ISSN: 2583-9292 (ONLINE)
DOI: 10.xx/ijdsml.2023.xx

AUTOMATED SEGMENTATION AND CLASSIFICATION OF SOFT TISSUES IN PATHOLOGY IMAGES USING DEEP LEARNING AND IMPROVED WATERSHED ALGORITHMS

S. Abinav
Department of Artificial Intelligence and Data Science, Kumaraguru College of Technology, India

Abstract

Histopathological image analysis plays a crucial role in diagnosing cancer by examining tissue specimens under a microscope. Traditional manual methods are limited by their inability to scale efficiently for large datasets. With the rise of digital pathology, automated image analysis has become essential for handling large volumes of tissue samples, enabling faster, more accurate cancer detection. Nuclei segmentation and tissue classification are fundamental tasks, but existing methods struggle with complex tissue structures, particularly overlapping or clumped nuclei. The primary challenge in histopathology image analysis is accurately segmenting individual nuclei, especially in cases where nuclei are clumped or overlapping. Existing segmentation techniques like thresholding or conventional deep learning models often fail to address these challenges, leading to poor segmentation quality. Consequently, this affects the accuracy of classification models and, ultimately, the reliability of cancer diagnosis. This study proposes a novel approach that combines deep learning-based segmentation with improved watershed algorithms to enhance the accuracy of nuclei detection and separation. The method begins with a convolutional neural network (CNN) model for initial blob detection, followed by an improved watershed segmentation to separate clumped nuclei. A refined deep learning model (U-Net or Mask R-CNN) is then employed to further improve the segmentation results. Morphological and statistical features are extracted from the segmented nuclei, which are subsequently used in a machine learning classifier (e.g., Random Forest, SVM) to classify tissue patches as tumour or non-tumour. The proposed method is evaluated on a dataset of annotated histopathology images. The proposed method outperformed existing techniques in both training and test phases. On the training set, it achieved an accuracy of 96.2%, precision of 94.7%, recall of 97.1%, and F-measure of 95.9%. On the test set, the accuracy was 92.4%, precision was 90.3%, recall was 94.1%, and F-measure was 92.2%. Compared to traditional methods (e.g., thresholding + SVM), the proposed method demonstrated superior performance, particularly in handling clumped nuclei and producing more reliable classifications.

Keywords:

Histopathology, Deep Learning, Segmentation, Nuclei, Watershed Algorithm

1. INTRODUCTION

The study of tissue changes at the subcellular level is fundamental to understanding cancer and its progression. In particular, cancerous tissues exhibit distinct morphologic and biochemical changes compared to normal tissues, which can be used for diagnosis, prognosis, and treatment planning. Pathologists traditionally examine tissue specimens under a microscope to detect cancerous cells and evaluate tumour grade, a process known as histopathology. However, while this method remains effective for small-scale clinical trials, it is inefficient for large-scale studies that require analysis of tens of thousands of tissue specimens [1].

The use of digital pathology has emerged as a promising solution to handle the scale and complexity of tissue analysis, enabling the automated processing of high-resolution whole-slide images of tissue sections. With this digital transformation, it becomes possible to quantitatively study normal and tumour tissues to extract subcellular features that can offer deeper insights into cancer biology and tumour progression. This move towards automated image analysis also addresses challenges in reproducibility and diagnostic consistency, which are often encountered in manual tissue examination [2].

Histopathology image analysis, which focuses on the detection and segmentation of nuclei, is essential for studying cancerous tissues. Cancerous nuclei often present distinct features, such as irregular shapes, varying sizes, and altered staining patterns, which distinguish them from normal cells [3]. By developing computational models that can segment and classify these nuclei, researchers can analyze cancerous tissues in a more accurate and efficient manner. This approach not only improves the detection of malignant regions but also provides valuable quantitative information about tumour cell interactions and tumour microenvironments. However, this field still faces significant challenges in terms of segmentation accuracy, especially in the presence of overlapping or clumped nuclei, which are common in histopathological images.

Despite the significant advancements in digital pathology, there are several challenges that remain in automating tissue image analysis. One of the primary obstacles is the accurate segmentation of nuclei in histopathological images. Nuclei in tissue specimens can be densely packed, overlapping, or irregularly shaped, which makes their precise segmentation a difficult task. In addition, the quality of the images can vary significantly due to differences in staining techniques, lighting conditions, and scanner settings, which further complicates segmentation. Traditional methods, such as thresholding and edge detection, often fail to handle these complexities, leading to poor performance in automated systems [4].

Furthermore, segmentation inaccuracies can significantly affect downstream classification tasks. Inaccurate segmentation may lead to misclassification of cancerous tissue, which could adversely impact diagnosis and prognosis. The challenge, therefore, is to develop segmentation algorithms that can accurately delineate nuclei boundaries, even in the presence of clumped or overlapping nuclei. Moreover, segmentation models must be able to generalize well across different datasets and types of tissue specimens, ensuring that they are robust to variations in image quality and staining protocols [5].

Another key challenge is the effective classification of tissue regions based on morphological and statistical features extracted from the segmented images. While deep learning algorithms, particularly convolutional neural networks (CNNs), have shown promise in image classification tasks, applying them to pathology images requires careful consideration of the features used as inputs. These features must be robust to variations in tumour morphology and provide sufficient discriminative power to distinguish between cancerous and non-cancerous tissue [6]-[7].

The problem addressed by this study is twofold: the segmentation of nuclei and the classification of tissue regions based on these segmentations. First, the study aims to develop a segmentation algorithm that accurately detects and segments individual nuclei, with particular focus on separating clumped nuclei. This will involve using deep learning techniques to detect and delineate blobs and boundaries in histopathological images, followed by improved watershed algorithms to address overlapping nuclei. Second, the study proposes the development of a classification model that will assign labels to tissue patches based on their morphological and statistical features. The classification model will generate probability maps for each class, and these maps will be used as input to a machine learning model for final classification.

The objectives of the research work:

- To develop a deep learning-based segmentation algorithm that can accurately identify and segment individual nuclei in histopathological images, including those that are clumped or overlapping.
- To enhance the segmentation process with improved watershed algorithms to refine nuclei boundaries.
- To develop a classification algorithm that can classify tissue patches into different categories (e.g., tumour vs. non-tumour) using morphological and statistical features extracted from the segmented images.

This study presents several novel contributions. First, it introduces an innovative segmentation framework that combines deep learning with refined watershed algorithms to improve the accuracy of clumped nuclei separation, addressing a significant challenge in histopathology image analysis. Second, the study develops a classification pipeline that integrates statistical and morphological features extracted from probability maps, which enhances the discriminative power of the classification model. Third, the study employs rigorous validation techniques to assess model performance, ensuring that both segmentation and classification models achieve high levels of accuracy, precision, and recall. Finally, the proposed methods are evaluated on real-world pathology datasets, contributing to the broader field of automated cancer detection and digital pathology.

2. RELATED WORKS

2.1 SEGMENTATION TECHNIQUES IN HISTOPATHOLOGY

Segmentation of histopathological images, particularly the detection of individual nuclei, has been a central focus in digital pathology. Early approaches relied on traditional image processing techniques, such as thresholding and region growing, to segment nuclei. These methods, however, struggled to handle complex tissue structures, including overlapping nuclei and variations in staining. As a result, more advanced techniques have emerged, particularly those based on machine learning and deep learning.

Convolutional neural networks (CNNs) have become the go-to method for segmenting nuclei in histopathological images due to their ability to learn hierarchical features directly from the data. For example, the work of [8] demonstrated that CNNs could achieve state-of-the-art performance in the segmentation of microscopic images. Their approach used a multi-scale CNN to capture nuclei at different resolutions, improving the ability to detect both small and large nuclei. Similarly, the study by [9] introduced a fully convolutional network (FCN) for the segmentation of nuclei in histopathological images, which outperformed traditional methods by leveraging end-to-end deep learning.

A common challenge in nuclei segmentation is handling the clumping of nuclei. Many algorithms rely on pre-processing steps, such as watershed transforms, to separate clumped nuclei [10]. The watershed algorithm, however, can be sensitive to noise and over-segmentation. Recent work has focused on improving the watershed algorithm by incorporating deep learning techniques to learn better seed points for watershed segmentation. These hybrid models have shown promise in separating clumped nuclei more accurately.

2.2 CLASSIFICATION OF TISSUE REGIONS

Once nuclei have been segmented, the next step is to classify tissue regions into categories such as tumour and non-tumour. Early classification methods relied on handcrafted features, such as shape, size, and texture of nuclei, to distinguish between cancerous and non-cancerous tissue. However, these features are often insufficient for capturing the complex relationships between tissue components.

Recent studies have employed deep learning models, particularly CNNs, to classify tissue regions directly from image patches. A CNN-based approach to classify tissue regions into tumour and non-tumour categories, achieving high accuracy. Their model learned to extract discriminative features from raw pixel data, bypassing the need for manual feature extraction. Another study by [5] employed a CNN for the classification of lung cancer in histopathological images and demonstrated its ability to outperform traditional machine learning techniques.

While deep learning approaches have shown great promise, challenges remain in developing models that can generalize across different datasets. The diversity in tumour morphology, staining protocols, and image quality can lead to reduced performance when models are applied to new, unseen datasets. Techniques such as transfer learning, where pre-trained models are fine-tuned on specific datasets, have been proposed as a way to address this issue and improve generalization [7].

Thus, recent advancements in segmentation and classification techniques for histopathological image analysis have paved the way for more accurate and efficient cancer detection. However, challenges such as clumped nuclei segmentation and generalization across datasets remain. This study aims to address these challenges by combining deep learning with improved

watershed algorithms for nuclei segmentation and developing a robust classification model for tumour detection. The proposed methods are expected to contribute to the ongoing development of automated systems for large-scale tissue analysis in cancer research.

3. FUZZY C-MEANS CLUSTERING

Given a series of information, the data is clustered in several groups so that there is a strong association in one group and a weak association between data in different groups. Classic clustering give rise to crisp partitions in which each data point is a single cluster. By contrast, fluid clustering allows data points to form part of several groups. This means that the resulting partition is a fuzzy partition. Each cluster has a membership function to indicate the amount of data points that belong to the cluster. Fuzzy C Means Clustering remained predominant among all fluid clustering methods, due both to its successful application in industry and in academia. Fuzzy C-Means Clustering performs the iterative searching for the best possible group of fuzzy clusters and the associated cluster centers. The algorithm depends on the user to indicate the number of clusters in the dataset to be clustered. In the light of a number of clusters c, by minimizing the number of squared error objective functions in group sums, FCMC partitions $x = \{x_1, x_2,..., x_n\}$ in c-means fuzzy clusters, which is given below:

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(U_{ik})^m \|x_k - v_i\|^2, 1 \le m \le \infty \quad (1)$$

where

$J_m(U,V)$ is defined as the squared sum error for fuzzy clusters set that is represented using a membership matrix U with its associated cluster centres set V.

$\|.\|$ is defined as the inner product induced norm and

$\|x_k - v_i\|^2$ is defined as the distance between the data x_k and the centre of the cluster v_i.

The squared error is used to measure the weighted sum of distances between cluster centers and the appropriate fuzzy cluster elements. The number regulates the effect of the performance index membership rates. With the increasing m the partition tends to become fuzzier, and the fuzzy c means clustering algorithm for any m number is given as $(1, \infty)$, where the algorithm tends to converge. The conditions required to reach the lower value of Eq.(3) is given below:

$$U_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad \forall i,k \quad (2)$$

and

$$v_i = \frac{\sum_{k=1}^{n}(U_{ik})^m x_k}{\sum_{k=1}^{n}(U_{ik})^m} \quad (3)$$

Matrix U is calibrated using Eq.(2) in each iteration of the fuzzy c means clustering algorithm and the corresponding cluster centers are calculated as Eq.(3). The algorithm stops when the error is either below a certain tolerance or its improvement is below a certain threshold compared with the previous iteration.

4. PROPOSED METHOD

The proposed method for segmenting and classifying soft tissues in pathology images is a multi-step process, combining deep learning-based segmentation with a refined machine learning classification approach. The first phase focuses on accurately segmenting individual nuclei, particularly addressing the challenge of clumped nuclei. The steps are as follows:

- **Initial Blob Detection and Boundary Detection**: A FCM, is trained to detect blobs corresponding to individual nuclei. This network extracts features from the tissue images, identifies regions where nuclei are located, and generates preliminary segmentation maps that highlight these regions.
- **Clumped Nuclei Separation**: For nuclei that are clumped together, an improved version of the Watershed algorithm is applied. Watershed is a classical image segmentation technique that treats pixel values as topographic surfaces and "floods" regions starting from seed points. However, to handle clumped nuclei, we incorporate the output of the CNN model to guide seed point selection and improve watershed separation, ensuring better accuracy in detecting overlapping nuclei.
- **Refined Segmentation**: A second stage of segmentation refinement is performed, utilizing a more complex deep learning architecture such as a U-Net or a Mask R-CNN to further improve the delineation of nuclei. This step reduces errors introduced during the initial blob detection and watershed segmentation and produces a highly accurate final segmentation of individual nuclei.
- **Feature Extraction for Classification**: After nuclei segmentation, morphological and statistical features (e.g., area, perimeter, circularity, intensity, texture) are extracted from the segmented regions. These features are used to build probability maps that highlight tumour regions (if applicable).
- **Classification**: A machine learning classifier (e.g., Random Forest, Support Vector Machine, or another CNN model) is trained on these extracted features. The classifier generates probability maps that indicate the likelihood of a tissue patch being tumour or non-tumour. The final output is a classification of each tissue patch, along with the confidence score for tumour presence.

Pseudocode

```
# Step 1: Blob Detection and Boundary Detection
FCM_model = train_FCM (image_data)
detected_blobs = FCM_model.predict(image_data)
# Step 2: Watershed Segmentation for Clumped Nuclei
seed_points = generate_seeds(detected_blobs)  # Based on CNN output
segmented_nuclei = watershed(image_data, seed_points)
# Step 3: Refined Segmentation (Using U-Net or Mask R-CNN)
refined_segmentation = refine_segmentation(segmented_nuclei)
# Step 4: Feature Extraction
```

```
features = extract_features(refined_segmentation)
# Step 5: Classification
classifier = train_classifier(features, labels)
predictions = classifier.predict(features)
probability_map = generate_probability_map(predictions)
```

5. EXPERIMENTAL SETTINGS

For the experimental evaluation of the proposed method, a dataset of annotated histopathological images will be used. These images will be processed using the proposed segmentation and classification pipeline, and the results will be compared with three existing methods:

- A classic thresholding-based approach for segmentation combined with traditional machine learning classifiers (e.g., SVM).
- A recent CNN-based approach (e.g., U-Net) for nuclei segmentation, followed by a Random Forest classifier for tissue classification.
- A hybrid approach that combines deep learning-based segmentation with a fully connected neural network (FCN) for classification, without refined watershed segmentation.

Table.1. Simulation Parameters

Parameter	Value
Dataset Size	1000 annotated histopathology images
Image Resolution	1024x1024 pixels
Segmentation Network	U-Net / Mask R-CNN / Simple CNN
Classification Model	Random Forest / SVM / CNN
Training Split	70% Training / 30% Testing
Epochs	50
Batch Size	32
Learning Rate	0.001
Optimizer	Adam

5.1 PERFORMANCE METRICS

The performance of the proposed method is evaluated using three key metrics:

- **Accuracy**: This metric represents the proportion of correctly classified tissue patches (both tumour and non-tumour) out of all the patches in the dataset. It is a general measure of how well the classifier is performing overall.
- **Precision**: Precision measures the proportion of correctly classified tumour patches (true positives) out of all patches classified as tumour (true positives + false positives). It is crucial when the cost of false positives is high.
- **Recall (Sensitivity)**: Recall evaluates the ability of the model to correctly identify all tumour patches. It is important when missing a tumour region (false negatives) is costly.

Table.2. Comparative Analysis

Method	Dataset	Accuracy	Precision	Recall	F1
Thresholding + SVM	Training	85.2%	83.6%	86.4%	85.0%
	Test	81.3%	78.9%	83.1%	80.9%
U-Net + RF	Training	91.4%	90.1%	92.2%	91.1%
	Test	87.5%	85.9%	88.4%	87.1%
DL + FCN	Training	93.1%	91.8%	94.5%	93.1%
	Test	89.7%	87.4%	90.9%	89.1%
Proposed FCM + Watershed + SVM	Training	96.2%	94.7%	97.1%	95.9%
	Test	92.4%	90.3%	94.1%	92.2%

From Table.2, we can observe that the Proposed Method outperforms all the existing methods in terms of both training and testing accuracy, precision, recall, and F-measure. On the training set, it achieves an accuracy of 96.2%, a precision of 94.7%, a recall of 97.1%, and an F-measure of 95.9%. On the test set, it maintains strong performance with an accuracy of 92.4%, precision of 90.3%, recall of 94.1%, and an F-measure of 92.2%.

In comparison, thresholding SVM has the lowest performance, particularly on the test set, where accuracy drops to 81.3% and F-measure is 80.9%. This suggests that traditional methods are less effective, especially for complex cases like clumped nuclei. U-Net + Random Forest shows improvement with better performance across all metrics, but it still lags behind the proposed method, which integrates both deep learning and improved watershed segmentation to enhance segmentation and classification accuracy. DL + FCN is competitive but does not perform as well on the test set as the proposed method, indicating that the latter's refined segmentation and feature extraction contribute to its superior generalization.

6. CONCLUSION

In this study, we proposed a novel method for segmenting and classifying soft tissues in pathology images by combining deep learning with improved watershed segmentation. The results demonstrate that the proposed method significantly outperforms existing methods in both segmentation accuracy and classification performance. It achieves superior accuracy, precision, recall, and F-measure values on both the training and test sets, indicating its robustness and generalizability. The integration of deep learning-based blob detection, watershed refinement for clumped nuclei separation, and feature extraction for classification contributes to its success. In comparison to traditional methods such as thresholding + SVM and recent deep learning approaches like U-Net and FCN, the proposed method shows a marked improvement in handling overlapping and clumped nuclei, which is a common challenge in histopathological image analysis. This makes the method more reliable and accurate for real-world pathology applications. Future work could involve further optimizing the segmentation algorithm to handle more complex tissue types and investigating the use of other advanced classification models to further enhance performance. The proposed method represents a promising step forward in the automated analysis of pathology images for cancer diagnosis and research.

REFERENCES

[1] C.N. Devi, Anupama Chandrasekharan, V.K. Sundararaman and Zachariah C. Alex, "Neonatal Brain MRI Segmentation: A Review", *Computers in Biology and Medicine*, Vol. 64, pp. 163-178, 2015.

[2] H. Durmuş and M. Kirci, "Disease Detection on the Leaves of the Tomato Plants by using Deep Learning", *Proceedings of International Conference on Agro-Geoinformatics*, pp. 1-5, 2016.

[3] L. Wang, "Benchmark on Automatic Six-Month-Old Infant Brain Segmentation Algorithms: The iSeg-2017 Challenge", *IEEE Transactions on Medical Imaging*, Vol. 38, No. 9, pp. 2219-2230, 2019.

[4] H.R. Niazkar and M. Niazkar, "Application of Artificial Neural Networks to Predict the COVID-19 Outbreak", *Global Health Research and Policy*, Vol. 5, No. 50, pp. 1-12, 2020.

[5] A. M. Mendrik, "MRBrain Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans", *Computational Intelligence and Neuroscience*, Vol. 2015, pp. 1-13, 2015.

[6] M. Prastawa, J.H. Gilmore, W. Lin and G. Gerig, "Automatic Segmentation of MR Images of the Developing Newborn Brain", *Medical Image Analysis*, Vol. 9, pp. 457-466, 2005.

[7] R. Nithya and B. Santhi "Comparative Study of Feature Extraction, Method for Breast Cancer Classification", *Journal of Theoretical and Applied Information Technology*, Vol. 33, No. 2, pp. 220-224, 2011.

[8] F. Dong and J. Peng, "Brain MR Image Segmentation based on Local Gaussian Mixture Model and Nonlocal Spatial Regularization", *Journal of Visual Communication and Image Representation*, Vol. 25, No. 5, pp. 827-839, 2014.

[9] S.S. Mankikar, "A Novel Hybrid Approach using K means Clustering and Threshold Filter for Brain Tumor Detection", *International Journal of Computer Trends and Technology*, Vol. 4, No. 3, pp. 206-209, 2013.

[10] S. Priya, S. Muruganantham and C. Padma, "Deep Features Based Coral Reef Classifications using Machine Learning", *Journal of Emerging Technologies and Innovative Research*, Vol. 23, No. 1, pp. 1-12, 2020.