

Pattern Imaging Analytic of Cancer: A survey

Nikita Sharma^{1*}, Hari Mohan Rai², Archika Verma¹, Sunidhi Tomar¹, Kartik Thakur¹, Ayush Bhatia¹,
Rishabh Nigam¹, Kapil Budlakoti¹

¹Department of Computer Science Engineering, Dronacharya group of institutions, Greater Noida

²Department of Electronics & Communication Engineering, Dronacharya group of institutions, Greater Noida

Abstract

This is meant to serve as an overview of the technical specification, which will be provided later. High-Resolution Microscopy Images of Soft Tissue Specimens Can Provide Precise Information Regarding the Morphology of Both Healthy and Diseased Tissue. Because of the complexity of tissue morphology and tumor heterogeneity, the development of accurate and efficient algorithms for these tasks is a challenging problem. This is a problem that must be solved. Captk is a software platform for the analysis of radiographic cancer images, with a primary focus at this time on cancers of the brain, breast, and lungs. Captk is an integration of advanced and validated tools that perform various aspects of medical image analysis. These tools have been developed in the context of active clinical research studies and collaborations with the goal of addressing real clinical needs. Captk is a component of the Captk platform. Captk aims to facilitate the rapid translation of advanced computational algorithms into routine clinical quantification, analysis, decision making, and reporting workflow. With an emphasis placed on its use as a very lightweight and efficient viewer, and with no prerequisites requiring a substantial background in computation, Captk does not require any prior knowledge of computational theory. Its long-term objective is to provide technology that is widely used and leverages the value of advanced imaging analytics in cancer prediction, diagnosis, and prognosis, as well as in better understanding the biological mechanisms that contribute to the development of cancer.

Keywords: Pattern Analysis, Biomedical Imaging, Image Processing, Detection, Segmentation

1. Introduction

Cancer is the abnormal growth of the human cells in uncontrolled manner, it may be arises in any parts of the body [1]–[3]. It can be cured if it is detected at the initial stage but in the final stage it becomes deadly. The most common types of cancers are skin cancer, breast cancer, brain cancer, lung cancer, mouth cancer, and any others. The hallmarks of cancer include spatial and temporal heterogeneity at many scales, which contributes to an unfavorable prognosis and treatment failure [4]. Through the acquisition of a variety of different modalities, clinical imaging provides the opportunity to shed light on the complex phenotypic aspects of the structure and physiology of cancer. When it comes to determining the heterogeneity of the underlying cancer, semantic features such descriptors of size, morphology, and location, which are often measured from radiologic images, have limited accuracy. Emerging in the field of quantitative analysis of oncologic multiparametric imaging is the field of cancer imaging phenomics, abbreviated as Ciph. Ciph provides a broad spectrum of phenotypic imaging signatures, which has the potential to bring increased precision to the diagnosis, prognosis, and prediction of response to therapy [5]. This is accomplished through mathematical measurements of the features that were mentioned above, which are more commonly known as radiomic features[6]. Captk provides a systematic quantification platform for designing clinical research studies using radiomics and radiogenomics methods, from harmonized data preprocessing and extraction of rich sets of ciph features that represent tumor characteristics to integration of features using appropriate machine learning (ml) methods for precision diagnosis and prediction captk provides a systematic quantification platform for designing clinical research studies using radiomics and radiogenomics methods captk provide Glioblastoma multiforme (Gbm) is an aggressive and genetically diverse neoplastic malignancy that has a poor response to available therapies, rapid progression, inevitable recurrence, short median progression-free survival (Pfs), limited treatment options, and a median overall survival (Os) of 15 months. In this article, we examine the application of captk in Gbm. Gbm is an aggressive and genetically diverse

neoplastic mal The improvement in a patient's overall survival or progression-free survival (as a surrogate marker of overall survival in patients with GBM) is the most objective way to measure the efficacy of novel treatment strategies. Pfs can be difficult to objectively define; however, when it is used as an end point, it has the advantage of helping to complete trials more quickly and more rapidly determining which interventions may be helpful for patients. This is despite the fact that it can be difficult to objectively define Pfs [7], [8]. Figure 1 presents a flow diagram of the process of pattern recognition used in the detection of cancer.

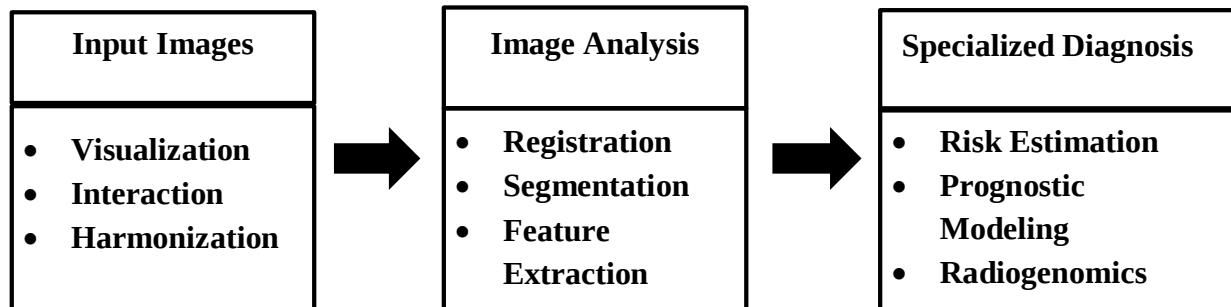


Fig.1: The flow diagram of the pattern recognition process

2. Analysis and Discussion

It is observed from the review table shown in table 1, that the best performance in the cancer detection techniques is achieved by authors in [9], where they have used K-means clustering technique and achieved 99.85 of classification accuracy. K means clustering is a method of vector quantization that seeks to partition n observations into k clusters, with each observation belonging to the cluster that has the nearest mean (cluster centers or cluster centroid), which serves as the prototype for the cluster. K means clustering is an example of a hierarchical clustering technique. As a consequence, the data space is partitioned into Voronoi cells. K-means clustering is able to minimize within-cluster variances (squared Euclidean distances), but it is unable to minimize regular Euclidean distances, which is the more difficult Weber problem. The geometric median is the only measure that can reduce Euclidean distances, while the mean is able to optimize squared errors. For instance, by using k -medians and k -medoids, it is possible to generate superior Euclidean solutions. The best possible outcome, with 99.8 percent accuracy, 100 percent sensitivity, and 99.5 percent specificity. The removal of the area of interest in the picture is dependent on the image being segmented using k -means clustering prior to the use of the suggested approach. The process of removing features was carried out across a wide variety of feature descriptors, namely LBP, HOG, and BoVW. Several categorization strategies were used in order to evaluate the performance of the features. The performance is maximized using the approach that was suggested.

In a similar vein, the authors in [8] attain one of the lowest levels of accuracy, despite the fact that they use several methodologies and data sets. In this study, the CaPTk program was used in order to construct and carry out the radiomic analysis for the purpose of PFS and RP prediction. CaPTk was developed with a three-level capability for radiomic analysis in mind when it was first designed. Activities related to the preparation of photos, such as transferring pictures from one format to another, make up the first level of the game. It is possible to perform a variety of things with photos, including smoothing, segmentation, registration, and format conversion. The second level consists of procedures for the extraction of features, the selection of features, and machine learning. These routines are used by CaPTk in order to do specialized tasks. Additionally, they may be used for individualized analytic pipelines as well as closed-form applications.

Table 1: The Review table of the cancer detection and segmentation Techniques

Literature	Technique	Dataset Used	Model Name	Accuracy
[9]	K-Means Clustering	HAM10000	Vector machines classifier	99.8%
[10]	CNN	-	Deep Learning Algorithms	95%
[11]	Semi Supervised Deep Learning	Patt-Patch Level Training and Testing	SSL,SL	93%
[5]	CNN	-	Deep Learning	90%
[6]	Brain-Captk	BRATS	Extensive Radiomic	89.92%
[12]	CNN	-	Patch Level	81%
[13]	Data Sampling Machine Learning	-	Pattern Based Hyper Conceptual Sampling	76%(NV) 82%(SVM)
[14]	Captk	-	Extensive QIP Panel	70.18%
[15]	Statistical Techniques	-	Animal Model of Dietary Pattern	--
[16]	Anti-Mannan Antibodies, D-Glucan Detection, and Other Molecular Technique	Retrospective Cohort Study	ML	--
[2]	Statistical Technique	SPSS (Version 25)	Statistical Model	-
[8]	Cancer Imaging Phenomics Toolkit (Captk)	Canicing Imaging Phenomics	Multi - Institutional phonemics	71%
[17]	Machine Learning Technique		GLCM	84%

Abbreviations: NV – Navye Bayes, SVM – Support Vector Machine, CNN – Convolutional Neural Network, GLCM – Gray Level Co-Occurrence Matrix, SL – Supervised Learning, SSL – Semi Supervised Learning

This level in particular aims to extract multiple features that capture diverse components of local, regional, and global imaging patterns. This will result in a broad feature panel that complies with the requirements of the Image Biomarker Standardization Initiative³¹, which will ensure that the radiomic features are reproducible and comparable. Additionally, it uses machine learning to construct prediction and diagnosis models by synthesizing features in order to differentiate between bigger feature sets and smaller feature subsets that are more relevant. The third level of CaPTk focuses on combining these characteristics into specialized applications through the use of machine learning algorithms to accomplish specific goals. These goals include making precise diagnoses, determining the risk of developing cancer, and generating models that predict a patient's response and survival. Subject to patients following normal first-line chemoradiation treatment, imaging signatures of presurgical MP-MRI scans reveal pretty strong prediction of timing and

location of GBM recurrence. This finding is accurate to a degree of 0.71 percent. Through its graphical user interface, CaPTk makes it simple to access powerful computational algorithms for determining imaging signatures that are predictive of clinical outcome. Additionally, CaPTk has the potential to be utilized for a wide variety of radiomic and radiogenomic research applications.

IMAGING PROCESSING: The CaPTk software was used to preprocess multiparametric magnetic resonance imaging (MP-MRI) data, which included T1-weighted (T1w) contrast-enhanced (T1CE), T2-weighted (T2w), T2-fluid-attenuated inversion recovery (T2-FLAIR), diffusion tensor imaging (DTI), and dynamic susceptibility contrast-enhanced MRI (DSC-MRI) images (rCBV). There were a few different types of diffusion derivatives, including axial diffusivity (AX), trace (TR), radial diffusivity (RAD), and fractional anisotropy (FA). An experienced radiologist named S.M. approved or adjusted the automatic segmentation that was performed by the DeepMedic32 module in order to identify tumoral subregions such as the enhanced tumor (ET), the nonenhancing part of the tumor core (NC), and the peritumoral edema using CaPTk (ED). Akbari et al¹¹ and Bakas et al³³ propose definitions. Two more tumor subregions were constructed in order to accurately depict the tumor core (TC, which was defined as the union of the ET and NC subregions) and the total volume of the tumor (WT, estimated as the union of all three tumour subregions [ie, ET, NC, and ED]).

RADIOMETRIC FEATURE EXTRACTION : The feature extraction panel in CaPTk was used in order to complete the processing of the preprocessed pictures. Relevant imaging features were computed for each patient from the five tumoral regions (ET, NC, ED, and TC) and all modalities in order to capture phenotypic aspects of short versus long PFS and close versus distant RP. This was done in order to better understand the relationship between these two outcomes. Extracted variables include volume, shape, and size, in addition to intensity, histogram, and gray-level co-occurrence matrix (GLCM) variables. Using these criteria, you will be able to conduct a quantitative investigation of the morphologic and geographic heterogeneity aspects of tumoral areas. A total of 1,980 features were obtained from the data for institution 1, but only 1,016 features were obtained from the data for the multi-institutional study.

3. Conclusion

In the study paper, an overview of several strategies utilized for the detection and segmentation of cancer was offered. After doing the research, we found that using K-means clustering to the HAM10000 datasets yields the best results, with a classification accuracy of 99.8 percent. There are also other methods of detecting cancer and segmenting images, but the program that is used for detecting cancer is called CapTk, and it has an approach that is entirely its own. It has an accuracy rate of over 70 percent when applied to huge picture datasets. Therefore, in order to identify and segment cancer, the dataset that is being used should be authenticated and benchmarked so that the correctness of the model can be confirmed. Additionally, not only must it be accurate, but it must also be both rapid and efficient in order for the process of cancer diagnosis to be carried out in a timely and effective manner.

4. Future Scope

In the future, we may incorporate a greater number of study articles to evaluate and investigate the cancer detection method. In addition, we shall limit ourselves to the methodology of a single kind of cancer detection, such as that of the lung, the skin, the brain, etc. We will make an effort not only to evaluate, but also to present our very own outcomes in an effort to raise the degree of accuracy by integrating the optimization approach with the deep CNN model.

5. References

- [1] J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," *Complex Intell. Syst.*, 2021, doi: 10.1007/s40747-021-00563-y.
- [2] C. I. Owobu *et al.*, "Pattern of Cancer in Irrua Specialist Teaching Hospital," *Int. J. Trop. Dis. Heal.*, vol. 42, no. December 2020, pp. 14–21, 2021, doi: 10.9734/ijtdh/2021/v42i730468.
- [3] T. Saba, "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges," *J. Infect. Public Health*, vol. 13, no. 9, pp. 1274–1289, 2020, doi: 10.1016/j.jiph.2020.06.033.

- [4] M. Hasan, S. Das Barman, S. Islam, and A. W. Reza, "Skin cancer detection using convolutional neural network," *ACM Int. Conf. Proceeding Ser.*, no. March 2020, pp. 254–258, 2019, doi: 10.1145/3330482.3330525.
- [5] S. Wang *et al.*, "Artificial intelligence in lung cancer pathology image analysis," *Cancers (Basel)*, vol. 11, no. 11, pp. 1–16, 2019, doi: 10.3390/cancers11111673.
- [6] S. Rathore *et al.*, "Brain Cancer imaging phenomics toolkit (brain-CaPTk): An interactive platform for quantitative analysis of glioblastoma," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10670 LNCS, pp. 133–145, 2018, doi: 10.1007/978-3-319-75238-9_12.
- [7] I. Cabria and I. Gondra, "MRI segmentation fusion for brain tumor detection," *Inf. Fusion*, vol. 36, pp. 1–9, 2017, doi: 10.1016/j.inffus.2016.10.003.
- [8] A. Fathi Kazerooni *et al.*, "Cancer Imaging Phenomics via CaPTk: Multi-Institutional Prediction of Progression-Free Survival and Pattern of Recurrence in Glioblastoma," *JCO Clin. Cancer Informatics*, no. 4, pp. 234–244, 2020, doi: 10.1200/cci.19.00121.
- [9] K. Iqtidar, A. Iqtidar, W. Ali, S. Aziz, and M. U. Khan, "Image Pattern Analysis towards Classification of Skin Cancer through Dermoscopic Images," *Proc. - 2020 1st Int. Conf. Smart Syst. Emerg. Technol. SMART-TECH 2020*, no. January 2021, pp. 208–213, 2020, doi: 10.1109/SMART-TECH49988.2020.00055.
- [10] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA - J. Am. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, 2017, doi: 10.1001/jama.2017.14585.
- [11] G. Yu *et al.*, "Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images," *Nat. Commun.*, vol. 12, no. 1, pp. 1–13, 2021, doi: 10.1038/s41467-021-26643-8.
- [12] Q. D. Vu *et al.*, "Methods for segmentation and classification of digital microscopy tissue images," *Front. Bioeng. Biotechnol.*, vol. 7, no. APR, 2019, doi: 10.3389/fbioe.2019.00053.
- [13] T. Salahuddin *et al.*, "Breast cancer image classification using pattern-based Hyper Conceptual Sampling method," *Informatics Med. Unlocked*, vol. 13, no. May, pp. 176–185, 2018, doi: 10.1016/j.imu.2018.07.002.
- [14] C. Davatzikos *et al.*, "Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome," *J. Med. Imaging*, vol. 5, no. 01, p. 1, 2018, doi: 10.1117/1.jmi.5.1.011018.
- [15] S. E. Steck and E. A. Murphy, "Dietary patterns and cancer risk," *Nat. Rev. Cancer*, vol. 20, no. 2, pp. 125–138, 2020, doi: 10.1038/s41568-019-0227-4.
- [16] vinod k ramani, "Analysis of Bloodstream Infections and Their Antibiotic Sensitivity Pattern (Pre- and Post-COVID Lockdown in an Indian Cancer Hospital): A Record-Based Retrospective Cohort Study," *Eurasian J. Med. Oncol.*, 2022, doi: 10.14744/ejmo.2022.18855.
- [17] T. T. Tang, J. A. Zawaski, K. N. Francis, A. A. Qutub, and M. W. Gaber, "Image-based Classification of Tumor Type and Growth Rate using Machine Learning: a preclinical study," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-48738-5.