INTELLIGENT MEDICAL DIAGNOSTIC SYSTEM FOR HEPATITIS B USING MACHINE LEARNING ALGORITHMS

A Thesis

Submitted in partial fulfillment of the requirements for the

Award of the degree of

DOCTOR OF PHILOSOPHY

In

COMPUTER SCIENCE AND ENGINEERING

By

Dalwinder Singh

(41800159)

Supervised By

Dr. Deepak Prashar

Associate Professor

Department of Computer Science and Engineering

Lovely Professional University.

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

2021

DECLARATION

I hereby certify that the work that is being presented in the thesis entitled "Intelligent Medical Diagnostic System for Hepatitis B Using Machine Learning Algorithms" submitted to the Department of Research at Lovely Professional University, Phagwara is an authentic record of my own work carried out during a period from August 2018 to July 2021 under the supervision of Dr. Deepak Prashar, Associate Professor, Department of Computer Science & Engg., LPU. The matter presented in this thesis has not been submitted in any other University/Institute for the award of PHD degree.

Signed : **Dalwinder Singh**

Date : **26.9.2021**

Place : **Phagwara**

CERTIFICATE

This is to certify that the thesis entitled “Intelligent Medical Diagnostic System for Hepatitis B Using Machine Learning Algorithms”, which is being submitted by Mr. Dalwinder Singh for the award of the degree of Doctor of Philosophy in Computer Science and Engineering from the Faculty of Technology and Sciences, Lovely Professional University, Punjab, India, is entirely based on the work carried out by him under my supervision and guidance. The work reported embodies the original work of the candidate and has not been submitted to any other university or institution for the award of any degree or diploma, according to the best of my knowledge.

Dr. Deepak Prashar
Associate Professor
Department of Computer Science and Engineering
Lovely Professional University,
Phagwara, Punjab-144411
(India)

ABSTRACT

Hepatitis B is a virus, which can attack the liver of a human body and leads to major disease of the liver. This virus is a possibly deadly infection of liver. The hepatitis infection mainly disturbs all the operations done by the liver. Moreover, this infection now becomes the world's most severe kind of virus among all hepatitis as out of 12 individuals; one is resulted as positive for hepatitis B disease globally. Hepatitis B is a short term disease for few individuals. But, sometimes, for few patients, it will become a chronic as well as long term disease. This severe infection will lead to various deadly diseases that can infect the liver of an individual completely. The most life threatening disease caused by it is a chronic liver disease which further leads to cancer as well as cirrhosis, and as a result, the life of a patient will put at significant risk. Therefore, it is very crucial and become a necessity to detect or identify this Hepatitis B virus at the very first stage or at an introductory stage. By doing so, the life of an individual can be saved for good. The biomarkers are also used to identify this deadly infection of the liver. If this infection is not cured until six months, then it will become chronic. However, the treatment and diagnosis of this life threatening disease is very expensive and can cause serious side effects. Hence, it is vital to develop a diagnostic system that reduces the cost of treatment and diagnosis.

The various artificial intelligence approaches have been used in the medical domain, which assists the doctor as well as experts in the diagnosis process. There are many techniques that are used to interpret the acquired data and further used for the process of diagnosis of any disease. However, the interpretation done by a specialist become very complicated as there is a considerable amount of data that has to be processed to make an accurate and appropriate decision. This data is range from the clinical symptoms of a particular disease to its biochemical data. Hence it is very crucial to make a model which helps the professionals or human expert to make the correct decision in a specific condition as the life of an individual is dependent on the decision taken by the doctor. Several medical diagnostic systems are already proposed by the distinct authors, but there is a huge need to propose as well as develop a medical diagnostic system that has less probability of error and has more accuracy of

classification. Due to the presence of technology, the process of diagnosis become very easy as the technology has many advantages such as its effectiveness, repeatability and has strong immunity, which helps to do those diagnoses as well where a human expert cannot do due it his or her specific factors.

There is a significant impact of each and every research work done by the various researchers, especially in the field of medicine, as the health of entire human kind is dependent on the result of particular researches. The primary intent of this research study is to propose as well as develop a medical diagnostic system which aids the doctors to make accurate decisions and has a low cost of treatment also with no side effects. Hence, in this research work, a medical diagnostic system to diagnose the Hepatitis B virus infection is developed. A multi-layered fuzzy inference system has been proposed by using the Mamdani fuzzy. In this proposed system, there are two layers, and both layers have different input variables. This system classifies hepatitis B patients and non-hepatitis B patients into different classes. The input variables used for layer 1 are jaundice, dark urine, abdominal pain and vomiting. Similarly, layer 2 uses the input variables such as HBsAg, Anti - HBs or HBsAb, Anti - HBc or HBcAb, HBV DNA and Anti – HbcAg - IgM. Layer 1's output is also used as the input for layer 2. Layer 1's output is either yes or no, as this layer only detects whether the individual suffering from this virus or not. Likewise, layer 2 gave the output as no HBV, acute disease or chronic disease. The accuracy of the system is also evaluated by test cases. The model has been proposed by using MATLAB R2014a software.

Additionally, this research effort is proposed an intelligent system that assists in the identification as well as diagnosis of the Hepatitis B virus in stage 1. This system is proposed by utilizing the neuro-fuzzy technique. An ANFIS is the collection of 2 soft computing methodologies named fuzzy inference system and ANN. The abilities of an ANN, as well as the capabilities of fuzzy logic, are combined together to build a new hybrid system. The developed hybrid system has 5 layers, i.e. fuzzification layer, rule layer, normalization layer, defuzzification layer and summation layer. The input variables or linguistic variables that are used in this study are HBsAg, Anti - HBs or HBsAb, Anti - HBc or HBcAb, HBV DNA and Anti – HbcAg - IgM. Similarly, the

output variables of this system are no HBV, acute disease or chronic disease. This hybrid system has been developed by using software named as MATLAB.

The diverse parameters that assist in the evaluation of performance are also determined by using the observed values from the proposed system for both developed models. The classification accuracy of a multi-layered fuzzy inference system is 94%. The accuracy with which the developed medical diagnostic system by using ANFIS classifies the result corresponding to the given input is 95.55%. These proposed systems assist the experts as well as professionals to diagnose the hepatitis B virus easily and also with minimal cost.

In order to figure out which developed model is best for the diagnosis of hepatitis B disease, a comparative study has been done among the fuzzy inference system as well as adaptive neuro-fuzzy inference system. By doing so, it is observed that which model has the capability to classify the provided inputs into accurate classes. The obtained result from both, i.e. fuzzy inference system as well as neuro-fuzzy inference system, are also compared with the results provided by the expert or professional doctor of hepatitis B disease. A comparison of both developed models on the basis of their performance parameters has been made. The obtained observation from is that the neuro fuzzy technique based diagnostic system has better accuracy in classifying the infected patients and non-infected patients as compared to the fuzzy diagnostic system. In other words, the adaptive neuro-fuzzy inference system has more capability to classify the offered inputs adequately than the fuzzy inference system.

Moreover, the classification accuracies of the already developed systems or existing systems are also compared with these two developed models. It is observed that the developed models are more accurate as compared to other existing ones as the accuracy of the classification of these two developed models i.e. multi-layered fuzzy inference system and adaptive neuro fuzzy inference system are more whereas the existing models have less accuracy of classification. After the comparative study between both the developed models and between the developed models with existing system, it is analysed that the most accurate model that can be used as the medical diagnostic system for the diagnosis of hepatitis B virus is adaptive neuro-fuzzy inference system.

ACKNOWLEDGMENTS

First of all, I pay my gratitude to the Almighty for conceding me an opportunity and strength to successfully accomplish this assignment.

I would like to thank my esteemed supervisor, **Dr. Deepak Prashar** for his valuable supervision, support and tutelage during the course of my PhD degree. I owe my deep sense of gratitude to the dynamic academician and generous guide, who all along has been an immense source of encouragement to me in completing this work. He really spared time for academic interactions with me.

My heartfelt gratitude and indebtedness go to my loving parents **(Sh. Kirpal Singh and Smt. Santosh Kumari)** for their blessings and encouragement. The efforts of my wife, **Er. Sheetal Katal** deserve a great appreciation for understanding and cooperation. I also appraise my children **(Agrima Salaria and Abhiraj Singh Salaria)** who adjusted themselves when I was busy overnights doing this thesis work.

Finally, I very sincerely acknowledge the cooperation of my friends and colleagues, who helped me a lot during this work. I am thankful to **Dr. Sahil Verma, Dr. Jimmy Singla, Dr. Harjit Singh, Dr. Arun Malik, Mr. Janpreet Singh and Mr. Kewel Krishan** for timely help and friendly guidance. I would like to thank all the authors whose studies have been quoted in this study.

Dalwinder Singh

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter – 1

INTRODUCTION

Hepatitis B is a deadly and severe infection of the liver. This infection is mainly caused by a virus, namely hepatitis B virus or HBV [1]. This virus spread from one person to another by various means such as semen, blood or other fluids of the body. The disease spread from person to person when these fluids transfer from the infected body to the uninfected body. The transfer of the hepatitis B infection is done due to sharing the needles, sexual contact [2], sharing the injections during the consumption of any drug, or it can be transferred from the mother to her children at the time of birth. It is not mandatory that the symptoms of this virus will be shown in each and every newly infected person. However, if the person is infected, then the symptoms will be pain in the stomach, fatigue, jaundice, poor appetite as well as nausea.

Hepatitis B is a short term disease for few individuals. But, sometimes, for few patients, it will become a chronic as well as long term disease. Then, this will result in a serious condition of health or even can be considered a life-threatening disease such as liver cancer or cirrhosis. The several stages and symptoms are there, which aids to classify the disease [3]. The biomarkers are also used to identify this deadly infection of the liver. If this infection is not cured until six months, then it will become chronic. This chronic stage of infection might lead to liver failure or liver cancer. The adults can survive this infection even if the infection is severe. But if the infection occurred in children and infants, then there are more chances to get it worst or can lead to the chronic stage [3]. Hence, the risk of this infection becoming chronic is also related to age. It is observed that approximately 90% of the patients are newly born in which this infection developed to the chronic stage, and only 2 to 6 percent of people are adults who are chronically infected by this virus. The only method to prevent the hepatitis B virus is to get vaccinated at the right time [5].

This disease is also a major health issue globally. If this disease is not diagnosed at a time, then this will lead to chronic infection, and also, the probability of death of that

particular patient increases due to liver cancer or cirrhosis [4]. In the past years, this disease grabs attention at a wider level as this disease has an extreme risk of cancer as well as liver diseases on the health of human. The hepatitis B virus infected more than 240 million people at a global level. Hence, the treatment of this disease became crucial [7].

1.1 STRUCTURE OF HBV

The systematic structure of the HBV is shown in figure 1.1. All the particles which are infectious, as well as non-infectious, are represented accurately in this figure. Mainly, the three kinds of particles were observed by the researchers in the serum of a patient who is infected by this virus. The first particle is enveloped nucleocapsid, which is consisting of Ribonucleic acid (RNA) or Deoxyribonucleic acid (DNA). The second particle is known as Sub viral particles, which are consisting of spherical and filament particles. And, third is Naked Nucleocapsid. These three particles are also known as the Dane particles [8].

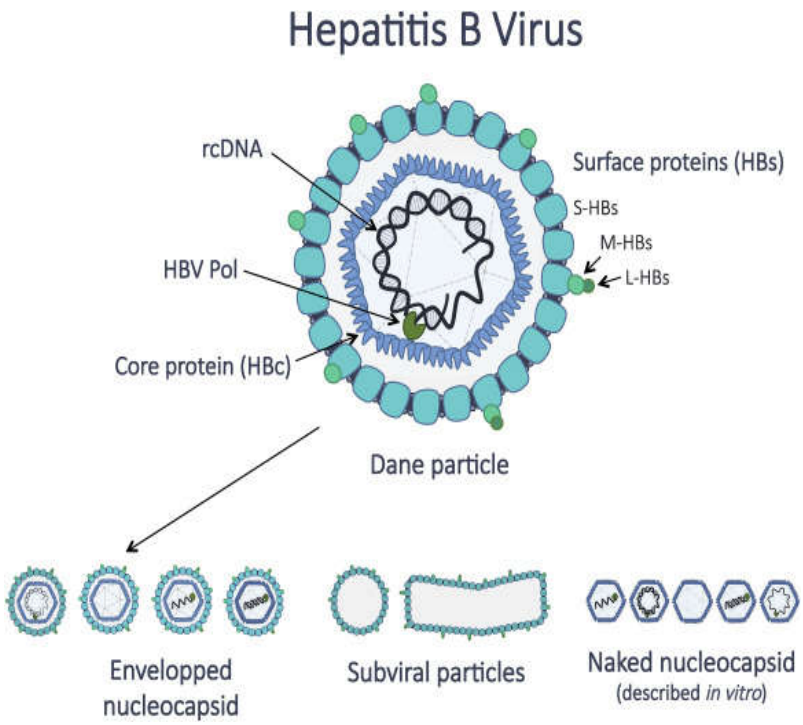

Figure 1.1: Systematic structure of Hepatitis B Virus or HBV [8]

1.2 THE LIFE CYCLE OF HBV

The entire life cycle of the hepatitis B virus is shown in figure 1.2. The key processes of this life cycle are first to generate the Hepatitis B virus in DNA, and then the host cell will viral this virus into the RNA of the body. After this, the second process is to reverse transcription, which happens within the viral virus, to form CCC DNA (Covalently Closed Circular DNA).

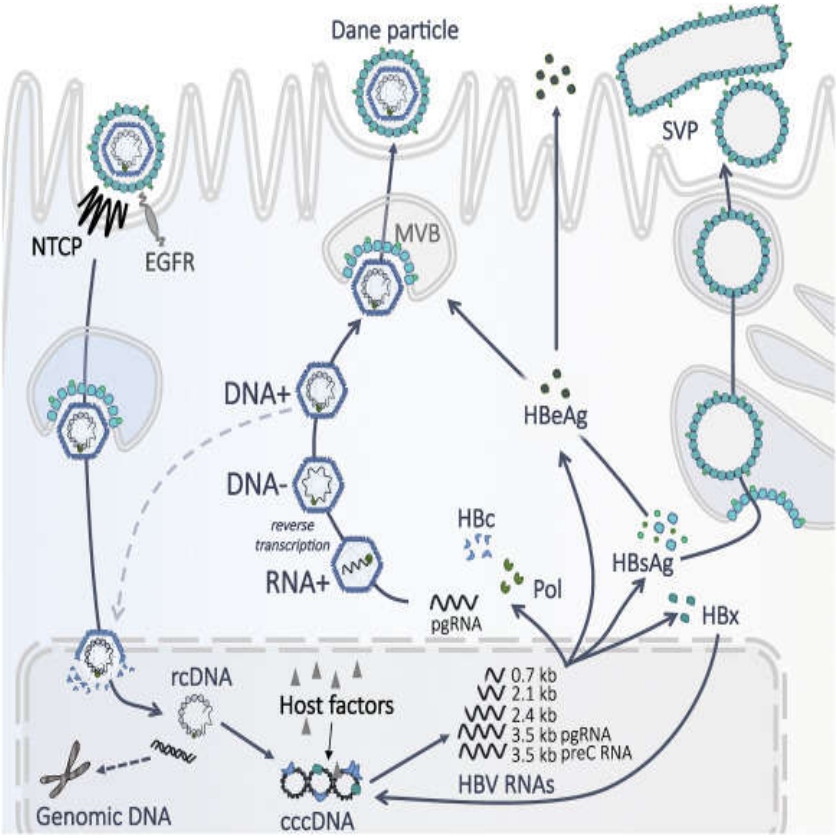

Figure 1.2: Life cycle of HBV [8]

1.3 SYMPTOMS OF HBV

The range of the symptoms of hepatitis B virus varies from mild conditions to severe condition. These signs are not observed at the initial stage [9]. It takes approximately one to four months to appear after any individual have been infected with HBV. Even though, these can be noticed after the two weeks as well. Some

patients, such as adults, do not have any kind of signs or symptoms at all. The various symptoms or signs of Hepatitis B virus are shown in figure 1.3 and also, given by [4], [10]:

- Fever
- Pain in joints
- Fatigue
- Weakness
- Pain in Abdominal
- The colour of urine is dark than normal
- Whiteness in the eyes
- Yellowing of skin due to Jaundice
- Vomiting
- Nausea

All the above symptoms can be judged physically and mentally. Fever, fatigue, weakness and vomiting are the symptoms that one can notice easily. Whiteness of eyes and yellow skin due to jaundice can be recognised by others easily. All these symptoms are very crucial to identify in the beginning to avoid worst effect of Hepatitis-B.

Figure 1.3: Signs or Symptoms of Hepatitis B Infection

1.4 CAUSES OF HBV

As it's mentioned above, the Hepatitis B infection is mainly spread by a virus, namely hepatitis B virus or HBV. This deadly virus can transfer from an infected individual to a non-infected individual through some kind of fluids of a person's body, such as semen or blood. Coughing or sneezing is not able to spread this virus [11]. The standard ways that contribute to the expansion or transmission of hepatitis B virus are described below [4]:

- **Needle Sharing:** The hepatitis B virus extends easily when a non-infected person shares the syringe or needle with an infected person as it contaminates the blood. Sharing the needle or injection needle for drugs might put an immense risk of HBV.
- **Sexual Contact:** The hepatitis B virus also spread when an individual has sex with an infected person without any protection. This is because, during the sex, there will be the transmission of saliva, blood, vaginal secretions or semen from one person to another and hence, virus.
- **Unintentional Stick of Needle:** Hepatitis B is also a major perturb for employees of medical services and another individual who interacts with human blood.
- **Mother to Child:** During childbirth, if a mother is infected by this virus, then this virus will automatically pass to the child. However, the vaccine can be provided to these new-born babies to prevent them from this life-threatening virus and also it works in many cases. Hence, first, monitor this virus by consulting with a specialist or doctor if any woman is pregnant.

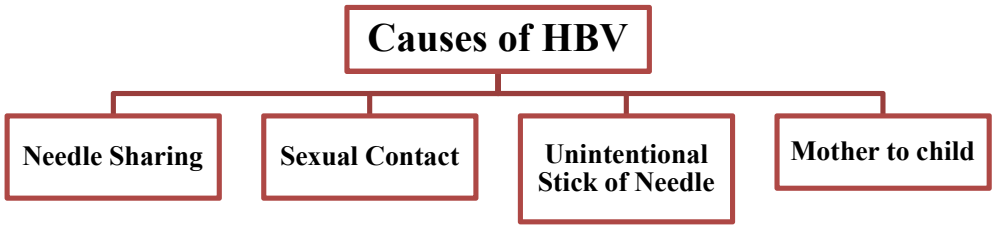

Figure 1.4: Causes of Hepatitis B Virus

1.5 PREVENTION OF HBV

The only prevention of hepatitis B virus is the vaccine. The complete vaccination of hepatitis B includes 3-4 injections in a period of six months [12]. As we know that there is no cure of this disease hence, it is mandatory to prevent this disease. These hepatitis B vaccines are recommended to newly born babies, children who are not vaccinated after their birth, the persons who lived with the infected individual, patients of chronic liver disease as well as kidney disease etc.

It is crucial to detect this virus in the early stages so that the vaccination can be provided to the needed patient, which leads to an increase in the life span of that particular infected individual. The biomarkers for this disease has been identified and also compared with each other by using the multimodal biomarkers [13]. The mining mutation hotspots are used to identify or detect the best clinical biomarkers. The feature tree also assists in finding the contribution of the biomarkers in causing disease. These biomarkers are further assists in the detection and treatment of the HBV virus in a particular patient at early stages [14].

The Granular Computing theory has been used to categorize the life-threatening disease by using the symptoms and stages [3]. There are numerous cells that cause this Hepatitis B. These host cells are identified by using mathematical models, and also, this model helps to understand the dynamic nature of cells. The different symptoms are accounted, and according to those symptoms, Hepatitis B is diagnosed [9]. There are five different viruses that can be considered as hepatitis infection. These are HAV, HEV, HBV, HCV and HDV. The classifiers such as Naïve Bayes, J48 and K-Star have been used to classify these different viruses into distinct classes [10]. The HBV RNA is also a biomarker for hepatitis B infection that circulate in the patient body [11].

The data mining approaches such as clustering, classification and others are also used to diagnose Hepatitis B infection. The meaningful information from the experimental results is also extracted by using data mining techniques [12]. The missing values and normalization on the experimental data are also done by using data mining techniques [13]. The K-means classification is also used for the classification of different DNAs that causes this deadly disease [14], [15]. The Bayesian inference based methodology

has been used for the identification of several viruses like HBV, hepatitis C virus, as well as HIV. This detection basically has been done according to the mutation patterns [16]. The random forest technique with Bayesian has been used for the prediction of risk factors of this severe disease [17]. The decision tree algorithms are also used to propose a system that can analyze the disease from its symptoms and having better accuracy [18], [19].

An expert system by utilizing the Generalized Regression Neural Network has been developed for the early detection of Hepatitis B. This proposed system classified the infected patients and non-infected patients into two different classes [20], [21]. The other artificial intelligence approaches such as SVM and neural networks were also applied to propose a system that aids in the diagnosis of hepatitis B infection. This developed system classifies the patient into six distinct classes [22], [23].

The hybrid system like adaptive neuro-fuzzy inference system, which is a collection of two different approaches of artificial intelligence, i.e. fuzzy logic and neural network, is also used for the recognition of hepatitis B. Additionally, the comparison of the Support vector machine and the neural network has been made [24]. The multi-layered fuzzy inference system has been developed by using numerous risk factors as the input variables and classify the disease into distinct classes [25].

The detection of the Hepatitis B virus can also be done with an expert system by using fuzzy logic [32] as well as by using an ANFIS. The expert system constructed by utilizing this methodology can make their own decisions identical to the human decisions. Fuzzy logic conserves the several characteristics or properties of classical logic system [33]. The primary idea of this expert system is to acquire the knowledge or information from the specialists or experts about the particular disease or problem and then transfer this knowledge to the computer machine [34]. The fuzzy logic expert systems are not only used to identify the diseases but in several other domains as well [35]. The saved knowledge within the system will be used to perform a particular operation by the user. Later, the developed system will generate an outcome after numerous computations. It is basically a program of computer which is derivative of a well-known concept of artificial intelligence. Additionally, this system is also superior

for any real-time application or field [36], [37]. It is analysed that approximately 70 percent of domains used this methodology to deal with authentic applications [38].

There are several risk factors that can be used for the identification of Hepatitis B disease. These symptoms can be clinical symptoms or laboratory symptoms. The clinical symptoms are general symptoms that can be felt by any particular patient. In contrast, the laboratory symptoms are those which are detected after doing some sort of laboratory tests of the patient by the doctor [39]. The factors, i.e., clinical symptoms and laboratory tests are considered in this research work, are given in figure 1.5 and 1.6, respectively, and a detailed explanation of these factors will be given in another chapter.

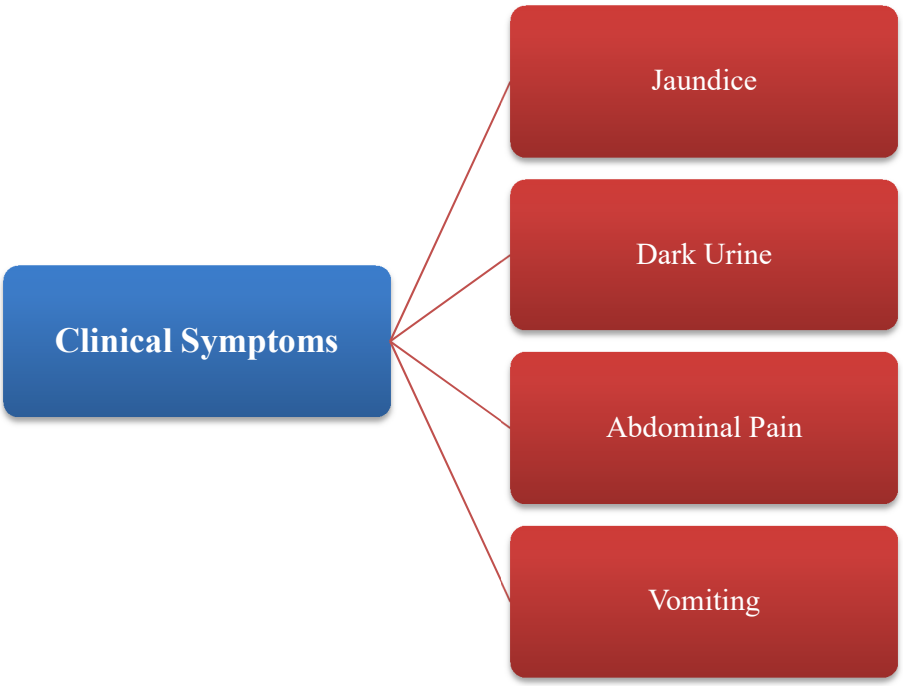

Figure 1.5: Clinical Symptoms

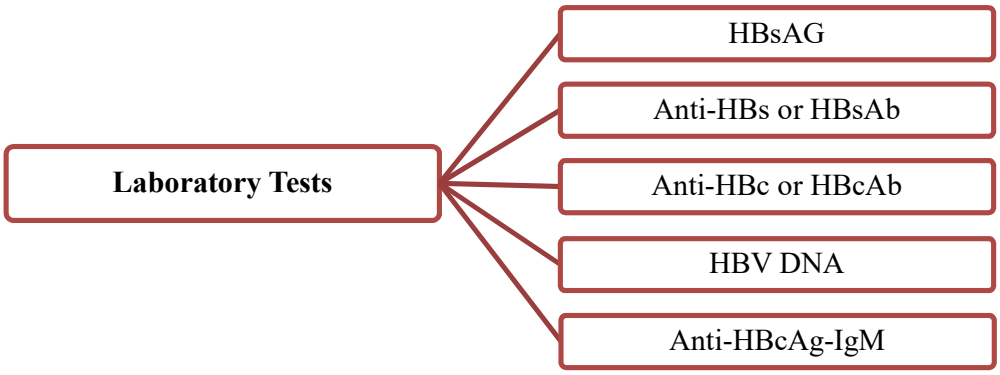

Figure 1.6: Laboratory Test

1.6 ARTIFICIAL INTELLIGENCE

In the modern era, technology is grooming rapidly, and each and every individual adopting it in a different manner day by day. One of the vastly growing computer science technologies is Artificial Intelligence, and this technology is already prepared to develop a new rebellion by proposing intelligent systems in this world. Artificial intelligence is a technology that is popular everywhere. Presently, this technique is used in numerous fields, starting from the general tasks of any domain to the specific operations. For instance: playing chess game or music, self-driving vehicles, to prove any theorems etc. [40].

Artificial intelligence has a tendency to develop a system that can operate identical to a human being. This fascinating property of this technology makes it a universal domain of computer science and immense scope in future as well [41].

As anyone can observe that there are two words in the name of this technique, i.e. “artificial” and “intelligence”. Both words have their own unique meaning, which ultimately defines the technique appropriately. The meaning of the first letter, artificial, means man-made, and similarly, the meaning of the second word, Intelligence, means the power of thinking. Therefore, by merging the meanings of these words, as a result, the connotation of artificial intelligence will generate that is man-made power of thinking [42].

Hence, artificial intelligence can be defined as it is an approach used to develop a system, a product, or a robot that can think as smart as a human being can [43]. In other words, when a system is much intelligent and can mimic a human being to operate any operation as well as to make adequate decisions as similar to the brain of a human, then that intelligence of a particular system is known as artificial intelligence [44]. It is an integrated field that is composed of many domains, for example, mathematics, neuroscience, computer science etc.

A system, which is created based on human skills like solving problems, reasoning and learning, then in that system, an artificial intelligence exists. With this delightful technology, it is not required to pre-program a system for any specific tasks. In spite of that, a machine only required a programmed algorithm that assists the system to do operations by using its own intelligence, and this is the best as well as amazing part of artificial intelligence technique. This technique is intangible and consists of various parts given below, as well as in figure 1.7.

- **Reasoning:** Reasoning is a process in which the inference is drawn to a particular condition or situation. The reasoning can be done by using an inductive method or deductive method.
- **Learning:** Learning is a process of artificial intelligence in which the particular system acquires knowledge as well as skills by using its experience or given data set. There are numerous kinds of methods that are used in artificial intelligence for the learning process. The most simple and easy process is the trial and error method that is also used by various models to grab the knowledge and reduce the probability of error.
- **Problem Solving:** In artificial intelligence, problem solving may be defined as the process in which the search is conducted systematically by using all possible actions and possibilities to accomplish the required as well as desired objective or intent.
- **Perception:** The process of interpretation of touch, smell as well as vision is known as perception. It helps the machine to gather the information from the sensors and act as identical as a human being.

- **Linguistic Intelligence:** Linguistic intelligence is the capability of the machine of artificial intelligence by which the machine can learn seen as well as unseen languages very quickly and further implement the learned language to fulfil any desired task.

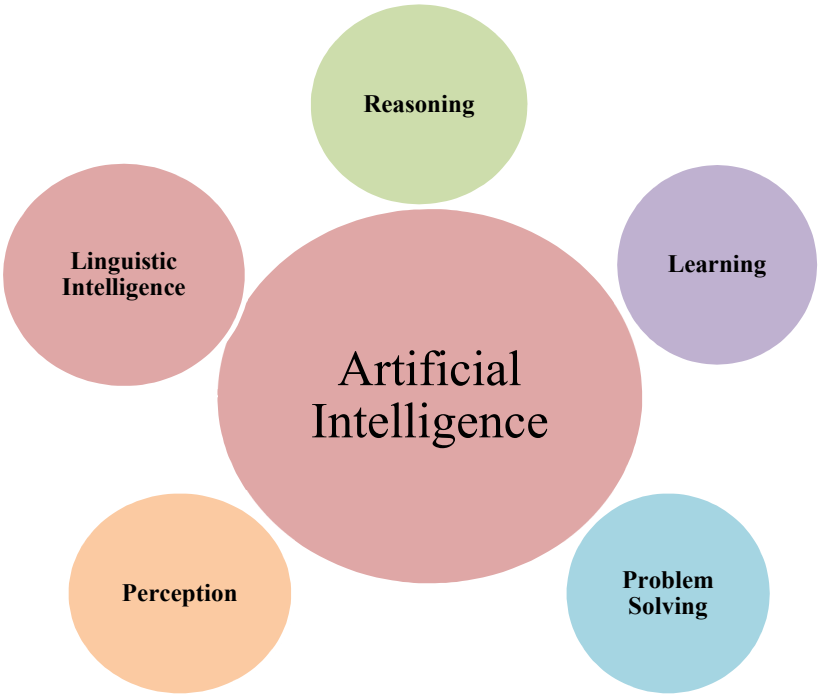

Figure 1.7: Composition of AI [43]

To accomplish the mentioned factors for a system developed by using artificial intelligence needs the below-given disciplines:

- Mathematics
- Statistics
- Computer Science
- Biology
- Neurons Study
- Sociology
- Psychology

1.6.1 Why Artificial Intelligence?

It is crucial to know the importance of artificial intelligence before starting the research work by using this technology. There are several domains in which artificial intelligence assists to solve complex as well as complicated problems. It can help to develop such systems which can figure out the problems of real-world like issues in traffic [45], healthcare [46], marketing [47] etc. simply as well as accurately. The systems created by using this approach can perform their task appropriately where a human cannot do work due to some sort of risk. This approach also offers new opportunities, devices as well as latest and useful technologies. This technique is not only suitable for solving complicated problems but also can be used in data handling [42].

The significant goals of this approach are the representation of knowledge, reasoning, learning, planning, realization, natural language processing and the capability to manipulate as well as to move things. Artificial intelligence also has long term objectives that are fruitful in future for the sector of intelligence [43].

1.6.2 Process of Artificial Intelligence

There are basically three steps that are performed in the process of artificial intelligence [48].

1. **Learning**, in this step, the needful knowledge or information will be acquired from the specialist or expert for a particular problem, and after that, rules will be generated by using this gathered information. This information is stored in the knowledge base of the system. All the generated rules are also saved in the knowledge base block of the machine.

2. **Reasoning**, in this step, the generated rules in step 1 will be fired. The rules will be mapped first corresponding to the input given to the system, and then the system will give an outcome accordingly. Basically, this step is done by the inference engine of the system.

3. **Self-correction**, in this step, the intelligence system will correct itself to reduce the error as well as to give output accurately from experience. Hence the capabilities of the machine will enhance.

1.7 INTELLIGENT SYSTEMS

Intelligent systems or IS offers a methodological approach that is standardized and provide a solution to complicated problems as well as generate logical and decisive outcomes with time [49]. In other words, this system is a machine that has a tendency to gather the needful data and then determine it, due to which the particular machine can interact with other machines as well. Expert systems are fundamental components of artificial intelligence. The expert system is a program of computer which has the capability to deal with unclear information. The expert systems acquire the knowledge from their experiences and then use this knowledge to predict the result for any specific problem by using the natural evolution process. It is observed that approximately 70% of real-world problems are figured out by using these expert systems.

There are three fundamental components of the intelligent or expert system as given by [50]:

- Dialog Structure
- Inference Engine
- Knowledge Base

The first part of an expert system is “dialog structure”. The dialog structure is an interface which allows the communication or interaction of users with the expert system. The most popular and widely used interface which supports effective communication is the graphical user interface or GUI.

The next significant component of expert systems is the “inference engine”. The inference engine is the most crucial part of the expert system as it helps the system to test the considered hypothesis by using various strategies and leads to the required outcome. The problems are processed in this component of an expert system. This part utilizes the rules and facts for debugging and reasoning capabilities. Additionally, the

inference engine is also responsible for the generated final output or conclusion by the expert system with respect to the provided input.

The last component is “knowledge base”. The knowledge base is used to save the facts as well as rules about a specific domain. It is assumed that the knowledge base holds the power of the expert system. Hence, by looking at the importance of this component, it is necessary that the knowledge base should be accurate, complete as well as consistent.

1.7.1 Simple Rule-Based Intelligent Systems

The simple rule-based intelligent systems are an easier form of AI [51]. The rule-based system is a manner of encoding the knowledge of a human expert into a system that is automated. This system is constructed by using the collection of assertion as well as rules that decide how to respond to the set of assertions. The IF-THEN statements are used to express the rules. Hence, these are also known as production rules or IF-THEN rules [50].

For instance:

IF….. input 1 AND Input 2……. THEN …….. Output

These rule-based expert systems are also used to solve issues in healthcare, such as for the detection of any specific disease. The gathering of the information or knowledge from the doctors or experts is very crucial in that case as the rules will generate according to the acquired data.

1.7.1.1 Elements of simple rule-based intelligent systems

The simple rule-based expert or intelligent system composed of three fundamental as well as simple components as described below [50]:

- Set of facts
- Set of rules
- Criteria for termination

1. **Set of Facts:** The facts can be anything that is suitable for the system's starting state. The assertions are basically known as the facts.

2. **Set of Rules:** The collection of rules consists of all the operations that ought to be performed within a particular problem scope. The association of rules, as well as fact, is done in the IF part that will aid to perform some sort of operations in the THEN part. There must be clear and accurate rules in the expert system as these are responsible for generating the output. The inaccurate rules can affect the performance of any expert or intelligent system.

3. **Criteria for Termination:** The criteria for termination offers a condition that investigates whether a solution to a specific problem has been found or not. There must be a termination condition for each and every expert system; otherwise, the system will go into the infinite loop.

1.8 FUZZY LOGIC

Fuzzy logic is a methodology that has the main objective to do reasoning and make rational decisions as a human being in an uncertain as well as imprecise environment [52]. The reasoning done by the developed system by using this methodology is dependent on the knowledge stored in the system by acquiring the answers to certain questions from the experts or specialists. It is also considered as Multivalued logic as it gives a result between 0 and 1 or between true or false [33].

1.8.1 Fuzzy Expert System

A fuzzy expert system is a machine that can make rational decisions as accurate as a human with imprecise data by using the fuzzy logic methodology [53]. This system also helps in the diagnosis of various deadly as well as life-threatening diseases. The fuzzy expert system is basically the set of membership function and the rules generated from the gathered data. The fuzzy intelligent systems are developing tools and uses values that are in crisp form and transform them into fuzzy values. These systems are highly recommendable as well as suitable for the issues where the information about

the problem is incomplete and indefinite. Additionally, the fuzzy expert systems can also be used to the problem, which is hard to solve and where other tools fail to conclude the result with available information [54].

The most common terms used in this methodology are crisp set and fuzzy sets. These both terms are defined below:

- **Crisp Sets**

  The classical set or crisp set is defined as the collection or set of elements $y \in Y$ that can be finite. For instance: to explain the collection of elements are greater than 10. The membership function representation for the above-given example will be like this:

  $Y = \{y | y > 10\}$;

  where, Y is the collection of integers which are positive.

- **Fuzzy Sets:** The fuzzy set can be defined as if Y is the collection of elements designated by y, then a set of fuzzy values X in Y is the collection of ordered pair and is expressed by [55]:

  $$X = \{(y, \mu_X(y)) / \, y \in Y\}$$

  Where, $\mu_Y(x)$ is the x's membership function in Y.

  Now, the architecture of the fuzzy logic system is shown in figure 1.8. There are four components of fuzzy logic as given by [56]:

  - **Fuzzifier:** In this component of fuzzy logic architecture, the transformation of crisp sets into fuzzy values will be done. The input provided to the system is in the form of non-fuzzy values; hence the Fuzzification process converts that input into the fuzzy set.
  - **Knowledge Base:** The knowledge base basically works as a store of information. The facts that are initially available to solve a problem and the rules that are created from the gathered data from an expert are stored in this

component. The inference engine also uses the knowledge base to map the rules.

- **Inference Engine:** The inference engine is completely responsible for the generation of the final outcome. It maps the stored rules as well facts corresponding to the provided input to the system. The outcome generated by the inference engine is in the form of fuzzy values.

- **Defuzzifier:** In the process of defuzzification, the transformation of fuzzy values into the crisp set has been done. The output given by the inference engine is in the form of fuzzy values then passes to this component. After that, the defuzzifier converts the output into the crisp set, and this set is considered as the final outcome or final solution to a specific problem.

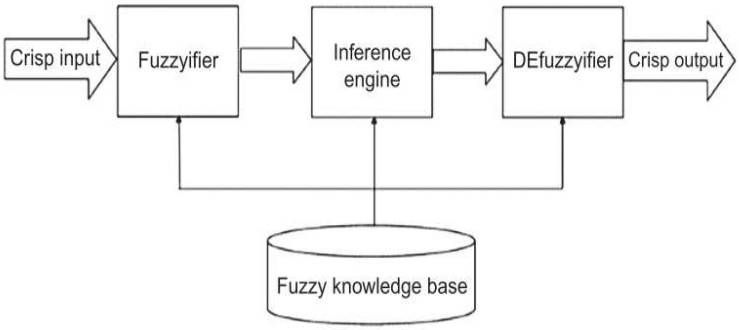

Figure 1.8: Fuzzy Expert System Architecture

1.9 NEURAL NETWORK

The idea of the artificial neural network was inspired by the refined functionality of the brain of human being in which several number of neurons that are interconnected with each other process the data in parallel [57],[58],[59]. The artificial neural network used to deal with various complicated mathematical tasks as well as large signal processing [60],[61]. By getting the motivation from this, the researchers try to construct a system with silicon that can work similar to the human brain and have intelligence. The developed system identical to the nervous system of human is known

as an artificial neural network. Several examples are used to train the network so that it can give accurate output.

Through the training phase or process, the neural networks are prepared to solve a specific problem or an issue. The neurons are interconnected with each other with weights. During the process of training, the weights are adjusted to reduce the chances of error. The formula to compute the error is subtracting the obtained value from the actual value. The ANN includes three layers that are one input layer, one or more hidden layers and, at last, an output layer. The architecture of the ANN is demonstrated in figure 1.9.

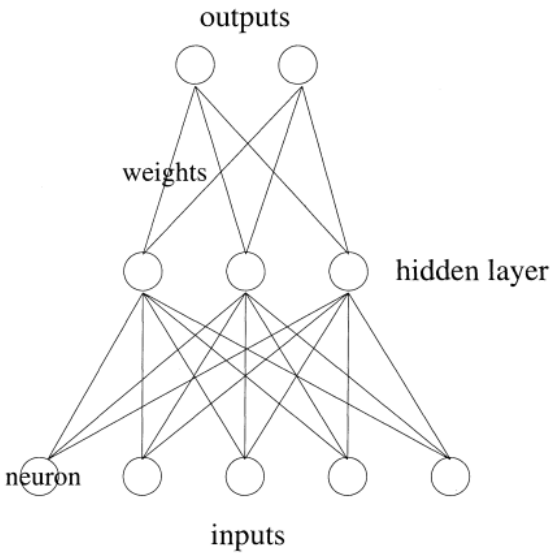

Figure 1.9: Artificial Neural Network [57]

- **Input Layer:** The collection of various neurons that are provided as the input to the network builds the input layer of the artificial neural network. The input, i.e., the initial information, is offered to the network by this layer. Moreover, this layer also has the tendency to support the distinct forms of inputs provided by the programmer or any user.

- **Hidden Layer:** The next layer of the artificial neural network is the hidden layer. The hidden layer situated between the input layer of the network and the output

layer of the artificial neural network. The network can also have one or more hidden layers. This layer is composed of numerous hidden neurons. Further, these hidden neurons do all the computations as well as calculations to figure out the features of the provided input to the input layer. Hence, the output of any artificial neural network depends upon the hidden neurons.

- **Output Layer:** The next layer of an artificial neural network is the output layer. This layer is composed of one or more output neurons. Although the input provided by the user to the network passes through first input layer then the computations of the hidden layer to calculate the output, but the final outcome is always generated by the output layer of the network.

Figure 1.9 displayed an architecture of a neural network in which the lines are associating the neurons. Every connection is related to a number which is numeric and known as weight. Assume that x_j are the inputs that given to the neurons of the input layer, N is the number of neurons of an input layer, V_{ij} are assumed as the weights, and b is bias or threshold provided to the hidden neurons. Also, the output layer used an activation function which helps to generate the outcome. Hence, consider σ as the activation function of the output layer. Now, the final output O_i of this network is given by:

$$O_i = \sigma\left(\sum_{j=1}^{N} V_{ij}x_j + b\right)$$

The artificial neural network is also utilized to the problems of classification by examining the existed information. It also has a tendency to investigate the complicated patterns which are complex to analyze by a human brain or any other methodology.

1.10 ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

It is found that the fuzzy inference system is relevant for the applications of medical-related, but some problems are huge complex that cannot be solved by fuzzy inference system; hence it fails to solve those particular problems. The main reason behind failure is that sometimes the rules, as well as membership functions of the expert

system, become exhausted. To solve this problem of exhaustion, the learning algorithm in the fuzzy expert systems is required. This need can only overcome by using the learning algorithm of the neural network in the fuzzy inference or expert system.

Thus, a hybrid system is proposed to figure out these issues that are not solved by a fuzzy inference system. An adaptive neuro-fuzzy inference system is the collection of 2 soft computing methodologies named fuzzy inference system and artificial neural network. The abilities of an artificial neural network, as well as capabilities of fuzzy logic, are combined together to form a new hybrid system. The developed hybrid system has 5 layers as displayed in figure 1.10. The ANFIS is a system that is very useful to solve distinct issues in several domains such as classification of data, analysis of data, recognition of complex patterns etc. The ANFIS is also used to detect or identify the disease with accurate and correct outcomes as compared to fuzzy inference or expert system or neural network.

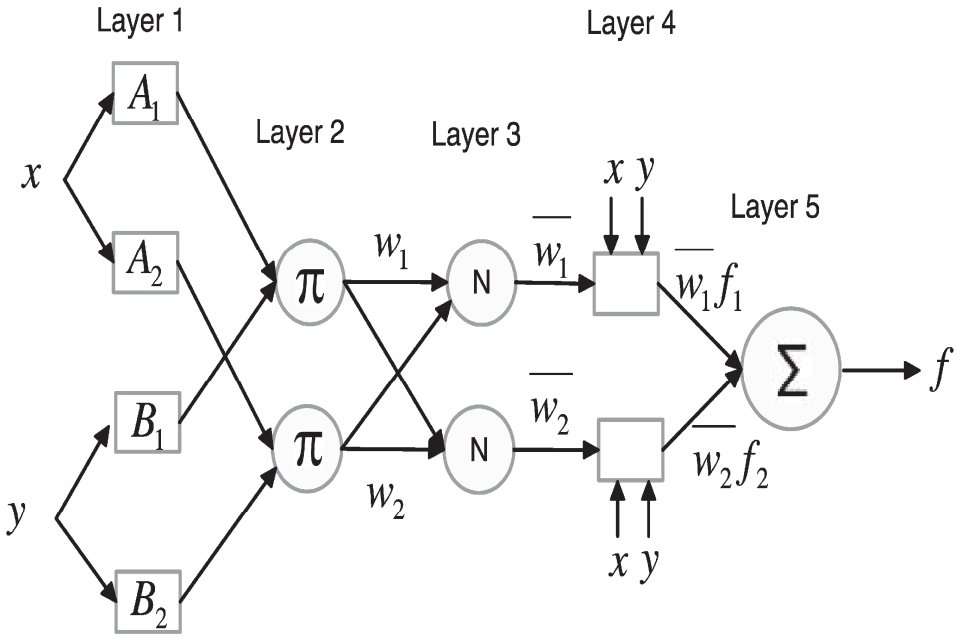

Figure 1.10: ANFIS Five Layer Structure

The description of these five layers of ANFIS is given by [62]:

- **Layer 1**

Layer 1 is known as the Fuzzification layer. In this layer, the fuzzification of all the provided inputs by the user has been done. The form of inputs when it is provided to the membership function is crisp. This layer converts that form of inputs into fuzzy values. In the membership functions, all the nodes of the network reside.

- **Layer 2**

Layer 2 is acknowledged as the Rule layer. 'Π' or the algebraic product is the representation of this layer. The intensity of firing rules according to the provided inputs to the system is evaluated by this rule layer.

- **Layer 3**

The third layer of ANFIS is known as the normalization layer. In this layer, all the nodes are fixed. The representation of the normalization layer has been done by symbol 'N'. This Layer elaborated the normalization occupation to the intensity of firing a rule in the second layer. That is why this layer is known as the normalization layer. In simple wording, the calculation of normalized weights is done in this normalization layer.

- **Layer 4**

This layer is acknowledged by the name of the defuzzification layer. The defuzzification layer comprises several linear operations for each and every input signal. The nodes of this layer are adaptive in nature.

- **Layer 5**

The last layer of the ANFIS is the summation layer. The representation of this layer is by summation sign that is '∑'. The primary operation of this layer is to sum all the signals coming from the preceding layers, and these calculations are responsible for the overall outcome of the system according to the provided inputs.

1.11 SUMMARY

Hepatitis B is a lifestyle disease which is very critical disease and it may result in liver cancer or cirrhosis, if not diagnosed on time. So, it becomes very essential to diagnose this disease during its early stage so that its right treatment can be started on time and life of the patient can be saved. For this purpose, there is a need of expert systems which can assist the physicians in timely diagnosis of this disease. Expert systems are based on the techniques of artificial intelligence (AI). In early days, simple rule-based intelligent systems were developed which do not tell the probability that diagnosis is close to reality. Fuzzy expert systems are also used for the same purpose where the problem is to choose the membership functions and to define the rules. Neural networks are also being used to develop the expert systems. Now a days, to get better results, hybrid techniques like neuro-fuzzy are also being used for developing such expert systems.

Chapter – 2

LITERATURE REVIEW

In this chapter, the work done on Hepatitis B disease by several researchers who used the various methodologies for detection, identification and diagnosis of this deadly disease is elaborated in an effective as well as efficient manner. The researchers figured out numerous gaps and tried to fill those gaps by their research work. Some of the researchers' work that is considered in this thesis work is explained below and also compared with each other in terms of used methodology and performance of developed models in table 1.

Emon (2019) [16] used data mining techniques to classify hepatitis B disease into different classes. The detection of Hepatitis disease is done by using relevant tests, symptoms of the disease as well as the mode of transmission. The author also enlightened that this deadly disease does not only affect the liver, but it also has the ability to infect other sites of the human body. Any age of people can be influenced by this life-threatening disease. The major five viruses that cause hepatitis disease are also mentioned in this paper. These viruses are HEV, HAV, HDV, HBV and HCV. The used techniques are K-star, Naïve Bayes as well as J48 classifier with 10 fold cross-validation. The author utilized the WEKA software to evaluate the outcome of this research work to develop a computer-aided diagnosis system that can detect hepatitis disease with high performance. Additionally, the classification accuracies, error, true positive rate, true negative rate and precision has been observed from these methods and also compared with one another. The accuracy of results generated by the Naïve Bayes algorithm is 93.8%. 98.6% is the classification accuracy given by the J48 classifier. Similarly, the K star algorithm gave the output with a classification accuracy of 97.2%. Therefore, this paper suggested that the J48 classifier has the maximum accuracy among these three methods used in this research paper.

Cheng et al. (2011) [14] discovered several clinical symptoms which cause Hepatitis B disease at the chronic stage. This investigation has been done by using the covariance

network with a mining mutation hotspot. In this research work, the author did one year study on this disease by examining 1694 clones from the 23 patients who are HBeAg positive chronic hepatitis B. These patients were analysed monthly by doing various tests. This study found that several individuals are consistently affected by the hepatitis B virus, and also the host immune response, as well as the serum viral load, is also diversifying from one individual to another. The various representative covariance networks were also identified of every patient group by using the integration of point mutation rule as well as covariance network. A feature tree of clinical biomarkers has been constructed according to the coverage rate, which further assists in the prediction of hepatitis B disease at the early stages. Figure 2.1 displays the systematic overview of the entire research work done by the authors.

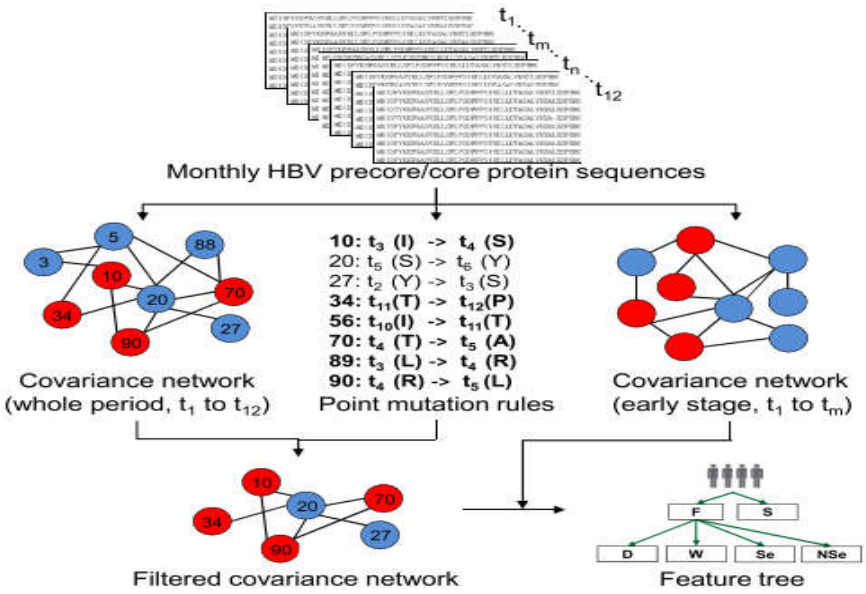

Figure 2.1: Systematic Overview of Research Work [14]

Leung et al. (2011) [18] classified the hepatitis B DNA by using data mining techniques by identifying the most difficult makers that are a genomic marker in HBV, which are correlated with liver cancer. This investigation has been done by equating the entire HBV genomic sequences of patients which were having HCC and the patients who are not having HCC. In this research work, the author introduced various data mining techniques such as classifier learning, molecular evolution analysis,

classification, feature selection and clustering. The required information is acquired from the large dataset with includes approximately 200 patients having either genotype C or B. In the clustering as well as molecular evolution analysis, the identification of three subgroups in genotype C has been made, and these subgroups were categorized by using the clustering method. The next process used in this study was feature selection. In this, the markers were selected on the basis of information gain. After that, based on the evolutionary algorithm, the Rule learning algorithm was used to generate the accurate and correct rules. In the next step, the classification has been done by using fuzzy measures, which provides better performance. Hence, by using this methodology, the author succeeded to classify the infected patients and non-infected patients with 70% accuracy and 80% sensitivity.

Mahmudy (2019) [20] used the Hierarchical k-Mean clustering to enhance the efficiency of the classical approach of clustering by using K-means. The methodology used by the author in his research work assisted in determining the initial cluster center on the basis of the result of the mean provided by the hierarchical clustering method. The data of the DNA sequence of hepatitis B has been taken, and according to that data set, the clustering has been done. The utilized dataset includes the DNA sequence of HBx in genotype A to G. However, the genotype E was not included in it. Additionally, the dataset also comprised of four different kinds of nucleotides which are Cytosine, Adenine, Guanine and Thymine. The hierarchical clustering method used in this study by the researcher has four steps. In the very first step, the input is provided to the system of the object. The second step is all about normalization by utilizing the scale method. Further, the dissimilarities, as well as similarities among all the objects pairs, were computed in the 3rd phase of the process. In the last step, the linkage is applied to the group objects into the hierarchical clustering. This method also requires more time because of the high computation for the massive amount of data. Hence, the specific method is only concentrated to generate the initial cluster centre in the k-means algorithm. The experiment result of this research work observed that the efficiency of hierarchical K-means clustering is better than traditional k-means.

Wang & Liu (2017) [23] developed a system that assists in identifying the various Hepatitis B's risk factors, which reactivate even after the radiotherapy of liver cancer of a patient. This system is proposed with the help of random forest and Bayesian classification. The developed system is basically a prediction model which predicts the reactivation of risk factors in the patient's body. This model assists the doctors or experts to analyse the factors which are risky and that will lead to a decrease in the disease incidence. In this research study, firstly, the author develops a random forest method in order to choose primary features as well as to build a classification prediction model by using the selected primary subset. The selected features are sorted with respect to their significance. In this particular work, the 5 important features were selected, which further merged into the latest feature subset. By using this new subset, the author built the Bayesian classification and random forest model. The result of this work done is that the proposed prediction model is the best system to figure out the classification prediction problem. The accuracy with which the developed model did classification with 5 fold cross-validation is 85.15%. In contrast, if the model used 10 fold cross-validation, then this classification accuracy was reached 84.57 percent. Similarly, if the value of k in k cross-validation is considered as 3 then the measured classification accuracy was 83.83 percent.

Chen (2012) [24] constructed a discriminant diagnosis model. The development of this model has been done by first selecting the attributes, decision tree C5.0 algorithms as well as then discrimination analysis. In this proposed model, two phases are there. In the first phase, the required attributes are selected from a large dataset of attributes. To gather these attributes, the decision tree algorithm has been used. The logical diagram of the process which has been used to identify the critical attributes is demonstrated in figure 2.2.

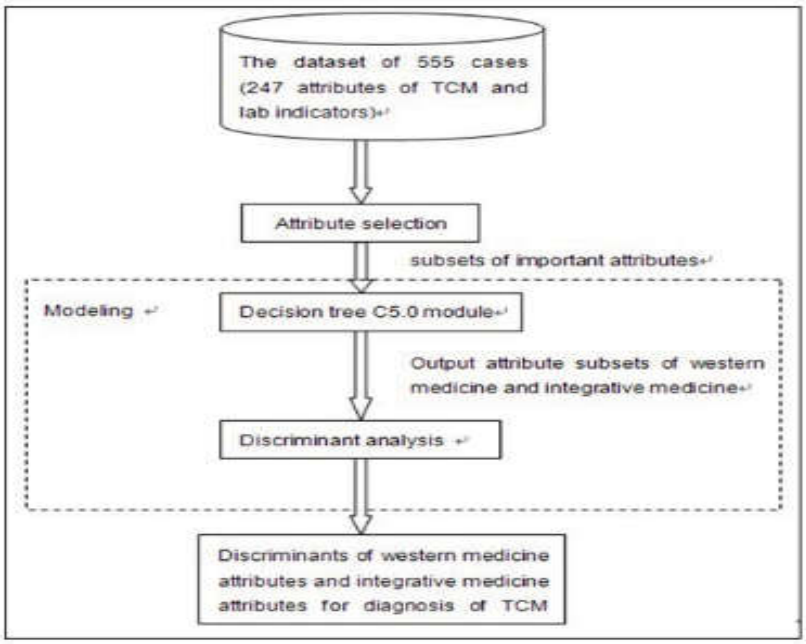

Figure 2.2: Logical Diagram for Feature Identification [24]

In the second phase, the syndromes are acquired, and this phase is also called as the modelling phase. The developed model used the two kinds of inputs first lab indicators and second clinical symptoms. The classification accuracy was also compared by the researcher with different methodologies and by using the different inputs to the proposed system. It is observed that the accuracy with which the decision tree and discriminant analysis classified the disease is 96.94 and 94.4%, respectively. These accuracies were considered as roughly the same by the researcher.

Kiruthika (2014) [26] proposed an expert system to diagnose the Hepatitis B by using the Generalized Regression Neural Network. This expert system did the classification of the patients who are suffering from this deadly disease and those who are not. The two classes are infected patients and non-infected patients. If the system shows the infected patient, then the system will also evaluate how much severe the disease is according to the intensity rate. This paper effectively explained the generalized regression neural network with its diagram of each layer (figure 2.3) and in the mathematical form.

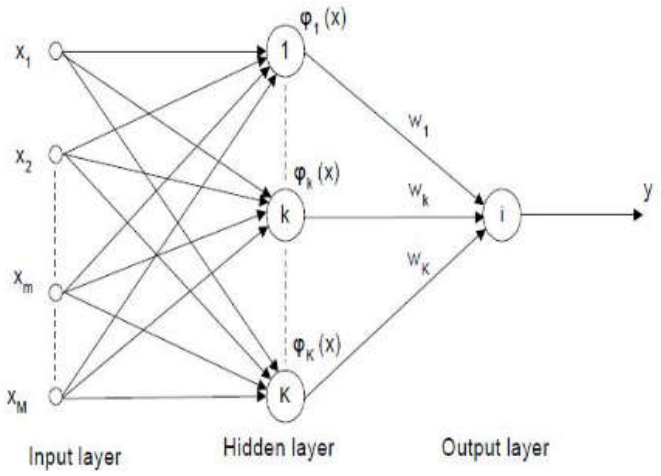

Figure 2.3: Architecture of Generalized Regression Neural Network [26]

In this paper, the author enlightened that the transmission modes of the hepatitis B virus are identical to HIV. Additionally, this virus can also survive for at least a week, even if it is outside the human body. In this time period, the virus can also infect individuals who are not properly vaccinated. The graphical user interface also constructed by the researcher for the developed expert system.

Rouhani (2009) [28] diagnosed Hepatitis B by using the Support Vector machine as well as neural networks. The proposed network in this research work assists in classifying diseased patients into six diverse classes. These different classes are two phases of hepatitis B, two phases of hepatitis C, non-viral hepatitis and no hepatitis. To achieve the decided objective, the researchers developed several networks in this work. These networks were Radial Basis Function, Generalized Regression Neural Network, probabilistic neural network, learning vector quantization and support vector machine. These distinct five networks were trained to detect hepatitis disease. The evaluation of whether a particular patient is infected by HBV or not and also which kind of hepatitis that specific patient is suffering from was done on the basis of clinical examination. The dataset utilized in this research study is comprised of 250 distinct cases, that were evaluated by the experts carefully. The performance of these five networks is also compared on the basis of classification accuracy. It is observed that the radial basis function generate more accurate results as compared to SVM, GRNN, PNN and LVQ. The accuracy for the RBF network is 96.4% for testing as well as training of the dataset.

Uttreshwar (2009) [27] presented a predicted model by using the logical inference with the GRNN (Generalized Regression Neural Networks) to identify the Hepatitis B disease. An expert system is proposed on the basis of logical inference. The logical inference assisted in making a decision about for a patient what kind of hepatitis is probable to occur as well as whether the disease is HBV or not. Further predictions with respect to this disease have been made by an artificial neural network. The GRNN is utilized for predictions by using the dataset of hepatitis disease. The dataset has been provided to the system which was used for the training phase. The used artificial intelligence technology assist the system in the prediction of Hepatitis B disease and give the outcome with the severity of the disease. An effective graphical user interface is also developed by the author, which makes the efficient use of this expert system more easy. The author also presented that the developed expert system for the diagnosis of HBV has high accuracy in classifying the infected as well as non-infected patients. Figure 2.4 represented the flow diagram of diagnosis that is used by the researcher in this research work.

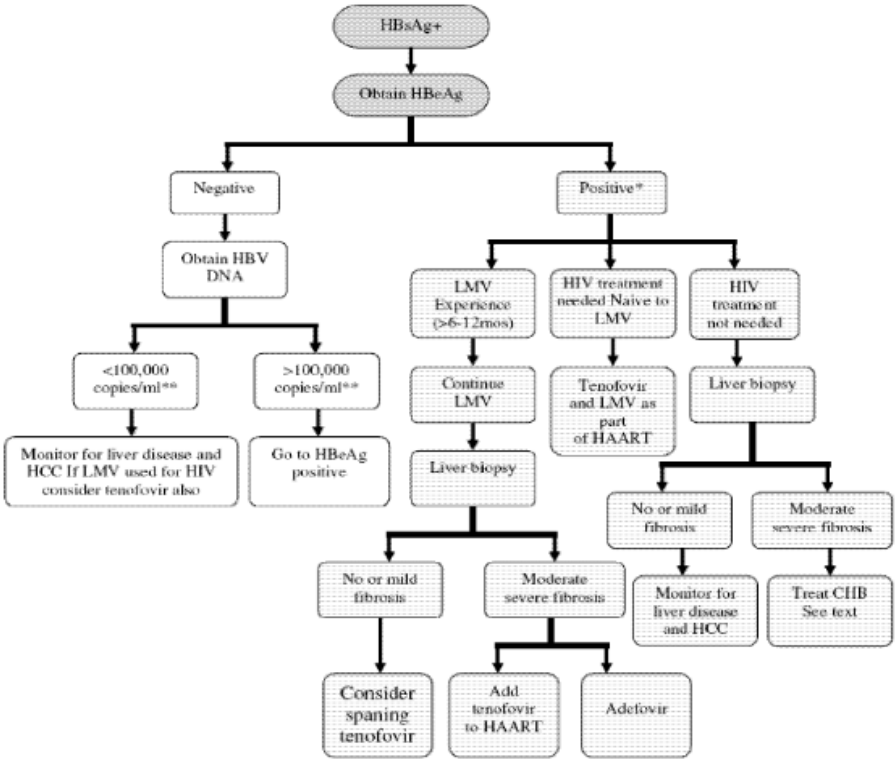

Figure 2.4: Flow Diagram for Diagnosis [27]

Uhmn, Kim, & Kim (2007) [25] utilized the machine techniques such as support vector machine, decision tree and decision rule to predict the harm done by liver diseases, chronic hepatitis from Single nucleotide polymorphism (SNP). These techniques are used to predict liver disease, i.e., chronic stage of hepatitis. Additionally, to acquire better results from the system, the backtracking technique was used for the selection of required features from the feature dataset. Also, a comparison is carried out between these three methodologies that are used in this research work. From the obtained results, it is obtained that the decision rule can identify chronic hepatitis effectively with an accuracy of 73.20%, whereas the SVM as well as decision tree have the least accuracy to classify these diseases as compare to the decision rule with an accuracy of 67.53% and 72.68% respectively. The paper also showed that the decision rule, as well as the decision tree, are powerful tools that can be used for the prediction of susceptibility from SNP data to chronic hepatitis.

Areghan and Konyeha (2019) [63] designed a system to diagnose hepatitis B disease by using genetic neural network methodology. The clinical symptoms were used by the researchers as input in the developed diagnostic system. Additionally, the UGC repository has been used to acquire the dataset, which further utilized in the training phase of the system. The designed diagnostic system integrated neural network as well as genetic algorithm; hence it is a hybrid system. It also helps to eradicate the problems that the latest diagnostic systems are facing. The selection of features, as well as extraction of the dataset, has been done by the genetic algorithm before the acquired data utilized in the training phase. It also assisted in the optimization of the dataset by doing feature section and extraction tasks. Further, the multilayer perception neural network was also utilized in the system training. The used network is the feed-forward neural network. This paper explained all the working of the developed system effectively and appropriately with all mathematical functions. Additionally, the number of GUI designed for each and every operation by the researchers is also shown in this particular research work. The performance of the designed diagnostic system to diagnose the Hepatitis B disease has been measured on the basis of classification accuracy, and the observed accuracy is 99.14%.

Thakur (2018) [64] developed an expert system to diagnose the Hepatitis B virus by using fuzzy base IF-THEN rules. This paper stated that there is an extreme inadequacy

of knowledge among the individuals except experts or professionals about causes of disease, symptoms of the disease, as well as remedial therapies that can be used to diagnose this life-threatening disease. In this research work, the primary aim of the author is to recognize several HBV genetic markers with their clinical information. This information was further utilized to detect if a patient is suffering from the infection of this virus and need therapy or not. The designed expert system for the detection of HBV generates the output in terms of possibility. Figure 2.5 displays the block diagram of the designed expert model.

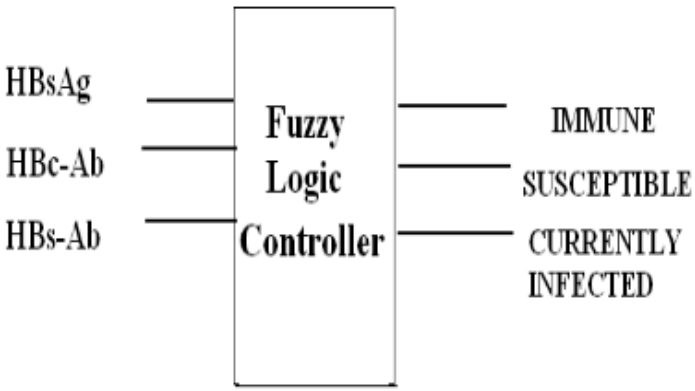

Figure 2.5: Block Diagram of a Developed Expert System [64]

The developed system is a powerful tool that can be used even in the absence of an expert or specialist to predict the probability of a patient being infected by the hepatitis B virus. The various symptoms of this disease, such as HBsAb, HBsAg and HBcAb, are used as the input variables to the system and corresponding to these inputs, the system will give the output. The output variables used in the system are immune, susceptible and currently infected.

Chen et al. (2017) [65] developed a system that assisted in making a decision by using the four different classical classifiers. The implemented system helps to enhance the performance for the diagnosis of hepatitis B. The classifiers that have been utilized by the researchers in this work are Random Forest, SVM, K-Nearest Neighbor and Naïve Bayes. The input fed to the system was RTE pictures of the liver. The required features are extracted from these images, and according to those attributes, the detection, as well as diagnosis of hepatitis B disease, has been done. The methodology used in this

research work has 4 major phases. In the very 1st step, the colored images were acquired by using the RTE equipment. In the second step, 11 features were extracted from the respective images, and each parameter has a distinct range. Further, in the next step, i.e., the 3rd step, the data processing has been done on the 11 extracted features in the previous step plus 2 features of basic information about the patient that is sex and age. The implementation of prediction was done in the last step. The classifiers were trained as well as tested by using the given dataset. Moreover, all the classifiers are compared with each other, and according to the obtained result, the accuracy of the random forest classifier is more as compared to other classifiers.

After figuring out the limitations in the prediction of cirrhosis as well as hepatic fibrosis, Wei et al. (2018) [66] used ML techniques such as gradient boosting (GB), decision tree (DT) and radial basis (RB) for its detection. The developed system has mainly four steps, as shown in figure 2.6.

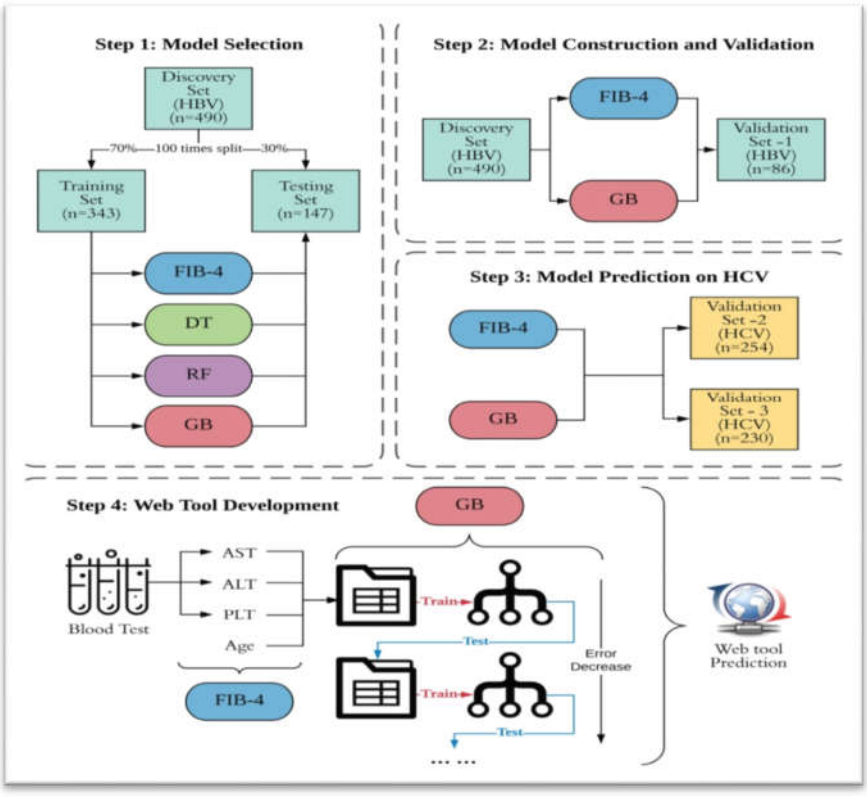

Figure 2.6: Flow Chart of the Developed System [66]

Step 1 is model selection. In this step, the researchers did testing as well as training by splitting the respective dataset 100 times. The used methodologies, i.e., decision tree, redial basis and gradient boosting, were trained in this step and compared the results of them with FIB-4. 70 percent of the given data was utilized in the training phase, whereas the 30% of the dataset was considered as the testing dataset. Step 2 is model construction as well as validations. In this step, the final model has been constructed and again compared with FIB-4. Additionally, the system validated on the hepatitis B virus validation set. Model prediction on HCV is the next step. In this step, the risk was predicted for HCV by using the FIB-4 and GB model. The last step is the Web tool development. In this step, a user-friendly web tool developed by the researchers for the experts or professionals. After comparing all models, it is observed that the GB model has more accurate as well as adequate result than other models. This research work also assisted in enhancing the specificity and sensitivity of the classification methods used in the existing system.

Akbari (2018) [67] used the non-linear iterative partial least square method to make the reduction of data, and the clustering task has been done by using the self-organizing map and a hybrid system, i.e., neuro-fuzzy inference system for the detection of HBV disease. The required and important attributes for the prediction of this disease have been extracted from the selected dataset by using the decision tree. The author used a real-world dataset to perform this research work. Additionally, the comprehensive review of research work done by other authors is also well explained in this paper. The comparison of the developed methodology is also made with the existing techniques. The result of this comparison stated that the performance of the designed model is accurate than the performances provided by other models such as SVM, neural network, k-nearest neighbour and ANFIS. This paper systematically illustrated the entire methodology by using neat and clean diagrams, flowcharts and tables. The accuracy of the proposed model was 93.06%, and this was calculated by using Receiver Operating Characteristic (ROC).

Ali, Bilal and Saleem (2018) [68] presented the investigation of hepatitis B virus disease in the blood serum of the human body by using the combination of pattern recognition technique and Raman spectroscopy. In this scrutiny, the dataset was

gathered from the Pakistan Atomic Energy Commission general hospital consist of approximately 119 positive patients of HBV disease. The classification of infected as well as non-infected patients has been done by using a support vector machine. The two distinct kernels used with SVM are Gaussian RBF and Polynomial function. Additionally, the performance of the system was also evaluated with the different problem of optimization that is the least square and quadratic programming. In this research work, the polynomial kernel of order 2 achieved the best performance when contrasted with others. The accuracy, precision, sensitivity and specificity for the diagnosis of hepatitis B disease of the developed model are 98%, 97%, 100% and 95%, respectively.

Ye et al. (2003) [69] constructed a molecular signature for the very first time by using the machine learning algorithm. It assisted in the classification of patients suffering from metastatic hepatocellular carcinoma and also helped in the identification of relevant genes to metastasis. The tree view and cluster software was used by the researchers to do unsupervised hierarchical clustering analysis. The used dataset has been collected from Liver Cancer Institute and Zhongsham Hospital. It includes 107 paired samples of patients. The primary objective of the research was to recognize the genes that categorize primary tumors and intrahepatic metastatic lesions. However, during this research work, the researchers found that these lesions were similar to primary tumors, despite patient age, size of tumor and encapsulation. Additionally, the supervised machine learning algorithm, i.e. compound co-variate predictor, was used to define the set of genes correctly as well as accurately. Researchers applied this classification algorithm to a dataset that includes 50 samples of HCC. But, again, the researchers figured out that there is no any major difference between the matched metastatic lesion and primary HCCs.

A comprehensive survey is presented by Alan et al. (2019) [70] in which the diversity of hepatitis B virus in the population of Asia and Europe is explained in the most effective manner. Also, the development of a model is being done by using the machine learning approach. This model aid in the identification of viral variants, which further do the classification of HBeAg status accurately. The novel patterns of these viral variants have been identified, which are related to the status of HBeAg. This research

work used two datasets named as dataset A and dataset B. Dataset A has been gathered from Erasmus Medical Centre, Rotterdam, The Netherlands, and it consists of plasma samples. Similarly, dataset B was collected from the Ethic Committee of Xiamen Centre for Disease Control and Prevention, and this dataset consists of samples of infected patients from chronic HBV. The respective work fails to make the inference about the methodology as what the developed system means corresponding to the outcome of patients. Additionally, this system only works for the respective dataset that is used for training as well as testing of the model. However, the system classifies the dataset with high accuracy. Many researchers has developed their models for diagnosis of Hepatitis B. A comparison of all these existing models is represented in table 2.1.

Table 2.1: Comparison of Existing Models Developed by Different Researchers for Diagnosis of Hepatitis B Virus

Authors	Year	Methodology	Input variables	Output Variables	Remarks
Emon [16]	2019	K- star, Naïve Bayes and J48 algorithm	Clinical symptoms, laboratory tests and transmission modes	Detection of different viruses, i.e. HEV, HAV, HDV, HBV and HCV	J48 algorithm gave accurate output than the other three used methods with 98.6% accuracy.
Cheng et al. [14]	2011	Covariance neural network	Symptoms of HBV	Detection of hepatitis B disease	Prediction of hepatitis B disease at the early stages successfully done by the researchers
Leung et al. [18]	2011	Data mining techniques	Markers of Genotype B and genotype C	Classified the HBV patient and non- HBV patients	Classification of infected patients and non-infected patients with 70% accuracy and 80% sensitivity
Mahmudy [20]	2019	Hierarchical k-Mean clustering	Four different kinds of nucleotides are Cytosine, Adenine, Guanine and Thymine	Generate the initial cluster centre in the k-means algorithm	The efficiency of hierarchical K-means clustering is better than traditional k-means

Authors	Year	Methodology	Input variables	Output Variables	Remarks
Wang & Liu [23]	2017	Random Forest and Bayesian classification	Five features: HBV DNA levels, Tumor stage TNM, V10, V20 and outer margin of radiotherapy	Classification of HBV stages	The accuracy of the developed model is 85.15% with five-fold cross-validation.
Chen [24]	2012	Decision tree C5.0 algorithm and discriminant analysis	Attributes from the dataset	Classification of HBV disease	Decision tree and discriminant analysis classified the disease with accuracy 96.94% and 94.4%, respectively
Kiruthika [26]	2014	Generalized Regression Neural Network	Disease symptoms	Classification of infected patients and non-infected patients	Successfully proposed an expert system to diagnose HBV.
Rouhani [28]	2009	RBF, Generalized Regression Neural Network, probabilistic neural network, learning vector quantization and SVM	Features from patients dataset	Classified diseased patients into six diverse classes, i.e. two phases of hepatitis B, two phases of hepatitis C, non-viral hepatitis and no hepatitis	The accuracy for the RBF network is 96.4% and is more accurate than other approaches

Authors	Year	Methodology	Input variables	Output Variables	Remarks
Uttreshwar [27]	2009	Logical inference with the GRNN	HBV Disease symptoms	Prediction of HBV	Developed expert system to diagnose HBV with high accuracy to classify the infected and non-infected patients
Uhmn, Kim, & Kim [25]	2007	Support vector machine, decision tree and decision rule	Features from the patient	Prediction of chronic stages of liver disease	A decision rule is a powerful tool that can be used for the prediction of hepatitis with an accuracy of 73.20%
Areghan and Konyeha [63]	2019	Genetic neural network	Clinical Symptoms of the disease	Diagnosis of Hepatitis B disease	The classification accuracy of this developed model is 99.14%
Thakur [64]	2018	Fuzzy logic	HBV disease symptoms: HBsAb, HBsAg and HBcAb	Classified health of patient into three categories: immune, susceptible and currently infected	Developed an expert system that can be used even in the absence of a specialist.

Authors	Year	Methodology	Input variables	Output Variables	Remarks
Chen et al. [65]	2017	Random Forest, SVM, K-Nearest Neighbour and Naïve Bayes	RTE pictures of liver	Detection of hepatitis B disease	The accuracy of random forest classifier is more as compared to other classifiers
Wei et al. [66]	2018	Gradient boosting (GB), decision tree (DT) and radial basis (RB)	Features from the acquired dataset	Prediction of HCV	Effectively developed a model which assists in the detection of HCV and also the designed user interface makes it easy to use by experts as well as others.
Akbari [67]	2018	Self-organizing map and neuro-fuzzy inference system	Features from the acquired dataset	Detection of HBV	The accuracy of the proposed model was 93.06%, and this was calculated by using Receiver Operating Characteristic (ROC)

Authors	Year	Methodology	Input variables	Output Variables	Remarks
Ali, Bilal and Saleem [68]	2018	Pattern recognition technique	Symptoms of infected patients	Detection of HBV	The accuracy, precision, sensitivity and specificity to diagnose HBV disease by using the developed model are 98%, 97%, 100% and 95%, respectively
Ye et al. [69]	2003	Unsupervised hierarchical clustering	Selected attributes from the gathered datasets	Difference between the metastatic lesion and primary HCCs	There is no any major difference between the matched metastatic lesion and primary HCCs
Alan et al. [70]	2019	Machine learning approach	Patterns of viral variant	Classification of HBeAg status	The identification of viral variants which further do the classification of HBeAg status accurately

2.1 SUMMARY

In this chapter, we have seen that different researchers have implemented different techniques for developing the expert systems to diagnose HBV diseases. The popular used techniques are Covariance neural network, Hierarchical k-Mean clustering, Random Forest, Bayesian classification, Decision tree, Generalized and Probabilistic regression neural network, Support vector machine (SVM), Fuzzy logic, Gradient boosting (GB), Radial basis (RB), Self-organizing map, Neuro-fuzzy etc. The different researchers has taken different parameters and has obtained different results. In the literature survey, the authors has generated membership functions and rules from the data set manually. In the proposed work, the rules and membership functions will be generated automatically in the training phase for the diagnosis of Hepatitis B. It is expected that the results of proposed systems will be better in terms of performance parameters such as accuracy, precision, sensitivity and specificity.

Chapter – 3

PROBLEM FORMULATION AND OBJECTIVES

In this chapter, we will discuss the basis behind problem formulation and will define very crisp objectives of the proposed research work.

3.1 PROBLEM FORMULATION

Hepatitis B disease seized the attention from the past few years at an extensive level because the infection spread by this disease increases the probability of occurrence of liver cancer as well as more liver diseases, which are more dangerous and deadly to the health of a human. This disease is preventable as well as treatable by using the proper vaccine, but till now, there is no any appropriate cure for this disease, although researchers are on their keen efforts [71]. Several antiviral therapies can be utilized to control this deadly disease. These therapies assist in diminishing some complications of it for the long term. However, this is not an effective cure. The researchers are identifying various methods and techniques in order to eliminate the hepatitis B virus from all around the world. Thus, these approaches help in the identification, diagnosis and treatment of this virus to prevent further expansion. A number of researchers developed different strategies that reduce or kill the infected cells from the infected patient's body [72]. Additionally, there is also a need for awareness about hepatitis B disease to the individuals as the infected patients are unfamiliar with this deadly disease. If the patients do not know about the disease, then patients are definitely unable to take appropriate prevention as well as treatment [73].

The treatment of this infection is imperative, as the number of people who are suffering from this disease is approximately 240 million across the world [6]. The number of infected patients is increasing rapidly, and it is found that in a year, about one million deaths are happening due to liver cancer as well as liver failure globally. Moreover, 40% of liver cancer is being caused by the HBV and become the second most life-threatening disease worldwide [74]. By using the latest and advanced technology, the

researcher might invent a single drug that can be used as a permanent cure for hepatitis B disease [75]. However, before giving the prescription of that cure, it is crucial to identify the disease at its introductory stage. Hence, a strategic methodology must be created after understanding the biology of the disease accurately and using the existing resources and tools effectively that assist the experts and specialists to identify, prevent as well as cure this harmful disease [76].

Therefore, the work done in this research is all about evaluation as well as the organization of infected patients from the Hepatitis B virus and non-infected patients in the accurate class. This classification is done by giving the biomarkers, symptoms and laboratory test report as the input to the developed models. This research work is also concentrated on the diagnosis of the Hepatitis B virus before it reaches the chronic stage. In this, intelligent systems are proposed by using the adaptive neuro-fuzzy technique and fuzzy logic, which gives accurate and effective results for the identification of the Hepatitis B virus.

3.2 OBJECTIVES OF THE STUDY

The objectives of the presented study are written below:

1) To study and review various existing intelligent medical diagnostic systems.

2) To reduce the dimensions of the dataset collected from hospitals using data-reduction technique and to apply fuzzy model based on data set.

3) To develop an intelligent medical diagnostic system for Hepatitis B disease using the neuro-fuzzy algorithm.

4) To compare the results in terms of performance parameters obtained from the proposed systems with existing systems.

3.3 SUMMARY

Hepatitis B is a kind of lifestyle disease which is spreading at an alarming rate. Approximately 240 million people across the globe are suffering from this disease. This

disease is treatable if it is diagnosed during its early stage, thus the early diagnosis of this disease is very essential to save the lives of the patients. The proposed research work concentrates on timely diagnosis of Hepatitis B before it reaches the chronic stage. For this, an intelligent system is proposed using the Fuzzy logic and adaptive neuro-fuzzy technique.

Chapter – 4

DIMENSIONALITY REDUCTION

As we know that the available data sets are very exhaustive consisting of many rows and columns. The parameters are represented in columns and the records of the patients are represented in rows. There may be few parameters which do not contribute much in the result and such parameters can be removed from the available dataset. So, there is need of data reduction. In this chapter, we will discuss data reduction using principal component analysis (PCA).

4.1 WHAT IS DIMENSIONALITY REDUCTION?

In the era of advanced technology, massive amounts of data are gathered as well as produced. But, having a huge dataset in the case of machine learning is considered as a bad thing. The dimensions or extra features are diminished till the data is generalized and produce better accuracy. Hence, the curse of dimensionality is having a huge number of features in the dataset [77].

The dimensionality reduction is a procedure in which the considered various random variables are reduced from a dataset [78]. In other words, the manner by utilizing which the complexity of a model is reduced and also over fitting is avoided is known as dimensionality reduction. The two main parts of this process are the selection of features and extraction of features [79].

In this research work, an algorithm that is used to reduce the considered data into the lower dimension feature subspace is Principal Component Analysis (PCA). The main objective of doing this is to maintain only that information that is important and relevant for a particular task.

4.2 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis is an unsupervised learning technique that is massively utilized in the distinct domains for the dimensionality reduction as well as extraction of required features [80]. The idea behind the principal component analysis algorithm is data dimensionality reduction as well as maintaining the required information as much as possible [81].

Principal component analysis assists in recognizing the patterns from the dataset on the basis of the relationship among the features. The main intent of this algorithm is to figure out the maximum variance direction in the high-dimensional dataset as well as proposed that particular dataset into a new subspace having fewer or equal dimensions as the original dataset had. There is also a constraint that the new feature axes must be orthogonal to one another, as shown in figure 4.1. In this figure, X_1 and X_2 are feature axes that are original and PC1, and PC2 are orthogonal axes or principal components. The result of applying PCA is the transformation of original data of d-dimensional into the new subspace having k-dimensions and mostly the $k << d$.

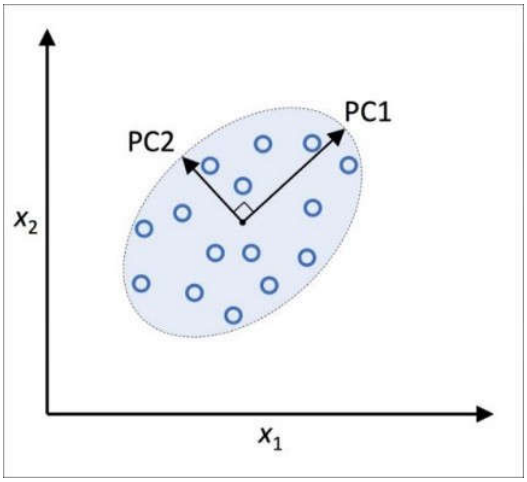

Figure 4.1: Orthogonal Principal Components [80]

The steps that have to be followed while applying the principal component analysis for the dimensionality reduction are illustrated below:

1) The first step is to standardize the considered dataset of d-dimension.

2) Now, create as well as compose the covariance matrix.

3) In this step, the decomposition of the constructed covariance matrix in the above step is being done into the eign values as well as eign vectors.

4) In order to rank the equivalent eign vector, sort the figured out eign values in the descending order.

5) Choose k eign vectors that are equivalent to the k largest eign values. Here, the "k" is the new feature subspace dimensions.

6) Create a matrix, say was the projection matrix from the k eign vectors which are on the top.

7) Convert the input data of d-dimensions into the new feature subspace having k-dimensions by using the projection matrix W.

4.3 PRINCIPAL COMPONENT ANALYSIS (PCA) ON CONSIDERED DATASET

In this research work, the used dataset is acquired from the hospitals as well as clinical trials including 11 different features of infected and non-infected people. The features in the original dataset are HBsAg, Anti-HBs, Anti-HBc, HBV DNA, Anti-HBcAg-IgM, Achy muscles, Jaundice, Dark Urine, Abdominal Pain, Vomiting and lack of energy. Minitab 14 software is used to apply the principal component analysis on the dataset.

The number of steps that are followed in this research work to reduce the dimensionality of the considered dataset is given by:

Step 1: Upload the original data

In this step, the original data that is considered for analysis is uploaded to the Minitab software. The columns of data must be representing the distinct calculations. The software removes those columns automatically, which has missing values.

Step 2: Enter the number of components to calculate

In this step, enter the number of principal components in the displayed textbox. If an individual does not know about the number of principal components that should be calculated, then leave this textbox blank.

Step 3: Choose the graphs

In this step, select the graphs that are required to analyze the result from applied principal component analysis. In this research work, the scree plot and loading graph are selected to represent the data and analyze the outcome.

Step 4: Evaluate the number of principal components by using the eign values and scree plot.

Table 4.1 and 4.2 represent the eigenanalysis of the correlation matrix and eign vectors, respectively. In figure 4.2, the scree plot is displayed.

Step 5:

Now, in terms of original values, every principal component is interpreted. Figure 4.3 represented the loading graph of the data. It can be evaluated that the two features that are Achy muscles and lack of energy, have large negative loadings as compare to other features. Hence, the new dataset that is obtained from the result given by principal component analysis has only nine features that are HBsAg, Anti-HBs, Anti-HBc, HBV DNA, Anti-HBcAg-IgM, Jaundice, Dark Urine, Abdominal Pain, and Vomiting.

Table 4.1: Eigenanalysis

Eigen value	3.1482	1.5791	1.2610	1.2399	0.9655	0.8487	0.7352	0.6248	0.5069	0.0647	0.0260
Proportion	0.286	0.144	0.115	0.113	0.088	0.077	0.067	0.057	0.046	0.006	0.002
Cumulative	0.286	0.430	0.544	0.657	0.745	0.822	0.889	0.946	0.992	0.998	1.000

Table 4.2: Eigenvectors of Correlation Matrix

Variable	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**	**PC11**
HBsAg	0.534	0.163	-0.082	-0.058	0.072	0.025	-0.020	-0.003	0.117	-0.686	0.433
Anti-HBs	0.544	0.117	-0.085	-0.076	0.095	0.037	0.010	0.011	0.082	-0.020	-0.812
Anti-HBc	0.165	-0.012	0.402	-0.543	-0.315	0.036	-0.363	-0.264	-0.460	0.042	0.027
HBV DNA	0.543	0.063	-0.097	0.013	0.060	0.053	-0.005	0.053	0.118	0.720	0.386
Anti-HBcAg-IgM	0.065	0.370	0.029	0.531	0.095	-0.526	-0.362	-0.013	-0.399	0.015	-0.014
Achy muscles	-0.085	0.247	0.692	-0.090	0.228	-0.158	-0.173	0.144	0.562	0.030	-0.010
Jaundice	0.075	-0.281	0.179	0.499	-0.035	0.627	-0.484	0.031	0.031	-0.054	-0.032
Dark Urine	0.189	-0.128	0.541	0.268	0.128	0.081	0.654	-0.016	-0.362	-0.024	0.017
Abdominal Pain	0.147	-0.545	0.048	-0.073	-0.134	-0.349	-0.113	0.716	-0.068	-0.057	-0.007
Vomiting	0.139	-0.540	0.041	0.166	-0.095	-0.413	-0.024	-0.620	0.308	-0.015	-0.012
Lack of energy	-0.053	-0.271	-0.076	-0.228	0.882	0.017	-0.179	-0.089	-0.217	0.005	0.034

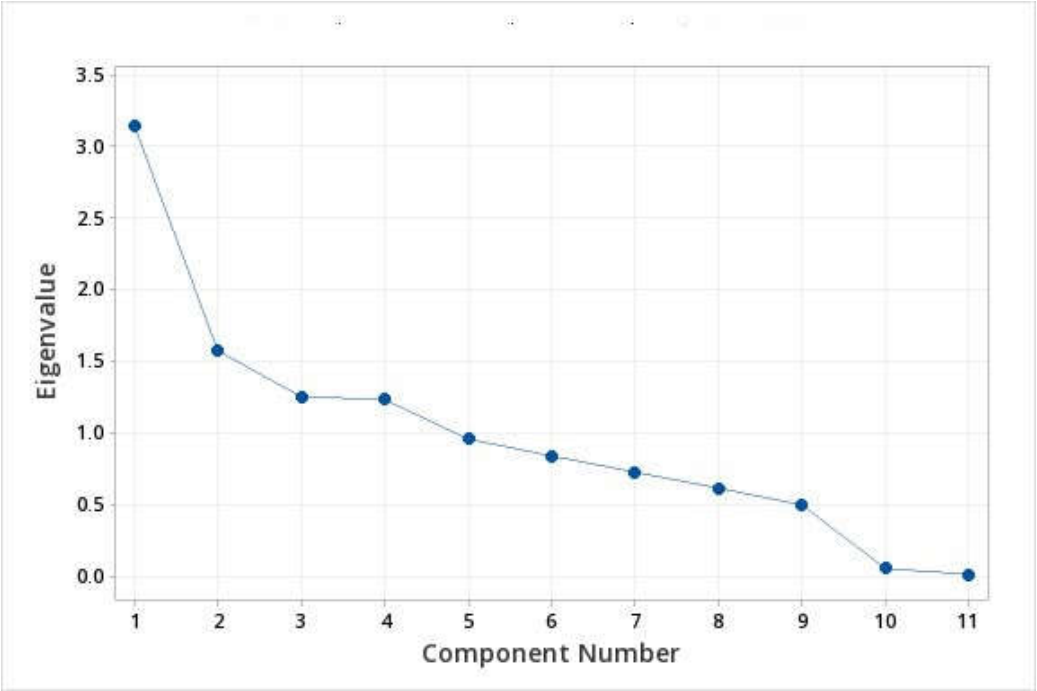

Figure 4.2: Scree Plot

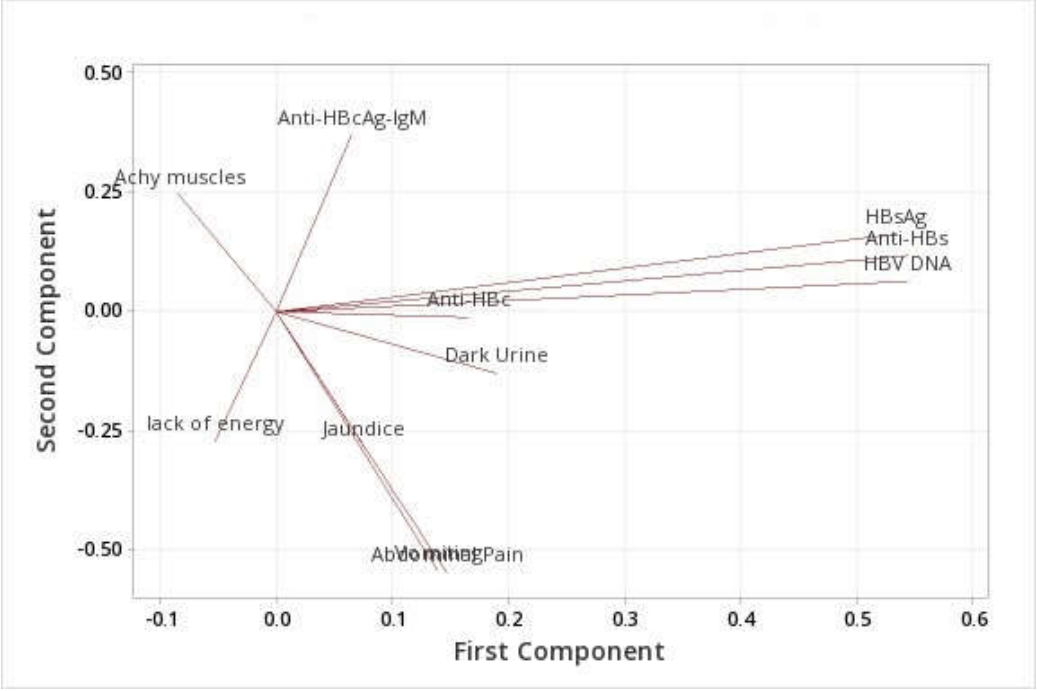

Figure 4.3: Loading Graph

4.4 SUMMARY

Principal component analysis (PCA) is used for data reduction. The main objective of using PCA is to hold only that information which is important and relevant for this research work. Data reduction is done using the five steps of PCA. There are total 11 features in the original dataset and they are HBsAg, Anti-HBs, Anti-HBc, HBV DNA, Anti-HBcAg-IgM, Achy muscles, Jaundice, Dark Urine, Abdominal Pain, Vomiting and lack of energy, After applying PCA, two features namely achy muscles and lack of energy are found to be irrelevant and hence are removed from the actual data set to get better accuracy.

Chapter – 5

EXISTING SYSTEMS

Different researchers has porposed or implemented various systems for the daignosis of Hepatitis using different techniques. In this chapter, we will discuss few major existing systmes for the daignosis of Hepatitis.

5.1 USE OF SVM, DECISION TREE AND DECISION RULE

Uhmn, Kim, & Kim (2007) depyoyed support vector machine (SVM), decision tree and decision rule to predict the chronic stage of hepatitis. SVM is one of the important and powerful learning systems which can be used to solve the problems related to non-linear classification, density estimation and function estimation. SVM separates hyperplane between two classes and then finds an optimal maximal margin. The main objective here is to maximize the margin and reduce the amount of error as much as possible. Here, a non-linear function is used to map the data to a high dimensional feature space. When given a training set, SVM makes a model which is applied to test set. This model maps the input either to positive class (+1) or to negative class(-1) output value. Due to few reasons, decision trees are also widely used for such type of applications. These reasons include:

1) Decision trees has intuitive representation due to which humans can assimilate the resulting model easily.
2) There is no need of parameter setting by the user.
3) It is possible to construct the decision trees faster than other methods.
4) Both, numerical and non-numerical values are allowed for feature values.

To get better results, selection of the required features from the data set was done using backtracking technique. After the comparision of results, it was found that the accuracy to classify the disease using decision rule, SVM and decision tree was 73.20%, 67.53%

and 72.68% respectively. The major observation from the comparision is that decision rule and the decision tree are powerful tools for the prediction of chronic hepatitis.

5.2 USE OF RANDOM FOREST AND BAYESIAN CLASSIFICATION

Wang & Liu (2017) created a system to identify various Hepatitis B's risk factors. Random forest and Bayesian classification techniques are used for developig the prediction model which predicts the reactivation of risk factors in the patient's body. This model is very much helpful for the doctors to analyze risky factors which will result in decrease of disease incidence. Firstly, random forest method is developed for choosing primary features and building a classification prediction model. Selected primary subset is used to build this classification model. The sorting of the selected feactures is done on the basis of their significance. The researcher has selected five important features which were merged to form a latest subset of features. This new subset was used to build the Bayesian classification and random forest model.

Random forest is an algorithm of machine learning which uses supervised learning technique. Classification and regression problems can be well solved by rendom forest technique. To improve the predicting accuracy of the used data set, random forest includes many decision trees on various data sets of the used data set and then takes the average. In other words, we can say that random forest does not rely on single decision tree rather it considers the prediction from all the trees. The final prediction output is generated on the basis of majority prediction votes. It is assumed that to get higher degree of accuracy and to avoid overfitting, higher number of decision trees in the forest should be used. Few decision trees may produce incorrect output whereas few can produce correct output but, when taken all together, average of all the trees leads to correct output. Bayesian classifier is one of the important classification algorithm which does prediction on the basis of prabability of objects. Followings are the steps used in bayesian classifier:

1) First, the given data set is converted into frequency tables.

2) Second, likelihood table is generated by finding the probabilities of the given features.

3) Finally, posterior probability is calculated by using Bayes theorem.

The main finding of this work is that the proposed prediction model is the best system to figure out the classification prediction problem. The accuracy obtained by using the random forest technique with 5 fold cross-validation is 85.15%. On the other side, bayesian classifier model with 10 fold cross-validation produced the classification accuracy of 84.57%.

5.3 USE OF SELF ORGANIZING MAP AND NEURO-FUZZY SYSTEMS

Akbari (2018) used the non-linear iterative partial least square method to make the reduction of data, and the clustering task has been done by using the self-organizing map and a hybrid system, i.e., neuro-fuzzy inference system for the detection of HBV disease. Decision tree is used to extract important features from the whole data set for predicting the hepatitis disease. A real-world data set is taken by the researcher to do his work. Self-organizing map is a type of artificial neural network. It uses unsupervised type of learning and makes a 2-dimensional map of a problem space. Unlike other techniques, SOM uses competitive learning whereas other techniques use error-correction learning. When input data is given to SOM, Euclidean distance between nodes is computed. Nodes are assigned with a weight. As the neural network start moving through the problem set, the weights starts looking more like the actual data. Here, the neural network trains itself to see the patterns in the data just like the way a human sees. Fuzzy logic and neural networks are different approaches which have their own pros and cons. Fuzzy inference system is relevant for the applications of medical-related, but some problems are huge complex that cannot be solved by fuzzy inference system; hence it fails to solve those particular problems. The main reason behind failure is that sometimes the rules, as well as membership functions of the expert system, become exhausted. To solve this problem of exhaustion, the learning algorithm in the fuzzy expert systems is required. This need can only overcome by using the learning algorithm of the neural network in the fuzzy inference or expert

system. Thus, there is a need of combining these two techniques. Neuro-fuzzy system is the combination of two soft computing methodologies named fuzzy inference system and artificial neural network. Neuro-fuzzy system becomes a new hybrid system which combines the abilities of an artificial neural network, as well as capabilities of fuzzy logic. The results of this research work shows that the performance of the proposed model is more accurate than the other used models such as SVM, neural network, k-nearest neighbor etc. The accuracy of the proposed model was 93.06%.

A comparison of all the three existing models is shown in the following table.

Table 5.1: Comparison of Discussed Existing Models

Sr. No.	Methodology	Classification Accuracy
1.	Support Vector Machine, Decision Rule and Decision Tree [25]	73.20%
2.	Random Forest and Bayesian Classifier [23]	85.15%
3.	Self-organizing map and neuro fuzzy inference system [67]	93.06%

5.4 SUMMARY

Mainly, three existing systems for diagnosing Hepatitis B are discussed here. Uhmn, Kim, & Kim has used support vector machine (SVM), decision tree and decision rule to predict the chronic stage of hepatitis. Accuracy obtanied to classify the disease using decision rule, SVM and decision tree was found to be 73.20%, 67.53% and 72.68%. Wang & Liu developed a system to identify various Hepatitis B's risk factors by using Random forest and Bayesian classification techniques. The accuracy obtained by using the random forest and bayesian classifier was 85.15% and 84.57%. Akbari used self-organizing map and neuro-fuzzy hybrid system for the detection of HBV. The accuracy obtained using nuero-fuzzy inference was 93.06% which was better than other used techniques.

Chapter – 6

DEVELOPMENT OF MULTILAYERED FUZZY INFERENCE SYSTEM

6.1 THE METHODOLOGY USED IN THE DEVELOPMENT OF A MULTI-LAYERED FUZZY INFERENCE SYSTEM

Figure 6.1 shows the methodology used to develop a multi-layered diagnostic system by using fuzzy logic to diagnose HBV.

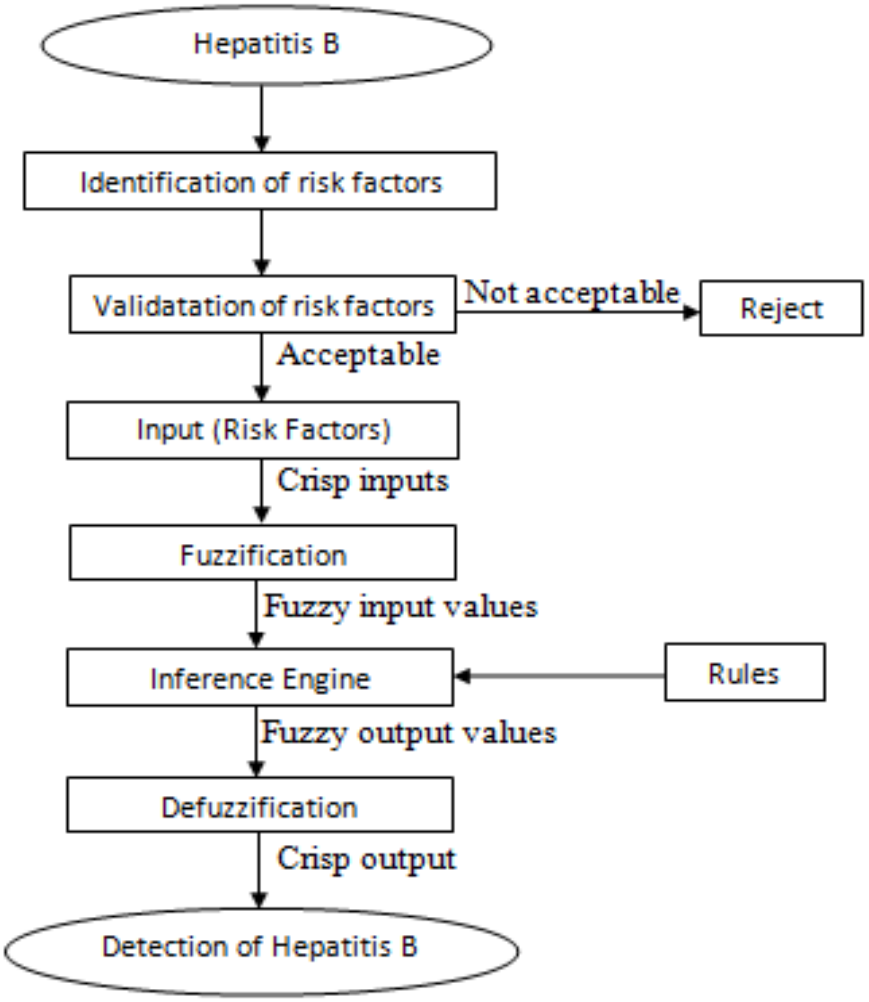

Figure 6.1: Methodology used to develop a System

The number of steps that have to be followed while the development of multilayered fuzzy inference system for the detection of hepatitis B virus is briefly explained below:

Step 1: Identification of risk factors

In the very first step, the various risk factors of the hepatitis B virus are identified. It is necessary to acquire all the causes, symptoms or factors due to which this deadly disease occurred. If the gathered risk factors are accurate, then the result of the system will be definitely correct, effective as well as accurate.

Step 2: Validation of risk factors

The validation of the acquired risk factors is very crucial as it is not mandatory that the gathered factors have much influence on the causes of disease. The risk factors which has a greater impact and has a high probability of causing this harmful disease are approved, and other factors are rejected. These collected risk factors are further used to train the system; hence only choose accurate risk factors in order to achieve maximum accuracy of classification.

Step 3: Input (Risk Factor)

The accepted risk factors in step 2 are used as the input for the system. The considered input set till this step is in crisp values form.

Step 4: Fuzzification

The input variables given to the system in the above step is in the form of crisp values. Crisp input is the number expressing measurement of a variable. The fuzzification step transforms these values into the fuzzy set. The key component of the fuzzy inference system that is the fuzzifier is used for the process of fuzzification.

Step 5: Inference Engine

The inference engine assists in the mapping of suitable rules to the given input and provides the output in a fuzzy set. All the generated facts, as well as rules, are acquired in the knowledge base. The respective rules are generated by acquiring knowledge from an experienced doctor or professional or experts of this disease. The whole output depends on these rules. Hence, the rules must be accurately generated and stored so

that there will be no any conflict while the system is evaluating the final outcome. The inference engine gives an output in the fuzzy set corresponding to the given fuzzy input.

Step 6: Defuzzification

The output computed by the inference engine is in the form of a fuzzy set. This step transforms the output provided by the inference engine in the fuzzy set into the crisp output. Crisp output is the number expressing measurement of a variable.

Step 7: Detection of Hepatitis B Virus

The final outcome is generated by the defuzzifier, and it tells about the health of the patient. The displayed output depends on the output variables that are provided to the system. As in the respective conducted study, the final output of the model will be either no HBV or acute or chronic which depicts the health of a patient.

Hence, these above-mentioned steps are followed during the development of a multilayered fuzzy inference system for the detection of the hepatitis B virus.

6.1.1 Algorithm

The algorithm of the developed methodology that can be used for the diagnosis of the deadly disease that is hepatitis B is given by and also displayed in figure 6.2.

Algorithm:

a) Describe the variables as well as linguistic terms that will be utilised in the developed diagnostic system

b) For all the defined variables in the above step, now, create and design the membership functions.

c) Use the acquired knowledge to generate the rules and construct a knowledge base to store the generated rules as well as facts.

Layer 1

d) In the fuzzification process, convert the provided crisp input data to the system into fuzzy sets by using the created membership functions.

e) After the transformation of data, the mapping of the input with the generated facts and rules will be done by the inference engine.

f) Integrate the outcomes that are evaluated by each and every rule stored in the knowledge base.

g) The output generated by the above step is in the form of a fuzzy set. Hence, now, convert the evaluated output from fuzzy values to crisp values.

Layer 2

h) The inputs for layer 2 of the developed system are decided, and the output of layer 1 is also used as the input to layer 2.

i) Transform the decided inputs into the fuzzy values from the crisp values by using a fuzzifier.

j) Calculate the output according to the stored rules as well as facts.

k) Combine the outcomes obtained from each and every rule and evaluate a final result.

l) The output generated by the above step of layer 2 is in the form of a fuzzy set. Hence, now, convert the evaluated output from fuzzy values to crisp values. This obtained result is considered as the final result given by the developed diagnostic system

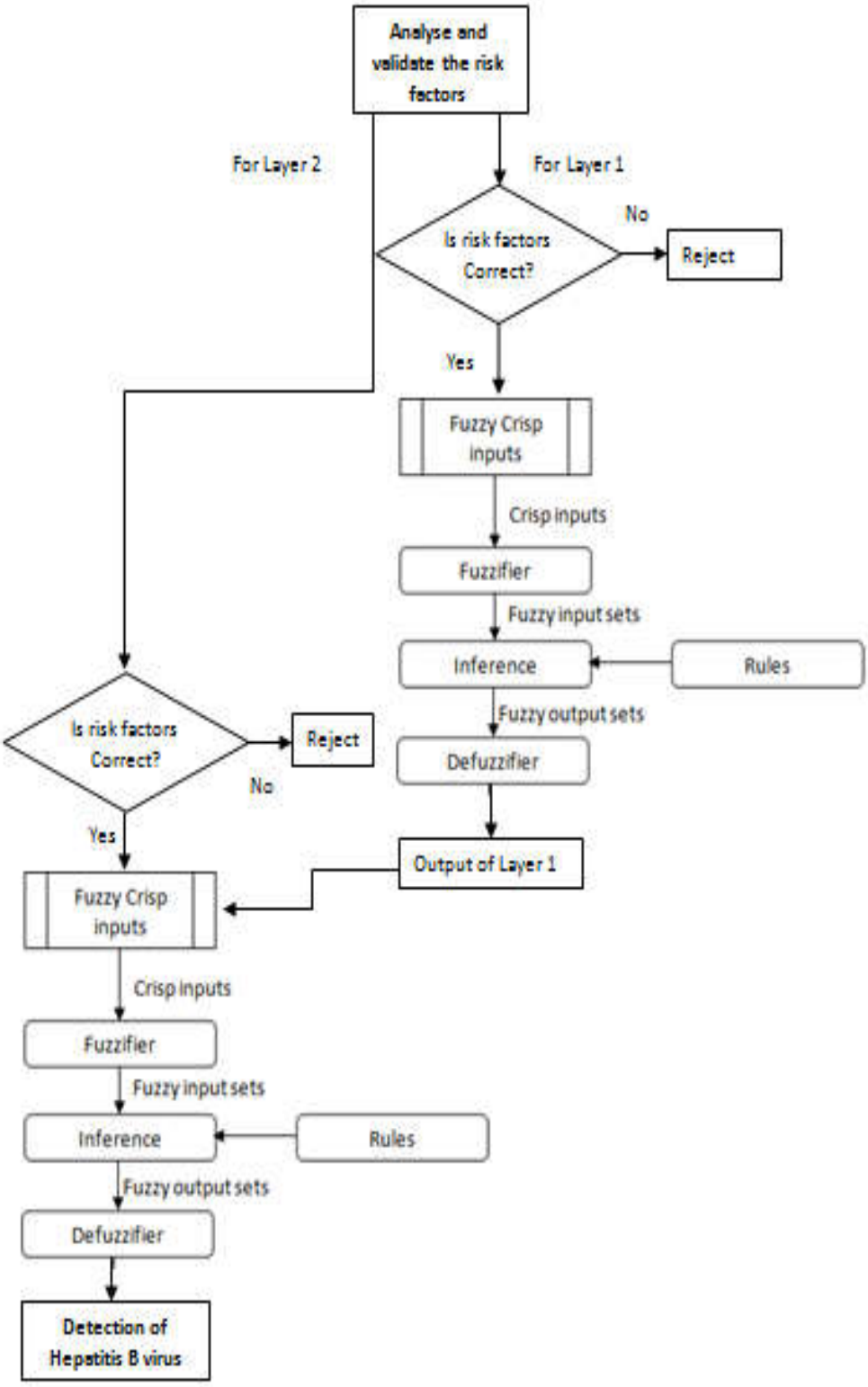

Figure 6.2: Structural diagram of the developed Algorithm

6.2 DEVELOPMENT OF MULTILAYERED FUZZY INFERENCE SYSTEM

Figure 6.3 presents the multi-layered expert system to diagnose Hepatitis B. In this proposed system, the first layer, i.e. layer 1 used clinical symptoms as the input variables and give the output either yes or no, which means this layer only detects whether the patient is suffering from the hepatitis B disease or not. Likewise, in the second layer, i.e., layer 2, the laboratory test values are used as the input variables, which will detect the condition of the patient, either no hepatitis B or chronic or acute. The several input variables used in both layers are decided on the basis of dataset obtained from the hospital and are explained below:

Layer 1: Clinical symptoms

- **Jaundice:** Jaundice is an ailment in which the mucous membrane, skin as well as whiteness of eyes turns into a yellow color. This happens due to the high level of bilirubin. Bilirubin is a bile pigment having yellow-orange color. In this research work, there are two input or linguistic variables for jaundice, i.e. yes and no with different ranges.

- **Dark Urine:** The second symptom that is considered as the input variable for the developed system is dark urine. The main reason for the dark urine is dehydration. The dark urine represents that there are dangerous, unusual as well as excess waste products that are flowing in an individual's body. It is the main symptoms of liver diseases. The input variables used for this clinical symptom in this research work are depicting whether the patient's urine is dark or not.

- **Abdominal Pain:** Abdominal pain is a pain that arises in either the outer muscle wall or inside of the abdomen. This pain can be a normal pain or severe, and if severe, then in that condition, the patient might need emergency care. It is also known as generalized pain that the patient feels pain in half or more area of his or her belly. The patient of hepatitis B feels discomfort in their abdominal. The input variables used in the research work for this clinical symptom are the range of the pain, such as whether the pain is low, moderate or high.

- **Vomiting:** When a patient suffering from hepatitis B virus, then this disease also cause vomiting. Hence, this clinical symptom can also be considered for the detection of HBV disease. In case of vomiting, the body of an individual forcefully throws out the harmful content of the stomach through the mouth of that particular patient. In this current research work, the input variables that are taken under consideration are whether the patient has the symptom of vomiting or not.

Layer 2: Laboratory test

- **HBsAg:** HBsAg is Hepatitis B Surface Antigen. It is basically a blood test in order to identify whether an individual is infected with hepatitis B virus or not. There are some particular antigens; if those antigens are observed in an individual's body, then that patient is considered an infected patient. In this research work, the linguistic variables are taken according to the test range of this test. The variable name "negative" means there is no any antigen found in the patient's body. The "borderline" variable means that some antigens are observed. Similarly, the variable name "positive" means that the patient is infected and antigens are found in his or her body along with the virus.

- **Anti-HBs:** Anti-HBs is an antigen that is present in the body of an individual who takes a vaccine successfully against HBV. Hence, the presence of this antigen indicates immunity as well as recovery from the infection of the hepatitis B virus. If the test result gives output as "positive" Anti-HBs, then it indicates that the particular person is immune enough to beat the hepatitis B virus. Moreover, this person is considered as a non-infected person and also can not able to spread this deadly virus to another person. Similarly, if the result of an individual observed as "negative" anti-HBs, then the patient does not have enough immunity to protect his or her body from this life-threatening disease.

- **Anti-HBc or HBcAb:** The HBsAb is Hepatitis B Core Antibody. The presence of this antibody in a particular body represented the ongoing or previous infection occur due to the hepatitis B virus within an ambiguous time period. The linguistic variables used for this input variable in this research work are "positive",

"borderline", and "negative". The meaning of these linguistic variables varies. If the test result of a patient is positive, then it represents the presence of this antibody in the body of that specific patient and has an acute infection. Similarly, borderline variable indicates the presence of these antibodies is acceptable, but there is a huge need to take care of it so that the antibodies will not increase in future and the patient can be protected from the deadly virus. Further, the negative variable shows the absence of the core antibodies of the hepatitis B virus in an individual's body.

- **HBV DNA:** HBV DNA is a blood test in which the hepatitis B virus is investigated in the DNA of a patient. It is basically the evaluation of the HBV viral load in the blood of an individual. There are two linguistic variables that are taken in this research work for this input variable, i.e., "positive" and "negative". The positive HBV DNA indicates the presence of a deadly virus in the DNA of an individual. The DNA levels of a patient are monitored for 30 days regularly to track the infection, and this might increase if proper treatment will be given to the patient or might decrease and then disappear after diagnosis appropriately. Similarly, the negative variable represents the absence of HBV in the DNA of an individual. It means that that particular patient is a non-infected patient and cannot able to spread this life-threatening virus to another individual.

- **Anti-HBcAg-IgM:** The Anti-HBcAg-IgM is the initial response of a particular body to the infection spread by the hepatitis B virus. Anti-HBcAg-IgM is the antibodies that directly fight against a portion of this virus known as the core. These antigens appear automatically in the body in a very short period of time after the symptoms. The taken linguistic variables for this input variable are "positive" and "negative". Both variables have different ranges according to which the results are observed about whether the Anti-HBcAg-IgM is positive in the blood of a specific patient or negative. The positive result of this laboratory test depicts the recent infection of hepatitis B virus from which a patient is suffering. The presence of this antibody in the blood of a patient represents the acute infection. Similarly, a negative variable indicates that the infection is not present in the person's body.

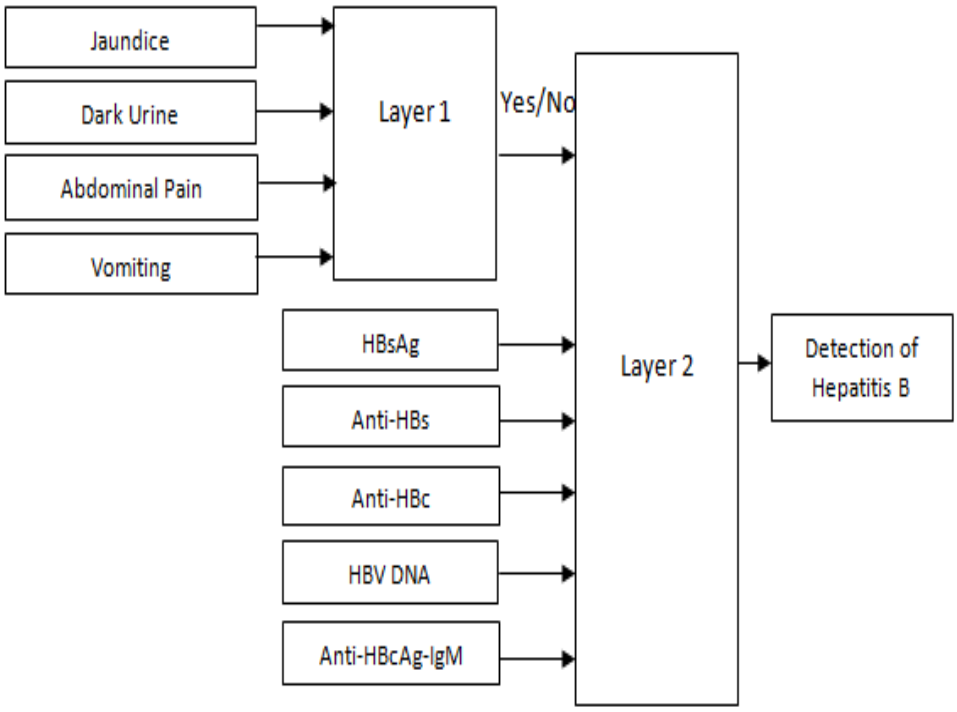

Figure 6.3: Development of Multi-layered Fuzzy Inference System to diagnose HBV.

6.2.1 Input Variables

The statistical values are used as the input to the developed model to diagnose the hepatitis B virus. The number of input variables that are used in this research work is nine. From these nine input variables, four input variables are used as the input for layer 1 and the other five input variables are utilized as the input for layer 2. The different ranges of taken membership functions are according to the test ranges and the knowledge acquired from specialists. These ranges, as well as semantic signs that are taken in the development of multilayered fuzzy inference system for layer 1 as well as layer 2 are written in table 6.1 and table 6.2 respectively. The abbreviations that are used in table 6.1 and table 6.2 are described in table 6.3. Additionally, in table 6.4 and table 6.5, the mathematical representation of each membership function of layer 1 and layer 2 are well-explained.

Table 6.1: Input Variables for Layer 1

Sr. No.	Input variable	Ranges (1/n)	Semantic Sign
1.	Jaundice	LT<4.5	No
		GT>3.4	Yes
2.	Dark Urine	LT<0.5	No
		GT>0.3	Yes
3.	Abdominal Pain	LT<0.3	Low
		B/W 0.2-0.7	Moderate
		GT>0.6	High
4.	Vomiting	LT< 0.4	Yes
		GT>0.3	No

Table 6.2: Input Variables for Layer 2

Sr. No.	Input variable	Ranges	Sematic Sign
1.	HBsAg (signal per cutoff (s/c))	LT<0.85	Negative
		B/W 0.81-0.99	Borderline
		GT>0.93	Positive
2.	Anti-HBs (IU/mL)	LT<10	Negative
		GT>8	Positive
3.	Anti-HBc (IU/mL)	LT<0.87	Positive
		B/W 0.8-1.1	BorderLine
		GT>1.02	Negative
4.	HBV DNA (IU/mL)	LT< 10	Negative
		GT>8.5	Positive
5.	Anti-HBcAg-IgM (IU/mL)	LT<0.85	Negative
		B/W 0.8-1.1	BorderLine
		GT>1.02	Positive

Table 6.3: Abbreviations used in table 4 and table 5

Abbreviation	Meaning
LT	Less Than
B/W	Between
GT	Greater Than

Table 6.4: Mathematical Representation of Membership Variables used in Layer 1

Sr. no.	Input Variables	Membership functions	Mathematical representation of membership functions
1.	Jaundice (J)	Trapezoidal	$\mu_{No}(j) = \begin{cases} 0, & j > 4.5 \\ \frac{0.45 - j}{0.3}, & 0.15 \le j \le 0.45 \\ 1, & j < 0.15 \end{cases}$ $\mu_{Yes}(J) = \begin{cases} 0, & j < 0.34 \\ \frac{j - 0.34}{0.26}, & 0.34 \le j \le 0.6 \\ 1, & j > 0.6 \end{cases}$
2.	Dark Urine (DU)	Trapezoidal	$\mu_{No}(du) = \begin{cases} 0, & du > 4.5 \\ \frac{0.5 - du}{0.23}, & 0.27 \le du \le 0.5 \\ 1, & du < 0.27 \end{cases}$

Sr. no.	Input Variables	Membership functions	Mathematical representation of membership functions
			$\mu_{Yes}(du)$ $= \begin{cases} 0, & du < 0.3 \\ \frac{du - 0.3}{0.25}, & 0.3 \leq du \leq 0.55 \\ 1, & du > 0.55 \end{cases}$
3.	Abdominal Pain (AP)	Trapezoidal	$\mu_{Low}(ap)$ $= \begin{cases} 0, & ap > 0.3 \\ \frac{0.3 - ap}{0.1}, & 0.2 \leq ap \leq 0.3 \\ 1, & ap < 0.2 \end{cases}$ $\mu_{Moderate}(ap)$ $= \begin{cases} 0, & ap < 0.2 \\ \frac{ap - 0.2}{0.2}, & 0.2 \leq ap \leq 0.4 \\ 1, & 0.4 \leq ap \leq 0.5 \\ \frac{0.7 - ap}{0.2}, & 0.5 \leq ap \leq 0.7 \\ 0, & ap > 0.7 \end{cases}$ $\mu_{High}(ap)$ $= \begin{cases} 0, & ap < 0.6 \\ \frac{ap - 0.6}{0.2}, & 0.6 \leq ap \leq 0.8 \\ 1, & ap > 0.8 \end{cases}$
4.	Vomitting (V)	Trapezoidal	$\mu_{No}(v) = \begin{cases} 0, & v > 0.4 \\ \frac{0.4 - v}{0.24}, & 0.16 \leq v \leq 0.4 \\ 1, & v < 0.16 \end{cases}$

Sr. no.	Input Variables	Membership functions	Mathematical representation of membership functions
			$\mu_{Yes}(v) = \begin{cases} 0, & v < 0.3 \\ \dfrac{v - 0.3}{0.2}, & 0.3 \leq v \leq 0.5 \\ 1, & v > 0.5 \end{cases}$

Table 6.5: Mathematical Representation of Membership Variables used in Layer 2

Sr. No.	Input Variables	Membership functions	Mathematical representation of membership functions
1.	HBsAg (hbs)	Trapezoidal	$\mu_{Negative}(hbs) = \begin{cases} 0, & hbs > 0.85 \\ \dfrac{0.85 - hbs}{0.15}, & 0.7 \leq hbs \leq 0.8 \\ 1, & hbs < 0.7 \end{cases}$ $\mu_{Borderline}(hbs) = \begin{cases} 0, & hbs < 0.81 \\ \dfrac{hbs - 0.81}{0.04}, & 0.81 \leq hbs \leq 0.85 \\ 1, & 0.85 \leq hbs \leq 0.95 \\ \dfrac{0.99 - hbs}{0.04}, & 0.95 \leq hbs \leq 0.99 \\ 0, & hbs > 0.99 \end{cases}$ $\mu_{Positive}(hbs) = \begin{cases} 0, & hbs < 0.93 \\ \dfrac{hbs - 0.93}{0.17}, & 0.93 \leq hbs \leq 1.1 \\ 1, & hbs > 1.1 \end{cases}$

Sr. No.	Input Variables	Membership functions	Mathematical representation of membership functions
2.	Anti-HBs (ahbs)	Trapezoidal	$\mu_{Negative}(ahbs)$ $= \begin{cases} 0, & ahbs > 10 \\ \frac{10 - ahbs}{2}, & 8 \leq ahbs \leq 10 \\ 1, & ahbs < 8 \end{cases}$ $\mu_{Positive}(ahbs)$ $= \begin{cases} 0, & ahbs < 8 \\ \frac{ahbs - 8}{2.8}, & 8 \leq ahbs \leq 10.8 \\ 1, & ahbs > 10.8 \end{cases}$
3.	Anti-HBc (ahbc)	Trapezoidal	$\mu_{Positive}(ahbc)$ $= \begin{cases} 0, & ahbc > 0.87 \\ \frac{0.87 - ahbc}{2}, & 0.7 \leq ahbc \leq 0.87 \\ 1, & ahbc < 0.7 \end{cases}$ $\mu_{Borderline}(ahbc)$ $= \begin{cases} 0, & ahbc < 0.8 \\ \frac{ahbc - 0.8}{0.05}, & 0.8 \leq ahbc \leq 0.85 \\ 1, & 0.85 \leq ahbc \leq 1.05 \\ \frac{1.1 - ahbc}{0.05}, & 1.05 \leq ahbc \leq 1.1 \\ 0, & ahbc > 1.1 \end{cases}$ $\mu_{Negative}(ahbc)$ $= \begin{cases} 0, & ahbc < 1.02 \\ \frac{ahbc - 1.02}{0.18}, & 1.02 \leq ahbc \leq 1.2 \\ 1, & ahbc > 1.2 \end{cases}$

Sr. No.	Input Variables	Membership functions	Mathematical representation of membership functions
4.	HBV DNA (hbd)	Trapezoidal	$\mu_{Negative}(hbd)$ $= \begin{cases} 0, & hbd > 10 \\ \frac{10 - hbd}{2}, & 8 \leq hbd \leq 10 \\ 1, & hbd < 8 \end{cases}$ $\mu_{Positive}(hbd)$ $= \begin{cases} 0, & hbd < 8.5 \\ \frac{hbd - 8.5}{2.3}, & 8.5 \leq hbd \leq 10.8 \\ 1, & hbd > 10.8 \end{cases}$
5.	Anti-HBcAg-IgM (ahbi)	Trapezoidal	$\mu_{Negative}(ahbi)$ $= \begin{cases} 0, & ahbi > 0.89 \\ \frac{0.89 - ahbi}{0.19}, & 0.7 \leq ahbi \leq 0.89 \\ 1, & ahbi < 0.7 \end{cases}$ $\mu_{Borderline}(ahbi)$ $= \begin{cases} 0, & ahbi < 0.8 \\ \frac{ahbi - 0.8}{0.05}, & 0.8 \leq ahbi \leq 0.85 \\ 1, & 0.85 \leq ahbi \leq 1.05 \\ \frac{1.1 - ahbi}{0.05}, & 1.05 \leq ahbi \leq 1.1 \\ 0, & ahbi > 1.1 \end{cases}$ $\mu_{Positive}(ahbi)$ $= \begin{cases} 0, & ahbi < 1.02 \\ \frac{ahbi - 1.02}{0.18}, & 1.02 \leq ahbi \leq 1.2 \\ 1, & ahbi > 1.2 \end{cases}$

The input variables with their membership functions used in layer 1 and layer 2 of the multi-layered fuzzy expert system to diagnose Hepatitis B are shown in the figures. The used membership function is trapezoidal function because as per the literature, this membership function gives better results for this kind of problems. Figure 6.4-6.7 present the input variables of layer 1, and similarly, figure 6.8-6.12 displays layer 2's input variables of the medical diagnostic system. Additionally, the whole structure of this developed inference system is shown in table 6.6.

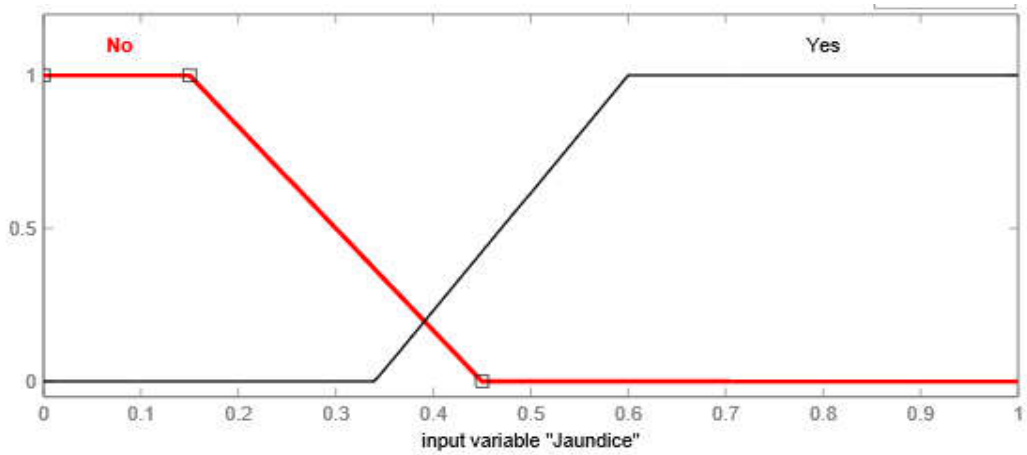

Figure 6.4: Input Variable of Layer 1, "Jaundice."

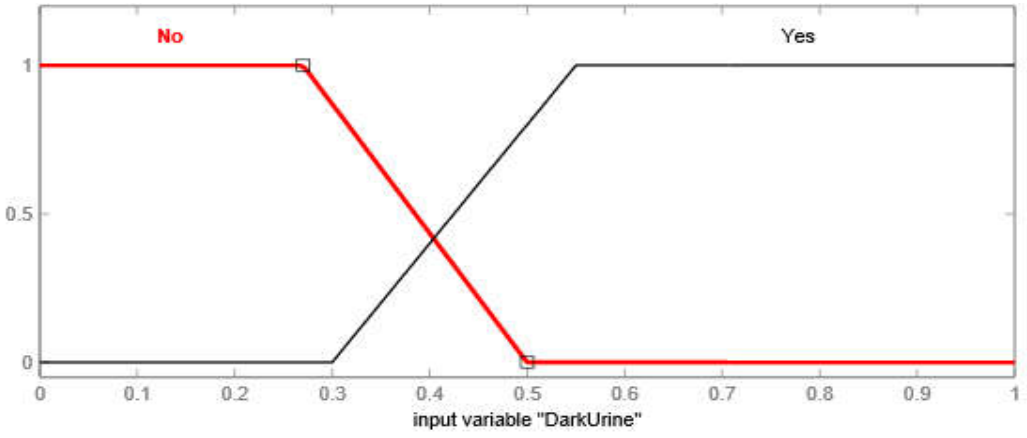

Figure 6.5: Input Variable of Layer 1, "Dark Urine."

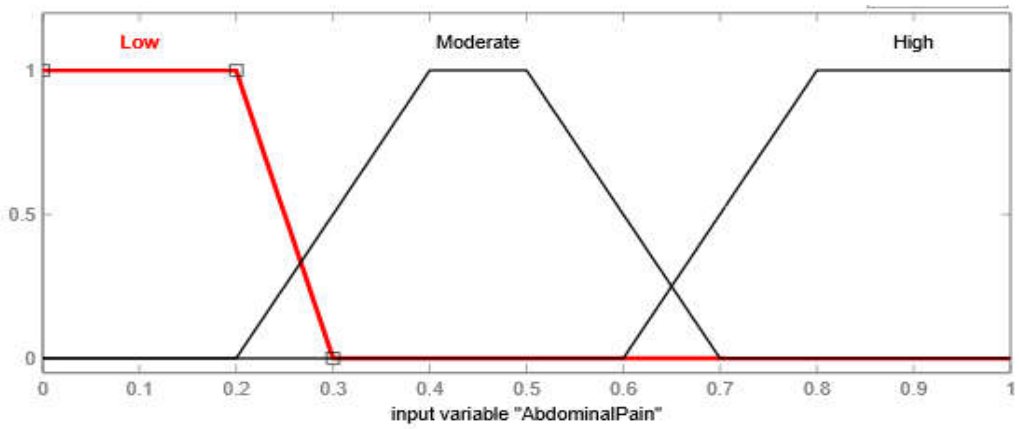

Figure 6.6: Input Variable of Layer 1, "Abdominal Pain."

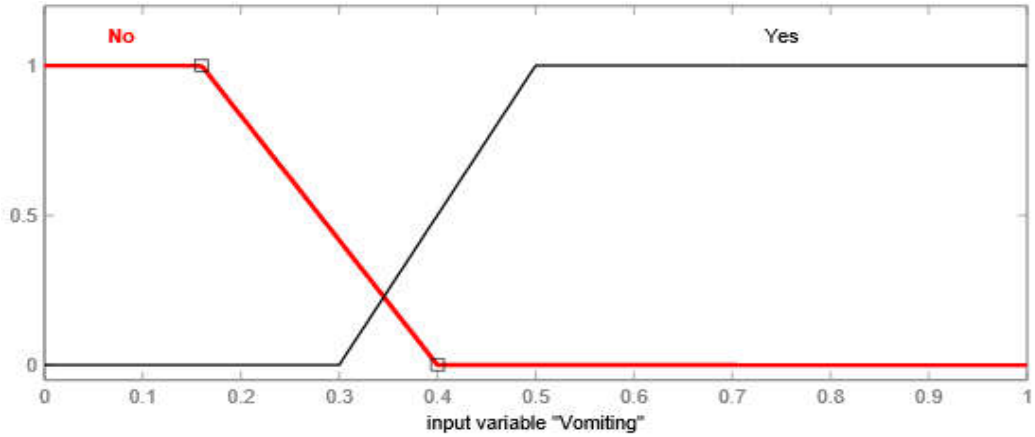

Figure 6.7: Input Variable of Layer 1, "Vomiting."

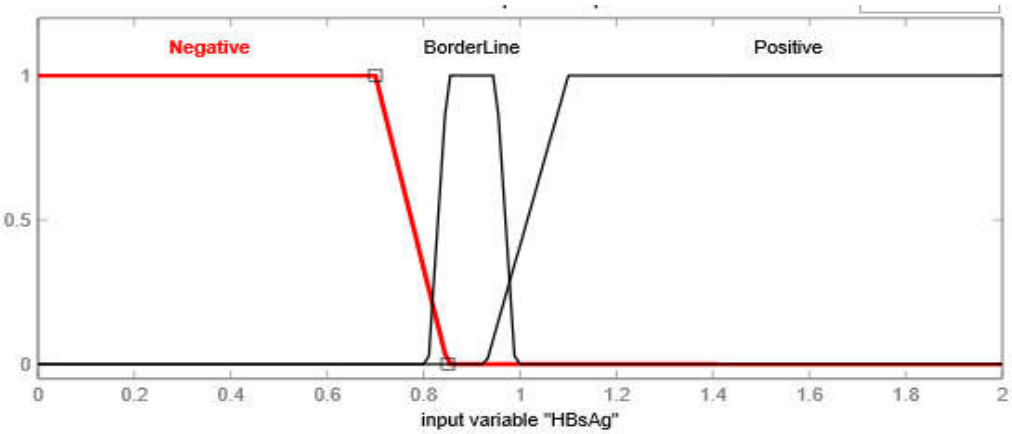

Figure 6.8: Input Variable of Layer 2 "HBsAg."

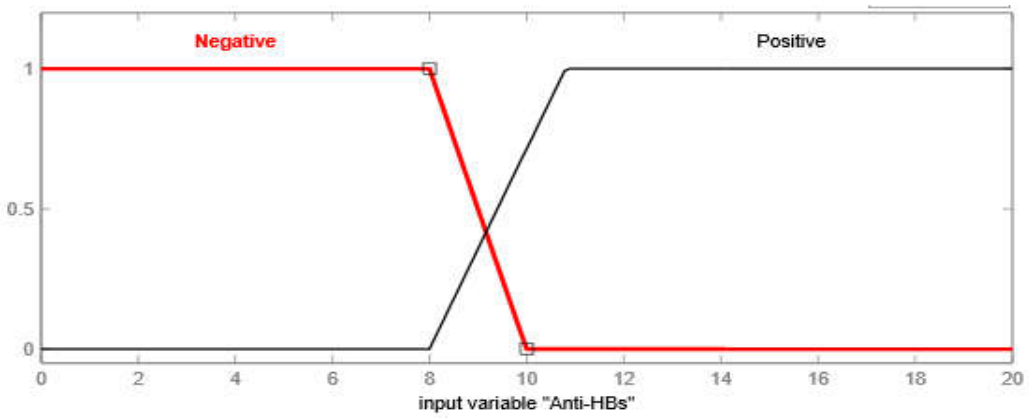

Figure 6.9: Input Variable of Layer 2 "Anti-HBs."

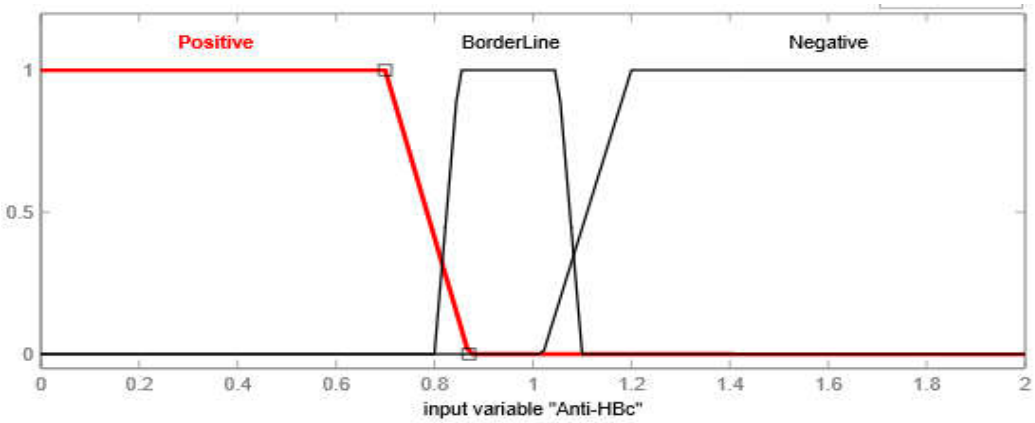

Figure 6.10: Input Variable of Layer 2 "Anti-HBc."

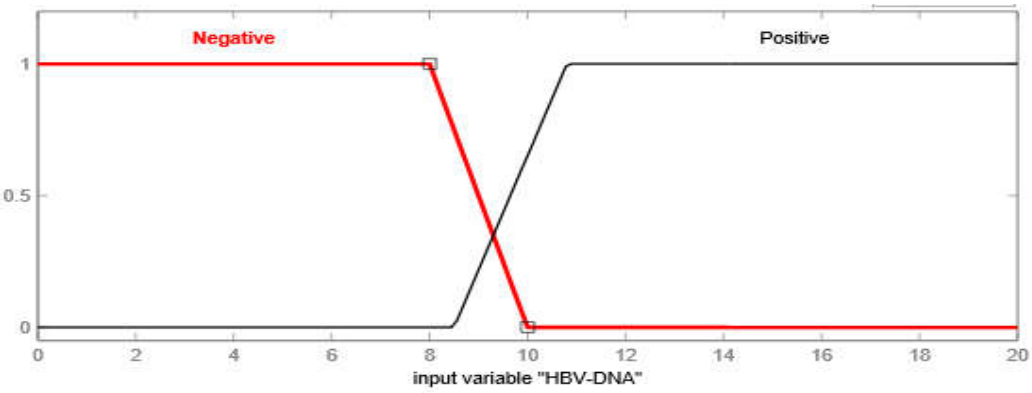

Figure 6.11: Input Variable of Layer 2 "HBV DNA."

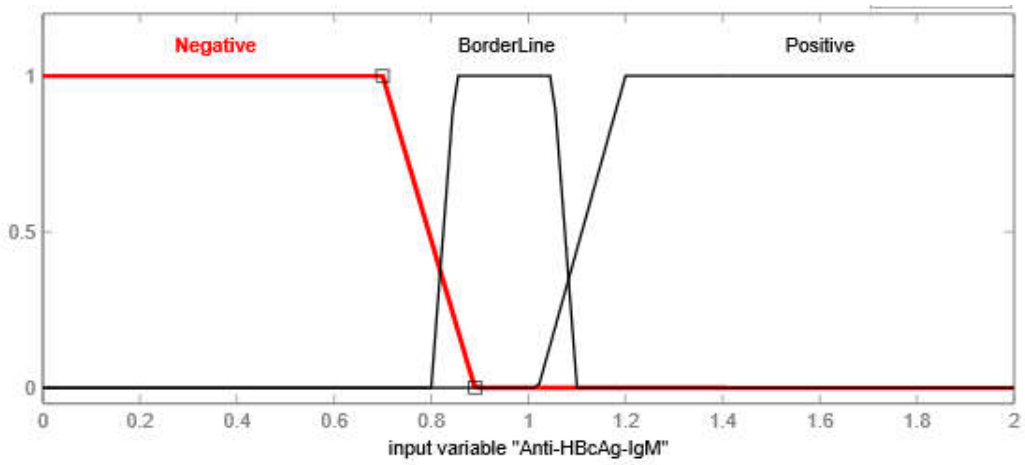

Figure 6.12: Input Variable of Layer 2 "Anti-HBcAg-IgM."

Table 6.6: Structure of developed Multilayered Fuzzy Inference System to diagnose HBV

Structure	**Information of Structure**
Number of layers in Fuzzy inference system	2
Semantic name of layers	Layer 1 Layer 2
Number of input variables for layer 1	4
Number of input variables for layer 2	5
Membership function type of input variables for layer 1	Trapezoidal
Membership function type of input variables for layer 2	Trapezoidal
Number of rules for layer 1	24
Number of rules for layer 2	108
Total number of rules	24 + 108 =132
Type of rules	IF-THEN
Number of membership functions for output variables for layer 1	2
Number of membership functions for output variables for layer 2	3

Membership function type of output variables for layer 1	Trapezoidal
Membership function type of output variables for layer 2	Trapezoidal
Method of defuzzification	Centroid

6.2.2 Rules

The knowledge base of an expert system is used to store the rules. The information is gathered from the professionals or experts of the disease, and according to that information, the rules are generated to detect the disease. The aggregation of these stores various rules have been made by disjunction or conjunction. The form of the generated rules is the IF-THEN form in the case of the Mamdani system of fuzzy logic. The framework of rules for layer 1 and layer 2 is represented in figure 6.13 and 6.14 respectively.

These rules are stored in below mentioned manner

If (input variable (a))

And (input variable (b))

And (input variable (c))

. . . . . .

Then output is (x)

Where a, b, c are linguistic variables of input variables and x is a linguistic variable of the output variable.

The instances of generated and stored rules for layer 1 are given by:

Instance 1: Inference Rule no. 5

If Jaundice is No

And Dark Urine is No

And Abdominal Pain is High

And Vomiting is No

Then Output is No (Means hepatitis B virus is absent)

Instance 2: Inference Rule no. 10

If Jaundice is No

And Dark Urine is Yes

And Abdominal Pain is Moderate

And Vomiting is Yes

Then Output is Yes (Means hepatitis B virus is present)

The instances of generated and stored rules for layer 2 are given by:

Instance 1: Inference rule no. 2

If HBsAg is Negative

And Anti-HBs is Negative

And Anti-HBc is Positive

And HBV DNA is Negative

And Anti-HBcAg-IgM is Borderline

Then HBV is No HBV

Instance 2: Inference rule no. 42

If HBsAg is Borderline

And Anti-HBs is Negative

And Anti-HBc is Positive

And HBV DNA is Positive

And Anti-HBcAg-IgM is Positive

Then HBV is Acute

Instance 3: Inference rule no. 102

If HBsAg is Positive

And Anti-HBs is Positive

And Anti-HBc is Borderline

And HBV DNA is Positive

And Anti-HBcAg-IgM is Positive

Then HBV is Chronic

The generated rules and fact has an essential role in the development of the medical diagnostic model. The accuracy of the classification and other performance parameters directly depends on these generated input-output rules. Hence, the knowledge should be acquired accurately from the expert in order to make correct, appropriate as well as adequate input-output rules and facts. The number of rules of an inference system is dependent on the count of linguistic variables taken in the input variables and output variables in that particular inference model.

For layer 1 of this developed model, the number of inputs is 4, and each input variables are either two membership functions or three membership functions. Therefore, the number of rules that can be generated is calculated as follow:

Total number of rules: number of membership functions in input variable 1 * number of membership functions in input variable 2 * number of membership functions in input variable 3 * number of membership functions in input variable 4

= Jaundice (2) * Dark Urine (2) * Abdominal Pain (3) * Vomiting (2)

= 2 * 2 * 3 * 2

= 24 rules

Similarly, for layer 2 of the developed model, the number of inputs are 5, and each input variables are either having two membership functions or three membership functions. Hence, the number of rules that must be generated can be calculated as follow:

Total number of rules: number of membership functions in input variable 1 * number of membership functions in input variable 2 * number of membership functions in input variable 3 * number of membership functions in input variable 4 * number of membership functions in input variable 5

= HBsAg (3) * Anti-HBs (2) * Anti-HBc (3) * HBV DNA (2) * Anti-HBcAg-IgM (3)

= 3 * 2 * 3 * 2 * 3 =108 rules

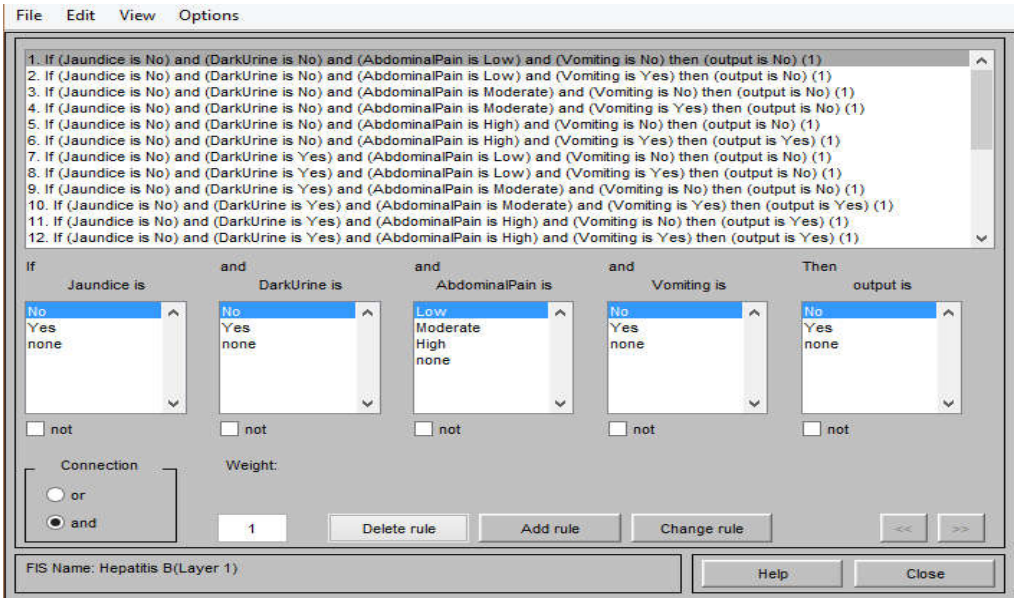

Figure 6.13: Framework of Rules for Layer 1

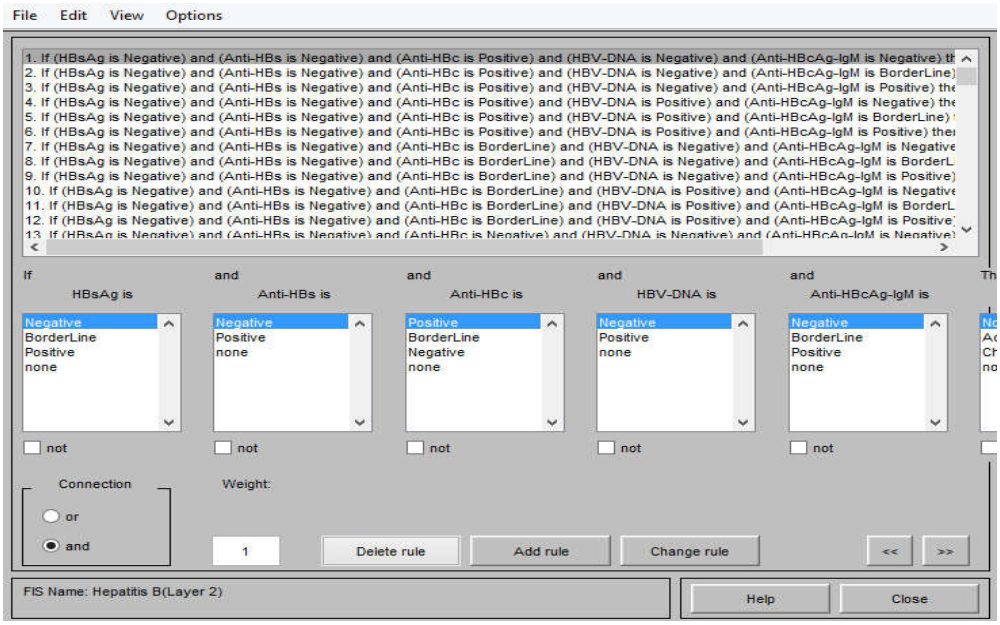

Figure 6.14: Framework of Rules for Layer 2

6.2.3 Output Variables

The last step of the fuzzy inference system is defuzzification. The defuzzifier is used to transform the output provided by the inference engine in the fuzzy set into crisp values. The different output variables are used for both layers of the developed multilayered inference system to diagnose HBV.

For layer 1, the name of the output variable is “output”, and it has two membership functions. The name of membership functions is “No” and “Yes”. This layer depicts that whether the patient has hepatitis B virus in his or her body or not. If the result of this layer is observed as “yes” then the patient will go through another layer, i.e. layer 2, for further laboratory tests. The trapezoidal membership functions are used for both taken variables.

Likewise, for layer 2, the name of the output variable is “HBV”, and it has three different membership functions. The name of these membership functions is “No HBV”, “Acute” and “Chronic”. This layer of the developed model indicated the stage of the virus in a specific patients’ body. Additionally, the membership functions of this layer are also trapezoidal membership functions.

Table 6.7: Output Variables for Layer 1 and Layer 2

Sr. no.	Layer	Output Variables	Ranges	Semantic Sign
1.	Layer 1	Output	LT<0.5	No
			GT>0.3	Yes
2.	Layer 2	HBV	LT<0.4	No HBV
			B/W 0.3-0.6	Acute
			GT>0.5	Chronic

These ranges, as well as semantic signs that are taken in the development of a multilayered fuzzy inference system for layer 1 as well as layer 2 are written in table 6.7. The abbreviations that are used in these tables are the same as explained in table 6.3. Additionally, in table 6.8, the mathematical representation of each membership function used for outputs of layer 1 and layer 2 are well-explained.

In the proposed medical diagnostic system, every layer has its individual output. The name of outputs is “output” and “HBV” for layer 1 and layer 2, respectively. The different membership functions that are considered in the development of a multilayered inference system to diagnose HBV are shown in figures. These figures represent the ranges of membership functions, type of membership function and semantic name of membership functions for both layer 1 as well as layer 2. The centroid method is used for defuzzification in this research work.

Table 6.8: Mathematical Representation of Output Membership Variables used in Layer 1 and Layer 2

Sr. No.	Layer	Output Variable	Membership Functions	Mathematical Representation of membership functions
1	Layer 1	Output (o)	Trapezoidal	$\mu_{No}(o) = \begin{cases} 0, & o > 0.5 \\ \frac{0.5 - o}{0.2}, & 0.3 \leq o \leq 0.5 \\ 1, & o < 0.3 \end{cases}$ $\mu_{Yes}(o) = \begin{cases} 0, & o < 0.3 \\ \frac{o - 0.3}{0.2}, & 0.3 \leq o \leq 0.5 \\ 1, & o > 0.5 \end{cases}$
2	Layer 2	HBV (hbv)	Trapezoidal	$\mu_{NoHBV}(hbv) = \begin{cases} 0, & hbv > 0.4 \\ \frac{0.4 - hbv}{0.2}, & 0.2 \leq hbv \leq 0.4 \\ 1, & hbv < 0.2 \end{cases}$ $\mu_{Acute}(hbv) = \begin{cases} 0, & hbv < 0.3 \\ \frac{hbv - 0.3}{0.1}, & 0.3 \leq hbv \leq 0.4 \\ 1, & 0.4 \leq hbv \leq 0.5 \\ \frac{0.6 - hbv}{0.1}, & 0.5 \leq hbv \leq 0.6 \\ 0, & hbv > 0.6 \end{cases}$ $\mu_{Chronic}(hbv) = \begin{cases} 0, & hbv < 0.5 \\ \frac{hbv - 0.5}{0.2}, & 0.5 \leq hbv \leq 0.7 \\ 1, & hbv > 0.7 \end{cases}$

Figure 6.15 displays the output variables that are used in layer 1 of the developed medical diagnostic system. Similarly, figure 6.16 indicates the output variables that are

used in layer 2 of the developed multilayered fuzzy inference system to detect the hepatitis B virus from a patient's body.

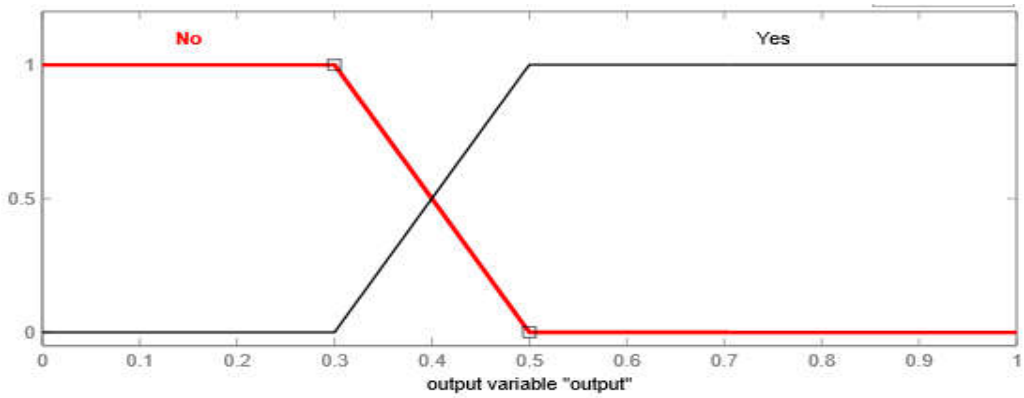

Figure 6.15: Output Variables of Layer 1

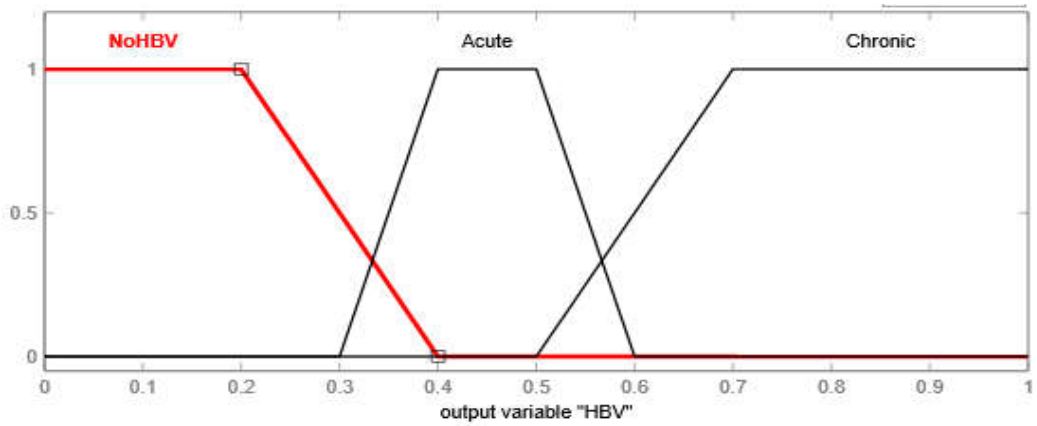

Figure 6.16: Output Variables of Layer 2

6.3 SUMMARY

In this chapter, the developed multilayered fuzzy inference system is explained. The system has two layers, layer1 and layer2. Clinical symptoms like jaundice, dark urine, abdominal pain and vomiting are used as inputs to layer1 whereas laboratory test parameters like HBsAg, Anti-HBs, HBsAb, HBV DNA, Anti-HBcAg-IgM are used as inputs to layer2. The layer1 tells that whether the patient is suffering from the hepatitis B disease or not and layer2 tells the stage of hepatitis B. Total 24 rules for layer1 and 108 rules for layer 2 are generated. All the membership functions are also represented with the help of various figures.

Chapter – 7

DEVELOPMENT OF MEDICAL INTELLIGENT SYSTEM FOR THE DIAGNOSIS OF HEPATITIS B BY USING NEURO-FUZZY TECHNIQUE

In this chapter, the entire methodology that is used for the development of the medical intelligent system to diagnose HBV is effectively discussed. Additionally, the description of the collected dataset, empirical study of the experimental setup, structure of the developed intelligent system, input variables, generated rules, output variables, as well as the training and testing of the developed model is well explained.

7.1 DATA DESCRIPTION

The dataset is the set of diverse data stored in a well-structured manner. The dataset that is used in this research work is collected from the hospitals as well as various trials are done by doctors or experts, or professionals. In this dataset, some people are infected from hepatitis B virus disease, and some are non-infected. Additionally, the dataset does not contain any kind of missing values. The considered input variables in the conducted study are known as attributes of the dataset. Each and every sample in the dataset represent the symptoms or risk factors of hepatitis B virus disease, and the outputs are also calculated with respect to provided samples. The various risk factors that are taken in this research work for the development of an intelligent system by utilizing the neuro-fuzzy technique are HBsAg, Anti-HBs or HBsAb, Anti-HBc or HBcAb, HBV DNA, and Anti-HBcAg-IgM. Each risk factors have different values that are extracted from laboratory tests of real patients. Furthermore, the output has also different ranges. This range depicts the health of the patient, such as whether the patient is suffering from the deadly as well as life-threatening disease, i.e., hepatitis B virus or not. All data in the dataset is in the form of crisp values.

7.2 ALGORITHM

Step 1: Collect the data set of Hepatitis B disease from physicians.

Step 2: Make adaptive-neuro fuzzy inference system by making calls to the function ANFIS.

Do

Load the data set.

Train the dataset

Set the epochs.

Repeat the steps for testing and checking.

Draw the structure of ANFIS that consist of 5 layers.

Layer 1 is used for the acceptance of fuzzy membership functions by using the formulae:

$$O\ 1, i = Ai(x) \quad \text{for } i = 1, 2$$

$$O\ 1, i = Bi\text{-}2(x) \quad \text{for } i = 3, 4$$

Where x is the input node. Variable i is a linguistic label associated with this node. O 1 is the output of the ith node of the layer l.

Second layer of ANFIS contains fixed nodes labelled as $\prod$. Each node represents the fire strength of the rule. This layer deals with T-norm operator that perform the AND operations. The output is the product of all the incoming signals and is represented as:

$$O2, i = wi = \mu Ai(x) \,.\, \mu Bi(y), \quad i = 1, 2$$

Third layer of ANFIS contains fixed nodes labelled as N. It plays the normalization role on the rules coming from the previous layer. It is also known as normalized firing strengths. The results are represented by the following equation:

$$O3, i = \overline{w}i = WiW1+W2, \quad i = 1, 2$$

In the fourth layer, every node is an adaptive node with a node function:

$$O4, i = \overline{w}ifi = \overline{w}i\ (pix + qiy + ri), \quad i=1,2$$

Where wi is the normalized firing strength from layer 3 and {pi, qi, ri} is the parameter set of this node. These are referred to as consequent parameters.

The single node in fifth layer is a fixed node labelled sum, which computes the overall output as the summation of all incoming signals until all inputs are classified correctly.

$$O5, i = F = \Sigma i\ \overline{w}ifi = \Sigma iwifi\Sigma iwi$$

Step3: Evaluate the accuracy, sensitivity, specificity and precision parameter.

7.3 EXPERIMENTAL SETUP

The Matlab R2014a software has been used to perform the implementation of the medical intelligent system to diagnose HBV with ANFIS classifier. The flow in which the whole procedure of development is done is illustrated in figure 7.1.

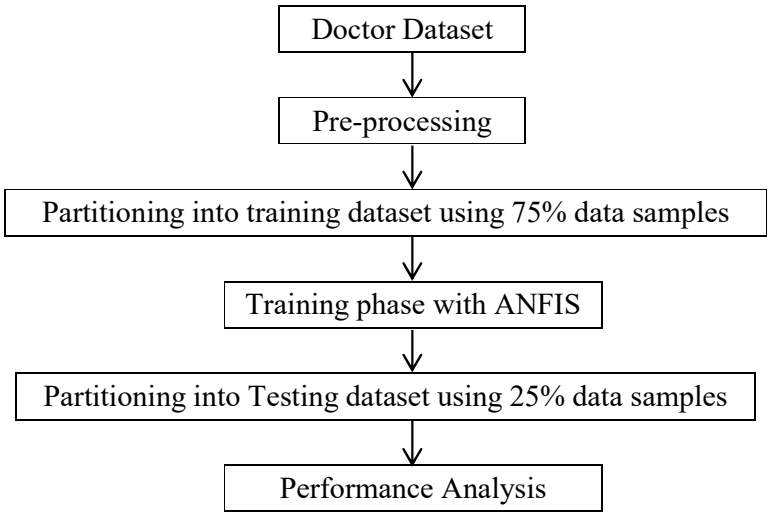

Figure 7.1: Flow Diagram of Model Development

The very first step is the partition of the acquired dataset. The dataset is partitioned into two different parts, which are further used for training of the system and testing of the model. In this research work, 75% of data is being utilized in the phase of

training of the developed model and the other 25% data is utilized to test the developed model. This dataset contains details of patients suffering from different stages of HBV like Acute, chronic or No HBV.

The training data, as well as testing data for a model to develop, must be diverse to see if the system has the capability to classify the unseen data samples as well or not. Hence, never used those data samples during the testing phase of the system, which are already used to train that specific model. The training of the developed model is being done by utilizing the hybrid technique, i.e. ANFIS.

7.4 THE METHODOLOGY USED IN THE DEVELOPED NEURO-FUZZY MODEL

An intelligent medical inference system to diagnose Hepatitis B is developed by using the adaptive neuro-fuzzy inference system. The various input variables used in the development of the system are HBsAg, Anti-HBs or HBsAb, Anti-HBc or HBcAb, HBV DNA, and Anti-HBcAg-IgM. The dataset is acquired from the professionals about the patients of hepatitis B disease as described in the previous sections of this chapter. The variables assigned to the input variables of the developed system are input1, input2, input3, input4 and input 5. Likewise, the output of the proposed system has been presented by the output variable. The structure of the whole model is displayed in table 7.1.

The structure of the proposed system using the adaptive neuro-fuzzy technology to diagnose Hepatitis B disease is displayed in figure 7.2. In this figure, the five different input variables that are selected to develop an intelligent model to diagnose HBV is effectively shown. Moreover, this figure also depicts the model used in this development, and there is one output as discussed in the above sections of this chapter. The different ranges of these inputs as well as output variables to create membership function are being generated automatically during the period of the training phase. The diverse ranges are according to the provided dataset to the developed model and also described in a commendable manner in previous sections of the respective study.

Table 7.1: Structure of developed Neuro-Fuzzy Model for Hepatitis B Disease

Structure	Information of Structure
Number of layers in ANFIS	5
Number of input variables in the developed model	5
Considered input variables	1. HBsAg 2. Anti-HBs or HBsAb 3. Anti-HBc or HBcAb 4. HBV DNA 5. Anti-HBcAg-IgM
Name of input variables	1. HBsAg as input1 2. Anti-HBs or HBsAb as input2 3. Anti-HBc or HBcAb as input3 4. HBV DNA as input4 5. Anti-HBcAg-IgM as input5
Type of membership functions for input variables	Trapezoidal
Number of rules	108
Number of output variable in the developed model	1
Name of the output variable	Output

All the linguistic variables or membership function are generated by the model instead of generating it manually as did during the development of the model by using fuzzy logic.

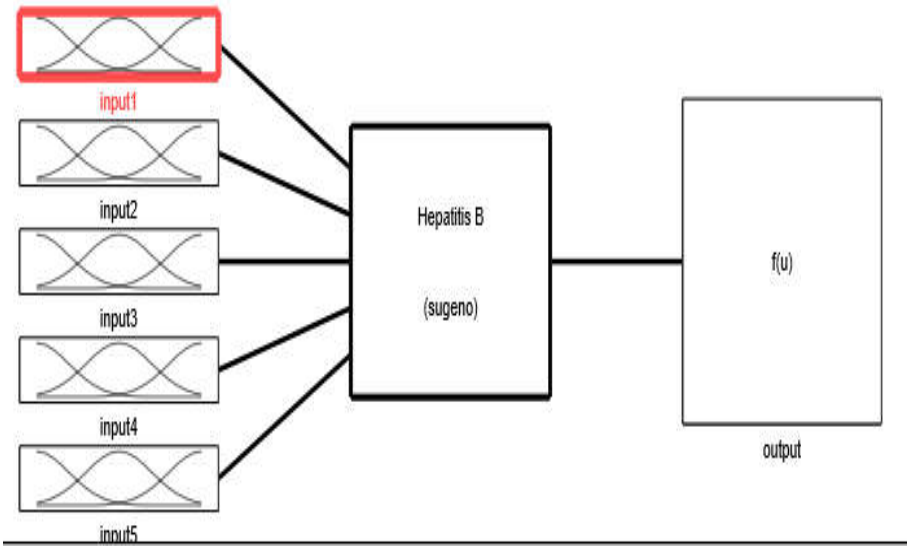

Figure 7.2: Structure of Medical System for Diagnosis of Hepatitis B

7.5 MEMBERSHIP FUNCTION

The membership function used in the development of the system is a trapezoidal membership function. The input variables such as input1, input3 and input5 have been divided into three linguistic variables hence has three trapezoidal membership functions. Likewise, the input variables such as input2 and input4 have two linguistic variables hence have two trapezoidal membership functions. For instance, the first input variable, i.e., input1, is grouped into negative, borderline and positive. In the developed model for the diagnosis of hepatitis B, these linguistic variables are represented as in1mf1, in2mf2 and in3mf3, respectively. These categories of membership function have been automatically generated by ANFIS according to the provided training data set.

All the input variables, along with the name of input variables, the number of linguistic variables, group of input variables and semantic name of those selected linguistic variables, are described in table 7.2.

Figure 7.3 to figure 7.7 displays the membership functions of input variables used in the developed system to diagnose hepatitis B disease, respectively.

Table 7.2: Input Variables for Neuro-Fuzzy Intelligent Model

Sr. No.	Input variable	Name of the input variable	Number of linguistic variables	Groups of input variables	Semantic name of linguistic variables
1.	HBsAg	Input1	3	Negative	In1mf1
				Borderline	In1mf2
				Positive	In1mf3
2.	Anti-HBs or HBsAb	Input2	2	Negative	In2mf1
				Positive	In2mf2
3.	Anti-HBc or HBcAb	Input3	3	Postive	In3mf1
				Borderline	In3mf2
				Negative	In3mf3
4.	HBV DNA	Input4	2	Negative	In4mf1
				Positive	In4mf2
5.	Anti-HBcAg-IgM	Input 5	3	Negative	In5mf1
				Borderline	In5mf2
				Positive	In5mf3

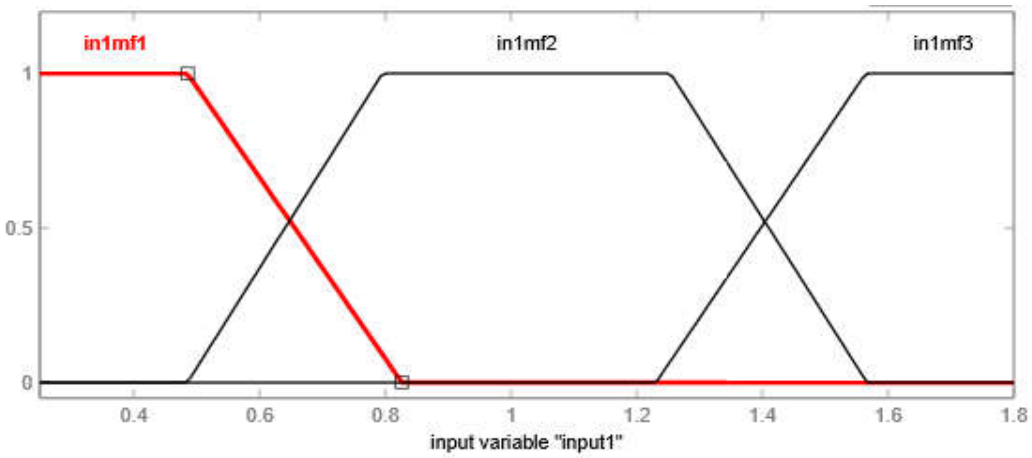

Figure 7.3: Membership Function of Input1, i.e., HBsAg

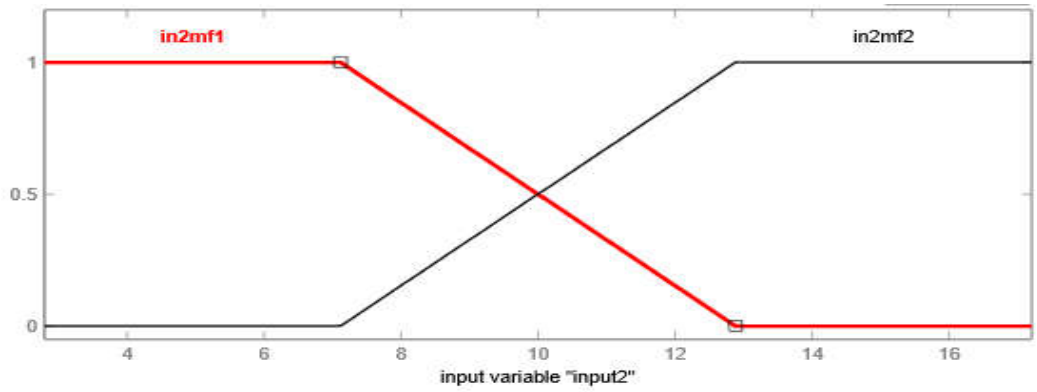

Figure 7.4: Membership Function of Input2, i.e., Anti-HBs or HBsAb

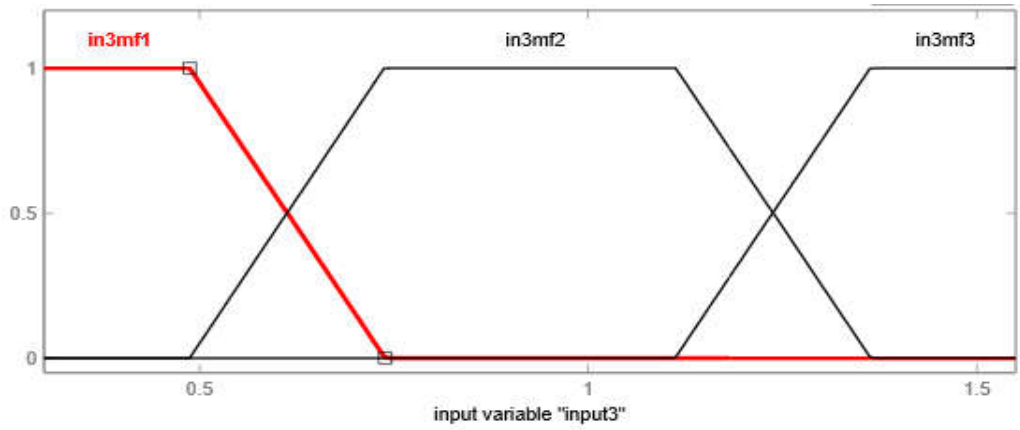

Figure 7.5: Membership Function of Input3, i.e., Anti-HBc or HBcAb

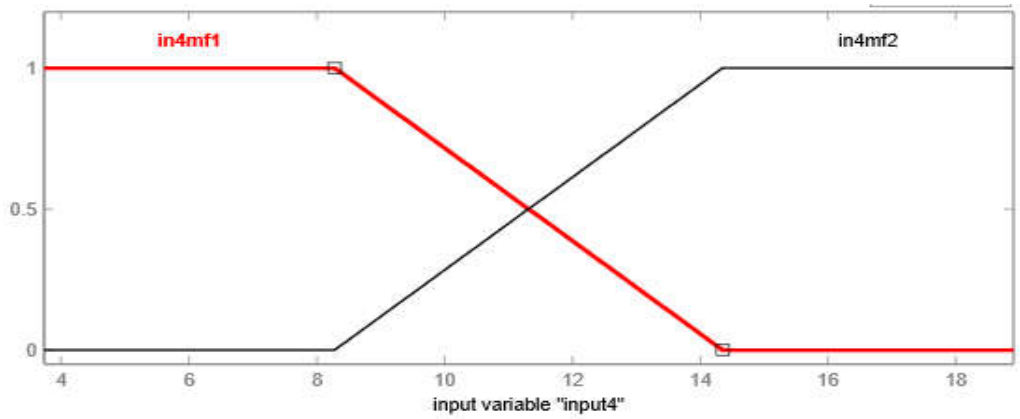

Figure 7.6: Membership Function of Input4, i.e., HBV DNA

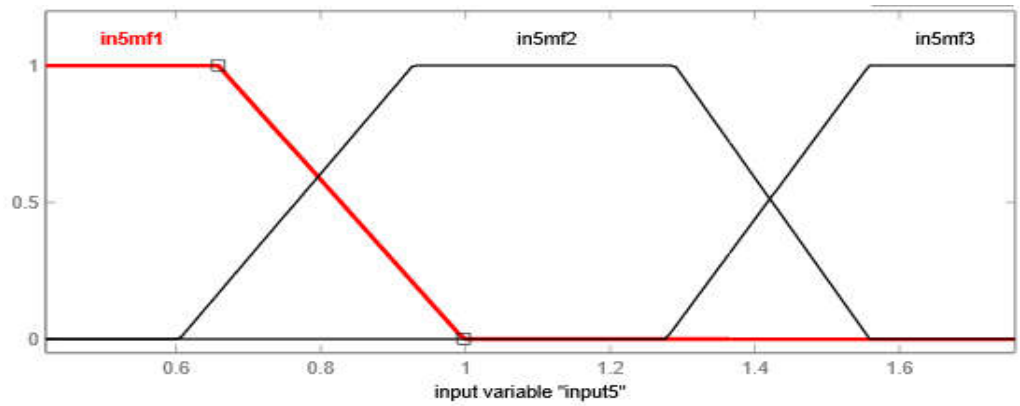

Figure 7.7: Membership Function of Input5, i.e., Anti-HBcAg-IgM

7.6 RULES

The rules are automatically generated by the adaptive neuro-fuzzy inference system from the training data set. Therefore, all the construction of membership functions is done by the ANFIS itself. This is the advantage of using ANFIS over a fuzzy inference system that the user does not define the rule as well as membership functions manually. Regardless, the number of rules, the ANFIS develops them from the dataset provided during the training phase. The aggregation of these stores various rules have been made by disjunction or conjunction. The form of the generated rules is the IF-THEN form in the case of the Mamdani system of fuzzy logic. The framework of generated rules is shown in figure 7.8.

These rules are stored in below mentioned manner

If (input variable (a))

And (input variable (b))

And (input variable (c))

And (input variable (d))

. . . . . .

. . . . . .

. . . . . .

Then output is (x)

Where a, b, c, d are linguistic variables of input variables and x is a linguistic variable of the output variable.

The automatically generated rules have an essential role in the development of an intelligent diagnostic system by using ANFIS. The accuracy of the classification and other performance parameters directly depends on these generated input-output rules. Hence, the data samples should be acquired accurately from the expert in order to make correct, appropriate as well as adequate input-output rules. Additionally, the data partitioning must be done accurately to acquire maximum accuracy. The number of rules of an intelligent system depends upon the number of linguistic variables taken in the input variables in that particular inference model.

The number of rules that are generated by the system by itself can be calculated as given below:

Total number of rules: number of membership functions in input variable 1 * number of membership functions in input variable 2 * number of membership functions in input variable 3 * number of membership functions in input variable 4 * number of membership functions in input variable 5

= input1 (3) * input2 (2) * input3 (3) * input4 (2) * input 5(3)

= 3 * 2 * 3 * 2 * 3 =108 rules

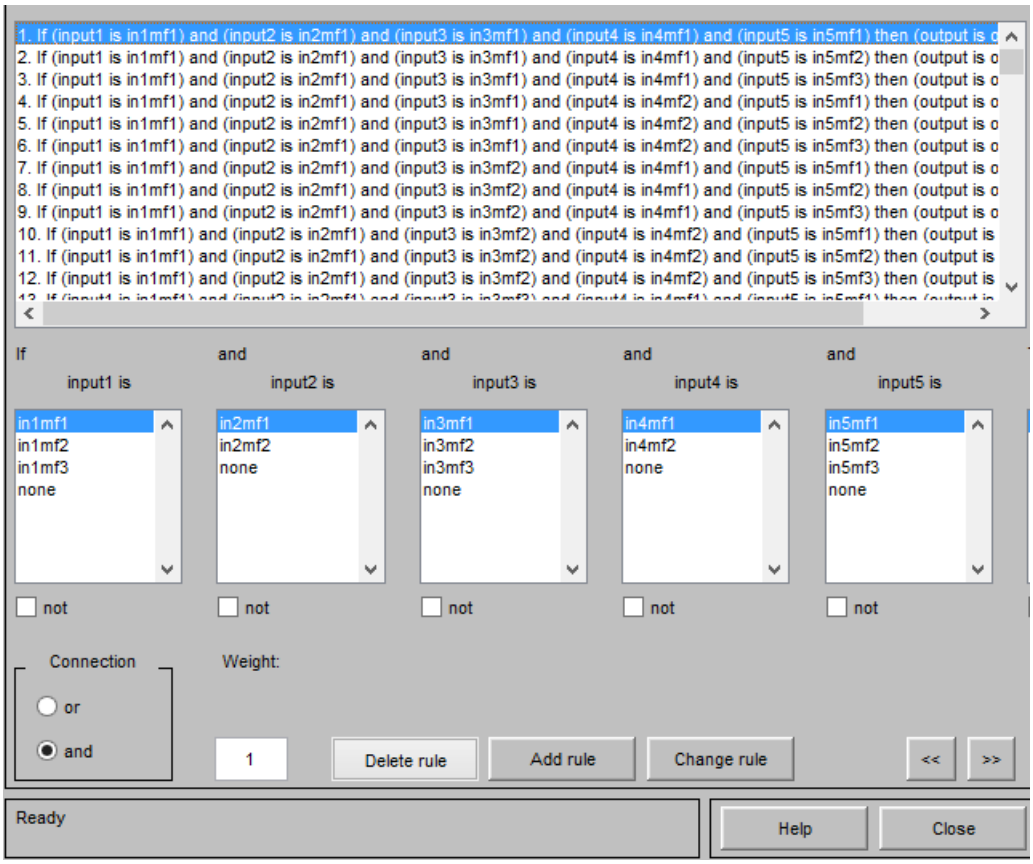

Figure 7.8: Framework of Rules in Developed Intelligent System

7.7 TRAINING

The training section elaborated the training process of a developed medical intelligent system by using ANFIS. Figure 7.9 demonstrates the error at 10 epochs. The original dataset has been categorized into two sets, namely, training dataset as well as testing dataset. There are several methods to categorize the corresponding dataset into the required sets. In this work, the k-fold cross-validation is utilized to group the data. Seventy five percent data from the original dataset is utilized in the phase of training to train the model. The training process assists the system in predicting the disease accurately. The speed of the diagnosis of the disease is also enhanced by using hybrid algorithms. Hence, as a result, the disease can be identified at an introductory stage, and the probability of saving the life of a patient also increased.

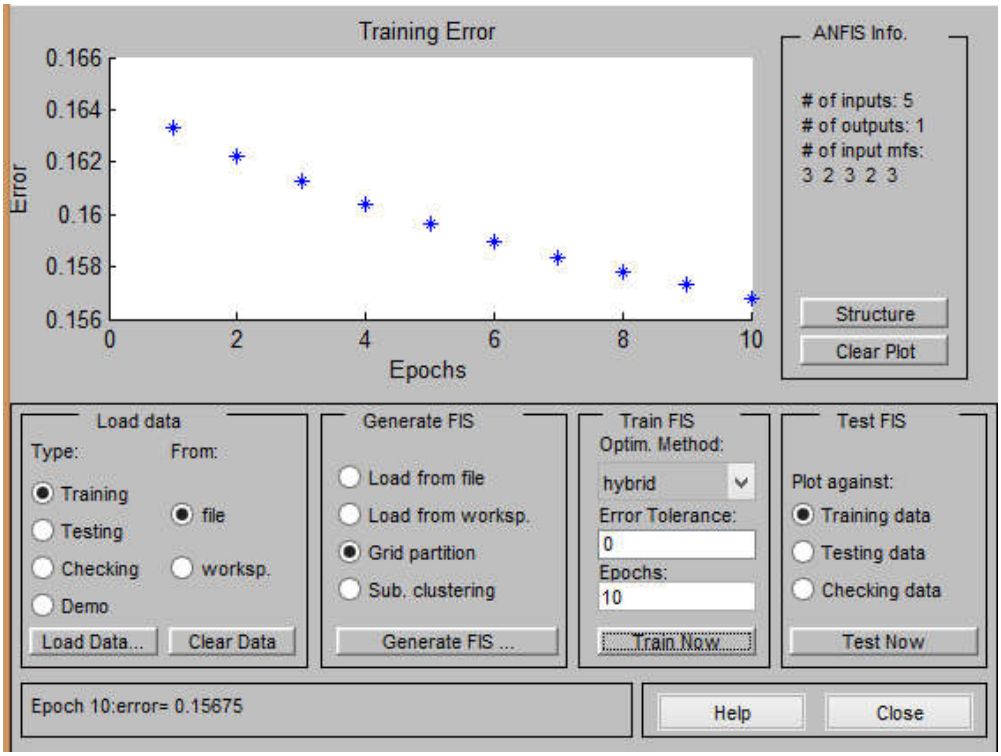

Figure 7.9: Training Error at 10 Epochs

7.8 TESTING AND VALIDATION

The testing of the proposed system has been done after its training process. It is necessary that the testing data and training data of the system are different from one another to validate the efficiency of the proposed system. Hence, the remaining 25% of data from the original dataset are considered as the testing dataset and also used to test as well as validate the model. In the validation process, the output of the developed system has been validated whether it classifies the given input into the accurate class or not. The proposed system must be capable of classifying the unseen data into the correct class.

7.9 SUMMARY

In this chapter, the developed adaptive neuro-fuzzy inference system (ANFIS) is explained. Neuro-fuzzy is a hybrid approach which is a combination of two soft computing techniques namely neural network and fuzzy logic. The developed system

is trained by using 75% of the available data set and remaining 25% data is used for testing the system. Total 108 rules are generated and all the membership functions are also represented with the help of various figures. During the validation process, the output of the developed system has been validated to check that whether it classifies the given input into the accurate class or not.

Chapter – 8

RESULTS AND DISCUSSION

Before evaluating the results for both developed models, i.e., medical diagnostic system by using fuzzy logic and medical intelligent system by using a neuro-fuzzy technique, it is necessary to understand some crucial terms and performance parameters first carefully. Hence, in this respective chapter, first the terms, as well as considered parameters to evaluate the performance, are discussed effectively and later the results for both developed system has been calculated.

8.1 PERFORMANCE EVALUATION

The performance of the developed model has been computed by considering various performance parameters. Before calculating this, there are some terms that are important to understand in order to do performance evaluation of any specific model. Those terms are well explained in this section.

8.1.1 Important Terms

Some of the important terms that are essential to comprehend for the evaluation of performance are explained below:

- **True positive (TP)**

The True Positive is defined as it is an output in which the developed system identify the positive class accurately as well as correctly. In this work, the test is considered as a true positive if the developed model accurately recognizes the infected person or patient.

- **False Negative (FN)**

The false negative is defined as an output in which the developed system identifies the negative class inaccurately or incorrectly. For example, if the developed fuzzy

inference system to diagnose HBV gives the output of a non-infected person as an infected person.

- **False Positive (FP)**

The false positive is defined as it is an output generated by the developed model in which the positive class is classified into an incorrect class. For instance, in this research work, if the model classifies an infected person into a non-infected class, then that particular test is considered into the false positive or FP.

- **True Negative (TN)**

The true negative is defined as it is an output provided by the developed system in which the negative class is predicted incorrectly or in the wrong class. The tests done on the developed inference system in which the non-infected patients are incorrectly classified are considered as the true negative or TN.

8.1.2 Performance Parameters

The various performance parameters that are taken under consideration to evaluate the performances of developed fuzzy expert or inference system or medical intelligent system by using a neuro-fuzzy technique for the diagnosis of hepatitis B disease are effectively elaborated below:

- **Sensitivity**

The sensitivity is the capability of a test of a developed model in order to recognize the infected patients with a disease accurately as well as correctly. The sensitivity of any particular model can be calculated by using below given formula.

$$Sensitivity = \frac{(TP)}{(TP + FN)}$$

Where TP is True positive

FN is False Negative

- **Specificity**

Specificity is the capability of a specific test of a developed system in order to recognize the non-infected patients accurately as well as correctly. The specificity of any particular model can be calculated by using below given formula.

$$Specificity = \frac{(TN)}{(TN + FP)}$$

Where TN is True Negative

FP is False Positive

- **Precision**

The precision of a developed model is a ratio of true positive with the sum of true positive tests and false positive tests. Mathematically, the precision can be defined as the below-written equation:

$$Precision = \frac{(TP)}{(TP + FP)}$$

Where TP is True Positive

FP is False Positive

- **Classification Accuracy**

The classification accuracy is the capability of a particular system to classify the infected as well as non-infected patients into the correct class. The classification accuracy of any system can be computed by using the below written formula:

$$Classification\ Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Where TP is True Positive

TN is True Negative

FP is False Positive

FN is false Negative

8.2 EXPERIMENTAL RESULTS FOR DEVELOPED FUZZY INFERENCE SYSTEM

To analyze whether the developed fuzzy inference system to diagnose HBV is classifying the provided data into accurate as well as appropriate classes, the different 50 tests are carried out. These tests are totally different from one another and also unseen by the system before this analysis. Meanwhile, these tests are also investigated by a group of professionals or expert doctors of this deadly disease and offered the correct result with respective samples. Later, the results provided by the developed fuzzy inference system and the evaluation done by the doctors are compared with one another to check whether the obtained output from the developed model is matched with the outcome offered by professionals or not. After this comparative study, it is observed that 47 tests out of 50 tests are accurately identified by the constructed fuzzy inference or expert system. In other words, the system shows 3 errors in the entire test phase.

Table 8.1 shows the confusion matrix of different test samples. As illustrated above, the total number of test samples is 50, and out of these, 47 are accurately classified into the considered classes. The 1st column represented the cases of No HBV. There is a total of 27 samples of those patients who are non-infected, and out of these samples, only one case is observed as classified into incorrect class. Likewise, the 2nd column represents the acute infection and having total 11 test cases. These 11 test cases are positively classified, and there is not any single case of acute infection which is classified incorrectly. Moreover, the thirds column indicates the chronic infection and having a total of 12 test samples. Out of these 12 samples, 2 are classified inaccurately, whereas 10 test cases are classified into the correct class by the developed model.

Table 8.1: Confusion Matrix

No HBV	Acute	Chronic	Class Name
26	1	0	**No HBV**
0	11	0	**Acute**
2	0	10	**Chronic**

Table 8.2 represents the confusion matrix, but the dimensions of it are reduced. The class name No HBV is considered into no class, and similarly, the other two classes, i.e. acute and chronic, are considered into yes class. Hence, the dimensionality of the matrix is reduced.

Table 8.2: Confusion Matrix with Reduced Dimensionality

No	Yes	Class name
26	1	**No**
2	21	**Yes**

From the above table, the various performance parameters that are discussed above are figured out.

Therefore, from table 8.2,

The value of True Positive, i.e., TP, is 21

The value of True Negative, i.e., TN, is 26

The value of False Negative, i.e., FN, is 2

The value of False Positive, i.e., FP, is 1

$$Sensitivity = \frac{(TP)}{(TP+FN)} = \frac{21}{21+02} = 91.20\%$$

$$Specificity = \frac{(TN)}{(TN+FP)} = \frac{26}{26+01} = 96.29\%$$

$$Precision = \frac{(TP)}{(TP+FP)} = \frac{21}{21+01} = 95.45\%$$

$$Classification\ Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} = \frac{21+26}{21+01+26+02} = 94\%$$

The calculated performance parameters, i.e., sensitivity, specificity, precision and classification accuracy, according to tests that have been done on the developed medical diagnostic system, are described in table 8.3. Further, the graphical representation of these parameters of performance is also shown in figure 8.1.

Table 8.3: Evaluated Performance Parameters for the Medical Diagnostic System by using Fuzzy Logic

Disease:	Hepatitis B
Model:	Multi-layered fuzzy logic
Classification Accuracy (%):	94 %
Specificity (%):	96.29 %
Sensitivity (%):	91.20 %
Precision (%):	95.45 %

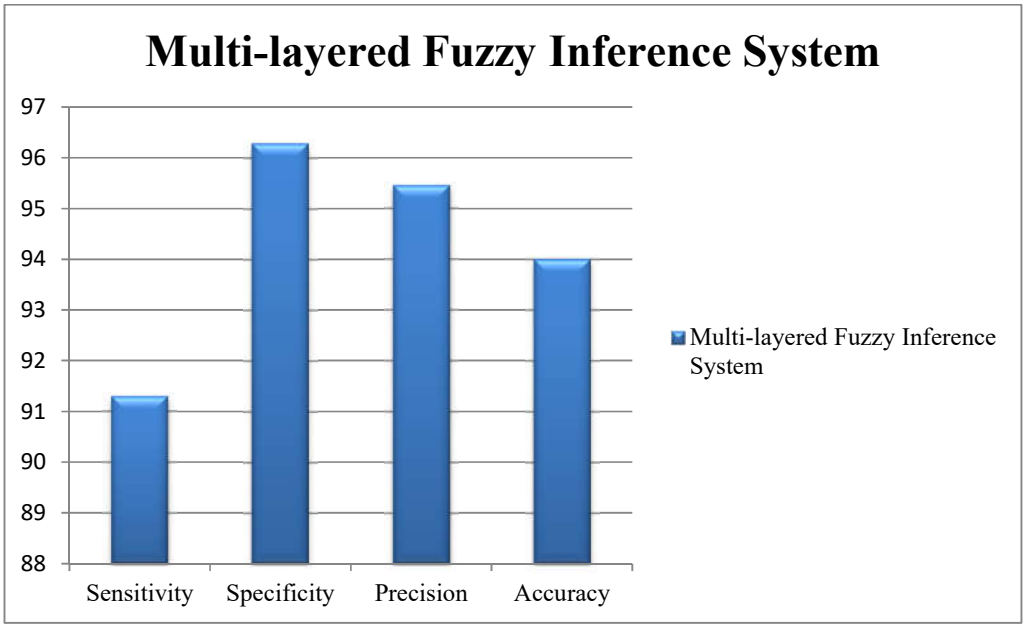

Figure 8.1: Graphical Representation of Performance Parameters of Multilayered Fuzzy Inference System

8.3 EXPERIMENTAL RESULTS FOR THE DEVELOPED MEDICAL INTELLIGENT SYSTEM BY USING A NEURO-FUZZY TECHNIQUE

The accuracy of the developed medical intelligent system for the diagnosis of hepatitis B disease has been evaluated by comparing the observed value from the system itself with the target values given by the professionals or experts. If the observed value and the target value are the same or approximately the same, then the proposed system is considered as correct, which means the system categorizes the infected and non-infected patient into accurate class otherwise not. After the investigation of the accuracy of the developed system to identify the hepatitis B disease, it is observed that the system classify the given input into the correct class.

The several parameters that assist in calculating the performance of the developed model are also determined, as shown in table 19 and figure 8.2 displays the graphical representation of it. The same performance parameters that are being calculated for multilayered fuzzy inference system are also evaluated to calculate the performance of ANFIS. The considered performance parameters are:

- Sensitivity
- Specificity
- Precision
- Classification Accuracy

The mathematical equations illustrated in the very first section of this chapter are used to calculate the value of each and every performance parameter.

Table 8.4 shows the confusion matrix of different test samples. As explained above, the total number of test samples is 50, and out of these, 48 are accurately classified into the considered classes. The 1st column represented the cases of No HBV. There is a total of 20 samples of those patients who are non-infected, and out of these samples, only one case is observed as classified into incorrect class. Likewise, the 2nd column

represents the acute infection and having total 19 test cases. Out of these 19 samples, 1 are classified inaccurately, whereas 18 test cases are classified into the correct class by the developed model. Moreover, the thirds column indicates the chronic infection and having a total of 9 test samples. These 9 test cases are positively classified, and there is not any single case of acute infection which is classified incorrectly

Table 8.4: Confusion Matrix

No HBV	Acute	Chronic	Class Name
20	1	0	**No HBV**
1	19	0	**Acute**
0	0	9	**Chronic**

Table 8.5 represent the confusion matrix, but the dimensions of it are reduced. The class name No HBV is considered into no class, and similarly, the other two classes, i.e. acute and chronic, are considered into yes class. Hence, the dimensionality of the matrix is reduced.

Table 8.5: Confusion Matrix with Reduced Dimensionality

No	Yes	Class name
20	1	**No**
1	28	**Yes**

From the above table, the various performance parameters that are discussed above are figured out.

Therefore, from table 8.5,

The value of True Positive, i.e., TP, is 28

The value of True Negative, i.e., TN, is 20

The value of False Negative, i.e., FN, is 1

The value of False Positive, i.e., FP, is 1

$$Sensitivity = \frac{(TP)}{(TP+FN)} = \frac{28}{28+01} = 96.55\%$$

$$Specificity = \frac{(TN)}{(TN+FP)} = \frac{20}{20+01} = 95.23\%$$

$$Precision = \frac{(TP)}{(TP+FP)} = \frac{28}{28+01} = 96.55\%$$

$$Classification\ Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} = \frac{20+28}{20+01+28+01} = 96\%$$

The values of all these parameters are presented in table 8.6 and the graphical representation of the same is shown in figure 8.2.

Table 8.6: Evaluated Performance Parameters for Medical Intelligent System by using a Neuro-Fuzzy Technique

Disease	Hepatitis B
Model:	Adaptive Neuro-Fuzzy Inference System (ANFIS)
Membership function:	Trapezoidal membership Function
Classification Accuracy (%):	96%
Sensitivity (%):	96.55%
Specificity (%):	95.23 %
Precision (%):	96.55%

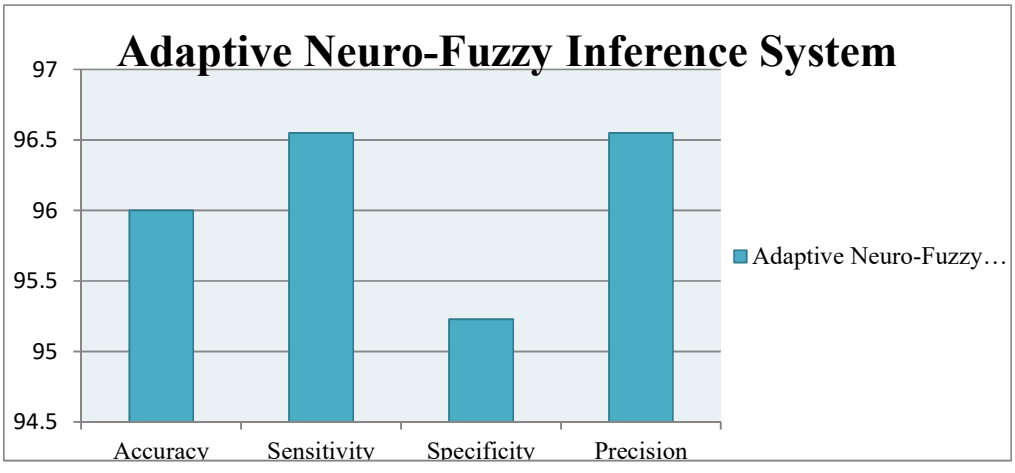

Figure 8.2: Graphical Representation of Performance Parameters of Adaptive Neuro-Fuzzy Inference System (ANFIS)

In the fuzzy logic based medical diagnostic system, the rules and membership functions are generated manually with the help of dataset but in case of neuro-fuzzy based medical diagnostic system, the rules and membership function are generated automatically in the training phase. These automatically generated rules and membership functions helps in producing better performance by neuro-fuzzy based medical diagnostic system as compared to fuzzy logic based medical diagnostic system.

8.4 SUMMARY

In this chapter, the results obtained from both the developed systems are discussed in detail. From the values of the confusion matrix, different parameters are calculated for both the developed systems. The accuracy, specificity, sensitivity and precision obtained using multilayered fuzzy inference system are 94%, 96.29%, 91.20% and 95.45% respectively. The percentage of same parameters using ANFIS are 96%, 96.55%, 95.23% and 96.55% respectively. The obtained results are also represented graphically.

Chapter – 9

COMPARISON OF RESULTS OF BOTH DEVELOPED INTELLIGENT SYSTEMS

In this respective chapter, the comparison of results of both developed intelligent system for the diagnosis of hepatitis B disease has been made. The performance parameters are compared with another in order to find out which developed intelligent system is better for the diagnosis of this life-threatening disease and offers accurate as well as appropriate results to the given corresponding inputs.

9.1 COMPARISON OF RESULTS

In order to figure out which developed model is best for the diagnosis of hepatitis B disease, a comparative study has been done among the fuzzy inference system as well as adaptive neuro-fuzzy inference system. The performance has been matched or compares with one another to make this comparison. By doing so, it is observed that which model has the capability to classify the provided inputs into accurate classes. The obtained result from both, i.e. fuzzy inference system as well as neuro-fuzzy inference system, are also compared with the results provided by the expert or professional doctor of hepatitis B disease.

After doing this comparative study, it is analyzed that the model that is developed by using the neuro-fuzzy technique has more accurate as well as correct results as compared to the system that is being developed by using fuzzy logic. In other words, the adaptive neuro-fuzzy inference system has more capability to classify the provided inputs adequately than the fuzzy inference system.

All the parameters of performance that are taken under consideration are also compared with each other of both systems. This comparison of the different parameters is made in table 9.1. The graphical representation of this comparative study makes it easy to understand the comparison between the performances of two developed systems. The two different charts are used to represent the analysis, i.e. column graph and bar chart. Figure 9.1 and 9.2 shows the graphical representation of this comparative study by using column graph and bar chart respectively.

Table 9.1: Comparison of Performances of Two Developed Models

Parameters / Models	Fuzzy Inference System	Adaptive neuro-Fuzzy inference system (ANFIS)
Classification Accuracy	94 %	96%
Sensitivity	91.20 %	96.55%
Specificity	96.29 %	95.23 %
Precision	95.45 %	96.55%

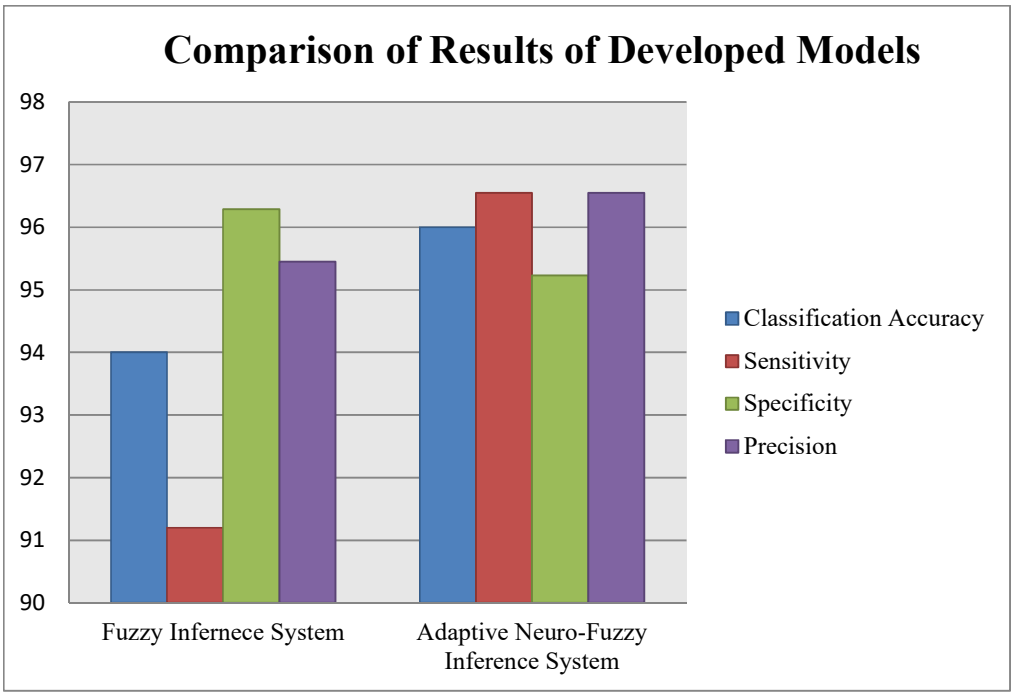

Figure 9.1: Graphical Representation of Comparison of Two Models by Column Graph

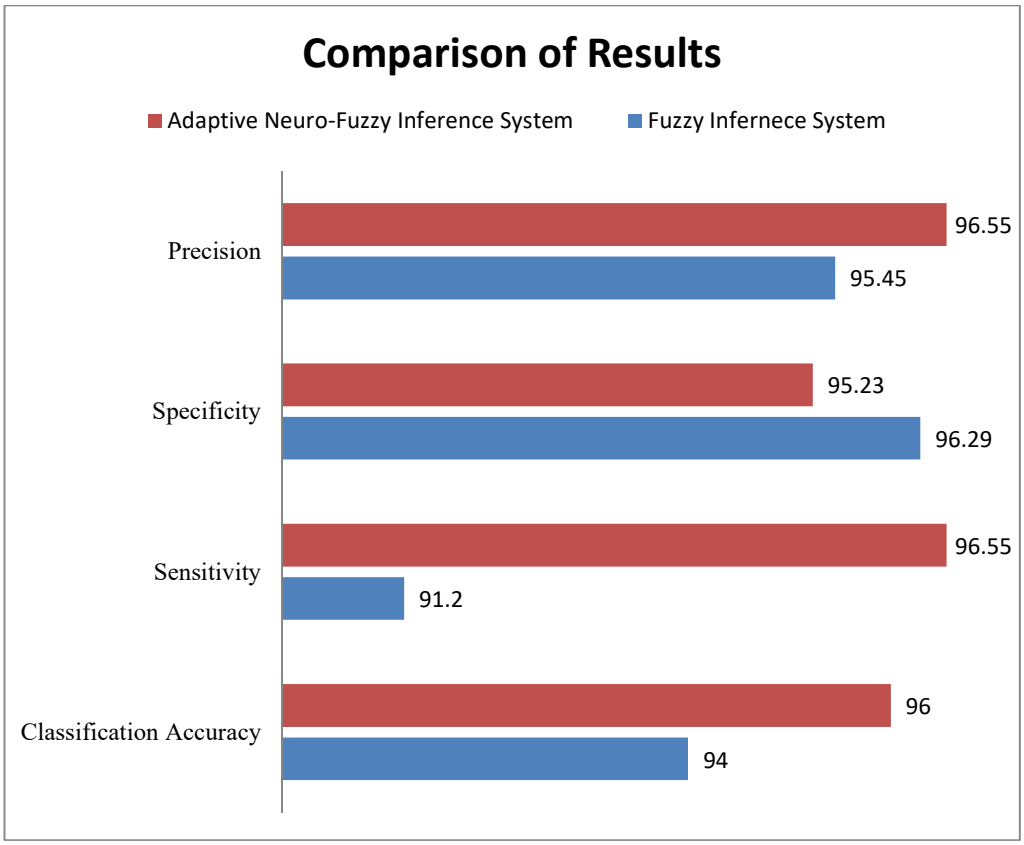

Figure 9.2: Graphical Representation of Comparison of Two Models by Bar Chart

Form the above comparative study; it is crystal clear that the adaptive neuro fuzzy technique has better accuracy whereas the fuzzy logic has least accuracy when compared with ANFIS. Now the accuracies of the already developed systems or existing systems are compared with these two developed models.

In table 9.2, the comparison of the developed models with existing ones is shown. From the table, it is observed that the developed models are more accurate as compared to other existing ones as the accuracy of the classification of these two developed models i.e. multilayered fuzzy inference system and adaptive neuro fuzzy inference system are more whereas the existing models have less accuracy. Among the proposed techniques, adaptive neuro-fuzzy technique gives more accuracy than fuzzy logic.

Table 9.2: Comparison of Existing Models with Developed Models

Sr. No.	Methodology	Classification Accuracy
1.	Support Vector Machine, Decision Rule and Decision Tree [25]	73.20%
2.	Random Forest [23]	85.15%
3.	Self-Organizing Map and Neuro-Fuzzy Inference System [67]	93.06%
4.	Proposed Multilayered Fuzzy Logic	**94%**
5.	Proposed Adaptive Neuro-fuzzy Technique	**96%**

Figure 9.3 represents the graphical representation of the comparison among developed and existing models. Here, the performance of the proposed system is high as compared to other systems because the proposed system is generated through training on real time dataset taken from the hospitals.

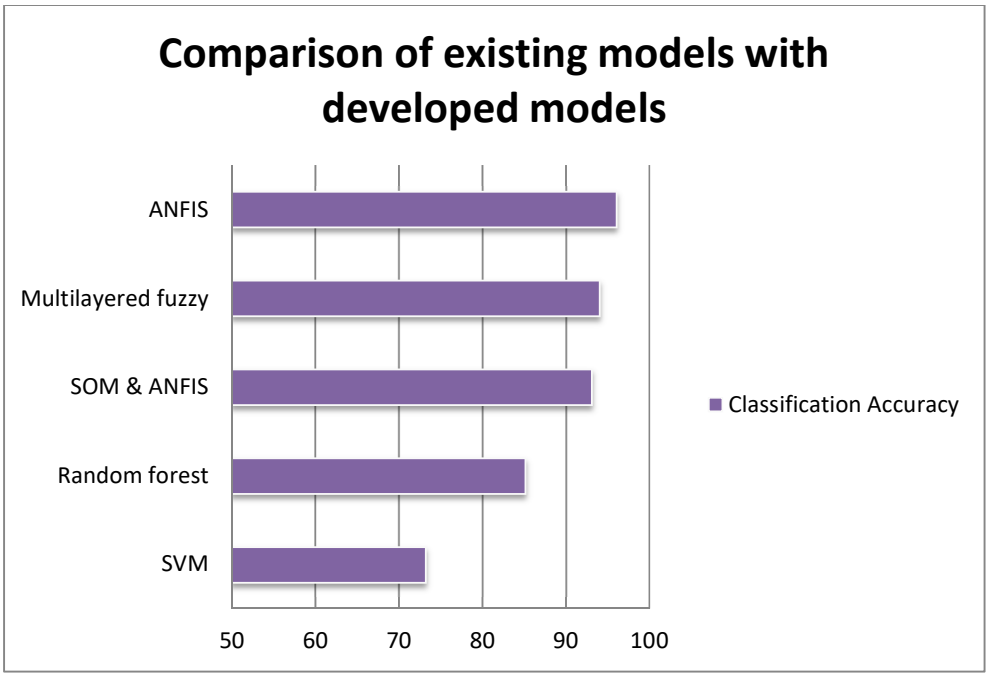

Figure 9.3: Comparison of Existing Models with Developed Models

9.2 SUMMARY

In this chapter, a detailed comparison between two proposed systems and three existing systems is done. The accuracy of three existing systems is 73.20%, 85.15% and 93.06% respectively. The accuracy obtained using FIS and ANFIS is 94% and 96% respectively. From the discussion, it is found that the proposed ANFIS produced the better accuracy as compared to FIS and other three existing systems. For better understanding, comparison is also represented using a graph.

Chapter – 10

CONCLUSION AND FUTURE SCOPE

In this chapter, we will discuss the conclusion of the whole research work and its future prospective.

10.1 CONCLUSION

The entire research study depends on the development of a medical diagnostic system to diagnose HBV by utilizing soft computing technologies, like fuzzy logic and neuro-fuzzy technique. Initially, the multilayered fuzzy inference system is proposed and developed. This system has two different layers for predicting hepatitis B disease. Both layers have different risk factors that are used to detect the selected deadly disease. The system classifies the non-infected as well as infected patients into different classes. This system can be used by the fresher doctors as well as the experts.

Further, an intelligent system has been proposed as well as developed by using the neuro-fuzzy technique. This developed model can assist the doctors as well as naïve individuals to identify the disease by themselves. The proposed system is initially trained in the training phase by utilizing the relevant dataset. After that, the testing and validation phase has been conducted for the system to evaluate the observed output provided by the system.

Additionally, the performance parameters are determined according to the outcome of the models. The considered parameters are classification accuracy, sensitivity, specificity and precision. A comparative study is also done to figure out which developed system is more accurate and which one is inaccurate as compare to another. The classification accuracy of the developed fuzzy inference system is 94%, and likewise, the classification accuracy of the adaptive neuro-fuzzy inference system is 96%. Hence, this performance evaluation concluded that the outcome given by the developed medical diagnostic system by using ANFIS is accurate and correct as compared to the developed fuzzy inference system and also can be used in hospitals

for the diagnosis of Hepatitis B disease. In other words, the adaptive neuro-fuzzy inference system has more capability to classify the provided inputs adequately than the fuzzy inference system. The main findings of this research work are:

1. A fuzzy inference system (FIS) for diagnosis of hepatitis B which has accuracy of 94%.
2. An adaptive neuro-fuzzy inference system (ANFIS) for diagnosis of hepatitis B which has accuracy of 96%.
3. The main finding is that the proposed ANFIS has produced the highest accuracy as compared the proposed FIS and three existing systems.

10.2 FUTURE SCOPE

In future, more biomarkers can be identified as well as other artificial intelligence methodologies can be applied to develop a decision-making model which assists the doctors as well as inexpert to diagnose the deadly and life-threatening hepatitis B disease. Moreover, the input variables can be changed by researching the risk factor more accurately. The membership functions can also be replaced by other types of membership functions in order to increase the performance as well as the accuracy of the system.

REFERENCES

[1] C. Long, H. Qi and W. Peng, "System Dynamics of Chronic Hepatitis B - Modeling the Virus and Immune System of Chronic Hepatitis B," 2010 2nd International Workshop on Intelligent Systems and Applications, Wuhan, China, 2010, pp. 1-4, doi: 10.1109/IWISA.2010.5473320.

[2] T. Inoue and Y. Tanaka, "Hepatitis B virus and its sexually transmitted infection – an update," *Microbial Cell*, vol. 3, no. 9, pp. 419–436, 2016.

[3] W. Shen, Z. Wei, Y. Li, H. Zhang and P. Liu, "Multiple Granular Analysis of TCM Data with Applications on Diagnosis of Hepatitis B," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 2015, pp. 2920-2925, doi: 10.1109/SMC.2015.508.

[4] M. C. Marketplace, Hepatitis B. (2020, September 04). Retrieved September 22, 2020, from https://www.mayoclinic.org/diseases-conditions/hepatitis-b/symptoms-causes/syc-20366802

[5] S. Das, K. Ramakrishnan, S. K. Behera, M. Ganesapandian, A. S. Xavier, and S. Selvarajan, "Hepatitis B Vaccine and Immunoglobulin: Key Concepts," *Journal of Clinical and Translational Hepatology*, vol. 7, no. X, pp. 1–7, 2019.

[6] Y.-F. Liaw and C.-M. Chu, "Hepatitis B virus infection," *The Lancet*, vol. 373, no. 9663, pp. 582–592, 2009.

[7] L. S. Tang, E. Covert, E. Wilson, and S. Kottilil, "Chronic Hepatitis B Infection," *JAMA*, vol. 319, no. 17, p. 1802, 2018.

[8] S. Tsukuda and K. Watashi, "Hepatitis B virus biology and life cycle," *Antiviral Research*, vol. 182, pp. 104–925, 2020.

[9] T. J. Liang, "Hepatitis B: The virus and disease," *Hepatology*, vol. 49, no. S5, 2009.

[10] P. A. Higgins, "Hepatitis B virus," *Journal of the American Academy of Physician Assistants*, vol. 29, no. 8, pp. 48–49, 2016.

[11] V. M. Taylor, J. H. Choe, Y. Yasui, L. Li, N. Burke, and J. C. Jackson, “Hepatitis B Awareness, Testing, and Knowledge among Vietnamese American Men and Women,” *Journal of Community Health*, vol. 30, no. 6, pp. 477–490, 2005.

[12] Y. Qiu, J. Ren, and J. Yao, “Healthy adult vaccination: An urgent need to prevent hepatitis B in China,” *Human Vaccines & Immunotherapeutics*, vol. 12, no. 3, pp. 773–778, 2015.

[13] Y. Bai, X. Chen, C. Dong, Y. Liu and Z. Zhang, "A comparison of multimodal biomarkers for chronic hepatitis B assessment using recursive feature elimination," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 2016, pp. 2448-2451, doi: 10.1109/EMBC.2016.7591225.

[14] C. Cheng et al., "Discovering Clinical Biomarkers of Chronic Hepatitis B by Mining Mutation Hotspots," 2011 International Conference on Technologies and Applications of Artificial Intelligence, Chung Li, Taiwan, 2011, pp. 51-56, doi: 10.1109/TAAI.2011.17.

[15] L. Du, D. Huang and Q. Xie, "A mathematical model for acute hepatitis B virus infection," 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China, 2010, pp. 1109-1113, doi: 10.1109/BMEI.2010.5639617.

[16] S. U. Emon, T. I. Trishna, R. R. Ema, G. I. H. Sajal, S. Kundu and T. Islam, "Detection of Hepatitis Viruses Based on J48, KStar and Naïve Bayes Classifier," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944619.

[17] S. Liu, B. Zhou, J. D. Valdes, J. Sun, and H. Guo, “Serum Hepatitis B Virus RNA: A New Potential Biomarker for Chronic Hepatitis B Virus Infection,” *Hepatology*, vol. 69, no. 4, pp. 1816–1827, 2019.

[18] K. Leung et al., "Data Mining on DNA Sequences of Hepatitis B Virus," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 8, no. 2, pp. 428-440, March-April 2011, doi: 10.1109/TCBB.2009.6.

[19] H. Yasin, T. A. Jilani, and M. Danish, “Hepatitis C Classification using Data Mining Techniques,” *International Journal of Computer Applications*, vol. 24, no. 3, pp. 1–6, 2011.

[20] L. Muflikhah, Widodo, W. F. Mahmudy and Solimun, "DNA Sequence of Hepatitis B Virus Clustering Using Hierarchical k-Means Algorithm," 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 2019, pp. 1-4, doi: 10.1109/ICETAS48360.2019.9117565.

[21] A. Bustamam, H. Tasman, N. Yuniarti, Frisca, and I. Mursidah, “Application of k-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV),” *AIP Conference Proceedings*, vol. 1862, no. 1, 2017.

[22] B. Liu, S. Feng, X. Guo and J. Zhang, "Bayesian analysis of complex mutations in HBV, HCV, and HIV studies," in Big Data Mining and Analytics, vol. 2, no. 3, pp. 145-158, September 2019, doi: 10.26599/BDMA.2019.9020005.

[23] H. Wang, Y. Liu and W. Huang, "Random forest and Bayesian prediction for Hepatitis B virus reactivation," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, 2017, pp. 2060-2064, doi: 10.1109/FSKD.2017.8393087.

[24] X. Chen, L. Ma, N. Chu and Y. Hu, "Diagnosis based on decision tree and discrimination analysis for chronic hepatitis b in TCM," 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), Atlanta, GA, USA, 2011, pp. 817-822, doi: 10.1109/BIBMW.2011.6112478.

[25] S. Uhmn, D. Kim, S. W. Cho, J. Y. Cheong and J. Kim, "Chronic Hepatitis

Classification Using SNP Data and Data Mining Techniques," 2007 Frontiers in the Convergence of Bioscience and Information Technologies, Jeju, Korea (South), 2007, pp. 81-86, doi: 10.1109/FBIT.2007.64.

[26] C. Mahesh, K. Kiruthika and M. Dhilsathfathima, "Diagnosing hepatitis B using artificial neural network based expert system," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 2014, pp. 1-7, doi: 10.1109/ICICES.2014.7033938.

[27] G G. S. Uttreshwar and A. A. Ghatol, "Hepatitis B Diagnosis Using Logical Inference And Generalized Regression Neural Networks," 2009 IEEE International Advance Computing Conference, Patiala, India, 2009, pp. 1587-1595, doi: 10.1109/IADCC.2009.4809255.

[28] M. Rouhani and M. M. Haghighi, "The Diagnosis of Hepatitis Diseases by Support Vector Machines and Artificial Neural Networks," 2009 International Association of Computer Science and Information Technology - Spring Conference, Singapore, 2009, pp. 456-458, doi: 10.1109/IACSIT-SC.2009.25.

[29] H. Wang, Y. Liu and W. Huang, "The application of feature selection in Hepatitis B virus reactivation," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 893-896, doi: 10.1109/ICBDA.2017.8078767.

[30] S. Shariati and M. M. Haghighi, "Comparison of anfis Neural Network with several other ANNs and Support Vector Machine for diagnosing hepatitis and thyroid diseases," 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), Krakow, Poland, 2010, pp. 596-599, doi: 10.1109/CISIM.2010.5643520.

[31] G. Ahmad, M. A. Khan, S. Abbas, A. Athar, B. S. Khan, and M. S. Aslam, “Automated Diagnosis of Hepatitis B Using Multilayer Mamdani Fuzzy Inference System,” *Journal of Healthcare Engineering*, vol. 2019, pp. 1–11, 2019.

[32] H. Ahmadi, M. Gholamzadeh, L. Shahmoradi, and M. Nilashi, “Computer Methods and Programs in Biomedicine Diseases diagnosis using fuzzy logic methods : A systematic and meta-analysis review,” *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 145–172, 2018.

[33] V. Novák, “Which logic is the real fuzzy logic?,” *Fuzzy Sets and Systems*, vol. 157, no. 5, pp. 635–641, 2006.

[34] S. Das and P. K. Ghosh, “Hypertension Diagnosis : A Comparative Study using Fuzzy Expert System and Neuro Fuzzy System,” *IEEE International Conference on Fuzzy Systems*, no. 2005, pp. 1–7, 2013.

[35] I. Fernando, F. Henskens, and M. Cohen, “A domain specific expert system model for diagnostic consultation in psychiatry,” *Proceedings - 2011 12th ACIS International Conference on Software Engineering, Artificial Intelligence Networking and Parallel Distributed Computing, SNPD 2011*, pp. 3–6, 2011.

[36] S. Sivathasan, F. Cecelja, and W. Balachandran, “ECG diagnosis using neural network and fuzzy expert system,” *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference [Cat. No. 00CH37066]*, vol. 2, pp. 988–992, 2000.

[37] I. B. De Medeiros, M. A. Soares Machado, W. J. Damasceno, A. M. Caldeira, R. C. Dos Santos, and J. B. Da Silva Filho, “A Fuzzy Inference System to Support Medical Diagnosis in Real Time,” *Procedia Computer Science*, vol. 122, pp. 167–173, 2017.

[38] E. Rovenţa and G. Roşu, “The diagnosis of some kidney diseases in a small Prolog Expert System,” *Proceedings - 2009 3rd International Workshop on Soft Computing Applications, SOFA 2009*, vol. 620, pp. 219–224, 2009.

[39] G. Fattovich, “Natural history of hepatitis B,” *Journal of Hepatology*, vol. 39, pp. 50–58, 2003.

[40] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, “Brain Intelligence: Go beyond

Artificial Intelligence," *Mobile Networks and Applications*, vol. 23, no. 2, pp. 368–375, 2017.

[41] A. McGovern, K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, "Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather," *Bulletin of the American Meteorological Society*, vol. 98, no. 10, pp. 2073–2090, 2017.

[42] R. Clarke, "Why the world wants controls over Artificial Intelligence," *Computer Law & Security Review*, vol. 35, no. 4, pp. 423–433, 2019.

[43] M. Selvamanikkam, "Introduction to Artificial Intelligence", *IBM Watson Analytics*, 2019.

[44] S. Dick, "Artificial Intelligence," *Harvard Data Science Review*, vol. 1, no. 1, 2019.

[45] M. Bielli, *Artificial intelligence applications to traffic engineering*. Utrecht, The Netherlands: VSP, 1994.

[46] R. Shah and A. Chircu, "IOT AND AI IN HEALTHCARE: A SYSTEMATIC LITERATURE REVIEW," *Issues In Information Systems*, vol. 19, no. 3, pp. 33–41, 2018.

[47] H. Conick, "The Past, Present and Future of AI in Marketing," *American Marketing Association*, 12-Dec-2019. [Online]. Available: https://www.ama.org/marketing-news/the-past-present-and-future-of-ai-in-marketing/. [Accessed: 24-Jul-2021].

[48] A. Haleem, R. Vaishya, M. Javaid, and I. H. Khan, "Artificial Intelligence (AI) applications in orthopaedics: An innovative technology to embrace," *Journal of Clinical Orthopaedics and Trauma*, vol. 11, pp. SS80–SS81, 2020.

[49] I. J. Rudas and J. Fodor, "Intelligent systems," *International Journal of Computers, Communications & Control*, vol. 3, no. 3, pp. 132–138, 2008.

[50] J. Liebowitz, "Expert systems: A short introduction," *Engineering Fracture*

Mechanics, vol. 50, no. 5-6, pp. 601–607, 1995.

[51] C. Grosan and A. Abraham, “Rule-Based Expert Systems,” *Intelligent Systems Reference Library*, vol. 17, pp. 149–185, 2011.

[52] L. A. Zadeh, "Fuzzy logic," in Computer, vol. 21, no. 4, pp. 83-93, April 1988, doi: 10.1109/2.53.

[53] L. A. Zadeh, “Is there a need for fuzzy logic?,” *Information Sciences*, vol. 178, no. 13, pp. 2751–2779, 2008.

[54] M. Hellmann, “Fuzzy logic introduction,” *Université de Rennes*, vol. 1, pp. 1–9, 2001.

[55] K. Tanaka, *An introduction to fuzzy logic for practical applications*. New York: Springer, 1997.

[56] O. Terrada and O. Bouattane, “A fuzzy medical diagnostic support system for cardiovascular diseases diagnosis using risk factors,” *2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pp. 1–6, 2018.

[57] S.-C. Wang, “Artificial neural network,” in *Interdisciplinary computing in java programming*, Boston, MA: Springer, 2003, pp. 81–100.

[58] S. E. Dreyfus, “Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure,” *Journal of Guidance, Control, and Dynamics*, vol. 13, no. 5, pp. 926–928, 1990.

[59] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.,” *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[60] S. Mangrulkar, “Artificial neural systems,” *ISA Transactions*, vol. 29, no. 1, pp. 5–7, 1990.

[61] J. M. Zurada, “Introduction to Artrificial Neural Network” , vol. 8, St. Pole: West, 1992.

[62] B. B. Ekici and U. T. Aksoy, “Expert Systems with Applications Prediction of building energy needs in early stage of design by using ANFIS,” *Expert Systems With Applications*, vol. 38, no. 5, pp. 5352–5358, 2011.

[63] E. Areghan and S. Konyeha, “A Genetic-Neural System Diagnosing Hepatitis B,” *Global Journal of Computer Science and Technology*, 2019.

[64] M. Thakur, “DESIGNING OF APPLIED ARTIFICIAL INTELLIGENCE BASED INTELLIGENT SYSTEM FOR DIAGNOSIS OF HEPATITIS B,” *Journal on Intelligent Systems & Robotics Insights & Transformations [ISSN: 2581-5636 (online)],* vol. 2, no. 1, 2018.

[65] Y. Chen *et al.*, “Machine-learning-based classification of real-time tissue elastography for hepatic fi brosis in patients with chronic hepatitis B,” *Computers in biology and medicine*, vol. 89, no. January, pp. 18–23, 2017.

[66] R. Wei *et al.*, “Clinical prediction of HBV and HCV related hepatic fi brosis using machine learning,” *EBioMedicine*, 2018.

[67] M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, and E. Akbari, “A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique,” *Journal of Infection and Public Health*, vol. 12, no. 1, pp. 13–20, 2019.

[68] S. Khan, R. Ullah, A. Khan, R. Ashraf, H. Ali, M. Bilal, and M. Saleem, “Analysis of hepatitis B virus infection in blood sera using Raman spectroscopy and machine learning,” *Photodiagnosis and Photodynamic Therapy*, vol. 23, pp. 89–93, 2018.

[69] Q.-H. Ye, L.-X. Qin, M. Forgues, P. He, J. W. Kim, A. C. Peng, R. Simon, Y. Li, A. I. Robles, Y. Chen, Z.-C. Ma, Z.-Q. Wu, S.-L. Ye, Y.-K. Liu, Z.-Y. Tang, and X. W. Wang, “Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning,” *Nature Medicine*, vol. 9, no. 4, pp. 416–423, 2003.

[70] A. J. Mueller-Breckenridge, F. Garcia-Alcalde, S. Wildum, S. L. Smits, R. A. de Man, M. J. van Campenhout, W. P. Brouwer, J. Niu, J. A. Young, I. Najera, L. Zhu, D. Wu, T. Racek, G. B. Hundie, Y. Lin, C. A. Boucher, D. van de Vijver, and B. L. Haagmans, "Machine-learning based patient classification using Hepatitis B virus full-length genome quasispecies from Asian and European cohorts," *Scientific Reports*, vol. 9, no. 1, 2019.

[71] M. Walker, "Hep B Cure Research: The Long Road Ahead," *Medical News*, 03-Jan-2020. [Online]. Available: https://www.medpagetoday.com/infectiousdisease/hepatitis/84211. [Accessed: 22-May-2021].

[72] P. A. Revill *et al.*, "A global scientific strategy to cure hepatitis B," *The Lancet Gastroenterology & Hepatology*, vol. 4, no. 7, pp. 545–558, 2019.

[73] J. Mugisha, J. Mokaya, D. Bukenya, F. Ssembajja, D. Mayambala, R. Newton, P. C. Matthews, and J. Seeley, "A study of knowledge, experience and beliefs about hepatitis B virus (HBV) infection in south western Uganda," *Front Public Health*, vol. 7, 2019.

[74] "Finding a cure for hepatitis B: are we close?," *World Health Organization*, 12-Jun-2018. [Online]. Available: https://www.who.int/hepatitis/news-events/hbv-cure-overview/en/. [Accessed: 22-May-2021].

[75] Hepatitis B: Is a Cure Possible?," *Hepatitis B Foundation | Baruch S. Blumberg Institute*, 2016. [Online]. Available: https://www.hepb.org/news-and-events/free-newsletters/hepatitis-b-is-a-cure-possible/. [Accessed: 22-May-2021].

[76] "Strategic plan for trans-NIH research to cure Hepatitis B," 2019. [Online]. Available: https://www.niaid.nih.gov/sites/default/files/Trans-NIH-Hep-B-Strategic-Plan-2019.pdf. [Accessed: 2021].

[77] T. Yiu, "The Curse of Dimensionality," *Medium*, 20-Jul-2019. [Online]. Available: https://towardsdatascience.com/the-curse-of-dimensionality-50dc6e49aa1e. [Accessed: 22-May-2021].

[78] A. Uberoi, “Introduction to Dimensionality Reduction,” *GeeksforGeeks*, 08-Feb-2018. [Online]. Available: https://www.geeksforgeeks.org/dimensionality-reduction/. [Accessed: 22-May-2021].

[79] J. T. Raj, “Dimensionality Reduction for Machine Learning,” *Medium*, 14-Mar-2019. [Online]. Available: https://towardsdatascience.com/dimensionality-reduction-for-machine-learning-80a46c2ebb7e. [Accessed: 22-May-2021].

[80] L. Li, “Principal Component Analysis for Dimensionality Reduction,” *Medium*, 27-May-2019. [Online]. Available: https://towardsdatascience.com/principal-component-analysis-for-dimensionality-reduction-115a3d157bad. [Accessed: 22-May-2021].

[81] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016.

ANNEXURE

LIST OF PUBLICATIONS

S.No.	Title of paper with author names	Name of journal / conference	Status	Issn no/ vol no, issue no	Indexing in Scopus/ Web of Science
1.	***A Comprehensive Review of Intelligent Medical Diagnostic Systems***	Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)	Published	IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0	Scopus
2.	A Neuro-fuzzy based Medical Intelligent System for the Diagnosis of Hepatitis B	2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)	Published	978-1-7281-9491-2	Scopus
3.	A Multi-layered Fuzzy Inference System for the	Turkish Online Journal of Qualitative Inquiry (TOJQI)	Published	Volume 12, Issue 4, Month 2021: -582-592	Scopus

S.No.	Title of paper with author names	Name of journal / conference	Status	Issn no/ vol no, issue no	Indexing in Scopus/ Web of Science
	Diagnosis of Hepatitis B				
4.	A Statistical Review on Machine Learning based Medical Diagnostic Systems	*Journal of Cardiovascular Disease Research*	Published	ISSN:0975-3583,0976-2833 VOL12,ISSUE05,2021	Scopus
5.	Intelligent Medical Diagnostic Systems for Hepatitis B	Journal of Healthcare Engineering, Hindawi	Reviewed and under process	NA	SCI