



L OVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

**DEVELOPMENT OF LIVER CANCER ONTOLOGY (LCO)
USING ONTOLOGY BASED TOOLS**

ADISSERTATIONSUBMITTED

BYSTUDENT

Parminder Kaur

RK2213A06

TO

Department of Computer Science

In partial fulfilment of the Requirement for the

**AWARD OF THE DEGREE OF
MASTER OF TECHNOLOGY
IN COMPUTER SCIENCE**

**UNDERTHEGUIDANCEOF
ADITYA KHAMPARIA**

(MAY 2015)

School of: Computer Science and Engineering

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the student : Parminder Kaur
 Batch : 2013-2015
 Session : 2014-2015

Registration No : 41200386
 Roll No : A13
 Parent Section :

Details of Supervisor:

Name : Adivya khamparia
 UID : 17862

Designation : Assistant Professor
 Qualification : M.Tech
 Research Exp. : 2 years

Specialization Area: Networking (Artificial Intelligence)

Proposed Topics:-

1. Development of Liver cancer ontology (LCO) using ontology based tools.
2. Ontology development on liver cancer
3. Performance analysis and development on education ontology

Adivya Khamparia
 Signature of supervisor

PAC Remarks:
Topic 1 is Approved
(Signature) 11/5/14

APPROVAL OF PAC CHAIRMAN

(Signature) 11/01/14
 Signature: _____ Date: _____

- *Supervision should finally encircle one topic out of three proposed topics and put up for an approval before Project Approval Committee (PAC).
- *Original copy of this format after PAC approval will be retained by the student and must be attached in the Project/Dissertation final report.
- *One copy to be submitted to supervisor.

ABSTRACT

Ontology is the study which depicts the nature of human beings and their relatives. Ontology is the study about what kinds of things exist, what entities are there in the universe. It helps in understanding those imperative things or some domain which can be systems. It helps in innovation of advanced and better information systems in health care. It also plays a vital role in revealing and sharing patient details. SPARQL queries are used for the diagnosis of liver cancer ontology through protégé. In data mining with the help of weka tool we can classify the liver based diseases using random tree algorithm. Classification is based on five parameters: Neurological, Psychiatric, Pathological, Physical, and Cognitive. The results are shown in terms of random decision tree with true positive rate, false positive rate and ROC curve.

CERTIFICATE

This to certified that Parminder Kaur has completed M.Tech dissertation titled Development of Liver Cancer Ontology (LCO) Using Ontology based tools

Under my guidance and supervision. To the best of my knowledge, the present work is there result of her original investigation and study. No part of the dissertation has ever been submitted for any other degree or diploma.

The dissertation is fit for the submission and the partial fulfilment of the conditions for the award of M.Tech Computer Science & Egg.

Dated:

Signature of Advisor

Name:

ACKNOWLEDGEMENT

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life. I acknowledge with deep sense of gratitude and most sincere appreciation, the valuable guidance and unfailing encouragement rendered to me by Mr. Aditya Khamparia for his proficient and enthusiastic guidance, useful encouragement and immense help.

I would like to place on record my deep sense of gratitude to (Dept. of Computer Science), LPU, Phagwara, Punjab for his generous guidance, help and useful suggestions.

.

DECLARATION

I here by declare that the dissertation entitled, Development of Liver cancer Ontology (LCO) Using Ontology Based tools submitted for the M.Tech Degree is entirely my original work and all ideas and reference have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Dated:-

Investigator

Regn No.

TABLE OF CONTENTS

1. INTRODUCTION

1.1	Introduction.....	1
1.2	Types of Ontologies.....	1-2
1.3	Data Mining.....	2
1.4	Liver cancer.....	3
1.5	Types of Liver Cancer.....	3
1.6	Risk factors of liver Cancer.....	4
1.7	Symptoms.....	4
1.8	Diagnosis.....	5
1.9	Treatment.....	6
1.10	Investigation.....	7
1.11	Prognosis.....	8
1.12	TNM Stages.....	8-9
1.13	Protege Tools.....	9-12
1.14	Applications of Protege.....	12-13

2. REVIEW OF LITERATURE

2.1	Literature Survey.....	14-27
-----	------------------------	-------

3. PRESENT WORK

3.1	Problem Formulation.....	28
3.2	Objectives.....	28-29
3.3	Research Methodology.....	29-31

4. OBSERVATIONS & DISCUSSIONS

4.1	Tools.....	32-33
4.2	Problem Implementation.....	34-48

5. CONCLUSION	
5.1 Conclusion.....	49
5.2 Future Scope.....	49
6. REFERENCES	
6.1 research papers.....	50-52
6.2 web sites.....	53
7. APPENDIX	
7.1 list of abbreviations.....	54

LIST OF TABLES

Table 1.1 Liver Cancer Survivability Rate.....	8
Table 4.1 Problem Description Table.....	36
Table 1.3 Detailed Description of Dataset.....	39

LIST OF FIGURES

Fig 1.1 Risk factors of Liver Cancer.....	4
Fig 1.2 Symptoms of Liver Cancer.....	4
Fig 1.3 Diagnosis of Liver Cancer.....	5
Fig 1.4 Treatment of Liver Cancer.....	6
Fig 1.5 Flow Diagram for Liver Cancer Diagnosis.....	7
Fig 4.1 Liver Cancer Diseases.....	3
Fig 4.2 Flow Diagram for Liver Based Diseases.....	34
Fig 4.3 Hierarchical Correlation of Sign and Symptoms of Liver Cancer.....	35
Fig 4.4 Dataset of Liver Cancer.....	36
Fig 4.5 Pre-process of Liver Cancer.....	37
Fig 4.6 Random Tree Generates Rules.....	39
Fig 4.7 Random Decision Tree Generated By Weka.....	40
Fig 4.8 Confusion Matrix.....	41
Fig 4.9 TPR and FPR Values.....	42
Fig 4.10 ROC Curve.....	42
Fig 4.11 High Level Terminology of Liver Cancer Ontology.....	44
Fig 4.12 Taxonomical Hierarchy of Concepts in Liver Cancer Ontology.....	44
Fig 4.13 Classes and Subclasses of Liver Cancer.....	45
Fig 4.14 Onto graph of Liver Cancer Ontology.....	46
Fig 4.15 Query related Wilson Disease.....	47
Fig 4.16 Query Related Tobacco.....	47

1.1 INTRODUCTION

Ontology is the study about what kinds of things exist, what entities are there in the universe. It is used for sharing and common understanding of some domain that can be communicated between people and application systems. Ontology helps us to build more powerful and interoperable information systems in health care.

Ontology which describes the relations of organizations is the paramount thing for making health systems in real medical scenario. Medical errors can be prevented by taking relevant precautions. They occur due to human negligence, inaccurate information and improper flow of communication in the clinic or hospital. Ontology eradicates various coding disorders, builds proper communication channel in the hospital, and helps in development of data related to project.

1.2 TYPES OF ONTOLOGIES

- ❖ **Domain Ontology:** Domain ontology consists of in-depth information for specific area (for an instance: medical).
- ❖ **Metadata Ontology:** Metadata ontology depicts the vocabulary for revealing what is written or described in online information.
- ❖ **Generic Ontology:** Generic ontology captures general knowledge about time, space and event.
- ❖ **Representational Ontology:** Representational ontology provides representational entities without stating what is represented.

There are different types of ontology tools are:

Ontology tools are important for ontology development process like building, merging, annotation, visualization etc which is used in different applications.

Types of ontology tools are:

Ontology tools are classified into development tools, merge and alignment tools, evaluation tools, annotation, querying and learning tools.

1. **Ontology Development Tool:** - Ontology development tool is helpful in building ontology in its initial stage and these tools are also redefines or reusing the existing ontology in an artificial environment.
2. **Ontology Merge and Alignment Tool:-** Ontology Merge and Alignment tools are used for solving the problem of merging or aligning different ontology on same and different domain.
3. **Ontology Evaluation Tool:-** Ontology Evaluation tools are used to evaluate the content of ontology. There are different types of evaluation and methodology which reduces problem difficulty and manages to reintegrate and reuse frequently.
4. **Ontology Annotation and Querying Tools:-** These tools allow the users to maintain ontology based mark up in web languages and querying tools .It enables information retrieval in ontology very easily with the help of their syntactic and semantic behaviour.
5. **Ontology Learning Tools:-** These tools allow the users to derive ontology from natural language texts, semi structured sources and databases through the machine learning and natural language techniques.

OWL is a knowledge representation of languages. OWL is used for sharing semantic web contents. It is implemented through the protégé OWL. Classes are the most important classes. The main property of OWL is annotation property that is used to add information i.e. data about data.OWL is a network of classes, properties and individuals.

1.3 Data mining

Data mining consist of two operations such as predictive and descriptive. Predictive include regression, classification and descriptive include deviation detection and clustering.

Random tree is used for classification and regression problems. It is a collection of tree predictors that is called forest. There is no need for any accuracy estimation procedures in random trees, such as cross-validation or bootstrap. Random tree is a class used for constructing a tree that considers k randomly chosen attributes at each node. Random tree provides a powerful technique for classification .Various random tree are available to classify the data, including ID3, C4.5, C5, J48, CART.

We develop the liver cancer ontology through protégé knowledge representation tool on the basis of symptoms, risk factors, causes, diagnosis, prevention and treatment with the onto graph. With the help of weka tool we classify the diseases on the basis of attributes high, medium, low, medium/low. We also generate the decision tree with the classification and through the attributes of liver cancer.

1.4 LIVER CANCER

The liver cancer is the sixth most common cancer in the world. Liver is a single organ. It is divided into two parts such as left and right hemi liver. Liver is the largest abdominal organ. Liver consist of primary and secondary liver cancer. Primary liver cancer is a cancer that originates in the liver. Secondary liver cancer is a cancer that spreads to all other parts of the body. With neurological, Psychiatric, pathological, physical and cognitive disordering occur some inflammation, cirrhosis, fibrosis and disruptions in the liver that create some problems like memory loss, vision, failing voice, enlarged liver, weight loss etc. If the liver fails, it causes many problems like malnutrition, jaundice, fatigue; there are different types of liver cancer such as fatty liver, cholestatic, autoimmune, inherited, Wilson.

1.5 TYPES OF LIVER CANCER

- ❖ **Fatty liver diseases:** Fatty liver builds up of fat in the liver. It is caused by alcoholism, drugs, viruses, excess body weight. The sign and symptoms of FLD are FG, WTL, WK, PRA, and LOC.
- ❖ **Cholestatic liver diseases:** Cholestatic diseases are of two types: primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC). The sign and symptoms of CLD are FG, IS, JA, SC, PRA, FV.
- ❖ **Autoimmune diseases:** Autoimmune diseases are diseases in which immune system attacks your liver cells. The sign and symptoms of AID are EL, DU, ABVS, AD, FG, VM, and JP.
- ❖ **Inherited liver diseases:** Inherited is one of the most common systematic disorders. It is a pathologic impact on the liver. The sign and symptoms of IHD are JP, FG, WTL, AP, DI, CC, JA, and SAF.
- ❖ **Wilson disease:** Wilson is an inherited disorder. It strikes in between the ages of 12 and 23. The sign and symptoms of WD are SS, ML, FV, BC, DP and FG.

Liver cancer is likely to be caused by a combination of risk factors. The liver cancer is described in its risk factors which consist of preventable and non-preventable risk factors.

1.6 RISK FACTORS

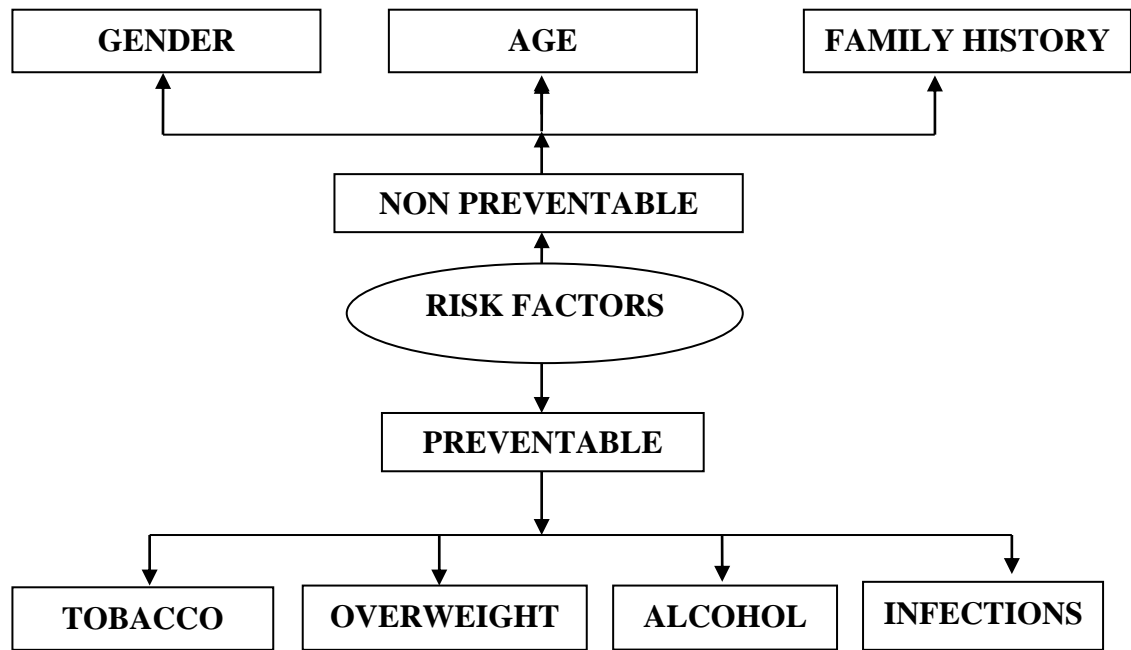


Figure 1.1: RISK FACTORS OF LIVER CANCER

1.7 SYMPTOMS

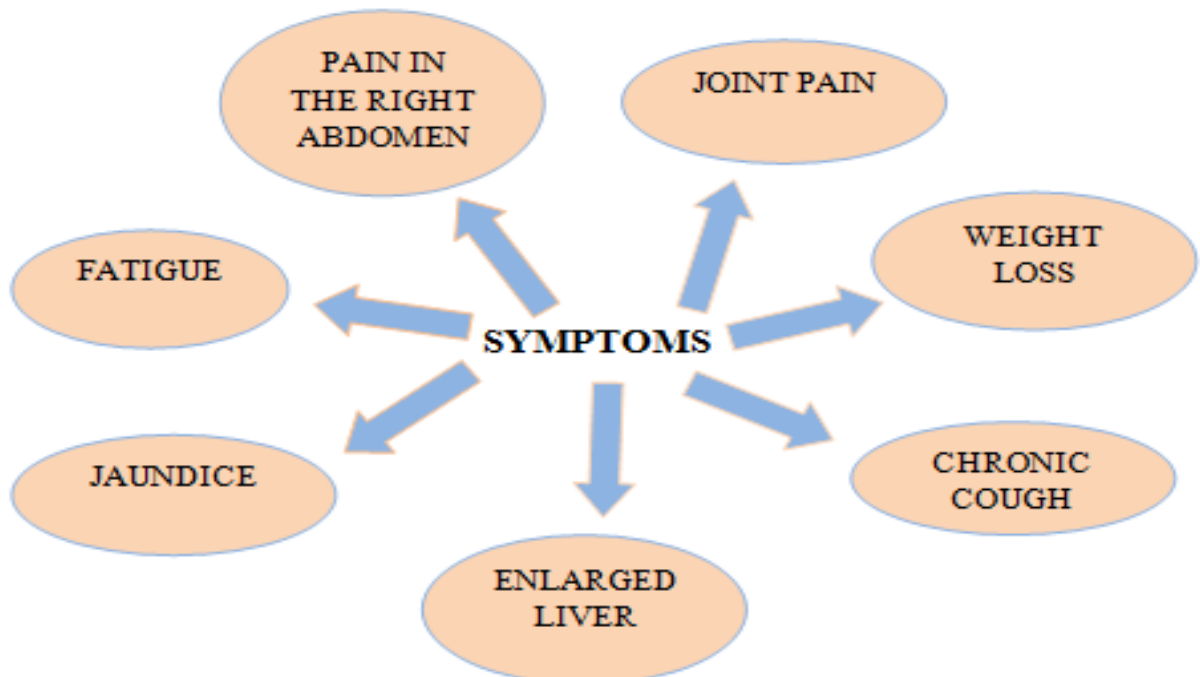


Figure1.2: SYMPTOMS OF LIVER CANCER

On the basis of symptoms we will diagnose the patient. Liver cancer can be found after symptoms appear, but for many cases early liver cancer has no symptoms.

1.8 DIAGNOSIS

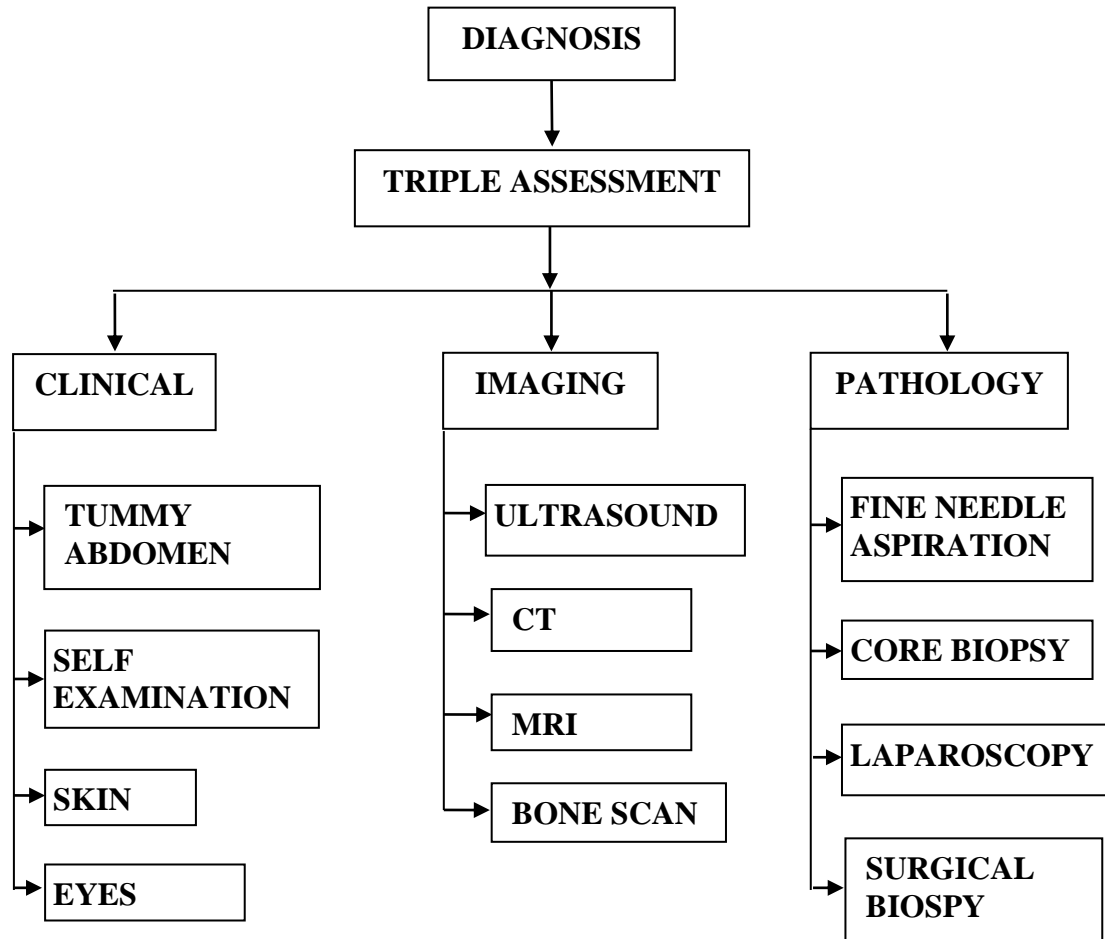


FIGURE 1.3: DIAGNOSIS OF LIVER CANCER

Hence , after performing screening tests before any symptoms develop is very essential. Early detection of liver cancer in the localized stage, increases the 5-year survival rate to 28%.

The tests can be performed are classified into a triple assessment routine. Liver Cancer can be divided into clinical, imaging and pathology assessment.

1.9 TREATMENT

The treatment of liver cancer depends on the type of liver cancer. The treatment of liver cancer can be divided into two parts: Surgical and non- Surgical treatment. Liver Cancer diagnosed at a later stage requires a different treatment than when diagnosed in its early stages. A patient may have one treatment or a combination. Treatment of liver cancer can be summarised as shown below:-

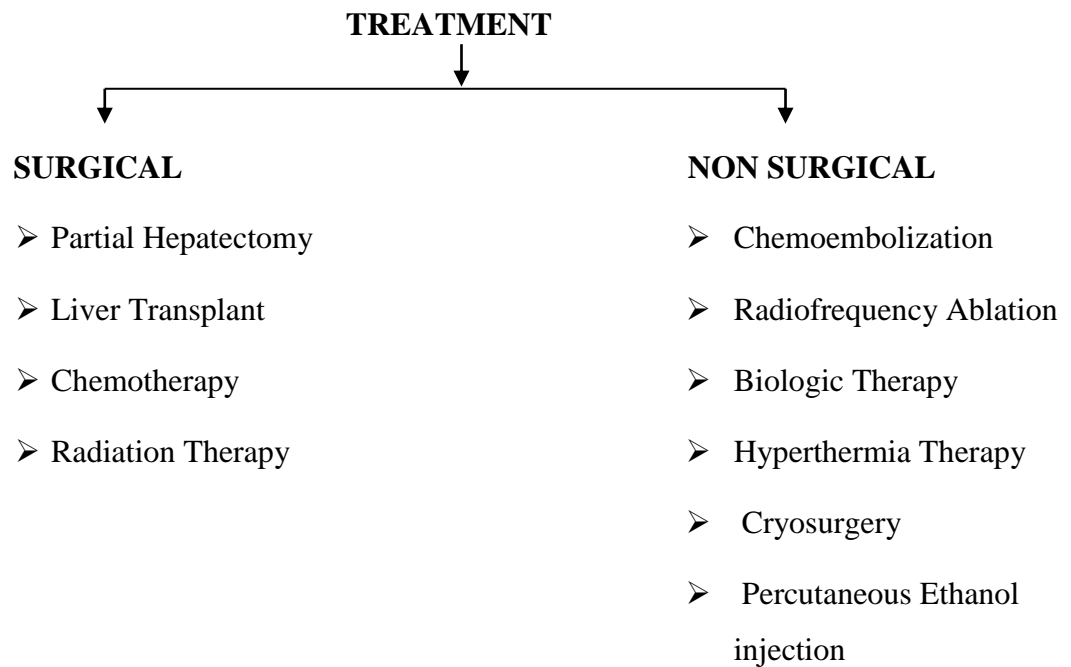


FIGURE 1.4: TREATMENT OF LIVER CANCER

1.10 INVESTIGATION OF LIVER CANCER

After being diagnosed with liver cancer, a patient will have many investigations like CT, MRI, and ultrasound. Biopsy is needed for the investigation of liver cancer.

FLOW DIAGRAM FOR LIVER CANCER DIAGNOSIS

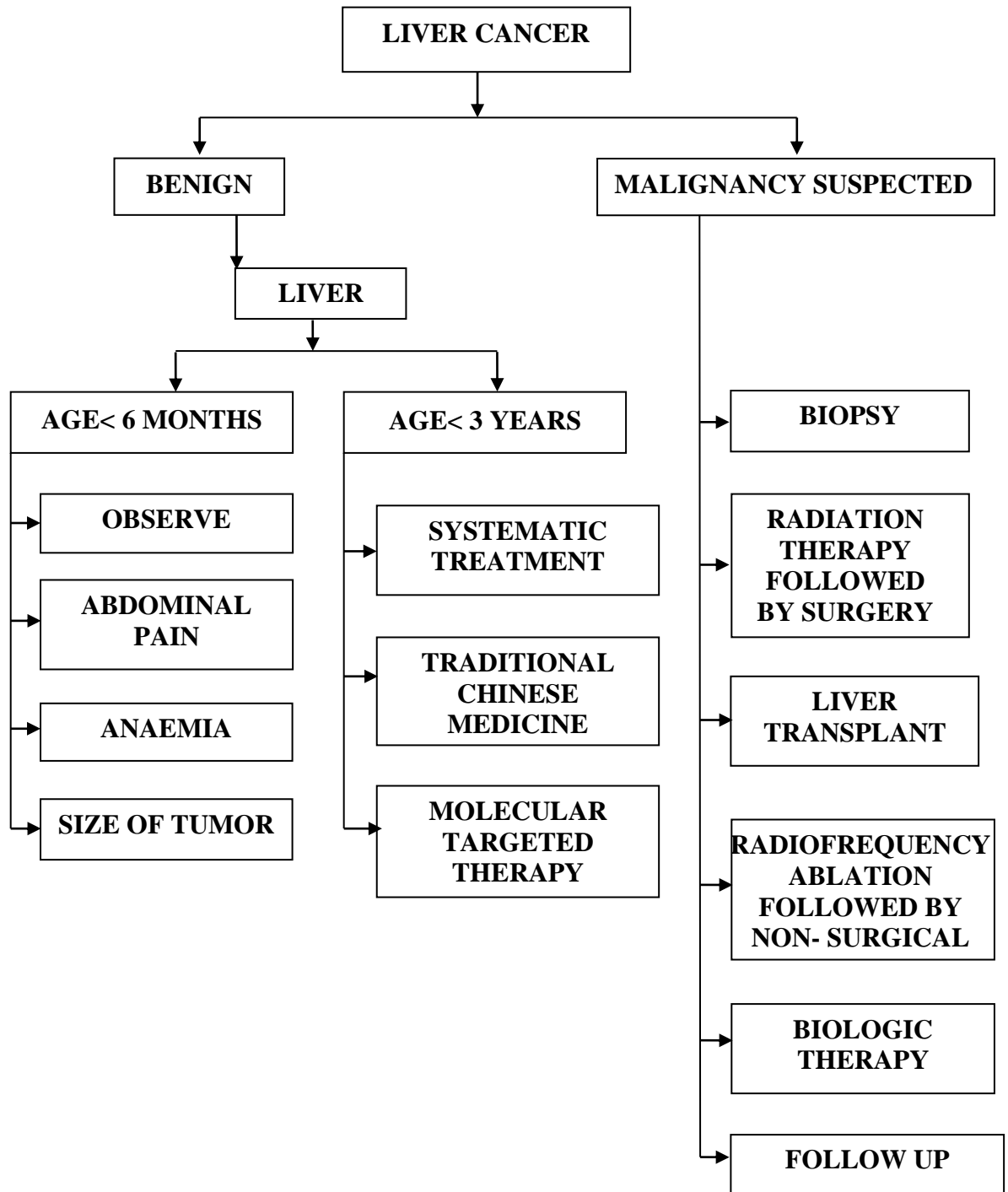


FIGURE1.5: FLOW DIAGRAM FOR LIVER CANCER DIAGNOSIS

1.11 PROGNOSIS OF LIVER CANCER

The prognosis of liver cancer include size of the liver, number of lesions, the health of the surrounding liver tissue and the general health of the patient. According to TNM there are several stages of liver cancer of liver cancer such as localized, regional and distant.

Stage	5-year relative survival rate
Localized	28%
Regional	7%
Distant	2%

TABLE 1: LIVER CANCER SURVIVABILITY RATE

1.12 TNM STAGES OF LIVER CANCER

The doctors use the TNM system to find where the liver tumor is located. The doctors determine three factors to find the liver tumor:

- ❖ Where the primary tumour is located and how large is it?
- ❖ Tumour spread to the lymph nodes or not.
- ❖ Cancer spreads to other parts of body or not.

There are different types of T groups consist of: Firstly, the primary tumour which cannot be assessed is known as TX. Secondly, when the primary tumour has no evidence is known as T0. Thirdly, when a single tumour has not grown into blood vessels is known as T1. Fourthly, when there is more than one tumour but no tumour is larger than 5 cm is known as T2. Fifthly, when there is more than one tumour with at least one tumour is larger than 5 cm is known as T3a. Where there is at least one tumour which is grown into a main branch of a large vein of the liver is known as T3b. When the tumour has grown into a nearby organ is known as T4.

There are different types of N groups consist of: Firstly, when all the nearby lymph nodes cannot be assessed is known as NX. Secondly, when the cancer has not spread to the nearby lymph nodes is known as N0. Thirdly, when the cancer has spread to the nearby lymph nodes is known as N1.

There are different types of M groups consist of: Firstly, when the cancer has not spread to distant nearby nodes. Secondly, when the cancer spreads to nearby nodes. Liver cancer mostly spreads to the lungs and bones.

The different types of stage groups include 1. When there is a single tumour that has not grown into blood vessels and the cancer has not spreads to the nearby lymph nodes are known as stage I (T1, N0, and M0). When there is a single tumour that has grown into blood vessels and the tumour has not spreads to the nearby lymph nodes known as stage II (T2, N0, and M0). When there is more than one tumour and its size is larger than 5 cm and the tumour has not spreads to the nearby lymph nodes is known as stage III a (T3a, N0, M0). When there is at least tumour which is grown into a branch of main vein of the liver and the tumour has not spreads to the nearby lymph nodes are known as stage III b (T3b, No, M0). When the tumour is grown into a nearby organ and the tutor has not spreads to the nearby lymph nodes are known as stage III c (T4, N0, and M0). When the tumour can be any size or any number and may have grown into a number of blood vessels and the tumour has spreads to the nearby lymph nodes. The cancer has not spreads to distant sites is known as stage IV a (Any T, N1, M0). When the cancer has spreads to other parts of the body is known as stage IV b (Any T, Any N, and M).

There are several other staging systems that have been developed such as Barcelona Clinic Liver Cancer (BCLC) system, the Cancer of the Liver Italian Program (CLIP) system, The Okuda system.

1.13 PROTEGE TOOLS

Protégé is open source ontology construction tool developed at Stanford University. Protégé allows user for creation, visualization and manipulation of ontology in representation format such as RDF, OWL etc. Protégé OWL is implemented in java and it runs on a broad range of hardware platforms and it is becoming the de-facto standard for auditing OWL editor. It provides a user friendly java based graphical user interface to create and edit ontology. It helps in solving multimedia queries, engines, inference and supports logical parameters.

There are different types of tools are:-

1. **Neon:-** Neon toolkit is based on eclipse platform and it provides complete support for ontology engineering life cycle. Neon provides multi-platform environment with additional features such as annotation on ontology matching, reasoning interface and

reuse. In this user interaction is easy to develop complex ontology and functionality can be enhanced with power plug-in.

2. **SWOOP:** - SWOOP is scalable OWL ontology browser and editor written in java. Apart from this strong reasoning support SWOOP provides creation, editing, comparison and merging of ontology. It also imports ontology from OWL, XML, RDF and text formats.
3. **Onto Analyzer:** - Onto Analyzer is a graphical tool which is used for representation of ontology in graphical form. It is also useful in creation, updating and representation of semantic ontology to show relationship between two classes in a hierarchical form.
4. **Onto Clean:** - Onto Clean tool has been implemented in java technology. It enables cleaning of class entities which is present in ontology to modify the existing ontology with less overhead. This tool also supports all platforms and enables flow control among different entities.
5. **Onto Lingua:** - onto Lingua was created by Knowledge systems laboratory at Stanford University. This was development tool which provides form based interface to user to build ontology. It also provides distributive and collaborative editing, browsing and editing ontology.
6. **Web Onto:-**Web Onto is a popular ontology editing tools which is developed by knowledge media institute of Open University. Its main features are Collaborative browsing, creation and management of new ontology using graphical user interface, automatic generation of instance editing forms and consistency checking.
7. **Oiled:** - It allows user to develop ontology in OIL and DAML + OIL languages , which is developed by Manchester University. It offers frame like paradigm for modelling the rich expressiveness of DAML + OIL ontology. It helps in creation, editing and maintenance of ontology.
8. **OntoSaurus:** - OntoSaurus developed by California University which comprises of Ontology editor and browser. It allows user to implement ontology in LOOM language. The browsing module generates HTML pages dynamically for existing ontology and allows user to search and browse ontology. It also provides HTML forms to edit ontology.

9. **Web ODE:** - It has been developed by ontological engineering group from University of Madrid. It provides support to most of the activities involved in ontology development process such as edition, navigation, documentation, merging and reasoning. It provides import and export services for ontology of XML format. It also provides translation services to other languages.
10. **Onto Edit:** - Onto Edit provides an ontology engineering environment for different platform and it is developed by Karlsruhe. It helps the user to maintain, create and edit their ontology very easily and in efficient manner. It also supports collaborative frame of maintaining ontology.
11. **Top Braid:** - Top Braid is helpful in building ontology for the semantic web applications. It provides support to develop and maintain ontology. It also includes a flexible and extensive framework with published API for developing semantics.
12. **Apollo:**-Apollo is user friendly knowledge modelling application. Apollo system is modelled according to OKBC protocol. Ontology can inherit their other ontology and use classes of inherited ontology as its own. It provides consistency check but not supports graphical view. In this information merging and extraction is difficult in Apollo due to the lack of collaboration.
13. **RDFedt:** - The RDFedt helps to build complex and structured RDF document. It helps to perform aggregation, notation, content, threading and set of label styles in XML, to generate code from tree RDF and finally to generate and print the tested code in RDF tree. It is not platform independent it works only on windows platform.
14. **KAON:**- KAON is open source ontology management system targeted for business applications. It allows ontology creation, management and provides a framework for building ontology based applications. It provides framework for navigating and searching ontology through web browsers. KAON provides two API for RDF and ontology language. It enables efficient reasoning and extends RDF with symmetric relations.
15. **ICOM:** - ICOM is an advanced case tool which allow user to design multiple extended entity relationship diagrams with inter and intra schema constraints. It is simple, freeware conceptual modelling tool that demonstrates use of novel and powerful knowledge representation based technologies for database and ontology

design. It employs SHIQ description logic to encode schemas and to express view and constraints.

16. **DOE:** - DOE allows users to build ontology according to taxonomies of concepts and relations. DOE not supported by many ontology constructional activities. It also provides interoperability without merging support and concepts cannot be defined implicitly with constraints.
17. **K-Infinity:**-It is tool developed by intelligent views for creation, maintenance and use of knowledge network. With the help of knowledge builder, engineers can create, edit and rename both objects and relations as well as to relate to each other according to defined relations. Relations can be expressed with help of concept editor and graph. It has K-Organization which provides administration, navigation search and query formulation to classify the distinct objects.
18. **Link Factory:** - Link factory Workbench is originally designed for very large Biomedical Ontology. It invokes client application on user side to manage Link Base ontology, which is an ontology base. It also acts as server interface, receiving and answering user requests, holding business logic. It provides merging of ontology which provides consistency checking with extraction of relevant information.

1.14 APPLICATIONS OF PROTEGE

Protégé is useful for building ontology in OWL and its applications are:-

- ❖ **GUI and API:** - Protégé OWL is built upon the protégé frame based knowledge model and use protégé GUI for editing classes, slots, individuals and instances. It provides an API that allows developers to integrate protégé OWL in their applications.
- ❖ **Graphical Editor for logical OWL expressions:**-It allows users to quickly resemble applications with the mouse or keyboard. It uses a graphical object oriented that display primitive and defined classes. In this the graphical editor supports copy/paste and drag/ drop.
- ❖ **Wizards to streamline Complex tasks:** -It supports Common Ontology like engineering patterns such as creating group of classes, making a set of classes disjoint and the creating, creating matrix properties in order to set many matrix properties values and value partitions.

- ❖ Direct access to reasoners:-The protégé OWL provides direct access to the high performance classifiers. The user interface supports three types of reasoning. First is consistency checking? Second is classification and third is instance classification.

The following features are also available in protégé OWL is:-

- ❖ Form Generation: - Protégé OWL can automatically generate a user interface to acquire instance data from class definitions; it is a useful feature that supports knowledge acquisition.
- ❖ Multi-User Support:-Protégé OWL provides multi- user support for synchronous knowledge entity.
- ❖ Support for multi-storage formats:- Protégé OWL can be extended with back ends for alternative file formats. Alternative file formats include clips, XML, RDF and OWL.

SPARQL is the standard language in which we used query graph data which is types of SPARQL queries are select, ask, construct and describe. SPARQL can also be used to query and update RDF data. SPARQL requires an interface which allows the user to enter the queries and to display the results. SPARQL retrieves and manipulates the data and answering the queries related to diseases or any other.

2.1 LITERATURE REVIEW

Sumblay.et.al (2014) in this paper the author discussed diagnosis of breast cancer using treatment. Breast cancer is a very common in real world. Breast Cancer is of two types: benign and malignant. It also generates a decision tree using J48 algorithm for classification of whether a patient had benign or malignant tumour. Decision tree is used for classification and prediction of breast cancer diagnosis. In this Wisconsin breast cancer datasets is used to differentiate the benign and malignant cancer. Benign cancer is a cancer which doesn't spread to other tissues. Malignant cancer is a cancer which is spread to other parts of the body. 80% cases are malignant cancer in the world. In one year 40,000 women die from breast cancer. Breast is mainly found in females. The breast anatomy of large breast includes ducts, lobules, nipple, fat, chest wall, dilated section of duct to hold milk, pectorals' major muscle. Breast cross include normal duct cells, ductal cancer cells, basement membrane and centre of duct. The paper works with different alternative approaches that include neural networks, digital mammography, Naive Bayes classifiers. Neural networks are network which input layer represents the dataset and output layer represent one node. Digital mammography is used to detect tumour. Naive Bayes perform analysis of survivability rate. It also represents a confusion matrix using the J48 algorithm in data mining.

Ali.et.al (2014) this paper introduced various classifier techniques for classification of liver cancer, and an efficient framework which is inspired by the human interpretation of US images is presented. The main objective of finding solutions to find accurate result using supervised and un-supervised machine learning algorithms in classification. The regions can be classified into two types such as benign findings and malignancy. In this paper disease progression is used to check the patient status. It helps in predicting the disease progression and according to this treatment can be suggested. Active contour model which is used for the segmentation of region of interest. To evaluate performance of various classifiers on LESH features, this is obtained from normalised craft and SVM classifier; we found that weka provides better performance as compared to other various classifiers. It also generate ROC curve such as true positive rate(TP) , false positive

rate(FP), precision, recall, F measure and Receiver operative curve (ROC) in numeric value . We can compare errors during classifiers and find out that SVM is the best and MLP is the second best. The future work is that disease is predicted by using US and magnetic resonance imaging.

Upadhyay&Wasson (2014) this paper mainly discussed about the various techniques for liver segmentation such as region growing, threshold based, level set method, statistical model, clustering algorithm, gray level methods, histogram based approach and genetic algorithm. But the author uses genetic algorithm for liver cancer detection. GA has several advantages as compared to other algorithms such as: Firstly, representation is in form of bit strings. Secondly selection is proportional. Thirdly, the main genetic operator is considered to be crossover. In this paper, first a liver image will be selected and then noise will be added to that image .After that image will be denoised. With the help of fuzzy coded binary map the region of interest will be selected out of denoised. The binary map values are selected and their inverse is mixed with denoised values. The results are segmented liver images when these values are given to an empty matrix. In this way using GA the cancer can be detected. With the help of X-ray, CT scan, images the doctor can detect the cancer but it is difficult to analyse it correctly due to the reasons of same intensity of images of other body parts. Liver segmentation is an important tool that is necessary for the diagnosis in computer aid. To analyse the patient dataset the doctor examines the physical exam .the liver size, volume, shape, structure of its vessels, tumour size and locations. With the help of liver segmentation we evaluate the pros and cons of liver transplantation and treatment method of liver tumour.

Kundra & Pandey (2014) in this paper they presented a J48 algorithm for the classification of diseases. The paper facilities with neurological and physiological disordering that occurs inflammation and disruptions in the brain that creates some problems which is based on two parameters: Psycho-physical and EEG signal characteristics. Psycho-physical is again divided into psychological, physical and cognitive. EEG is again divided into brain lobes and brain waves. These attributes are categorised into data type high, medium, low. J48 is an algorithm that can handle continuous and discrete attributes. It also generates pruned tree, confusion matrix, and decision tree and sensitivity rate.J48 can also handle the data with missing attribute values. It generates J48 rules and decision tree that helps us to diagnose the disease on the basis of taken symptoms or attributes of EEG based diseases such as dementia, ADHD,

mood disorder. There are various techniques that are used in EEG based diseases that includes regression, clustering, classification. It compares two operative characteristic of ROC curve. ROC is also used for diagnostic test evaluation. The sensitivity of all these diseases lies in the range of 70-100% and lies in the range of 94-100%.

Gangwar.et.al (2013) this paper introduced RBR (rule based reasoning), CBR (case based reasoning) and ANN (artificial neural networks) methods for diagnosis of neuropsychiatric disease in a heuristic manner. In heuristic method case is matched with cases being. CBR is used for relative and absolute diagnosis in neuropsychiatric disease. RBR is used for correlate the sign and symptoms of the disease hierarchal. ANN is used for matching and selection process in CBR. There are different types of models that are used in this disease: knowledge-based and rule-based. The author discussed the Schizophrenia disease based on two parameters such as: physio-psycho and signal. All the parameters are represented by binary numbers. Binary number consists of 0 and 1. Physio-psycho is again divided into muscular, cognitive and psychological. Signal is divided into EEG characteristics and image characteristics. Muscular physiological include motor activity and muscular, Cognitive include speech disability and forgetting memory, psychological include abnormal behaviour, delusion and social withdrawal. EEG characteristics include temporal and frontal, image characteristics include temporal, frontal, parietal, anterior cingulated cortex. In this ANN is implemented by using MATLAB .Neuropsychiatric its treatment depends on its accurate diagnosis.

Jensen.et.al (2013) this paper describes neurological disease ontology which is caused by genetic or environmental reasons. Genetic reasons may be due to complex of gene. Environmental reasons may be due to stress and family conditions. Neurological is concerned with diseases that are related to the nervous system. We can classify the neurological disease pathological or symptom logy. This ontology has four branches: types, causes, symptoms and treatment. In this ontology more than one drug treatment is possible for particular disease. This ontology using tool named OWL2 and protégé 4.x. This ontology is based on foundational model of anatomy. The neurological disease is developed as an extension OGMS (Open General Medical Science) that provides high level terms including disorders, disease. The main goal of neurological disease represents molecular, genetic and environmental origins and include signs, symptoms, diagnosis, treatment and testing methods. We present the types of neurological diseases in clinical literature and classify them within the ontology. In this ontology top down and bottom up

approaches are used. ND terms are annotated by using metadata such as label, definition source, and textual definition. The most initial application of neurological disease includes analysis of large data sets patient records for multiple sclerosis, Alzheimer's disease and stroke. Disease Ontology covers conditions that are deviated from a healthy status less frequently diagnostics criteria which is used for these conditions.

Bodenreider.et.al (2013) this paper mainly discussed about the signs and symptoms associated with the diseases. This ontology is based on unified medical language system (UMLS). These are generally biomedical ontology which describe domain of illness, disorders. This paper explains annotation of diseases in biomedical datasets. It is used for aggregation purposes as well as clinical decision support. Decision ontology is used for decision support as well as aggregation purposes. The main characteristics of disease ontology are phenotypes in which organisms' results from the genetic makeup of particular organisms. Both sign and symptoms and phenotypes are partial overlap. Biomedical ontology consist of biomedical entities and its relations with other biomedical entities. The main use of disease ontology supports the annotation of diseases in biomedical datasets, which include the knowledge bases curation, electronic health records of the clinical documentation. There are different types of ontology are discussed such as disease ontology, diagnostic and statistical manual of mental disorders, human phenotype ontology, international classification of diseases, international classification of primary care, infectious disease ontology, logical observation identifier names and codes ,MEDCIN, MedDRA, medical subject headings, NCI thesaurus, path base pathology ontology etc..

Satria.et.al (2012) this paper has been implemented using protégé 4.1and OWL DLS .Tropical diseases are infectious diseases which is found in tropical and sub tropical regions. The paper facilities with different types of tabs that are used to build the ontology such as: classes, data properties, individuals, annotations properties, data types and DL query. DL query syntax is used to describe classification and scoring process for inters-individual relationships. The tropical disease ontology covers the disease treatment, control, prevention to drive health policy strategy based on particular disease. Two types of restrictions are defined in this ontology: some and only. It uses ontology based approach to represent entities and events with their properties and relations in the form of knowledge representation. The developed ontology provides structured information of tropical disease which is useful for health care professionals. Now, it is an online system

the user accessed anywhere or anytime and support for better medical decision. IDO was also designed to make infectious diseases that is derived from different sources which are comparable .Visualization tool was used in tropical disease ontology by using ontology browser. Browser allows OWL and RDF files. In tropical disease environmental conditions represent climatic, social and economic factors. These both functions lie in Tomcat Apache 5.5 server environment. Therefore the author installed Tomcat Apache 5.5 server. In this medical ontology have been used in semantic based statistical analysis of medical data..

Moawad.et.al (2012) this paper introduced different viruses: hepatitis a, hepatitis b, hepatitis c and hepatitis d viruses which is found in both males and females. Hepatitis A is a infectious disease. Hepatitis B is a serious liver disease. Hepatitis is due to infection of the liver. Hepatitis D is also an infectious disease. This ontology has been implemented in protégé OWL tool. The ontology of biomedical reality is divided into two classes: biological continuant and biological Occur rent. Both are inherited from basic formal ontology. The biological continuant is divided into Organismal Continuant , extra Organismal biological continuant and the biological Occur rent is divided into Organismal Occur rent and Extra –Organismal biological Occur rent .The paper facilities with various methodologies that are used to build the viral hepatitis ontology such as: VH ontology extraction phase, validation phase and VH ontology OWL phase. After that it is integrated with OBR. It also represents OBR viral hepatitis disease ontology. Bottom up approach has been used in designing the ontology. Intelligent systems and Physicians can share and exploited this ontology in different ways. The result of viral hepatitis is shown in three classes such as symptoms, sign and lab result .Viral hepatitis represent the classes and relations classification tree in OWL and then it is merged with OBR in OWL. Then protégé –OWL is implemented this phase.

Herdiani.et.al (2012) introduced a new technique development of life cycle method for dengue hemorrhagic fever ontology. Dengue is an infectious disease and mosquito borne viral disease in the world. Entities of Infectious Disease Ontology (IDO) are relevant to biomedical and clinical aspects of all infectious disease .Development life cycle include specification, knowledge acquisition, implementation and evaluation. There are different types of schemas that are used to build the schema such as dengue diagnosis, dengue therapy, dengue prevention and control, endemic diseases: prevention and control, fluid therapy, diagnosis, differential, disease outbreaks: prevention and control, mosquito

control. This paper has been implemented in OWL tool. The dengue hemorrhagic ontology provides semantic web infrastructure for DHF research and treatment. The author uses interoperable platform which is used for sharing and reuse of knowledge related to dengue. The DDSS (Decision Dengue Support System) project is used for implementation of dengue and dengue vector control programs. The DHFO hierarchical conceptual schemas which is based on knowledge acquisition. DHFO integrates other DHF dimensions by using relevant ontology which includes IDO and then it is submitted to bio-ontology portal which can be assessed and evaluated by anyone.

Alfonse.et.al (2012) in this paper they presented TNM system (Tumour node metastasis). The paper facilities with various types that are used to build the liver cancer ontology such as: top level ontology, domain ontology, task ontology, application ontology. The components of ontology are classes, attributes and relations. Ontology are now used in e-health, medicine, bioinformatics and many information system enterprises. In this ontology different subtypes of primary liver cancer are explained in which 90% are hepatocellular carcinomas, 9% are cholangio carcinoma and 1% are angiosarcoma. Liver Cancer has three main classes: diseases, medical intervention, references. Disease contains the types .Medical intervention contains the staging, diagnosis and treatment classes. References contain the symptoms, stage, risk factors and TNM classes. It is web based ontology. It uses protégé OWL tool. It also provides a semantic representation of liver cancer over the web. The liver cancer consists of risk factors, symptoms, treatment, diagnosis and prevention.

Kalaivani.et.al (2011) this paper has been implemented using the tool named protégé version 4.0. We constructs poultry ontology in web and generates resource description framework schemas (RDF) and XML scripts. The author started the paper with semantic web technology and then classifies the poultry science with the help of protégé. It also defines the classes, properties, features. It also used reasoners to check semantic consistency. The main issue of reasoners is that OWL is not able to handle expressivity. The paper also shows graphical view of classes which is linked with existing web resources into the semantic web. Two types of properties are defined: object properties and data type properties .Object properties are properties in which there is a link between individual to individual. Data type properties are properties in which there is a link between individual and data type values.

Adams.et.al (2011) introduced Mendelian disease which regulate the immune system and develop maintenance. The main feature of PIDO is that the canonical phenotypes of PIDs i.e. sign and symptoms associated with the disease. The primary immunodeficiency disease ontology has been implemented using protégé OWL tool. It uses biomarkers for further classification of symptoms (Phenotypes). Biomarker is an imaging biomarker that is used for observing the projection, radiography and computed tomography scanning. It is based foundational model of anatomy. This ontology is based on the characteristics of quality, function, structure, process that are interoperable with other biomedical ontology. In this ontology we developed the Pathfinder which consists of a set of phenotypes with a suspected primary immunodeficiency disorder which are observed in a patient. PIDO uses the general formal ontology (GFO). GFO represent the processes, integration of objects and provide axioms in OWL version. It is used for the development of medical applications for the domain of primary immunodeficiency and immunology diseases.

Almeida.et.al (2011) in this paper they presented gene ontology, cell type ontology pertaining to the OBO foundry framework. OBO foundry framework includes OBO scope, protein ontology, cell type ontology, gene ontology, upper level ontology etc. There are three main axes which are used in this ontology that represent the knowledge organization, knowledge acquisition from experts and texts, visualization tool. The paper facilities with various BLO management that are used in blood ontology such as: quality management, blood utilization management, donor selection, blood collection, control of transfusions and blood testing. BLO helps in interoperable system which is used for educational purposes as a knowledge base. Blood banks and transfusion services helps in covering the issues related to the official documentation of interest. It provides the basic information on Haematological interest. The BLO core focuses on physiological aspects of blood. Abnormalities can develop in blood during illness, in diagnosis can detect them. This blood ontology uses protégé OWL tool and is based on the components of blood. BLO reuse the data available in other ontology, organization of hierarchy and relationships. The different purposes of BLO represent as a base for computational inference, knowledge base for educational purposes and as a tool which is used for diagnosis.

Mungall.et.al (2011) introduced a reasoned tool which is a kind of germ cell differentiation and is based on cl classification. In this OBO foundry helps in solving the problem of ontology maintenance and correctness. The evolution of gene ontology is

integrates with other OBO ontology. The gene ontology provides structured annotations for gene and gene products which is represented in terms of molecular function, biological processes and cellular component. This can also be used for making cross ontology queries which is derived from gene ontology. This paper uses relational graph and OWL tool. In this ontology cross products are used for visualization. The author also uses the internal cross products logical definitions for logical and probabilistic inference of annotations. The most common use of gene ontology is that class enrichment analysis in which gene set is statistically analyzed to find over represented classes. We can also use the cross product definitions of logical and probabilistic inference of annotations. In gene ontology reasoners can also be used for the development of ontology, the classes can automatically be placed in a hierarchy on the basis of their properties. The partitioned of gene ontology includes mutually exclusive cross product sets and candidate ontology include chemical entities, proteins. This helps in reusing, maintainability and evolution of gene ontology. The limitation of gene ontology is fixed structure of pre-composed classes.

Schneider & Brochhausen (2011) this paper mainly discussed about the methodology and principles of chronic ontology in the EU funded project. It describes the telemedical platform for monitoring the general health status of patients with chronic conditions. This paper is based on basic formal ontology which is widely used in biomedical domain in which higher classes are concerned. It has been implemented using the tool named OWL. Chronious ontology was built on top upper ontology. Two domain ontology are defined such as Chronic Obstructive Pulmonary Disease (COPD) and Chronic Kidney Disease (CKD). In Chronious ontology middle layer ontology for clinical care (MLOCC) include classes for objects such as cells, tissues, organs, chemical substances. In this BFO class process is sub-divided into MLOCC class processes which include intentional and natural processes. The intentional process relates to human and social activities and natural relates to organism and chemical process. In clinicians it also provides decision support for treating patients. This ontology also represents a thought of chronic disease and pathological conditions. The Chronious literature system is user friendly for health care professionals. The best option for implementing the Chronious ontology is realist ontology. Realist ontology are ontology that represent the reality independent of mental representation of end users.

Topalis.et.al (2011) in this paper they presented micro application which is drive a dedicated related database on insecticide resistance which was adopted by the world health organization. This paper is based on basic formal ontology and OBO foundry. The paper works with various aspects that are used in the vector borne diseases such as clinical aspects, therapy, physiological processes of the vectors, prevention and control. Vector- borne diseases represent the physiological processes of mosquitoes which are involved in disease transmission. To control the vector borne diseases, the author construct the decision support system (DSS) and worked out in the frame of IDO (Infectious disease ontology). Medical workers and health agencies can also used DSS in remote areas. This ontology has been modified using the IDOMAL. IDOMAL is a ontology which describes malaria .It is based on BFO and IDO reference ontology. IDOMAL and IDO have same meaning but its name should be different .IDOMAL will switch to IDO's ID number. The IDOMAL covers the immunology of malaria and immune states of the vertebrate hosts of the parasites. The novel approach of vector – borne diseases is information technologies as a complement of biomedical techniques which is based on the study of biology of vector diseases.

Chen.et.al (2011) the author discussed a survival prediction of liver cancer using the data mining tracking period, surgery, radiotherapy, chemotherapy which is related to liver cancer survival. With the help of ANN we predict the accurate result and assist patients to understand their survival and treatment of liver cancer. ANN is more accurate than CART for survival prediction. We can also generate decision tree and classification of liver cancer with the help of attributes.

Qiu.et.al (2011) introduced a new technique the fuzzy enhancement algorithm which was applied to computer aided diagnosis of liver cancer ultrasound images. The pathological measurement of liver cancer diseases includes hepatitis and liver cirrhosis. The three sets of ultrasonic liver images include normal liver, hepatoma and cirrhosis. In this fuzzy operator tool is used for enhancement. This paper uses three methodologies to find the accuracy of three classifiers: K-means, back propagation neural network and SVM. SVM finds the accurate result as compared to other methods. It also used features from first order statistic (FOS), gray- level occurrence matrix (GLCM), gray level dependent matrix (GLDM), gray level co-occurrence matrix (GLCM) and laws. The methods which are used for image enhancement are divided into two broad categories such as frequency-domain methods and spatial domain methods. Frequency domain method is based on

modifying the image and spatial domain method is the direct manipulation of pixels is adopted. Fuzzy enhancement improves the classification accuracy of three classifiers. The main use of fuzzy enhancement in the computer aided diagnosis of liver cancer is an efficient pre-processing procedure. The classification accuracy with fuzzy enhancement, liver cancer increases from 96.67% to 100%, haemangioma increases from 80% to 93.33%. The future work of liver tumour diagnosis includes more subject images could be used to evaluate the efficiency of the algorithm.

Massieh.et.al (2010) in this paper they discussed about how the liver structure is segmented through the original CT image. In this automatic hepatic lesion is the most important step for diagnosis and surgery planning of liver cancer. The paper presents the 3-D consistency performs which are based on 3-D dimensional information that a lesion mass does not appear in a single slice. Firstly, Liver can be segmented manually, semi-automatically and automatically from original CT image. Secondly, it enhanced the segmented slices as liver parenchyma and tumour tissue have grey levels. Thirdly, eliminate the segmented tumour. The proposed automatic method for liver tumour detection can be classified as contrast stretching, add slice to itself, Gaussian blurring, isodata threshold and morphological filter. The contrast enhancement is applied on the segments of liver tumour and adds each image itself to the white image with pepper noise and tumour as dark grey spots. After that Gaussian smoothing is applied and then used isodata threshold to turn the image into binary with tumour as black spots. The results are shown with the help of test reported on abdominal datasets. It uses five metrics to evaluate the quality of segmentation such as volumetric overlap, relative absolute volume difference, average symmetric absolute surface distance, and symmetric RMS surface distance, maximum symmetric absolute surface distance. These metrics can be applied to these five datasets to detect liver from CT scans.

Wennerberg.et.al (2009) in this paper they presented statistical approach for modularizing large biomedical ontology. All the modules can be identified in statistical analysis in domain corpus. The most important concept of statistical analysis is chi – square scores which is used for computed nouns and adjectives. In MEDICO we use case focuses on patients who is suffering from lymphoma in the neck area. Lymphoma is a cancer in lymphocytes and it is a systematic disease with sign and symptoms of multiple organs. The major use of MEDICO is to provide specific domain knowledge about lymphoma. On the basis of CT scan, X- ray and MRI its diagnosis and treatment should

be done. Biomedical ontology helps in identifying the most relevant ontology. It supports clinical applications such as medical image search. In biomedical ontology we observe that these ontology are complex and comprehensive. The level of detail and complexity of biomedical knowledge identify the modules. This paper uses unified medical language system and foundational model of anatomy (FMA). Foundational model of anatomy is a machine which processed resource on human anatomy. This machine is used for very large and complex models. One drawback of biomedical is that the ontology module can be large. It is difficult to identify the module. Biomedical knowledge can be used in text search and semantic medical image. The module should be identified in the form of sub-hierarchies that is accessed through the UMS tree browser.

Mizoguchi.et.al (2009) in this paper they discussed anatomical structure of organs and abnormal states in human body. Organ consists of internal organ, tissue portion and cells, body parts, and generic structure component. Internal organ consists of ordinary organs like tissue portion, heart and cells. Body parts consist of structural parts such as arms, legs, face. Generic structure consists of tabular structure such as blood vessels and oesophagus. According to the context of disease abnormal states are also pathological states. A new operator p was introduced in ontology building tool Hozo. Hozo is a mutual dependency between the whole and its parts. In order to make more articulate the clinical ontology introduced generic or disorder components. The author developed some ontology like SNOMED-CT, GALEN, ICD-10etc.The two types of advanced issues in clinical ontology is-a and part-of relations. No single hierarchy of diseases doesn't work well for all stakeholders so a new technique reorganization of is-a hierarch is introduced. This paper is based on EAV triple and P operator automatically generates a generic concept which represents all parts of thing. It enables semantic interoperability of medical information. In this we also use foundational model of anatomy which reduces virtual classes and ends up with partial solution of the problem.

Daramola & Fatumo (2009) the paper has been implemented as an OWL ontology using the protégé 3.4 tool .It is a semantic based statistical analysis for malaria research and treatment. Malaria is a vector borne diseases which is found in tropical and sub- tropical regions of the world that effect humans. Ontology is a domain which is used for sharing and reuse of knowledge. With the help of creating interoperable platforms, Ontology is also used for sharing and reuse of biomedical knowledge across heterogeneous platforms. For querying the business logic, the HMCO was implemented as an Enterprise java

beans. The OWL DL reasoners is used for semantic web reasoning on the classes and individuals in the HMCO. The paper facilities with seven dimensions that are used to build the ontology such as: malaria vectors, malaria types, malaria parasites, malaria symptoms, malaria treatment, malaria prevention and malaria control. This ontology also contains semantic description of features of a domain which is readable by both man and machine. To control human malaria initiatives HMCO provides interoperability support for the knowledge management. The main benefits of malaria are to provide an interoperable platform which is used for accessing malaria epidemiology information over the web. It also provides information support which is used for controlling malaria initiatives.

Denney.et.al (2009) in this paper they discussed a biomarker tool in translational medicine ontology. The paper facilities with different types of entities that includes disease, drug, patient, target, gene, risk, pathway, population, sign and symptoms, treatment. It is based on basic formal ontology. It provides interoperability between domain ontologies. Translational medicine draws on existing domain ontology in which a high level, patient centric ontology was developed. This ontology allows the integration of data through the drug development process. It facilities data integration from diverse areas such as clinical research, discovery research, hypothesis management. To develop a tailored drugs, first manufacturers need to identify biomarkers, that indicate how a given patient will respond to a particular treatment. In translational medicine a patient centric application ontology was developed between collaborative effort of industry and academic to highlight methodology and work to date. It also provides framework for the modelling of patient centric information which is essential for tailoring drugs. The translational medicine ontology requires traditional datasets from early drug discovery which is used for patients in clinical settings be integrated, queries and analysed data integration from diverse applications such as discovery research, track of doctors and track of medications.

Bonevich.et.al (2009) this paper mainly discussed about routine properties was collected like cancer tumour stage and mutation stage. In every relevant dataset controlled list of allowable terms was created and applied. Then we organize this vocabulary into a hierarchy called the compendia ontology. In this rule engine which was created to iteratively apply definitions to partition samples into distinct classes and then filter these sample groups, to reduce multi- variate effects. To analyze the metadata we summarize

the gene activity and meaningful Meta analysis is co mine. To explore gene expression results on co mine concepts edition was used. These results can be measured in affymetrix w133 platform which is related to clinical outcomes. This ontology uses tool OWL and it is a knowledge based framework ontology.

Hastings.et.al (2009) in this paper they introduced a new technique known as CheBI .CheBI is a description logistics which provides a candidate technology for automatic classification of entities. It is used for annotation of chemicals within biological datasets, text mining and data integration. As the complexity and size of ontology increases the efficiency of a available reasoning technology will need to be assessed. Molecular and chemical entities are also represented in this ontology.

Mokkarala.et.al (2008) this paper introduced MI (Myo Car dial) Infarction which was the wrong dosage of medicine due to verbal direction given by ICU cello. The error persistent next day through team rounds. This ontology has been implemented by using the protégé OWL tool in which we represent the classes and its relationships in a reference ontology and it is integration based development. There are different approaches that are used to build the ontology such as simple ontology approach, multiple ontology approach and hybrid approach. But the most important approach that we are used in medical error ontology is hybrid approach. The ontology development process is divided into seven processes such as identify source ontology, align source ontology, merge source ontology, validate and refine reference ontology, map source ontology, validate medical error ontology and build knowledge base. It also represents the simple queries on this ontology. In this languages level heterogeneity was introduced which means avoiding translating in different languages. An error occurs while searching literature and identification of candidate ontology. Another error occurs in the ICU or the operating room. One approach of medical error is to solve a problem. It is useful in error reporting systems. The main goal of medical error is for preventing the adverse events in health sector. The medical error ontology presents the scope and requirements for the ontology, the development approach, the steps used in the development process, and the plans for future work. This paper is based on integration based development that involves reuse of knowledge from existing taxonomies by merging individual taxonomies to create reference ontology.

Wang.et al (2005) this paper discussed about protégé OWL plug in which we designed an architecture of World Wide Web especially for the semantic web. The Semantic Web

allowing data to be shared and then reuse among applications, enterprises. Protégé OWL. Reference to a domain expert in medicine Blood pressure ontology is used to reveal the body functions and relationships between entities in which blood pressure is semantic web. OWL is a network of entities and their relationships .In this OWL builds on RDF and RDF schema and adds more vocabulary to describe properties and classes among relations. Ontology are presented in a logic base language; they represent sound and meaningful relationships for class, functions, properties and relations. Ontology are easy to understand. This ontology takes sufficient time, energy savings and it is a user friendly interface which makes editing less confusing. The OWL is used for specific user communities. The Blood ontology is based on partition.

Jovic.et.al (2004) the author discussed that heart failure ontology is based on knowledge. The main application of heart failure is intelligent decision support system (DSS). DSS is a complement expert system for patients related to warnings, suggestions Building knowledge based systems in the heart failure domain as well as unambiguous communication between professionals. This ontology is very precise, voluminous, portable representation of heart failure domain. Heart failure ontology is reusable in various patient data transformations in diverse applications. It also provides interoperability among the applications which enables easier verification and comparison. This ontology has been implemented using the protégé OWL tool. OWL has been developed for web applications. OWL supports objects and data type properties and it integrates with Semantic Web Rule Language (SWRL).It includes heart failure concepts, properties, diagnosis and treatment. The main goal of heart failure ontology is that it is represented in both computer and human readable. The main result of the development of heart failure ontology is EU Heartfaid project. HF ontology consists of congestive heart failure, medical synonyms and types of classification. The most motivating factor of heart failure is knowledge representation.

3.1 PROBLEM FORMULATION

There are many patients who are suffering from liver cancer in the world. Liver cancer can be found after symptoms appear, in early stages liver cancer has usually no symptoms. But in later stages the patient suffers from many liver problems like jaundice, fatigue, enlarged liver, weight loss, chronic cough, joint pain, pain in the right upper abdomen, vomiting, malnutrition. Jaundice is the main problem in liver cancer. The patient diagnosed the liver cancer disease in its early stage with non-surgical treatment like molecular targeted therapy, traditional Chinese medicine but in later stages they faced any problems in later stages like surgery, liver transplant, chemotherapy, radiation therapy and Partial Hepatectomy .The automatic classification of liver cancer is an important real world medical problem.

3.2 OBJECTIVES

The main objective of liver cancer is to develop the liver cancer ontology using ontology based tools. We used random tree algorithm for the classification of liver based diseases such as: fatty liver, cholestatic, autoimmune, inherited, Wilson diseases with the help of weka tool in data mining. Decision tree is used for diagnosis of liver cancer by using the attributes of liver cancer in data mining. We also represent the rules generated by the random tree algorithm .The attributes neurological, psychiatric, pathological, physical, cognitive are categorised into data types: high, medium, low, and medium/low. On the basis of attributes high, medium, low, medium/low we can conclude whether the disease is fatty liver, Wilson , autoimmune, inherited, cholestatic . On the basis of analysis derived from weka tool we also represent the confusion matrix, TPR and FPR values which is based on true positive rate and false positive rate. We also generate the ROC curve with the help of weka tool in data mining. We can also develop the liver cancer ontology through protégé tool which is a knowledge representation .Ontology consist of classes, individuals, data type properties, reasoners, Sparql queries. Liver cancer ontology consists of risk factors, symptoms, diagnosis, treatment and prevention. We can also used onto graph to represent the graph of liver cancer ontology. In liver domain Sparql queries retrieves and manipulates query data based on querying of various

diseases. Sparql queries are used for diagnose the liver cancer diseases such as fatty liver, Wilson, cholestatic, Autoimmune and inherited diseases. Query retrieval related to risk factors, symptoms, diagnosis, and treatment of liver cancer ontology by using the protégé tool. Protégé is a knowledge representation tool. They government has taken steps for the poor people who do not have enough money to diagnose our cancer. They have taken Rs.1,50,000 to those needy people who have cancer problem.

3.3 METHODOLOGY

We used Weka Protégé in development of liver cancer ontology using ontology based tools.

3.3.1Weka is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. Weka is a open source software which is issued under GNU (General Public License). Weka is also a bird which is found only in the islands of New Zealand. It implements algorithms for data pre-processing, classification, regression. Visualization tool was also used in weka tool. The weka GUI chooser has four applications such as explorer, experimental, knowledge flow and simple CLI. Weka contains nominal, numeric, string and relational attributes. In weka we can import data node various formats like CSV, C4.5, binary and ARFF. In weka we can apply algorithms directly or through java code.

How to launch a weka explorer

Click on windows 'start'-> 'taskbar'->'weka 3.4

In data mining we use weka tool to generate a decision tree with the help of classification. Classification predicts the nominal or numeric quantities. Classification include decision trees, instances and logistic regression.

- ❖ Click on windows->taskbar->weka 3.4.
- ❖ Click on choose button in the classifier box.
- ❖ Then click on the weka tool.

3.3.2 Protégé is a free, knowledge representation system and open source ontology editor. It is an extensible, platform-independent environment for creating and editing ontologies.

Protégé is easy to create, upload, modify and share ontologies for viewing and editing ontologies. The different formats available for protégé are turtle, OWL/XML, OBO, and

RDF/XML. The protégé includes object properties, data type properties and annotation properties. Object properties are properties in which an individual linked to an individual. Data type properties are properties in which an individual linked to an XML. Annotation properties are properties in which we add information to classes, individuals, object properties and data type properties. Weka is also used for loading the text data. The main features of protégé are: Firstly, it is easy to use .secondly, its small user interface new applications are linked with a knowledge base a protégé tab .Thirdly, protégé database is used for caching free memory and back end load frames are on demand only. Fourthly, we can extend the protégé with plug-in tailored for your domain and tasks. Fifthly, it is easily used by people who are not data mining specialists.

In weka decision tree is a machine learning method used for classification. Decision tree consist of decision nodes and prediction nodes. Decision nodes are nodes which specify the predicate condition in weka tool. Prediction nodes are nodes which contain a single number. In weka confusion matrix represents the column and rows. Each column represents the instances in a predicted class in a matrix. Each row represents the instances in a actual class in a matrix. Confusion matrix is also known as contingency table or matrix table. Confusion matrix specifies the classes of obtained results in data mining. In weka ROC is used as a measure of quality of a probabilistic classifier. The ROC curve is obtained from a given dataset set of points. ROC curve corresponds to a binary classifier. In weka we can calculate classification accuracy and other quality measures from the ROC curve. The main aspect of ROC performance analysis is the ability to assign weights to positive and negative errors in weka tool. The ROC curve is used to evaluate the quality of classification models. In weka there are true positive which means test outcome is positive and the condition is positive , false positive means test outcome is positive and the condition is negative, true negative means test outcome is negative and the condition is negative, false negative means test outcome is negative and the condition is positive values. In TPR and FPR values 1 stands for true and 0 stands for false. The F measure is a combined measure of recall and precision. The precision is the matrix in which diagonal elements divided by the sum over the relevant column. The recall is the ratio of relevant documents found in the search related to the total of all relevant documents. Both these measures are useful for comparing classifiers with the help of weka tool.

The main applications of protégé API are tab and standalone. The pros of protégé tab application are that it is simple, iteration, reuse of protégé GUI components. The cons of

protégé tab application are that protégé must be installed. It stucks with protégé toolbar and menus. The pros of standalone application are that it is no need to install protégé. A user doesn't need to know anything about protégé. Some tab will work. The cons of standalone application are that iteration is more difficult than in tab. The main description of standalone application is it can be embedded as a tab.

4.1 TOOLS

Weka is a data mining tool which is used for classification. We use random tree algorithm for the classification of liver based diseases such as: fatty liver, cholestatic, autoimmune, inherited, Wilson diseases. Random tree is used for regression and classification problems. Random tree is used for constructing a tree in which we considers k randomly chosen attribute at each node. Decision tree is used for diagnosis the liver based diseases. The algorithm for conducting the procedure of liver cancer diseases is as follows:

ALGORITHM: LIVER DIAGNOSIS

INPUT: Civil hospital data set pre-processed to satisfy prerequisites of the data mining technique.

OUTPUT: Random decision tree with cancer either fatty liver, cholestatic, inherited, autoimmune or Wilson.

PROCEDURE:

1. Acquire dataset from civil hospital cancer datasets from UCI machine learning repository.
2. Pre-process data for applying random decision tree data mining technique.
3. Pre-processed dataset uploaded in WEKA toolkit for analysis.
4. Random tree algorithm applied in weka in liver cancer dataset.
5. Decision tree random tree algorithm implemented, generating a tree with fatty liver, cholestatic, inherited, autoimmune, Wilson.
6. Diagnosis of new patients is done by classification of liver cancer diseases in weka.

From detailed study of liver cancer it has been observed that according to doctors of Fortis hospital every year due to liver disease Over 2 lakh people lose their lives in India and close to 20,000 people need liver transplantation to survive.

The percentage usage of different liver diseases are autoimmune liver disease is of 25%, Cholestatic liver disease is of 13%, fatty liver disease is of 22%, Wilson disease is of

33%, inherited liver disease is of 7%. Fatty liver include symptoms such as pain and swelling in abdomen, decreased appetite, nausea, fatigue, vomiting, dry mouth & increased thirst, yellow colour in the skin or mucus membrane or eyes, small, red spider-like veins on skin, itching and redness on feet or hands. Autoimmune liver disease is four times more common in women than in men. Cholestatic liver disease is of two types: primary sclerosing cholangitis (PSC) and primary Biliary Cirrhosis (PBC). Females mostly suffer from PBC and males more effects from male population. Wilson's disease is a long term liver disorder resulting in excess copper deposition in liver, brain, kidney and cornea of eye. It strikes in between the ages of 5-40 years.

LIVER CANCER

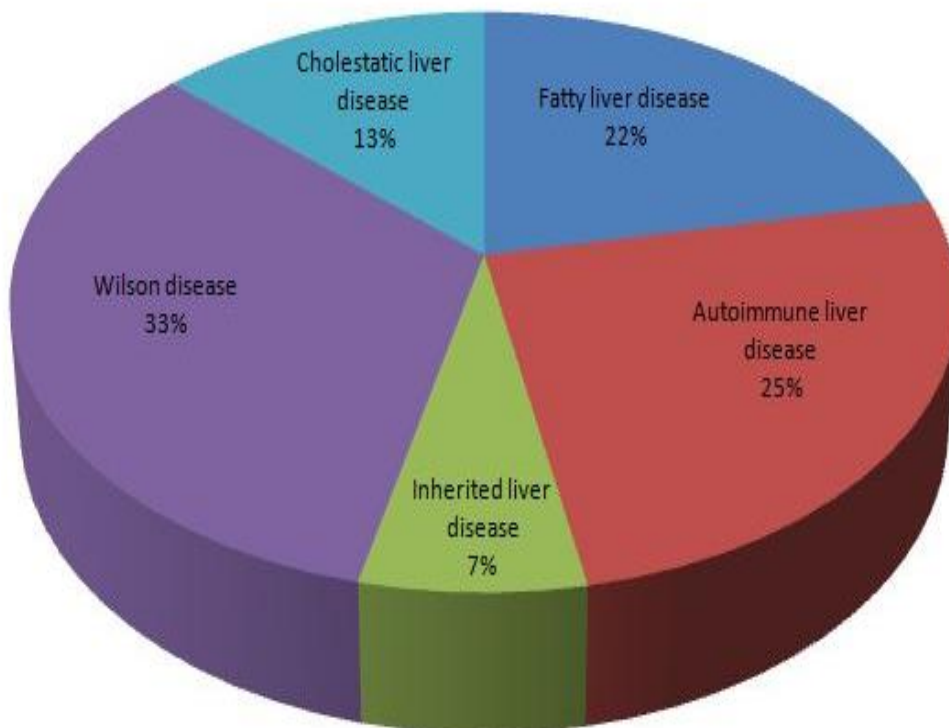


FIGURE 4.1: LIVER CANCER DISEASES

4.2 PROBLEM IMPLEMENTATION

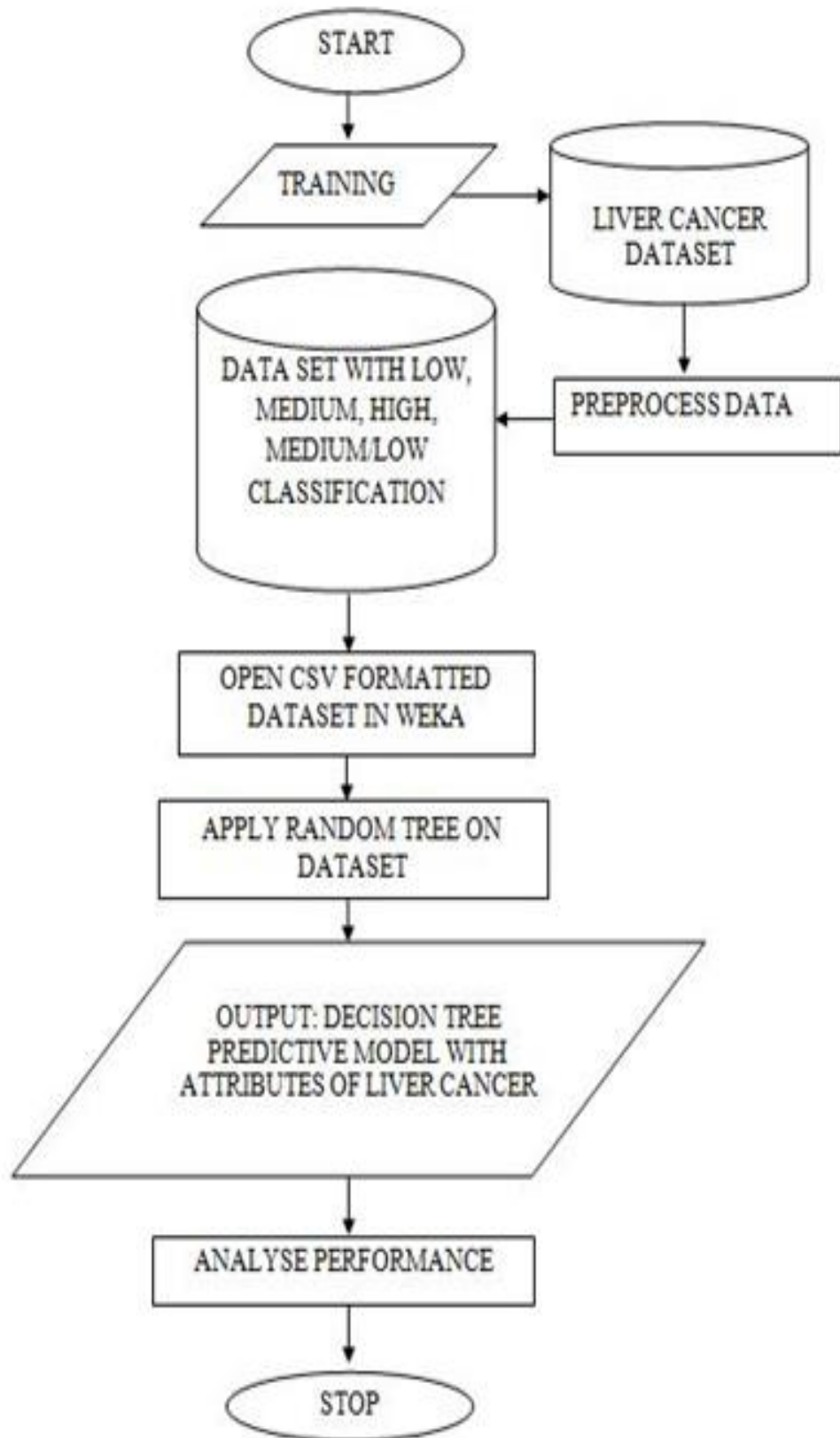


FIGURE 4.2: FLOW DIAGRAM FOR LIVER BASED DISEASES

The liver cancer is divided into two parameters: Neurological- Psychiatric and Pathophysical – cognitive. Neurological- Psychiatric is divided into neurology and Psychiatric and Pathophysical- cognitive is divided into pathological, physical and cognitive.

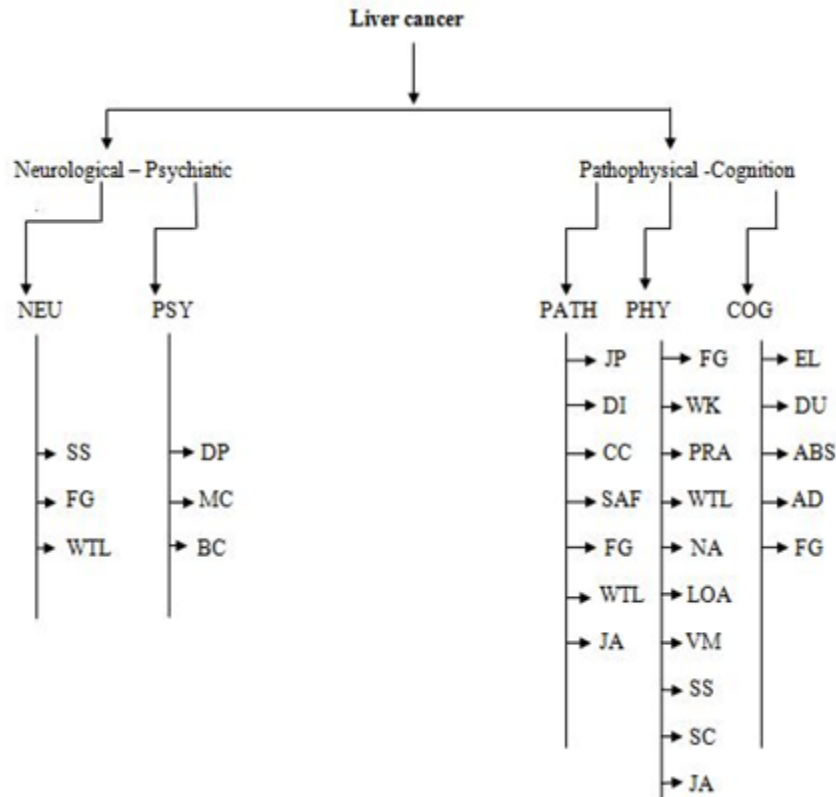


FIGURE 4.3: HIERARCHIAL CORRELATION OF SIGN AND SYMPTOMS OF LIVER CANCER

As shown in table 4, it describes diseases with their two important parameters. Pathological-Physical cognition and Neurological Psychiatric. Pathological-physical cognition parameter are further divided into three parts 1) pathological (PATH) parameter are joint pain (JP), diabetes (DI), chronic cough (CC), swelling of the ankles and feet (SAF). 2) The physical (PHY) parameter are fatigue (FG), weakness (WK), pain in the right abdomen (PRA), weight loss (WTL), nausea (NA), loss of appetite (LOA), vomiting (VM), jaundice (JD).3) The cognition parameters are enlarged liver (EL), dark urine (DU), abnormal blood vessels on skin (ABVS), abdominal distention(AD). Neurological-Psychiatric characteristics is divided into two parts: 1) Neurological is observed in different ways such as slurred speech(SS), failing voice(FV).2) Psychiatric is

observed in Behavioural changes(BC), memory loss(MC), depression(DP). In Table1 row represents the disease and column represents their respective parameters. The sub columns of neurology, psychiatric, pathology, physical, cognitive parameter contains “Y” if the symptom is present in the disease shown in respective row. For example, liver cancer has neurology symptoms such as: failing voice. Therefore these columns contain “Y” as in Table 4 . Similarly, the pain in the right abdomen and loss of appetite sub-column of physical contains “Y”. Finally the table is formed with five rows corresponding to five diseases and 24 columns corresponding to the parameters of the diseases.

TABLE 1. Problem Description Table (PDT)

DISEASES	NEUROLOGICAL		PATHOPHYSICAL-COGNITIVE																				
	PSYCHIATIC																						
	NEUROLOGY (NEU)		PSYCHIATIC (PSY)			PATHOLOGICAL (PATH)					PHYSICAL (PHY)								COGNITIVE (COG)				
	SS	FV	DP	BC	MC	JP	DI	CC	SAF	FG	WK	PRA	WL	NA	LOA	VM	JA	SS	SC	EL	DU	AVS	AD
FLD	N	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
WILSON	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	N	N	N	N	N	Y	N	N	N	N	N	N
CHOLESTATIC	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	N	Y	N	N	N	N	N	N
INHERITED	N	N	N	N	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	N
ALD	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	Y	N	N	Y	Y	Y	Y

TABLE 4.1: PDT

First, we enter data into excel and save it with .csv extension. We can enter the data with missing values also in random algorithm.

	A	B	C	D	E	F	G
1	PATH	NEU	PHY	COG	PSY	PM	
2		HIGH			HIGH	WD	
3		HIGH			MEDIUM/LOW	NOD	
4		LOW	HIGH			FLD	
5		LOW	MEDIUM/LOW			CLD	
6		MEDIUM		HIGH		AID	
7		MEDIUM		MEDIUM/LOW		NOD	
8	HIGH	MEDIUM/LOW				IHD	
9	MEDIUM/LOW	MEDIUM/LOW				NOD	
10							
11							

FIGURE 4.4: DATASET OF LIVER CANCER

After that open weka and then click on pre-process and then open your file

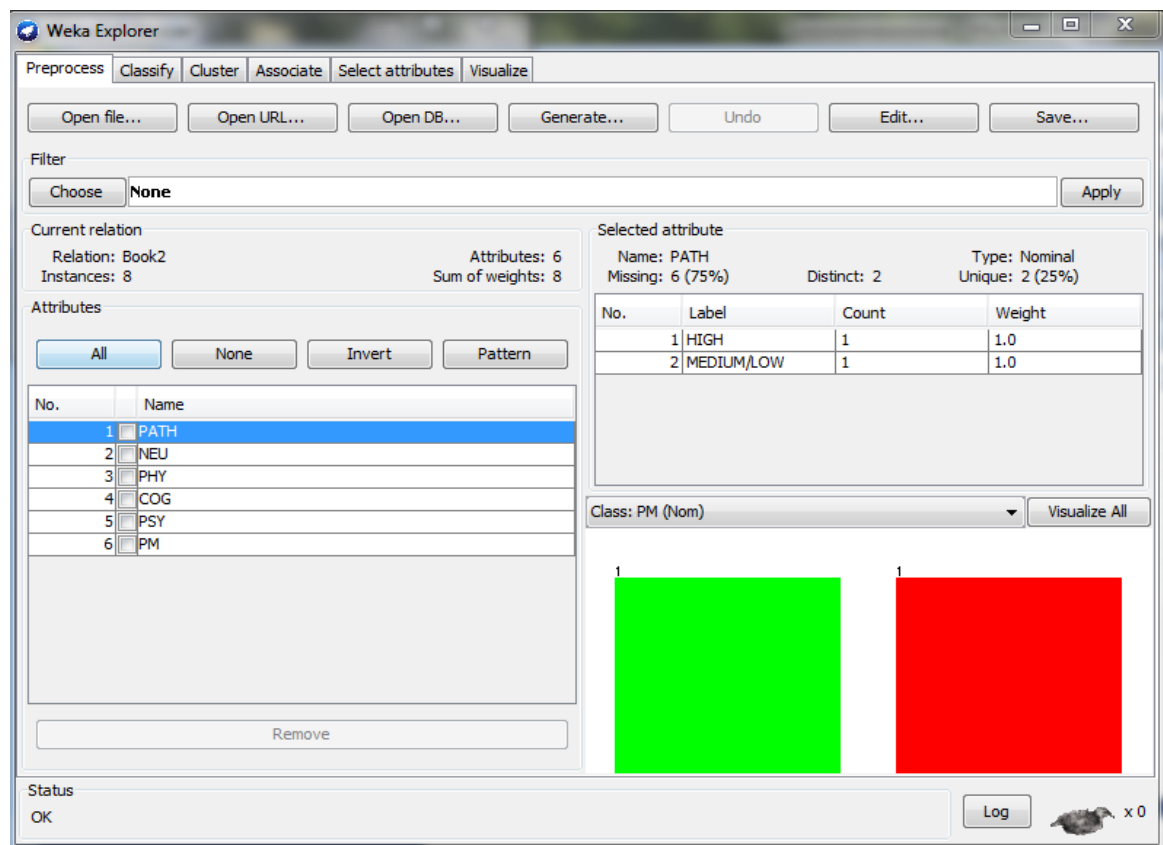


FIGURE 4.5 PREPROCESS OF LIVER CANCER

The liver cancer diseases are classified using random tree algorithm using weka tool in data mining. We categorised the different attributes Neurological, pathological, Physical, Cognitive, Psychiatric into data type high, medium, low, medium/low. The liver cancer data is collected from Civil Hospital Nawashar, Dayan and medical Hospital Ludhiana and the Max hospital Mohall. The dataset consist of 12 patients and 5 attributes. Detailed description of data set is given below:

ATTRIBUTES	DATA TYPE
Neurological	Categorical(high ,medium, low, medium/low)
Psychiatric	Categorical(high ,medium, low, medium/low)
Pathological	Categorical(high ,medium, low, medium/low)
Physical	Categorical(high ,medium, low, medium/low)
Cognitive	Categorical(high ,medium, low, medium/low)

TABLE 4.2: DETAILED DESCRIPTION OF DATA SET

In data mining weka tool first of all open the weka tool. After that click on pre-process and then open file. You open the files which you save in excel with .csv format. Then click on classify choose random tree and select cross- validation and then click on start button. Then Random tree generates rules is shown.

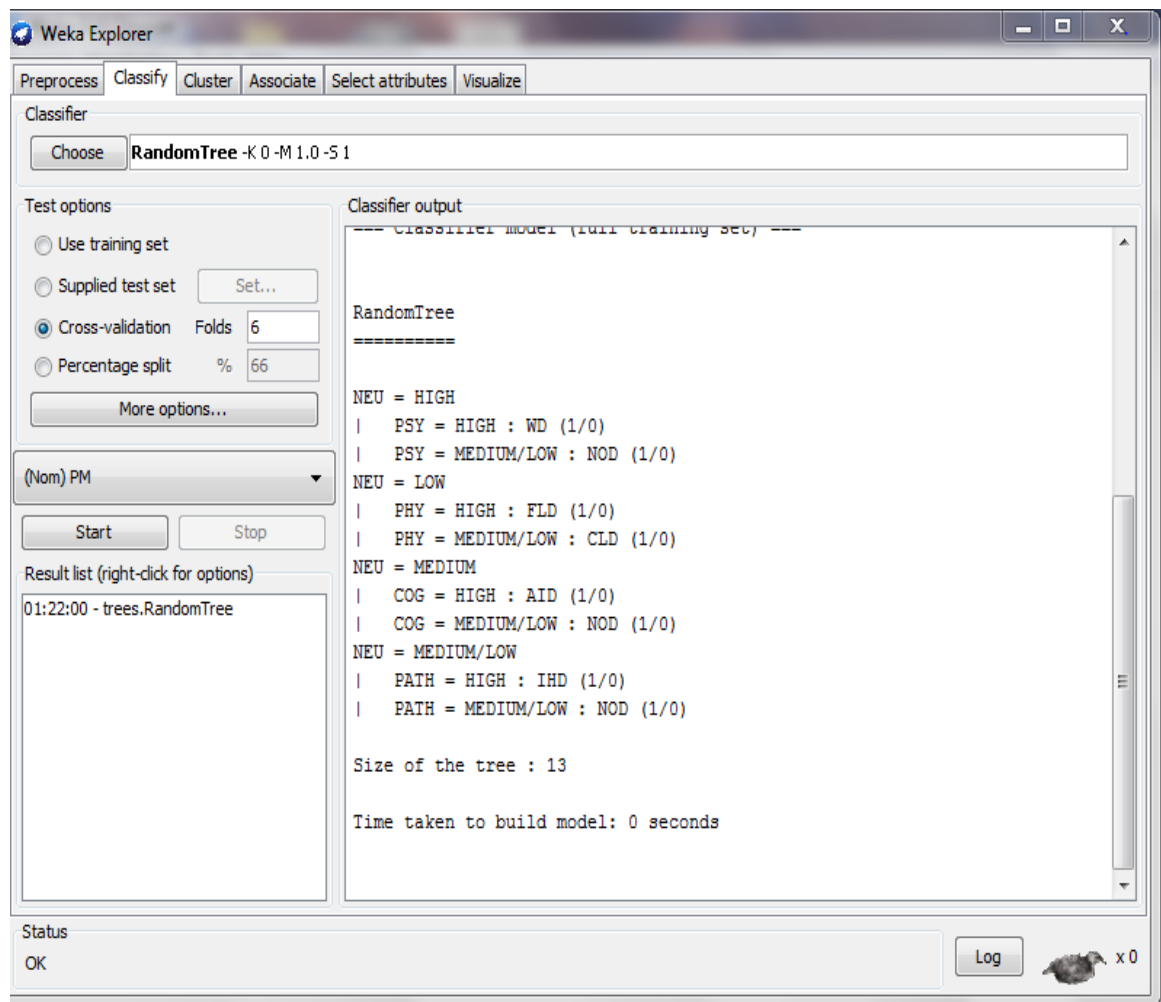


FIGURE 4.6: RANDOM TREE GENERATES RULES

Rules generated by random tree

Rule 1: If neurological (NEU) symptoms and Psychiatric (PSY) symptoms are high, then disease is “Wilson”.

Rule 2: If neurological (NEU) symptoms are high and Psychiatric (PSY) symptoms are medium/low, then disease is “NO disease”.

Rule 3: If neurological (NEU) symptoms are low and physical symptoms are high, then disease is “Fatty Liver”.

Rule 4: If neurological (NEU) symptoms are low and physical (PHY) symptoms are medium/low, then disease is cholestatic.

Rule 5: If neurological (NEU) symptoms are medium and cognitive (COG) symptoms are high, then disease is autoimmune.

Rule 6: If neurological (NEU) symptoms are medium and cognitive (COG) symptoms are medium/low, then disease is “No disease”.

Rule 7: If neurological (NEU) symptoms are medium/low and pathological (PATH) symptoms are high, then disease is “Inherited”.

Rule 8: If neurological (NEU) symptoms are medium/low and cognitive (COG) symptoms are medium/low, then disease is “No disease”.

Following are the points in order to obtain more usual representation of the same random tree.

- ❖ Click the right mouse button on the model type trees.
- ❖ Random trees in the result list frame and select the item visualize tree.
- ❖ Resize a new window with graphical representation of the tree.
- ❖ After that click with the right mouse button to the space in this screen and in the pop up menu and then click on the fit to screen.

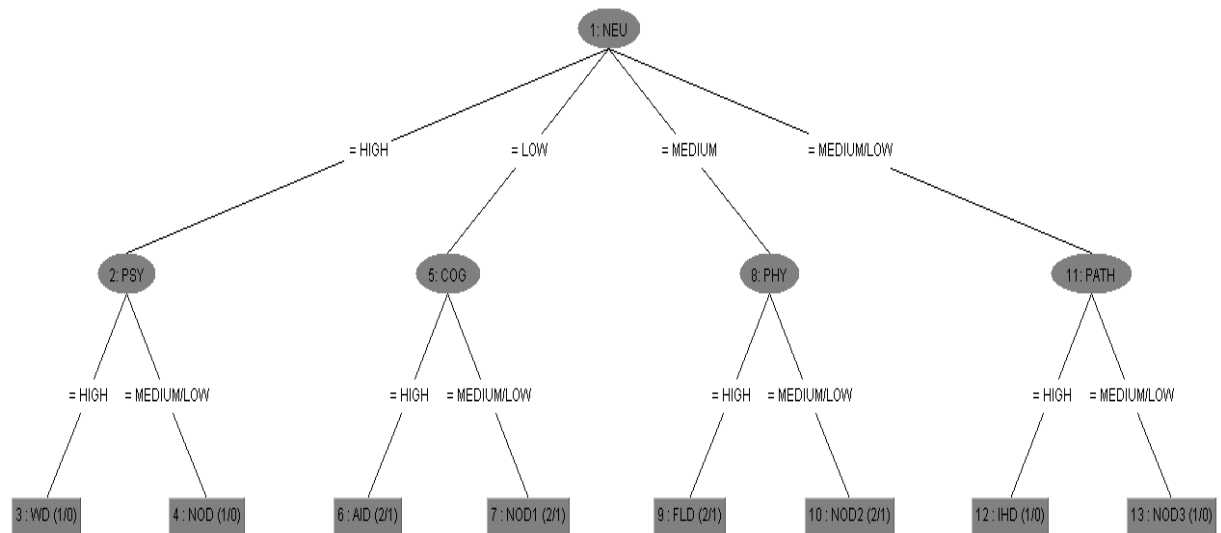


FIGURE 4.7: RANDOM TREE GENERATED BY WEKA

Click on pre-process in weka tool. Open yourfile which you saved in excel with .csv format. Then click on classify and choose random tree and cross validation folds and then click on start . Random tree is applied on the dataset in wekatool and a confusion matrix is generated.

```

=== Confusion Matrix ===

  a b c d e f  <-- classified as
  0 1 0 0 0 0 | a = WD
  1 0 0 0 1 1 | b = NOD
  0 0 0 1 0 0 | c = FLD
  0 0 1 0 0 0 | d = CLD
  0 1 0 0 0 0 | e = AID
  0 1 0 0 0 0 | f = IHD

```

FIGURE 4.8: CONFUSION MATRIX

From the above confusion matrix, true positive value for class a = "WD" is 1 while false positive values are 0,0,0,0. Whereas for class b = "NOD", true positive is 3 while false positive values are 0, 0, 0, 0. For class c = "FLD", true positive is 1 while false positive values are 0, 0, 0, 0. For class d = "CLD" true positive is 1 while false positive values are 0, 0, 0, 0. For class e = "AID" true positive is 1 while false positive values are 0, 0, 0, 0. In this diagonal elements of the confusion matrix represents the positive (TP) values and rest of the values represents the false (FP) values. Different operative characteristics are defined below:

True positive (TP) = When test outcome is positive and condition is positive.

False positive (FP)= When test outcome is positive and condition is negative.

True negative (TN) = When test outcome is negative and condition is negative.

False negative (FN) = When test outcome is negative and condition is positive.

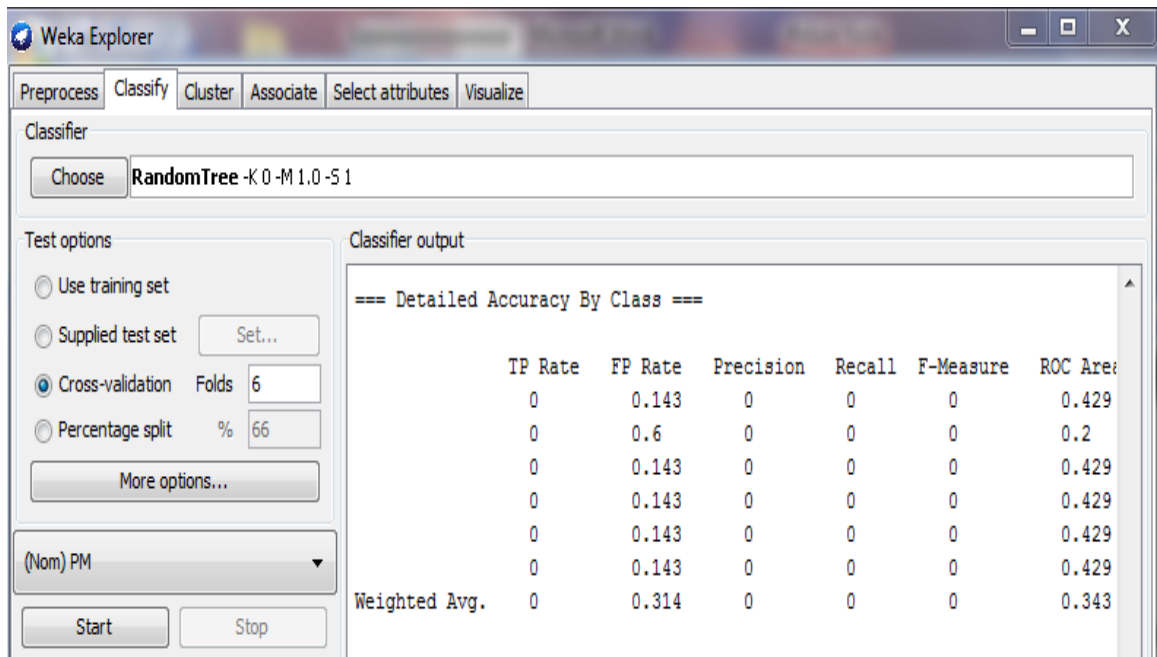


Figure 4.9: TPR AND FPR VALUES

The ROC curve obtained from random tree algorithm. It compares two operating characteristics that are true positive rate and false positive rate. Receiver operating characteristics curve (ROC) and area under curve (AUC) are calculated by threshold curve class in weka.



FIGURE 4.10: ROC CURVE

We develop the liver cancer ontology through protégé which is a knowledge representation using ontology based tools. Protégé consists of re:-classes, subclasses, data type properties, individuals, annotation properties, reasoners, Sparql queries. We also

used onto graph to represent a tree with the help of protégé. The basic blocks of OWL ontology .

- ❖ **Classes:-**Classes allow to group objects that have something in common. Classes are the basic structure of every ontology. Class hierarchy provides three buttons to manipulate the classes: First button allows users to create a subclass and second button creates a sibling and third button is used for delete a class in protégé. All the three buttons work with respect to selected class in the hierarchy tree in the protégé tool.
- ❖ **Individuals:-**OWL has its own name for members of a class are called individuals. Individuals contained in a class form a class extension .class extension does not restrict any other individual.
- ❖ **Properties:-** Properties express the relationship in a ontology. Two types are properties are defined: object properties and data type properties. Data type properties are properties that do not link two individuals. Object properties are properties that can defined their domain and ranges. Ranges of data properties are called data types. There are two types of object properties are top object property and bottom object property. There are two types of data properties are top data property and bottom data property.

There are three categories of tools that help in OWL ontologies:-

- ❖ **Editors:** - It is used to design classes, properties and individuals.
- ❖ **Reasoners:** - A reasoners is ready to check whether the ontology was designed as expected and its assertions are consistent. Reasoners can be selected and are started with appropriate menu that are located in the reasoners menu in the protégé tool.
- ❖ **Visualization:-**Visualization can be used to obtain better understanding of the ontology. Onto graph is also used for visualization. To see the nodes and its relationships in protégé tool. OWLViz visualizes the class structure of an ontology. It is a part of protégé-OWL and it is based on graph viz.

The taxonomy definition of graphical presentation of liver cancer ontology is given below. Diagnosis of evidence-based from main source and reuse of liver cancer ontology gives 5 selected candidate terms. We identified the terms in an unstructured list and its summary is given below.

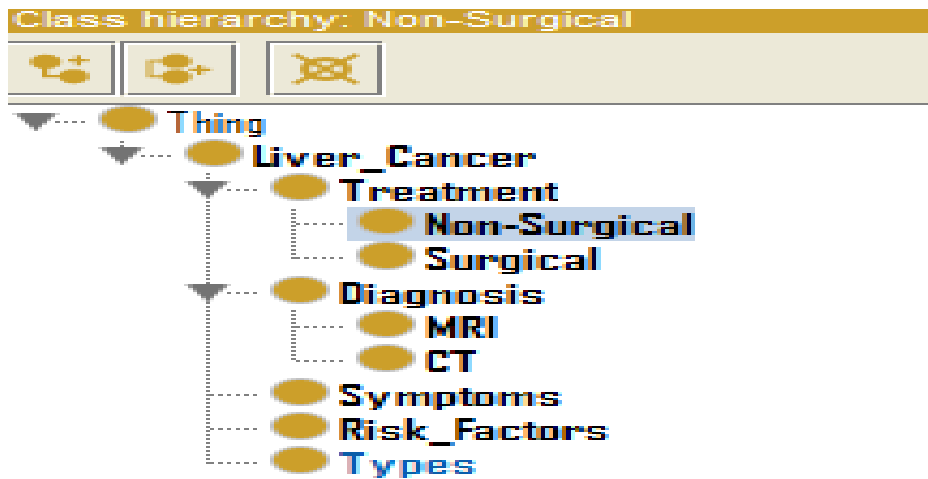


FIGURE4.11: HIGH LEVEL TERMINOLOGY OF LIVER CANCER ONTOLOGY

The candidates which are recognised for the high stage are normalised and grounded to the canonical names. This structure creates a taxonomy hierarchy which is organised by super sub concept relationship.

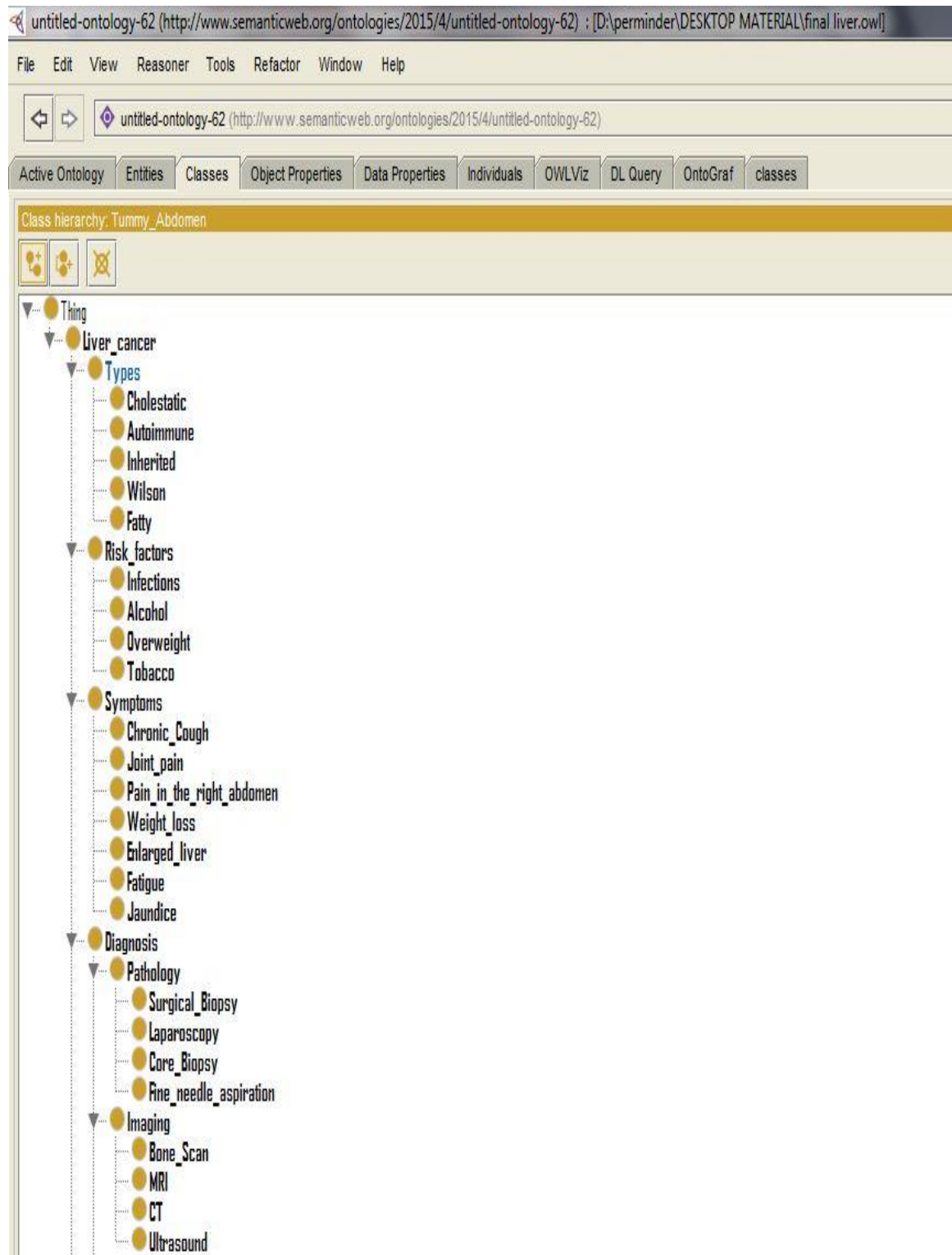


FIGURE 4.13: TAXONOMICAL HIERARCHY OF CONCEPTS IN LIVER ONTOLOGY

Onto Graph give support for interactively navigating the relationship of your OWL ontologies. In Onto graph various layouts are supported for automatically organizing the structure of your ontology. The onto graph supports different relationships that include classes, individuals, range properties, data type properties and equivalence. Relationships and node types can be filtered to help you for creating the ontology. We can zoom the graph larger or smaller to see more nodes and relationships.

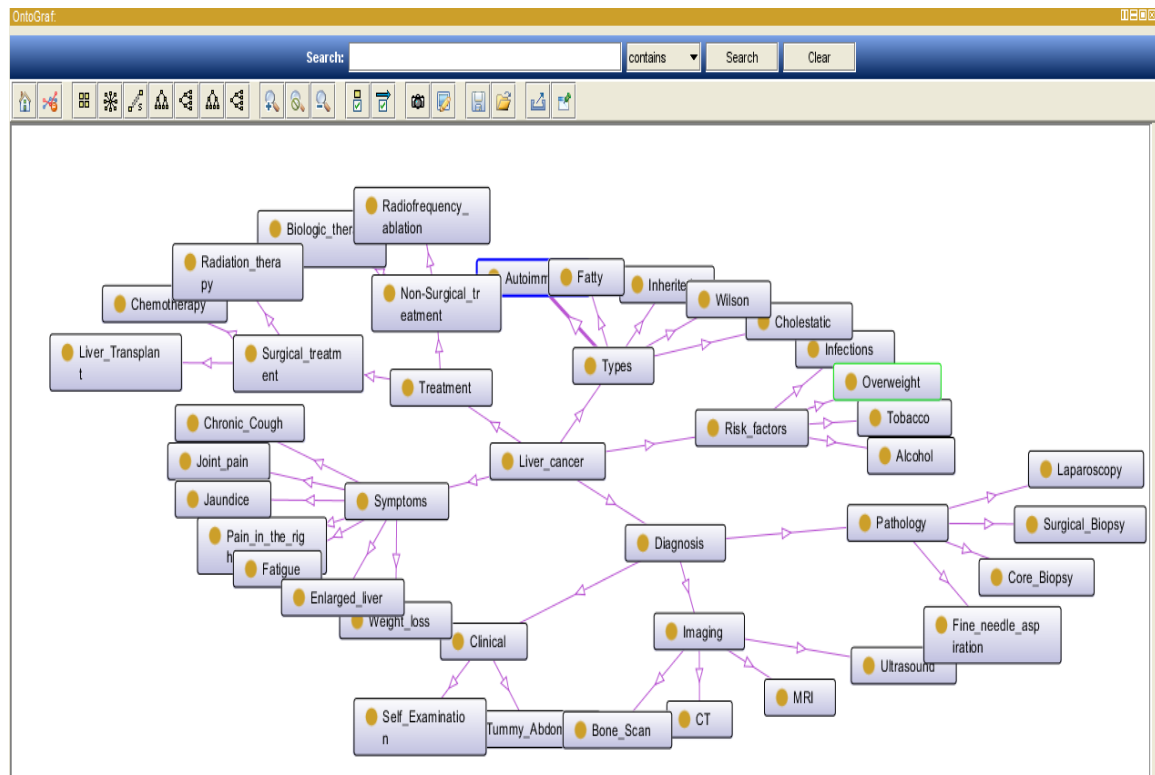


FIGURE 4.14: ONTO GRAPH OF LIVER CANCER

In liver cancer ontology SPARQL retrieves and manipulates query data of liver domain based on querying of various diseases. Various object and data properties with query editor can be expressed from liver ontology. Query retrieval related to risk factors, symptoms, diagnosis and treatment with patients. Different queries are grouped into different types according to number of classes they enter and the number of classes having common characteristics. Different query types are presented with brief descriptions. We use different SPARQL queries are ask, select, construct and describe. The main application of SPARQL queries are PSM. SPARQL consist of triple patterns, conjunctions, disjunctions and optional patterns. SPARQL provides a set of analytical query operations such as join, sort and aggregation.

1. To identify the symptoms which are responsible for liver cancer type 'Wilson'

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?Symtoms
WHERE
{?Liver_Cancer LC: hastype "WILSON"}
```

FIGURE 4.15: QUERY RELATED WILSON DISEASE

2. What are the various liver cancer types for which risk factors become Tobacco.

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?factor_type
WHERE
{
{?Symptom ab: hastype "Fatty"}
.
{?Risk_factor ab:hastype "Tobacco"}
}
```

FIGURE 4.16: QUERY RELATED TOBACCO

5.1 CONCLUSION

In data mining with the help of weka tool we generate a random decision tree with the help of classification to reveal the attributes of liver cancer. The attributes is categorised into high, medium, low, medium/low. On the basis of the analysis derived from weka tool we conclude that the disease is fatty liver, cholestatic, autoimmune, inherited, Wilson We also create the liver cancer ontology with the help of protégé tool on the basis of symptoms, types, diagnosis, risk factors, treatment. Protégé is a knowledge representational tool.

5.2 FUTURE SCOPE

The future scope of liver cancer ontology is that this work can be carried forward by analyzing the classification of liver cancer and developing the liver cancer ontology using ontology based tools on some other parameters also apart from the two parameters and random algorithm that are already discussed. i.e. Neurological- Psychiatric and Pathophysical –Cognitive. New algorithm can be developed for this purpose and input and output would also change accordingly. The ontology can also be developed by using other liver based diseases.

Reference for an article:

- [1]. Sumbaly, Vishnusri, Jeyalatha (2014)” Diagnosis of Breast Cancer using Decision Tree Data Mining Technique “,*International Journal of Computer Applications*, Volume 98- No.10.
- [2]. Ali, Hussian, Zakir (2014),” Intelligent Image Processing Techniques for Cancer Progression Detection, Recognition and Prediction in the Human Liver”, *IEEE*.
- [3]. Upadhay,Wasson (2014) “ Application of Genetic Algorithm for Liver cancer Detection”, *International Journal of research in Engineering & Technology*, Volume 03, Issue 05.
- [4]. KundraPandey (2014), “ Classification of EEG based Diseases using Data Mining”, *International Journal of Computer Applications*, Volume 90.
- [5]. Gangwar, Mishra, Yadav, Pandey(2013),” Intelligent Computing Methods for the Interpretation of Neuropsychiatric Diseases based on Rbr – Cbr –Ann Integration” *International Journal of Computers & Technology*, Vol 11.
- [6]. Jensen, Cox, Pandhey (2013), “ The Neurological disease Ontology”, *Journal of Biomedical Semantics*, Department of Neurology ,USA.
- [7]. Bodenreider (2013),” Disease Ontologies”, *Encyclopedia of Systems Biology: Springer*, US National Library of Medicine, Maryland, USA.
- [8]. H.Satria, R.S. Priya, L.H. Ismail, E.Supriyanto (2012),” Building and Reusing Medical Ontology for Tropical Diseases”, Management, *International Journal Of Education And Information Technologies*, Issue 1, Volume 6.
- [9]. I.F. Moawad, G.AL Marzoqi, A- badeeh M. Saleem,(2012)” Building OBR- based OWL Ontology for Viral Hepatitis”, *Faculty of Computer and Information science, Ain shams University, Egyptian Computer Science Journal, ECS*, Vol.36
- [10]. A Herdiani1, L. Fitria2, H. Hayurani3, W.C. Wibowo4 and S. Sungkar5 (2012),”Hierarchial Conceptual Schema for DengueHemorrhagic Fever Ontology”, *IJCSI International Journal of Computer Science Issues*,Vol.9, Issue 4.

- [11]. M. Alfonse, M. Aref, M. Salem (2012), "Ontology –Based Knowledge Representation for Liver Cancer", *Computer Science Department Faculty Of Computer & Information Sciences Ain Shams University*, pp.18-20 .
- [12]. Kalaivani, Anandaraj, Raja (2011), "An Ontology Construction Approach For The Domain Of Poultry Science Using Protege", *International Journal Of Information Technology and Management Sciences*, Volume 1, Issue 2.
- [13]. Adams, Hoehndorf, Gkoutos (2011), "The Primary Immunodeficiency Disease Ontology", *Bioinformatics Advanced Access*, The Oxford University.
- [14]. Mungall, Bada, Beradini (2011), "Cross Product Extensions of the Gene Ontology", *Journal Of Biomedical Informatics*.
- [15]. Schneider, Brochhausen (2011), "The Chronious Ontology Suite: Methodology and Design Principles", *International Conference On Biomedical Ontology*, USA.
- [16]. Topalis, Dialynas, Mitraka, Deliyanni (2011), "A Set Of Ontologies to drive tools for the Control of Vector- Borne Diseases", *Journal Of Biomedical Informatics*, pp.42-47.
- [17]. Wu Qiu , Rui Wang, Fengxiao (2011), "Research on Fuzzy Enhancement in the Diagnosis of Liver Tumor from B-mode Ultrasound Images", *IEEE*.
- [18]. H. Massieh, M. Hadhoud, M. Amin (2010), "Automatic Liver Tumor Segmentation from CT Scans with Knowledge- based Constraints", *IEEE*, Egypt.
- [19]. Oezdenwernerberg, Sonja Zillner (2009), "Towards Context Driven Modularization Of Large Biomedical Ontologies", Munich, Germany.
- [20]. Mizoguchi, Kozaki (2009), "An Advanced Clinical Ontology", *International Conference on Biomedical Ontology*, Buffalo.
- [21]. Daramola, Olawande, F. Segun (2009), "Developing Ontology Support for Human Malaria Control Initiatives", *Department of Computer and Information Sciences, Covenant University, Ota, Nigeria*.
- [22]. Christine Denney¹, Colin Batcher², Olivier Bodenreider (2009), "Creating a translational Medicine Ontology", *Proceedings of the First International Conference on Biomedical Ontology*, 2009:164.
- [23]. Bonevich, Dull, Gates, Steck (2009), "Discovery Of Novel Cancer Genes Through Application Of Clinical Metadata Ontology", *Compendia Bioscience*, United States.
- [24]. Janna Hastings, Paula De Matos (2009), "Towards Automatic Classification Of Entities with the ChEBI Ontology", *International Conference On Biomedical Ontologies*, USA.

- [25]. PallaviMokkarala, Julie Brixey(2008),” Development Of a Comprehensive Medical Error Ontology”, *Department Of Biomedical Informatics, Kansas*.
- [26]. Wang, D Haynes, M.D (2005) “*AN ONTOLOGY OF BLOOD PRESSURE: A CASE STUDY USING PROTÉGÉ OWL* “
- [27]. Alan Jovic, Gamberger(2004),”Heart Failure Ontology” .
- [28]. Holmes, A. Donkin, I. H. Witten. WEKA: A Machine Learning Workbench. *In Proceedings of the Second Australian and New Zealand Conference on Intelligent Information systems*. 357-361.1994.
- [29]. I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation. *Morgan Kaufmann publishers*, 1999.
- [30]. DraganGašević; DraganDjurić; VladanDevedžić (2009). *Model Driven Engineering and Ontology Development* (2nd ed.). Springer.p. 194. ISBN 978-3-642-00282-3
- [31]. Hebel, John; Fisher, Matthew; Blace, Ryan; Perez-Lopez, Andrew (2009). *Semantic Web Programming*. Indianapolis, Indiana: John Wiley & Sons. p. 406. ISBN 978-0-470-41801-7.
- [32]. Segaran, Toby; Evans, Colin; Taylor, Jamie (2009). *Programming the Semantic Web*. O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. p. 84. ISBN 978-0-596-15381-6.
- [33]. Bruix, J; Sherman, M; American Association for the Study of Liver, Diseases (March 2011). "Management of hepatocellular carcinoma: an update". *Hepatology (Baltimore, Md.)* **53** (3): 1020–2. doi:10.1002/hep.24199. PMID 21374666.

REFERENCE TO WEB PAGE

- [1]. [http://www.obitko.com/tutorials/ontologies semantic -web/what-is-ontology.html](http://www.obitko.com/tutorials/ontologies%20semantic%20-web/what-is-ontology.html)
- [2]. <http://en.wikipedia.org/wiki/Ontology>
- [3]. <http://www.cancer.org/cancer/livercancer/detailedguide/liver-cancer-staging>
- [4]. <http://www.cancer.org/cancer/livercancer/overviewguide/liver-cancer-overview-what-is-liver-cancer>
- [5]. <http://www.kucancercenter.org/cancer-information/specialties-and-treatment/liver-cancer>
- [6]. <http://weka.sourceforge.net/doc.stable/weka/classifiers/trees/package-summary.html>
- [7]. <http://www.gastroliverspecialist.com/liver-disease.html>
- [8]. <http://www.slideshare.net/OpenDataSupport/introduction-to-rdf-sparql>
- [9]. http://protege.stanford.edu/conference/2006/submissions/slides/AppDevelopmentTutorial_Part1.pdf
- [10]. <http://protege.stanford.edu/plugins/owl/index.html>
- [11]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560433/>
- [12]. <http://en.wikipedia.org/wiki/SPARQL>
- [13]. http://www.w3.org/standards/techs/owl#w3c_all
- [14]. <http://protegewiki.stanford.edu/wiki/OntoGraf>

7.1 LIST OF ABBREVIATIONS

TNM = Tumor Node Metastasis.

PDT = Problem Description table

SWOOP= Semantic Web Ontology Overview and Perusal

DOE= Differential Ontology Editor

PSM= Problem Solving Method

GUI = Graphical User Interface.