LIVER CANCER PATIENT AGE PREDICTION USING CLASSIFICATION COMPARISON AND FEATURE SELECTION

This Report Presented in Partial Fulfillment of the Requirements for The Degree of Bachelor of Science in Electrical and Electronics Engineering

By

MD. SABBIR AHAMED
ID: 183-33-4865

Supervised By

MR. JAHEDUL ISLAM
Lecturer Department of EEE
Daffodil International University

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
FACULTY OF ENGINEERING

DAFFODIL INTERNATIONAL UNIVERSITY

JANUARY 2022

The project and thesis entitled “**LIVER CANCER PATIENT AGE PREDICTION USING CLASSIFICATION COMPARISON AND FEATURE SELECTION**” submitted by **Md. Sabbir Ahamed**, ID No: 183-33-4865, Session: Fall 2018 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of **Bachelor of Science in Electrical and Electronic Engineering** on 27 September 2022.

BOARD OF EXAMINERS

-- Chairman

-- External Member

-- Internal Member

-- Coordinator

DECLARATION

I hereby declare that this thesis was fulfilled under the supervision of **Mr. Jahedul Islam, Lecturer, Department of EEE** at Daffodil International University. I also declare that this thesis and its portions have not been submitted to any other university to issue a degree.

Submitted by:

Md. Sabbir Ahamed
ID: 183-33-4865
Department of EEE
Daffodil International University

Supervised by:

Mr. Jahedul Islam
Lecturer
Department of EEE
Faculty of Science and Engineering
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I sincerely thank and praise Allah Almighty for his divine mercy, which also helped me complete the thesis I was working on for my final year.

I am immensely thankful and wish to express my sincere gratitude to **Mr. Jahedul Islam, Department of EEE,** Daffodil International University. His unending patience, constant encouragement, and also friendly guidance enabled me to finish this thesis.

I also wish to express my sincere respect to **Dr. Md. Shahid Ullah, Professor, and Head of the Department of EEE,** in addition to the other faculty and staff members from the EEE department, for their assistance in completing my thesis.

In Addition, I want to express my thankfulness to every one of my classmates who took part in this discussion when performing the prescribed coursework.

Finally, I want to thank my parents for their unwavering support and patience.

ABSTRACT

Cancer is one of the deadliest diseases in the world. Cancer spreads to the liver, bones, breast, lungs, and other organs. Hepatitis B and C, liver cirrhosis, and every one of these factors contribute to the growth of liver cancer. The deadliest type of cancer is liver cancer, which is a lifetime condition. Based on WHO studies, there are about 30 cases of liver cancer per 100,000 individuals, with the majority of them originating in Asian and African nations first. The rate where the cancer patients increasing is so worrying, particularly in Bangladesh. The most popular form of liver cancer, hepatocellular carcinoma, or HCC affects more men than women. It has become a frequent condition in modern times. A growth called liver cancer forms in the liver tissue as a result of the liver cells' uncontrolled cell cycle. In this paper, we have explored the age prediction system for patients with hepatocellular carcinoma, a type of liver cancer. Based on the corresponding liver cancer diagnosis and other pertinent factors, we develop a neural network model which primarily focuses on the recognition of a certain age group.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

CHAPTER 1

INTRODUCTION

1.1 Introduction

The second most significant cause of death worldwide is cancer, accounting for 9.6 mortality globally in 2018. Every year, approximately 1.5 lakh people in Bangladesh develop cancer. Outside of Dhaka, the country has just four specialized cancer hospitals [1].

Inside the upper right quadrant of your belly, below the abdomen and above the stomach, is an organ called the liver. The liver is vulnerable to developing a variety of cancer types. Hepatocytes, a healthy type of liver cell, are also where hepatocellular carcinoma, a much more prevalent form of liver cancer, begins [2]. Cancer generally spreads to the liver more often than it does when it first develops in the liver's cells. HCC risk is 100 times higher than that of an uninfected person. Smoking and alcohol consumption and hepatitis B infection both increase the risk of HCC. Aflatoxin and inherited conditions are two less common risk factors.

There are approximately 15 lakh cancer survivors in Bangladesh, while 1.5 lakh people pass away from the disease each year, according to estimates provided by the World Health Organization.

In Bangladesh, 26% of households consumed a catastrophic amount of treatment in the last two years. The cancer had the highest rate of CHE (50%), followed by hepatic disorders (49.2%) and paralysis (43.6%) [3]. Although it is not possible to completely eradicate cancer from a nation, data mining can significantly lessen its severity. This study's main objective is to determine the age range in Bangladesh at which the risk of developing cancer is highest.

The essay is structured as follows for the remaining portions. Explanations of the analysis of related works are provided in Section II. Again, Sections III and IV respectively offer our working method, experimental findings, and observations Finally, in Section V, we offer our concluding thoughts as well as the breadth and directions of our next research.

1.2 Motivation

With 142 million inhabitants, Bangladesh is the seventh most densely populated nation in the world. In Bangladesh, there are between 13 and 15 million cancer patients, with over 2 lakh persons getting a cancer diagnosis each year. Because it is constantly growing, it poses a severe threat to our nation.

However, scientifically oriented cancer research in Bangladesh is inadequate. Therefore, we were encouraged to work on it in "Machine Learning." We hope to assist people by raising awareness of cancer through this research. Furthermore, it will enhance the medical field.

1.3 Research Question

The age of the patient is the primary focus of our research. Age is the single most significant cancer risk factor. Half of all cancers develop at age 66 and older, while the risk gradually increases after age 50. The median age at diagnosis for liver cancer is 67 years for men and 72 years for women.

1.4 Research Methodology

We used two types of feature selection and three types of Classifications.

Feature Selection is:

1. Feature Importance Ranking Measure
2. Recursive Feature Elimination

Classifications are:

1. Logistic Regression
2. Support Vector Machine
3. Convolutional Neural Network

1.5 Research Objective

1.5.1 Liver Cancer prediction in Bangladesh.

1.5.2 People can know the possibility of Liver cancer through Cause, Age, and Gender.

1.5.3 As our goal is to build a model which can predict Liver cancer so people can take precautionary steps to protect themselves.

1.6 Report Layout

This section briefly summarizes each progression's elements, which we used in this thesis.

Chapter 1: Introduction

Chapter 2: Background

Chapter 3: Research Methodology

Chapter 4: Result & Observation

Chapter 5: Conclusion and Future work

CHAPTER 2

BACKGROUND

2.1 Review Works

In 2017 Prakash and Ramkumar used the WEKA tool and the Bayes theory to work on a Conditional Probability Bayes theorem. When applied to the data set values and the graph above, the results reveal that while the 17 patients who underwent the test did not develop liver cancer, the three people who drink more alcohol will almost certainly develop liver cancer [4].

In the same year in 2017, another team works on liver cancer. Using genomic data, they design a neural network to identify patients with liver cancer into high-risk and low-risk subgroups. Their method offers a fresh way to categorize large data sets utilizing models of neural networks. The data is preprocessed using wavelet analysis and singular value decomposition before compression [5].

In 2018 in Taiwan, the rate of incidence of liver cancer from 1997 to 2014. In 2004 (157.6 cases per 100,000 population). The rate is expected to fall by 22.2% between 2014 and 2025, and by 37.3% between 2014 and 2035 [6].

After 1-year Atrayee and Aditya developed an Image Processing Method for Detecting Liver Cancer [7]. In 2019 Arjmand and T. Angelis worked on Liver Biopsies using Convolutional Neural Networks and their performance eventually had a classification accuracy was 95% [8].

In the same year, Huang and Chow work for HCC Detection Using Convolutional Network, on the HCC tumor detection task, their approach achieved 91% mIOU [9]. Later in 2020, Saroj reviewed methods using deep learning and machine learning to identify liver cancer, and they discovered that the most effective algorithms provide up to 92% accuracy when based on CT scan images [10].

In the past year, Rajesh and Choudhury have been using machine learning to identify HCC liver cancer. They also attempt to apply additional machine learning algorithms and deep learning techniques like ANN and CNN to obtain the best outcomes [11]

2.2 Bangladesh Perspective

Bangladesh is a densely populated country. In Bangladesh, there are 13 to 15 lakh cancer patients, with roughly half a lakh people receiving a cancer diagnosis each year. As enough research is not available in this field, cancer prediction can be a big deal in the engineering field from Bangladesh's perspective.

2.3 Government Goals and Regulation

The government goal should be to decrease cancer's growth rate gradually every year. Through regulation, education, and support programs, governments will promote an environment where people maintain sensible amounts of physical exercise, a healthy body weight, and a sensible diet. Cancer interference and therefore the creation of a culture of health is a vital mission of government, on the far side that of the normal health-focused departments like health ministries; it's within the domain of government agencies concerned with environmental protection, activity safety, and transportation. Cancer interference and health promotion are within the realm of the board, the board of education, and therefore the board of health.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Working procedure

In this research, we try to identify the age group that is most likely to develop liver cancer for a wide range of reasons. We use a genuine dataset obtained from "Ahsania Mission Cancer and General Hospital" to accomplish this task. Our dataset contains information on 117 patients, including their age, gender, liver cancer type, pertinent cause, and other crucial details. Three distinct parts make up our total working method. The stages are, in order, feature selection, data classification, and data preprocessing. Figure 1 illustrates the general specifics.

Table-3.1: Range of Values Related to the Dataset

Data Variable	Observed Values
Gender	Female, Male
Area	Tongi, Kamarpara, Abdullahpur etc.
Alpha_1-Antitrypsin Deficiency_(AATD)	No, Yes
Non-Alcoholic_Steatohepatitis_(NASH)	No, Yes
Hepatitis_B_Virus_(HBV)	No, Yes
Smoking_Tobacco	No, Yes
Due_to_Genetic_Disorders	No, Yes
Hepatitis_C_Virus_(HCV)	No, Yes
Presence_of_Arsenic	No, Yes
Type_2_Diabetes	No, Yes
Due_to_Obesity	No, Yes
Exposure_to_Certain_Chemicals	No, Yes
Due_to_Advancing_Age	No, Yes
Glycogen_Storage_Diseases	No, Yes
Presence_of_Epstein_Barr_Virus	No, Yes
Abdominal_Swelling	No, Yes
Hereditary_Hemochromatosis	No, Yes
Transjugular_Intrahepatic_Portosystemic_Shunt_(TIPS)	No, Yes
Autoimmune_Hepatitis_(AIH)	No, Yes

Data Variable	Observed Values
Due_to_Tyrosinemia	No, Yes
Hepatic_Encephalopathy	No, Yes
High_Alcohol_Consumption	No, Yes
Presence_of_Antineoplastic_Agents	No, Yes
Presence_of_Iron_Storage_Disease	No, Yes
Cirrhosis	No, Yes
Wilson_Disease	No, Yes
Due_to_Gallstones	No, Yes
Primary_Sclerosing_Cholangitis_(PSC)	No, Yes
Primary_Biliary_Cirrhosis_(PBC)	No, Yes
Shortage_of_Aflatoxin_B1	No, Yes
Use_Anabolic_Steroids	No, Yes
Variceal_Hemorrhage	No, Yes
Porphyria_Cutanea_Tarda	No, Yes
Due_to_Autoimmune	No, Yes
Diagnosis	Non-Alcoholic Steatohepatitis (NASH), High Alcohol Consumption, Hepatitis B Virus (HBV), Hepatitis C Virus (HBV), etc.
Age	1 to 9, 10 to 19, 20 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, Above 80

3.1.1 Flow chart of working procedure

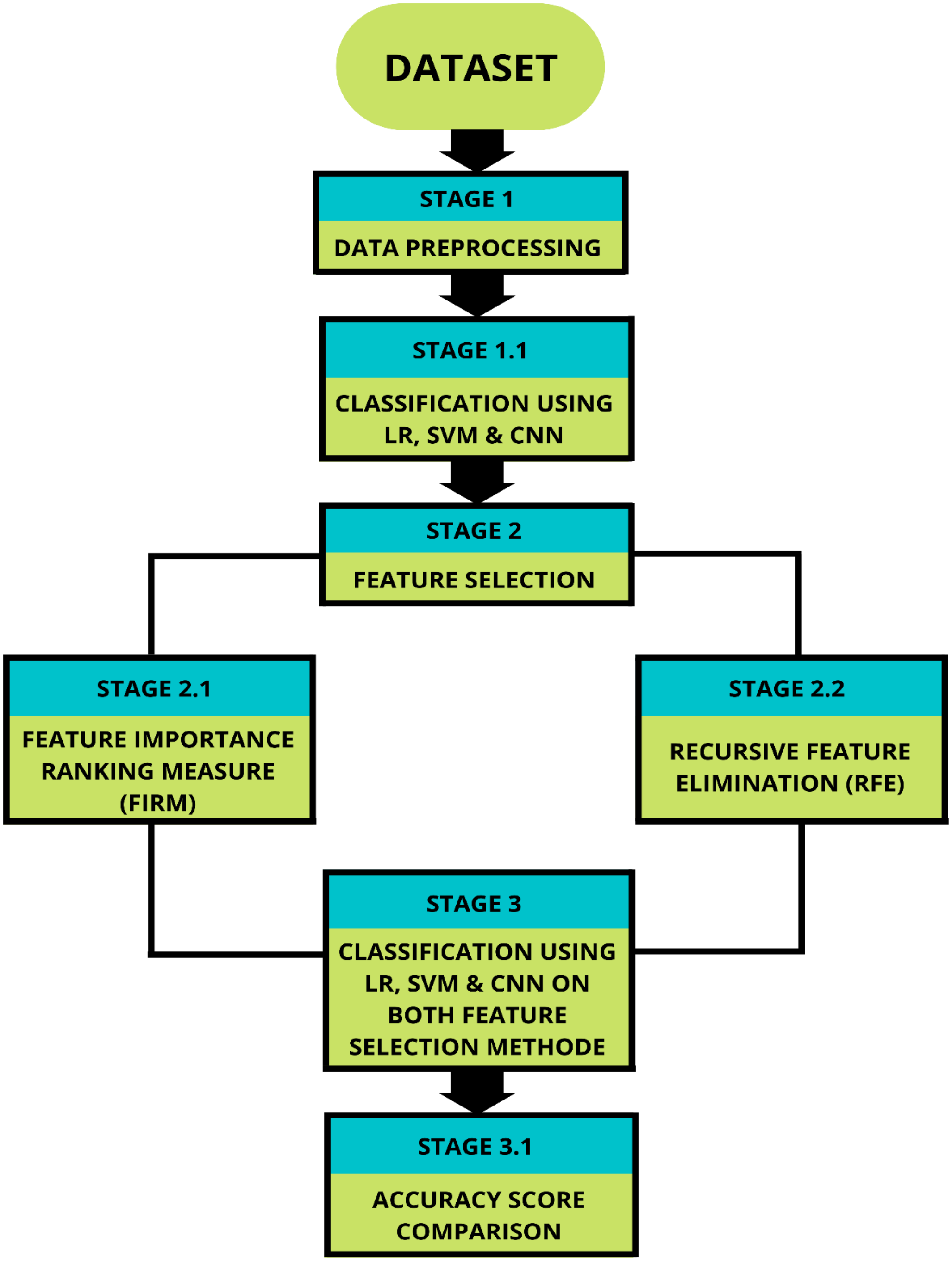

Figure-3.1: Stages of Working Procedure

3.1.2 Preprocessing of Data

Data preprocessing system the data into a format that is more effectively processed. In order to assure reliable findings, seven different classes make up the output class; these are highlighted in table III.

3.1.3 Feature Selection

By using only significant data and making an effort to eliminate noise from the data, the feature selection method reduces the input variable to our model. Depending on the kind of issue our attempting to resolve, it entails automatically selecting features for our machine learning model that are pertinent. Our main feature selection techniques in this study were the Recursive Feature Elimination and Feature Importance Ranking Measure.

3.1.3.1 Feature Importance Ranking Measure

Accuracy is generally obtained via learning algorithms with complex feature spaces. The Feature Importance Ranking Measure takes into account the features underlying correlation structure (FIRM). As a result, it may find the most pertinent characteristics even when noise entirely prevents their emergence in the training data [12]. In feature significance techniques, each input feature in a certain model is given a score that simply indicates the "importance" of that feature. If a character has a higher score, it suggests that it will have a greater influence on the prediction model for that particular variable.

3.1.3.2 Recursive Feature Elimination

This is a method for choosing features that reduce the complexity of a model by selecting significant features and eliminating weaker ones. The selection process eliminates these less important features one by one until the optimum number is reached to ensure peak performance. RFE uses filter-based feature selection internally and has a wrapper-style approach to feature selection. It begins with all of the characteristics in the training dataset and then identifies a subset of them. Prioritizing features, removing the least important characteristics, and re-fitting the model are the methods used to achieve this.

3.1.4 Classification

Support Vector Machine and Logistic Regression are two baseline classifiers that we used in our research together with Convolutional Neural Networks (CNN). These classifiers were used both during the pre-feature and post-feature selection stages.

3.1.4.1 Logistic Regression

Predictive analytics and categorization usually employ this statistical model, commonly known as the logit model. The likelihood of an occurrence, either voting or not voting, is assessed using logistic regression and a variety of independent factors [13]. Considering that the output is a probability, this dependent variable's value is 0 to 1. The odds in logistic regression, which are the probability of success divided by the likelihood of failure, are subject to the logit transformation. This is known as either log odds and or odds' natural logarithm [14]. Figure 2 depicts the use of the diagrammatic technique in logistic regression. Here's how the logistic function is defined:

$$logistic(\eta) = \frac{1}{1+exp\ (-\eta)} \qquad \text{—— (1)}$$

Logistic regression follows linear regression, it is easy to transition. Equation 2 illustrates how we used a linear equation in the regression model to express the connection between the features and the outcome.

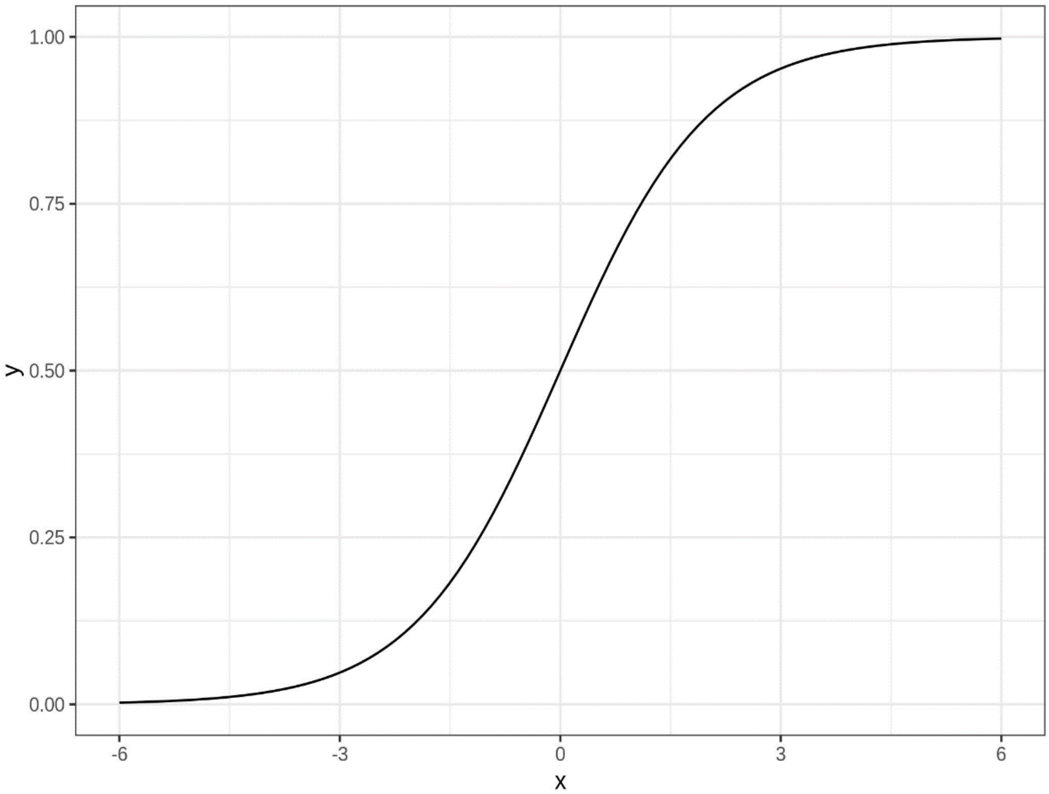

Figure-3.2: Logistic Regression.

$$\hat{y}^{(i)} = \beta_0 + \beta_1 x_1{}^{(i)} + \ldots + \beta_p x_p{}^{(i)} \quad \text{—— (2)}$$

We use the logistic function to integrate the appropriate part of the equation since we prefer variations between 0 and 1 for classification. The output is constrained to only take numbers between 0 and 1.

$$P(y^{(i)} = 1) = \frac{1}{1 + exp\ (-(\beta_0 + \beta_1 x_1{}^{(i)} + \ldots + \beta_p x_p{}^{(i)}))} \quad \text{—— (3)}$$

3.1.4.2 Support Vector Machine

A separating hyperplane function as the proper definition of the classifier recognized as a Support Vector Machine. In other words, the algorithm generates the appropriate hyperplane for sorting fresh examples given training data with labels (supervised learning) [15]. A hyperplane is a two-dimensional line that divides a plane into two sections, with each category being on each side of the line. Boundaries known as hyperplanes are useful for categorizing information points. If a data point is placed along either side of the hyperplane, it will be split into entirely distinct categories. Position and direction both of the hyperplanes are impacted by support vectors that are closer to it. We attempt to increase the classifier's margin when we use support vectors.

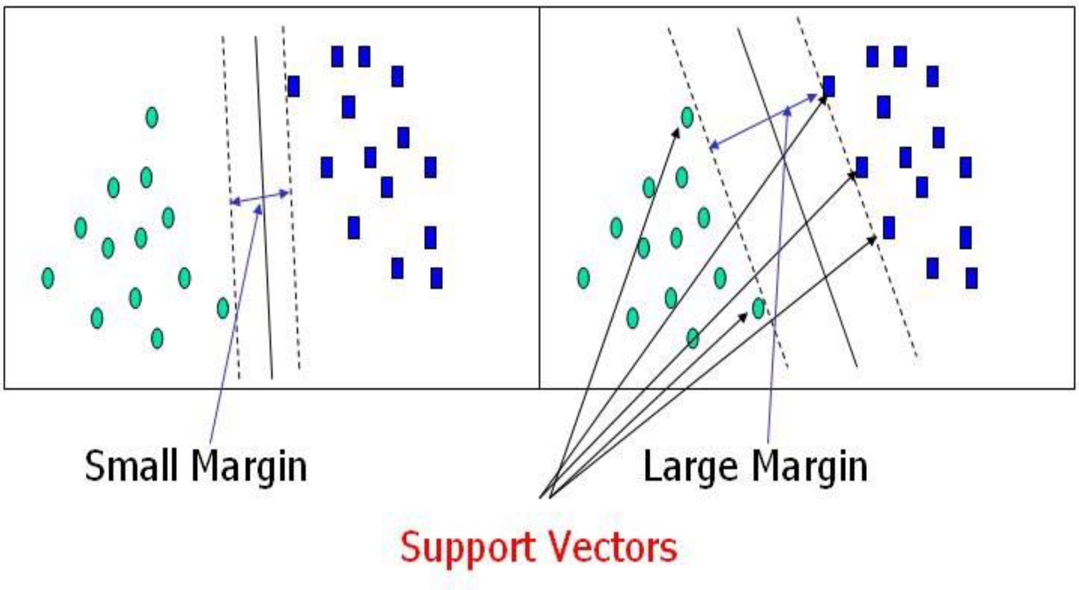

Figure-3.3: Support Vector Machine.

3.1.4.3 Convolutional Neural Network

Deep learning algorithms are based on neural networks, a type of machine learning. There are several types of neural nets, each with its own set of use cases and data types. CNN (or convolutional neural networks) are increasingly being used for classification and computer vision tasks [16].

Neurons that are associated with local areas or receptive fields in the input are generated in the convolution layer. Max pooling is one of the subsampling techniques. Its goal is to obtain an input representation by reducing its proportions, which aids in decreasing overfitting [17].

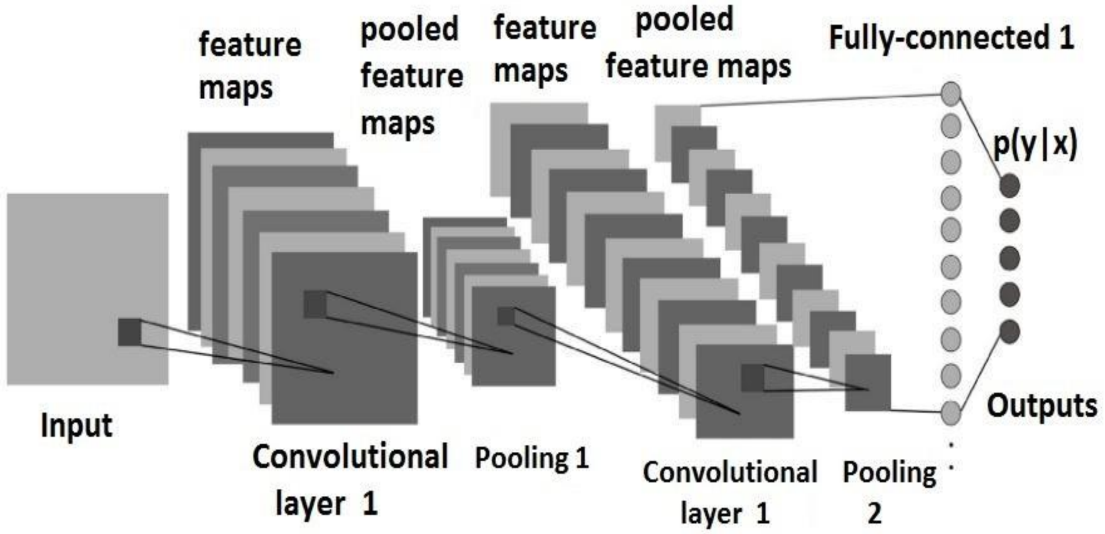

Figure-3.4: Convolutional Neural Network.

CHAPTER 4

RESULT AND OBSERVATIONS

4.1 Experimental Analysis

We evaluated our experimental findings using the most common measures for classifier performance. Accuracy, recall, F1-score, assist, and accuracy are just a few examples. Each of these metrics is represented mathematically as follows:

$$Precision = \frac{truepositive}{truepositive + falsepositive} \quad \text{——(4)}$$

$$Recall = \frac{truepositive}{truepositive + falsenegative} \quad \text{——(5)}$$

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{——(6)}$$

$$Accuracy = \frac{truepositive + truenegative}{(truepositive + truenegative + falsepositive + falsenegative)} \quad \text{——(7)}$$

4.2 Comparative Analysis

Before using feature selection techniques, we implement Logistic Regression, Support Vector Machine, and Convolutional Neural Networks. In these cases, we find that LR performs both SVM and CNN. The outcome analysis before the feature selection stage is thoroughly broken down in table number II.

Table-4.1: Classification Report before Feature Selection

Classifiers	Class	Precision	Recall	F1-score	Support	Accuracy
LR	0	0.00	0.00	0.00	0	59.04
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.50	0.50	0.50	2	
	4	0.67	0.86	0.76	8	
	5	0.52	0.53	0.51	7	
	6	1.00	0.49	0.64	2	
SVM	0	0.00	0.00	0.00	0	45.90
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.00	0.00	0.00	2	
	4	0.50	0.73	0.59	8	
	5	0.57	0.57	0.57	7	
	6	0.00	0.00	0.00	2	
CNN	0	0.00	0.00	0.00	0	50.00
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.00	0.00	0.00	2	
	4	0.56	0.88	0.73	8	
	5	0.57	0.57	0.57	7	
	6	0.00	0.00	0.00	2	

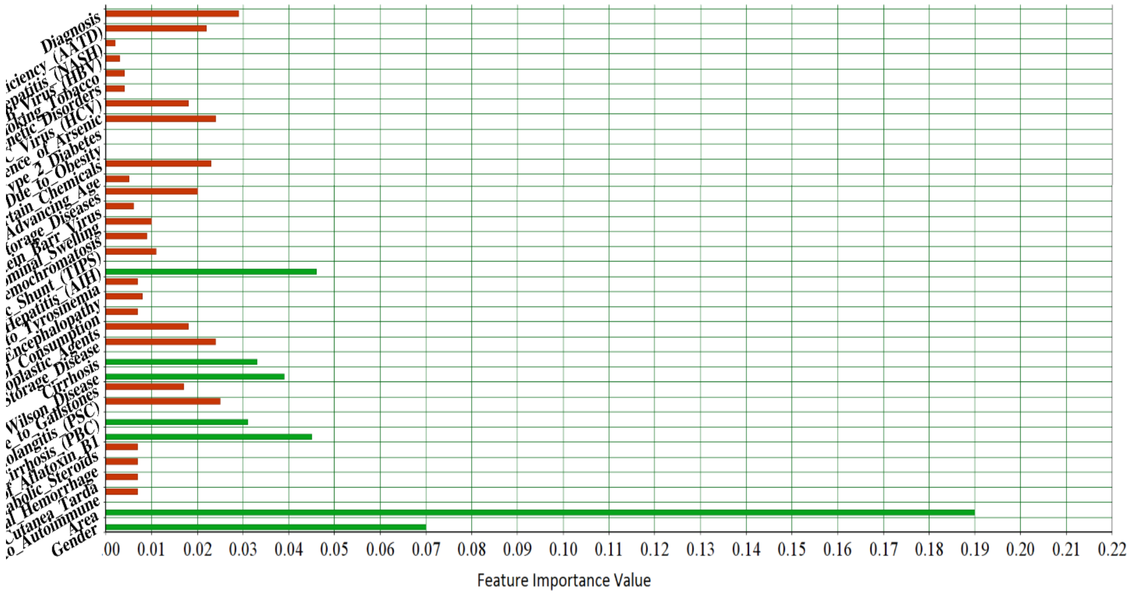

Figure-4.1: Features Selected after Applying FIRM.

In terms of accuracy, CNN outperforms LR and SVM when the FIRM feature selection method is used. During FIRM application, extracted features are shown in Figure 5, and a thorough explanation of the outcome is shown in Table III. Last and not least, as seen in table IV, the RFE method, every single classifier outperformed the FIRM technique. As a result, the RFE-extracted features are much more relevant to furthering the cause.

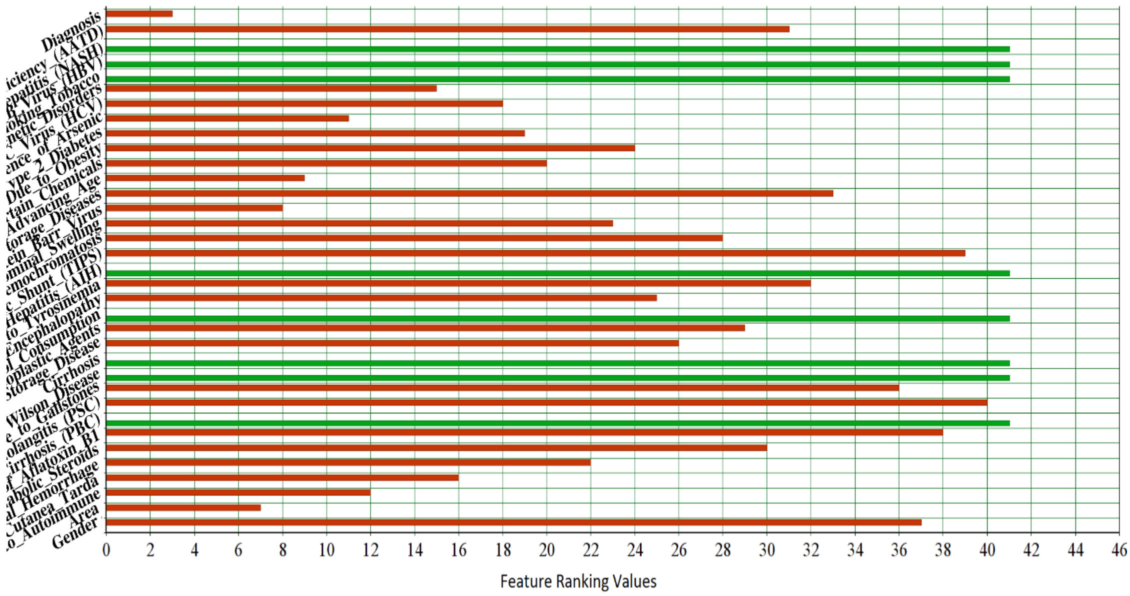

Figure-4.2: Features Selected after Applying RFE.

Finally, after implementing the RFE technique, all of the classifiers significantly outperform the FIRM technique.

Table-4.2: Classification Report after FIRM

Classifiers	Class	Precision	Recall	F1-score	Support	Accuracy
LR	0	0.00	0.00	0.00	0	45.47
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.41	0.84	0.57	2	
	4	0.59	0.42	0.51	8	
	5	0.00	0.00	0.00	7	
	6	0.00	0.00	0.00	2	
SVM	0	0.00	0.00	0.00	0	40.97
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.00	0.00	0.00	2	
	4	0.50	0.63	0.50	8	
	5	0.50	0.56	0.54	7	
	6	0.00	0.00	0.00	2	
CNN	0	0.00	0.00	0.00	0	50.00
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0. 00	0.00	0.00	2	
	4	0.43	0.86	0.59	8	
	5	0.81	0.56	0.68	7	
	6	0. 00	0.00	0.00	2	

Table-4.3: Classification Report after RFE

Classifiers	Class	Precision	Recall	F1-score	Support	Accuracy
LR	0	0.00	0.00	0.00	0	54.49
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.50	0.50	0.50	2	
	4	0.63	0.87	0.76	8	
	5	0.60	0.43	0.50	7	
	6	0.25	0.51	0.32	2	
SVM	0	0.00	0.00	0.00	0	45.38
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.00	0.00	0.00	2	
	4	0.63	0.87	0.72	8	
	5	0.34	0.56	0.42	7	
	6	0.00	0.00	0.00	2	
CNN	0	0.00	0.00	0.00	0	59.07
	1	0.00	0.00	0.00	2	
	2	0.00	0.00	0.00	1	
	3	0.50	0.50	0.50	2	
	4	0.69	0.88	0.79	8	
	5	0.68	0.57	0.63	7	
	6	0.25	0.51	0.32	2	

Table IV illustrates this. As a result, the RFE-extracted features play a larger role in supporting the claim. Non-Alcoholic Steatohepatitis (NASH), Hepatitis B Virus (HBV), Smoking Tobacco, Autoimmune Hepatitis (AIH), High Alcohol Consumption, Cirrhosis, Wilson Disease, Primary Biliary Cirrhosis (PBC) are the extracted features here.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In Bangladesh, where 49.2% of liver cancer patients reside, if we did not diagnose before a particular stage, liver cancer is unavoidable. By increasing awareness of cancer’s early warning signs, there may be a reduction in the number of liver cancer sufferers.

Using data mining techniques, we attempted to identify the signs of various liver cancer types in Bangladesh for distinct age groups. To do this, we employed two feature selection techniques: Three separate classification techniques are used: 1. Feature Importance 2. Recursive Feature Elimination 3. Vector Machine Neural Network with Convolutions benefit from Logistic Regression. The Convolutional Neural Network provided the best performance.

5.2 Future Work

Our purpose is to investigate the regional variations in the situation and intensity of liver cancer. Furthermore, trying to manage a real-time dataset was difficult for us. As a result, we intend to evaluate the model's accuracy on a huge dataset in the future.

REFERENCES

[1] Chandan, M., 2022. Cancer treatment: Still a long way to go. [online] The Daily Star. Available at: <https://www.thedailystar.net/opinion/perspective/news/cancer-treatment-bangladesh-still-long-way-go-1696912> [Accessed 10 September 2022].

[2] Mayo Clinic. 2022. Liver cancer - Symptoms and causes. [online] Available at: <https://www.mayoclinic.org/diseases-conditions/liver-cancer/symptoms-causes/syc-20353659> [Accessed 10 September 2022].

[3] Tajmim, T., 2022. When health cost is a catastrophe for a family. [online] The Business Standard. Available at: <https://www.tbsnews.net/bangladesh/health/when-health-cost-catastrophe-family-485750> [Accessed 10 September 2022].

[4] Ramkumar, N., Prakash, S., Kumar, S.A. and Sangeetha, K., 2017, January. Prediction of liver cancer using Conditional probability Bayes theorem. In 2017 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.

[5] Zeinalzadeh, A., Wenska, T. and Okimoto, G., 2017, May. A neural network model to classify liver cancer patients using data expansion and compression. In the 2017 American Control Conference (ACC) (pp. 2135-2139). IEEE.

[6] Su, S.Y., Chiang, C.J., Yang, Y.W. and Lee, W.C., 2019. Secular trends in liver cancer incidence from 1997 to 2014 in Taiwan and projection to 2035: An age-period-cohort analysis. Journal of the Formosan Medical Association, 118(1), pp.444-449.

[7] Dutta, A. and Dubey, A., 2019, April. Detection of liver cancer using image processing techniques. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0315-0318). IEEE.

[8] Arjmand, A., Angelis, C.T., Tzallas, A.T., Tsipouras, M.G., Glavas, E., Forlano, R., Manousou, P. and Giannakeas, N., 2019, July. Deep learning in liver biopsies using convolutional neural networks. In 2019 42nd International Conference on Telecommunications and Signal Processing (TSP) (pp. 496-499). IEEE.

[9] Huang, W.C., Chung, P.C., Tsai, H.W., Chow, N.H., Juang, Y.Z., Tsai, H.H., Lin, S.H. and Wang, C.H., 2019, March. Automatic HCC detection using convolutional network with multi-magnification input images. In 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS) (pp. 194-198). IEEE.

[10] Maurya, B., Hiranwal, S. and Kumar, M., 2020, December. A Review on Liver Cancer Detection Techniques. In 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-5). IEEE.

[11] Rajesh, S., Choudhury, N.A. and Moulik, S., 2020, December. Hepatocellular carcinoma (HCC) liver cancer prediction using machine learning algorithms. In 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-5). IEEE.

[12] Zien, A., Kraemer, N., Sonnenburg, S. and Raetsch, G., 2022. The Feature Importance Ranking Measure. [online] arXiv.org. Available at: <https://arxiv.org/abs/0906.4258> [Accessed 10 September 2022].

[13] Ibm.com. 2022. What is Logistic regression? | IBM. [online] Available at: <https://www.ibm.com/topics/logistic-regression> [Accessed 11 September 2022].

[14] Molnar, C., 2022. 5.2 Logistic Regression | Interpretable Machine Learning. [online] Christophm.github.io. Available at: <https://christophm.github.io/interpretable-ml-book/logistic.html> [Accessed 11 September 2022].

[15] Gandhi, R., 2022. Support Vector Machine — Introduction to Machine Learning Algorithms. [online] Medium. Available at: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47> [Accessed 11 September 2022].

[16] O'Shea, K. and Nash, R., 2022. An Introduction to Convolutional Neural Networks. [online] arXiv.org. Available at: <https://arxiv.org/abs/1511.08458v2> [Accessed 11 September 2022].

[17] Sharma, A., 2022. [online] DataCamp. Available at: <https://www.datacamp.com/tutorial/convolutional-neural-networks-python> [Accessed 11 September 2022].