf algorithm." Journal of ambient intelligence and humanized computing (2023): 1-11.

[25] Alisha Prasad, Syed Mohammad Abid Hasan, "DNA microarray analysis using a smartphone to detect the BRCA-1 gene", DOI: https://doi.org/10.1039/C8AN01020J, Issue 1, 2019.