



Critical Review and Analysis for Early Diagnosis of Uterus Cancer Detection

¹Anjali Deswal, ²Sanjeev Dhawan, ³Kulvinder Singh
^{1,2,3}Department of Computer Science and Engineering
University Institute of Engineering and Technology (UIET),
Kurukshetra University, Kurukshetra-136119, Haryana,
INDIA

Abstract: Nowadays, Health care has become one of the superior cares in the domain of researchers and cancer is amongst one of them. Cancer growth describes a certain danger problem in today's world. In developing countries, early diagnosis of uterus cancer represents an essential part in the treatment of cancer detection and diminishing deaths affected by cancer. Patients undergoing from this condition can simply be spared if and only if it is discovered in an early stage (stage I and stage II). If else identified in the final stage (stage III and stage IV) then have a very shorter possibility of survival. With the aim of early diagnosis of uterus cancer detection methods like machine learning and data mining have the forward systems for predicting classification which can efficiently detect the appearance of cancer extension and can preserve the many lives. In this paper, work done by the many researchers has been presented an attempt has been made to diagnosis the early detection for uterus cancer by introducing a new methodology.

Keywords: Uterus Cancer, Early Detection, Machine Learning, WEKA

I. Introduction

Cancer is the most basic problem nowadays amongst the worldwide which have created human life brutally in the past few cycles. With its numerous types, the state of uterus cancer has seen in the amount of large of reported cases since the past few years thus which is growing an essential element of current interest in the healthcare area. Uterus Cancer [1] has always been dangerous among the public health issue. Every year approximately about 4.4 million women suffer from this disease. It is highly recommended to the women to have a regular screening test to the ease of discovering it. There are various kinds of Screening and testing techniques exists which are introduced to diagnosis the early-stage cancer detection. Technologies[2][3] like Data mining and machine learning which are judged to be the most powerful learning for the patients, which dependably give more explicit prediction and cost-effective explication and are used by many organizations which concentrate in various specialities such as medical, healthcare, higher education, telecommunication etc. Cancer can happen due to unnatural changes, or seldom family past is also accountable for the maturity of the cancer cell.

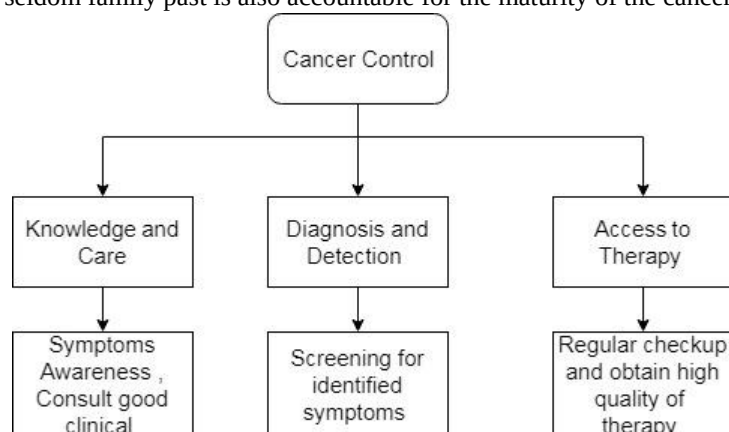


Figure 1 Flow Diagram for Cancer Diagnosis

II. Causes of Uterus cancer

A. Genetic factors

- Dominant: Uterus cancer is dominant in females and recessive in male.
- Family history and genetic factors: Seldom, family past is also capable of been diagnosed with cancer.
- Menstrual and reproductive history: Beginning discharge (before age 12), postponed menopause (After age 45), producing your first child at an older age, or nevermore conceived an issue can also raise the venture of uterus cancer.
- Certain genes: This can be determined by the genetic circumstances a personage with a natural gene modification by moving it onto their offspring.

B. Environmental and lifestyle risk factors

- Lack of physical activity: A disturbing lifestyle with physical motions can decline the risk of cancer.
- Poor diet: Deficiency of high riches of healthful nutrition can diminish the chance of uterus cancer.
- Obesity: This can raise the danger of destructive growth.
- Absorb a Large Quantity of Alcohol: On the Annual premise of the consumption of liquor can hazard the danger of fetal growth.

III. Types of uterus cancer

Uterus destructive growth expansion can happen in any layer of the uterus there is no term for cancer which is consist of three layers and others layers too.

Types of uterus cancer include:

- **Endometrium layer:** This layer is firm for implantation, which is in care of the mensuration cycle and built up of connective membranes, endocrine processes, columnar epithelium connective muscles, endocrine glands, columnar epithelium.
- **Myometrium layer:** This layer is the middlemost thickness of the uterus which is also termed as the outside layer and formed of soft muscle, and reliable for childbirth.
- **Perimetrium layer:** This layer contributes support to the uterus part and moreover also known as the external utmost layer of the uterus.

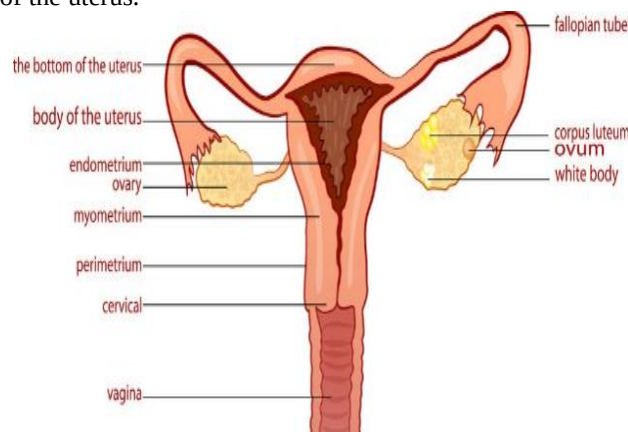


Figure 2 Female Reproductive System [9]

IV. Related work

The women death rate is expanding step by step due to uterus malignant growth. It is a tumour that forms in the cells of the uterus; the current techniques give away to group cancerous and non-cancerous tumour in uterus cells. By impacting the use of advanced techniques medical experts can recognize whether an individual will be influenced by uterus disease sickness or not. Different researchers used several detection algorithms to find an early detection for uterus cancer and classifiers to identify an early diagnosis. In this literature, research papers related to uterus cancer and other cancers diagnosis are also explored. The present review of the literature has been carried out from the year 2002 to 2019 in chronological order.

Similarly, D. B. Fogel *et al.*[4] Provided an overview of the recent works for neural networks and linear classifiers to the evolution of breast cancer. Computer-assisted mammography can also be used to implement an impression which may produce the specific for the diagnosis.

Likewise, Z. Zhang *et al.* [5] in this paper ANN classifier performed on the small portion (N=174) of the total accessible data and tested result shown more self-sufficient for the ANN classifier. The results produce by mathematical calculation is more predictive power which assures the ANN classifiers may perform better on upcoming future data. M. Krishnan *et al.* [6] proposed a method for extracting a new texture based on the parameters in spectra of colon images also suggested this method can be applied for the multiple an array of abnormalities. Seki, Kazuhiro *et al.* [7] Introduced a method which can identify protein names in the biomedical field along with the protein name boundary and method utilized a simple heuristics technique by model and combined with the word classes to avoid data problem. In the mean, V. Gautam *et al.* [8] improved existing classifiers of KNN distance-based and further compared the methods with real and synthetic datasets. This paper presented [9] an association method which was applied to the medical image. The result showed the accuracy of the joining associative classifier which reaches 77.48% instead of 69.11% and executes the Apriori algorithm itself. Future work can improve the results by extracting various types of features. While. BHill *et al.* [10] Genetic algorithm were successfully expanded for the multiclass classification and Experiments has given efficient results with other used classifiers, ANN, SVM and KNN.

Additionally, M. Chen *et al.* [11] Provided the prediction accuracy and capability for the decision tree by performing logistic regression technique. To apply logistic and decision have shown effected result for breast cancer from the prediction models by a discovered method with a tree structure. Consequently, S. Abouet *et al.* [12] proposed a system which was used to identify automatic for early diagnosis of liver cancer detection with an accuracy of 96.125%. The system can be used for its maximum accuracy and its computation simplicity. Likewise, Z. Chen *et al.* [13] proposed a support vector machine which can rely on many kernels in order to obtain the marker gene. Moreover, this method has achieved classification accuracy using the discovered marker genes. The future aim is to improve its performance. This paper [14] have achieved method, better in capabilities by including the global information from the data record and used Dynamic Programming for reduction of the complexity in the proposed methodology. I. El Naqa [15] presented a summary for the statistical modelling methods which can be utilised to combine different features also intended for producing robust PET-based models for output. However, the domain is yet in its opening to provide interest from multiple different societies of stakeholders in atomic medicine, oncology, radioactivity biology and bioinformatics to succeed popular tests and recognised the potential for decision-producing in oncology. Yomna *et al.* [16] proposed an LCPS system which focuses on the patients suffering from the lung cancer gives the doctors and patients room to forecast and define a selected targeted treatment process to the patients care and can improve the quality of treatments for the patients. LCPS can accept a large number of datasets and can employ several data mining algorithms to find health signs and probable outcomes. Thus, the system has decreased lung cancer death cases. This paper introduced [17] a system which predicts cancer or non-cancer ineffective way also encourage doctors to focus on the appropriate treatment for many patients. Where the Decision tree algorithm and the K-means have shown results in the proposed system. Similarly, Y.E Kurniawati *et al.* [18] Obtained datasets from various data records to perform Pap smear test results. Based on the achievement metrics Random forest tree was found best classifier amongst distinct classifiers. Meanwhile, S. G. Sutar [19] build an integrating method to obtain therapy side outcome symptom, medication names, and data is obtained from a cancer forum. Further work is to improve clustering efficiency and improve medication recommendation. M. T. Scholar conducted [20] a survey on accuracy test for breast cancer and diabetes using WEKA tool. Where J48 gave 72.28% accurate results on breast cancer and SMO gave 76.80% accurate result on diabetes also researchers said WEKA tool is considered the best tool for data mining classification. B. Kaur [21] upgraded of the ovarian disease patients with a specific end to improve the quality and able to distinguish between affected area or non-affected area. The result is obtained by using a genetic algorithm. Proposed [22] an automated detection of effective lymph in the mammograms and for its classification has been developed. Where false positive values are greatly reduced and higher accuracy for the classification has been achieved. Similarly, F. Harirchi *et al.* [23] image features with the significant dataset have been classified by using SVM classifier where the image contains 25 innocuous images and 25 malignant images by proposed algorithm efficiency has been adopted up to 94%. P. Hamsagayathri [24] focused on the execution of J48 and priority-based decision tree algorithm for breast cancer analysis. The result was shown on priority based classifier with 98.5%. This paper [25] presented a survey was introduced a new system which overcomes some issues from the previous system also concluded that most automatic cancer prediction systems based on machine learning concepts by including classification and clustering algorithm.

Subsequently, G. Tamilmani *et al.* [26] studied various techniques for cancer discovery with different events which put advanced with an innovative scheme. A future scope is shown in a combination of gene-based algorithm and biological. Efficiency and least RMSE of 0.0833 also recommended an algorithm which occupies less time to build the also diminishes complexity, cost and memory usage.

In the meantime, V. Jpdlo *et al.* [27] performed methods where the dataset of breast cancer was taken from UCI, results were shown with 97.901% on K-nearest neighbours and 96.6035% on a multilayer perceptron. This paper

also calculated the Kappa statistic, ROC area, MCC, Mean absolute error measure. ROC has shown the highest percentage with 99.9% and multilayer perceptron have shown 98.01% of ROC area. Consequently, P. Singhal [28] diagnosed an algorithm for breast cancer performed using the back-propagation algorithm and output was analysed by model, obtained seven hidden neurons has the maximum precision as compared with the other models also found lesser the mean square error more is the accuracy of the model and feed-forward backpropagation algorithm consider the best algorithm Primary. A. Sanjay *et al.* [29] proposed a new hybrid feature selection approach for the early-stage forecast of breast cancer. Data mining used this system which was extremely effective and the result was received from different features choice methods which helped to diagnose cancer in the beginning stage. Moreover, D. S. Jacob *et al.* [30] performed a remembrance of classification and clustering algorithm where the classification algorithm was considered the genuine algorithm for prognosticating breast cancer based on the precision of the algorithm. S. Bakken *et al.* [31] proposed an application for data science which has shown successful results same as methods used by nurses and provided various challenges and set targeted roles for primary responsibilities for using data science application in the advancement of health management.

V. Existing Methodology

Yomna *et al.* [16] proposed an LCPS system which focuses on the patients suffering from the lung cancer gives the doctors and patients room to forecast and define a selected targeted treatment process to the patients care and can improve the quality of treatments for the patients. LCPS can accept a large number of datasets and can employ several data mining algorithms to find health signs and probable outcomes. Thus, the system has decreased lung cancer death cases.

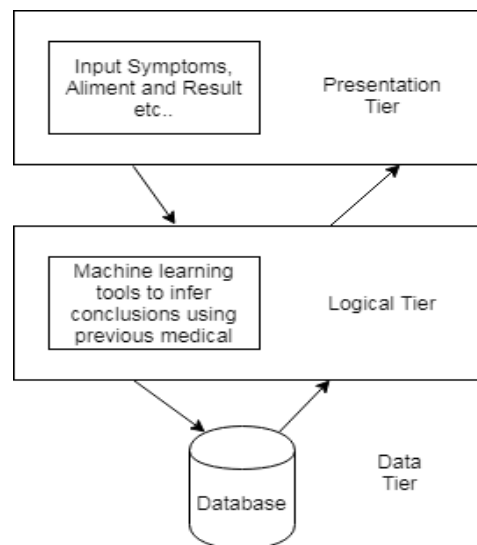


Figure 3 Three Tier System Architecture for Uterus Cancer Diagnosis [16]

- Presentation Tier: This layer consists of input symptoms, an ailment which further produces an efficient result.
- Logical Tier: This layer is responsible for performing the logical operations where machine learning infers to the conclusion using the previous medical history.
- Database Tier: The database is the collection of the related information which can be recorded and can have a certain importance. A database can be of any size and complexity.

VI. Proposed Methodology

Aim of proposing methodology is to contribute to cancer society where several patients suffering from serious disease like uterus cancer, which needs proper treatment along with accurate efficiency. In the proposed methodology, uterus cancer diagnosis algorithm will be applied to the dataset to diagnose cancer based on determining parameters. Further the resultant dataset to the WEKA classifier to check the efficiency of the algorithm. The efficiency of the algorithm will be based on performance metrics such as TP matrix, Confusion matrix, and ROC curve.

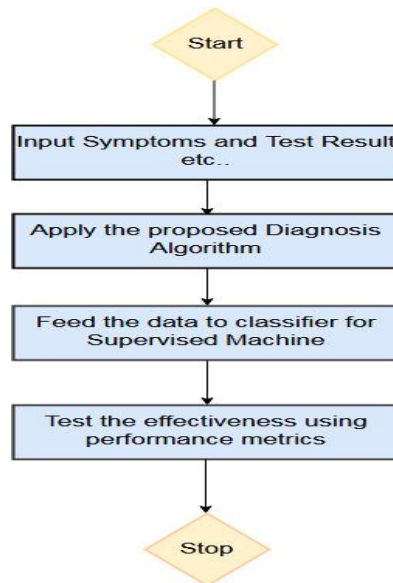


Figure 4 Proposed Methodology Uterus Cancer Detection

VII. Issues and Challenges

This section focuses on the improvements for the existing uterus cancer detection system available.

A. Screening for cancer

The screening term shows all the potentialities before the manifestations shows. The screening mass relates to the group of screening where personalities belong to the related kind of age group or same medicinal form with most probabilities of cancer illness which requires tests like needle test for a cancer cell, urine and some imaging, Pap smear. But the difficulty begins in the screening is the necessity of exactness. So the screening must be followed by any of the superior ways that can be perfect and make fewer faults.

B. Tests in Cancer Detection

Early-stage discovery for cancers still have a durability possibility of more than 80% but the techniques available now for early-stage discovery is faulty. The modern classifications have conferred some particular procedures like bio marking based exposure, high-grade delicate Signings for most kinds of ovarian malignancy.

C. Overview of Existing Detection Techniques

Contemporary time detection ways are more forward than in the past time. It has four mild methods are as follows: Pre-process, Specialty extraction, Segmenting and analysis of the cell. Proper utilization for diagnosing of breast infection mode is called as Mammogram. For skin cancers, Pseudo medial algorithm is applied to divide the tone of the skin representation. Liver damage area can be viewed by the CT classification by side method but for Uterus Cancer still, major difficulties are waiting to receive an exact sequence.

VIII. Conclusion

This literature review and the studies have given an overall relationship of the early diagnosis of uterus cancer still endures an essential firm and aims to utilize data mining and machine learning have the facilities to give numerous precise prediction and cost-effective clarification for the better approach. Moreover, research will undoubtedly be helpful to unleash the various techniques to the analysis of uterus cancer in an early stage.

IX. References

- [1] Sudhamony, S & Nandakumar, K & P J, Binu & Niwas S, Issac. (2008). Telemedicine and telehealth services for cancer-care delivery in India. Communications, IET. 2. 231 - 236. 10.1049/iet-com:20060701
- [2] D. Chauhan, "An efficient data mining classification approach for detecting lung cancer disease," *2016 Int. Conf. Commun. Electron. Syst.*, pp. 1–8.
- [3] R. Radha, "Using K-Means Clustering Technique To Study Of Breast Cancer," *2014 World Congr. Comput. Commun. Technol.*, pp. 211–214, 2014.
- [4] D. B. Fogel et al., "Using Evolutionary Computatio to Learn About Detecting Breast Cancer," pp.1749-1754.
- [5] Z. Zhang, H. Zhang, and R. C. B. Jr, "An Application of Artificial Neural Networks in Ovarian Cancer Early Detection," pp. 107–112, 2000.
- [6] M. Krishnan and R. Doraiswami, "Based Evaluation of colonoscopic images," pp. 1710–1713, 2002.
- [7] Seki, Kazuhiro and Javed Mostafa. "A Probabilistic Model for Identifying Protein Names and their Name Boundaries."

- Proceedings. IEEE Computer Society Bioinformatics Conference 2 (2003): 251-8.
- [8] V. Gautam and K. Ismo, "Outlier Detection Using k-Nearest Neighbour Graph," pp. 4–7.
- [9] J. Yun, L. Zhanhuai, W. Yong, and Z. Long, "Joining Associative Classifier for Medical Images," no. 60373108.
- [10] B. Hill and L. Liverpool, "Extending Genetic Programming for multi-class classification by combining K-nearest neighbour " Liang Zhang, Lindsay B. Jack and Asoke K. Nandi, Senior Member, IEEE Signal Processing and Communications Group, Department of Electrical Engineering and Electronics," pp. 349–352, 2005.
- [11] M. Chen, H. Liao, and C. Huang, "Predicting Breast Tumour via Mining DNA Viruses with Decision Tree," pp. 3585–3589, 2006.
- [12] S. Abou, M. W. Fakhr, A. Farag, and A. Mohamed, "Automatic Diagnosis of Liver Diseases from Ultrasound Images," pp. 313–319.
- [13] Z. Chen, J. L. Ψ, and L. Wei, "Maker Gene Identification : a Multiple Kernel Support Vector Machine Approach," no. 1, pp. 276–279, 2007.
- [14] A. K. Alqallaf and A. H. Tewfik, "Maximum likelihood principle for DNA copy number analysis," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 2009, pp. 501-504.doi: 10.1109/ICASSP.2009.4959630
- [15] I. El Naqa, "The role of quantitative PET in predicting cancer treatment outcomes," no. 4, 2014.
- [16] Y. Omar, A. Tasleem, M. Pasquier, and A. Sagahyroon, "LungCancerPrognosisSystemusingData Mining Techniques," vol. 5, no. Biotech, pp. 361–368, 2018.
- [17] R. Kaviarasi, "Recognition and Anticipation of cancer and non- Cancer Prophecy using Data Mining Approach," *2016 Int. Conf. Emerg. Trends Eng. Technol. Sci.*, pp. 1–4, 2016.
- [18] Y. E. Kurniawati and A. E. Permanasari, "Comparative Study on Data Mining Classification Methods for Cervical Cancer Prediction Using Pap Smear Results," *2016 1st Int. Conf. Biomed. Eng.*, pp. 1–5, 2016.
- [19] S. G. Sutar, "Intelligent Data Mining Technique of Social Media for Improving Health Care," pp. 1356–1360, 2017.
- [20] M. T. Scholar, "using Data mining classification Techniques," *2017 Int. Conf. Intell. Sustain. Syst.*, no. Iciss, pp. 533–538, 2017.
- [21] B. Kaur, "Convolutional Neural network," no.Access, pp. 855–859, 2017.
- [22] I. Conference and I. Systems, "Performance Analysis of Mammogram," pp. 1–5, 2017.
- [23] F. Harirchi, P. Radparvar, H. A. Moghaddam, F. Dehghan, and M. Gita, "Two-Level Algorithm for MCs detection in mammograms using Diverse- Adaboost-SVM," pp. 269–272, 2010.
- [24] P. Hamsagayathri, "Priority-Based Decision Tree Classifier for Breast Cancer Detection," pp. 4–9, 2017.
- [25] S. Shandilya and C. Chandankhede, "Survey on Recent Cancer Classification Systems for Cancer Diagnosis," pp. 2590–2594, 2017.
- [26] G. Tamilmani and S. Sivakumar, "A Survey on Various Data Mining Methods for DetectingCancer Cells," vol. 34, no. August, pp. 242 245, 2017.
- [27] Bharati, Subrato & Atikur Rahman, Mohammad &Podder, Prajoy. (2018). Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA. 10.1109/CEEICT.2018.8628084.
- [28] P. Singhal, "Artificial Neural Network for Prediction of Breast Cancer," *2018 2nd Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud) (I-SMAC)I-SMAC (IoT Soc. Mobile, Anal. Cloud) (I-SMAC)*, 2018 2nd Int. Conf., pp. 464–468, 2018.
- [29] A. Sanjay, H. V. Nair, S. Murali, and K. S Krishnaveni, "Method On Real-Time Data," *2018 Int. Conf. Adv. Comput. Commun. Informatics*, pp. 2437–2440, 2018.
- [30] D. S. Jacob, R. Viswan, V. Manju, L. Padma Suresh, and S. Raj, "A Survey on Breast Cancer Prediction Using Data mining techniques," *2018 Conf. Emerg. Devices Smart Syst.*, no. March, pp. 256–258, 2018.
- [31] S. Bakken and T. A. Koleck, "Big Data Challenges from a Nursing Perspective," pp. 3–16, 2019.