چهارمین کنگره
توسعه علمی و فناوری
دانشجویان زیست
شناسی و شیمی
4th Congress of
Scientific and Technological
Development of Biology and
Chemistry Students
www.bcs.cdsts.ir

K-Means Clustering and Dense–Sparse–Dense Optimized LSTM for Prostate Cancer

Aryan Jalaeianbanayan [1,*] Gohar Javadzadeh [2*] ,and *Maryam Darvari* [3, 4]

[1] Department of Computer Science, University of Verona, Verona, Italy, Email: aryan.jalaeianbanayan@studenti.univr.it
[2] Department of Molecular and cellular biology, University College of Rouzbahan, Sari, Iran, Email: javadzadeh.gohar@gmail.com
[3] Department of Cellular and Molecular Biology, Islamic Azad University, Ghaemshahr branch, Ghaemshahr, Iran, Email: Maryam.Darvari@gmail.com
[4] Health Reproductive Research Center, Sari Branch, Islamic Azad University, Sari, Iran

ABSTRACT

Prostate cancer is one of the most common cancers affecting men worldwide. Accurate segmentation of prostate images plays a crucial role in diagnosis and treatment planning. This study presents an optimized segmentation method that combines the k-means clustering algorithm with the elbow method to automatically determine the optimal number of clusters. The approach was applied to two datasets: near-infrared (NIR) prostate images obtained using PSMA-targeted photodynamic therapy (PDT) agents, and histological prostate cancer images. Experimental results show that the method consistently identifies four as the optimal number of clusters, producing clearer and more reliable segmentation compared to conventional methods. Statistical analysis of clustered pixels confirmed the effectiveness of the proposed technique in highlighting tumor regions. The method's simplicity and accuracy make it a useful tool to support radiologists in prostate cancer analysis and diagnosis, although further improvements are needed to address variations in illumination, contrast, and outliers.

Keywords: *Prostate cancer, Image segmentation, K-means clustering, Elbow method, Near-infrared (NIR) imaging*

1. *INTRODUCTION*

Prostate cancer is one of the most prevalent malignancies among men and represents a major health concern worldwide. According to the American Cancer Society, nearly 192,000 new cases and over 33,000 deaths were reported in the United States in 2020 alone. Similarly, national cancer registries in Saudi Arabia rank prostate cancer among the top five cancers in men, reflecting its global significance. Early detection and accurate diagnosis are therefore critical for improving treatment outcomes and reducing mortality *[1, 2]*.

Recent advances in medical imaging have provided clinicians with powerful tools for cancer detection and treatment monitoring. Among them, **photodynamic therapy (PDT)** combined with **prostate-specific membrane antigen (PSMA) targeting agents** has emerged as a promising approach for both noninvasive treatment and image-guided surgery *[3]*. However, reliable image segmentation remains a major challenge. Prostate images are often characterized by heterogeneity, weak boundaries, and low contrast

between cancerous and normal tissues, making manual interpretation by radiologists time-consuming, subjective, and prone to error *[4]*.

Over the last decade, various segmentation techniques have been explored, including deformable models, deep learning methods, and convolutional neural networks (CNNs). While these approaches show potential, they often require large training datasets, complex computations, or suffer from sensitivity to variations in illumination and tissue appearance. Traditional clustering algorithms, such as **k-means**, offer a simpler alternative for unsupervised segmentation, yet their performance strongly depends on selecting an appropriate number of clusters (K). An improper choice of K can lead to poor clustering and inaccurate segmentation *[5]*.

To address this issue, this study proposes an **optimized segmentation approach** that integrates the k-means clustering algorithm with the **elbow method**. The elbow criterion provides an automatic and reliable way to determine the optimal number of clusters, thereby improving segmentation accuracy and reducing manual effort. The method is validated on two datasets: near-infrared (NIR) prostate images acquired using PSMA-targeted PDT agents and histological prostate cancer images. The results demonstrate that the optimized approach produces more accurate segmentation, enhancing visualization of cancerous regions and supporting radiologists in prostate cancer diagnosis *[1]*.

2. Methods and Materials

Datasets

Two datasets were used to validate the proposed segmentation approach.

- **Dataset 1** consisted of four near-infrared (NIR) prostate images obtained from previous work. These images were generated using PSMA-1-IR700 and PSMA-1-PC413 conjugates, with and without light irradiation.
- **Dataset 2** contained four histological prostate images (Gleason pattern 5) collected from the Johns Hopkins University pathology dataset.

Together, these datasets provided representative samples for testing segmentation on both NIR and histological prostate cancer images.

K-Means Clustering

K-means is an unsupervised clustering algorithm that partitions data into *K* groups by minimizing the **sum of squared errors within clusters (SSEWC)**. For an image, pixels with similar intensity or color features are grouped into the same cluster. The algorithm proceeds iteratively:

1. Initialize *K* cluster centroids.
2. Assign each pixel to the nearest centroid.

3. Recalculate centroids based on assignments.
4. Repeat steps 2–3 until assignments stabilize.

The goal is to maximize separation between clusters while minimizing variation within each cluster.

Elbow Method for Optimal Cluster Selection

A major limitation of k-means is the need to predefine *K*. To overcome this, the elbow method was applied:

- For each candidate *K*, the SSEWC was computed.
- SSEWC decreases as *K* increases, but after the optimal *K* the rate of decrease flattens.
- The "elbow point" in the curve indicates the best cluster number, balancing accuracy and complexity.

Proposed Segmentation Approach

The optimized method combines k-means clustering with the elbow technique to automatically determine the most suitable number of clusters for each image. The workflow was:

1. Input NIR or histological prostate image.
2. Apply k-means clustering with varying *K*.
3. Use the elbow method to select the optimal *K*.
4. Segment the image into distinct pixel clusters.
5. Analyze cluster distributions and pixel ratios for diagnostic interpretation.

This approach enables robust segmentation of prostate images while reducing reliance on manual parameter selection.

3. Results and Discussion

The optimized k-means clustering method combined with the elbow technique was applied to two datasets of prostate cancer images.

Elbow Curve Analysis

For both datasets, the elbow plots consistently indicated that **four clusters (K=4)** represented the optimal segmentation. As shown in Figures 4 and 5, the sum of squared errors (SSE) decreased sharply up to K=4, after which further increases in cluster number produced only marginal improvements. This confirmed that four clusters provide a balance between segmentation accuracy and computational efficiency.

Segmentation Performance on NIR Images

Dataset 1 (NIR images) was segmented into four distinct pixel clusters corresponding to light, near-light, near-dark, and dark regions (Figure 6 and Figure 8). The segmented regions were well separated with minimal overlap, highlighting the ability of the method to distinguish between irradiated and non-irradiated areas of prostate tissue. Pixel distribution analysis (Table 1) showed cluster ratios ranging from ~15% to ~34%, with two dominant clusters capturing the majority of pixel intensities.

Segmentation Performance on Histological Images

Dataset 2 (histological Gleason pattern 5 images) also yielded four stable clusters (Figure 7 and Figure 9). The method accurately separated cancerous regions from surrounding tissue. Pixel distribution (Table 2) revealed balanced cluster ratios, indicating that the segmentation captures subtle differences in tumor morphology across samples.

Comparison with the Previous Method

Figure 12 compares the proposed approach with a previous Hopfield neural network-based segmentation. The optimized k-means approach produced cleaner segmentation boundaries and fewer artifacts, enabling improved visualization of malignant regions.

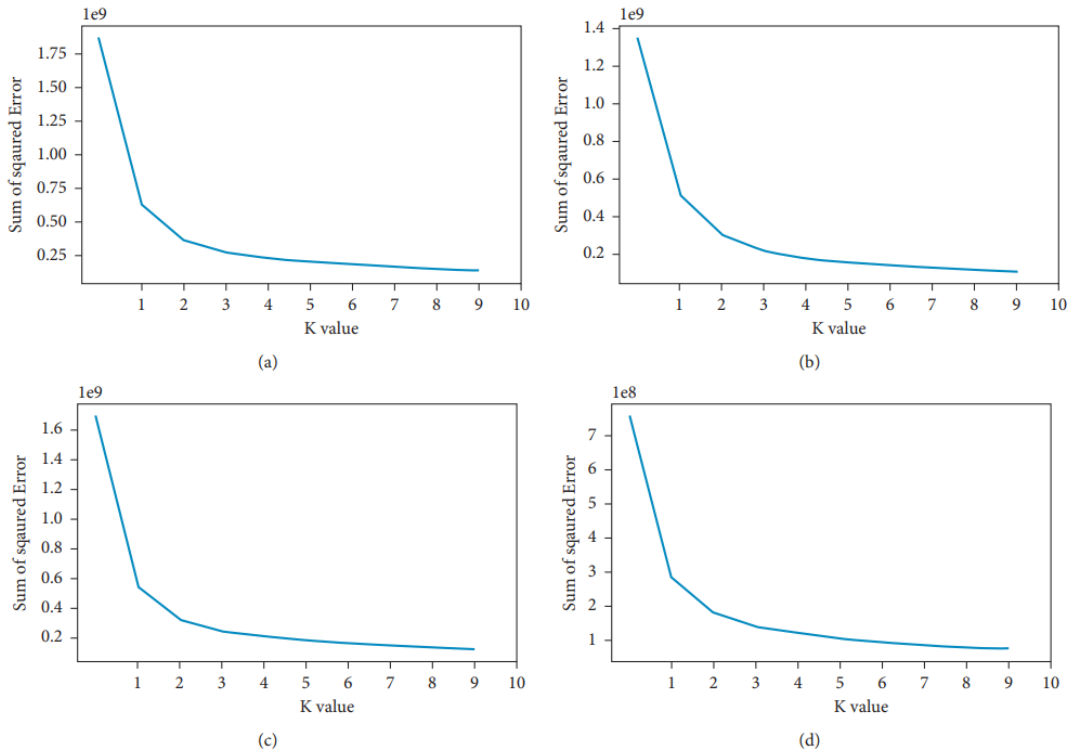


Figure 4. Elbow criterion curves for NIR prostate images (Dataset 1). The highest curvature at *K=4* indicates the optimal number of clusters.

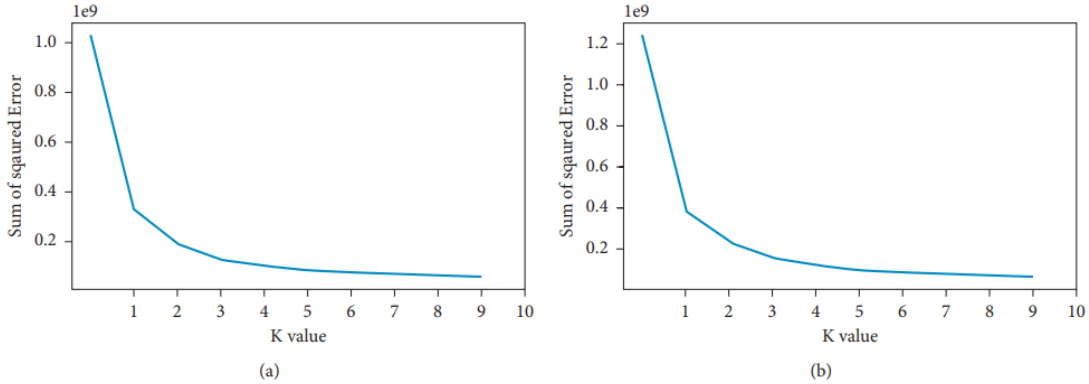


Figure 5. Elbow criterion curves for histological prostate images (Dataset 2). Consistent with Dataset 1, *K=4* provided the best balance between intra-cluster compactness and inter-cluster separation.

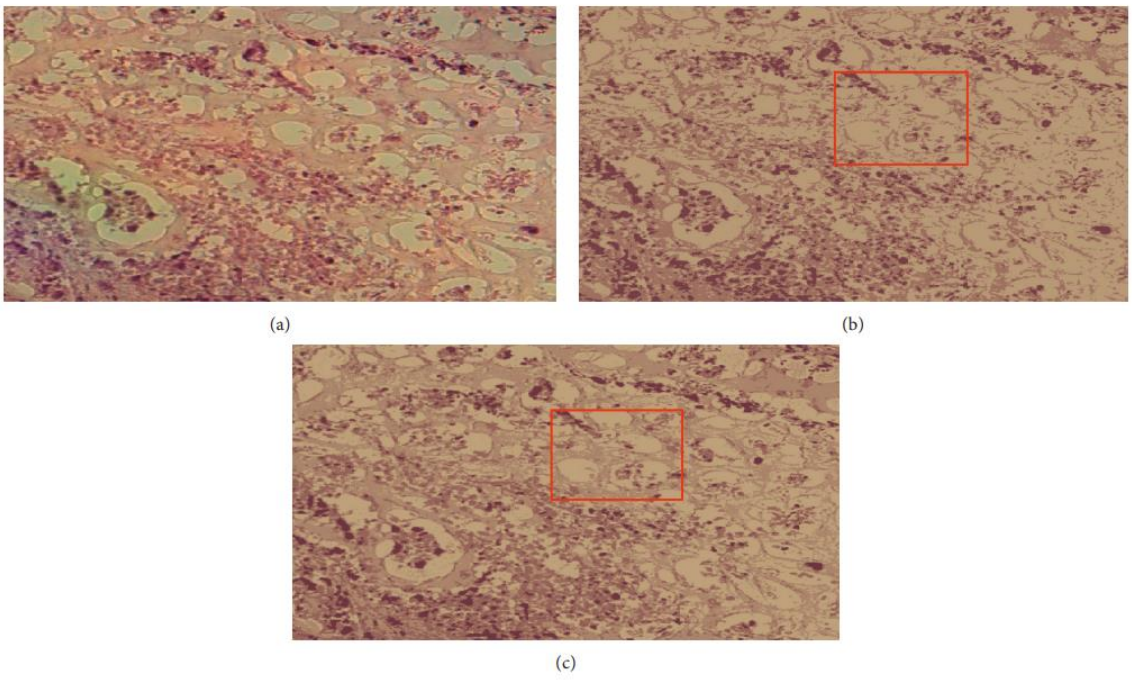


Figure 6. Example segmentation of an NIR prostate image (Dataset 1). (a) Original image; (b) segmentation with *K=3*; (c) segmentation with *K=4*. The optimized approach (c) produces clearer boundaries.

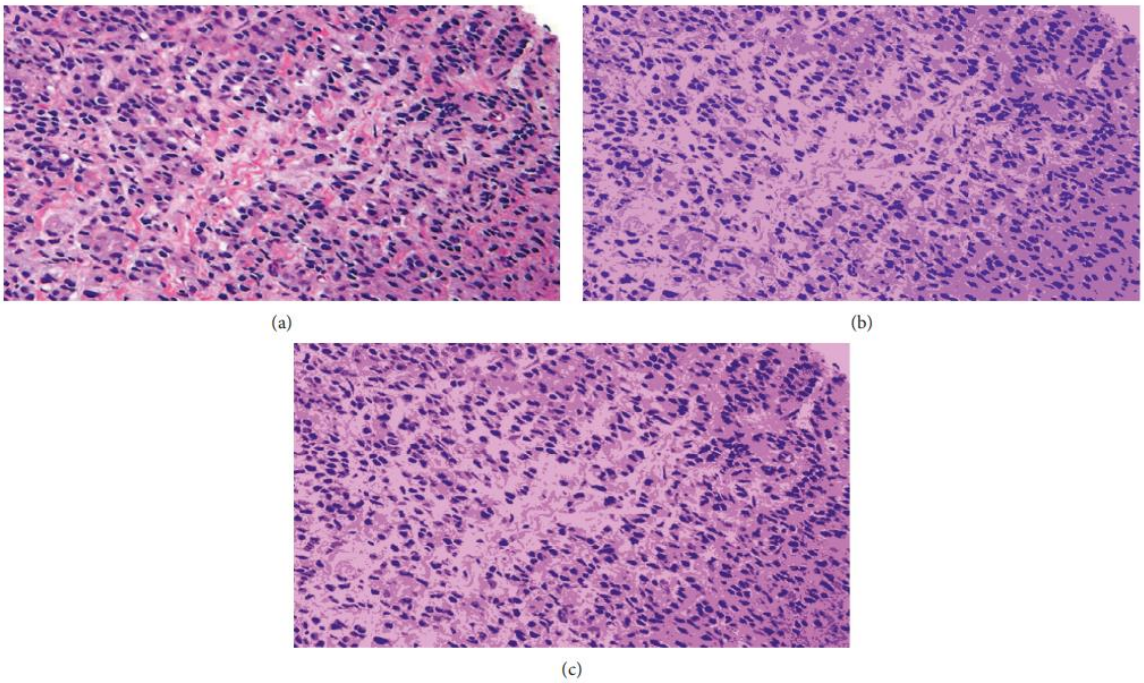


Figure 7. Example segmentation of a histological prostate image (Dataset 2). (a) Original image; (b) segmentation with *K=3*; (c) segmentation with *K=4*.

REFERENCES

1. Davoodi Nik, B., et al., *Differential expression of ion channel coding genes in the endometrium of women experiencing recurrent implantation failures.* Scientific Reports, 2024. **14**(1): p. 19822.
2. Hashemi Karoii, D., et al., *Alteration of the metabolite interconversion enzyme in sperm and Sertoli cell of non-obstructive azoospermia: a microarray data and in-silico analysis.* Scientific Reports, 2024. **14**(1): p. 25965.
3. Shams, A.A., et al., *Paternal trans fatty acid and vitamin E diet affect the expression pattern of androgen signaling pathway genes in the testis of rat offspring.* Theriogenology, 2025. **231**: p. 1-10.
4. Niazi Tabar, A., et al., *Testicular Localization and Potential Function of Vimentin Positive Cells during Spermatogonial Differentiation Stages.* Animals, 2022. **12**(3): p. 268.
5. Hashemi Karoii, D. and H. Azizi, *A review of protein-protein interaction and signaling pathway of Vimentin in cell regulation, morphology and cell differentiation in normal cells.* Journal of Receptors and Signal Transduction, 2022. **42**(5): p. 512-520.