

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Mishra, Sushruta; Thakkar, Hiren Kumar; Mallick, Pradeep Kumar; Tiwari, Prayag; Alamri, Atif  
**A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection**

*Published in:*  
Sustainable Cities and Society

*DOI:*  
[10.1016/j.scs.2021.103079](https://doi.org/10.1016/j.scs.2021.103079)

Published: 01/09/2021

*Document Version*  
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

*Published under the following license:*  
CC BY-NC-ND

*Please cite the original version:*  
Mishra, S., Thakkar, H. K., Mallick, P. K., Tiwari, P., & Alamri, A. (2021). A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustainable Cities and Society*, 72, Article 103079. <https://doi.org/10.1016/j.scs.2021.103079>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# **A Sustainable IoHT based Computationally Intelligent Healthcare Monitoring System for Lung Cancer Risk Detection**

**Abstract-** A sustainable healthcare focuses on enhancing and restoring public health parameters thereby reducing gloomy impacts on social, economic and environmental elements of a sustainable city. Though it has uplifted public health, yet the rise of chronic diseases is a concern in sustainable cities. In this work, a sustainable lung cancer detection model is developed to integrate the Internet of Health Things (IoHT) and computational intelligence, causing the least harm to the environment. IoHT unit retains connectivity continuously generates data from patients. Heuristic Greedy Best First Search (GBFS) algorithm is used to select most relevant attributes of lung cancer data upon which random forest algorithm is applied to classify and differentiates lung cancer affected patients from normal ones based on detected symptoms. It is observed during the experiment that the GBFS-Random forest model shows a promising outcome. While an optimal accuracy of 98.8% was generated, simultaneously, the least latency of 1.16 seconds was noted. Specificity and sensitivity recorded with the proposed model on lung cancer data are 97.5% and 97.8%, respectively. The mean accuracy, specificity, sensitivity, and f-score value recorded is 96.96%, 96.26%, 96.34%, and 96.32%, respectively, over various types of cancer datasets implemented. The developed smart and intelligent model is sustainable. It reduces unnecessary manual overheads, safe, preserves resources and human resources, and assists medical professionals in quick and reliable decision making on lung cancer diagnosis. **Keywords:** Sustainable healthcare, Internet of Health Things (IoHT), Heuristics, Random forest, Greedy Best First Search (GBFS), Lung cancer, Classification, Computational Intelligence.

## 1. Introduction

Sustainability is a trending concept laid out throughout the world due to global warming and environmental change in urban regions. To reciprocate the harm caused by humans to the environment and enhance human lifestyle, the idea of a sustainable city has come into play. Individual vital pillars associated with sustainability, as highlighted in figure 1, include resource access, greenery, public safety, smart computing, conserve with the preserve, and computational intelligence [1].

**Resource Access:** Right to public resources is an essential foundation of a sustainable city. Some of these necessities include education, healthcare access, timely transport, good quality air and water, safety, and proper disposal of wastes.

**Greenery:** A significant concern in urban regions is the lack of vegetation and isolation from nature, which is a primary factor for health disorders. Thus, building green spaces around urban surroundings is an important factor for a sustainable city to get natural shade, good quality air, and noise pollution.

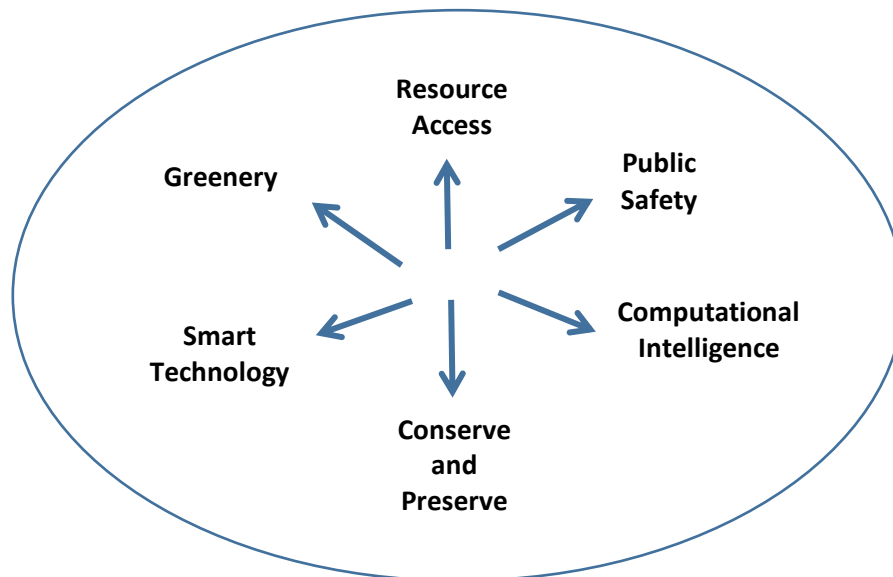
**Public safety:** Ensuring the public's safety and well-being is a primary concern of a sustainable city. Integrating technology in providing public safety can achieve efficiency in a sustainable environment. As an example, deploying power-efficient lighting grids reduces irrelevant expenditures and helps in the conservation of energy.

**Smart Computing:** Integrating IoT based technologies in the public domain across cities can ensure all-time connectivity to several public and private applications like in emergencies.

**Conserve and preserve:** A sustainable city must be inclined to use green power and conservation of water resources. It also refers to enhancing solar, wind, and nuclear energy usage to ensure that the public has access to all kinds of resources.

### ***Computational Intelligence***

With the rise in population in sustainable cities, the complexities in life also increase. To make things easier and faster, computational intelligence methods can be used to forecast and predict future events to help in effective decision making. Numerous machine learning-based algorithms can be used for this purpose.



**Figure 1. Pillars of Sustainable City**

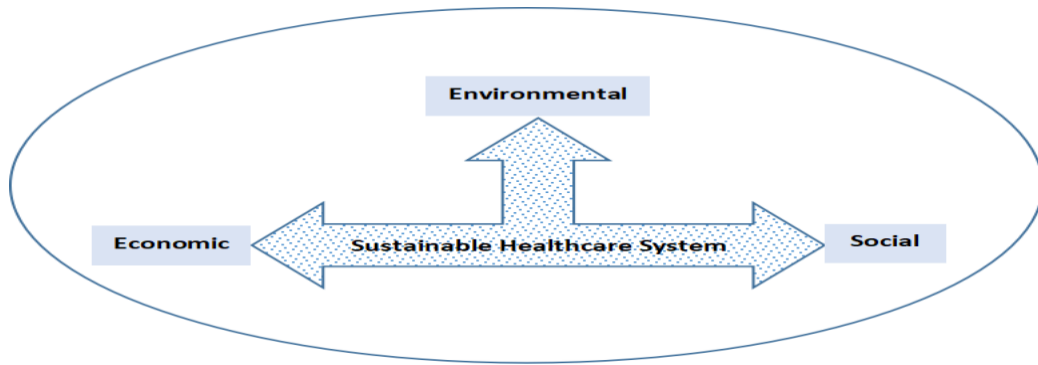
With the consistent rise in pollution, heat waves, and stress in populated places, it is essential to adopt a sustainable approach in urban regions to enhance residents' health and well-being. The overall health of people is vastly improved in a sustainable city, and it is determined by three elements of human health, which include:

- ❖ ***Physical health:*** It is improved by deploying sports field and engaging in outside sports events.
- ❖ ***Mental health:*** It is improved through surrounding greenery along with the soothing nature of trees and plants.
- ❖ ***Social health:*** It is improved by creating spaces for encouraging social contacts and sharing information on a daily basis.

Though a sustainable city offers lots of health-related benefits, it also leads to chronic related diseases, especially in a dense urban population [40]. Many lives are lost regularly due to these chronic diseases, and several people's normal activities are affected due to it. The rise in chronic disorders and disease complexities, increasing drug expenditures with technological cost, harmful waste generation, excess manpower usage, unnecessary wastage of resources, and restricted usage of data are recognized to be potential concerns in determining healthcare effectiveness and sustainability system [41]. Thus, there is a need to develop a more advanced, reliable, public friendly and efficient healthcare system that can help in improving, maintaining and restoring public health, reducing harmful effects on the environment. Such healthcare services will eventually benefit both presents as well as the future generation. Hence a sustainable healthcare model can fulfil the requirements of a more advanced medical service in a smart and sustainable environment [42].

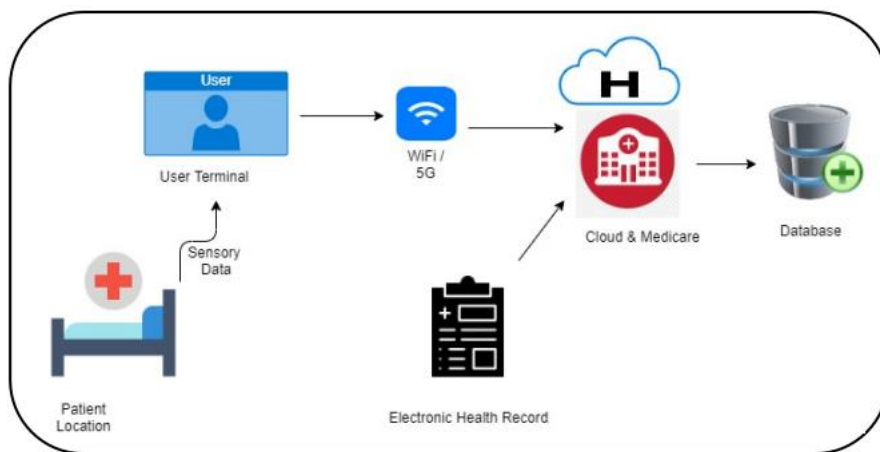
A sustainable healthcare model can be imagined to be bounded with three resources: social, financial, and environmental, as shown in Figure 2. It is achievable through providing quality medical help without unnecessary wastage of natural resources. Some of the measures through which a healthcare system can be made sustainable include the following.

- Taking necessary precautions about hazardous chemicals safety.
- Strictly adhere to protocols for disposing of wastes generated.
- Limit efforts of the workforce and avoiding unnecessary workloads.
- Restrict and regulate the usage of energy and carbon emission.
- Preserve and manage the amount of water usage.
- Automate the system functioning using technological advancements



**Figure 2. Resource Boundaries in a Sustainable Healthcare System**

Gradually the modern healthcare is embracing IoT technology in sustainable cities. The Internet of Health Things (IoHT) concept is viewed as the application of IoT in e-health [2]. It provides a technically and financially feasible solution allowing interconnectivity of medical services with patients. Numerous healthcare conditions are supported by IoHT, which includes pediatric care, elderly supervision, chronic disease diagnosis, and managing overall fitness and well being of society health. IoHT solutions not only offer greater patient care in less time but also ensures patient safety.

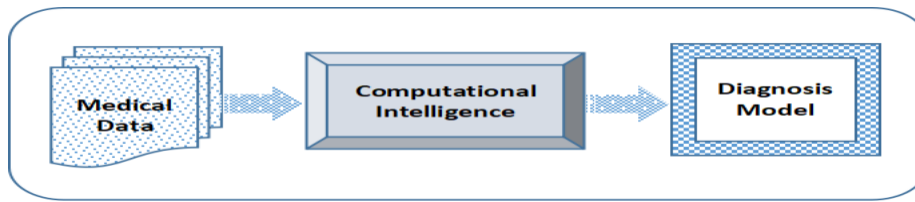


**Figure 3. A simple IoHT model**

A simple demonstration of IoHT comprising various elements is shown in Figure 3. Required wearables are embedded in sensors based on the patient's health status to collect readings. Sensors are associated with routers through radio transceivers. The

patient's aggregate data is handled by integrated cloud service in medical centers that act as administrators. These medical centers are interlinked with the local node where healthcare staff can access digital records. These records are securely stored in database servers for future access.

With the constant rise in population in sustainable cities, healthcare complexities are also increasing every day. Massive healthcare-related raw data are regularly accumulated from different heterogeneous sources in real-time. Many patients related information gets routinely collected. But due to lack of skilled workforce, delay in functionalities, and traditional manual procedures, the disease diagnosis task gets affected to a more considerable extent. Effective diagnosis of chronic diseases is a big concern in such scenarios. If medical professionals get the physiological and genetic factors of patients suffering from chronic diseases beforehand, diagnosis becomes more effective. Here computational intelligence can offer great help in building models that interrelate various features with disease risk. These intelligent algorithms provide some significant benefits in disease diagnosis tasks. It includes discovering factors associated with a disease, early and precise diagnosis of a disease, and limiting and scheduling healthcare unit visits as per the patient's need. To make things easier and faster, computational intelligence methods can be used to forecast and predict disease risks to help in effective decision making. A sample computational intelligent model for disease diagnosis is shown in Figure 4 where medical data records are used to generate a predictive diagnostic model through computational intelligence approach. Medical data may be disease risk attributes and symptoms or healthcare record samples in the form of either textual or image based samples.



**Figure 4. A sample disease diagnosis with computational intelligence**

Heaps of data are generated from numerous devices by IoT. Computational intelligence is enabled by these data which provides deep insight from it. Using previous data instances, computational intelligence approach helps in identifying trends which can be applied to develop future pattern prediction model. Business benefits from this integration of computational intelligence and IoT in performing prediction functionality with test cases thereby enabling it with superior automation ability [43-44]. A hybrid model that integrates IoHT along with computational intelligence approach can provide sustainable healthcare service and can effectively address concerns of chronic disease diagnosis in a sustainable urban environment. A suitable IoHT model helps in continuous tracking and collecting required data from patients, thereby reporting any medical staff ambiguities. A computationally intelligent-based smart model can help integrate massive data gathered from different heterogeneous sources, reduce unnecessary workload, preserve essential resources, and thereby facilitate patient diagnosis in a fast and cost-effective procedure.

Computational Intelligence forms a replica of human intelligence denoted through aggregation, acquisition, and interpretation of informative knowledge in computer systems. It deals with rational building agents involved in performing specific search algorithms in the background to accomplish their tasks. In general, searching is a well-defined procedure to determine a series of steps required in solving a particular problem at hand. Rational agents in these techniques act as a goal-oriented agent that usually uses these search methodologies and algorithms to determine an optimal solution to a



problem. Transformation of start state to goal state leads to a search problem and is achieved through search algorithms. The essential components of a search problem are depicted in Table 1.

**Table 1. Constituents of a Search Problem**

State space	Series of all feasible states possible
Start State	State which initiates the search procedure
Goal State	Function which analyses the current state and determines if it is the goal state or not.

State space represents the encoding of all feasible states which are relevant to a specific problem. Rather than including all details, only vital data should be encoded in the state space else unnecessary information will be accumulated making the search process tedious. The search agent needs to be informed about the initial state before the search begins in order to monitor all states throughout the procedure. It is crucial to have a definite goal state which determines when the solution to the problem is achieved. Thus solution to a problem in hand is denoted by computing a series of actions leading to a desired goal. Every action alters the state where the purpose is to determine the set of actions and states which map start state onto goal state.

The purpose of AI is to offer an adequate solution to a problem. In many problems there are many observed results which offer a combination of outcomes providing not so feasible solutions and also there are certain outcomes which are acceptable as feasible solutions. In general, instead of more feasible solutions, we get more results. Thus a solution set is more likely to be a subset of result set. It infers that a specific problem requires a search method to explore relevant solution set. Here a search algorithm acts as an AI agent that works in background and helps in achieving the ultimate goal after analyzing all feasible alternatives. Without these algorithms, an AI system is incapable

of exploring the most appropriate solution. The search algorithms can be categorized into blind search and heuristic search algorithms based on search problems in hand. Blind search adopts a brute-force approach and does not have prior knowledge of any domain-specific data like closeness or any goal related data. Still, heuristic search utilizes knowledge of the application domain, and complete information regarding the problem is available to guide the search procedure. Table 2 highlights the comparative study between heuristic search and blind search.

**Table 2. Heuristic Search vs. Blind Search**

Heuristic Search	Blind Search
Makes use of domain knowledge in searching process.	No domain knowledge is available during the searching process.
Reaches goal state in a quick time.	Comparatively takes more time to reach the goal state.
Provides direction at every phase during searching procedure.	No suggestion is provided regarding solution aspects.
Implementation is short and precise.	Implementation is quite lengthy.
Operates with relatively low cost overhead.	Operates with a higher cost overhead.
Highly effective.	Moderately effective.

There are some productive benefits that the heuristic offers while solving a problem. Some common advantages include less computational and implementation latency while providing a creative means of approaching a problem at hand. Figure 5 illustrates some basic advantages that the heuristic approach offers while solving a problem.

The heuristic function is employed in heuristic search to determine the most promising route to find a solution. The agent's current state acts as the input to heuristic, and an estimate of the agent's closeness to the goal state is generated as output. A.I. makes use of this heuristic in search of an efficient solution space. A heuristic algorithm in A.I. is

an effective way of problem-solving that provides an immediate and short term goal state in less time with good efficiency.

Admissibility concerning a heuristic function can be represented in equation 1 as:

$$h(n) \leq h^*(n) \quad (1)$$

Here  $h(n)$  denotes the heuristic cost while  $h^*(n)$  can be viewed as the estimated cost.

Thus estimated heuristic cost should not exceed the estimated cost.

Nodes are expanded based on the heuristic function  $h(n)$ . In general, two lists are maintained, which include OPEN and CLOSED list. Already expanded nodes are placed in the CLOSED list while nodes yet to be grown are available in the OPEN list. In every round, a node with the least  $h(n)$  value is expanded, thereby disclosing all its successors, and finally, node 'n' is pushed to the CLOSED list. The procedure is repeated till a goal state is reached.

The paper is organized as follows. The first section introduces the topic and addresses the importance of sustainability in modern times. The need for sustainable healthcare is discussed using IoHT and computational intelligence approach to tackle chronic disease risks in a sustainable environment. Further, the use of heuristics is defined. Section 2 describes the problem statement and gives importance to the prevailing lung cancer, and explains developing a computational efficient lung cancer detection model for sustainable cities. Section 3 deals with the relevant background work being done with respect to the domain. Section 4 introduces the lung cancer dataset used in the study. Section 5 presents the heuristic-based greedy best-first search method used in attribute optimization, and Section 6 computes the attribute's importance in a random forest algorithm. Section 7 discusses the IoHT based proposed lung cancer detection model for sustainable cities in detail. Section 8 gives the results and analysis outcome of the implementation of the proposed model. Finally, Section 9 concludes the research.

## **2. Problem Statement**

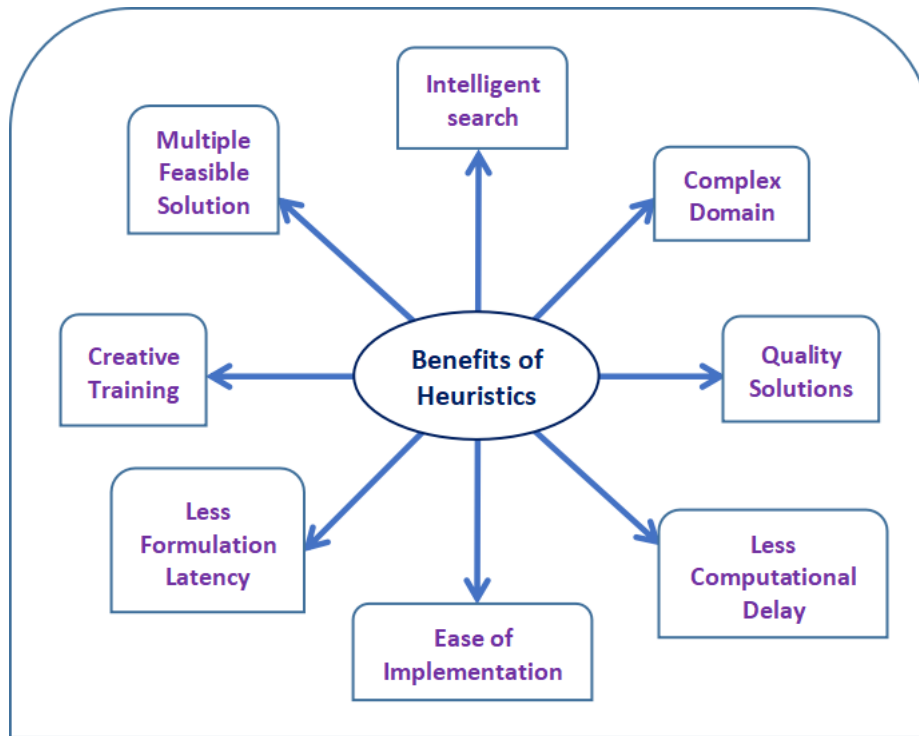
Cancer is a more general term that can be regarded as a group of disease disorders that affect various body organs of human beings like kidneys, liver, hair, skin, and lungs. Cancer diseases are on the constant rise in sustainable cities. At present, there are around 17 million cancer-related cases worldwide, and it is predicted that this figure will rise to approximately 23 million new cases every year by 2030. Lung cancer is being perceived as a dangerous and rapidly spreading chronic health disorder in a sustainable environment. It is caused by a community of cancer cells that develop rapidly in the lung tissue forming a malignant tumor to take shape. These cells exhibit abnormal behavior and interfere in the normal functioning of the lung. It has a 12.8% overall presence among all cancer reported throughout the world. Around 18% of deaths caused by cancer are due to lung cancer [3]. The male population affected by this represents 38.6%, while it is 5.2% in affected females. A majority of patients are diagnosed in an advanced stage of lung cancer. Even if by treating it through sophisticated technologies, the survival rate of patients is highly reduced. It is observed that even if after diagnosis of lung cancer, the maximum lifetime of a person is 5 years. Besides this, misdiagnosis is another worrying factor.

In some cases, it is noticed that a benign type is detected as malignant and vice versa by medical experts. It puts the patient's life in a risky and uncertain situation. Hence, it is highly recommended that if it is detected at an early stage of growth, the patient's survival probability is improved. With the recent advancements in computational intelligence techniques and smart computing, it is feasible to develop an automated IoHT based intelligent lung cancer detection model for sustainable cities. Such a model can help clinical personnel to identify the disease risks associated with lung cancer at

an early stage. Patient's data can be continuously monitored and collected using the IoHT module. These data can be used to extract relevant information about the patient, which can be later used to generate hidden patterns using computational intelligence methods. This can help detect lung cancer in patients, which can be useful for medical professionals. Lung cancer data collected from various sources may contain inconsistencies, and some features may be of significant help during the treatment process. These less relevant parameters need to be dropped from the data samples. In such scenarios, heuristic techniques can help detect these less significant features from the dataset and eliminate them, thus generating a more refined data record.

Classification with a refined and optimal dataset generates a very high accuracy and efficiency. It is observed that many classifiers tend to suffer from the over-fitting issue, and its variance is reduced. Some classifiers fail to handle both numerical and categorical values, while few are susceptible to outliers. Random forest algorithms can be helpful to overcome these pitfalls of general classifiers. This research study uses Greedy Best First Search (GBFS) algorithm as the heuristic approach to optimize the lung cancer dataset's parameters and features.

Furthermore, a random forest algorithm is applied to detect the presence of risk disorders in patients, thereby helping in the classification of lung cancer patients. Using the GBFS algorithm and a random forest classifier, the classification performance can be enhanced, which can be extremely informative for healthcare experts. Medical experts can take advantage of heuristic benefits in developing an effective, sustainable, reliable, and intelligent classification model that can assist them in the treatment of prominent, widespread cancers in urban populated regions.



**Figure 5. Graphical view illustrating prime benefits of heuristic approach**

### **3. Background Study and related works**

Sustainable cities offer several benefits to society, and the healthcare standard is also uplifted in a sustainable environment. But with the adoption of sustainability, the visible rise of chronic diseases cannot be ignored. It poses a huge challenge to deal with chronic disease risks in densely populated sustainable cities. Lung cancer is one leading chronic disease being seen in many sustainable cities. Various IoHT and intelligent models have adopted different advanced technologies to handle lung cancer in these scenarios. In this section, a range of background studies is presented where several relevant works have been undertaken about the classification and prediction of lung cancer using computational intelligence methods.

Lung cancer occurs in the tissues of the lungs, and it is the prime source of tumor in human beings. Tobacco is being highlighted as the chief source of lung cancer, responsible for around 85% of death cases. An uninhibited development of abnormal cells affecting the lung around the line air division region causes lung cancer [3]. The

survival rate in diabetes patients can vary slightly. As per the observations, diabetic patients with higher usage of insulin are at greater risk [4]. Some studies also inferred that lung cancer would be affected in patients having diabetes mellitus [5]. Regular smokers have a nicotine effect on insulin action and secretion in diabetes [6]. Avoiding smoking is crucial to regulate diabetes and reduce diabetic issues [7] [8]. EHRs can be used to help patients to manage personalized care, and also medical care performance can be coordinated [9]. Various computational techniques have been using in previous research works on lung cancer treatment. Setio et al.. [10] developed a 3D convolution neural network model for F.P. minimization in lung nodule categorization. It was used to analyze the 3D nature of C.T. scans to decrease fault diagnosis, and a weighted sampling method was applied to enhance results. Jiang Hongyang et al.[11] developed a community-oriented pulmonary nodule identification model using a multi patches scheme applying frangi filter to improve prediction performance. It showed a sensitivity value of 80.06% and an F.P. rate of 15.1 units. Kattan and Bach [12] proposed an analysis on the change in lung cancer risk disorders observed among smokers based on several physiological factors. It was observed that older people beyond 68 years of age group had a higher risk of lung cancer, and youths addicted to smoking over 28 years were more vulnerable to lung cancer. Authors in [13] developed a prevention model for different disease risks including hepatitis and lung cancer based on knowledge mining techniques where distinct data samples with numerous factors were used. Authors in [14] used classifiers like Bayes trees and decision trees on heart disease dataset with many samples to predict risk factors associated with heart risks. Manikandan et al. [15] developed a hybrid neuro-fuzzy model to predict lung cancer using 11 symptoms on a dataset of 271 samples. Arulananth et al. [16] defined various symptoms to be used for lung cancer forecasting. Symptoms distinguished between diagnostic factors like age,

gender, family history, etc., and cancer presence was considered. Senthil and Ayshwaya [17] applied neural networks and evolutionary techniques to determine the degree of lung cancer risk based on several risk factors. Lung cancer data from the UCI data repository was used for the computational purpose. Markaki et al.[18] developed a medical risk prediction prototype for lung cancer based on symptoms related to smoking. Mohapatra et al.[19] observed that sustainable green computing plays a vital role in developing an effective environment friendly disease diagnosis prototype. Krishnaiah V, Narsimha G, Subhash Chandra N [20] developed an accurate framework for detection of lung cancer disease risks using symptoms like age, gender, wheezing, shortage of breath, and chest pain, among others. It helped in predicting the probability likelihood of patients affected with lung cancer. Prashant Naresh [21] used a pattern prediction model for lung cancer prediction where a patient's predisposition for lung cancer is detected. Machine learning models proposed by Gao et al. [22] and Wang et al.[23] yielded a classification accuracy of 86%, while other frameworks like Guo et al. [24] and Zhao et al. [25] generated a relatively higher accuracy of 92% and 95.6%, respectively. In all these classification models, texture-based attributes were used for the analysis and categorization of lymph nodes [38-39]. Table 3 highlights the relevant existing works carried out on lung cancer detection using computational intelligence approaches.

**Table 3. Existing works on lung cancer detection using machine learning**

<b>Authors</b>	<b>Model used</b>	<b>Discussion</b>
Gao et al. [22]	SVM & RBF kernel	86% accuracy rate is produced.
Wang et al. [23]	CNN	86% mean accuracy rate is the outcome.
Guo et al. [24]	Support Vector Machine	Augmented accuracy generated.
Zhao et al. [25]	SVM & RBF	Sensitivity of 95.6%.



Eric Bair et al. [26]	Semi-Supervised Methods	Supervised principal component Analysis generates optimum error rate of 0.06.
Shraddha et al.[27]	Fuzzy Min-Max Neural Network.	94% accuracy obtained by pruned FMM model with 8 hyper-boxes.
V.Krishnaiah et al. [28]	B.N., D.T., ANN	81% accuracy with N model and it is 5% more than accuracy of ANN and D.T.
Kourou et al. [29]	ANN, B.N., SVM, and DT	ANN gave highest accuracy of 83%.
Joseph A. Cruz et al.[30]	Random forest model	Latest computational methods utilized to predict outcome of autistic disorders.
Xiao Zang et al. [31]	Statistical models	In testing samples, the ADC and SCC patients are classified into high and low risk instances estimated by morphological features.
Xueyan [32]	D.T. and R.F.	80% accuracy is outcome with random forest.
Jung Hun Oh et.al [33]	Classification algorithms in the machine learning.	Kernel SVMs generates MCC of 89% that is 5% more than linear SVM.
Yang Xie et.al [34]	DTNB hybrid intelligent model	99% prediction accuracy with hybrid model is outcome.
Wenqing et.al [35]	CNN, DBNs, SDAE	80% accuracy achieved in CNN model which is an improvement over DBN and SDAE of 2%.
Stephan Wienert et.al [36]	"minimum model "cell detection technique to detect tumor	Presented method provides precision and recall of 0.908 and 0.859 respectively

An analytical survey concerning the use of computational predictive models in lung cancer detection and classification was undertaken. Various relevant research works are discussed and their result outcomes are highlighted in this section. From machine learning to deep learning models, several significant research analysis are presented. It is observed that most of the existing models used simple classification machine learning

models. Even if some attribute optimization methods were used in few cases, still irrelevant features persist in the data samples after applying those methods. As a result, there was less impact on the lung cancer detection accuracy. It is also noted that heuristic optimization approach is seldom used in lung cancer analysis. Apart from this, sustainability issue is untouched till present studies. Deploying a smart IoHT enabled disease diagnosis framework is rarely implemented in existing works.

#### 4. Lung cancer dataset used in research

There are numerous factors associated with lung cancer disorders in patients. The systematic analysis suggested that apart from medical symptoms, statistical factors are also interrelated with lung cancer. Some common risk factors observed in lung cancer include cough, chest pain or back pain, weight loss, shortage of breath, etc. In this research, the lung cancer dataset was collected from the UCI machine learning repository. As noted as many as 16 distinct symptoms are taken into consideration in this dataset. It is a standardized dataset which is mostly used in research analysis. The raw dataset comprises only 32 instances. In our research, data samples are enhanced to 488 through data augmentation approach. Common factors like age, chest pain, overweight and gender are considered in the data. Besides these, some complex symptoms like genetic disorders, alcohol use and smoking habit among others are also utilized in the data instances for research. The domain range of all attributes are labeled on a scale of 1 to 10 and all attributes are integer data type. A sample feature set of the data sample is highlighted in table 4.

**Table 4. Lung Cancer dataset details [37]**

<b>Feature</b>	<b>Details</b>	<b>Range</b>
Age	Age of the lung disease patient	( Age between 13 – 80)

Sex	Gender of the patient	Male-1/ Female -2/Others-3)
Air pollutants	Different air pollutants detected	( Domain: 1 -10)
Use of alcohol	Whether patient is alcoholic	( Domain: 1 -10)
Occupational trait	Job concern details of patient	( Domain: 1 -10)
Hazardous element	If any poisonous element detected	( Domain: 1 -10)
Gene disorders	Inheritance and genetic risks	( Domain: 1 -10)
Lung chronic risks	Chronic lung disorders observed	( Domain: 1 -10)
Balanced food intake	Diet and food habits of patient	( Domain: 1 -10)
Overweight	Obesity issue of patient	( Domain: 1 -10)
Smoke symptom	Smoking pattern of patient	( Domain: 1 -10)
Non-active smoker	Regular or occasional smoker	( Domain: 1 -10)
Pain in chest	If patient suffering from chest pain	( Domain: 1 -10)
Blood in cough	Blood trace present while coughing	( Domain: 1 -10)
Loss in weight	Constant weight loss of patient	( Domain: 1 -10)
Snore symptom	Snoring habit of patient observed	( Domain: 1 -10)

## 5. Greedy Best First Search Heuristic Method

Heuristic approaches are quite popular and widely used in various domains. In this study, Greedy Best First Search (GBFS) method is used for generating an optimal feature subset of the lung cancer dataset. Greedy Best-First Search (GBFS) algorithm chooses the route which seems best at that instance. It can be viewed as a combination of depth-first search and breadth-first search methods. A heuristic function is used to guide the searching procedure. A priority queue is implemented in this algorithm to store all processed nodes. At every step, the most promising node is selected. The node closest to the goal node is always expanded first, which is computed with the help of heuristic function  $h(n)$  given in equation 2 as:

$$f(n) = h(n) \quad (2)$$

Where  $h(n)$  denotes the cost estimated from a node 'n' towards the goal state. The pseudocode of GBFS is shown in table 5. It uses two empty lists (INIT and CLOSE)

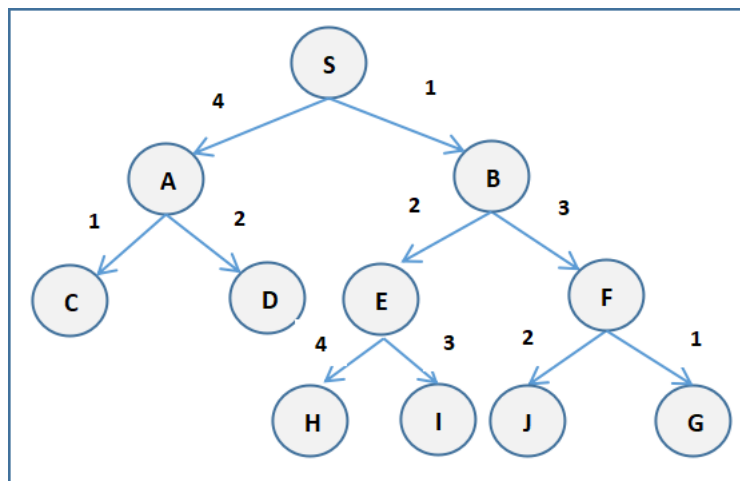
*INIT*- It stores all generated nodes which are not yet visited.

*CLOSE* - It stores all visited nodes.

**Table 5. Pseudocode for GBFS method**

Step 1:	Two empty lists are created (INIT and CLOSE).
Step 2:	Start from the first node (say 'A') and place it in the INIT list.
Step 3:	Subsequent steps are repeated until the goal node is reached.
Step 4:	Exit loop and return 'fail' if the INIT list is found empty.
Step 5:	The first node 'A' is selected from the INIT list and moved to the CLOSE list.
Step 6:	If 'A' is the goal node, then shift it to the CLOSE list. Loop is exited returning 'true'. The solution is calculated by backtracking the route.
Step 7:	If 'A' is not a goal node, then 'A' is expanded to produce all 'immediate' next nodes interlinked with 'A'.
Step 8:	All those interlined nodes are added to the INIT list.
Step 9:	Nodes are rearranged in the INIT list on the basis of evaluation function $h(n)$ .

S: Initial state, G: goal.



**Figure 6. Graphical illustration of working of GBFS heuristic method**

Graphical representation of the GBFS method is shown in Figure 6. 'S' is assumed to be the initial node, while 'G' denotes the goal node. The distance units between any two nodes are specified and taken as the heuristic function which is highlighted in table 5. The distance units of source node 'S' to all other intermediary nodes are computed with visiting the next immediate node upon traversal.

**Table 5. Heuristic estimation**

node	h(n)	node	h(n)	node	h(n)
<b>A</b>	11	<b>E</b>	4	<b>I, J</b>	3
<b>B</b>	5	<b>F</b>	2	<b>S</b>	15
<b>C, D</b>	9	<b>H</b>	7	<b>G</b>	0

Table 6 presents the overall function of the GBFS method for the example taken into consideration in Figure 2. Individual steps are highlighted in different loop counts. The graphical view of the operational steps is illustrated in Figure 7.

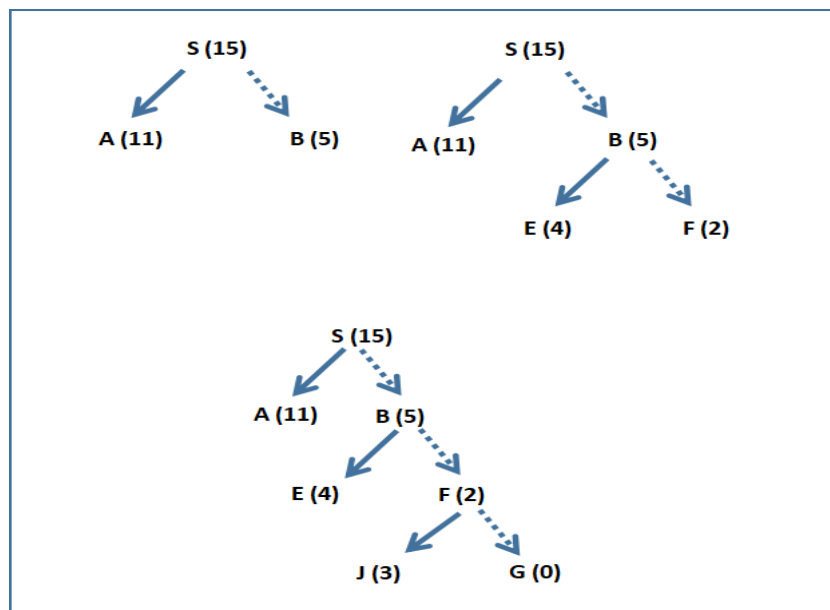
**Table 6. The functioning of the GBFS Heuristic method**

Phase	Operation
<b>Initialization</b>	S (15)
	Nodes of S expanded and S added to CLOSED
	Successors of 'S' added to INIT and find f(n).
<b>Loop Count 1</b>	Optimal node is chosen (min h(n))
	Drop node from INIT and insert it to CLOSE.

---

	Successors of 'B' added to INIT and find f(n).
<b>Loop Count 2</b>	Optimal node chosen (min h(n))
	Drop node from INIT and insert it to CLOSE.
	Successors of 'F' added to INIT and find f(n).
<b>Loop Count 3</b>	Optimal node chosen (min h(n))
	Drop node from INIT and insert it to CLOSE.

---



**Figure 7. Illustration of working of GBFS method for Figure 2 example**

Shortest path route from the source node to the goal node as denoted in figure 6 is represented here. Initialization of source node 'S' takes place and its immediate neighbors 'A' and 'B' are explored. The currently explored nodes are placed in INIT list while already visited node is placed in CLOSE list. The process is continued in three

loops till all nodes are explored. Finally the least path distance from 'S' to goal node 'G' is through intermediate nodes 'B' and 'F'. The overall procedure is shown in table 7.

Shortest route from S → G is S → B → F → G

Total cost:

<https://www.blogger.com/blogger.g?blogID=2953467764241060843><https://www.blogger.com/blogger.g?blogID=2953467764241060843><https://www.blogger.com/blogger.g?blogID=2953467764241060843>  
 $1+3+1=4$

**Table 7. Computation of shortest path route using GBFS method**

Loop Count	INIT	CLOSE
Initialization	A, B	S
Loop Count 1	A	S, B
Loop Count 2	F, E, A	S, B
	E, A	S, B, F
Loop Count 3	J, G, E, A	S, B, F
	J, E, A	S, B, F, G

GBFS method of heuristic search utilizes less memory and time requirement thereby providing a promising performance even if the search space is huge. GBFS method generates an optimal solution set for a specific problem comprising multiple solutions.

### **6. Determination of attribute importance in random forest classifier**

Here information gain is used to split the dataset using entropy measure. It is determined to reduce entropy while splitting the dataset upon a specific attribute, as shown in equation 3.

$$\text{Gain}(V, A) = \text{Entropy}(V) - \text{Entropy}(V, A) \quad (3)$$

$V$  = Target variable, and 'A' denotes the splitting attribute.

Entropy ( $V,A$ ) = Computed entropy after splitting dataset on attribute 'A'.

During the training phase, random forest builds several individual decision trees, and predictions among these trees are aggregated to make the cumulative prediction. An attribute's importance is determined by reducing node impurity, taking the weight of probability to reach that node. Thus as the value is more, the feature becomes more important. For every decision tree, a node's importance is calculated with a Gini significance score, as depicted in equation 4.

$$np_q = W_q D_q - W_{\text{left}(q)} D_{\text{left}(q)} - W_{\text{right}(q)} D_{\text{right}(q)} \quad (4)$$

$np_q$  = importance of node  $q$ .

$W_q$  = weighted instances reaching node  $q$ .

$D_q$  = impurity value of node  $j$ .

$\text{left}(q)$  = child after left split on node  $q$ .

$\text{right}(q)$  = child after right split on node  $q$ .

Then an attribute's importance is computed in equation 5.

$$\text{Importance}(p)_p = \frac{\sum_{q:\text{node } q \text{ split attribute } p} p^{mp}_q}{\sum_{l \in \text{all nodes}} m^{pl}} \quad (5)$$

$\text{Importance}(p)_p$  = importance of attribute  $p$ .

$p^{mp}_q$  = importance of node  $q$ .

It is followed by the normalization of values between 0 and 1 by dividing the addition of all attributes' importance values as shown in equation 6.

$$\text{Normal}(A)p_p = \frac{A p_p}{\sum_{q \in \text{all attributes}} A p_q} \quad (6)$$

The final attribute importance of random forest represents the mean of all the individual trees, which is calculated as denoted in equation 7 as follows:

$$\text{R. F. } (A)_p = \frac{\sum_{q \in \text{all trees}} \text{Normal}(A)p_{p,q}}{T_s} \quad (7)$$



$R. F. (A)_p$  = importance of attribute  $p$  determined from all individual trees in random forest ensemble.

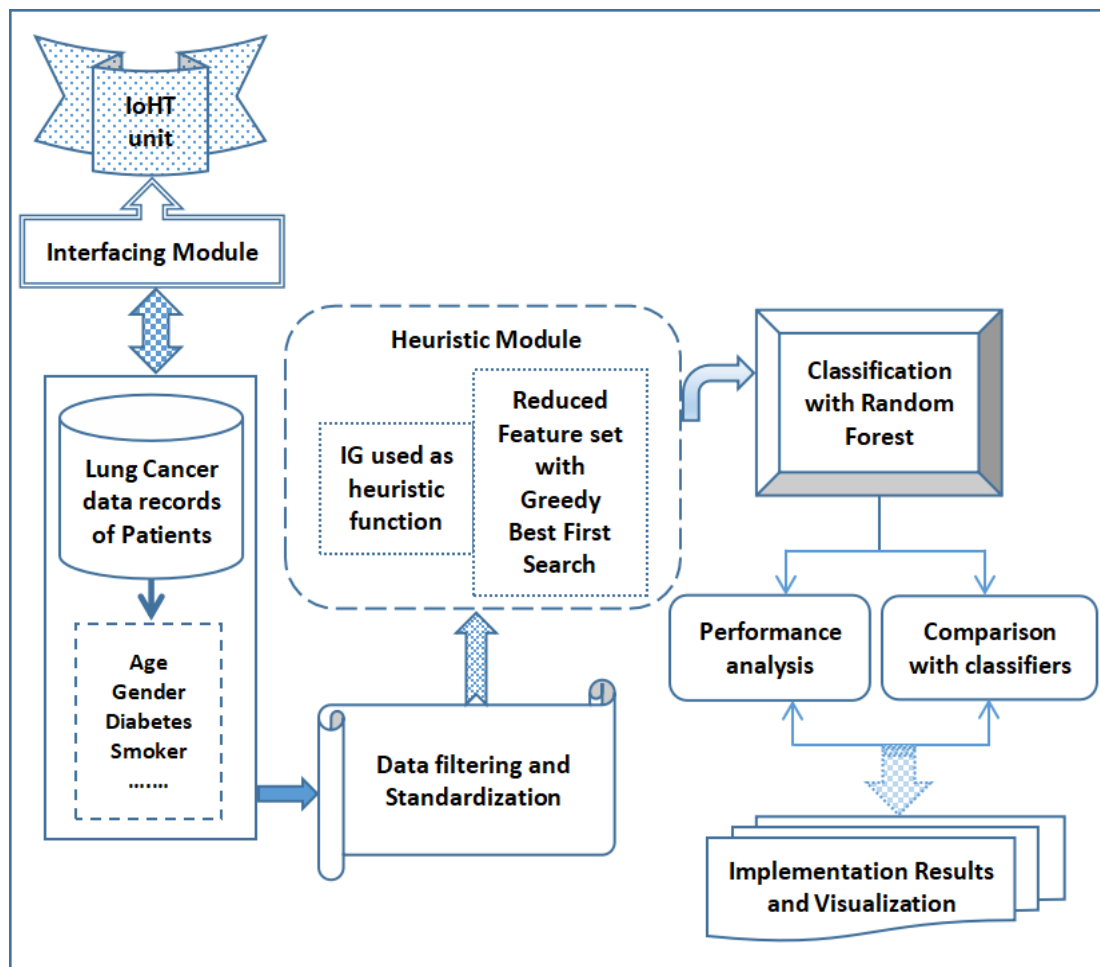
$Normal(A)_{p,q}$  = normalized attribute importance of  $p$  in tree  $q$ .

$T_s$  = quantity of tree count.

## **7. Proposed Methodology for lung cancer detection**

The proposed lung cancer detection model for sustainable healthcare in urban cities is discussed in figure 8. It integrates both IoHT and computational intelligence approaches in developing a smart and intelligent sustainable lung cancer model [45-46]. It presents an intelligent and precise analysis of lung cancer data samples processed and later classified using a heuristic-based GFS algorithm and Random forest classifier. Patient-related data are tracked continuously through an IoT unit. Variety of sensors are integrated in the IoHT unit of the model to track several health related patterns. Some vital sensors include optical heart rate sensor, respiratory sensor, lighter sensor, gyroscope for smoke detection, alcohol detector and accelerometers among others. It monitors and collects the attribute values of the patient under consideration. Information like pulse rate, blood pressure, smoking information, and other relevant information required for analysis are aggregated through the IoHT unit and is passed to the interfacing module, which acts as the cloud based interface between the IoHT unit and computational intelligence unit. It is the storage unit where all patient's related health data and lung cancer data samples are accumulated and stored for usage. Open end platform Thingspeak is used for the purpose. Apart from data storage, it facilitates scheduling, application integration and visualization functionalities. Lung cancer dataset details are retrieved from the UC Irvine Machine Learning repository [46]. After obtaining data from the IoHT unit, all features are examined and verified for

completeness and suitability for usage. A count of numerical and categorical features are noted. The raw unstructured data was further pre-processed and filtered to remove inconsistencies. The data records' null values were detected and replaced with the mean value of that feature column vector. Since some columns contained varying scale data, hence re-scaling of the dataset was done to set the range of data values in an identical scaling range. A binary threshold parameter conversion is used to convert the domain values beyond threshold into '1', and other values below or equal to the threshold are assigned '0'. Data standardization is applied to the dataset where the data instances center around the average mean with a unit standard deviation. Here the average mean of a feature column is zero while a unit standard deviation is obtained for the final data distribution.



**Figure 8. Proposed GBFS-Random Forest-based Sustainable Model for Lung Cancer Analysis**

Once the data is standardized and filtered, it is ready to be input into the heuristic module. This module uses the Information Gain (I.G.) method as a heuristic function and Greedy Best First Search (GBFS) method as the heuristic search technique. IG computes the entropy reduction and is utilized in building random forest from a training samples by determining the gain in information for every feature under consideration. The feature with maximum information gain reduces the overall entropy and is selected for splitting the data samples for classification in random forest technique. The I.G. method extracts the entropy information for all features in the dataset. The more the I.G. value, the more is the reduction in entropy of a feature. I.G. value of all features are computed, which served as the heuristic function  $h(n)$ . Using this  $h(n)$  value, GBFS determines the optimal set of features in the lung cancer data records. The resultant dataset is a pre-processed data sample with significant feature values. The least essential features which contributed less to the performance of classification are dropped. The random forest algorithm is then applied to the optimal dataset for classification and to figure out the disease stage and risk factors associated with a lung cancer patient. The classification performance is further tested using evaluation parameters like accuracy, latency, and error rate. Finally, the classification model is evaluated and compared with other existing classifiers to determine its consistency and effectiveness in general. The proposed IoHT based heuristic computational intelligence model is demonstrated using a python programming language. Results of implementation are visualized and analyzed in the form of graphs and tables.

**8. Implementation and Result Analysis**

The work comprised the development of a smart and intelligent lung cancer detection model for a sustainable environment. It uses a heuristic-based technique on lung cancer dataset to select an optimal feature set fed to a random forest algorithm to classify patients detected with lung cancer symptoms. GBFS method was the heuristic method used in the proposed work with the I.G. method as the heuristic function. I.G. method helped in determining the information content of features in the lung cancer dataset. This record was used as a heuristic measure in guiding the search space for the GBFS method in finding an optimal feature set. The entire research work was implemented using the Python programming language. Results obtained were visualized and arranged for graphs and tables for comparison with other evaluation parameters. Different hyper parameters of both random forest and GBFS method were tuned during implementation. Table 7 highlights the hyper parameters of random forest used in study and table 8 depicts that of GBFS method.

**Table 7. Hyper parameters of Random forest used in study**

<b>Parameter</b>	<b>Detail</b>	<b>Range</b>
max_depth	Longest route between leaf node and root node	[60, 80, 100]
min_sample_split	Least samples to partition an internal node	[2, 4, 6]
max_leaf_nodes	Optimum number of leaf nodes	[2, 4]
min_samples_leaf	Least samples requirement at leaf node	[2,3,4]
n_estimators	Decision tree count in random forest	[60,100,150,180]
max_sample	Optimal samples in training each decision tree.	20% of data instances
max_features	Optimum features to split a node	['auto', 'sqrt']

**Table 8. Hyper parameters of GBFS approach used in study**

<b>Parameter</b>	<b>Detail</b>
------------------	---------------

---

INIT list	Tracks current nodes for traversal
CLOSED list	Tracks already traversed nodes
start_node	Initial node to begin traversal
goal_node	Target goal node to reach
open.append	Include the nodes one by one to open list
closed.append	Include the considered node to closed list
open.sort()	Open list sorted to retrieve least cost node
current_node.position	Identify the position of node under consideration

---

Various performance metrics were used for the implementation analysis, and these are derived from the confusion matrix. Performance parameters used in our study are presented here.

Accuracy rate refers to the ratio of the accurate disease predictions to the total number of predictions, and it is shown in equation 8.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

Specificity is represented by the capability of a prediction decision to accurately detect individuals without the disease risks shown in equation 9.

$$\text{Spec} = \frac{\text{T.N.}}{\text{T.N.} + \text{F.P.}} \quad (9)$$

Sensitivity is defined as a decision outcome's ability to detect individuals with the disease accurately and is computed in equation 10.

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

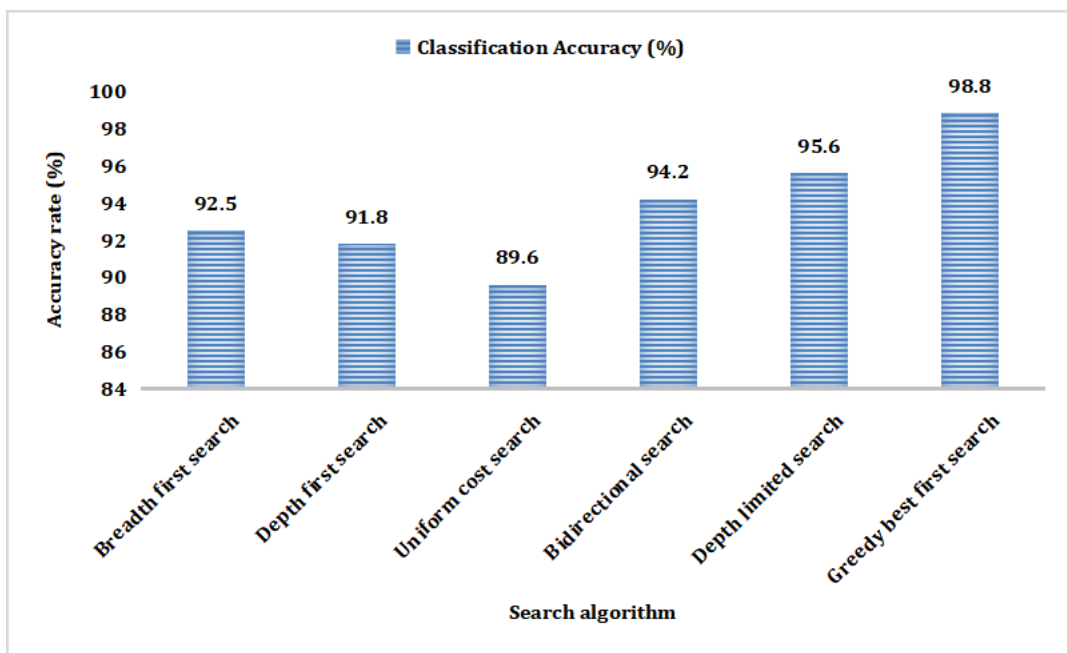
F-Score is the harmonic average mean between the specificity and sensitivity values, which is denoted in equation 11.

$$F - \text{Score} = \frac{2TP}{2TP + FP + FN} \quad (11)$$

Execution time latency is the cumulative time delay in training the computational intelligence model and testing it. It is shown in equation 12.

$$\text{Exec. time} = \text{Time}_{\text{train}} + \text{Time}_{\text{test}} \quad (12)$$

Where: TP = True positives, TN = True negatives, FP = False positives, FN = False negatives.

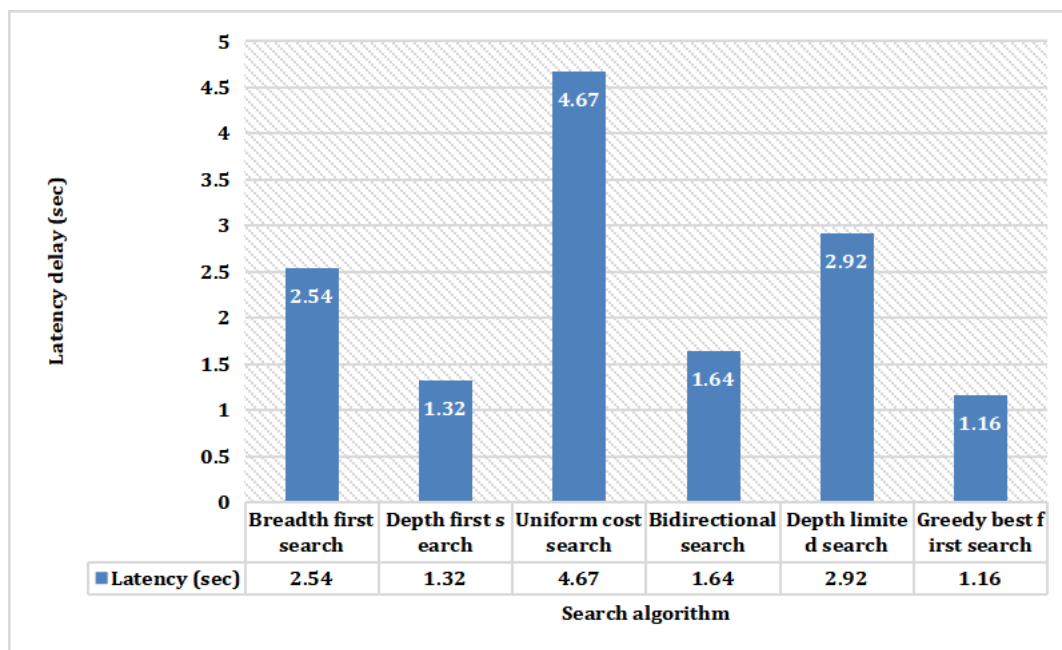


**Figure 9. Comparison of accuracy rate among heuristics and blind search methods**

Here the result obtained after implementing the proposed GBFS-Random forest model on the lung cancer dataset is presented. A comparative analysis was done using the GBFS heuristic method with some other popular blind search methods like depth-first search, breadth-first search, and bidirectional search, among others, as shown in Figure 9. Impressive classification accuracy of 98.8% was generated with the GBFS method. Among blind search depth, the limited search produced a perfect accuracy of 95.6%. The uniform cost search method gave the least accuracy of 89.6%. Random forests was

the classifier used for the evaluation. In general, heuristic approach generates better accuracy since it utilizes specified domain based information to do the search and also uses a predefined heuristic function.

Evaluation of the proposed heuristic-based classification model was performed with other blind search techniques in terms of the execution period. A significantly less latency of 1.16 sec was observed with the proposed model, while a relatively high 4.87 sec was noted with a uniform cost search method. Heuristics approach is usually quick as it does not need to store all unwanted solutions to reach the target so it required a less intensive memory. Figure 10 shows the overall execution time latency for the process.

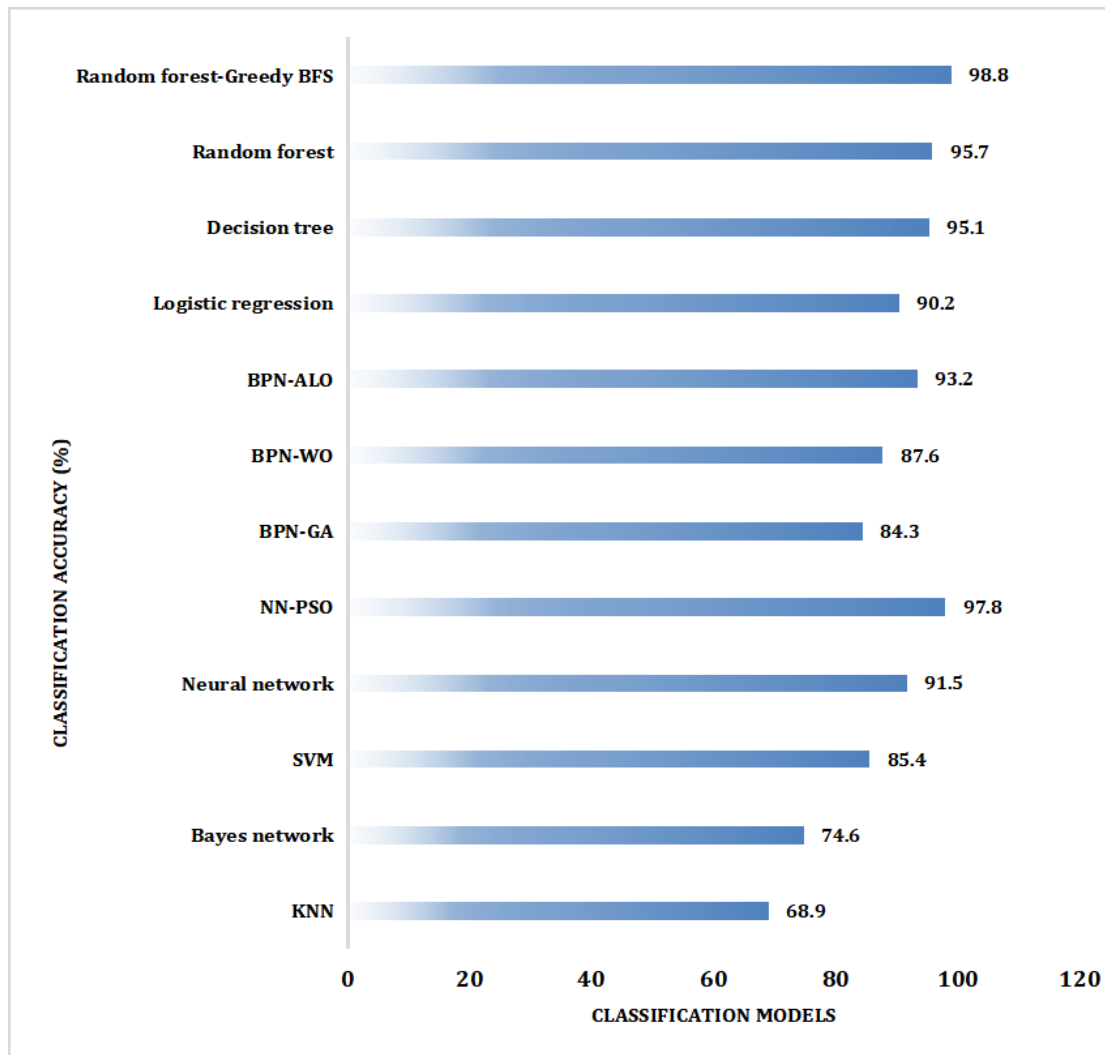


**Figure 10. Comparison of execution time among heuristics and blind search methods**

The efficiency of any machine learning model is determined by its ability to outperform other existing models. The proposed heuristic classification technique was evaluated using various other existing classification methods like SVM, Neural network, decision

tree, and regression analysis, among others using lung cancer data instances. KNN algorithm gave a very low accuracy of 68.9%, while 95.1% accuracy was noted with the decision tree classifier. Few hybrid models like Back Propagation Network-Genetic Algorithm (BPN-GA), Back Propagation Network-Whale Optimization (BPN-WO), and Neural Network- Particle Swarm Optimization (NN-PSO) produced an accuracy of 84.3%, 87.6%, and 97.8%, respectively. The enhanced classification accuracy is attributed due to the fact that overfitting and variance are reduced through random forest with embedded attribute optimization feature. Also it has an inbuilt attribute selection functionality during model training phase. When it combines with GBFS heuristic search, the attribute set is further optimized and gives an improved accuracy. The proposed GBFS-Random forest technique gave an optimum accuracy of 98.8%. Figure 11 depicts the comparative analysis.





**Figure 11. Comparison of classification accuracy of the proposed model with existing works**

A new predictive healthcare model is said to be efficient and reliable if it generates consistent outcome with heterogeneous disease datasets. Evaluation of the proposed IoHT enabled predictive model is effective only when the accuracy is good enough with variety of dataset samples. The proposed heuristic-based GBFS-Random forest model was tested against some cancer datasets with varying feature sets and instances, as highlighted in table 9. Skin cancer constituted the larger data samples with 1200 instances, while breast cancer had the least 286 data samples. It successfully reduced the feature set and generated an optimal feature set with all cancer datasets. The classification accuracy was also enhanced when the GBFS heuristic search method was

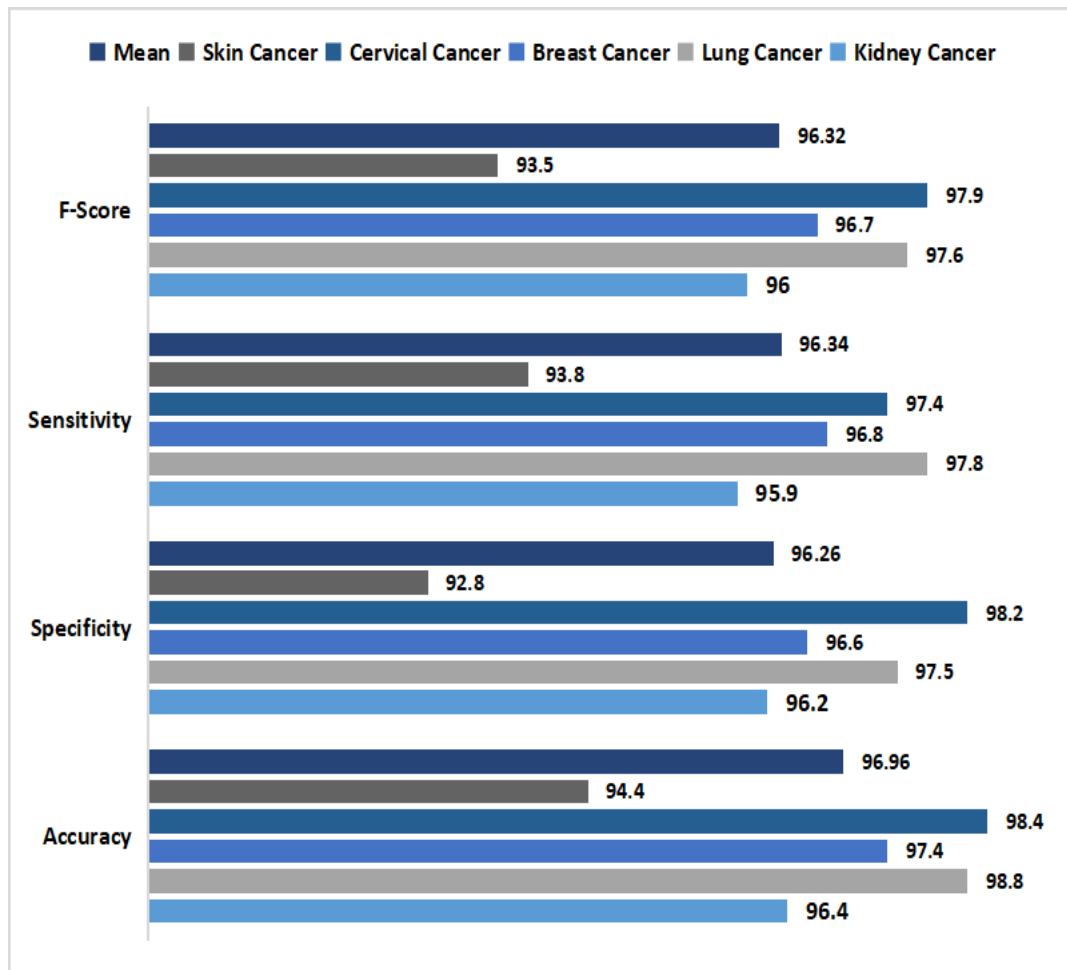
used in combination with a random forest classifier. Cervical cancer gave the highest accuracy of 98.4%, slightly less than 98.8% in lung cancer. Thus it is observed that the performance of the proposed IoHT model generates a consistent outcome with different cancer dataset with distinct samples taken into consideration.

**Table 9. Classification accuracy analysis of proposed model on different cancer datasets**

Cancer Dataset	Instances	Initial	Reduced	Classification	Classification
		feature Set	feature with GBFS Heuristic method	accuracy using Random Forest	accuracy using GBFS-Random Forest
Kidney Cancer	400	25	20	92.4	96.4%
Lung Cancer	346	16	12	95.1%	98.8%
Skin Cancer	1200	07	06	93.2	97.4%
Cervical Cancer	858	36	28	90.8	98.4%
Breast Cancer	286	9	07	91.5	96.96%

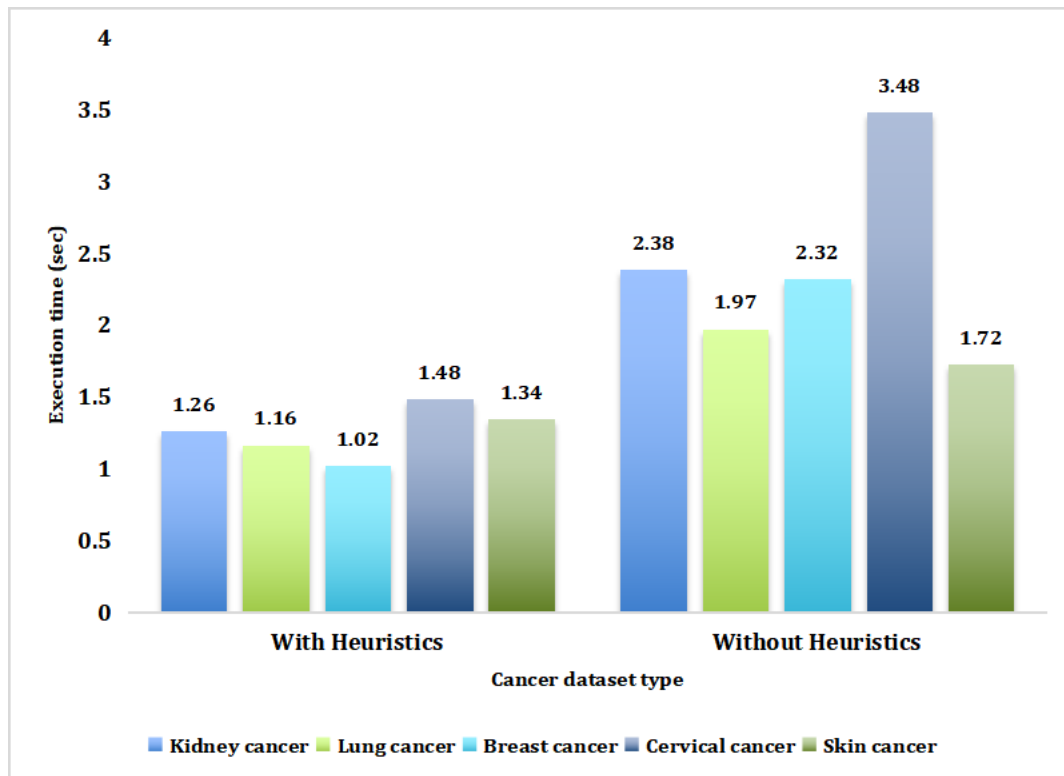
Various other performance indicators like specificity, sensitivity, and f-score metrics were used to determine the proposed hybrid heuristic classifier model's effectiveness. Five cancer datasets were taken into consideration for evaluation. It was observed that lung cancer classification generated the highest accuracy of 98.8%, while a slightly lower accuracy of 94.4% was noted with skin cancer data. 97.5% specificity value was noted with lung cancer, and a maximum specificity value of 98.2% was noted with cervical cancer. As far as sensitivity is concerned, lung cancer recorded the optimum value of 97.8%. A maximum value of 97.9% was generated with skin cancer data samples. Overall a very consistent classification performance was observed with the application of the proposed heuristic-based classification approach. The mean accuracy, specificity, sensitivity, and f-score value recorded was 96.96%, 96.26%, 96.34%, and

96.32%, respectively, over these cancer datasets. The results of the evaluation using these performance indicators are shown in Figure 12.



**Figure 12. Comparison of performance metrics on several cancer datasets**

An execution time latency analysis was conducted to compare classification with the heuristic approach and without heuristic over those cancer datasets, as depicted in figure 13. Classification with the heuristic based GBFS method proved to be very productive and fast. Execution time with heuristics-based GBFS method generated a very minimum delay of 1.26 sec, 1.16 sec, 1.02 sec, 1.48 sec, and 1.34 sec on kidney, lung, breast, cervical, and skin cancer, respectively.



**Figure 13. Comparison of the execution time of proposed model on several cancer datasets**

Results indicate the proposed heuristic-based classification model is very effective and accurate. It can help guide and assist medical professionals in the proper treatment of lung cancer and useful in detecting risk disorders at an early phase. The use of IoHT in smart cities to retain sustainability in the proposed model is immense. With the proposed model, real-time data healthcare providers can continuously monitor patients and thus they can identify any disease risks before it spreads further.

- ❖ Overall technological improvement: It avoids unnecessary medical visits and provides timely patient's health data processing and analysis.
- ❖ Enhanced accessibility: Doctors can access and monitor patient related data anytime and anywhere without leaving his place.
- ❖ Minimized errors: The proposed IoHT model permits accurate data accumulation, automation in workflow with reduced waste of resources thereby limits error risks.

- ❖ Cost effective: It is easy to manage and deploy. Patient visit to medical centers is reduced and lung cancer diagnosis and treatment is accelerated. Interconnected capability of the developed model further reduces patient stay at hospital and re-admissions. Hence the proposed IoT enabled model is cost effective.
- ❖ Remote care: Through our IoHT model, remote patient monitoring can be done effectively without the need of patient to meet medical experts face to face every time.

It significantly minimizes the unnecessary workload overhead that was done through manual procedures. It is fast and cost-effective too; thereby, it helps restrict resources and thus takes good care of the surrounding smart environment, making it a more reliable and sustainable model.

## **9. Conclusion**

Currently, the world is facing a tremendous urban transition. This global transition is one of the chief reasons for the change in environmental conditions, which impacts the sustainable human beings' health. Thus, the present scenario of this transition provides an excellent platform to live a healthy life. Sustainable cities play a vital role in this urbanization process and crucial for sustainability and health. Though this urbanization push has predominantly uplifted residents' health status, it brings some serious concerns with it. The rapid rise of chronic diseases in urban and sustainable cities is a significant issue. Handling of chronic disorders on a mass scale in sustainable cities is a prevailing challenge. In this study, an automated and computationally intelligent IoHT based lung cancer detection model is developed to be implemented in sustainable cities.

Lung cancer is seen as a predominant factor for deaths due to cancer in the current generation. Effective treatment of lung cancer is feasible if symptoms can be detected at early stages. The use of the latest technology through IoT and computational intelligence can help develop a sustainable prototype model for lung cancer treatment without harming the environment. It will reduce resource wastage, avoid unnecessary manual overloads, and offer faster lung cancer diagnosis with minimum manual intervention. In this research, a new hybrid machine learning model using the heuristics-based Greedy Best First Search (GBFS) algorithm is used for optimizing the lung cancer dataset. At the same time, a random forest classifier helps in classifying lung cancer patients based on their symptoms. The developed model was further implemented using python software programming language. Our IoHT based sustainable model's performance was evaluated against several metrics to determine the proposed model's effectiveness. It generated an optimal accuracy rate of 98.8% and a latency period of 1.16 seconds. Specificity and sensitivity recorded with the proposed model on lung cancer data were observed to be 97.5% and 97.8%, respectively. The evaluated model was validated with different cancer datasets collected from the UCI repository. The mean accuracy, specificity, sensitivity, and f-score value noted are 96.96%, 96.26%, 96.34%, and 96.32%, respectively, over these cancer datasets implemented. The results obtained is very satisfactory and it will be beneficial to society in developing a sustainable healthcare in smart cities. Patient can get real time treatment of lung cancer in cost effective manner with least latency and effort anytime and anywhere with more accuracy. It will uplift the healthcare standard of society as a whole. Thus, the proposed IoHT based computationally effective lung cancer detection model can be inferred as reliable and sustainable.

The developed model can be further upgraded and enhanced in future. It can be tested with larger and complex datasets too. The accuracy of lung cancer detection can be enhanced by using deep learning approach. It can be further tested with different disease datasets to form a homogeneous interface which can be deployed as a smart mobile application also in remote regions. Further the model can be optimized using deep learning hybrid models and deploying it as a smart phone application will make it even more convenient to use. Also the model can be made more secure so that it can operate in densely populated scenarios. Resource consumption is another factor which need to be optimized for future aspects.

## References

- [1] World Economic Forum. Sustainable health systems. Visions, strategies, critical uncertainties and scenarios. Geneva, Switzerland: World Economic Forum, 2013.
- [2] M. A. Rahman et al. 2020. Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach. IEEE Access. 8(2020), 205071-205087.
- [3] B. Sobolev, A. Levy, and S. Goring, Eds (2016).: "Health Services Data: Big Data Analytics for Deriving Predictive Healthcare Insights, in Data and Measures in Health Services Research". SpringerUS, 2016, 1(1)1–17, DOI: 10.1007/978-1-4899-7673-4\_2-1, ISBN 978-1-4899-7673-4.
- [4] Mishra, S., Tripathy, H. K., Mallick, P. K., Bhoi, A. K., & Barsocchi, P. (2020). EAGA-MLP—An Enhanced and Adaptive Hybrid Classification Model for Diabetes Diagnosis. Sensors, 20(14), 4036.
- [5] Mishra, S., Mallick, P. K., Tripathy, H. K., Bhoi, A. K., & González-Briones, A. (2020). Performance Evaluation of a Proposed Machine Learning Model for Chronic

Disease Datasets Using an Integrated Attribute Evaluator and an Improved Decision Tree Classifier. *Applied Sciences*, 10(22), 8137.

[6] Mishra, S., Mallick, P. K., Jena, L., & Chae, G. S. (2020). Optimization of Skewed Data Using Sampling-Based Preprocessing Approach. *Frontiers in Public Health*, 8.

[7] Mallick, P.K., Mishra, S. & Chae, G.S. Digital media news categorization using Bernoulli document model for web content convergence. *Pers Ubiquit Comput* (2020). <https://doi.org/10.1007/s00779-020-01461-9>

[8 ] Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society*, 54, 102010.

[9] Sohrabi, S., Khreis, H., & Lord, D. (2020). Impacts of Autonomous Vehicles on Public Health: A Conceptual Model and Policy Recommendations. *Sustainable Cities and Society*, 63, 102457.

[10] Setio, A.A.A.; Traverso, A.; de Bel, T.; Berens, M.S.N.; van den Bogaard, C.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B.; et al.. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med. Image Anal.* 2017, 42, 1–13.

[11] Jiang, H.; Ma, H.; Qian, W.; Gao, M.; Li, Y. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. *IEEE J. Biomed. Heal. Inform.* 2018, 22, 1227–1237.

[12] P. Bach, M. Kattan, M. Thornquist, et al., Variations in lung cancer risk amongsmokers, *J. Natl. Cancer Inst.* 95 (6) (2003) 470–478.

[13] P. Ramachandran, N. Girija, T. Bhuvanewari, Early detection and prevention ofcancer using data mining techniques, *Int. J. Comput. Appl.* 97 (13) (2014) 48–53.



- [14] P. Thangaraju, G. Barkavi, T. Karthikeyan, Mining lung cancer data for smokers and NonSmokers by using data mining techniques, *Int. J. Adv. Res. Comput. Commun.Eng.* 3 (7) (2014) 7622–7626.
- [15] T. Manikandan, N. Bharathi, M. Sathish, V. Asokan, Hybrid neuro-fuzzy system for prediction of lung diseases based on the observed symptom values, *J. Chem.Pharmaceut. Sci.* 8 (2017) 69–76.
- [16] H. Bharathi, T.S. Arulananth, A review of lung cancer prediction system using data mining techniques and self organizing map (SOM), *Int. J. Appl. Eng. Res.* 12 (10)(2017) 2190–2195.
- [17] S. Senthil, B. Ayshwarya, Lung cancer prediction using feed forward backpropagation neural networks with optimal features, *Int. J. Appl. Eng. Res.* 13 (1)(2018) 318–325.
- [18] M. Markaki, I. Tsamardinos, A. Langhammer, V. Lagani, K. Hveem, O.D. Røe, A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a hunt study, *EBioMedicine* 31 (2018) 36–46.
- [19] A. Cassidy, S. Duffy, J. Myles, T. Liloglou, J. Field, Lung cancer risk prediction: a tool for early detection, *Int. J. Canc.* 120 (1) (2006) 1–6.
- [20] Krishnaiah, V., Narsimha, G., SubhashChandra, N., Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, et al., *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 4 (1) , 2013.
- [21] Prashant Naresh, Early Detection of Lung Cancer Using Neural Network Techniques, *Journal of Engineering Research and Applications* Vol. 4, Issue 8, August 2014.

- [22] Gao, X., Chu, C., Li, Y., Lu, P., Wang, W., Liu, W. and Yu, L.: The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *European Journal of Radiology.*, Vol. 84, No. 2, pp. 312-317, 2015.
- [23] Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., Liu, W. and Yu, L.: Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images, *EJNMMI Research*, Vol. 7, No. 1, pp. 11, 2017.
- [24] Guo, N., Yen, R. F., El Fakhri, G. and Li, Q.: SVM based lung cancer diagnosis using multiple image features in PET/CT, in: *Proceedings of the Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, October 2015, IEEE, pp.1-4.
- [25] Zhao, J., Ji, G., Qiang, Y., Han, X., Pei, B., Shi, Z.: A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm, *Plos One*, Vol. 10, No. 4, pp e0123694, 2015.
- [26] B. Eric and T. Robert, —Semi-supervised methods to predict patient survival from gene expression data, in *PLoS Biol*, 2004, 2(4), pp. E108.
- [27] S. Deshmukh and S. Shinde, "Diagnosis of Lung Cancer using Pruned Fuzzy Min-Max Neural Network," *Int. Conf. Autom. Control Dyn. Optim. Tech. ICACDOT 2016*, pp. 398–402, 2017.
- [28] V. Krishnaiah, G. Narsimha, N. Subhash, —Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, 2013, 4 (1), 39 – 45.

- [29] K. Kourou, T.P. Exarchos, K.P. Exarchos, Karamouzis M.V. and Fotiadis, D.I., —Machine learning applications in cancer prognosis and prediction, in Computational and structural biotechnology journal, 2015, 13, pp.8-17.
- [30] A. Joseph, S. David, —Applications of Machine Learning in Cancer Prediction and Prognosis, 2 Issue published, 2006.
- [31] X. Luo, X. Zang, L. Yang, J. Huang, F. Liang, Rodriguez, —Comprehensive computational pathological image analysis predicts lung cancer prognosis, in Journal of Thoracic Oncology, 2017, 12(3), pp.501-509.
- [32] M. Xueyan, "Predicting Five-year Overall Survival in Patients with Non-small Cell Lung Cancer by ReliefF Algorithm and Random Forests", IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (2017).
- [33] J.H. Oh, R. Al-Lozi and I. El Naqa, —Application of machine learning techniques for prediction of radiation pneumonitis in lung cancer patients, In 2009 International Conference on Machine Learning and Applications (pp. 478-483). IEEE (2019).
- [34] Y. Xie, X. Wang and M. Story, —Statistical methods of background correction for Illumina BeadArray data, in Bioinformatics, 2009, 25(6), pp.751-757.
- [35] W. Sun, B. Zheng and W. Qian, —Computer aided lung cancer diagnosis with deep learning algorithms, In Medical imaging 2016: computer-aided diagnosis (Vol. 9785, p. 97850Z). International Society for Optics and Photonics, 2016.
- [36] S. Wienert, D. Heim, K. Saeger, —Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach, in Scientific reports, 2012.
- [37] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.

- [38] Jena, L., Mishra, S., Nayak, S., Ranjan, P., & Mishra, M. K. (2021). Variable Optimization in Cervical Cancer Data Using Particle Swarm Optimization. In *Advances in Electronics, Communication and Computing* (pp. 147-153). Springer, Singapore.
- [39] Milojevic-Dupont, N., & Creutzig, F. (2020). Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society*, 102526.
- [40] G. Muhammad, M. S. Hossain and N. Kumar (2021). EEG-Based Pathology Detection for Home Health Monitoring. *IEEE J. Sel. Areas Commun.* 39(2), 603-610,
- [41] M. A. Rahman and M. S. Hossain (2021). An Internet of Medical Things-Enabled Edge Computing Framework for Tackling COVID-19. *IEEE Internet of Things J.* (2021). 1-8, doi: 10.1109/JIOT.2021.3051080.
- [42] M. Shorfuzzaman, M. S. Hossain, and M. F. Alhamid (2021). Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic. *Sustainable Cities and Society*. 64(2021),102582, DOI: 10.1016/j.scs.2020.102582.
- [43] K.H. Abdulkareem et al., "Realizing an Effective COVID-19 Diagnosis System Based on Machine Learning and IOT in Smart Hospital Environment," in *IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2021.3050775.
- [44] Mutlag, A. A., Abd Ghani, M. K., Arunkumar, N. A., Mohammed, M. A., & Mohd, O. (2019). Enabling technologies for fog computing in healthcare IoT systems. *Future Generation Computer Systems*, 90, 62-78.
- [45] Lahoura V, Singh H, Aggarwal A, Sharma B, Mohammed MA, Damaševičius R, Kadry S, Cengiz K. Cloud Computing-Based Framework for Breast Cancer Diagnosis

Using Extreme Learning Machine. *Diagnostics (Basel)*. 2021 Feb 4;11(2):241. doi: 10.3390/diagnostics11020241. PMID: 33557132; PMCID: PMC7913821.

[46] Mutlag AA, Khanapi Abd Ghani M, Mohammed MA, Maashi MS, Mohd O, Mostafa SA, Abdulkareem KH, Marques G, de la Torre Díez I. MAFC: Multi-Agent Fog Computing Model for Healthcare Critical Tasks Management. *Sensors*. 2020; 20(7):1853. <https://doi.org/10.3390/s20071853>.