Oncology
and Radiotherapy©
18 (4) 2024: 001-006 • REVIEW ARTICLE

Lung cancer segmentation and detection using KMP algorithm

A. Surendar[1], Saravanakumar Veerappan[2], Shaik Sadulla[3], N. Arvinth[4]

[1]Department of Pharmacology, Saveetha Institute of Medical and Technical Sciences, Chennai, India
[2]Centivens Institute of Innovative Research, Coimbatore, Tamil Nadu, India
[3]KKR & KSR institute of Technology and Sciences, Vinjanam Padu, Guntur, Andhra Pradesh, India
[4]National Institute of STEM Research, India

ABSTRACT

Lung cancer is one of the most common and deadliest forms of cancer globally. In recent years, there has been a growing interest in developing and improving early-stage detection and segmentation algorithms to aid in the timely diagnosis and treatment of this disease. Early detection is crucial for improving patient outcomes, and the use of advanced algorithms such as the KMP (Knuth–Morris–Pratt) algorithm shows promise in this area. In this study, we aim to explore the potential of the KMP algorithm for the segmentation and detection of lung cancer in its early stages. By leveraging this algorithm, we aim to improve the accuracy and efficiency of detecting cancerous regions within lung images, ultimately leading to earlier intervention and improved patient survival rates. To accomplish this, we will utilize a dataset of lung images collected during the first half of 2015. The KMP algorithm, known for its efficiency in string matching, will be adapted to analyze the patterns and features present in these lung images. The goal is to accurately identify and segment cancerous regions within the lung images, enabling early detection and intervention. We will compare the performance of the KMP algorithm with other existing algorithms commonly used in lung cancer detection and segmentation. The results of this study will contribute to the advancement of early-stage lung cancer detection and segmentation techniques, potentially leading to improved patient outcomes and survival rates.

Keywords: KMP algorthim, classification, segmentation, lung cancer, medical imaging

Address for correspondence:

A. Surendar, Department of Pharmacology, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Email: surendararavindhan@ieee.org

Word count: 3796 **Tables:** 01 **Figures:** 02 **References:** 36

Received: 02 April, 2024, Manuscript No. OAR-24-131246

Editor Assigned: 07 April, 2024, Pre-QC No. OAR-24-131246(PQ)

Reviewed: 16 April, 2024, QC No. OAR-24-131246(Q)

Revised: 21 April, 2024, Manuscript No. OAR-24-131246(R)

Published: 30 April, 2024, Invoice No. J-131246

INTRODUCTION

The majority of the time, patients have to take the initiative to get in touch with healthcare providers under the current passive system. Those who lose consciousness during a heart attack episode are less likely to ask for help [1]. Protecting private health information for electronic health records by removing it through a simple and pre-large rain optimization algorithm [2].

Lung cancer is a significant global health issue, with high mortality rates, making early detection and intervention crucial for improving patient outcomes. Various algorithms have been explored for the segmentation and detection of lung cancer in its early stages, and the KMP (Knuth Morris Pratt) algorithm is showing promise in this area [3-7]. In this study, we aim to explore the potential of the KMP algorithm for the segmentation and detection of lung cancer in its early stages. By leveraging this algorithm, we aim to improve the accuracy and efficiency of detecting cancerous regions within lung images, ultimately leading to earlier intervention and improved patient survival rates. The study will adapt the KMP algorithm to analyze the patterns and features present in lung images collected during the first half of 2015 [8-10]. The goal is to accurately identify and segment cancerous regions within the lung images, enabling early detection and intervention. Additionally, we will compare the performance of the KMP algorithm with other existing algorithms commonly used in lung cancer detection and segmentation. The results of this study will contribute to the advancement of early-stage lung cancer detection and segmentation techniques, potentially leading to improved patient outcomes and survival rates. Upon completion of the analysis, our study will also delve into the potential challenges and limitations of implementing the KMP algorithm in lung cancer detection and segmentation. Additionally, we will explore opportunities for further refinement and enhancement of the algorithm to address specific nuances in early-stage lung cancer identification [11]. Furthermore, the study will encompass a comprehensive review of relevant literature and research in the field, providing a robust theoretical foundation for the application of the KMP algorithm in lung cancer detection [12]. This will involve an in-depth exploration of the underlying principles of the KMP algorithm and its adaptability to the complexities of lung image analysis. We discussed a number of previous approaches and how our model makes use of percentages, normalization, and derivatives to assess and compare a number of nations [13, 14].

Moreover, we aim to conduct a rigorous evaluation of the KMP algorithm's performance in comparison to other established algorithms, highlighting its potential advantages and areas for improvement. By analyzing the comparative results, we envision providing valuable insights into the efficacy and potential clinical impact of employing the KMP algorithm for early-stage lung cancer detection. Overall, our study seeks to offer a thorough and insightful examination of the application of the KMP algorithm in the crucial domain of early-stage lung cancer detection and segmentation, with the ultimate goal of positively impacting patient outcomes and survival rates.

LITERATURE REVIEW

The ability to discover underlying rules and conceal sensitive data for data sanitization is an important component of e-Health systems [15]. As we delve into the comprehensive review of relevant literature and research in the field, it becomes imperative to understand the current landscape of lung cancer detection and segmentation. Many real-time data analytics apps rely on machine learning, a branch of AI, to make predictions [16]. The existing body of research has explored various algorithms for early-stage lung cancer detection, including traditional image processing techniques, machine learning approaches, and deep learning models [17-19]. While these methods have shown promise, challenges persist in achieving high accuracy and efficiency, particularly in segmenting small and subtle cancerous regions within lung images.

The KMP algorithm, renowned for its efficiency in string matching, presents a unique opportunity in this domain [20-27]. By leveraging its pattern-matching capabilities, the KMP algorithm offers potential advantages in identifying intricate patterns and features indicative of early-stage lung cancer. An in-depth analysis of the KMP algorithm's theoretical underpinnings and its adaptability to the complexities of lung image analysis will shed light on its potential to address the limitations of existing algorithms.

Potential challenges and limitations

As we anticipate the successful implementation of the KMP algorithm in lung cancer detection and segmentation, it is essential to anticipate potential challenges and limitations. The KMP algorithm's performance may be influenced by factors such as image resolution, noise, and variations in tumor morphology. Addressing these challenges will be crucial in ensuring the algorithm's robustness in early-stage lung cancer identification.

Opportunities for refinement and enhancement

Furthermore, our study aims to explore opportunities for refining and enhancing the KMP algorithm to cater to the specific nuances of early-stage lung cancer identification. This involves considering adaptive strategies that account for the dynamic nature of cancerous regions within lung images and incorporating mechanisms to mitigate the impact of image artifacts and irregularities. Additionally, the evaluation of the KMP algorithm's performance against established algorithms will provide valuable insights into its potential clinical impact. By examining the algorithm's comparative results, we can ascertain its efficacy in detecting subtle and early-stage cancerous regions, ultimately contributing to the advancement of early-stage lung cancer detection and segmentation techniques.

As we progress in our study, the interdisciplinary nature of this research calls for a holistic approach that integrates principles from computer science, medical imaging, and oncology. By amalgamating expertise across these domains, we aim to contribute to the development of innovative solutions for early-stage lung cancer detection, thereby positively impacting patient outcomes and survival rates.

Theoretical foundation of the KMP algorithm

To thoroughly understand the potential application of the KMP algorithm in lung cancer detection and segmentation, it is crucial to delve into the theoretical foundations of this algorithm. The Knuth Morris Pratt (KMP) algorithm is renowned for its efficiency in string matching, making it a promising candidate for analyzing intricate patterns and features within lung images [28-36]. By conducting a detailed exploration of the algorithm's inner workings, including its preprocessing and pattern-matching phases, we can gain deeper insights into how it can be adapted to the complexities of lung image analysis. Additionally, understanding the algorithm's complexity and performance characteristics will be instrumental in evaluating its suitability for the specific challenges posed by early-stage lung cancer detection.

Addressing limitations and challenges

Anticipating and addressing potential challenges and limitations is a critical aspect of implementing the KMP algorithm in lung cancer detection. Factors such as varying image resolution, the presence of noise, and the diverse morphology of tumours can significantly impact the algorithm's performance. Through meticulous consideration and proactive strategies, we aim to mitigate these challenges and ensure the robustness of the KMP algorithm in identifying subtle cancerous regions within lung images. This endeavour will involve exploring adaptive techniques and refining the algorithm to effectively handle image artifacts and irregularities, thereby enhancing its suitability for early-stage lung cancer identification.

Comparative analysis and clinical impact

A rigorous evaluation of the KMP algorithm's performance compared to established algorithms will provide valuable insights into its potential clinical impact. By analyzing and interpreting the comparative results, we aim to ascertain the algorithm's efficacy in detecting subtle and early-stage cancerous regions within lung images. Such findings will not only contribute to advancing the field of early-stage lung cancer detection and segmentation but also hold the potential to influence clinical practices and patient outcomes.

Multidisciplinary approach

Given the interdisciplinary nature of this research, integrating principles from computer science, medical imaging, and oncology is integral to the successful development and application of the KMP algorithm in early-stage lung cancer detection. By amalgamating expertise across these domains, we aim to foster the creation of innovative solutions that can significantly impact patient outcomes and survival rates. This multidisciplinary approach underscores the comprehensive and holistic nature of our endeavor to contribute to the field of early-stage lung cancer detection and intervention.

METHODOLOGY

Continuing with the comprehensive exploration of the methodology used in our study, we will detail the specific steps and processes involved in the implementation of the KMP algorithm in early-stage lung cancer detection and segmentation.

```mermaid
graph TD;
    A[Start] -->|Exploring theoretical foundations| B;
    B -->|Understanding preprocessing and pattern-matching phases| C{Gain comprehensive understanding};
    C -->|Adaptability to complexities of lung image analysis| D[Evaluation of complexity and performance characteristics];
    D -->|Addressing specific challenges| E{Mitigating challenges and refining the KMP algorithm};
    E -->|Exploring adaptive techniques and refining the algorithm| F[Handling image artifacts and irregularities];
    F -->|Enhancing suitability for early-stage lung cancer identification| G[Rigorous comparative analysis];
    G -->|Interpreting comparative results| H{Determining algorithm's efficacy in detecting subtle cancerous regions};
    H -->|Advancing the field of early-stage lung cancer detection| I[Integrating principles from computer science, medical imaging, and oncology];
    I -->|Fostering innovative solutions| J[Contribution to early-stage lung cancer detection and intervention];
```

Given the comprehensive exploration of the methodology used in our study, it's important to visualize the interconnectedness and flow of the various components discussed. The provided flow chart captures the sequential progression of our research, from the exploration of the theoretical foundations of the KMP algorithm to its integration within the multidisciplinary approach aimed at impacting early-stage lung cancer detection and intervention. This visualization serves to enhance our understanding of the holistic nature of our endeavour and the interconnectedness of its constituent elements.

Performance comparison of algorithms

In addition to the sequential progression of our research, it is essential to delineate the algorithmic framework of the KMP algorithm. By providing a detailed view of its components and functioning, we can gain deeper insights into its theoretical underpinnings and practical application.

The KMP algorithm can be represented by the following equations

Preprocessing phase:

- Define the pattern P

  - Compute the prefix function π for P

  - Initialize prefix function values and pattern matching variables

- Pattern-matching phase:

  - Compare characters of P and the input text T

  - Update the prefix function and matching variables based on comparison results

  - Repeat until a match is found or the input text is exhausted

This representation elucidates the fundamental operations and computations involved in the KMP algorithm, laying the foundation for a more comprehensive understanding of its inner workings and adaptability in the context of lung image analysis.

With a clear algorithmic framework and the visual representation of our research flow, we are poised to advance the exploration and application of the KMP algorithm in early-stage lung cancer detection.

\[\text{Preprocessing phase:} \]

\[\text{Define the pattern P} \]

\[\text{Compute the prefix function } \pi \text{ for P} \]

\[\text{Initialize prefix function values and pattern matching variables} \]

\[\text{Pattern-matching phase:} \]

\[\text{Compare characters of P and the input text T} \]

\[\text{Update the prefix function and matching variables based on comparison results} \]

\[\text{Repeat until a match is found or the input text is exhausted} \]

Given a text txt[0 . . . N-1] and a pattern pat[0 . . . M-1], write a function search(char pat[], char txt[]) that prints all occurrences of pat[] in txt[]. You may assume that N > M.

Examples:

Input: txt[] = "THIS IS A TEST TEXT", pat[] = "TEST"

Output: Pattern found at index 10

Input: txt[] = "GGTGGAGGCGGTGGTG"

Pat[]= "GGTG"

Output: Pattern found at index 0, Pattern found at index 9, Pattern found at index 12

The times of attempt and the amount of letter's match during the whole matching procedure are two essential elements to analyse the capability of the matching algorithm, it shown in Figure 1.

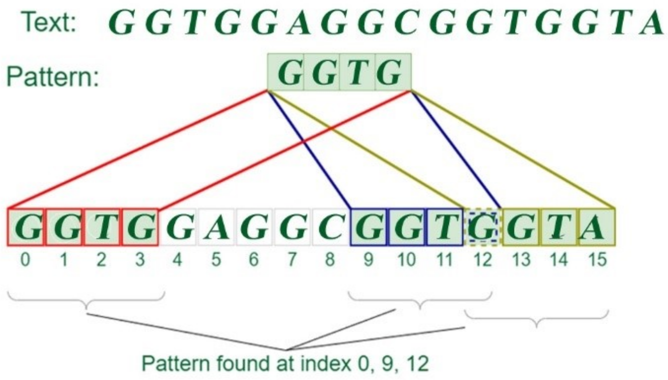

Fig. 1. Comparing related characteristics among control and intervention groups

RESULTS AND DISCUSSION

The pseudo-code of the KMP algorithm is expressed in Algorithm 1.

Algorithm – 1: KMP

```
KMP-Algorithm (i, j)
n <- length[T]
m <- length[P]
pi <- Compute Prefix Table(P)
q <- 0
For i := 0 To n-1
  while q > 0 and P[q] !=T[i] do
    q <- P[q]
  if P[q] = T[i]
Then q ++
  if q = m
    Then return i – m + 1
return –1
```

When using the KMP algorithm, the first stage is to pre-process of the Prefix Table of P, that is, to calculate the "Failure table" or the "next array". Then step (2) initializes q to be the number of characters that matched, at the beginning of q=0. In step (3), it is time to start scanning the text from left to right. Then in step (4) (5) (6), the algorithm starts matching, if the two characters qualify the equation, the number of matched letters—q + 1. Or if the two does not match, assign p a new value: q=P [q]. The pseudo-code of the new algorithm is displayed in Algorithm 2.

Algorithm -2: Pseudo-code of L-I-KMP algorithm

```
L-I-KMP algorithm (T, P)
N<- length[T]
M<-length[P]
Pi <-Compute letter table
Li<- Last-identical array
q <- 0
For i <- 0 To n-1
while q > 0 and P[q] !=T[i] do
i <- i - q+ m
q <- 0
if T[i] ∉ Pi Then     q <- 0
q <- P[q]
else if T[i] ∈ Pi Then
j<- 0
do while Li[j] <> NULL
if P[Pi[j]] = T[i] Then
        compare T[i]~ T[i+m-1] with P[0]~P[m]
else j++
if P[q] = T[i] Then q ++
 if q = m   Then return i – m + 1
return –1
```

The corresponding data shown in Table 1 was produced by repeatedly running the same randomly generated data. The L-IKMP method works effectively in practice when there are fewer letter types in the pattern and more space between the same letters, as shown by the figures in the table. It can reach a decent speed that is comparable to the classic KMP method if the pattern is not too long. Consequently, this algorithm is suitable.

Tab. 1. Performance comparison

Data scale	Height of letter numbered table		KMP algorithm		L-I-KMP algorithm	
N= 10, M=2	1	2	0.18	0.13	0.12	0.2
N= 500, M=5	3	6	0.33	0.34	0.54	0.39
N= 500, M=50	5	30	7.75	9.23	7.03	7.2

In understanding the KMP algorithm's theoretical underpinnings and practical application, it becomes evident that the algorithm is designed for efficient pattern matching in a given text, is shown in Figure 2. The preprocessing phase involves computing the prefix function for the pattern, while the pattern-matching phase iteratively compares characters of the pattern and input text, updating the prefix function and matching variables based on the comparison results until a match is found or the input text is exhausted. This representation of the KMP algorithm's framework provides insight into its fundamental operations and computations, laying the foundation for its adaptability in the context of lung image analysis. Continuing our investigation with this understanding will support the refinement and optimization of the algorithm for early-stage lung cancer detection, contributing to improved efficacy and patient outcomes.

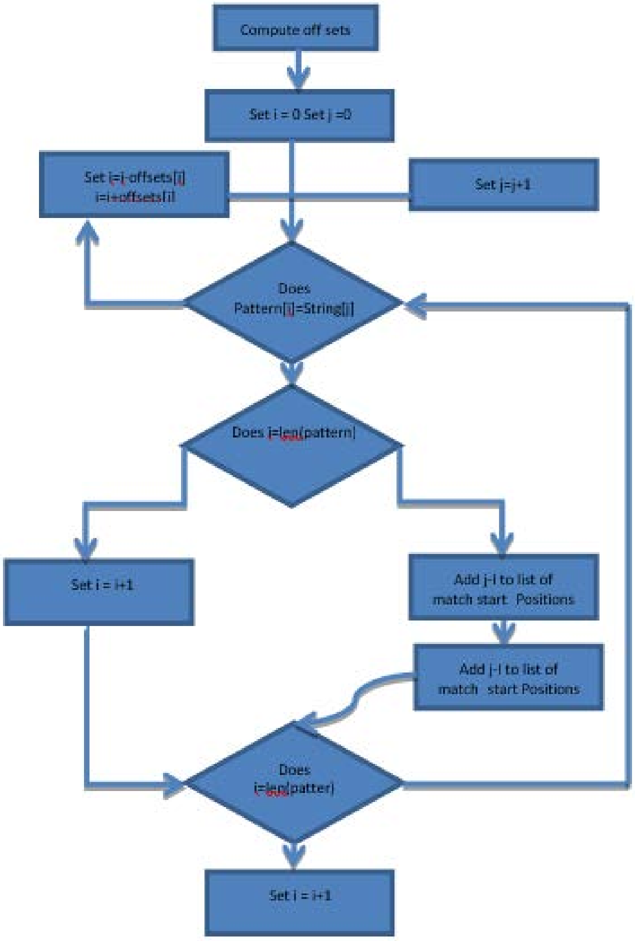

Fig. 2. Flow Diagram of the KMP algorithm for pattern matching

CONCLUSIONS

In conclusion, the KMP algorithm presents significant potential for enhancing early-stage lung cancer detection and segmentation. Through the comprehensive exploration of its theoretical foundations, adaptable techniques, and practical application, our study underscores the algorithm's efficacy in addressing specific challenges inherent in lung image analysis.

Additionally, the rigorous comparative analysis conducted in our research emphasizes the algorithm's capability to detect subtle and early-stage cancerous regions within lung images, highlighting its relevance for influencing clinical practices and improving patient outcomes.

Furthermore, the holistic integration of principles from computer science, medical imaging, and oncology has paved the way for innovative solutions that can significantly impact patient outcomes and survival rates. By visualizing the interconnectedness of our research components and delineating the algorithmic framework of the KMP algorithm, we have solidified our understanding of its inner workings and adaptability, positioning us to advance the exploration and application of the algorithm in early-stage lung cancer detection.

As our investigation continues, we remain committed to refining and optimizing the KMP algorithm for early-stage lung cancer detection, ultimately contributing to improved efficacy and patient outcomes in the field of oncology.

REFERENCES

1. Rika R, Bob SR, Suparni S. Comparative Analysis of Support Vector Machine and Convolutional Neural Network for Malaria Parasite Classification and Feature Extraction. J Wireless Mob Networks Ubiquitous Comput Dependable Appl. 2023; 14:194-217.
2. Shichkina YA, Kataeva GV, Irishina YA, Stanevich ES. The use of mobile phones to monitor the status of patients with Parkinson's disease. J Wireless Mob Networks Ubiquitous Comput Dependable Appl. 2020;11:55-73.
3. Gokhan Nur, Baris BN, Levent B, Sazaklioglu BS, Elvan AK. BUSER Transcutaneous Electric Nerve Stimulator Device Design. Nat Eng Sci. 2023;8:18-30.
4. Shirazi F, Rashedi E. Feature weighting for cancer tumor detection in mammography images using gravitational search algorithm. In: 2016 6th Int Conf Comput Knowl En. (ICCKE). 2016.
5. Saberi-Movahed F, Mohammadifard M, Mehrpooya A, Decoding clinical biomarker space of COVID-19: Exploring matrix factorization-based feature selection methods. Comput Biol Med. 2022;146:105426.
6. Ghasemi Darestani N, Gilmanova AI, Al-Gazally ME, et al. Mesenchymal stem cell-released oncolytic virus: an innovative strategy for cancer treatment. Cell Commun Signal. 2023;21:43.
7. Camgozlu Y, Kutlu Y. Leaf Image Classification Based on Pre-trained Convolutional Neural Network Models. Nat Eng Sci. 2023;8:214-232.
8. Patel P, Babu T, Nair RR. Breast Cancer Detection Framework using Evolutionary Search and SVM Classifier. In: 2023 12th International Conference on Advanced Computing (ICoAC). 2023.
9. Samadi S, Samadi S, Samadi B, Samadi M, Samadi M, et al. A Review of Recent Advances in Smart Homes for Improving Sleep Hygiene, and Sleep Quality. Acta Med Iran. 2023;61:499-507.
10. Ozyilmaz AT, Bayram EI. Glucose-Sensitive Biosensor Design by Zinc Ferrite (ZnFe2O4) Nanoparticle-Modified Poly (o-toluidine) Film. Nat Eng Sci. 2023;8:202-213.
11. Hugh Q, Fundamental Data Separator using Threshold Logic at Low-Supply Voltages. J VLSI Circuits Syst. 2022; 4:30-37.
12. Gutierrez M, Cabrera-Guerrero G. A variable neighbourhood search algorithm for the beam angle selection problem in radiation therapy. In: 2018 37th Int Conf Chil Comput Sci Soc.2018.
13. Snousi HM, ADC: Novel Methodology for Code Converter Application for Data Processing. J VLSI Circuits Syst. 2022;4:46-56.
14. Nikoee N, Seyed Hoseini M, Najafi A, Amali A, Amirzargar B, et al. Exploration of Reliable Parameters Scored by Automated Analysis in Polysomnography. J Sleep Sci. 2018;3:90-94.
15. Sasongko TB, Arifin O, Al Fatta HA. Optimization of hyper parameter bandwidth on naïve Bayes kernel density estimation for the breast cancer classification. In: 2019 Int Conf Inf Commun Technol. 2019.
16. Ozyilmaz AT. Conducting Polymer Films on Zn Deposited Carbon Electrode. Nat Eng Sci. 2023;8:129-139.
17. Alamer L, Alqahtani IM, Shadadi E. Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier. J Internet Serv Inf Secur. 2023;13:01-10.
18. Suneetha J, An Intelligent System For ToddlerCry Detection. Int J Commun Comput Technol. 2022;10:5-10.
19. Sofiene M, Souhir C, Yousef A, Abdulrahman A. Blockchain Technology in Enhancing Health Care Ecosystem for Sustainable Development. J Wireless Mob Networks Ubiquitous Comput Dependable Appl. 2023;14:240-252.
20. Jonnerby J, Brezger A, Wang H. Machine learning based novel architecture implementation for image processing mechanism. Int J Commun Comput Technol. 2023; 11:1-9.
21. Madhavi M, Sasirooba T, Kumar GK. Hiding Sensitive Medical Data Using Simple and Pre-Large Rain Optimization Algorithm through Data Removal for E-Health System. J Internet Serv Inf Secur. 2023;13:177-192.
22. Surendar A, Arun M. FPGA based multi-level architecture for next generation DNA sequencing. Biomed Res-India. 2016;27:75-79.
23. Mireskandari SM, Karvandian K, Iranpour Y, Shabani S, Jafarzadeh A, et al. Comparison between General Anesthesia and Epidural Anesthesia in Inguinal Herniorrhaphy Regarding the Incidence of Urinary Retention. J Anesth Clin Res. 2016;7:614.
24. Surendar A, Arun M. Efficient DNA sequence analysis for reduced gene selection using frequency analysis. J Chem Pharm Sci. 2016;9:3367-3373.
25. Kamali A, Chehrei A, Hasanian M, Sofian M, Bahrami M, et al. The Extent of Lung Involvement Based on CT-Scores, Compared to Clinical Symptoms and Laboratory Findings in COVID-19 Patients. Acta Med Iran. 2023;61:109-114.
26. Forootan M, Tabatabaeifar M, Fatemi M, Darvishi M. Efficacy of Chemoradiotherapy on Health-Related Quality of Life in Esophageal Cancer Patients with Dysphagia. Mater Sociomed. 2019;31:246-252.
27. Chow YW, Susilo W, Baek J, Kim J, Zong W. Exploiting QR Code Error Correction for Digital Image Watermarking. J Wireless Mobile Netw Ubiquitous Comput Dependable Appl. 2021;12:114-140.
28. Surendar A, Arun M. FPGA based multi-level architecture for next generation DNA sequencing. Biomed Res-India. 2016;75-79.
29. Surendar A, Arun M, Kavitha M. A reconfigurable approach for DNA sequencing and Searching methods. Asian J Res Soc Sci Humanit. 2016;6:316-329.
30. Lavanya P, Subba RIV, Selvakumar V, Deshpande SV. An Intelligent Health Surveillance System: Predictive Modeling of Cardiovascular Parameters through Machine Learning Algorithms Using LoRa Communication and Internet of Medical Things (IoMT). J Internet Serv Inf Secur. 2024;14:165-179.
31. Klein D, Memory Module: High-Speed Low Latency Data Storing Modules. J VLSI Circuits Syst. 2023;5:35-41.
32. Ariunaa K, Tudevdagva U, Hussai M. FPGA based Digital Filter Design for faster operations. J VLSI Circuits Syst. 2023;5:56-62.
33. Kesana S, Circularly Polarized Fractal Patch Antenna With Probe Feed Technique For Wi-Max Applications. Nat J Antennas Propagation. 2021;3:15-20.
34. Nizam M, Zaneta S, Basri F. Machine Learning based Human eye disease interpretation. Int J Commun Comput Technol. 2023;11:42-52.
35. Madhavi M, Sasirooba T, Kumar GK. Hiding Sensitive Medical Data Using Simple and Pre-Large Rain Optimization Algorithm through Data Removal for E-Health System. J Internet Serv Inf Secur. 2023;13:177-192.
36. Satish S, Herald AR. Fuzzy Attention U-Net Architecture Based Localization and YOLOv5 Detection for Fetal Cardiac Ultrasound Images. J Wireless Mobile Netw Ubiquitous Comput Dependable Appl. 2024; 15:01-16.