High-Performance Computing System

Manichithra PS[1], Dr. Sanjay Kumar Tiwari[2]

[1]Research Scholar, Sunrise University, Alwar
[2]Associate Professor, Sunrise University, Alwar

Abstract

Nowadays, researchers can't do their jobs without high performance computing (hpc). Research on how cloud and hpc systems compare in terms of their respective application profiles applications for high-performance computing in the cloud computational power at a high level recognizes speech in progress high-performance computational algorithm supercomputing on nvidia tesla as a consequence of computers' inability to mimic human thought processes, hpc is often replaced with lpc in situations when the former is called for. The use of high-performance computing in the cloud.

Keywords: Parallel Computing, High-Performance, Computing System, HPC and Computer.

1. INTRODUCTION

One can assume that a computer programmer is doing high-performance computing (HPC) if it is able to complete a job more quickly than a person could. However, this is erroneous since computers and humans approach problem-solving in fundamentally different ways. Since people tend to be "mono-task" focused, it may be challenging to study, design, and develop a multi-task nature tool, especially when considering the approach paradigm. This paper draws heavily on the author's experience with the IST Gravity Group1 Computer Cluster at Baltasar Sete S'ois for both inspiration and examples of practical use. 2 Baltasar is a member of Professor Vitor Cardoso's "DyBHo" ERC Starting Grant to investigate and comprehend dynamical black holes. Baltasar, a Cactus-based solution, was created to address a narrow need. 3 In order to do this, we needed at least 500MB/s of network storage capacity, 200 CPUs, and 4GB of memory per processor. It was intended for a particular use case, but it performed so well that the team is now considering it a general-purpose cluster. The most basic concepts of computing that are necessary for grasping the parallel computation paradigms are first covered in this overview. We also discuss some of the primary resources required for an effective HPC process. Finally, we make an effort to simplify the "beast" of using a shared computer cluster by providing illustrative instances.

Computers are multi-use machines that may be instructed to perform a limited range of arithmetic or logic tasks. The flexibility of a computer's processing pipeline allows it to address a wide variety of problems. A layperson in science-computer research doesn't need to completely comprehend the complicated architecture of the system, despite the system's seeming complexity. A computer is broken down into its processor, memory, and storage space in order to better grasp its function in high-performance computing. A programme may be compared to a workflow instruction manual, and a processor can be thought of as a worker. Instructions in a programme are executed by a processor in the order specified by the programme. Therefore, it follows that a computer processor can do only a single task at a time. The data set the software requires to perform its tasks is kept in memory, a volatile storage medium. Despite its instability, it may be accessed in a flash. There's no restriction on the types of datasets that may be stored in memory; they can all be either freshly produced by the software or imported from a data repository.

A data store is a repository for datasets that may be accessed at any time. However, this comes at a cost, since its access time is far longer than memory's. These days, most computers have many processors. Tightly connected computer system with two or more processing units (Several Processors) that share primary memory and peripherals and may run multiple applications in parallel. When considering a single application running on several processors concurrently, the benefits of such an architecture may seem evident. However, the reality is not that black and white. Although two or more processors sharing memory may work together to solve an issue, this coordination must be handled on a case-by-case basis. In the next section, "Parallel Computing," we'll go further into this topic.

Having access to top-tier computing resources provides modern businesses a leg up and makes it possible to transform ideas into products. Altair's cutting-edge HPC solutions provide a streamlined, user-friendly setting in which even the most complex workloads can be visualized, orchestrated, analyzed, and optimized to produce the next great breakthrough.

If you're looking for HPC solutions, go no further than Altair. Altair's sophisticated scheduling for CPUs and GPUs, along with optimizations for software licencing, I/O, and storage needs, are just a few of the many key characteristics of contemporary infrastructure that our clients depend on us to optimize. From the greatest throughput to the most highly parallel processes, our scalable HPC solutions span both local and remote resources. Our enterprise-level solutions are backed by a worldwide team of professionals to guarantee you maximum efficiency and effectiveness. No matter whether you're a Fortune 500 giant or a brand-new start-up, Altair's HPC technology and world-class employees will assist your team speed up time-to-market, reduce expensive downtime, maximize return on investment (ROI), and boost profits.

2. LITERARTURE REVIEW

Tziortziotis, I.; Laskaratos, F.-M.; Coda, S. (2021): Capsule endoscopy (CE) has been increasingly utilised in recent years as a minimally invasive tool to investigate the whole gastrointestinal (GI) tract and a range of capsules are currently available for evaluation of upper GI, small bowel, and lower GI pathology. Although CE is undoubtedly an invaluable test for the investigation of small bowel pathology, it presents considerable challenges and limitations, such as long and laborious reading times, risk of missing lesions, lack of bowel cleansing score and lack of locomotion. Artificial intelligence (AI) seems to be a promising tool that may help improve the performance metrics of CE, and consequently translate to better patient care. In the last decade, significant progress has been made to apply AI in the field of endoscopy, including CE. Although it is certain that AI will find soon its place in day-to-day endoscopy clinical practice, there are still some open questions and barriers limiting its widespread application. In this review, we provide some general information about AI, and outline recent advances in AI and CE, issues around implementation of AI in medical practice and potential future applications of AI-aided CE.

Choi J, Shin K, Jung J, Bae H J, Kim D H, Byeon J S, and Kim N (2020) The high prevalence of tuberculosis (TB) in Indonesia puts Indonesia in the second-highest national TB prevalence in the world after India. This high prevalence can cause a failure to deliver medical treatments to TB patients, which is exacerbated by the disproportionate distribution of doctors in Indonesia. To address this issue, an artificial intelligence (AI) system is necessary to help doctors in screening a large number of patients in a short time. However, to develop a robust AI for this purpose, we need a large dataset. This study aims to develop a database system for storing TB sputum sample images, which can be used as the dataset to train an AI system for TB detection. The developed system can help doctors and health workers to manage the images during their daily job.

Crowley C, Guitron S, Son J, and Pianykh OS (2020) Limited resources and increased patient flow highlight the importance of optimizing healthcare operational systems to improve patient care. Accurate prediction of exam volumes, workflow surges and, most notably, patient delay and wait times are known to have significant impact on quality of care and patient satisfaction. The main objective of this work was to investigate the choice of different operational features to achieve (1) more accurate and concise process models and (2) more effective interventions. To exclude process modelling bias, data from four different workflows was considered, including a mix of walk-in, scheduled, and hybrid facilities. A total of 84 features were computed, based on previous literature and our independent work, all derivable from a typical Hospital Information System. The features were categorized by five subgroups: congestion, customer, resource, task and time features. Two models were used in the feature selection process: linear regression and random forest. Independent of workflow and the model used for selection, it was determined that congestion feature sets lead to models most predictive for operational processes, with a smaller number of predictors.

De Lange T, Halvorsen P, and Riegler M (2018) Assisted diagnosis using artificial intelligence has been a holy grail in medical research for many years, and recent developments in computer hardware have enabled the narrower area of machine learning to equip clinicians with potentially useful tools for computer assisted diagnosis (CAD) systems. However, training and assessing a computer's ability to diagnose like a human are complex tasks, and successful outcomes depend on various factors. We have focused our work on gastrointestinal (GI) endoscopy because it is a cornerstone for diagnosis and treatment of diseases of the GI tract. About 2.8 million luminal GI (esophageal, stomach, colorectal) cancers are detected globally every year, and although substantial technical improvements in endoscopes have been made over the last 10-15 years, a major limitation of endoscopic examinations remains operator variation. This translates into a substantial inter-observer variation in the detection and assessment of mucosal lesions, causing among other things an average polyp miss-rate of 20% in the colon and thus the subsequent development of a number of post-colonoscopy colorectal cancers. CAD systems might eliminate this variation and lead to more accurate diagnoses. In this editorial, we point out some of the current challenges in the development of efficient computer-based digital assistants. We give examples of proposed tools using various techniques, identify current challenges, and give suggestions for the development and assessment of future CAD systems.

Borgli Hanna, Vajira Smedsrud, and Pia Hicks (2020) Artificial intelligence is currently a hot topic in medicine. However, medical data is often sparse and hard to obtain due to legal restrictions and lack of medical personnel for the cumbersome and tedious process to manually label training data. These constraints make it difficult to develop systems for automatic analysis, like detecting disease or other lesions. In this respect, this article presents HyperKvasir, the largest image and video dataset of the gastrointestinal tract available today. The data is collected during real gastro- and colonoscopy examinations at Bærum Hospital in Norway and partly labeled by experienced gastrointestinal endoscopists. The dataset contains 110,079 images and 374 videos, and represents anatomical landmarks as well as pathological and normal findings. The total number of images and video frames together is around 1 million. Initial experiments demonstrate the potential benefits of artificial intelligence-based computer-assisted diagnosis systems. The HyperKvasir dataset can play a valuable role in developing better algorithms and computer-assisted examination systems not only for gastro- and colonoscopy, but also for other fields in medicine

3. COMPARISON APPLICATION PROFILE OF CLOUD COMPUTING AND HIGH-PERFORMANCECOMPUTING

1. Multiple characteristics are needed in a cloud computing application.
2. The applications are not the parallel applications, or the thread applications.
3. Third, most software doesn't have high demands on system resources like RAM or processing power.
4. The applications rarely execute I/O (input/ output) instructions of computing.
5. the software is self-healing. If a programmer fails for whatever reason, the user will not notice since it may simply restart. Multiple characteristics are needed in a high-performance computing application.

- One common kind of HPC application is the serial application. Some forms of information exchange between the application's many processes are supported.
- the amount of information sent between processes may be negligible at times.
- there may be several data transmissions between processes at certain moments.
- Some software is designed to operate on a single node, and this includes serial apps and thread applications. One such example is the BLAST protocol (Bell Labs Layered Space time).
- Many IO operations may be carried out by both serial and parallel applications. As an example, FEA software packages like Ansys, Abaqus, and Nastran, etc.
- 6 A checkpoint, or a snapshot of the current computation, may be generated by certain programmers. When programmed failures occur for whatever reason, the system may start again from the most recent save point rather than from scratch. However, this functionality is not included in all programmers.

It would seem that high-performance and cloud-based apps are completely different. In spite of their differences, cloud apps and high-performance applications nonetheless have certain commonalities.

The high-performance computing programmed BLAST, for instance, does not rely on the nodes' ability to connect to one another or exchange data with one another. It also doesn't call for a lot of IO instructions to be carried out. The unique qualities of these traits make them excellent candidates for cloud computing. Finding such specialized programmers is challenging since most parallel applications nowadays can be executed entirely on a single node. Applications may be deployed in the cloud so long as the data set can be stored in the nodes without requiring frequent switches.

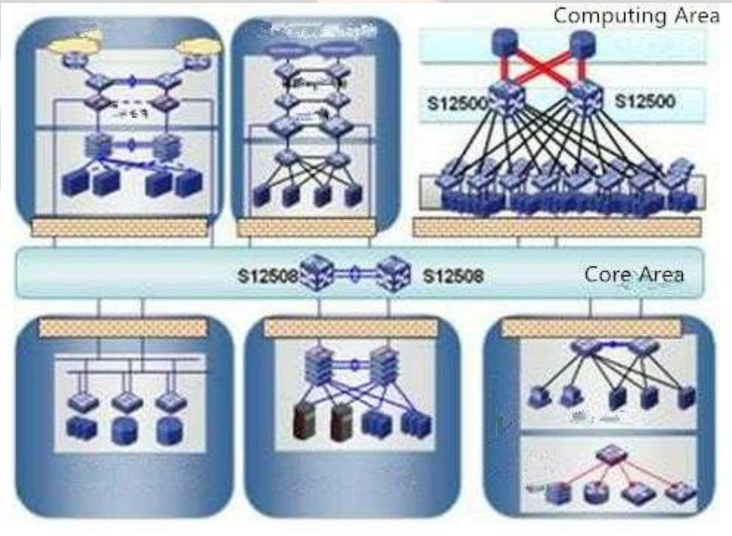
For the proper functioning of cloud-based programmers and data sets, high-performance computing capabilities are a must. It is essential, first and foremost, that the high-performance computing application and its data set reside on a single node. The second need for HPC is that the application has low input/output requirements. Third, in the event of a system failure, the high-performance computing programmed must execute quickly and be able to generate a checkpoint.

High Performance Computing Deployments in Cloud Computing Environments

High-performance computing has the potential to become a niche service offered by cloud providers to those who need it most: internet users. As such, HPC may be thought of as a subset of cloud computing's data center. Cloud computing has many advantages, but HPC is still extremely different. The following is a detailed description of the expected network and security parameters. High-performance computing is one area where a server cluster really shines. The server must be a system in order for this application to function properly, as seen below.

Before everything else, the high-performance computer cluster cannot support isomerism. Second, there are stringent demands placed on the quality of communication services provided inside the high-performance computing cluster. As a result, the system can't let additional users to use the commercial service channel. Finally, the high-performance computer cluster has top-notch security. Separation between the access area and the computing area, both logically and physically, is necessary. Fourthly, the virtual machine is generally nonexistent. As a result, the high-performance computing cluster's internal communication channels are less congested. In the end, the computational nodes' performance must be stellar to fulfil the specs.

Figure 1. Internet Solution of the Integration of High-Performance Computing and Cloud



Computing

The secure partition plan and the safe deployment of ends to ends are two essential strategies for the security solution of the computer system. The network of a supercomputer center performs a wide variety of functions and contains a wide variety of equipment, all of which are susceptible to varying degrees of attack. In order to provide the highest level of safety, the supercomputer network is segmented into multiple zones. High levels of data security and data service may be maintained thanks to the logical separation of the zones.

field of high-performance computing the logical separation of users' access to the supercomputer center may be preserved by deploying security measures between the endpoints of the network. The end-to-end separation of functions may improve data center safety and service quality.

Increased network dependability in supercomputers is possible with the help of IRF II (the next generation of Intelligent Resilient Framework) technology. Additionally, IRF II might lessen the time spent configuring and fixing your system. With IRF II in place, it is possible to:

1. To boost network performance, the L2 and L3 protocols were distributed for processing.

2. Each subset functioning as its own logical fabric. More effective configuration management is possible.

3. The device's software is easily upgradable thanks to the cluster's design.

4. For the high-end devices, multiple devices can be treated as a resource to manage, simplifying the network.

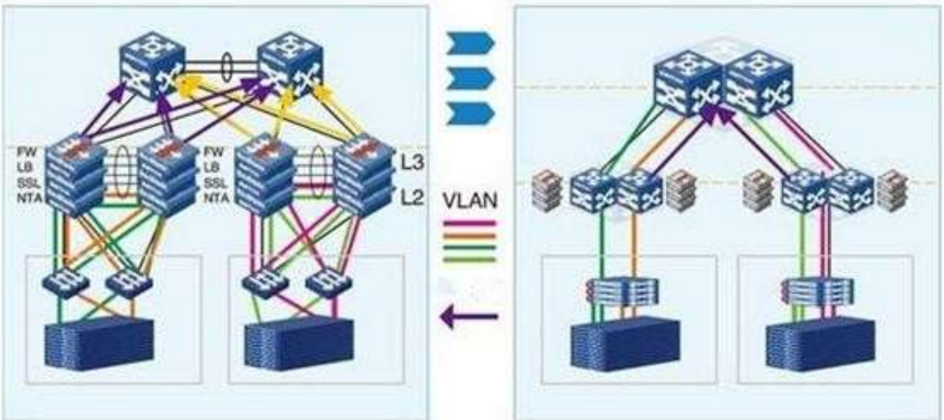


Figure 2. Deployment of the IRF II (The Second Generation of IntelligentResilient Framework)

4. HIGH PERFORMANCE COMPUTING INSPEECH RECOGNITION

Algorithm 1: High Performance Computing

High-performance computation in voice recognition is made possible by this unique approach. With this method, you may take use of a brand-new algorithm developed just for voice recognition. The findings of the experiment show that the aforementioned method yields effective results for precise recognition and matching. This method is helpful for the mission-critical issue when price is less of a concern but precision cannot be compromised. In this method, we utilize the notations and presumptions below. Coefficients discovered using feature extraction methods are stored in a M two-dimensional array. w = Wave File, N = Number of Feature Vectors (After Compression), th = Thread Length (in Size). The components of the retrieved characteristics are given in a compressed form by fun (). The number of separate words uttered is denoted by d. Pattern-matching probabilities are stored in a two-dimensional array denoted by the symbol R. In order to keep track of which pattern-matching strategies are available, we may use the array P to do so. Similarity index k. k is represented by the index value j. The array S maintains the maximum value index for each thread.

The following is the suggested algorithm for high-performance computing:

High_Performance_Computing_Algorithm (wave file)

Step 1: Take input string ‘w’ as wave file.

Step 2: Find out the feature by applying feature extraction technique, store features and design a database.

Step 3: L= length of array M.

Step 4: N = fun (w, L);

Step 5: Create different threads(N);

Step 6: Get th = length thread(N);

Step 7: for k = 1 to d step by one do

{Step 8: for i = 1 to th step by one do

{(a) R[k][i] = Pattern_ matching_ algorithm (P, i);

(b) Apply over all the thread simultaneously;}}

Step 9: for i=1 to th step by one do

{Step 10: for j= 1 to d step by one do

{k=max (R[j][i])

return (j); //where j is base index for value k}

S[i] = j;}

Step 11: Count the maximum occurrence of the index value in the S array;}

For each matched isolated word in the database, this method determines how many times the index values appear in array S. Standard pattern matching algorithms are used to a set of m words if they are all equally likely to be found in the database.

Analysis

Stepwise analysis of the proposed high performance computing algorithm is carried out based on the parameters and assumptions which are used in the design of the algorithm discussed in the previous section.

Step 1: time complexity: O (1).

Step 2: time complexity: depends on the feature extraction technique used.

Step 3: time complexity: O (1) // by direct method.

Step 4: time complexity: depends on the compression technique used.

Step 5: time complexity: O (1).

Step 6: time complexity: O (1) // by direct method.

Step 7: time complexity: O (d*th*p).

Where p is the time complexity of pattern matching algorithm used.

Step 7: time complexity: O (th*d).

Step 8: time complexity: O (th).

Total time complexity = O (d*th*p) … (5.1)

for small vocabulary (d< <p) the order of time complexity is O(th*p).

if n is the no. of threads,

then th = N/n;

if n=1;

then th = N;

Time complexity = O (d*N*p) … (5.2)

Mathematically if n= N then th =1;

Time complexity will be O (d*p)

(HPC) - SUPERCOMPUTING WITH NVIDIA TESLA

The NVIDIA® Tesla™ 20-series is designed from the ground up for high performance

5. CUDA Architecture CUDA is NVIDIA’s

design for parallel computation that significantly boosts speed by taking use of graphics processing units (graphics processing unit). Image and video processing, computational biology and chemistry, fluid dynamics modelling, CT image reconstruction, seismic analysis, ray tracing, and many more applications are just a few of the many that software developers, scientists, and academics are finding CUDA useful for. "Central processing" on the CPU is giving way to "co-processing" on the CPU and GPU. NVIDIA's CUDA parallel computing architecture, which is already shipping in GeForce, ION, Quadro, and Tesla GPUs, represents a sizable installed base for application developers and enables this new computing paradigm.

Products from Elemental Technologies, Motions, and LoiLo, Inc., to name a few, have already been or will soon be accelerated by CUDA in the consumer market. The scientific community has embraced CUDA with open arms. For instance, CUDA has improved the performance of AMBER, a molecular dynamics modelling tool used by over 60,000 scientists from universities and pharmaceutical businesses across the globe to hasten the development of novel drugs.

Acceleration in the use of Tesla GPUs for GPU computing is indicative of CUDA's widespread acceptance. More than 700 GPU clusters have been deployed in Fortune 500 organisations so far, from oil giants Schlumberger and Chevron to financial powerhouse BNP Paribas. NVIDIA's GeForce 400 Series [GF100] (GPU) 2010-03-27) comes equipped with the company's next-generation CUDA architecture (codename: "Fermi"), which was built from the ground up to natively support additional programming languages including C++. When compared to the previous version of Nvidia's Tesla GPU, this one is eight times more powerful in terms of peak double-precision floating-point performance.

Fig 3Tesla GPU Computing Processor The GPU Computing Processor transforms a standard workstation into a personal supercomputer

Fig 4 Tesla Personal Supercomputer the Tesla personal supercomputer delivers cluster level computing performance on your desk - 250 times faster than standard PCs and workstations.

Private, In-House Supercomputer Every scientist and engineer in need of computational resources would have access to them 24/7/365. Boost Your Workstation's Performance with a Cluster An evaluation of a cluster's efficiency in a desktop environment. The combined processing power of four Tesla GPUs is close to 4 Teraflops. Massively Parallel Several Cores Each GPU in the GPU architecture has 240 processing cores that can run thousands of tasks simultaneously. Take Advantage of Multiple GPUs to Tackle Huge Problems Using numerous GPUs, your application may tap into the processing power of thousands of cores, making short work of massively parallel computations. Memory bandwidth of 4 GB/GPU Allows bigger datasets to be stored locally for each CPU, taking advantage of the 102GB/s memory transfer rates and reducing the amount of data that must be sent throughout the system. The results provided by IEEE 754 Floating Point Precision (both single and double precision) are portable and up to par with professional standards. Floating-Point 64-Bit ALUs Math Equipped with 64-bit ALUs, it can handle the granularity of even the most complex programmers. To handle the most difficult computing issues, the GPU Computing System integrates easily into existing business server clusters.

6. CONCLUSION

When dealing with enormous datasets or complex data sets, it is helpful to have a processing system that can do it quickly. High-performance computing (HPC) and cloud computing are two concepts that have been presented to improve service quality. Experiment results demonstrate the efficacy of the aforementioned approach in producing reliable results for accurate identification and matching. If cost isn't as much of a concern as accuracy, this approach may be quite useful for solving a mission-critical problem. Following are some of the notations and assumptions used in this approach.

7. REFERENCES

1. Tziortziotis, i.; laskaratos, f.-m.; coda, s. Role of artificial intelligence in video capsule endoscopy. Diagnostics 2021, 11, 1192. Https://doi.org/10.3390/ diagnostics11071192
2. Choi j, shin k, jung j, bae h j, kim d h, byeon j s, and kim n (2020): convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy, clinical endoscopy, vol. No. 53(2), pp. 117–126.
3. Crowley c, guitron s, son j, and pianykh os (2020): modeling workflows: identifying the most predictive features in healthcare operational processes, plos one, vol. No. 15(6).
4. De lange t, halvorsen p, and riegler m (2018): methodology to develop machine-learning algorithms to improve performance in gastrointestinal endoscopy, world journal of gastroenterology, vol. No. 24(45), pp. 5057–5062.
5. Borgli hanna, vajira smedsrud, and pia hicks (2020): a comprehensive multi-class image and video dataset for gastrointestinal endoscopy, scientific data, vol. No. 7.
6. Heaton. J (2015): artificial intelligence for humans, deep learning and neural networks, heaton research inc. St. Louis, vol. No: 3.
7. Hemalatha, r j vijaybaskar, v and thamizhvani t r (2018): performance evaluation of contour-based segmentation methods for ultrasound images, advances in multimedia, vol. No: 2018, article id: 4976372.

8. Hetal j vala, and astha baxi (2013): a review on otsu image segmentation algorithm, international journal of advanced research in computer engineering & technology, vol. No 2(2), pp. 387-389.
9. Hirasawa t, aoyama k, tanimoto t, ishihara s, shichijo s, ozawa t, ohnishi t, fujishiro m, matsuo k, and fujisaki j (2018): application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images, gastric cancer, vol. No: 21, pp. 653–660.
10. I armeni, o sener, a r zamir, h jiang, i brilakis, and m. Fischer (2016): 3d semantic parsing of large-scale indoor spaces, proceedings of the ieee international conference on computer vision and pattern recognition.
11. I v gorbunov, s r subhankulova, i a hodashinsky, and a e yankovskaya (2016): comparative analysis of feature selection algorithms in construction of fuzzy classifiers, ieee 10th international conference on application of information and communication technologies, pp. 1-3.
12. Illavarason p, arokia renjit j, and mohan kumar p (2018): clinical evaluation of functional vision assessment by utilizing the visual evoked potential device for cerebral palsy rehabilitation, procedia computer science, vol. No: 132, pp. 128-140.
13. Indraswari rarasmaya, kurita takio, arifin agus zainal, suciati nanik, and astuti eha (2019): multi-projection deep learning network for segmentation of 3d medical images, pattern recognition letters, vol. No: 125, pp. 791-797.
14. Irum isma, shahid muhammad, and muhammad mudassar (2015): a review of image denoising methods, journal of engineering science and technology review, vol. No. 8, pp. 41-48.
15. Iwahori, y, a hattori, y adachi, m k bhuyan, and j robert (2015): automatic detection of polyp using hessian filter and hog features, procedia computer science, international conference in knowledge based and intelligent information and engineering systems, pp. 730–739.