

Diagnosis of Cervical Cancer using effective Hybrid Model in Bangladesh

by

Syed Saad-Ul-Kabir
17101363

Ashwaq Noor Neeha
19101679

Rifah Nanjiba Ibnat
17101422

Tanjila Haque Prottasha
17301193

Sadika Rahman Ananna
17101514

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2021

© 2021. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. We are the only authors of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other university or any Institution. This paper is fully our own work while completing degree at Brac University.
2. The thesis does not contain material previously published or written, except where this is appropriately cited through full and accurate referring.
3. The content of the thesis is the result of work which has been carried out since the date of approval of the research program.
4. All the ethics procedure and guidelines have been strictly followed while preparing thesis

Student's Full Name & Signature:

Syed Saad-Ul-Kabir
17101363

Rifah Nanjiba Ibnat
17101422

Ashwaq Noor Neeha
19101679

Tanjila Haque Prottasha
17301193

Sadika Rahman Ananna
17101514

Approval

The thesis/project titled “Diagnosis of Cervical Cancer using effective Hybrid Model in Bangladesh” submitted by

1. Syed Saad-Ul-Kabir (17101363)
2. Ashwaq Noor Neeha (19101679)
3. Rifah Nanjiba Ibnat (17101422)
4. Sadika Rahman Ananna (17101514)
5. Tanjila Haque Prottasha (17301193)

Of Spring, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on June 2, 2021

Examining Committee:

Supervisor:
(Member)



Md. Saiful Islam
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)



Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

Assigning an individual based on a heuristic provided by a Machine Learning model without any further analysis or evaluation might provoke ethical dilemmas. Hence, we assured the transparency of the evaluation process. In addition to that provide visual interpretation of the output provided by the Machine Learning Model.

Abstract

This paper discusses the various forms of restraints, regulations, struggles, and remedies for overcoming cervical cancer challenges. It is a life-threatening disease that affects women all over the world. The lack of adequate early treatment has been one of the leading causes of this. As a result, early diagnosis of cervical cancer is important. A substantial number of deaths in Bangladesh were caused by a shortage of specialists for cervical screenings. In addition, hospitals in many parts of Bangladesh do not have medical equipment for cervical screening, such as Pap smear, colonoscopy, and biopsy. An au-dominated system can therefore be extremely valuable for quick identification and treatment if the likelihood of cervical abnormalities is predictable. We employed hybrid machine education for this job. From Hybrid Machine Learning, we worked with K-Nearest Neighbor, Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, Stochastic Gradient Descent, Ada-boost, and Recurrent Neural Network to find the data classification. KNN uses architectures for easy and accurate detection including the classification of cervical cells. With the use of KNN, we use demographic data to determine whether or not a patient has an atypical cervix instead of picture data. The advantage of a decision tree is that all potential decisions will be analyzed and each path concluded. Moreover, we have used the Naive Bayes algorithm. It is a probabilistic machine that may be used for a variety of classification tasks. Random forest is a supervised learning algorithm designed to solve classification problems. It's just a set of decision trees whose outcomes are combined into a single final result. We have also used SVM in the prediction of cervical cancer. RNN is a type of artificial neural network in which nodes' connections form a directed graph that follows a temporal series. It can now show temporal hierarchical actions as a result of this. We have applied for RNN and got a good result. In each iteration, Ada-boosts are used to set the classifier weights and to train the data sample so that those odd observations are predicted accurately. Thanks to a unique training example processed by the network, SGD is easy to match to memory. It is quick as only one sample is analyzed simultaneously. These models and methods have achieved higher accuracy of the decisions that they recommend including deep understanding to make decision-makers easier to implement. We got better accuracy by learning implicit, non-implicit, and non-symbolic knowledge. Furthermore, in the future, we will use more algorithms to improve accuracy.

Keywords: Cervical Cancer, K-Nearest Neighbor, Hybrid Machine Learning, Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, Recurrent Neural Network, AdaBoost, Stochastic Gradient Descent, Demography.

Dedication

We would like to dedicate this thesis to our loving parents. As well as, all the amazing faculties we encountered and learnt from in the course of pursuing our Bachelors degree. It has been a journey worthwhile. . .

Acknowledgement

Firstly, we would like to acknowledge sincere gratitude to our supervisor Md. Saiful Islam for the continuous support of our thesis paper and for his patience, motivation, enthusiasm, and immense knowledge. His valuable suggestions, ideas, support and effort have significantly helped us in our research work. Beside our supervisor, we would like to thank our co-supervisor Wasif Shafaet Chowdhury, who encouraged and assisted us with our work in researching. His guidance helped us all the time. We are thankful to Almighty Allah for providing us the opportunity, guidance and confidence to work in our research. Last but not the least, we would like to express our genuine appreciation to our parents who made sure we read in one of the best institutions and without their constant source of inspiration, support, faith and prayers we could never accomplish success.

Table of Contents

Declaration	i
Approval	ii
Ethics Statement	iii
Abstract	iv
Dedication	v
Acknowledgment	vi
Table of Contents	vii
List of Figures	ix
List of Tables	x
Nomenclature	xi
1 Introduction	1
1.1 Overview	1
1.2 Importance/Usefulness :	2
1.3 Current scenario and Motivation :	2
1.4 Research Objectives :	2
1.5 Challenges Faced:	3
1.6 Thesis Outline:	3
2 Literature Review	5
2.1 Literature Review	5
3 Background Analysis	7
3.1 Decision Tree	7
3.2 Random Forest	9
3.3 Naive Byes:	10
3.4 RNN	11
3.5 AdaBoost	13
3.6 SVM:	14
3.7 KNN	15
3.8 SGD	15

4	Research Methodology	17
4.1	Dataset:	17
4.2	Project Overflow:	17
4.3	Feature Scaling:	18
4.4	Feature Selecting & Engineering:	18
4.5	Train-Test split	20
5	Model Implementation and Optimization	21
5.1	Work Flow Overview:	21
5.2	Evaluating and comparing Machine learning Models:	21
5.3	Model Optimization:	23
5.4	Hyperparameter Tuning :	23
6	Experimental Results and Analysis	25
6.1	Comparative Analysis of Supervised Models	25
6.2	Comparison on Training Data	25
6.3	Graphs	26
7	Conclusion and Future work	28
7.1	Conclusion	28
7.2	Future Work	28
	Bibliography	31

List of Figures

3.1	Decision Tree	8
3.2	Random Forest	10
3.3	Naive Bayes	11
3.4	RNN	11
3.5	Forward Pass	12
3.6	AdaBoost	14
3.7	SVM	14
3.8	KNN	15
3.9	SGD	16
4.1	Some columns of our dataset	18
4.2	Density Plots of Categorical features against Age	19
4.3	Heat-map of Numeric Features	20
5.1	Work-Flow Diagram	22
6.1	Model Loss	26
6.2	Model Accuracy	26
6.3	Graph Chart Mae	27
6.4	Graph Chart Accuracy	27

List of Tables

6.1	Comparison of models	25
-----	--------------------------------	----

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

CV Coefficient of Variation

DT Decision Tree

E Entropy

KNN K-Nearest Neighbours

P Probability

RF Random Forest

RNN Recurrent Neural Network

SDR Standard Deviation Reduction

SGD Stochastic Gradient Descent

SVM Support Vector Machine

Chapter 1

Introduction

1.1 Overview

Cervical cancer is an ovarian cancer. It is one of the most common cancer among women and each year a great number of women lose their life due to this cancer [25]. Cervical cancer is a sort of malignant growth. Exactly when introduced to HPV, the body's safe system generally holds the disease back from doing hurt. In somewhat level of people, regardless, the contamination gets by for a seriously long time, adding to the cycle that makes some cervical cells become illness cells [41].

Generally, cervical cancer affects aged women more than young women. Women who are aged over 30 are at greater risk. Enduring disease with specific sorts of human papillomavirus (HPV) is the primary driver of cervical malignant growth. HPV can be passed during sex [41]. This malignancy can influence the more profound tissues of their cervix and may spread to different pieces of their body (metastasis), regularly the bladder, rectum, vagina, lungs, and liver. At present women in our country are at high risk of getting this cancer due to their lack of knowledge about their body and due to their negligence towards their health. Even if some abnormality starts to show within their body, they ignore it and do not take proper medication. Thus they easily get this cancer and most of the time they do not even realize it until it's too late to recover. The chance of getting recovery depends on the stage of this cancer. The odds of living for in any event five years in the wake of being determined to have cervical malignant growth are, if it's in stage 1 then the chances are 80-99%. If it is in stage 2, then the chances are 60-90% and if it is in stage 3, the chances are 30-50%. At present, it is the second most common cancer In Bangladesh. The main purpose behind our research is to make cervical cancer familiar among general people and make awareness in individuals about it. Our research will help people to understand it. We are utilizing techniques that can anticipate the danger factors of cervical irregularity and thus helps to diagnosis the disease. Our paper will discuss the application of different supervised algorithms to predict the risk of getting diagnosed with cervical cancer. Our data set includes cervical cancer patient's history and information. We have used Decision tree, KNN, SVM, Naïve Bayes, RNN, Random Forest, AdaBoost and SGD to help us calculate the final result. A comparative study of each model will be made to choose the most ideal model. The rest of the paper incorporates a brief conversation about some other important work that has been done regarding cervical cancer. There will be an itemized portrayal of our proposed model and our informational collision. There

will also be a discussion about data analysis. We have mentioned the outcomes and test investigation in detail as well. We have also discussed our future work and thus the paper will conclude with closing remarks.

1.2 Importance/Usefulness :

Cervical cancer is a very common disease in Bangladesh. Many women in our country are suffering from this cancer but due to lack of knowledge and ignorance, they do not even realize it until it's too late to recover. According to the international research agency for cancer, more than 50 million women are at high risk of developing cervical cancer. Unfortunately, their health issue lacks priority or get ignored most of the time. As a result, most of the patients die or lose the ability to conceive a child. This disease can be cured if it can be diagnosed earlier and [19] by keeping these in mind we are analyzing the cervical patient's data and using hybrid models so that we can diagnose the disease or the chance of getting the disease. By using our method women can predict whether they are at risk or not and took the necessary precautions. Our research will help them know about the disease and how important it is to pay attention to their health. Women will be able to understand the abnormality in their bodies and thus take the necessary treatment. Our research will help us observe the abnormality in women's cervix and how cervical cancer affects the body's other functions. Since our country is a conservative one so till now most people are not comfortable discussing these issues with doctors, especially male doctors. Thus our conservative society prevents women from getting the proper treatment. Our research will help our conservative society to know more about the disease and its necessity to be cured. It will help us to lower the rate of cervical cancer patients in our country by rising peoples' awareness.

1.3 Current scenario and Motivation :

Right now cervical malignancy is the fourth most regular disease for ladies all through the world. In Bangladesh, the scenario is much more frustrating. Each year more than 17686 new cases of cervical cancer are found and it causes almost 10362 death per year [41]. Therefore on average each day almost 49 new cases are being registered and 28 death occur due to this. But the good news is, whenever analyzed at the forerunner stage, notwithstanding, cervical malignant growth is a condition that can be treated effectively [41]. Bangladesh is a poor country and it is almost impossible for poor people to get treatment for this disease but if the disease can be predicted early, then they might get a chance to get proper treatment and cure themselves. This is the main motivation behind our research and we want to ensure a better health condition for women. We want to make sure that women are conscious of their physical problems and taking proper care of their bodies [39].

1.4 Research Objectives :

Cervical cancer causes a psychosocial impact on both women and their partners. Women in our country are still not aware of this cancer. As a result, high death

rates had been seen in Bangladesh. This has an effect on people and the country's economy. Objectives for this paper are as follows:

- The purpose of this research is to create awareness in people about cervical cancer. It will help to work with the patient to know the forecast of the danger level of cervical malignant growth so that whenever related to high infection hazard level, patients can take necessary steps.
- This paper will help people to know about cervical cancer and help them understand the importance of it. Moreover, people will not hesitate to talk about this cancer.
- we aim to improve the physical and mental condition of cervical cancer patients and to provide the highest accuracy through this research.
- This study will enhance more legitimate and skillful broadcasting across the target women.
- Furthermore, methods that we are using, can detect and foresee the risk factors of cervical malformation, and it will furnish a better opportunity to the relevant contributor so that it tends to be extremely valuable for fast discovery and treatment.

1.5 Challenges Faced:

Initially, after a lot of brainstorming, we decided a topic on which we wanted to do our thesis. Since our topic was based on cancer patients, we decided to collect patient's data from the hospital.

Firstly, we had to find out all the cervical cancer hospitals. Then, we made a questionnaire. We went to several hospitals. We had to explain our motive and then gave the questionnaire to some of the hospitals. It was very time-consuming. We had to wait hours in the hospitals to get permission for taking patients' data. We had to explain the whole thing to them. Then they used to send us from one person to another. Finally, we used to get permission to collect data. Unfortunately, even after waiting for hours, the person we had to see for collecting data was not there so we had to come back. Then comes the Covid situation. After the lockdown started in March, we had to stop going to the hospitals. As the hospitals were the most threatening places to go. As a result, we had to drop the idea of collecting data from hospitals. After that, we were very confused. However, we decided to take an approach using the collected data set from Kaggle. Later, if the corona situation comes to normal and we get hospital data sets, we can just fit it into our developed models.

1.6 Thesis Outline:

The main goal of this thesis was to construct a model capable of predicting cervical cancer in the early stage based on patient's different symptoms. As well as to detect and predict the risk of being diagnosed. The aim of the authors was to construct the best model for achieving the task and enhance it for accuracy. In addition to it, provide visual reasoning behind the output.

To start with, in the first chapter (Chapter 1), an overview and the benefits of the model are discussed. The problem statement was to make awareness of cervical

cancer and introduce early detection along with prevention in our country, thus help avoid the loss of lives.

Secondly, in (Chapter 2) related work surrounding hybrid algorithm classification and its predictions were discussed. Indicating significant results achieved by researchers. Also, the lacking of existing methods were overviewed.

Thirdly, in (Chapter 3) background analyses of various supervised algorithms were discussed along with their implementation details. These algorithms were later used in the research pipeline.

In (Chapter 4) the Research Methodology and workflow were proposed. Details of data collection, Feature Scaling, Feature selection & Engineering conducted in the research were elaborated.

Furthermore, in (Chapter 5) selected models were optimized by hyperparameter tuning and cross-validation to improve the accuracy metric. Comparative analysis was executed between various algorithms. model optimization details were also given.

In (Chapter 6) Experimental results and analysis were conducted. Visual comparisons of the models were provided. Accuracy metrics were tabulated.

Finally, in (Chapter 7) Conclusions and further work were drawn.

Chapter 2

Literature Review

2.1 Literature Review

Cervical cancer is a prominent matter in the fields of medical science. Many researchers have studied and researched this cancer. For many years researchers have been used hybrid algorithms for studying cervical cancer. These Algorithms and Neural Network based designs make us ready of getting results quicker and with more noteworthy accuracy that we need in our advanced days. Dipti N. Punjani, Dr. Kishor H. Atkotiya elaborated on the concept of cervical cancer in their research. Their research paper “Cervical Cancer Prediction Using Data Mining” shows us the related determinants and also gives us a comprehensive simplification of data mining techniques, their categories, their uses, and their application [32] . In 2014, P.Ramachandran, N.Girija, and T.Bhuvanewari did research on cervical cancer where they used data mining techniques to detect cervical cancer. In their research paper ”Early Detection And Prevention Of Cancer Utilizing Data Mining Techniques” devise a strategy for ascertaining cancer risk factor where they used factors like family history and schooling among others and they utilized k methods and clustering as well [16]. Akshitha Shetty , Vrushika Shah studied cervical cancer where they used a machine learning approach for predicting the disease. In their research paper “Survey of cervical cancer Prediction using Machine Learning: A comparative approach” , they stated the strategy of comparing some extensively used machine learning algorithms including support vector machine, k-nearest neighbors, Artificial Neural Network, Classification and regression tree (CART), Random forest tree (RFT) and so on [37]. M.K. Soumya, K.Sneha, and C.Arunvinodh of the Department of Computer Science Engineering, Royal College of Engineering Technology, India used a different approach to classify and detecting cervical cancer. They utilized image preparation and texture analysis to foresee cervical cancer growth in the patient’s body [26]. Their research paper “Cervical Cancer Detection and Classification using Texture Analysis” proposes a characterization procedure utilizing Magnetic Resonance Images(MRI) to acquire the arranging of cervical disease patients and they also showed us the utilization of nonlinear SVM grouping models [26]. Their study inferred that texture highlights beats transformed highlights [26]. They executed nonlinear SVM grouping models and these models were built dependent on changing features of tumors for malignant growth and second order texture highlights as well [26]. Bargana Benazir and A.Nagarajan introduced an automated strategy in their paper “An Expert System for predicting the Cervical Cancer using

Data Mining Techniques” where they incorporate pre-handling methods [34]. In their research, they utilized a middle channel to eliminate the commotions in the cervical tissue cell gathered and at that point, its highlights are extricated utilizing Gray level co-occurrence matrix (GLCM). Hereditary calculation and Neural Organization Classifier are utilized to first, select an element subset and afterward arrange the cervical cells as typical or unusual [34]. Miao Zheilkui, Cao Yuzhen, Yin Wei, and Liu Rui published a research paper “An early stage Cervix Cancer Diagnosis System” where they depict the rule and plans of the system drive capacities, like auto-focusing, auto-exploring, and state analysis of the magnifying instrument [3]. “Prediction of Cervical Cancer using Hybrid Induction Technique: A Solution for Human Hereditary Disease Patterns” by R. Vidya and G. M. Nasira utilizes ideas of Information Mining. They constructed a decision tree using classification and Relapse Tree (CART) and Random Forest Tree (RFT) with mathematical datasets [27]. They used 500 records with 61 factors were utilized and tried with Regression tree calculation first with a precision of 83.87% and afterward Random Forest tree with 93.54% expectation exactness and lastly, K-Mean Method with the most noteworthy exactness of 96.77% was accomplished [27]. In 2014, “Application of machine learning to predict the recurrence-proneness for cervical cancer” by Chih-Jen Tseng, Chi-Jie Lu, Chi-Chang Chang, and Gin-Den Chen considered three machine learning approaches which are support vector machine, C5.0, and extreme learning machine [17]. They utilized this approach to find the important risk factors to predict the recurrence proneness for cervical cancer [17]. “Prediction of Recurrence in Patients with Cervical Cancer Using MARS and Classification” by Chi-Chang Chang, Sun-Long Cheng, Chi-Jie Lu, and Kuo-Hsiung Liao applied advanced data mining technique for recurrent cervical cancer in survival analysis [13]. The 367890-y gathered patient information with 12 indicator factors and applied two strategies on them which are C5.0 classifier and Multivariate Adaptive Regression Splines (MARS) where MARS showed a grouping pace of 86% though C5.0 showed an order uncommon of 96% which presumes that the C5.0 model has the best characterization capacity regarding normal right arrangement rate and it is better than SVM Model as well [13].

Chapter 3

Background Analysis

3.1 Decision Tree

Decision Tree being a swift learning algorithm, allows us to categorize and estimate a target variable by forming a tree structure. Decision Tree creates a training model based on decision rules. Being a tree-like structure, its training data set elements are managed to built the edges to proceed from one node to another till it gets to the leaf node which correlates to one of the output classes. The benefits of this algorithm are that it can handle a vast range of input data types and operate missing values. Besides, it can also attain a better initial accuracy with simple implementation. If there are N attributes in our data set, we have to use some criteria like Standard deviation, Entropy, Information Gain, Gain Ration, etc. to solve the attribute selection issue.

The key algorithm for forming a decision tree is known as ID3 by J. R. Quinlan [1]. It involves a top-down, greedy search through the volume of probable branches with no backtracking and splitting the data into subsets that hold samples with alike values(homogeneous). Standard deviation is used to compute the homogeneity of a sample instance. A sample is entirely homogeneous when the standard deviation is null.

$$StandardDeviation, \alpha = \sqrt{\frac{1}{N} \sum_{i=1}^N (xi - \mu)^2} \text{ where } (Mean), \mu = \frac{1}{N} \sum_{i=1}^N xi \quad (3.1)$$

$$CoefficientOfVariation(CV) = \frac{\partial}{\mu} \times 100 \quad (3.2)$$

The coefficient of variation (CV) is chosen as the criterion for stopping the recursion, i.e if the CV falls under a certain threshold, we stop the splitting process and allocate the average value at the leaf node for the subset.

The Standard deviation for two variables(Target, Feature):

$$S(T, X) = \sum_{c \in x} P(c)s(c) \quad (3.3)$$

$$E(T, x) = \sum_{c \in x} E(c)P(c) \quad (3.6)$$

Where T is the current state and X is the selected attribute
The Information Gain(IG) function would be

$$IG = Entropy(Before) - \sum_{j=1}^k Entropy(j, after) \quad (3.7)$$

Then we have to calculate a cost function named Gini Index to analyze the dataset splits.

$$Gini = 1 - \sum_{i=1}^c (p_i)^2 \quad (3.8)$$

Information Gain inclines to determine attributes with a large number of values. To overcome this biasness, a Gain Ratio is needed. Following Gini Index, we have to calculate the Gain Ratio.

$$GainRatio = \frac{InformationGain}{SplitInfo} \quad (3.9)$$

3.2 Random Forest

Random Forest has the ability to perform both classification and regression at the same time. By working with the principle, it utilizes a category of poor learners to build a strong learner. RF simply creates multiple decision trees and merges them to find the overall finest results [10]. In the same way, our possibility of creating the right forecasts enhances the quantity of uncorrelated trees in our model with the Random Forest algorithm. Decision trees are usually sensitive to the trained data. We can get remarkable unrelated tree structures with a slight modification of the training set. Random Forest gains an advantage from this and lets each tree casually sample from the dataset which results in different trees. This procedure is called bagging [38]. We train individual trees on a separate junk with bagging instead of sub-setting the training data into mini portions. We have a N size sample and yet we feed individual trees N size training data. We use an arbitrary sample with replacement instead of the original training data and that random sample is taken from N. Generally, at the time of splitting a node in a normal decision tree we consider all probable features and select the one that generates the greatest separation among the observations in the left node and the observations in the right node. On the other hand, a tree in a random forest can select one exclusively from a random subset of features. This compels to create more variety across the trees which causes lesser correlation and more diversification among the trees. Therefore, Random Forest uses bagging and feature randomness to predict more accuracy by the committee as forming each tree creates an uncorrelated forest. Moreover, in this algorithm, we get trees that use unique features to make decisions and due to bagging, those trees are trained on separate sets of data.

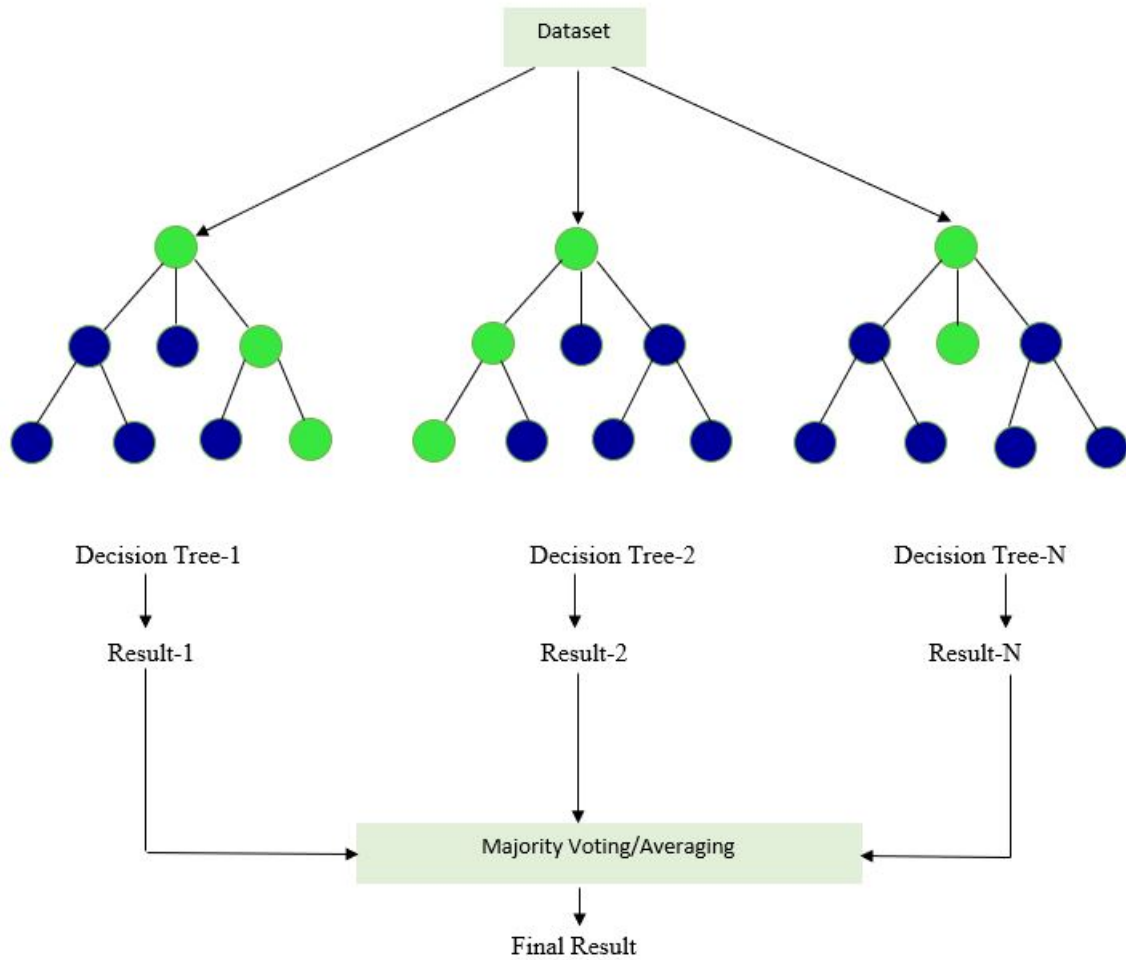


Figure 3.2: Random Forest

3.3 Naive Bayes:

One of the simplest, fastest, and eager learning classifiers that we can apply to our data is Naive Bayes. It is a powerful algorithm and its classification techniques are based on base theorem with an assumption of independence among predictors. Naïve Bayes is a very much useful and so much reliable algorithm. It has two parts. Naïve and Bayes. Naïve based model is pretty much simple and easy to build, especially for a large number of datasets. On the other hand, Bayes theorem tells us the probability of a situation, based on the previous fact of the condition that might be related to that situation. It is generally a way so that we could easily figure out the conditional probability. Naïve Bayes is the most straightforward and fast classification algorithm and It is very efficient in the case of large data [31] , [40]. Hypothesis H and the Evidence E , the relationship between them:

$$P(H|E) = \frac{(P(E|H).P(H))}{P(E)} \quad (3.10)$$

The Naïve Bayes algorithm is very popular in case of medical issues. And it gives a very good and effective output. In our task the algorithm used as follow:

1) First, we have loaded our data in CSV format.

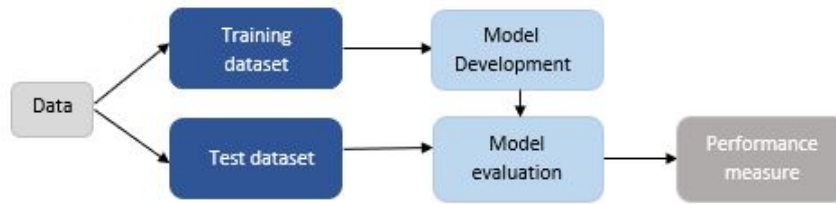


Figure 3.3: Naive Bayes

- 2) Then we have explored the dataset and split the dataset into training and testing datasets. And our ratio was 80:20
- 3) Then we have generated a model on the training dataset and making all the predictions by given the test data.
- 4) At the end, we have summarized the dataset.

3.4 RNN

Recurrent Neural Network processes a sequence of data with the time step-index. RNN is called recurrent as it executes the same task for each component of the series and the output of each element depends on the former computation.i.e. For every new step, the previous output is used as an input while their states are kept hidden. This neural network has the capacity to process input of any length but the input size does not affect the size of the model. When computing, historical informations are taken into account and weights are shared over time [12].

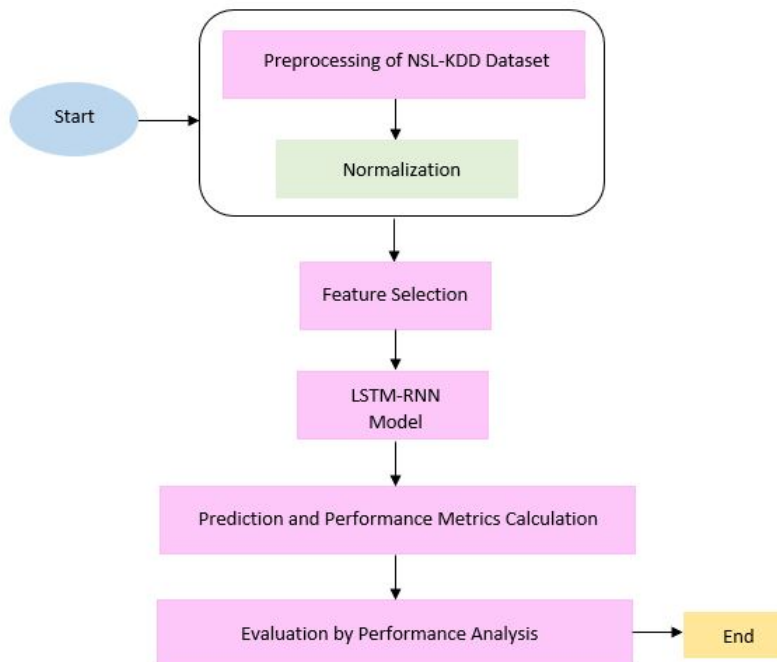


Figure 3.4: RNN

Loss function and Backpropagation:

The loss function in a recurrent neural network of all-time steps is defined by the loss at each step.

$$L(\hat{y}, y) = \sum_{t=1}^{T_y} L(\hat{y}^{<t>}, y^{<t>}) \quad (3.11)$$

While backpropagation is performed at each and every point within time. When the timestep is T, loss L deviation is as follows:

$$\frac{\partial L^{(T)}}{\partial W} = \sum_{t=1}^T \frac{\partial L^{(T)}}{\partial W} |_{(t)} \quad (3.12)$$

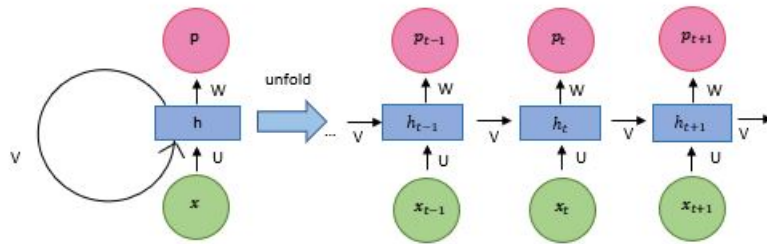


Figure 3.5: Forward Pass

Forward pass:

The below equations can be represented as the RNN forward pass:

$$\begin{aligned} f^{(t)} &= r + Wh^{(t-1)} + Ux^{(t)} \\ \text{hiddenstate}^{(t)} &= \cosh(f^{(t)}) \\ p^{(t)} &= s + Vh^{(t)} \\ \hat{y}^{(t)} &= \text{adam}(p^{(t)}) \end{aligned}$$

This example of RNN depicts an input series to an identical length of output series [28].

- The total losses over all the time steps is the result of the overall loss for a stated series of x values paired with a series of y values.
- As the adam function is needed to get vector \hat{y} of possibilities on the output, we suppose outputs $p^{(t)}$ are used.
- We assume loss L is the negative logarithm of target $y^{(t)}$ considering the input to date.

Backward pass:

Following the backward propagation pass that moves right to left, gradient computation requires carrying out a forward propagation pass that moves from left to right via graph [30].

- As the forward propagation graph is sequential, inherently; the individual time step is calculated only after the earlier one and as a result, the runtime $p^{(T)}$ cannot be decreased by parallelization.

- Until reused during backward pass, conditions calculated in the forward pass needs to stow if we want the memory cost to be $p^{(T)}$ as well.
- Backpropagation through time is the algorithm executed to the unrolled graph with the $p^{(T)}$ cost [14].
- The gradient at individual output relies on the computations of the current time step as well as on the previous time steps when the all-time steps within the network share the parameters.

Computing Gradients:

We compute the gradients from the given loss function L , for the bias r , s and three weight matrices U , V , W . Also, we need to upgrade it with a learning rate termed α . The gradient simply provides us the way of the loss changing concerning each and every weight parameter just like backpropagation. By upgrading the weights we can reduce loss as follows:

$$W \leftarrow W - \alpha \frac{\partial L}{\partial W} \quad (3.13)$$

3.5 AdaBoost

AdaBoost is basically used to boost the performance of a decision tree. AdaBoost is very popular in Data Science or Machine Learning field. By combining multiple classifiers that have low accuracy, it increases the accuracy higher. AdaBoost is an iterative method [8]. The concept of AdaBoost is very simple. Here we need to set weights and train the dataset so that it can fulfill the target result. The weights are included in the training dataset and each time while training a classifier, AdaBoost increases the weights on the inappropriate models and target to make up a bigger piece of the following classifiers preparing set with all that misclassified models so that the following classifier prepared will perform better on them. As it is an iterative process, so in every process it tries to give an outstanding fit to minimize the training error. [8], [15]

AdaBoost first selects a training subset and first, it chooses it randomly. Then iteratively it starts to train the AdaBoost Machine Learning model and try to predict the accurate prediction. Wrong classifiers get higher weight as at the end with the all the low accuracy it will create high accuracy model. It will continue until data doesn't fit without error.

The workflow of AdaBoost is given below:

- 1) We have first imported the libraries that are required.
- 2) Then we have loaded the dataset from the CSV file and split it into Train Test sets with the ratio of 80% and 20%.
- 3) Then we have created the AdaBoost model using Scikit -Learn and as default, it uses Decision Tree Classifier.
- 4) Now at the end we evaluated the model and predict until the whole dataset fits with minimizing error.

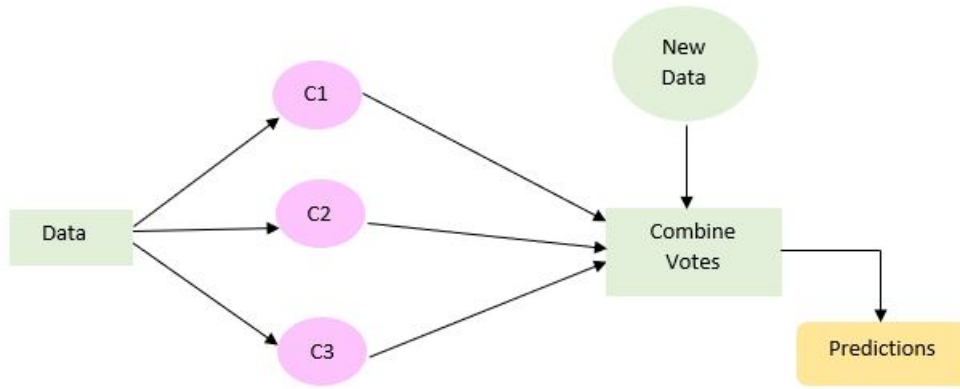


Figure 3.6: AdaBoost

3.6 SVM:

Support Vector Machine can be used as classification and regression as support vector regression known as SVR and support vector classification known as SVC [2]. SVM happens to be very effective in the higher dimension. It depends on detecting hyperplane which separates features into different separate domains. As the hyperplane happens to be affected by only support vectors thus outliers have less impact. Support vector points are the points nearest the hyperplane and margins are the space between the vectors and the hyperplane. The basic instinct to built here is that the probability of precisely classifying the points in their respective regions increases as the SV points move away from the hyperplane. These SV points seems to be very crucial in finding out the hyperplane because whenever the vector position changes, the hyperplane's position changes as well. In 2-D, hyperplane are the function classifying points within the higher dimension and a line is used when the function classifies between features [6].

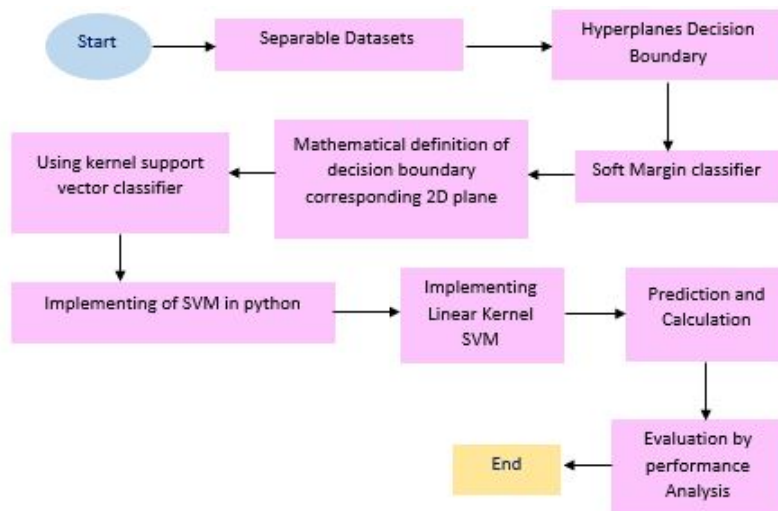


Figure 3.7: SVM

The action of computing two dot products of x and y vectors in feature spaces in very high dimensional is the Kernel. In consequences at times functions of Kernel is known as generalized dot product. In our model, we have used a linear Kernel

and it's calculation is given below:

$$F(x) = \text{sum}(a_i * (x, x_i)) + B(0) \quad (3.14)$$

3.7 KNN

KNN is an algorithm that stores every available case and makes a group of the new data or cases depend on the similarity measure. It is a very simple algorithm. KNN basically apply in such field where we are searching similar items. It suggests that if we are near to the neighbor then we are one of them. [5]

Here, k denotes the quantity of the closest neighbors that use as a voting class of the new data or the testing data, and k can be any of the numbers. If k=5 then the labels of that classes are checked and the most acceptance label is set as testing data. K calculates the least distance to measure the nearest neighbor. [5] , [33]

In our tasks the algorithm used as follow:

1. Open the dataset from CSV format and split it into test and train datasets.

Our ratio was 80:20

2. By using the Euclidean distance formula, we have computed the distance from the training example to the labeled example. Then calculate the distance between two data instances.

$$EuclideanDistance = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.15)$$

3. Locate K's most similar data instances. And generate a response from a set of data instances.

4. Lastly, we have summarized the accuracy of prediction.

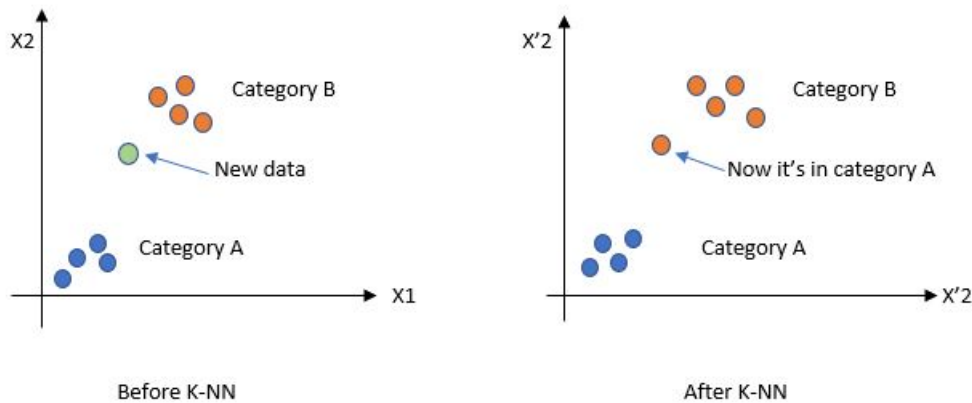


Figure 3.8: KNN

3.8 SGD

Gradient Descent is mainly used for solving the problems of optimization and to improve the learning of the model. And Stochastic Gradient Descent Algorithm

is one of them. The SGD is a bit modification to the Gradient Descent which calculates the gradient and changes or updates the weight matrix over the batches of the training data. It helps us to make faster changes by taking larger steps on the gradient that eventually leads to further speed up the convergence. The learning rate and df/dx are the functions of minimizing our loss.

$$f(m, b) = \frac{1}{N} \sum_{i=1}^n (y_i - (m x_i + b))^2 \quad (3.16)$$

First, we need to import the necessary libraries for our SGD like sklearn, pandas, NumPy, etc. After loading the dataset, we will split the data set with a train test split. For standardizing we use Standard Scalar. For avoiding being overfitted, we only fit the training data. For that our model won't be seeing this data before hand. We used SGD Regressor for viewing errors and SGD classifiers for accuracy.

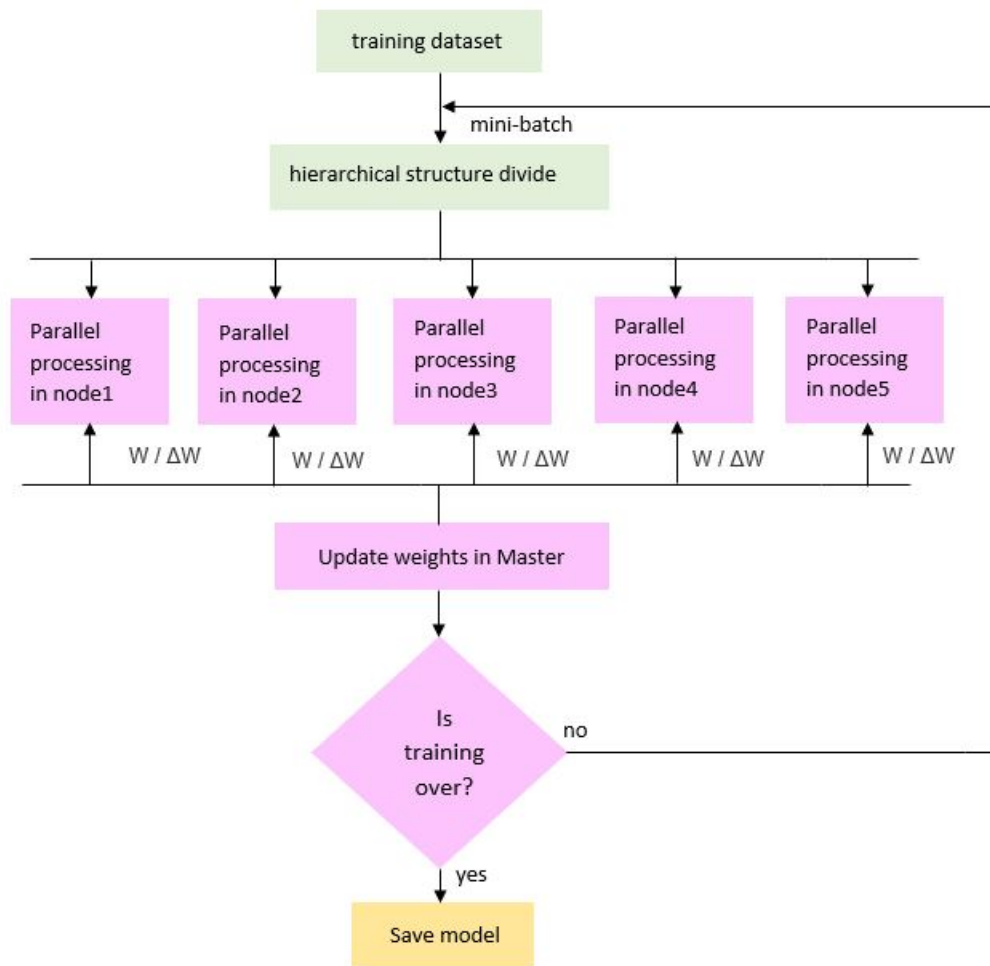


Figure 3.9: SGD

Chapter 4

Research Methodology

4.1 Dataset:

Even though the data was supposed to be retrieved from Cancer hospitals, due to unfortunate circumstances, the actual data has not yet been received. Therefore, to conduct our data analysis we opted for enough reliable dataset with 589 entries extracted from Kaggle [<https://www.kaggle.com/loveall/cervical-cancer-risk-classification/code>].

4.2 Project Overflow:

The goal of this workflow is to use Hybrid Algorithm in Machine Learning or the diagnosis of Cervical cancer data to show high prediction accuracy. That is why we need to train and test the dataset accurately with a suitable model. We have used 6 models to see which gives us the best accuracy and can predict the best result. For the training of dataset, we have input variable (x) and output variable (y) wherein our case (y) is DX: Cancer (we have changed (y) in some algorithm to get the higher accuracy) and we have used the algorithms from the input to the target to learn the mapping function. And this is a supervised regression machine learning problem.

* Supervised problem: We need to train the data where both the features and the targets are given. [2]

* Regression problem: It is a process of investigating the relationship between an independent and dependent value. The target is an incessant variable Machine Learning workflow [7] , [4].

Here is the process of implementations that we have applied for the models. Though some implementations may vary according to the specific models however in the cases of Machine Learning projects the format is almost constant. The general structure is given below:

- Data preprocessing
- Exploratory data analysis
- Train and test split
- Feature scaling
- Feature selection and engineering

- Training and prediction
- Complete hyperparameter tuning on the best model
- Evaluating the algorithm using metrics
- Evaluate which model in the training dataset

With the above discussion, it can be seen that it is not a linear workflow. This process is iterative as for evaluating something we sometimes need to visit one thing more than once, also it does not always maintain the same protocol.

	Age	Number_of_sexual_partners	First sexual intercourse	Num of pregnancies	Smokes	Smokes (years)	Smokes (packs/year)	Cc
0	18	4	15	1	0	0.0	0.0	
1	15	1	14	1	0	0.0	0.0	
2	52	5	16	4	1	37.0	37.0	
3	46	3	21	4	0	0.0	0.0	
4	42	3	23	2	0	0.0	0.0	

Figure 4.1: Some columns of our dataset

4.3 Feature Scaling:

Feature scaling is very important to evaluate all the features equally. In our pre-processing steps, we have done the feature scaling to normalize the range of the independent variable. Some Machine Learning Algorithms do not work properly without normalization. Moreover, it is also faster to converges the gradient descent algorithm [4] . For example, while doing Random Forest we have done Feature scaling using Scikit-Learn’s StandardScaler class.

Min-Max formula is used for scaling. By selecting the range, we have sued this formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4.1)$$

4.4 Feature Selecting & Engineering:

•Feature Selection: Basically, feature selection is used when there are a large number of input variables and need to reduce the irrelevant variables and only use those that are supposed to be the best for a model to predict the target variable [20]. The main technique is to use the target variables. The process of selecting the most useful data is not always the same. It depends on many factors. But the features that have intended to show the highest compatibility with the target that is selected and other features that are less relevant are removed that help the model to understand the new data.

•Feature Engineering: The process of where the raw data is transformed into features that understand the problems of the predictive models better and improve the model accuracy on unseen data [29] . There is much use of feature engineering techniques to improve the function quality of a model. Handling Outliers is used, though

some models are sensitive to outliers. But usually, Logarithm and square root are used which is one of the most common mathematical transformations in the case of feature engineering [29], [11]. In the case of numerical features of the dataset, some models require scaling 4.3.

Here, Age Density plots were plotted to show the relationship of categorical values for each categorical feature. The plotting shows how the Density changes for different categorical feature values on the age. It is known that for the large amount of data we need to redundant those features which values make less effect on the model 4.4. So, from the plots, we looked at those features which values did not make any change in the case of the distribution of the age. In this way, in the final model, we could give them less priority. The plotting of categorical features against age is given below:

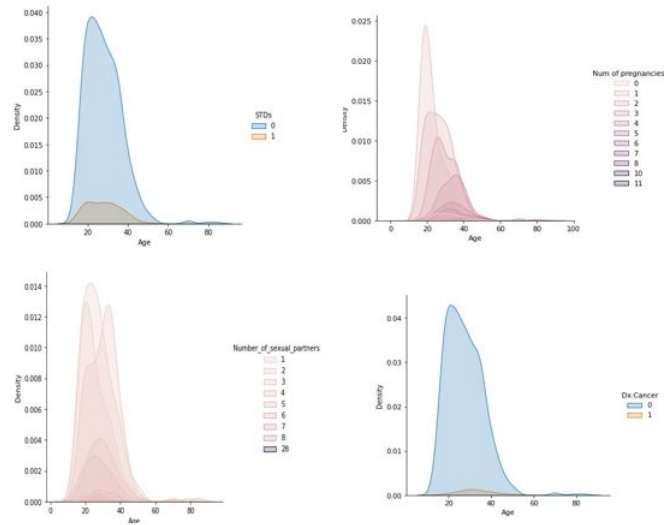


Figure 4.2: Density Plots of Categorical features against Age

From the above plotting, it is clear that the “Number of Pregnancy” did not make any changes as its values are almost close and don’t affect the final model. As a result, in our final model, we could give it less priority. But the values of other features affected the distribution on Age.

Feature engineering and selection is a continuous process that is why it needs several repetitions to get the right result. Different types of models require different types of feature engineering techniques. As an example, in the case of Random Forest, we can see the different accuracy before using feature scaling and after using. Using irrelevant variables and features or selecting the feature that is not best can give the model a poor performance. [23]

By applying the log and square roots technique, we need to avoid the features that are multi-linear. Because those features will not put any effect on improving the accuracy. So, there is a threshold that is maintain to understand how much co-linearity is acceptable. In different cases different threshold is acceptable. To find the multi-linearity, the correlations between the feature are generated by Heat-map illustrating 4.3

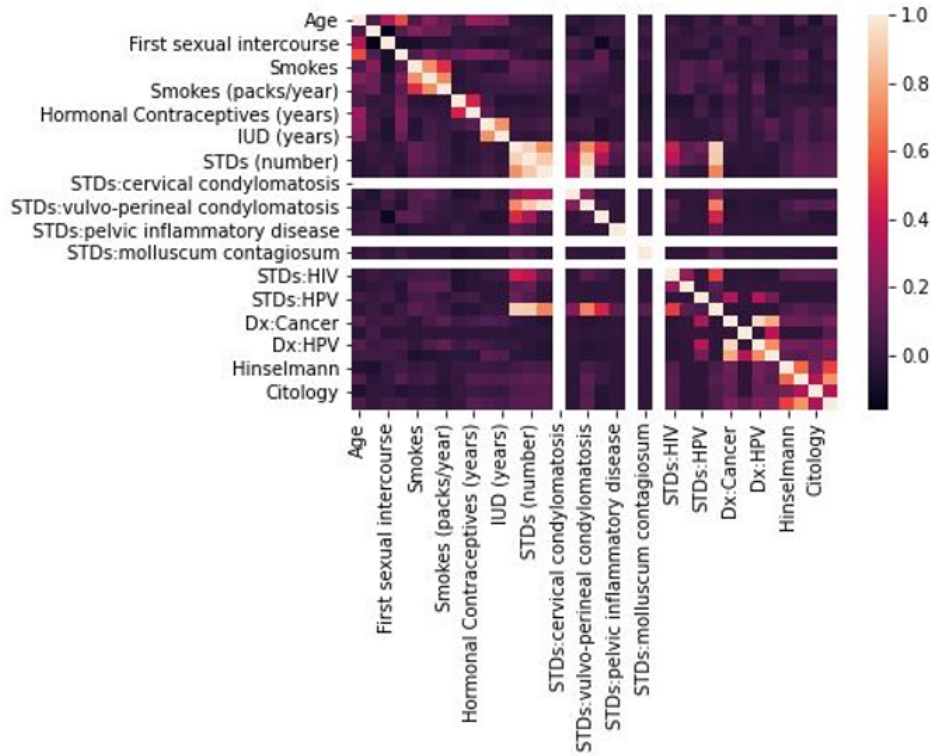


Figure 4.3: Heat-map of Numeric Features

4.5 Train-Test split

Splitting the dataset into training and testing is a very important thing of a dataset for the evaluation of a model and also helps to better understand the characteristics of models. Train dataset uses to make the machine learning model fit. On the other hand, the test dataset uses to assess that fi machine learning model. By using the library of Scikit-learn, we have split the dataset into Train Test. Normally though the standard ratio is 30% test set and 70% train set but in our algorithms, we have used 80% and 20% ratio for better performance. So, we have maintained the standard and select the features (X) and Target (Y = "Cancer"), then divided the dataset into Train Test split. Here the training dataset will prepare an unknown model or the trained model and an accuracy evaluation would be done by test dataset.

With the help of a training dataset, a prediction would make: From the prediction the Mean Absolute Error = 0.543.

Chapter 5

Model Implementation and Optimization

5.1 Work Flow Overview:

It is important to apply the accurate model space, given the training set in order to have the finest possible model. Thus, our initial goals are to aim towards minimizing the true error on the test set and to avoid over-fitting the data on the training set. Following is an overview of our Work-Flow 5.1. The details are as follows:

- Raw Dataset is cleaned and processed. This includes assigning missing values, identifying outliers, and feature scaling.
- Feature Selection and Engineering are carried out. Selection of important features using correlation and feature importance as reference. Performing recursive feature importance elimination. Feature Engineering is conducted by adding square roots and logs of numeric columns and removing multi-linearity in the features.
- Train Test split carried out in the ratio 80:20, where the test set is held out for evaluation in the final model.
- Models are optimized using Hyperparameter tuning and Cross-validation
- Final model estimated on hold out test set. Results obtained and analyzed. A comparative analysis was conducted between the models and the final model.
- * Randomized Search over hyperparameters of selected models is conducted to further filter models and parameters.
- * Models are optimized using Cross-Validation along with Randomized Search over the hyper-parameter space

5.2 Evaluating and comparing Machine learning Models:

Various machine learning methods are built, learned, and evaluated for a supervised regression task. The purpose is to figure out a model that holds the most promising further development for instance hyperparameter tuning. We also compare the models using the mean absolute error.

Scaling Features: It is important because features are in various units and we need to regularize the features so that the units do not influence the algorithm. RNN, AdaBoost, decision tree, random forest, Naïve Bayes do not require feature scaling

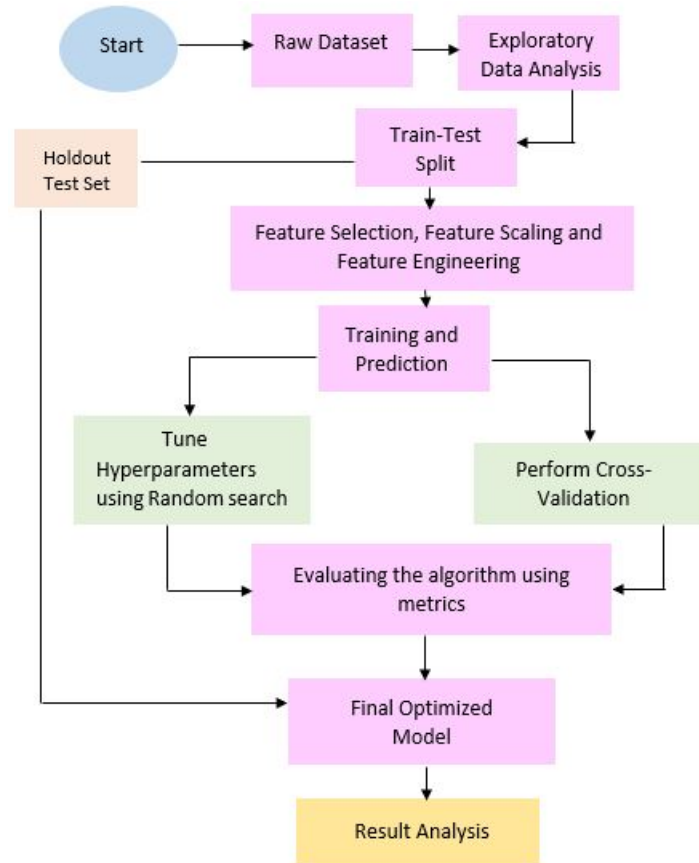


Figure 5.1: Work-Flow Diagram

but on the other hand methods like k-nearest neighbors and support vector machines do require it as they take the Euclidean distance between observations into consideration. As a result, it is best to scale features when we compare multiple algorithms. [22]

The two ways to scale features are given below:

- For each value, we divide by the standard deviation of the feature after subtracting the mean of the feature. This process is known as standardization and it results in a mean of 0 and standard deviation of 1 in each feature.
- For each value, we divide the maximum minus minimum for the feature (the range) after subtracting the minimum value of the feature. This guarantees that all the values for a feature are between 0 and 1 and is called scaling to a range or normalization.

The test and training sets were both scaled and normalized. 8 different machine learning models were trained and evaluated using the great Scikit-Learn library.

1. K-Nearest Neighbor Regression
2. Decision tree Regression
3. Random Forest Regression
4. Support Vector Machine Regression
5. Recurrent Neural Network Regression
6. Naïve Bayes Regression
7. Ada-Boost Regression
8. Stochastic Gradient Descent

The default models were trained on the training set and their mean absolute error on the test set was calculated.

5.3 Model Optimization:

Enhancing a model in machine language means figuring the best set of hyperparameters to solve a specific problem. The difference between model parameters and model hyperparameters are given below [24]

- Model parameters are configuration variables whose values are learned by the model during training. For example, the support vectors in a support vectors machine.
- Model hyperparameter is regarded as the best setting for the algorithms of machine learning that are tuned before training by the data scientists. Such as the number of neighbors used in the K Nearest Neighbors Regression. Also, the number of trees in the Random Forest can be an example of model hyperparameter. [35]

5.4 Hyperparameter Tuning :

Different limits, weights, and rates for different information patterns will be used to generalize the same type of machine learning model. These steps have to be tuned so that the algorithm can best solve the machine learning problem. The tests are called hyper-parameters. The optimization of hyperparameters considers a multitude of hyperparameters to have the optimum model that reduces the predefined loss function of data independently. The objective function takes some hyperparameters and returns the corresponding loss. We can determine the optimum hyperparameter for a model via random search and cross-validation. Random search establishes a range of values for hyper play and selects random mixes for model and score training. The search number is determined by time or money. It's amazingly strong. In a small fraction of the computer time, a random search across the same domain finds templates that are as good or better [18]. In far fewer iterations, it typically finds a close-sufficient value. The research space is best explored and therefore a strong mix of hyperparameters can normally be found in far less iterated areas. It should possibly be the first method of optimizing hyperparameters tested for their performance. Cross-validation is a re-examination technique used to approximate the capabilities of machine learning models. In applied machine learning, it is widely used in comparing and selecting a model for a particular problem in predictive modeling because it is easy to interpret, simple to apply, and results in ability estimations that usually have lower biases than other approaches. To test each hyperparameter value combination, a validation collection must be used to rank them. Test data cannot be used to adjust these hyperparameters [9] . Only once will we analyze the final model, we use the training results. Cross-validation is mostly used to evaluate a machine learning model's ability on invisible data. The process has a single parameter called k which corresponds to the number of groups in which the data sample is divided. Also, k-fold cross-validations are named. This should give us a greater approximation of the general results on the test set than a single validation set to preserve training details. Each discovery in the data collection is allocated to a certain group and stays inside the group for the rest of the operation. That

means any sample can be used for the set time and for training the model in the set time. This implies that each sample is used in the hold-out collection once and only used to train the model $k-1$ times. This approach is easy to grasp and is usually less flexible or positive than other approaches, e.g. a simple train or a test break. The random search hyperparameter optimization technique can be programmed and used for regression in the same way as it can be used for classification. For regression operations, Random Forests can be used. It is most suitable for model optimization. Random Forest Regression is an algorithm of controlled learning that employs the group's process of regression learning. Random forest works by building multiple decision trees and by predicting all trees during the training period and outputting the average groups. RF uses highly resistant overfitting decision trees [18]. RF chooses to generate a vast number of them depending on sacking to increase their precision. The basic idea is to repeatedly examine the data and train a new classifier for each dataset. Different classifiers overfit the data differently, and the discrepancies are measured by voting. It works the opposite way to the error-reducing mission. It takes the reverse approach to error reduction by lowering variance. The trees are not associated to optimize variance reduction, but the algorithm does not minimize distortion. The RF algorithm contains additional randomness in the quest for the best function when the tree splits. This results in high volatility and low bias, which is usually a stronger formula. Our mean absolute error of Random forest regression is 0.543

Chapter 6

Experimental Results and Analysis

6.1 Comparative Analysis of Supervised Models

After fitting our data in the existing models described in the literature review,

- Decision Tree Regression
- Naive Bayes Regression
- Random Forest Regression
- Recurrent Neural Network Regression
- Support Vector Machine Regression
- AdaBoost Regression
- Stochastic gradient descent Regression
- k-Nearest Neighbor Regression,

we get the following accuracy and mean absolute error.

Model	Mean Absolute Error	Accuracy
Stochastic gradient descent	0.010	99.78%
Random Forest Regression	0.543	99.75%
Decision Tree Regression	0.090	97.74%
K-Nearest Neighbor Regression	0.043	96.33%
AdaBoost Regression	0.079	94.49%
Naive Bayes Regression	0.841	94.91%
Recurrent Neural Network Regression	0.497	93.33%
Support Vector Machine Regression	0.564	92.79%

Table 6.1: Comparison of models

6.2 Comparison on Training Data

We have used 80:20 split on our data. For that, we have achieved the following Loss and Accuracy.

In the Recurrent Neural Network, with the increase of every epoch, the loss of the training data is decreasing. In graph 6.1, we see that when the epoch is 0.0, the loss is more than 0.13 and when the epoch is 17.5, the loss is below 0.10. So, it is confirmed that with the increase of epoch, the loss of the training data decreases.

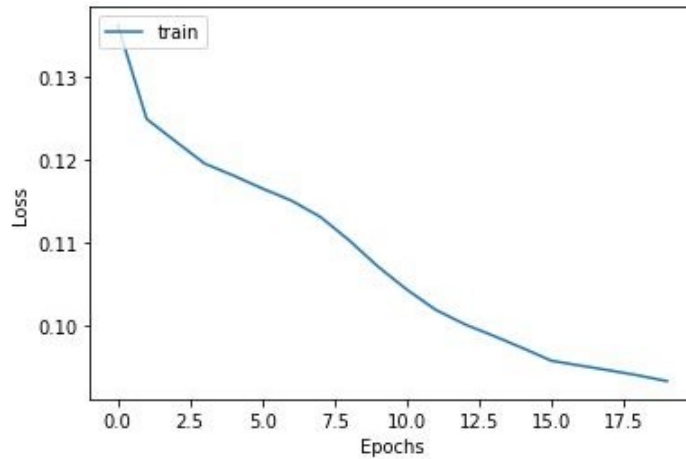


Figure 6.1: Model Loss

Also, in the Recurrent Neural Network, with the increase of every epoch, the accuracy of the model increases. In graph 6.2, when the epoch is 0.0, the accuracy is below 0.89 and when the epoch is 17.5, the accuracy of the model is 0.93

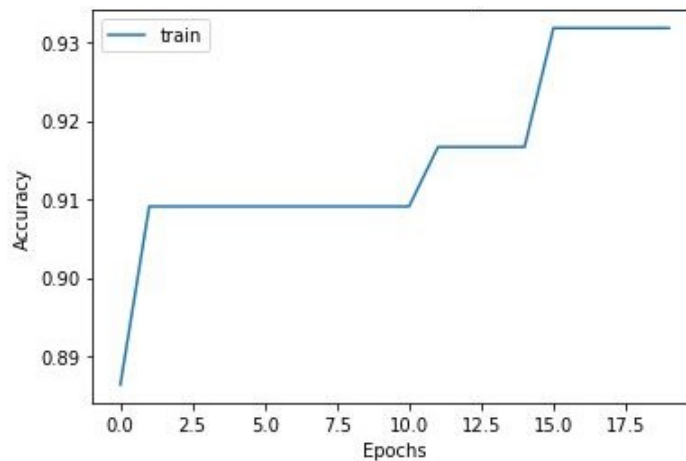


Figure 6.2: Model Accuracy

6.3 Graphs

From the below graph, we can see that we got a different mean absolute error for different models. Mean absolute error is very important for the quality of models. The model accuracy graph shows the accuracy of the models we got after we trained our data set. Stochastic gradient descent regression having the highest accuracy which is 99.78% and the Support Vector Machine Regression has the lowest accuracy is 92.79%. All the eight models that we have trained have accuracy above 90%

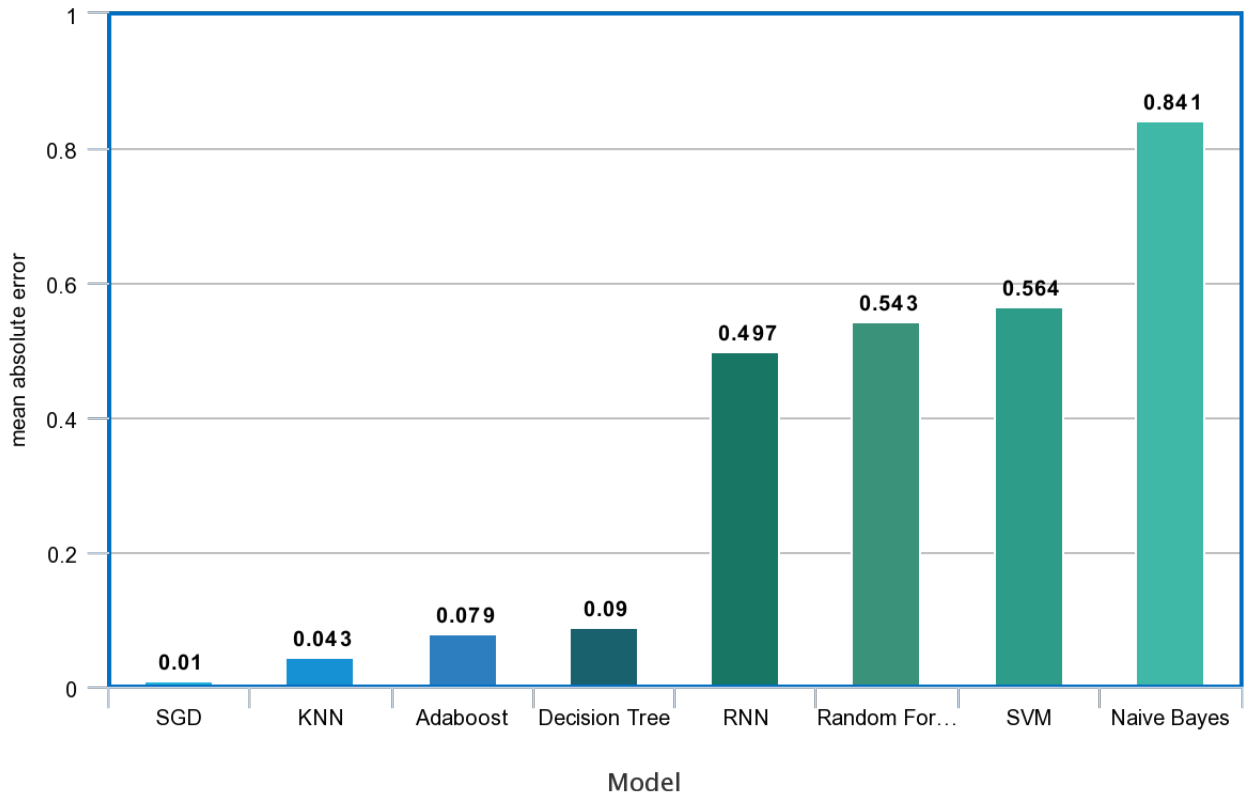


Figure 6.3: Graph Chart Mae

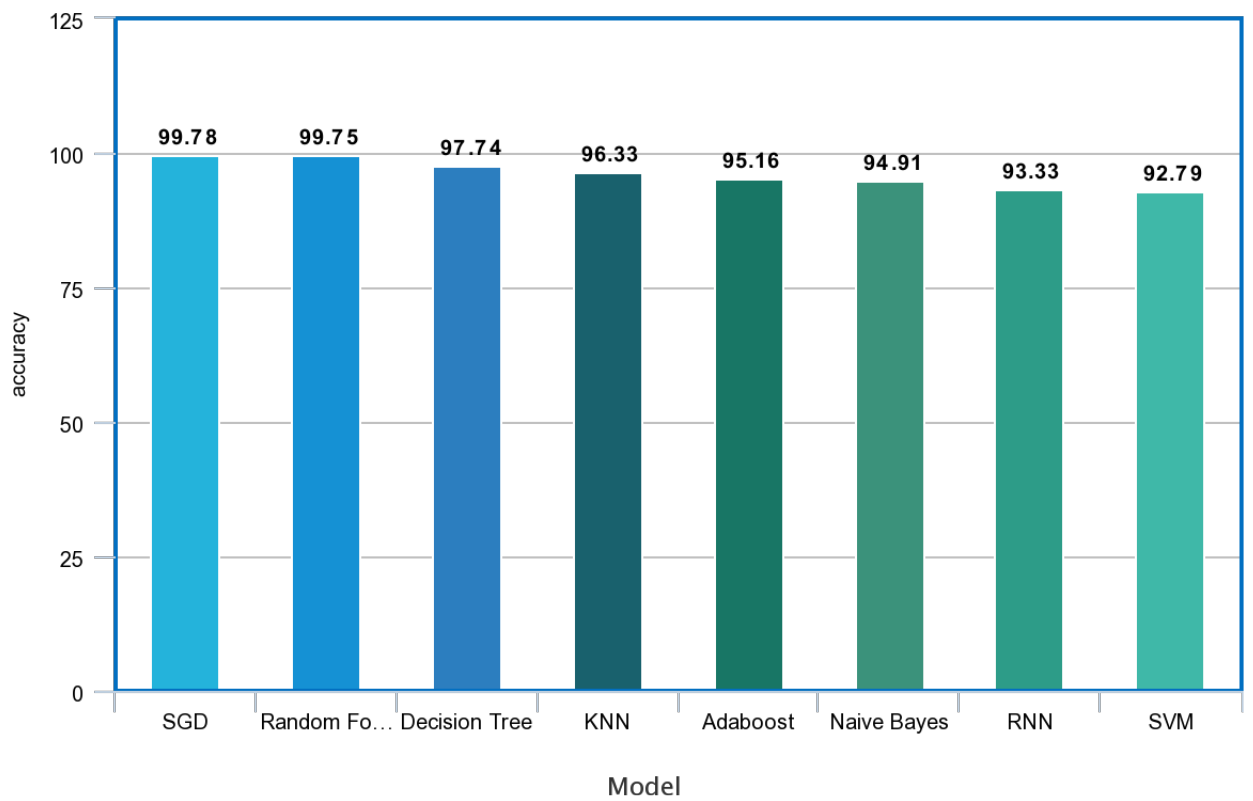


Figure 6.4: Graph Chart Accuracy

Chapter 7

Conclusion and Future work

7.1 Conclusion

At present cancer has become one of the main reasons for clinical demise. If we can predict it early, then it can be the string among life and demise. In this modern age, science has diminished us from many compromising sicknesses yet at the same time now it couldn't overcome cancer. Especially in a poor country like Bangladesh, cervical cancer is a threat for women with roughly 12,000 new cases identified each year, and more than 6,000 passing because of the seriousness of the illness [41]. Given the huge weight and high dismalness and mortality experienced by women with the cervical disease in this setting, directed research is fundamentally required, especially execution science examination to educate practical and manageable systems to expand the number of women who came to with administrations. Ideal analysis and legitimate therapy can fix cervical malignancy [21]. On the off chance that we can recognize cervical cancer growth at the beginning phase and take legitimate medicines in like manner, there is a high potential for recuperating. Accordingly by applying various calculations along with AI approaches we can make it attainable to right on time anticipate and recognize the side effects of cervical malignant growth in ladies body also, act likewise.

7.2 Future Work

From our above analysis, we got a great extent of accuracy from our hybrid model and our future goal is to extend the research and use it for a wide range of malignancies. Furthermore, this type of research and analysis can be used for other dangerous sicknesses as well. Further, the objective is to afford the patient the chance to have a healthy life as well as allowing cancer-affected people for a cure in the future.

Bibliography

- [1] J. R. Quinlan, “Induction of decision trees,” *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [2] S. R. Gunn *et al.*, “Support vector machines for classification and regression,” *ISIS technical report*, vol. 14, no. 1, pp. 5–16, 1998.
- [3] M. Zhenkui, C. Yuzhen, Y. Wei, and L. Rui, “An early stage cervix cancer diagnosis system,” in *SICE’99. Proceedings of the 38th SICE Annual Conference. International Session Papers (IEEE Cat. No. 99TH8456)*, IEEE, 1999, pp. 1059–1065.
- [4] B. C. Love, “Comparing supervised and unsupervised category learning,” *Psychonomic bulletin & review*, vol. 9, no. 4, pp. 829–835, 2002.
- [5] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “Knn model-based approach in classification,” in *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, Springer, 2003, pp. 986–996.
- [6] V. Jakkula, “Tutorial on support vector machine (svm),” *School of EECS, Washington State University*, vol. 37, 2006.
- [7] Z. Zhao and H. Liu, “Spectral feature selection for supervised and unsupervised learning,” in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 1151–1157.
- [8] J. Thongkam, G. Xu, and Y. Zhang, “Adaboost algorithm with random forests for predicting breast cancer survivability,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, IEEE, 2008, pp. 3062–3069.
- [9] T. Fushiki, “Estimation of prediction error by using k-fold cross-validation,” *Statistics and Computing*, vol. 21, no. 2, pp. 137–146, 2011.
- [10] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?” In *International workshop on machine learning and data mining in pattern recognition*, Springer, 2012, pp. 154–168.
- [11] Y. Xu, K. Hong, J. Tsujii, and E. I.-C. Chang, “Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries,” *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 824–832, 2012.
- [12] J. A. Bullinaria, “Recurrent neural networks,” *Neural Computation: Lecture*, vol. 12, 2013.
- [13] C.-C. Chang, S.-L. Cheng, C.-J. Lu, and K.-H. Liao, “Prediction of recurrence in patients with cervical cancer using mars and classification,” *International Journal of Machine Learning and Computing*, vol. 3, no. 1, p. 75, 2013.

- [14] J. Guo, “Backpropagation through time,” *Unpubl. ms., Harbin Institute of Technology*, vol. 40, pp. 1–6, 2013.
- [15] R. E. Schapire, “Explaining adaboost,” in *Empirical inference*, Springer, 2013, pp. 37–52.
- [16] P. Ramachandran, N. Girija, and T. Bhuvaneshwari, “Early detection and prevention of cancer using data mining techniques,” *International Journal of Computer Applications*, vol. 97, no. 13, 2014.
- [17] C.-J. Tseng, C.-J. Lu, C.-C. Chang, and G.-D. Chen, “Application of machine learning to predict the recurrence-proneness for cervical cancer,” *Neural Computing and Applications*, vol. 24, no. 6, pp. 1311–1316, 2014.
- [18] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, “Hyperopt: A python library for model selection and hyperparameter optimization,” *Computational Science & Discovery*, vol. 8, no. 1, p. 014 008, 2015.
- [19] S. Kaaviya, V. Saranyadevi, and M. Nirmala, “Pap smear image analysis for cervical cancer detection,” in *2015 IEEE International conference on engineering and technology (ICETECH)*, IEEE, 2015, pp. 1–4.
- [20] D. Li, B. Zhang, and C. Li, “A feature-scaling-based k -nearest neighbor algorithm for indoor positioning systems,” *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 590–597, 2015.
- [21] F. S. Papri, Z. Khanam, F. Islam, and M. M. Hakim, “Knowledge and awareness about risk factors of cervical cancer, its screening and vaccination among the women attending chittagong medical college hospital,” *Chattagram Maa-O-Shishu Hospital Medical College Journal*, vol. 14, no. 2, pp. 57–60, 2015.
- [22] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, “Investigation and improvement of multi-layer perceptron neural networks for credit scoring,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3508–3516, 2015.
- [23] M. R. Anderson and M. Cafarella, “Input selection for fast feature engineering,” in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, IEEE, 2016, pp. 577–588.
- [24] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, 2. MIT press Cambridge, 2016, vol. 1.
- [25] D. Kashyap, A. Somani, J. Shekhar, A. Bhan, M. K. Dutta, R. Burget, and K. Riha, “Cervical cancer detection and classification using independent level sets and multi svms,” in *2016 39th international conference on telecommunications and signal processing (TSP)*, IEEE, 2016, pp. 523–528.
- [26] K. Sneha, C. Arunvinodh, *et al.*, “Cervical cancer detection and classification using texture analysis,” *Biomedical and Pharmacology Journal*, vol. 9, no. 2, pp. 663–671, 2016.
- [27] R. Vidya and G. Nasira, “Prediction of cervical cancer using hybrid induction technique: A solution for human hereditary disease patterns,” *Indian Journal of Science and Technology*, vol. 9, no. 30, pp. 1–10, 2016.
- [28] N. Ketkar, “Recurrent neural networks,” in *Deep Learning with Python*, Springer, 2017, pp. 79–96.

- [29] Z. Li, X. Ma, and H. Xin, “Feature engineering of machine-learning chemisorption models for catalyst design,” *Catalysis today*, vol. 280, pp. 232–238, 2017.
- [30] A. Mallya, *Some rnn variants*, 2017.
- [31] M. Ontivero-Ortega, A. Lage-Castellanos, G. Valente, R. Goebel, and M. Valdes-Sosa, “Fast gaussian naive bayes for searchlight classification analysis,” *Neuroimage*, vol. 163, pp. 471–479, 2017.
- [32] D. N. Punjani and K. H. Atkotiya, “Cervical cancer prediction using data mining,” *Int. J. for Research in Applied Science & Eng. Tech.(IJRASET)*, vol. 5, no. 12, p. 1856, 2017.
- [33] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient knn classification with different numbers of nearest neighbors,” *IEEE transactions on neural networks and learning systems*, vol. 29, no. 5, pp. 1774–1785, 2017.
- [34] B. Benazir and A. Nagarajan, “An expert system for predicting the cervical cancer using data mining techniques,” *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 1971–87, 2018.
- [35] H. Ma, X. Yang, J. Mao, and H. Zheng, “The energy efficiency prediction method based on gradient boosting regression tree,” in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, IEEE, 2018, pp. 1–9.
- [36] S. Pathak, I. Mishra, and A. Swetapadma, “An assessment of decision tree based classification and regression algorithms,” in *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, IEEE, 2018, pp. 92–95.
- [37] A. Shetty and V. Shah, “Survey of cervical cancer prediction using machine learning: A comparative approach,” in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2018, pp. 1–6.
- [38] A. Arfiani and Z. Rustam, “Ovarian cancer data classification using bagging and random forest,” in *AIP Conference Proceedings*, AIP Publishing LLC, vol. 2168, 2019, p. 020 046.
- [39] C. Ifemelumma, C. Anikwe, B. Okorochukwu, F. Onu, J. Obuna, B. Ejikeme, and O. Ezeonu, “Cervical cancer screening: Assessment of perception and utilization of services among health workers in low resource setting,” *International journal of reproductive medicine*, vol. 2019, 2019.
- [40] H. Kamel, D. Abdulah, and J. M. Al-Tuwaijari, “Cancer classification using gaussian naive bayes algorithm,” in *2019 International Engineering Conference (IEC)*, IEEE, 2019, pp. 165–170.
- [41] M. R. Hoque, E. Haque, and M. R. Karim, “Cervical cancer in low-income countries: A bangladeshi perspective,” *International Journal of Gynecology & Obstetrics*, vol. 152, no. 1, pp. 19–25, 2021.