

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

## Supervised Deep Learning Approach To Identify the Lung Cancer

Ram Nikhilesh.M<sup>1</sup>, Indresh.M<sup>2</sup>, Vishal Vineeth<sup>3</sup>, Pranamita Nanda<sup>4</sup>

UG Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,  
Tamil Nadu, India<sup>1,2,3</sup>

Assistant Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,  
Tamil Nadu, India<sup>4</sup>

**ABSTRACT:** With the huge development of technology and rapid growth of data in various sectors leads to the need of extracting the knowledge from it. The knowledge could be achieved through the data analytics. The analytics has found its application in industry, health care, information technology, e-commerce, education, banking, etc. Since India is second in population, the diseases had coupled. This has necessitated the introduction of data analytics in health care for diagnoses of diseases. Of this cancer is a fatal disease, it is diagnosed at the stage 4. Among all, the people are affected more by lung cancer. It is estimated that 1.38 million cancer deaths per year worldwide. Lung cancer accounts for more deaths than any other type of cancer. The diagnosing of lung cancer at early stage is quite difficult. The cancer analysis needs complex equipment's like X-Rays, Computer Tomography (CT) Scan, Sputum analysis and other imaging process. They are not only complex but also costlier. These equipment's provide accurate result when the patient is at the last stages either stage 3 or stage 4. This could be overcome by developing a system which analyzes the past history of an individual's or a patient's behavior and also based on the result of simple diagnosing method (blood test) using Machine learning algorithm. This paper involves the process of identifying efficient algorithm to develop a system. Based on the result of the pre-diagnosis, the doctor can perform confirm the cancer on further analysis and the treatment can be started at early stages. The developed system might improve the survival rate of the patients.

**KEYWORDS:** lung cancer, data analytics, machine learning algorithm, diagnosis system

### I.LITERATURE SURVEY

**Juliet et.al**<sup>[1]</sup> does survey on various data mining techniques in predicting cancer and proposed a methodology of introducing a data mining (Data Integration, Data Selection, Data Cleaning, Data Transformation, Data Classification) instead of traditional diagnosis system. The author has considered the dataset which contains basic attributes, primary and secondary symptoms. Artificial Neural Network and the self organizing map (SOM) yields better performance even in the missing information. SOM is used for visualization. Data Cleaning is the challenging step.

**Fatma et.al**<sup>[2]</sup> used Hopfield Neural Network and Fuzzy C-Mean algorithm to segment the cancer cells in CAD (Computer Aided Diagnosis) system images. Hopfield Neural Network is one of the artificial neural networks to segment colored and gray images. Fuzzy C-Mean is a clustering algorithm. It physically clusters the logically similar objects. Fuzzy C-Mean uses reciprocal distance to calculate the fuzzy weights. Finally Hopfield Neural Network, has segmented the images better than the Fuzzy C-mean.

### II.EXISTING SYSTEM

In recent past, various disease prediction systems have been developed and in operation. The existing systems deploy a combination of machine learning algorithms which are reasonably accurate in predicting diseases. However the

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

limitation with the existing systems are varied. First, the existing systems are costlier only rich people could afford to such prediction systems. And also, when it comes to individuals, it becomes even more costlier. Second, the prediction systems are generic and vague so far. So that, a machine can predict a certain disease but cannot predict the sub types of the diseases and diseases caused by the occurrence of one disease. For instance, if a group of people are predicted with Diabetes, probably some of them might have higher risk for Heart diseases due to the existence of Diabetes. The existing systems fail to predict all possible conditions of the patient.

### III. PROPOSED SYSTEM

We begin with the concept that was not implemented by the predecessors. We implement Decision Tree machine learning algorithm for predicting disease as well as predicting all the other possible sub diseases. We employ MapReduce algorithm for partitioning the data such that a query would be analysed only in the specific partition, which will increase operational efficiency but reduce query retrieval time. In addition to that, we provide individual provisions for individual clients to check his/her condition. Thus making our application broadly accessible by all at affordable cost.

#### Subsystem

##### Data Collection

A login, logging in or logging on is the entering of identifier information into a system by a user in order to access that system (e.g., a computer or a website). It is an integral part of computer security procedures. A login generally requires the user to enter two pieces of information, first a user name and then a password. This information is entered into a login window on a GUI (graphical user interface) or on the command line in a console (i.e., an all-text mode screen), depending on the system and situation. A user name, also referred to as an account name, is a string (i.e., sequence of characters) that uniquely identifies a user. User names can be the same as or related to the real names of users, or they can be completely arbitrary.

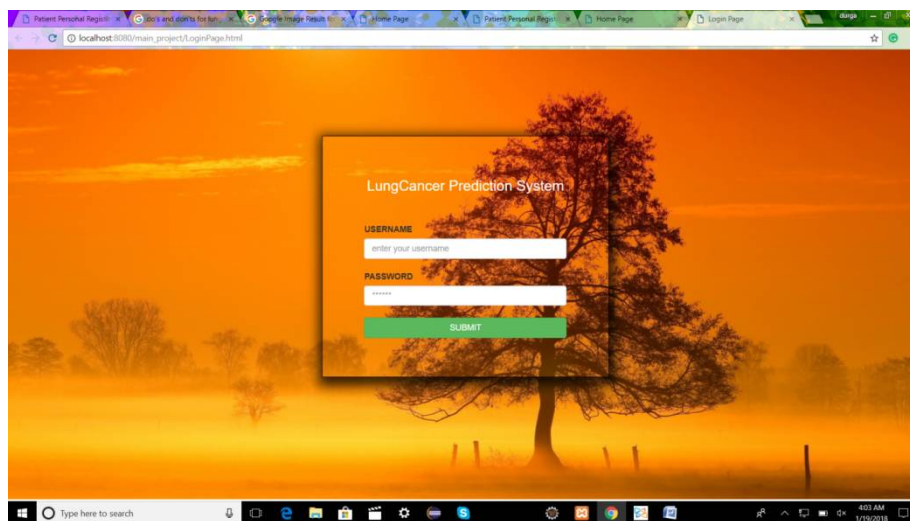


Figure 3.1 :DataCollection

##### Data Pre-processing

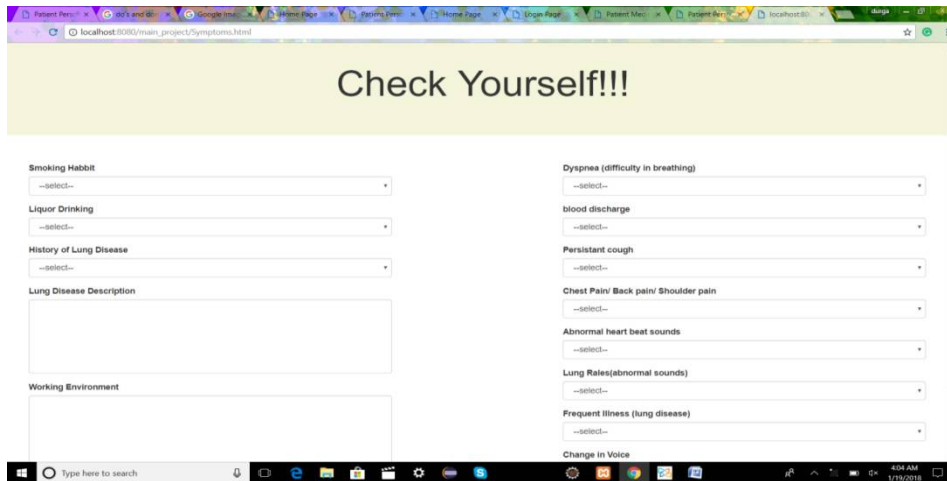
After successful login user can upload their medical information through this application. All user data encrypted initially then save into databases. Uploading module get the information of user age, living area, basic medical information like BP, sugar, Height, weight ,etc.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

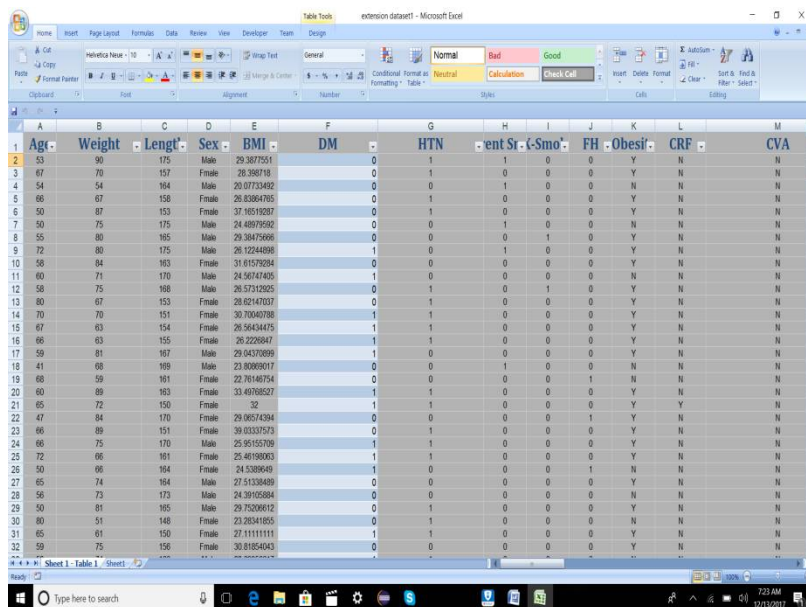
Vol. 7, Special Issue 2, March 2018



The screenshot shows a web browser window with a form titled "Check Yourself!!!". The form contains several dropdown menus and text input fields for medical history. The fields include: Smoking Habit, Liquor Drinking, History of Lung Disease, Lung Disease Description, Working Environment, Dyspnea (difficulty in breathing), blood discharge, Persistent cough, Chest Pain/ Back pain/ Shoulder pain, Abnormal heart beat sounds, Lung Rales(abnormal sounds), Frequent illness (lung disease), and Change in Voice.

## Module Training & Testing

Machine Learning (ML) provides methods, techniques, and tools that can help solving diagnostic and prognostic problems in a variety of medical domains. ML is being used for the analysis of the importance of clinical parameters and their combinations for prognosis, e.g. prediction of disease progression, extraction of medical knowledge for outcome research, therapy planning and support, and for the overall patient management. ML is also being used for data analysis, such as detection of regularities in the data by appropriately dealing with imperfect data, interpretation of continuous data used in the Intensive Care Unit, and intelligent alarming resulting in effective and efficient monitoring. It is argued that the successful implementation of ML methods can help the integration of computer-based systems in the healthcare environment providing opportunities to facilitate and enhance the work of medical experts and ultimately to improve the efficiency and quality of medical care. Below, we summarize some major ML applications in medicine.



	A	B	C	D	E	F	G	H	I	J	K	L	M
	Age	Weight	Length	Sex	BMI	DM	HTN	Sr.	c-Smo	FH	Obesi	CRF	CVA
2	53	90	175	Male	29.3877551	0	1	1	0	0	Y	N	N
3	67	70	157	Female	28.388718	0	1	0	0	0	Y	N	N
4	54	54	164	Male	25.07733492	0	0	1	0	0	N	N	N
5	68	67	158	Female	26.8384785	0	1	0	0	0	Y	N	N
6	50	67	153	Female	27.16519287	0	1	0	0	0	Y	N	N
7	50	75	175	Male	24.4897592	0	0	1	0	0	N	N	N
8	55	80	185	Male	29.38475666	0	0	0	1	0	Y	N	N
9	72	80	175	Male	26.1224486	1	0	1	0	0	Y	N	N
10	58	64	163	Female	31.61572654	0	0	0	0	0	Y	N	N
11	60	71	170	Male	24.56747405	1	0	0	0	0	N	N	N
12	58	75	168	Male	26.57312925	0	1	0	1	0	Y	N	N
13	80	67	153	Female	28.62147037	0	1	0	0	0	Y	N	N
14	70	70	151	Female	30.70040798	1	1	0	0	0	Y	N	N
15	67	63	154	Female	29.56534315	1	1	0	0	0	Y	N	N
16	66	63	155	Female	26.2226847	1	1	0	0	0	Y	N	N
17	59	81	167	Male	29.04370899	1	0	0	0	0	Y	N	N
18	41	68	169	Male	23.80086917	0	0	1	0	0	N	N	N
19	68	59	161	Female	22.76146754	0	0	0	0	1	N	N	N
20	60	69	163	Female	33.49305227	1	1	0	0	0	Y	N	N
21	65	72	150	Female	32	1	1	0	0	0	Y	Y	N
22	47	84	170	Female	29.06574394	0	0	0	0	1	Y	N	N
23	66	89	151	Female	39.03337373	0	1	0	0	0	Y	N	N
24	66	75	170	Male	25.95150799	1	1	0	0	0	Y	N	N
25	72	66	161	Female	25.46190983	1	1	0	0	0	Y	N	N
26	50	66	164	Female	24.5389649	1	0	0	0	1	N	N	N
27	65	74	164	Male	27.51338489	0	0	0	0	0	Y	N	N
28	58	73	173	Male	24.39105884	0	0	0	0	0	N	N	N
29	50	81	165	Male	29.75206612	0	1	0	0	0	Y	N	N
30	80	51	148	Female	23.28541855	0	1	0	0	0	N	N	N
31	65	61	150	Female	27.11111111	1	1	0	0	0	Y	N	N
32	59	75	156	Female	30.81854043	0	0	0	0	0	Y	N	N

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

## IV.RESULTS

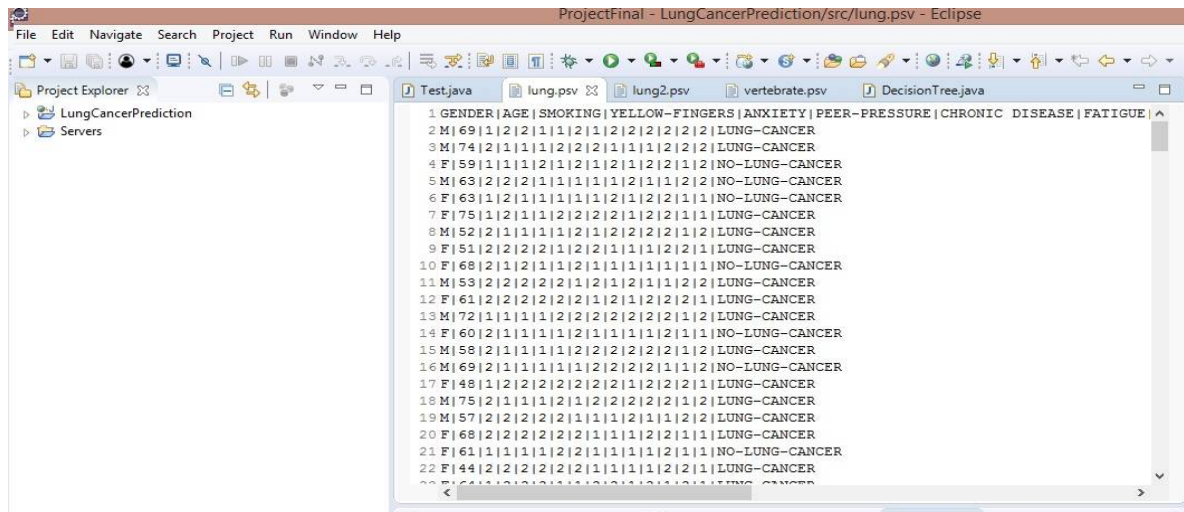


FIG 4.1 : Data sets

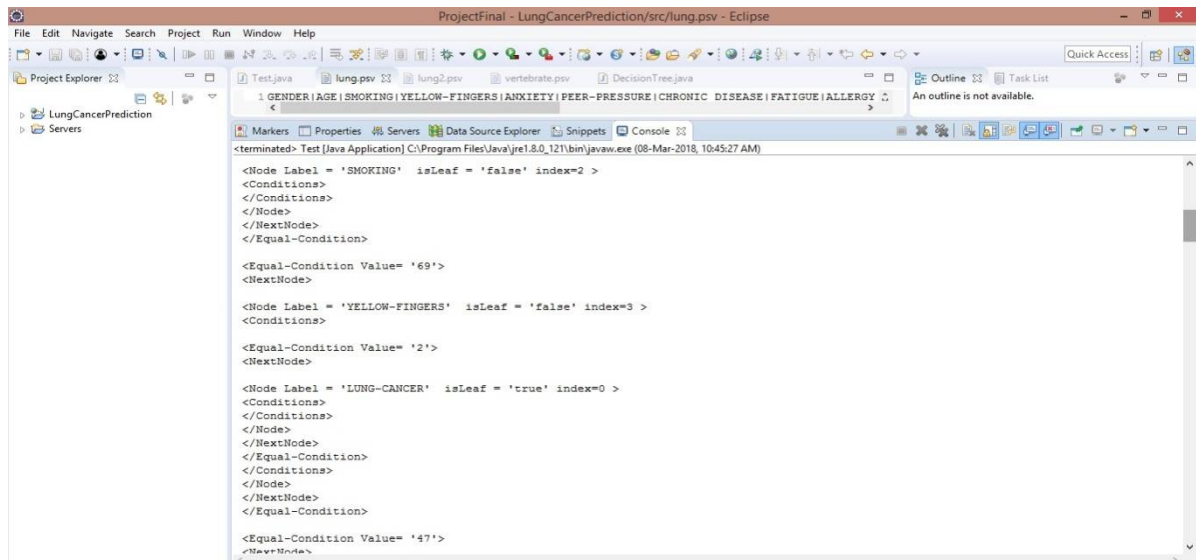


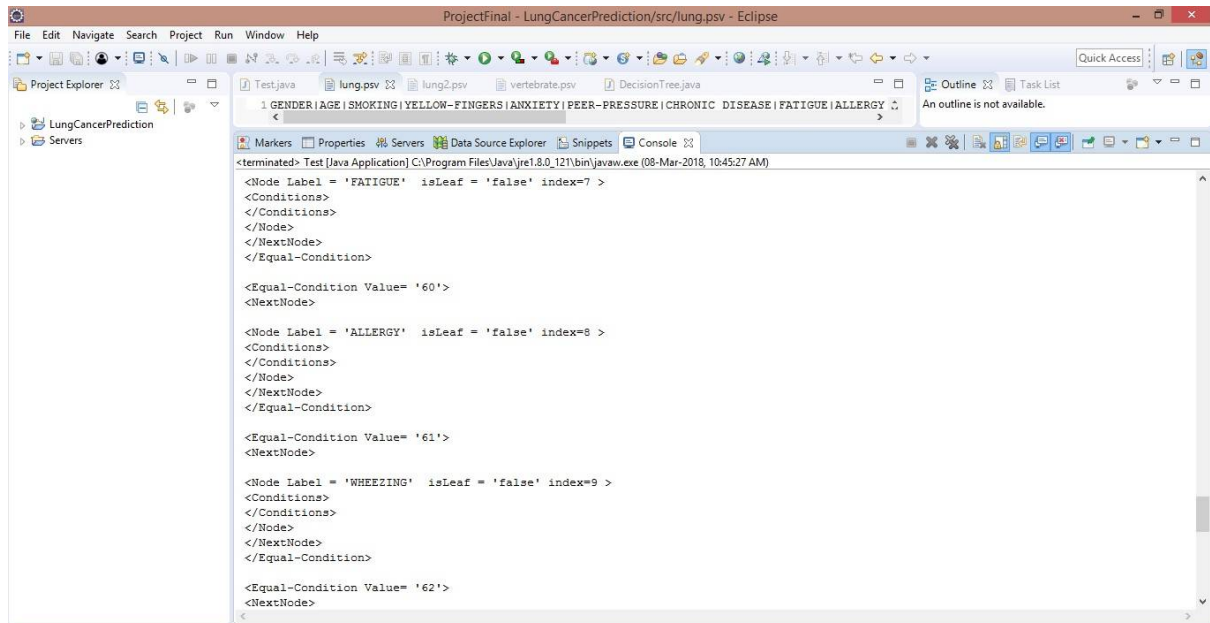
FIG 4.2 : Data Validation

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

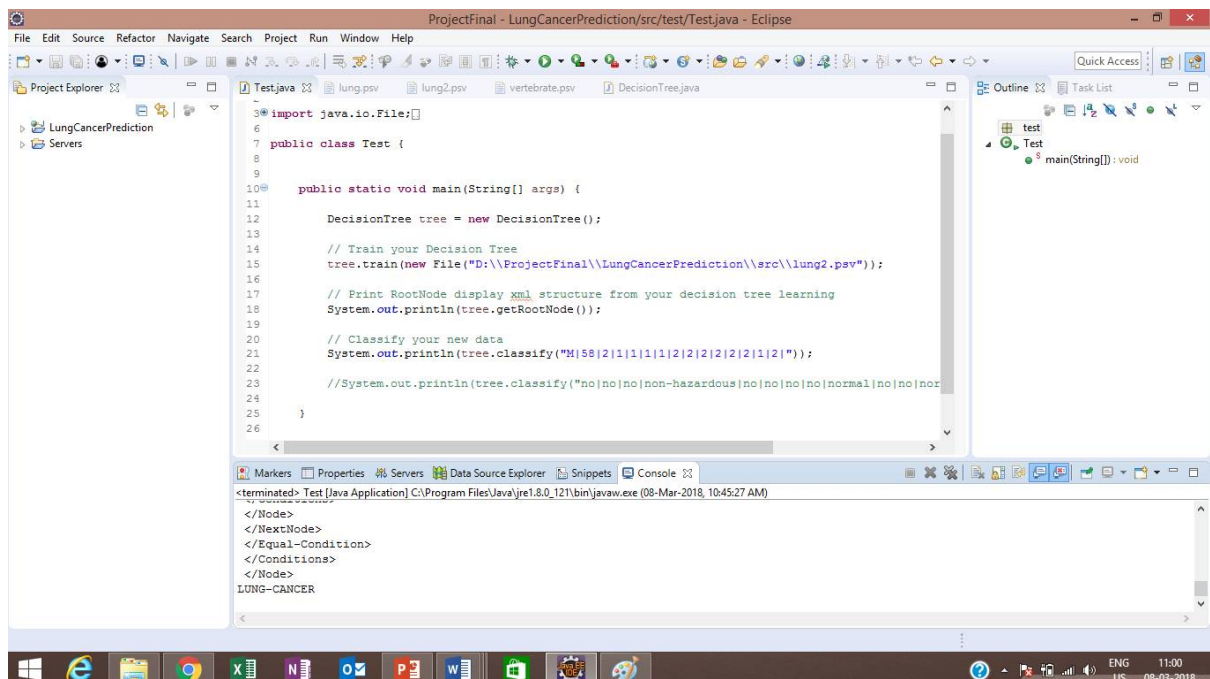


```

ProjectFinal - LungCancerPrediction/src/lung.psv - Eclipse
File Edit Navigate Search Project Run Window Help
Project Explorer
LungCancerPrediction
Servers
Test.java lung.psv lung2.psv vertebrate.psv DecisionTree.java
1 GENDER | AGE | SMOKING | YELLOW-FINGERS | ANXIETY | PEER-PRESSURE | CHRONIC DISEASE | FATIGUE | ALLERGY
<terminated> Test [Java Application] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (08-Mar-2018, 10:45:27 AM)
<Node Label = 'FATIGUE' isLeaf = 'false' index=7 >
<Conditions>
</Conditions>
</Node>
</NextNode>
</Equal-Condition>
<Equal-Condition Value= '60'>
<NextNode>
<Node Label = 'ALLERGY' isLeaf = 'false' index=8 >
<Conditions>
</Conditions>
</Node>
</NextNode>
</Equal-Condition>
<Equal-Condition Value= '61'>
<NextNode>
<Node Label = 'WHEEZING' isLeaf = 'false' index=9 >
<Conditions>
</Conditions>
</Node>
</NextNode>
</Equal-Condition>
<Equal-Condition Value= '62'>
<NextNode>

```

FIG 4.3 : Data Analyzing



```

ProjectFinal - LungCancerPrediction/src/test/Test.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Project Explorer
LungCancerPrediction
Servers
Test.java lung.psv lung2.psv vertebrate.psv DecisionTree.java
3 import java.io.File;
6
7 public class Test {
8
9
10 public static void main(String[] args) {
11
12     DecisionTree tree = new DecisionTree();
13
14     // Train your Decision Tree
15     tree.train(new File("D:\\ProjectFinal\\LungCancerPrediction\\src\\lung2.psv"));
16
17     // Print RootNode display xml structure from your decision tree learning
18     System.out.println(tree.getRootNode());
19
20     // Classify your new data
21     System.out.println(tree.classify("M|58|2|1|1|1|1|2|2|2|2|2|2|2|2"));
22
23     //System.out.println(tree.classify("no|no|no|non-hazardous|no|no|no|no|normal|no|no|nor
24
25 }
26
<terminated> Test [Java Application] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (08-Mar-2018, 10:45:27 AM)
</Node>
</NextNode>
</Equal-Condition>
</Conditions>
</Node>
LUNG-CANCER

```

FIG 4.4 : Final Output



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

## V. CONCLUSION

In this review we have attempted to explain, compare and assess the performance of different machine learning that are being applied to cancer prediction and prognosis. Specifically we identified a number of trends with respect to the types of machine learning methods being used, the types of training data being integrated, the kinds of endpoint predictions being made, the types of cancers being studied and the overall performance of these methods in predicting cancer susceptibility or outcomes. It is likely that the use of machine learning classifier will become much more commonplace in many clinical and hospital settings.

## REFERENCES

1. Juliet Rani Rajan and ChilambuChelvan C, "A Survey on Mining Techniques for Early Lung Cancer Diagnoses", 2013 IEEE..
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4991146/#sec1-1title> .
3. FatmaTaher and RachidSammouda, "Lung cancer detection by using artificial neural network and fuzzy clustering methods", 2011 IEEE GCC Conference and Exhibition (GCC), February 19-22, 2011.
4. RitikaAgarwal, AnkitShankhadhar and Raj Kumar Sagar, "Detection of lung cancer using content based medical image retrieval", 2015 Fifth International Conference on Advanced Computing & Communication Technologies, 2015 IEEE.
5. Hamid Reza Hassanzadeh, John H. Phan and May D. Wang, "A Multi-Modal Graph-Based Semi-Supervised Pipeline for Predicting Cancer Survival", 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
6. Devi Nurtiyasari, DediRosadi and Abdurakhman, "The Application of Wavelet Recurrent Neural Network/or Lung Cancer Classification", 2017 3rd International Conference on Science and Technology - Computer (ICST).
7. ThanhTrungGiang, Thanh Phuong Nguyen and Dang Hung Tran, "Stratifying Cancer Patients based on Multiple Kernel Learning and Dimensionality Reduction", 2017 IEEE 9th International Conference on Knowledge and Systems Engineering (KSE).
8. Xiaozhou Li, Rong Wang and Ming Lei "Analysis on data for detection of Lung cancer using serum auto-fluorescence", 2011 IEEE.
9. Weixing Wang and Shuguang Wu, "A Study on Lung Cancer Detection by Image Processing", 2006 IEEE Saranya P and Satheeskumar B, "A Survey on Feature Selection of Cancer Disease Using Data Mining Techniques", International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 713-719.
10. Dmitry Ignatov and AndreyIgnatov, "Decision Stream: Cultivating Deep Decision Trees", 3 Sep 2017 IEEE.
11. Kelvin KF Tsoi1, Yong-Hong Kuo and Helen M. Meng, "A Data Capturing Platform in the Cloud for Behavioral Analysis Among Smokers An Application Platform for Public Health Research", 2015 IEEE.
12. Gonzalez-Alonso P, VilarR, Lupiañez-Villanueva F, "Meeting Technology and Methodology into Health Big Data Analytics Scenarios", 2017 IEEE.
13. Balaji S, MeghanaPatil and Carolyn McGregor AM, "A Cloud based Big Data Based Online Health Analytics for Rural NICUs and PICUs in India: Opportunities and Challenges", 2017 IEEE.