# Natural Language Processing for Health Education and Behavior Change Interventions

**Ashu Nayak[1], Kapesh Subhash Raghatate[2]**

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.*

Cervical cancer is one of the main causes of cancer-related mortality among Indian women. Due to social taboos, early pregnancy, poor cleanliness, and low literacy, it might have a significant death rate in its mature phases. India does not have nearly as many pathologists as it needs, especially in rural areas. Cervical cancer can be identified more precisely and with less human resources by an automated method that makes use of image processing and machine learning techniques. The segmented CT image, features extracted from the segmented area, relevant and helpful features selected from the extracted feature set, and features classified into classes corresponding to normal or malignant conditions are all made easier by these algorithms for cervical cancer detection. This work aims to create, deploy, and improve the functionality of a cervical cancer detection system that classifies and recognises malignant cells in CT images using a natural language processing approach.

**Keywords:** Natural Language processing, Health care, detection, prediction.

## 1. Introduction

The many cell kinds that make up the human body. Every tissue is made up of cells. Cells in our bodies divide, grow, and die constantly. However, there are instances when cell division and growth spiral out of control. Tumour or cancer formation is the result of the body's cells proliferating out of control [1]. Cells typically perish when they age or sustain injury. However, cancer cells do not end their lives at the expected time [3]. Certain types of malignancies exhibit rapid cell proliferation, while slower cell division and growth occur in other cancer types [10]. When genetic alterations tamper with this well-organized process, cancer starts [9]. Benign and malignant are the two broad categories into which all cancers fall [7]. Benign cancers and tumours are confined to their original place and are not very aggressive. Malignant tumours, on the other hand, pose a greater risk to life, are more aggressive, and have the potential to spread to other body areas [8]. The World Health Organisation (WHO) projects that 9.6 million deaths worldwide will be attributable to cancer

in 2018 [2]. Globally, cancer is the second most common cause of death. Cancer accounts for 70% of cancer-related deaths in middle-income nations. Regardless of gender, lung cancer is the most prevalent type of cancer globally [13]. By reducing modifiable risk factors like alcohol and tobacco use, poor diet, and physical inactivity, many cancers can be avoided. Unhealthy eating practices along with a sedentary lifestyle are another significant factor [14]. Furthermore, a good number of tumours can be managed or treated with surgery, radiation, or chemotherapy, particularly if they are found early. Cutting edge, trustworthy, and secure diagnostic methods are accessible.

This is how the rest of the paper is structured. The lung segmentation algorithms whose capabilities are assessed in this work are described in Section 3. The results and discussion of the experiment are detailed in Section 4. Lastly, Section 5 provides the conclusion and next steps.

## 2. Literature Survey

These days, the analysis of medical images is a widely employed field of study that requires quick advancements [11]. In clinical diagnosis, the parameters of diagnosis system robustness, accuracy, and efficiency are critical [6]. It serves as a method for lung image tumour detection because of its spatially resolved tissue contrast [4]. CAD systems are used to analyse both high- and low-resolution CT images. Throughout the last two decades, the rapid growth of CAD systems has been visually represented. Using their unaided eyes, radiologists analyse lung images both quantitatively and qualitatively [5].

The human eye is unable to discern even the smallest details in the lung image's grayscale. Accurate and effective detection is essential as incorrect identification and categorization might be lethal. The segmentation algorithms should have faster convergence for real-time applications [12-7]. Several efforts have been taken in order to analyse medical lung images. In the field of medical image processing, lung tumour detection is a difficult task, and much research is being done in this area. It is said that pattern recognition is an extremely difficult procedure that relies on extracted feature vectors to make meaningful conclusions.

## 3. Methodology

unprocessed pictures from CT scanners and picture processing is a technique that takes a picture, digitalizes it, and applies certain operations to it to create an augmented image from which important information can be extracted. Similar to digital signal processing, it is somewhat. An image (which could be a still frame from a video or a snapshot in any format) captured during a test is the input image in image processing. The output could be another image or just a subset of the input image's features. An image is regarded as a two-dimensional signal in image processing systems. This technology is still in its infancy, yet it has applications across many different industries. Image processing is a thrust topic in other computer science fields and a fundamental research subject in engineering. Researchers are very interested in image processing because it allows for real-time applications and user access to the findings of image processing techniques..

The primary techniques covered in this part include CT preprocessing techniques like de-

noising, contrast enhancement, and noisy region removal. 3 x 3 is used for de-noising, while a modified color-based Histogram equalisation technique is used for improvement. The suggested CT yields outputs for each level and is based on mathematical morphology and related regions. To determine the region that has to be clipped, this approach makes use of the mathematical morphology's dilation attribute. This algorithm's accuracy is calculated manually with database images. The preparation time of 0.84 seconds on average indicates the speed of the suggested algorithm. No additional research has been done on the suggested method because the main goal is to create a reliable and accurate segmentation algorithm. Since the suggested approach has a high degree of accuracy, the segmentation quality will improve.

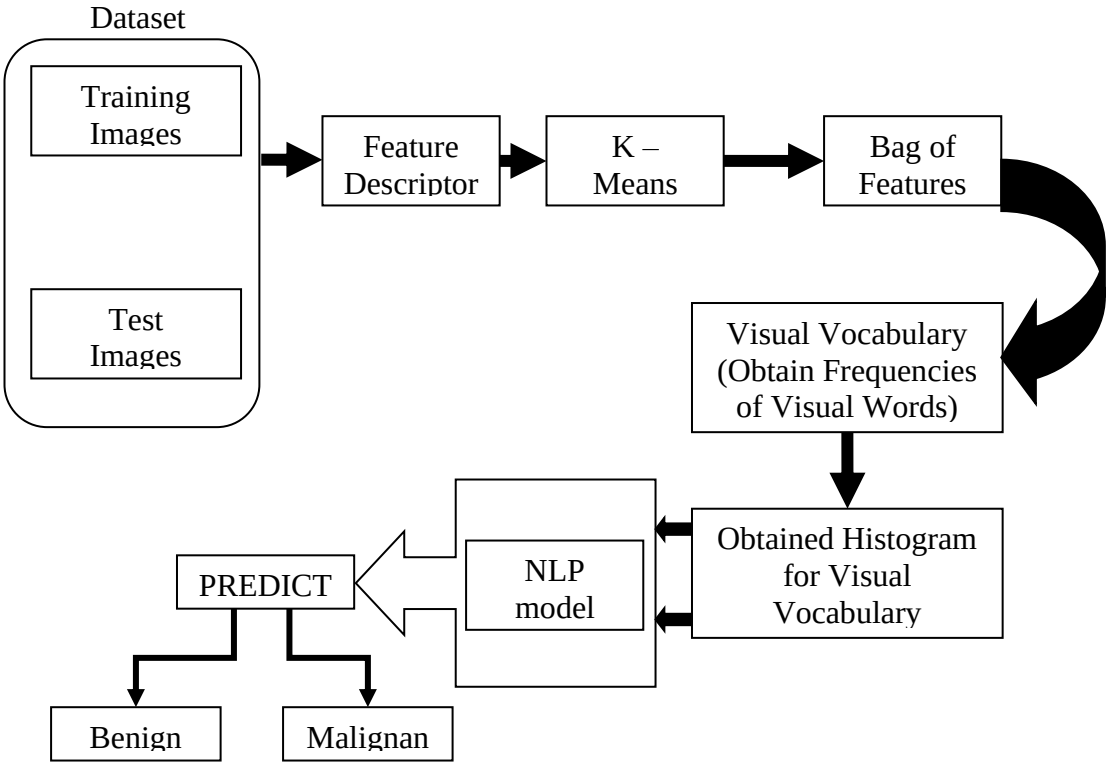


Figure 1: NLP Classifier

This is the NLP algorithm's expansion. The supervised learning model is called BOVW. Make a bag of visual terms and use the computer vision system toolbox tools for classifying images into categories. A histogram of visual word occurrences that depict an image is produced by the process. For the purpose of training an image category classifier, this histogram.

## 4. Results and Discussion

The input photos are extracted in the experimental setup. Images of the normal and abnormal cervix are captured for analysis during testing. The photos were captured from CT scans. Any

size can be used for the test image. Preprocessed photos make up the input. Subsequently, the improved pictures are divided. Within the tumours, the thick tissue areas are divided. Ultimately, the picture is categorised as either an aberrant or normal cervix. The pictures are displayed in Figure 2.

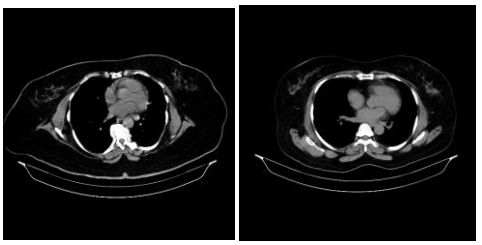

Figure 2: Input Images

Figure 3 displays the preprocessing output for the matching input image..

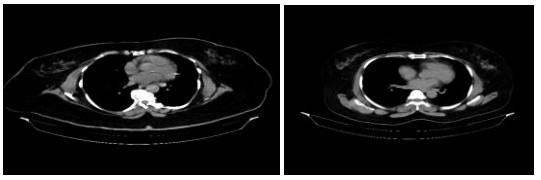

Figure 3: Preprocessing output

The output of segmentation is shown in figure 4.

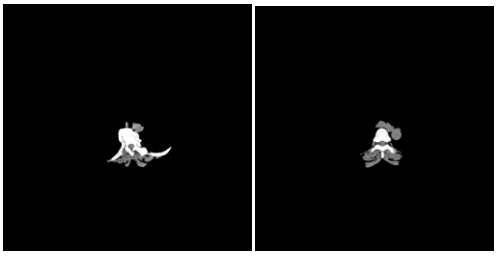

Figure 4: Segmentation output

This method of segmentation looks at the pixels that surround the original seed points and decides if such pixels should be included in the region or not. The procedure is repeated, much like standard data clustering techniques.

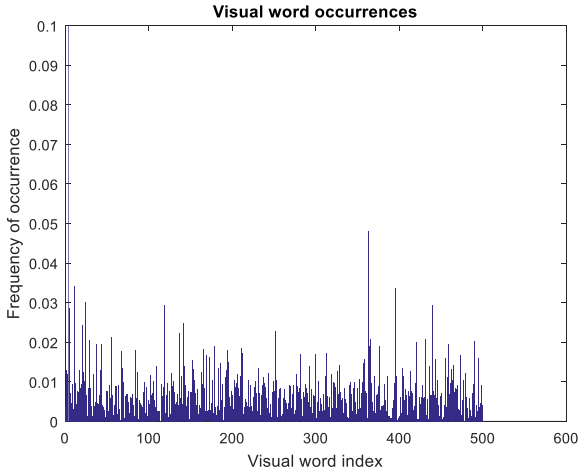


Figure 5: Visual vocabulary BOVW classifier

The bag of visual words categorization results, with an accuracy of 98.6%, are displayed in Figure 5. The confusion matrix for this method is formed when the picture bag of features is extracted before to classification. Once the confusion matrix has been formed in this BOVW classifier.

## 5. Conclusion

India is home to more than 1.35 billion people, making it the world's second-most populous country. Approximately 70% of Indians live in rural areas with inadequate access to healthcare, societal norms, low literacy rates combined with social taboos, and subpar personal hygiene and personal care facilities. Different issues come up when treating lung malignancies, which makes segmenting them challenging. The sizes and forms of these tumours vary. It can appear in any location and at various picture intensities. These methods provide a variety of lung imaging data. It is laborious to accomplish the automated segmentation using prior miniature or prior observation. Precisely dividing up the interior structures of the lung is crucial for tumour research and treatment. It works to improve the management of tumours by surgery or radiation therapy while reducing mortality. An evocative human lung model that can coordinate tumour information gathered from Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) data is also desirable in the field of lung oncology.

## References

1. Elayaraja, P., and M. Suganthi. "Automatic approach for cervical cancer detection and segmentation using neural network classifier." Asian Pacific journal of cancer prevention: APJCP 19, no. 12 (2018): 3571.
2. Bedell, Sarah L., Lena S. Goldstein, Amelia R. Goldstein, and Andrew T. Goldstein. "Cervical cancer screening: past, present, and future." Sexual medicine reviews 8, no. 1 (2020): 28-37.
3. Ramakrishnan, J., Ravi Sankar, G., & Thavamani, K. (2019). Publication Growth and Research in India on Lung Cancer Literature: A Bibliometric Study. Indian Journal of Information

Sources and Services, 9(S1), 44–47.
4. Pramanik, Rishav, Momojit Biswas, Shibaprasad Sen, Luis Antonio de Souza Júnior, João Paulo Papa, and Ram Sarkar. "A fuzzy distance-based ensemble of deep models for cervical cancer detection." Computer Methods and Programs in Biomedicine 219 (2022): 106776.
5. Xue, Peng, Jiaxu Wang, Dongxu Qin, Huijiao Yan, Yimin Qu, Samuel Seery, Yu Jiang, and Youlin Qiao. "Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis." NPJ digital medicine 5, no. 1 (2022): 19.
6. Alamer, L., Alqahtani, I. M., & Shadadi, E. (2023). Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier. Journal of Internet Services and Information Security, 13(1), 01-10.
7. Singh, Sanjay Kumar, and Anjali Goyal. "Performance analysis of machine learning algorithms for cervical cancer detection." In Research Anthology on Medical Informatics in Breast and Cervical Cancer, pp. 347-370. IGI Global, 2023.
8. Mehmood, Mavra, Muhammad Rizwan, Michal Gregus ml, and Sidra Abbas. "Machine learning assisted cervical cancer detection." Frontiers in public health 9 (2021): 788376.
9. Ramana, R.H.V., & Ravisankar, V. (2024). Precision in Prostate Cancer Diagnosis: A Comprehensive Study on Neural Networks. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 109-122. https://doi.org/10.58346/JOWUA.2024.I2.008
10. Sneha, K., and C. Arunvinodh. "Cervical cancer detection and classification using texture analysis." Biomedical and Pharmacology Journal 9, no. 2 (2016): 663-671.
11. Kashyap, Debashree, Abhishek Somani, Jatin Shekhar, Anupama Bhan, Malay Kishore Dutta, Radim Burget, and Kamil Riha. "Cervical cancer detection and classification using Independent Level sets and multi SVMs." In 2016 39th international conference on telecommunications and signal processing (TSP), pp. 523-528. IEEE, 2016.
12. Kundrod, Kathryn A., Chelsey A. Smith, Brady Hunt, Richard A. Schwarz, Kathleen Schmeler, and Rebecca Richards-Kortum. "Advances in technologies for cervical cancer detection in low-resource settings." Expert review of molecular diagnostics 19, no. 8 (2019): 695-714.
13. REYES, EFRAIM O., et al. "Energy Harvesting Digital Filter Implementation With Novel Architecutre." International Journal of communication and computer Technologies 11.1 (2023): 75-84.
14. Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. Natural and Engineering Sciences, 9(1), 72-83.
15. Shiraz, Aslam, Robin Crawford, Nagayasu Egawa, Heather Griffin, and John Doorbar. "The early detection of cervical cancer. The current and changing landscape of cervical disease detection." Cytopathology 31, no. 4 (2020): 258-270.