

Original Article

# KNN Model for Cancer Prediction Using Stem Cells

P. Mahalakshmi<sup>1</sup>, P. Sangeetha<sup>2</sup>, G. Harikrishnaveni<sup>3</sup>, C. Dhivya<sup>4</sup>

<sup>1,2,3,4</sup>Department of CSE & P. A. College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

Received Date: 28 May 2023

Revised Date: 30 May 2023

Accepted Date: 07 June 2023

**Abstract:** To develop a method for detecting and identifying the type of cancer using the k-nearest neighbor (KNN) algorithm. The KNN algorithm is a simple machine learning algorithm that classifies [1]-[3] an input sample based on the majority of its nearest neighbor in the training data. The KNN algorithm is trained on a large dataset of cancer patients and their corresponding diagnosis, and is then used to predict the type of cancer for a new patient based on their clinical and laboratory test results. The results of this study show that the KNN algorithm is able [4]- [13] to achieve high accuracy in detecting and identifying the type of cancer, making it a promising tool for supporting clinical decision-making in the early stages of cancer diagnosis.

**Keywords:** Classifies, Training Data, Diagnosis, Detecting and Identifying Cancer, KNN.

## I. INTRODUCTION

Cancer leads to the fourth highest number of deaths in female cancers, carrying high risks of morbidity and mortality. Nevertheless, the cancer is slow growing, so its progression through precancerous changes [2]- [5] provides opportunities for prevention, early detection, and treatment. The main challenge of cancer elimination centers on the low- and middle-income countries (LMIC s) where gender discrimination and extreme poverty severely limit a woman's choice to, to seek care resulting in over 88% of deaths from cancer. Prostrate cancer leads to the fourth highest number of deaths in male cancers, carrying high risks of morbidity and mortality. Nevertheless, the prostrate cancer is slow growing, so its progression through [5]- [10] precancerous changes provides opportunities for prevention, early detection, and treatment. The main challenge of prostrate cancer elimination centers on the low- and middle-income countries (LMIC s) where gender discrimination and [8]-[12] extreme poverty severely limit a woman's choice to seek care resulting in over 88% of deaths from Prostrate cancer.

## II. RELATED WORK

Pre-processing is the term for operations on images at the bottom level of abstraction. These operations do not increase image information content but they shrinkage it if entropy is an information measure. These iconic biomedical images are of the same kind as the unique data captured by the sensor, with an concentration of biomedical image usually represented by a matrix [7]-[9]of image function values. The resolve of pre-processing is to raise the image's quality so that it can be analyze and it more efficiently. Preprocessing allows us to abolish undesirable distortions and improve specific qualities that are essential for the application can be working on. Those characteristics could change depending on the application. An image must be preprocessed in order for software to function correctly and produce the correct results. Image processing methods have unlocked the opportunity to extract quantitative information from con crucial microscopy images of biological samples, dramatically growing the range of questions that can be addressed [2]-[9]experimentally in biology. Biologists intention to understand the cells behave and they genes do to build a normal animal, and they goes erroneous in disease or upon injury. It improves accuracy and reliability. Preprocessing data removes missing or unpredictable data values [21] resulting from human or computer error, which can increase the accuracy and quality of a dataset, making it more reliable.

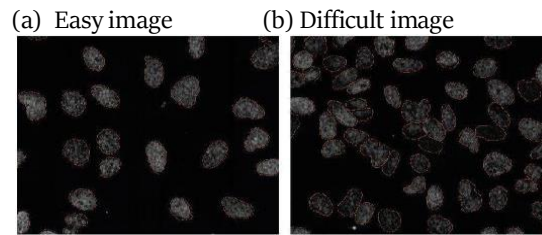
### A. Cell Segmentation

The identified background label along with the segmented nuclei issued in the seeded machine learning segmentation of the cell marker image. The approach allows for the identification and separation [17] of cells. For each nucleus the approach will identify a corresponding cell. Automatic and reliable characterization of cells in cell cultures is key to numerous applications such as cancer research and drug discovery.

The exact number of viable cells in a given cell suspension is determined by cell counting. This is done to determine the exact number of viable cells in a given cell suspension. Cell quantification is necessary to set up FACS analysis, determine correct bone marrow volume, and analyze differences between mice. A hematology analyzer is a blood test device that

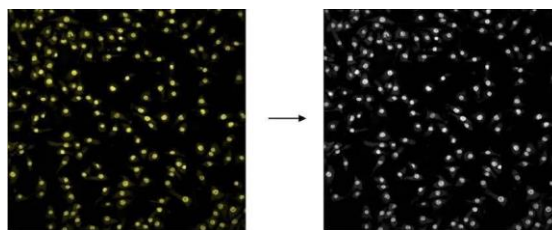


creates a complete [1] blood counter hemogram. Perform quantitative analysis of blood components such as red blood cells, white blood cells (white blood cells), and platelets (platelets).



**Figure 1: Cell Segmentation Cell counting**

Cellular quantification is required to set up FACS analysis, determine correct bone marrow volume, and analyze differences between mice. A hematology analyzer is a blood test device that creates a complete blood counter hemogram. Perform quantitative analysis of blood components such as red blood cells, white blood cells (white blood cells), and platelets (platelets).



**Figure 2: Cell Counting**

The figure 2 shows the counting the number of cells per unit of volume

### III. RESEARCH METHODOLOGY

The cancer detection and type identification using the KNN algorithm is a machine learning-based approach that leverages the poor of the k-nearest neighbor algorithm to accurately classify the type of cancer in patients.

#### A. Post Methodology

This approach has also attracted attention in various stem cell research activities, especially stem cell characterization and differentiation. Adopting a similar computational approach to study and assess the biosafety and bio efficacy risks of stem cells in clinical use is the next advance. Tumorigenicity is one of the major problems in stem cell therapy. The factors that value for k is chosen using cross-validation, and the trained model is tested on another set of test data. Prediction accuracy is evaluated using metrics such as accuracy, precision, recall, and F1 score. The system is simple, fast, effective, scalable, and adaptable, making it a valuable tool for early clinical decision-making in cancer diagnosis and treatment planning.

#### B. Image Segmentation

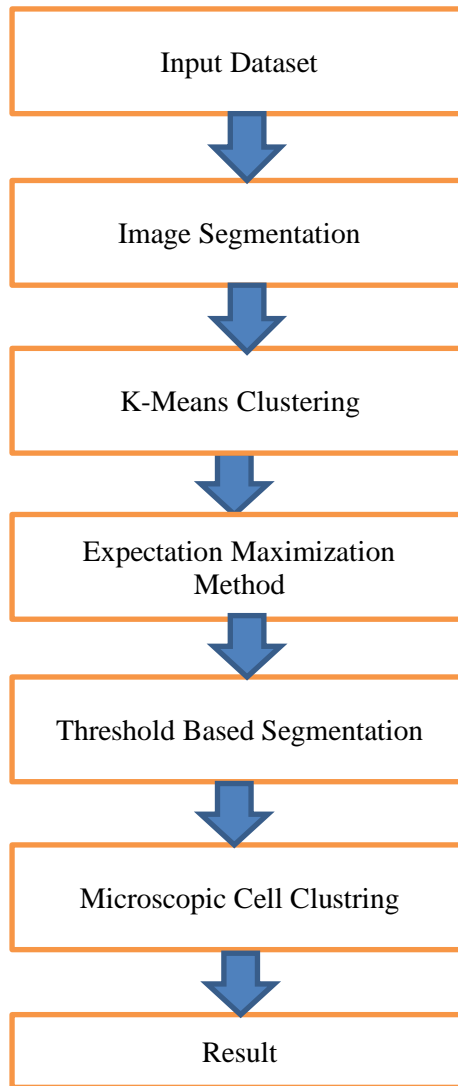
Influence stem cell tumor development is numerous and can be difficult to assess under conventional laboratory conditions. Due to the possible multifactorial etiology of tumorigenesis, defining a unified strategy to test for this risk in stem cells is not feasible and may compromise the safety and efficacy of stem cells in therapy have been achieved effectively.

#### C. Proposed Methodology

The proposed ANN algorithm-based cancer detection and type identification system is a machine learning-based approach that utilizes the k-nearest neighbor algorithm to accurately classify a patient's cancer type. The system begins by collecting large datasets of cancer patients, including demographic information, medical history, laboratory test results, and diagnoses. The collected data are processed again to remove missing or inconsistent values and standardize the input characteristics. The K algorithm is trained on each processed dataset using diagnostics as target labels. The best Although cell segmentation plays an important role in biological imaging studies, it still remains a challenge due to the complexity of many imaging scenes. Our approach utilizes a novel immersion simulation-based self-assembly (ISSO) transform, [14] [20] automated method. For image segmentation. Consider an image  $I$  of size  $r=m*n$  pixels. Each pixel can have  $L$  possible grayscale values in the range  $[0, L- 1]$ .  $h(x)$  is the normalized histogram of the image.

#### D. K-Means Clustering

Signal processing aimed at dividing the observations into k clusters is derived by k-means clustering. Each observation belongs to a cluster, with the nearest mean serving as the prototype of the cluster. This divides the data space into cells. K-means clustering minimizes within-cluster variance [16], but it does not minimize the usual Euclidean distance. This makes the Ber problem even more difficult. The mean optimizes the squared error and the geometric median minimizes the [12] Euclidean distance. For example, k- median and k-medoid can be used to find a better Euclidean solution. When using k-means clustering for image segmentation, the optimal threshold is set such that the image features of the pixels on either side of the threshold are closer to the mean of the features than the distance between them. Find the mean and threshold.



**Figure 3: Data Flow Diagram**

The figure 3 shows that dataflow diagram for module description

#### E. Expectation Maximization Method

Expectation-maximization (EM) algorithms assume that an image consists of a set of grayscale regions and can be described by a parametric data model. The histogram of gray levels is then considered as an estimate of the probability density function [19] and the parameters of the function can be estimated for each gray level range using the histogram.

#### F. Threshold-Based Segmentation

Threshold segmentation is a technique that divides an image into a number of meaningful regions based on a chosen threshold. If the image is a gray image, the threshold is an integer in the range [10] to [0, L-1], where L-1 is the maximum intensity value. Usually this method is used to divide the image into her two regions. Object with background [8] and threshold.

### G. Global Minimization of the ActiveContour Model (GMAC)

Select Global Minimization of Active Contour Model (GMAC) to analyze the implementation of active contours in cell image segmentation. This method is characterized by easy initialization and fast computation, avoiding deadlocks at unnecessary local minima. GMAC [15] is based on the function of Mumford and Shah (MS) and the active contour without edges (AC) model of Chan and Vase. GMAC improves AC by using a dual formulation of illuminated global variations and TV shapes, preserving the benefits of AC.

### H. Microscopic Cell Clustering Using K-means Methodologies

Detection of white blood cells (WBC) is the first step in diagnosing certain diseases. Conditions such as acquired immunodeficiency syndrome prostrate, and other blood- related disorders are usually examined under a light microscope by a pathologist. This process [18] is time- consuming, very time-consuming and expensive, and requires experienced professionals in the field. This is why computer-aided diagnostic systems that assist pathologists in the diagnostic process are so effective. The main goal of this study is to segment leukocytes from microscopic images. For our purpose to present the three-step combination of thresholding, K-means clustering, and a modified machine learning algorithm involving the segmentation of leukocytes from microscopic images.

- Extraction of nuclei from cell images
- Separation of overlapping cells and nuclei.

Evaluation results of the method proposed by [6] show that the similarity, accuracy and sensitivity of core segmentation are 92.07, 96.07 and 94.30%, respectively.

## IV. IMPLEMENTATION AND RESULT

The designed application is a machine learning approach to cancer prediction using an ANN application with cancer prediction capabilities. This system was developed in Mat lab. The application is divided into five main modules. The first module covers image segmentation and mainly focuses on the analysis of dividing an image into multiple parts or regions. The second module, k-means clustering, divides the n observations into k-clusters. The main purpose of the Expectation Maximization (EM) module is to estimate the maximum likelihood [13] of the parameters in a function. The fourth model uses threshold segmentation to split the image into her two sections: background and threshold. The final module of microscopic cell clustering using K-means clustering includes (1) segmentation of leukocytes from microscopic images [10] (2) extraction of nuclei from cell images (3) separation of overlapping cells and nuclei. includes three steps (1) segmentation of WBC [10] from microscopic image (2) extraction of nuclei from cell’s image (3) separation of overlapping cells and nuclei.

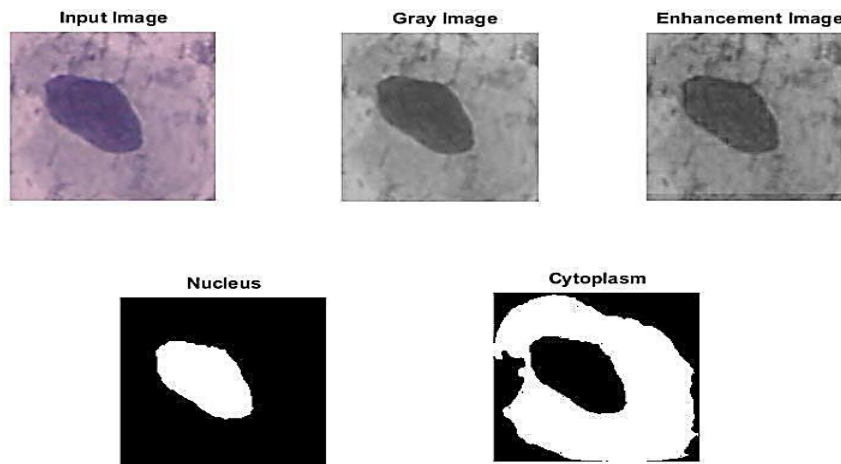
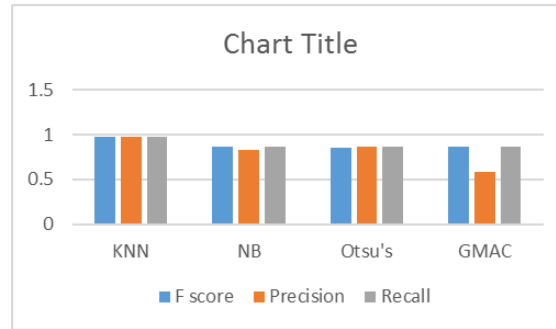


Figure 4: Shows the output of each module

Table 1: Average Measures of the Segmentation Methods Applied On High Quality Synthetic Cell Images

	F score	Precision	Recall
KNN	0.9745	0.9756	0.9765
NB	0.865	0.8267	0.865
Otsu’s	0.856	0.865	0.865
GMAC	0.865	0.586	0.865



## V. CONCLUSION

Machine learning is playing a key role in saving lives by enabling doctors to quickly identify diseases and suggest possible treatments. Cancer is any of a variety of diseases characterized by abnormal cell development and destruction of normal body tissues. New K-means and EM methods have been developed for cell segmentation in fluorescence microscopy images. This approach yielded satisfactory results [4]. This method is suitable for cell isolation and can make cell-by-cell characterization suitable for complex studies Analysis of virus infection. Extracted cells from the background using a machine learning algorithm. The first segmented [5] image was the input of his two-step algorithm for machine learning methods. Applies a split and join process based on machine learning transforms to correctly separate cells. The segmentation process identifies clustered cells based on cell fitness characteristics such as area and intensity, and computes a distance transform [9] to apply machine learning. The combining process uses area and eccentricity to identify over segmented regions and morphological operations to remove subdivision. employs morphological operations to eliminate the divisions.

## VI. REFERENCES

- [1] Jothi and V. M. A. Rajam, "A survey on automated cancer diagnosis from histopathology images," *Artif. Intell. Rev.*, vol. 48, no. 1, pp. 31– 81, Jun. 2020
- [2] Taha, J. Dias, and N. rghi, "Classification of prostrate -cancer using pap-smear images: A convolutional neural network approach," in *Proc. Annu. Conf. Med. Image Understand. Anal. Berlin, Germany: Springer, 2020*, pp. 261–272.
- [3] Zhang, S. Qi, P. Monkam, C. Li, F. Yang, Y.-D. Yao, and W. Qian, "Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images," *IEEE Access*, vol. 7, pp. 110358–110371, 2021
- [4] Li, D. Xue, X. Zhou, J. Zhang, H. Zhang, Y. Yao, F. Kong, L. Zhang, and H. Sun, "Transfer learning based classification of prostrate cancer immunohistochemistry images," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med. (ISICDM), 2021*, pp. 102–106.
- [5] C. Li, H. Chen, L. Zhang, N. Xu, D. Xue, Z. Hu, H. Ma, and H. Sun, "Prostrate histopathology image classification using multilayer hidden conditional random fields and akly supervised learning," *IEEE Access*, vol. 7, pp. 90378–90397, 2021
- [6] C. Li, H. Chen, X. Li, N. Xu, Z. Hu, D. Xue, S. Qi, H. Ma, L. Zhang, and H. Sun, "A review for prostrate histopathology image analysis using 43 machine vision approaches," *Artif. Intell. Rev.*, pp. 1–42, Feb. 2020, doi:10.1007/s10462-020-09808-7
- [7] C. Qian, Y. Yu, and Z.-H. Zhou, "Analyzing evolutionary optimization in noisy environments," *Evol. Comput.*, vol. 26, no. 1, pp. 1–41, Mar. 2020.
- [8] Goceri, B. Goksel, J. B. Elder, V. K. Puduvali, J. J. Otero, and M.N. Gurcan, "Quantitative validation of anti-PTBP1 antibody for diagnostic neuropathology use: Image analysis approach," *Int. J. Numer. Methods Biomed. Eng.*, vol. 33, no. 11, p. e2862, Nov. 2020.
- [9] Chollet, "Xception: Deep learning with depthwise separable convolutions," 2020, arXiv:1610.02357. [Online]. Available: <http://arxiv.org/abs/1610.02357>
- [10] Shoeleh and M. Asadpour, "Graph based skill acquisition and transfer learning for continuous reinforcement learning domains," *Pattern Recognit. Lett.*, vol. 87, pp. 104–116, Feb. 2021
- [11] Gautam, H. K. K., N. Jith, A. K. Sao, A. Bhavsar, and A. Natarajan, "Considerations for a PAP smear image analysis system with CNN features," 2020, arXiv:1806.09025. [Online]. Available: <http://arxiv.org/abs/1806.09025>
- [12] Komagata, T. Ichimura, Y. Matsuta, M. Ishikawa, K. Shinoda, N. Kobayashi, and A. Sasaki, "Feature analysis of cell nuclear chromatin distribution in support of prostrate cytology," *J. Med. Imag.*, vol. 4, no. 4, p. 1, Oct. 2021.
- [13] Su, X. Xu, Y. He, and J. Song, "Automatic detection of prostrate cancer cells by a two-level cascade classification system," *Anal. Cellular Pathol.*, vol. 2016, no. 4, 2020, Art. no. 9535027.
- [14] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. Nonhandcrafted features for computer vision classification," *Pattern Recognit.*, vol. 71, pp. 158–172, Nov. 2020.
- [15] L. i, Q. Gan, and T. Ji, "Prostrate cancer histology image identification method based on texture and lesion area features," *Comput. Assist. Surg.*, vol. 22, no. 1, pp. 186–199, Oct. 2021
- [16] L. Zhang, L. Lu, I. Noguees, R. M. Summers, S. Liu, and J. Yao, "DeepPap: Deep convolutional networks for prostrate cell classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1633– 1643, Nov. 2021.
- [17] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of prostate cancer using artificial neural networks," *Procedia Comput. Sci.*, vol. 89, pp. 465–472, 2021

- [18] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228–235, Apr. 2020.
- [19] M. Rahaman, C. Li, X. Wu, Y. Yao, Z. Hu, T. Jiang, X. Li, and S. Qi, "A survey for prostate cytopathology image analysis using deep learning," *IEEE Access*, vol. 8, no. 1, pp. 61687–61710, 2020.
- [20] N. Crossley, C. Tipton, T. Meier, M. Sudhoff, and J. Kharofa, "The value of hybrid interstitial tandem and ring applicators for organ at risk dose reduction in small volume prostate cancer," *Brachytherapy*, vol. 17, no. 4, p. S111, Jul. 2021.
- [21] P. Mahalakshmi, T. Manigandan, and D. Chitra, "Detecting Character in Video Frames for Traffic Vehicle" volume 4 Issue 9 – September 2017.
- [22] Nitin Sonaji Magar, Zafar Ul Hasan, Anand B.Humbe, 2023. "Design and Development of Optimized Cardiovascular Disease Prediction Model using Artificial Intelligence" *ESP International Journal of Advancements In Science & Technology (ESP- IJAST)* Volume 1, Issue 2 : 20-27
- [23] Doctor, A., B. Vondenbusch, and J. Kozak. "Bone segmentation applying rigid bone position and triple shadow check method based on RF data." *Acta of Bioengineering and Biomechanics*, 13.2 (2011): 3-11. [[Google Scholar](#)]
- [24] Muthukumaran Vaithianathan, Mahesh Patil, Shunyeek Frank Ng, Shiv Udkar, 2023. "Comparative Study of FPGA and GPU for High-Performance Computing and AI" *ESP International Journal of Advancements in Computational Technology (ESP-IJACT)* Volume 1, Issue 1: 37-46. [[PDF](#)]
- [25] Ayyalasomayajula, M. M. T., Chintala, S., & Sailaja, A. (2019). A Cost-Effective Analysis of Machine Learning Workloads in Public Clouds: Is AutoML Always Worth Using? *International Journal of Computer Science Trends and Technology (IJCST)*, 7(5), 107–115.
- [26] Naga Ramesh Palakurti, 2022. "AI Applications in Food Safety and Quality Control" *ESP Journal of Engineering & Technology Advancements* 2(3): 48-61.
- [27] Kalla, Dinesh, Nathan Smith, Fnu Samaah, and Kiran Polimetla. "Enhancing Early Diagnosis: Machine Learning Applications in Diabetes Prediction." *Journal of Artificial Intelligence & Cloud Computing*. SRC/JAICC-205. DOI: doi.org/10.47363/JAICC/2022 (1) 191 (2022): 2-7.
- [28] Chanthathi, Sasibhushan Rao. (2022). *A Centralized Approach To Reducing Burnouts In The It Industry Using Work Pattern Monitoring Using Artificial Intelligenc*. *International Journal on Soft Computing Artificial Intelligence and Applications*. Sasibhushan Rao Chanthathi. Volume-10, Issue-1, PP 64-69.[[LINK](#)]
- [29] Aparna Bhat, Rajeshwari Hegde, "Comprehensive Study of Renewable Energy Resources and Present Scenario in India," 2015 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, TN, India, 2015. [[Link](#)]
- [30] S. E. V. S. Pillai, A. A. ElSaid and W. -C. Hu, "A Self-Reconfigurable System for Mobile Health Text Misinformation Detection," 2022 *IEEE International Conference on Electro Information Technology (eIT)*, Mankato, MN, USA, 2022, pp. 242-247, doi: 10.1109/eIT53891.2022.9813840.